

Transcriptomics of the Bed Bug (*Cimex lectularius*)

Xiaodong Bai¹*, Praveen Mamidala¹*, Swapna P. Rajarapu¹, Susan C. Jones², Omprakash Mittapalli^{1*}

1 Department of Entomology, Ohio Agricultural and Research Development Center, The Ohio State University, Wooster, Ohio, United States of America, **2** Department of Entomology, The Ohio State University, Columbus, Ohio, United States of America

Abstract

Background: Bed bugs (*Cimex lectularius*) are blood-feeding insects poised to become one of the major pests in households throughout the United States. Resistance of *C. lectularius* to insecticides/pesticides is one factor thought to be involved in its sudden resurgence. Despite its high-impact status, scant knowledge exists at the genomic level for *C. lectularius*. Hence, we subjected the *C. lectularius* transcriptome to 454 pyrosequencing in order to identify potential genes involved in pesticide resistance.

Methodology and Principal Findings: Using 454 pyrosequencing, we obtained a total of 216,419 reads with 79,596,412 bp, which were assembled into 35,646 expressed sequence tags (3902 contigs and 31744 singletons). Nearly 85.9% of the *C. lectularius* sequences showed similarity to insect sequences, but 44.8% of the deduced proteins of *C. lectularius* did not show similarity with sequences in the GenBank non-redundant database. KEGG analysis revealed putative members of several detoxification pathways involved in pesticide resistance. Lamprin domains, Protein Kinase domains, Protein Tyrosine Kinase domains and cytochrome P450 domains were among the top Pfam domains predicted for the *C. lectularius* sequences. An initial assessment of putative defense genes, including a cytochrome P450 and a glutathione-S-transferase (GST), revealed high transcript levels for the cytochrome P450 (CYP9) in pesticide-exposed versus pesticide-susceptible *C. lectularius* populations. A significant number of single nucleotide polymorphisms (296) and microsatellite loci (370) were predicted in the *C. lectularius* sequences. Furthermore, 59 putative sequences of *Wolbachia* were retrieved from the database.

Conclusions: To our knowledge this is the first study to elucidate the genetic makeup of *C. lectularius*. This pyrosequencing effort provides clues to the identification of potential detoxification genes involved in pesticide resistance of *C. lectularius* and lays the foundation for future functional genomics studies.

Citation: Bai X, Mamidala P, Rajarapu SP, Jones SC, Mittapalli O (2011) Transcriptomics of the Bed Bug (*Cimex lectularius*). PLoS ONE 6(1): e16336. doi:10.1371/journal.pone.0016336

Editor: Frederic Marion-Poll, AgroParisTech, France

Received: October 11, 2010; **Accepted:** December 10, 2010; **Published:** January 19, 2011

Copyright: © 2011 Bai 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: This research was supported 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

These authors contributed equally to this work.

Introduction

"Good night, sleep tight, don't let the bed bugs bite!" This common nighttime verse now has become a precautionary catch phrase around the globe. Bed bugs (*Cimex lectularius* L.) are flightless, nocturnal, obligate blood-feeding ectoparasites that preferentially feed on humans. Bed bug infestations pose grave economic concerns and quality-of-life issues for households [1]. The resurgence of bed bugs poses an urgent situation as infestations are rampant globally, nationally, and locally. The control of these medicinally important insect pests in urban environments costs billions of dollars annually and typically requires the use of large quantities of pesticides/insecticides.

Individuals that are allergic to *C. lectularius* bites often experience itching and erythematous or papular urticaria-like dermatitis, which favors secondary infections like impetigo, ecthyma and lymphangitis [2–7]. *C. lectularius* infestations also result in anxiety, insomnia or worsening of an existing mental health condition [7–9]. However, the risk of transmission of human disease by *C. lectularius* is still not clear [10]. These

ectoparasites are an important public health issue affecting all socioeconomic classes.

The association of *C. lectularius* and humans dates back to 1350 B.C. or earlier, as evidenced by well-preserved bed bug remains recovered from the Workmen's Village at el-Amarna, Egypt [11]. Bed bugs are not native to North America but rather were introduced by the early colonists in the 17th century. *C. lectularius* were extremely common pests in the United States prior to World War II, however extensive use of dichloro-diphenyl-trichloroethane (DDT) and other long-lasting residual insecticides greatly reduced their numbers [12].

During the past decade or so, the resurgence of *C. lectularius* has been recorded across the globe including North America, Europe, Australia, and Eastern Asia with an estimated 100–500% annual increase in bed bug populations [13–17]. Survey by the National Pest Management Association and the U.S. Environmental Protection Agency (EPA) indicated that *C. lectularius* stress calls increased 81% during the last decade; the majority of bed bug complaints came from occupants of multi-unit apartment complexes. Furthermore, 76% of pest management companies

confirmed that *C. lectularius* were the most difficult pest to control (www.pestworld.org). Several hypotheses have been proposed to explain the sudden resurgence of *C. lectularius* worldwide which include, but are not limited to, frequent international travel (to/from areas where *C. lectularius* remained common), increased exchange of used furniture, a shift from usage of broad-spectrum insecticides to more specific/selective control tactics such as baits for other urban pests, and insecticide resistance within the insect [18–25].

Resistance to pyrethroids (e.g., deltamethrin and lambda-cyhalothrin) appears to be widespread within U.S. populations of *C. lectularius* [23]. Pesticide resistance in *C. lectularius* is purported to result from point mutations in the open reading frames of voltage-sensitive sodium channel genes compared to pesticide susceptible populations [26]. However, the role of cytochrome P450s and glutathione S-transferases (GSTs) has yet to be established in pesticide resistance of *C. lectularius*. In many insects, both cytochrome P450s and GSTs have been shown to metabolize synthetic chemicals (insecticides/pesticides) and host plant allelochemicals [27–31]. The cytochrome P450 and GST detoxification systems catalyze physiological reactions that modify toxic compounds into water-soluble, non-toxic compounds that are excreted by insects.

Despite the high-impact status of *C. lectularius*, very little is known about this blood-feeding insect at the molecular level. The next generation sequencing methods (Roche 454, Solexa/Illumina, etc.) provide a unique opportunity for genomic exploration in non-model insect species wherein little or no

molecular knowledge is available [32]. In particular, 454-sequencing technology based on the pyrosequencing principle has recently enabled the application of functional genomics to a broad range of insect species including *Melitaea cinxia* [33], *Zygaena filipendulae* [34], *Chrysomela tremulae* [35], *Aphis glycines* [36], *Manduca sexta* [37,38], *Laodelphax striatellus* [39], *Stomoxys calcitrans* [40], *Dermacentor variabilis* [41], *Erynnis propertius* and *Papilio zelicaon* [42], and *Agilus planipennis* [43]. In the current study we applied 454 technology to build a sufficiently large expressed sequence tag (EST) database for *C. lectularius*. Our results will allow for a better understanding of the physiology-driven molecular processes in *C. lectularius* and the identification of candidate genes potentially involved in insecticide resistance.

Results and Discussion

Transcriptomic analysis

Roche 454 pyrosequencing of adult *C. lectularius* yielded a total of 216,419 transcriptomic reads with 79,596,412 bp, which were assembled into 35,646 ESTs (3,902 contigs and 31,744 singletons) (Figure 1) using the Roche Newbler program. The length of the contigs varied from 60–4,615 bp with an average contig length of 759 bp and totaling 2,962,366 bp. The singletons ranged from 50–863 bp with an average length of 313 bp and totaling 9,919,703 bp. From the current *C. lectularius* transcriptomic database, 29.6% transcripts showed significant similarity (E value $<1e^{-5}$) to proteins in the GenBank nr database. As expected, the majority of the sequences (85.9%) were matched to insect proteins

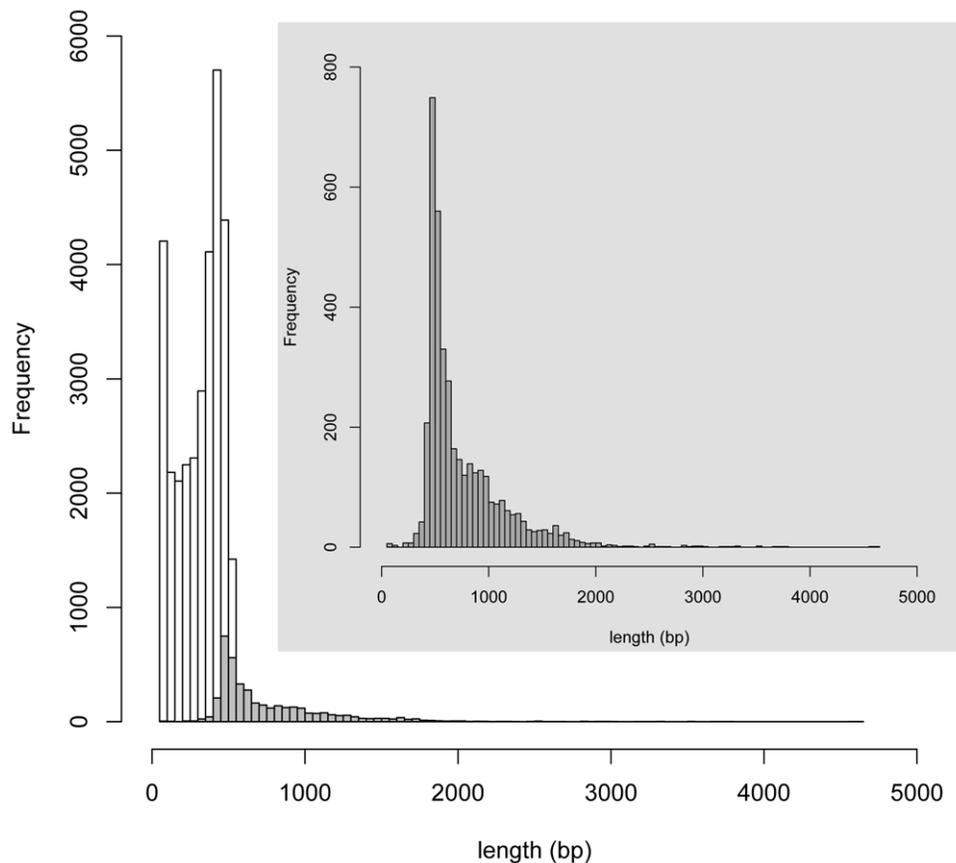


Figure 1. Summary of *Cimex lectularius* transcriptomic sequences. The contig sequences are represented by shaded bars and the singleton sequences by clear bars.

doi:10.1371/journal.pone.0016336.g001

and the remaining were matched to non-insect eukaryotes (11.16%), fungi (1.78%), bacteria (1.21%), viruses (0.04%), Archaea (0.02% sequences) and artificial sequences (0.03% sequences) (Figure 2).

Comparative analysis

The comparison of *C. lectularius* transcriptomic sequences to the draft protein sequences of three insect species [44,45] revealed that the majority of sequences (46.1%, 16,367 of 35,505) were similar to *Pediculus humanus* (body louse) followed by *Acyrtosiphon pisum* (pea aphid), (45%) and *Drosophila melanogaster* (fruit fly) (23.6%) (Figure 3). High sequence similarity of *C. lectularius* with *P. humanus* might be due to their similar diet, i.e., blood. A significant percentage of transcripts (44.8%) were found to be unique to *C. lectularius* and perhaps could be attributed to the presence of novel genes. Alternatively, the derived transcripts may be from the cDNA of untranslated regions, chimerical sequences (assemblage errors) and non-conserved areas of proteins where homology is not detected, which is in agreement with several other transcriptomic studies [43,46,47].

Gene Ontology assignments

In total 8,363 transcripts of *C. lectularius* were assigned for Gene Ontology (GO) terms based on BLAST matches with sequences whose function is previously known (Figure 4, Table S1). These transcripts were assigned for biological process (7,066 sequences, Figure 4a), cellular component (5,549 sequences, Figure 4b) and molecular function (6,290 sequences, Figure 4c). Among the molecular function assignments, a high percentage of genes were assigned for Binding (49.1%), predominantly heat shock proteins (Hsp). In a recent study of *C. lectularius*, the transcript levels for Hsp70 and Hsp 90 were observed to be elevated when bugs were subjected to various stress factors (heat, cold and dehydration) suggesting that these proteins may play an important role during environmental stress and could potentially play a role in control strategies [1,8,13,48]. The cellular component terms showed a significant percentage of genes assigned to cell part (53%) whereas the biological process terms were associated predominantly with cellular processes (32%) such as proteolysis, carbohydrate metabolic processes and oxidation reduction utilization. Similar observations for metabolic processes were reported in transcriptomic studies of other insects [38,43,49].

KEGG analysis

The KEGG metabolic pathways presented in the current EST database of *C. lectularius* were Nucleotide Metabolism (569 transcripts), Protein Metabolism (560), Lipid Metabolism (346), Alkaloid Metabolism (329), Carbohydrate Metabolism (295), Detoxification by cytochrome P450 (91), and Vitamin Metabolism (82) (Table S2). Taken together, the putative KEGG pathways identified in the current study shed light on specific responses and functions involved in the molecular processes of *C. lectularius*.

Protein Domains

A total of 6,752 protein domains were identified in 6,286 *C. lectularius* transcripts using HMMER3 software (Table S3). Among these domains, lamprin proteins were the highest with a total of 223 (Table 1). Lamprin proteins are a unique family of hydrophobic self-aggregating proteins consisting of GGLGY tandem pentapeptide repeats reported in lamprey cartilage proteins, mammalian and avian elastins, and various insects (silk moth chorion protein and spider dragline silk) [50–53]. Protein kinase (82) and protein tyrosine kinase (55) were among the other top Pfam domains in our study. Both proteins are involved in signal transduction pathways, development, cell division and metabolism in higher organisms [54,55]. Approximately 60 cytochrome P450 domains were predicted in the derived transcriptomic sequences of *C. lectularius*. Insect cytochrome P450s are reported in the metabolism of xenobiotics, wherein induced levels are correlated with resistance to synthetic insecticides and plant allelochemicals [56,57].

In total, 58 RNA recognition motifs (RRMs) were predicted in the *C. lectularius* sequences. These domains are also referred to as RNA-binding domain (RBD), consensus sequence RNA-binding domain (CS-RBD), ribonucleoprotein domain (RNPD), and RNP consensus sequence (RNP-CS). These proteins are involved in pre-mRNA processing and transport, regulation of stability and translational control [58,59]. RRM are reported to be involved in male courtship and vision in *D. melanogaster* [61,62]. Mutations in *D. melanogaster* RRM resulted in reduced viability, female sterility with abnormal wing and mechanosensory bristle morphology [58].

From the *C. lectularius* database we predicted 54 protein domains belonging to the Ras family, which are thought to be involved in insect development especially in cell differentiation and proliferation [62]. A high number of WD domains were identified in this

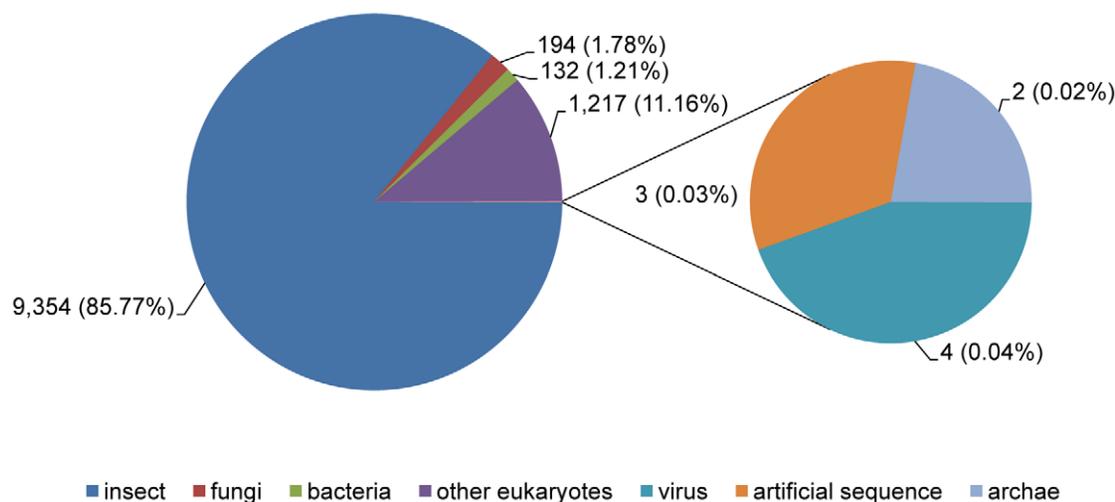


Figure 2. A pie chart showing species distribution of the top BLAST hits of the *Cimex lectularius* sequences to various insect species. doi:10.1371/journal.pone.0016336.g002

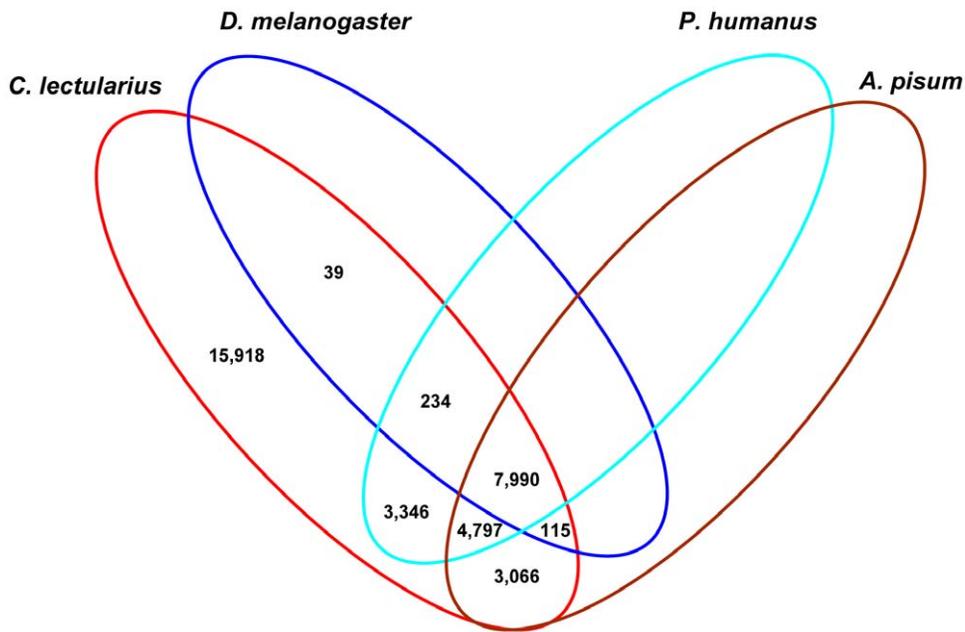


Figure 3. Comparative summary of *Cimex lectularius* transcriptomic sequences with the protein sequences of *Drosophila melanogaster*, *Pediculus humanus* and *Aphis pisum*.
doi:10.1371/journal.pone.0016336.g003

study (Table 1), which are primarily involved in protein-protein interactions [36]. Sugar transporters (52) that are associated with transport of nutrients, and domains of mitochondrial carrier proteins (47), which are primarily involved in transport of metabolite intermediates, were also among the top ten domains predicted in the *C. lectularius* sequences [63,64]. The later proteins are recognized by their unique signature motif P-X-[D/E]-X-X-[R/K] and the presence of six helical transmembrane segments made up of three tandem repeated sequences [65].

We predicted 45 insect cuticle proteins in the derived *C. lectularius* sequences. Insect cuticle is a complex structure consisting of chitin embedded in a protein matrix that lacks cysteine residues but is characterized with conserved R&R domain (G-x(7)-[DEN]-G-x(6)-[FY]-x-A-[DGN]-x(2,3)-G-[FY]-x-[AP] [66–68]. The R&R consensus is further classified into RR1 (soft cuticle), RR2 (hard cuticle) and extended R&R consensus chitin binding domain [66]. The other highly abundant domains identified in the present study include Miro-like protein (39), Major Facilitator Superfamily (38), ADP-ribosylation factor family (35), Immunoglobulin I-set domain (34) and TCP_1/cpn60 chaperonin family (33). Interestingly, we didn't find PAZ and PIWI domains, which are believed to be important components of the RNA induced silencing complex. The lack of these domains in our current database of *C. lectularius* could be attributed to insufficient coverage of the transcriptome.

Genes of Interest

We have mined the current transcriptomic database to obtain genes putatively involved in insecticidal resistance of *C. lectularius* (Table 2). Given that one of the factors responsible for *C. lectularius* resurgence is purported to be pyrethroid resistance during the last decade, we are specifically interested in genes that participate in generalized insect defense. Metabolic resistance in insects has been attributed to induced levels of cytochrome P450 monooxygenases (CYPs), glutathione *S*-transferases (GSTs), superoxide dismutases (SODs), catalases (CATs), glutathione peroxidases (GPXs), carboxyl choline esterases and ascorbate peroxidases [69,70]. Intriguingly,

the majority of the cytochrome P450s identified in the *C. lectularius* transcriptome database belonged to the CYP3 clade (includes CYP3, CYP6 and CYP9 members) compared to other CYP clades, which is in agreement with other insect systems [56].

Although pesticide resistance in *C. lectularius* is thought to be via point mutations in voltage-gated sodium channels [24,26], the role of the detoxification and antioxidant enzymes is poorly understood. Hence, from the current database, we profiled the transcript levels for a cytochrome P450 (CYP9) and a GST (Delta-epsilon) in different developmental stages (early-instar nymphs, late-instar nymphs and adults) of pesticide-susceptible and pesticide-exposed *C. lectularius* populations. Quantitative real-time PCR (qPCR) analysis of the CYP9 showed higher mRNA levels for all developmental stages in pesticide-exposed populations compared to pesticide-susceptible populations (Figure 5A). In particular, the highest transcript levels for CYP9 were observed in early instars of the pesticide-exposed population. Similar observations were reported in *Heliothis virescens* and *M. sexta* wherein CYP9A1 and CYP9A2 were over-expressed in response to insecticidal treatments [71,72]. More recently, CYP9M10 of *Culex quinquefasciatus* was shown to be involved in pyrethroid detoxification [31]. Based on these studies, the CYP9 profiled in *C. lectularius* could also be induced upon pesticide exposure; however further functional studies (gene expression and RNAi) are required to elucidate the role of CYP9 in *C. lectularius*.

GSTs are thought to be potential secondary detoxification agents and are majorly involved in DDT resistance [73]. Expression analysis of a candidate GST retrieved from the current EST database revealed highest mRNA levels in the late-instar nymphs from pesticide-exposed *C. lectularius* populations compared to those of the pesticide-susceptible populations (Figure 5B). However, there was no significant difference in GST activity for adults from the two populations, an observation previously reported for adult bed bugs [26].

Wolbachia. In the current transcriptome database of *C. lectularius*, we found 59 sequences showing similarity with *Wolbachia*

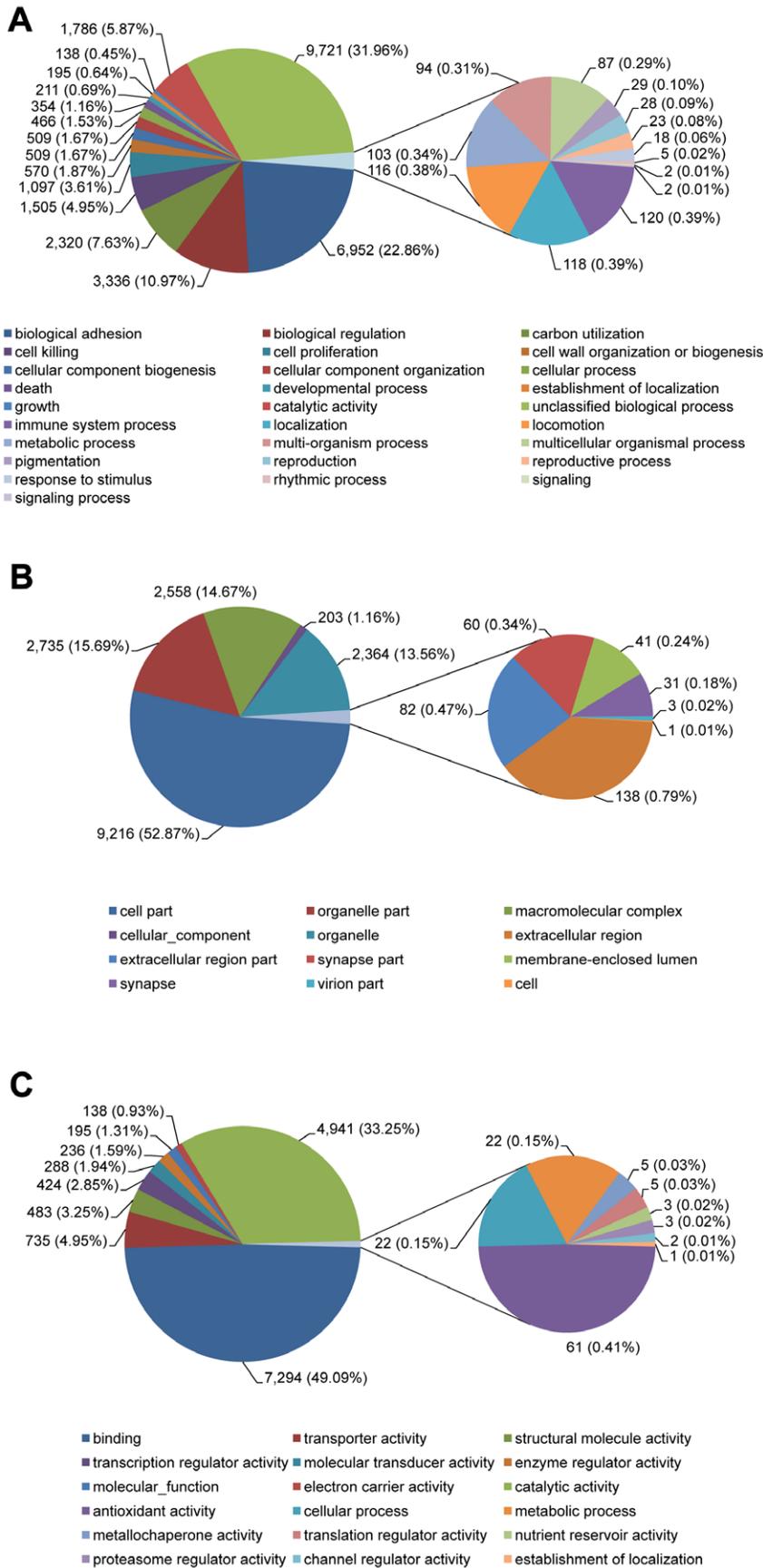


Figure 4. Gene ontology (GO) terms for the transcriptomic sequences of *Cimex lectularius*. (A) biological process, (B) cellular component and (C) molecular function.
doi:10.1371/journal.pone.0016336.g004

Table 1. Summary of top 15 domains predicted in *Cimex lectularius* sequences.

| Domain accession | Domain name | Domain description | Occurrence |
|------------------|---------------|--|------------|
| PF06403.4 | Lamprin | Lamprin | 223 |
| PF00069.18 | Pkinase | Protein kinase domain | 82 |
| PF00067.15 | p450 | Cytochrome P450 | 60 |
| PF00076.15 | RRM_1 | RNA recognition motif (a.k.a. RRM, RBD, or RNP domain) | 58 |
| PF07714.10 | Pkinase_Tyr | Protein tyrosine kinase | 55 |
| PF00071.15 | Ras | Ras family | 54 |
| PF00400.25 | WD40 | WD domain, G-beta repeat | 53 |
| PF00083.17 | Sugar_tr | Sugar (and other) transporter | 52 |
| PF00153.20 | Mito_carr | Mitochondrial carrier protein | 47 |
| PF00379.16 | Chitin_bind_4 | Insect cuticle protein | 45 |
| PF08477.6 | Miro | Miro-like protein | 39 |
| PF07690.9 | MFS_1 | Major facilitator superfamily | 38 |
| PF00025.14 | Arf | ADP-ribosylation factor family | 35 |
| PF07679.9 | I-set | Immunoglobulin I-set domain | 34 |
| PF00118.17 | Cpn60_TCP1 | TCP-1/cpn60 chaperonin family | 33 |

doi:10.1371/journal.pone.0016336.t001

(Table 2, Table S4). As an endosymbiont, *Wolbachia* is reported in nearly 70% of all insect species [74–76]. Besides their role in nutrition, these are thought to play an important role in

manipulating the host reproductive system through reproductive parasitism, i.e., feminization of genetic males, parthenogenesis and cytoplasmic incompatibilities, thereby increasing the frequency of

Table 2. Genes of interest in *Cimex lectularius*.

| Candidate genes | #Occurrence | Family members with corresponding numbers |
|-------------------------------------|-------------|--|
| Cytochrome P450 | | |
| <i>Cyp2 clade</i> | | |
| CYP18 | 01 | CYPXVIII1 |
| CYP307 | 03 | CYPCCCVIIA1 |
| CYP314 | 03 | CYPCCCXIVA1 |
| <i>CYP3 clade</i> | | |
| CYP3 | 05 | CYPPIIA1 (2), CYPPIIA12 (1), CYPPIIA13 (1), CYPPIIA31 (1) |
| CYP6 | 35 | CYPVIA2 (1), CYPVIA13 (5), CYPVIA14 (9), CYPVIA18 (2), CYPVIA21 (3), CYPVIB2 (1), CYPVIB4 (1), CYPVID2 (2), CYPVID4 (1), CYPVID5 (4), CYPVIJ1 (4), CYPVIK1 (2) |
| CYP9 | 05 | CYPIXE2 |
| <i>CYP4 clade</i> | | |
| CYP4 | 17 | CYPIVC1 (5), CYPIVF8 (1), CYPIVG1 (2), CYPIVG15 (9) |
| <i>Mitochondrial CYP clade</i> | | |
| CYP 301 | 04 | CYPCCCIA1 |
| Glutathione S-transferases | 14 | |
| Superoxide dismutases | 06 | |
| Catalases | 02 | |
| Peroxidases (GPX and PhGPX) | 03 | |
| Voltage gated sodium channel | 01 | |
| <i>Wolbachia</i> | 59 | |

doi:10.1371/journal.pone.0016336.t002

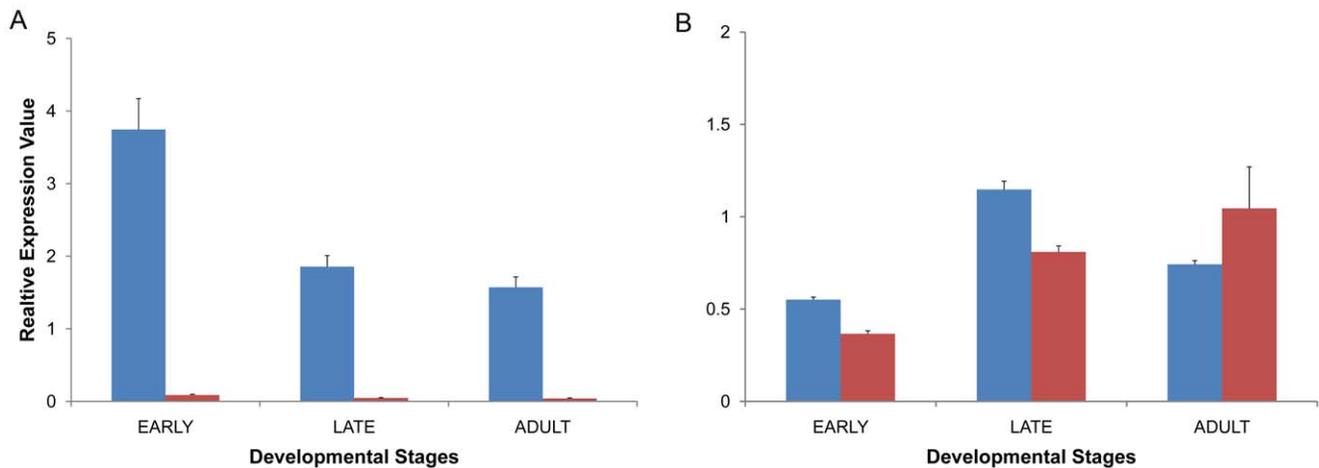


Figure 5. Quantitative PCR analysis of candidate defense genes in *Cimex lectularius* early- and late-instar nymphs and adults. (A) mRNA levels for a cytochrome P450 (CYP9) in pesticide-exposed (blue bars) and pesticide-susceptible populations (red bars). (B) mRNA levels for a glutathione-S-transferase (GST) in pesticide-exposed (blue bars) and pesticide-susceptible populations (red bars). An EF-1 α was used as the internal reference gene. Standard error of the mean for three technical replicates is represented by the error bars. doi:10.1371/journal.pone.0016336.g005

infected females in the host population [77–82]. In a recent study, *Wolbachia* was shown to be essential for *C. lectularius*' synthesis of B vitamins, which are deficient in blood meals. Antibiotic-supplemented blood meals for *C. lectularius* resulted in delayed adult emergence and egg deposition; however, normal adult emergence and egg development was restored when the blood meal containing antibiotic was supplemented with B vitamins [82].

Putative Molecular Markers

We predicted a total of 296 putative single nucleotide polymorphisms (SNPs) wherein 96 were transversions and 200 were transitions (Table 3, Table S5). Additionally, we identified 370 simple sequence repeats (SSRs or microsatellites), of which 69% were trinucleotide repeats, followed by 27% dinucleotide and 4% tetranucleotide repeats (Table 4, Table S6). Molecular markers identified in the current study could lay a platform for better understanding the adaptation/ecology of *C. lectularius* as reported in other insect systems [83]. However all the predicted molecular markers need to be validated to rule out false positives and sequencing errors.

Table 3. Predicted single nucleotide polymorphisms (SNP) in *Cimex lectularius* sequences.

| SNP type | Counts |
|---------------------|--------|
| Transition | 200 |
| A-G | 91 |
| C-T | 109 |
| Transversion | 96 |
| A-C | 21 |
| A-T | 31 |
| C-G | 16 |
| G-T | 28 |

doi:10.1371/journal.pone.0016336.t003

Conclusions

This study is the first to obtain fundamental molecular knowledge of *C. lectularius*. Some noteworthy results of this study are 1) a significant number of putative defense pathways were identified within the derived sequences; 2) a number of SNPs and microsatellite markers were predicted, which upon validation could facilitate the identification of polymorphisms within *C. lectularius* populations; and 3) high transcript levels for a cytochrome P450 (CYP9) in pesticide-exposed *C. lectularius* populations provide initial clues to metabolic resistance. These characteristic features along with the recovered sequences of *Wolbachia* provide new insights into the biology of *C. lectularius*.

Materials and Methods

Insect material

C. lectularius populations used in this study include the pesticide-susceptible strain “Harlan” which has been in laboratory culture since 1973 and hence has not experienced pesticide exposure for several decades. Pesticide-exposed bed bugs from Columbus, OH, were collected during 2009 and 2010 from an apartment that had undergone repeated insecticide treatments without successful bed bug control. These bed bug populations were reared in the laboratory as previously described [21]. Samples of the above-mentioned collections were transported to the Ohio Agricultural and Research Development Center (OARDC, Wooster, OH) and were categorized into different developmental stages as per Usinger [84].

RNA isolation, cDNA library construction and 454 sequencing

Total RNA was extracted using TRIzol[®] reagent (Invitrogen) from a total of 15 individual insects of various developmental stages (1st-instar nymph– adult) of the Harlan strain. Approximately 10 μ g of the extracted RNA was shipped to the Purdue Genomics Core Facility (West Lafayette, IN) for cDNA library construction and subsequent 454 sequencing. The cDNA library was constructed using the SMART cDNA library construction kit and following the manufacturer's protocol with a few modifications to enhance sequencing: i) a modified CDSIII/3' primer (5'-TAG AGG CCG

Table 4. Summary of microsatellite loci predicted in *Cimex lectularius* sequences.

| Number of repeats | Dinucleotide repeats | Trinucleotide repeats | Tetranucleotide repeats |
|-------------------|----------------------|-----------------------|-------------------------|
| 5 | - | 151 | 12 |
| 6 | - | 58 | 01 |
| 7 | - | 21 | 02 |
| 8 | 41 | 10 | - |
| 9 | 26 | 04 | - |
| 10 | 10 | 03 | - |
| 11 | 06 | 02 | - |
| 12 | 02 | 01 | - |
| 13 | 03 | - | - |
| 14 | - | 03 | - |
| 15 | 02 | 01 | - |
| 16 | 03 | 01 | - |
| 17 | 03 | - | - |
| 20 | 01 | - | - |
| 21 | 01 | - | - |
| 22 | 01 | - | - |
| 35 | 01 | - | - |
| Subtotal | 100 | 255 | 15 |

doi:10.1371/journal.pone.0016336.t004

AGG CGG CCG ACA TGT TTT GTT TTT TTT TCT TTT TTT TTT VN-3'; PAGE purified) and SuperScript II reverse transcriptase (Invitrogen, Carlsberg, CA) were used for first-strand cDNA synthesis, and ii) cDNA size fractionation was excluded and final products were cleaned and eluted using a QIAquick PCR purification kit (Qiagen, Valencia, CA). Following agarose gel electrophoresis and extraction of DNA from gels, DNA bands (500–800 bp) were purified, blunt ended followed by ligation with adapters and finally immobilized on beads. Single stranded DNA isolated from the beads was characterized for correct size using a LabChip 7500. The concentration and the proper ligation of the adapters were examined using qPCR. One-quarter of a pico-titer plate was sequenced following manufacturer's protocol using the Roche 454 GS FLX Titanium chemistry (Roche Diagnostics, Indianapolis, IN).

Bioinformatic analysis

The sequences were assembled using NEWBLER software package (a *de novo* sequence assembly software) after the removal of adapter sequences. For attaining better results, the contigs and singletons were renamed in the format of "BB454ONE000001" where "BB" stands for the bed bug species, "454" for 454 sequencing technology, "ONE" for the first trial, and "000001" for an arbitrarily assigned number. All the contigs and singletons of *C. lectularius* were analyzed using BLASTx algorithm [85] against GenBank non-redundant database at National Center for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov/>). Using BLASTx algorithm we also compared the sequences to the insect-specific protein sequences. To examine the protein domain all the sequences were searched against the Pfam database [86] by HMMER v3 program [87]. The Blast2GO software [88,89] was used to predict the functions of the sequences, assign Gene Ontology terms, and predict the metabolic pathways in Kyoto Encyclopedia of Genes and Genome [90–92]. SSRs were identified using Msafinder version 2.0.9 program [93] whereas SNPs were predicted using

gsMapper software (Roche Diagnostics) with an arbitrary criterion of at least 4 reads supporting the consensus or variant.

Gene mining and quantitative real time PCR

Total RNA was extracted from different development stages (early instars, late instars, and adults) using TRIzol[®] following the manufacturer's protocol. The total RNA obtained was re-suspended in 40 μ l of nuclease-free water and the concentration was measured using Nanodrop (Thermo Scientific Nanodrop 2000). About 0.5 μ g of total RNA was used as template to synthesize first-strand cDNA using Superscript II Reverse Transcriptase kit (Invitrogen) following the manufacturer's protocol. The resultant cDNA was diluted to 20 ng/ μ l for further use in qPCR. Genes of interest included CYP9 and a GST that were subjected to qPCR analysis. Primers were designed using Beacon Designer 7 software (primer sequences upon request). The cycling parameters were 95°C for 5 min followed by 40 cycles of 95°C for 10 s and 60°C for 30 s ending with a melting curve analysis (65°C to 95°C in increments of 0.5°C every 5 s) to check for nonspecific product amplification. Relative gene expression was analyzed by the $2^{-\Delta\Delta CT}$ method (User Bulletin #2: ABI Prism 7700 Sequence Detection System *vide supra* (<http://www3.appliedbiosystems.com/>)). An elongation factor 1-alpha (EF1- α) of *C. lectularius* was used as the internal reference gene, as has been used in other insect systems [94].

Data deposition

The Roche 454 reads of *C. lectularius* were submitted to NCBI Sequence Read Archive under the accession number of SRA024509.1.

Supporting Information

Table S1 Gene Ontology of *C. lectularius* sequences. (XLSX)

Table S2 KEGG summary of *C. lectularius* sequences. (XLSX)

Table S3 Pfam domain search of *C. lectularius* sequences. (XLSX)

Table S4 Predicted *Wolbachia* sequences of *C. lectularius*. (XLS)

Table S5 Putative SNPs in *C. lectularius* sequences. (XLSX)

Table S6 Putative microsatellite loci in *C. lectularius* sequences. (XLSX)

References

- Reinhardt K, Siva-Jothy MT (2007) The biology of bedbugs (Cimicidae). *Ann Rev Entomol* 52: 351–374.
- Stucki A, Ludwig R (2008) Images in clinical medicine: bedbug bites *N Engl J Med* 359: 1047.
- Scarupa MD, Economides A (2006) Bedbug bites masquerading as urticaria. *J Allergy Clin Immunol* 117: 1508–1509.
- Leverkus M, Jochim RC, Schad S, Brocker EB, Andersen JF, et al. (2006) Bullous allergic hypersensitivity to bed bug bites mediated by IgE against salivary nitrophenol. *J Invest Dermatol* 126: 91–96.
- Abdel-Naser MB, Lotfy RA, Al-Sherbiny MM, Sayed ANM (2006) Patients with papular urticaria have IgG antibodies to bedbug (*Cimex lectularius*) antigens. *Parasitol Res* 98: 550–556.
- Stibich AS, Carbonaro PA, Schwartz RA (2001) Insect bite reactions: an update. *Dermatology* 202: 193–197.
- Rossi L, Jennings S (2010) Bed bugs: a public health problem in need of a collaborative solution. *J Environ Health* 72: 34–35.
- Hwang SW, Svoboda TJ, De Jong IJ, Kabasele KJ, Gogosis E (2005) Bed bug infestations in an urban environment. *Emerg Infect Dis* 11: 533–538.
- Goddard J, deShazo R (2009) Bed bugs (*Cimex lectularius*) and clinical consequences of their bites. *J Am Med Assoc* 301: 1358–1366.
- Goddard J (2010) Bed bugs: Do they transmit diseases? In: Goddard J: *Infectious Diseases and Arthropods*. pp 177–181.
- Panagiotakopulu E, Buckland PC (1999) *Cimex lectularius* L., the common bed bug from Pharaonic Egypt. *Antiquity* 73: 908–911.
- Boase CJ (2001) Bedbugs: back from the brink. *Pest Outl* 12: 159–162.
- Doggett SL, Geary MJ, Russell RC (2004) The resurgence of bed bugs in Australia, with notes on their ecology and control. *Environ Health* 4: 30–38.
- Ter Poorten MC, Prose NS (2005) The return of the common bedbug. *Pediatr Dermatol* 22: 183–187.
- Reinhardt D, Kempke RA, Nay LOR, Siva-Jothy MT (2009) Sensitivity to bites by the bedbug, *Cimex lectularius*. *Med Vet Entomol* 23: 163–166.
- Lee IY, Ree HI, An SJ, Linton JA, Yong TS (2008) Reemergence of the bedbug *Cimex lectularius* in Seoul, Korea. *J Parasitol* 46: 269–271.
- Anderson AL, Lefler K (2008) Bedbug infestations in the news: a picture of an emerging public health problem in the United States. *J Environ Health* 70: 24–27.
- Potter MF (2005) A bed bug state of mind: emerging issues in bed bug management. *Pest Control Technol* 33: 82–85.
- Myamba J, Maxwell CA, Asidi A, Curtis CF (2002) Pyrethroid resistance in tropical bedbugs, *Cimex hemipterus*, associated with use of treated bed nets. *Med Vet Entomol* 16: 448–451.
- Gangloff-Kaufmann J, Hollingsworth C, Hahn J, Hansen L, Kard B, et al. (2006) Bed bugs in America: a pest management industry survey. *Am Entomol* 52: 105–106.
- Moore DJ, Miller DM (2006) Laboratory evaluations of insecticide product efficacy for control of *Cimex lectularius*. *J Econ Entomol* 99: 2080–2086.
- Karunaratne SHPP, Damayanthi BT, Fareena MHJ, Imbuldeniya V, Hemingway J (2007) Insecticide resistance in the tropical bedbug *Cimex hemipterus*. *Pest Biochem Physiol* 88: 102–107.
- Romero A, Potter MF, Potter DA, Haynes KF (2007) Insecticide resistance in the bed bug: a factor in the pest's sudden resurgence? *J Med Entomol* 44: 175–178.
- Zhu F, Wigginton J, Romero A, Moore A, Ferguson K, et al. (2010) Widespread distribution of knockdown resistance mutations in the bed bug, *Cimex lectularius* (Hemiptera: Cimicidae), populations in the United States. *Arch Insect Biochem Physiol* 73: 245–257.
- Wang C, Saltzmann K, Chin E, Bennett GW, Gibb T (2010) Characteristics of *Cimex lectularius* (Hemiptera: Cimicidae), infestation and dispersal in a high-rise apartment building. *J Econ Entomol* 103: 172–177.
- Yoon KS, Kwon DS, Strycharz JP, Hollingsworth CS, Lee SIH, et al. (2008) Biochemical and molecular analysis of deltamethrin resistance in the common bed bug (Hemiptera: Cimicidae). *J Med Entomol* 45: 1092–1101.
- Tomita T, Scott JG (1995) cDNA and deduced protein sequence of CYP6D1: the putative gene for a cytochrome P450 responsible for pyrethroid resistance in house fly. *Insect Biochem Mol Biol* 25: 275–283.
- Cohen MB, Schuler MA, Berenbaum MR (1992) A host-inducible cytochrome P-450 from a host-specific caterpillar: molecular cloning and evolution. *Proc Natl Acad Sci U S A* 89: 10920–10924.
- Scott JG, Liu N, Wen Z (1998) Insect cytochromes P450: diversity, insecticide resistance and tolerance to plant toxins. *Comp Biochem Physiol C Pharm Toxicol* 121: 147–155.
- Mittapalli O, Neal JJ, Shukle RH (2007b) Tissue and life stage specificity of glutathione S-transferase expression in the Hessian fly, *Mayetiola destructor*: Implications for resistance to host allelochemicals. *J Insect Sci* 7: 20.
- Komagata O, Kasai S, Tomita T (2010) Overexpression of cytochrome P450 genes in pyrethroid-resistant *Culex quinquefasciatus*. *Insect Biochem Mol Biol* 40: 146–152.
- Gibbons JG, Janson EM, Hittinger CT, Johnston M, Abbot P, et al. (2009) Benchmarking next-generation transcriptome sequencing for functional and evolutionary genomics. *Mol Biol Evol* 26: 2731–2744.
- Vera JC, Wheat CW, Fescemyer HW, Frilander MJ, Crawford DL, et al. (2008) Rapid transcriptome characterization for a non-model organism using 454 pyrosequencing. *Mol Ecol* 17: 1636–1647.
- Zagrobely M, Scheibye-Alsing K, Jensen NB, Moller BL, Gorodkin J, et al. (2009) 454 pyrosequencing based transcriptome analysis of *Zygaena filipendulae* with focus on genes involved in biosynthesis of cyanogenic glucosides. *BMC Genomics* 10.
- 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.
- Bai X, Zhang W, Ormats 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.
- Zou Z, Najjar F, Wang Y, Roe B, Jiang H (2008) Pyrosequence analysis of expressed sequence tags for *Manduca sexta* hemolymph proteins involved in immune responses. *Insect Biochem Mol Biol* 38: 677–682.
- 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.
- Zhang F, Guo H, Zheng H, Zhou T, Zhou Y, et al. (2010) Massively parallel pyrosequencing-based transcriptome analyses of small brown planthopper (*Laodelphax striatellus*), a vector insect transmitting rice stripe virus (RSV). *BMC Genomics* 11: 303.
- Olafson PU, Lohmeyer KH (2010) Analysis of expressed sequence tags from a significant livestock pest, the stable fly (*Stomoxys calcitrans*), identifies transcripts with a putative role in chemosensation and sex determination. *Arch Insect Biochem Physiol* 74: 179–204.
- Jaworski DC, Zou Z, Bowen CJ, Wasala NB, Madden R, et al. (2010) Pyrosequencing and characterization of immune response genes from the American dog tick, *Dermacentor variabilis* (L.). *Insect Mol Biol* 19: 617–630.
- O'Neil ST, Dzurisin JDK, Carmichael RD, Lobo NF, Emrich SJ, et al. (2010) Population-level transcriptome sequencing of nonmodel organisms *Erynnis propertius* and *Papilio zelicaon*. *BMC Genomics* 11: 310.
- 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. *PLoS One* 5: e13708.
- Kirkness EF, Haas BJ, Sun W, Braig HR, Perotti MA, et al. (2010) Genome sequences of the human body louse and its primary endosymbiont provides

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. The authors thank Joshua Bryant, Benjamin Diehl, Andrew Hoelmer, and George Keeney for assisting with the collection and maintenance of pesticide-exposed *C. lectularius* populations. We gratefully acknowledge Harold Harlan's contribution of pesticide-susceptible bed bugs. We thank Binny Bhandary and Andrew Hoelmer for assisting with RNA isolations.

Author Contributions

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

- insights into the permanent parasitic lifestyle. Proc Natl Acad Sci U S A 107: 12168–12173.
45. The International Aphid Genomics Consortium (2010) Genome sequence of the pea aphid *Acyrtosiphon pisum*. PLoS Biol 8: e1000313.
 46. Liang H, Carlson JE, Leebens-Mack JH, Wall PK, Mueller LA, et al. (2008) An EST database for *Liriodendron tulipifera* L. floral buds: the first EST resource for functional and comparative genomics in *Liriodendron*. Tree Genet Genom 4: 419–433.
 47. Wang J-PZ, Lindsay BG, Leebens-Mack J, Cui L, et al. (2004) EST clustering error evaluation and correction. Bioinformatics 20: 2973–2984.
 48. Benoit JB, Lopez-Martinez G, Teets NM, Phillips SA, Denlinger DL (2009) Responses of the bed bug, *Cimex lectularius*, to temperature extremes and dehydration: levels of tolerance, rapid cold hardening and expression of heat shock proteins. Med Vet Entomol 23: 418–425.
 49. Wang XW, Luan JB, Li JM, Bao YY, Zhang CX, et al. (2010) *Denovo* characterization of a whitefly transcriptome and analysis of its gene expression during development. BMC Genomics 11: 400.
 50. Hamodrakas SJ, Bosshard HE, Carlson CN (1998) Structural models of the evolutionarily conservative central domain of silk-moth chorion proteins. Protein Eng 2: 201–207.
 51. Robson P, Wright GM, Youson JH, Keeley FW (2000) The structure and organization of lamprin genes: Multiple-copy genes with alternative splicing and convergent evolution with insect structural proteins. Mol Biol Evol 17: 1739–1752.
 52. Bochicchio B, Pepe A, Tamburro AM (2001) On (GGLGY) synthetic repeating sequences of lamprin and analogous sequences. Matrix Biol 20: 243–250.
 53. Lewis RV (1992) Spider silk: the unraveling of a mystery. Acc. Chem. Res. 25: 392–398.
 54. Maier D, Nagel AC, Gloc H, Hausser A, Kugler SJ, et al. (2007) Protein kinase D regulates several aspects of development in *Drosophila melanogaster*. BMC Dev Biol 7: 74.
 55. Ahier A, Rondard P, Gougnard N, Khayath N (2009) A new family of receptor tyrosine kinases with a Venus Flytrap binding domain in insects and other invertebrates activated by amino acids. PLoS One 4: e5651.
 56. Feyereisen R (2006) Evolution of insect P450. Biochem Soc Trans 34: 1252–1255.
 57. Li X, Schuler MA, Berenbaum MR (2007) Molecular mechanisms of metabolic resistance to synthetic and natural xenobiotics. Annu Rev Entomol 52: 231–253.
 58. McNeil GP, Schroeder AJ, Roberts MA, Jackson FR (2001) Genetic analysis of functional domains within the *Drosophila* LARK RNA-binding protein. Genetics 159: 229–240.
 59. Sutherland LC, Rintala-Maki ND, White RD, Morin CD (2005) RNA binding motif (RBM) proteins: a novel family of apoptosis modulators? J Cell Biochem 94: 5–24.
 60. Rendahl KGN, Gaukhshteyn DA, Wheeler TA, Hall JC (1996) Defects in courtship and vision caused by amino acid substitutions in a putative RNA-binding protein encoded by the *no-on-transient A (nonA)* gene of *Drosophila*. J Neurosci 16: 1511–1522.
 61. Stanewsky R, Fry TA, Reim I, Saumweber H, Hall JC (1996) Bioassaying putative RNA-binding motifs in a protein encoded by a gene that influences courtship and visually mediated behavior in *Drosophila*: *in vitro* mutagenesis of *nonA*. Genetics 143: 259–275.
 62. Ogura T, Anjiang T, Tsubota T, Nakakura T, Shiotsuki T (2009) Identification and expression analysis of ras gene in silkworm, *Bombyx mori*. PLoS One 4: e8030.
 63. Rhodes JD, Croghan PC, Dixon AFG (1997) Dietary sucrose and oligosaccharide synthesis in relation to osmoregulation in the pea aphid, *Acyrtosiphon pisum*. Physiol Entomol 22: 373–379.
 64. Price DRG, Wilkinson HS, Gatehouse JA (2007) Functional expression and characterisation of a gut facilitative glucose transporter, NIHT1, from the phloem-feeding insect *Nilaparvata lugens* (rice brown planthopper) Insect Biochem Mol Biol 37: 1138–1148.
 65. Kunji ER (2004) The role and structure of mitochondrial carriers. FEBS Lett 564: 239–244.
 66. Rebers JE, Riddiford LM (1988) Structure and expression of a *Manduca sexta* larval cuticle gene homologous to *Drosophila* cuticle genes. J Mol Biol 203: 411–423.
 67. Rebers JE, Willis JH (2001) A conserved domain in arthropod cuticular proteins binds chitin. Insect Biochem Mol Biol 31: 1083–1093.
 68. Magkrioti CK, Spyropoulos IC, Iconomidou VA, Willis JH, Hamodrakas SJ (2004) cuticleDB: A relational database of Arthropod cuticular proteins. BMC Bioinformatics 5: 138.
 69. Mittapalli O, Neal J, Shukle RH (2007) Antioxidant defense response in a galling insect. Proc Natl Acad Sci U S A 104: 1889–1894.
 70. Barbehenn RV (2002) Gut-based antioxidant enzymes in a polyphagous and graminivorous grasshopper. J Chem Ecol 28: 1329–1347.
 71. Stevens JL, Snyder MJ, Koener JF, Feyereisen R (2000) Inducible P450s of the CYP9 family from larval *Manduca sexta* midgut. Insect Biochem Mol Biol 30: 559–568.
 72. Rose RL, Goh D, Thompson DM, Verma KD, Heckel DG, et al. (1997) Cytochrome P450 (CYP)9A1: the first member of a new CYP family. Insect Biochem Mol Biol 27: 605–615.
 73. Davies TGE, Field LM, Usherwood PNR, Williamson MS (2007) A comparative study of voltage-gated sodium channels in the Insecta: implications for pyrethroid resistance in anopheline and other neopteran species. Insect Mol Biol 16: 361–375.
 74. Werren JH, Windsor DM, Gao L (1995) Distribution of *Wolbachia* among Neotropical arthropods. Proc R Soc Lond B Biol Sci 265: 1447–1452.
 75. Jeyaprakash 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.
 76. Werren JH, Windsor DM (2004) *Wolbachia* infection frequencies in insects: Evidence of a global equilibrium? Proc R Soc Lond B Biol Sci 267: 1277–1285.
 77. Werren JH (1997) *Wolbachia* run amok. Proc Natl Acad Sci U S A 94: 11154–11155.
 78. Klasson L, Walker T, Sebahia M, Sanders MJ, Quail MA, et al. (2008) Genome evolution of *Wolbachia* strain wPip from the *Culex pipiens* group. Mol Biol Evol 25: 1877–1887.
 79. Laven H (1967) Eradication of *Culex pipiens fatigans* through cytoplasmic incompatibility. Nature 216: 383–384.
 80. Dobson SL, Fox CW, Jiggins FM (2002) The effect of *Wolbachia*-induced cytoplasmic incompatibility on host population size in natural and manipulated systems. Proc Biol Sci 269: 437–445.
 81. Saridaki A, Bourtzis K (2010) *Wolbachia*: more than just a bug in insects' genitals. Curr Opin Microbiol 13: 67–72.
 82. Hosokawa T, Koga R, Kikuchi Y, Meng XY, Fukatsu T (2010) *Wolbachia* as a bacteriocyte-associated nutritional mutualist. Proc Natl Acad Sci U S A 107: 769–774.
 83. Behura SK (2006) Molecular marker systems in insects: current trends and future avenues. Mol Ecol 15: 3087–3113.
 84. Usinger R (1966) Monograph of Cimicidae. Thomas Say Foundation, Vol. 7. College ParkMD: Entomological Society of America.
 85. 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.
 86. Cogill P, Finn RD, Bateman A (2008) Identifying protein domains with Pfam database. Curr Protoc Bioinform 2: 2.5.
 87. Eddy SR (1998) Profile hidden Markov models. Bioinformatics 14: 755–763.
 88. 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.
 89. 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.
 90. 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 35: D480–D484.
 91. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, et al. (2006) From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res 34: D354–357.
 92. Kanehisa M, Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 28: 27–30.
 93. Thurston MI, Field D (2005) Msatfinder: detection and characterization of microsatellites. Distributed by the authors at URL: <http://www.genomics.ceb.ac.uk/msatfinder/>.
 94. Van Hiel MBV, Wielendaele V, Temmerman L, Soest SV, Vuerinckx K, et al. (2009) Identification and validation of housekeeping genes in brains of the desert locust *Schistocerca gregaria* under different developmental conditions. BMC Mol Biol 10: 56.