Sequencing, *De Novo* Assembly and Annotation of the Colorado Potato Beetle, *Leptinotarsa decemlineata*, Transcriptome

Abhishek Kumar[1,2], Leonardo Congiu[1], Leena Lindström[3], Saija Piiroinen[3], Michele Vidotto[1¤], Alessandro Grapputo[1]*

1 Department of Biology, University of Padova, Padova, Italy, 2 Department of Genetics & Molecular Biology in Botany, Institute of Botany, Christian-Albrechts-University at Kiel, Kiel, Germany, 3 Centre of Excellence in Biological Interactions Research, Department of Biological and Environmental Science, University of Jyväskylä, Jyväskylä, Finland

Abstract

Background: The Colorado potato beetle (*Leptinotarsa decemlineata*) is a major pest and a serious threat to potato cultivation throughout the northern hemisphere. Despite its high importance for invasion biology, phenology and pest management, little is known about *L. decemlineata* from a genomic perspective. We subjected European *L. decemlineata* adult and larval transcriptome samples to 454-FLX massively-parallel DNA sequencing to characterize a basal set of genes from this species. We created a combined assembly of the adult and larval datasets including the publicly available midgut larval Roche 454 reads and provided basic annotation. We were particularly interested in diapause-specific genes and genes involved in pesticide and *Bacillus thuringiensis* (Bt) resistance.

Results: Using 454-FLX pyrosequencing, we obtained a total of 898,048 reads which, together with the publicly available 804,056 midgut larval reads, were assembled into 121,912 contigs. We established a repository of genes of interest, with 101 out of the 108 diapause-specific genes described in *Drosophila montana*; and 621 contigs involved in insecticide resistance, including 221 CYP450, 45 GSTs, 13 catalases, 15 superoxide dismutases, 22 glutathione peroxidases, 194 esterases, 3 ADAM metalloproteases, 10 cadherins and 98 calmodulins. We found 460 putative miRNAs and we predicted a significant number of single nucleotide polymorphisms (29,205) and microsatellite loci (17,284).

Conclusions: This report of the assembly and annotation of the transcriptome of *L. decemlineata* offers new insights into diapause-associated and insecticide-resistance-associated genes in this species and provides a foundation for comparative studies with other species of insects. The data will also open new avenues for researchers using *L. decemlineata* as a model species, and for pest management research. Our results provide the basis for performing future gene expression and functional analysis in *L. decemlineata* and improve our understanding of the biology of this invasive species at the molecular level.

Introduction

The Colorado potato beetle, *Leptinotarsa decemlineata* (Say), is the major defoliator of potato throughout the northern hemisphere [1–5]. Both larvae and adults feed on potato plants causing damage to potato fields and financial losses to farmers [6]. The beetle is native to Mexico and south-eastern USA [7], where it lives on wild solanaceous species such as *Solanum rostratum* and *S. angustifolium* [8]. The shift to potato occurred sometime before 1850 in the US [9], when potato farming reached the distribution range of the beetle [2]. The beetle spread rapidly throughout the US reaching the east coast before 1880 [2], and was accidentally introduced to Europe (in France) in the 1920s. In 50 years it

spread throughout Europe except for the UK and Scandinavia. Its current range covers 16 million km^2 in North America, Europe and Asia. It is currently spreading further eastwards and also towards higher latitudes [9–12].

Several factors have contributed to the beetle's high success as an invader and a pest species: adaptation to the potato host, high fecundity, the ability to synchronize its life cycle through diapause, and a high capability to evolve resistance to insecticide [1,13]. Diapause is a physiological state of dormancy that allows insects to escape unfavorable conditions, such as harsh winters or drought. The decision to enter diapause is mainly determined by day length, but is affected by food availability, temperature and moisture conditions [14–17]. Diapause is therefore a critical

component of insect phenology [18]. More importantly, insecticide resistance can interact with diapause in insects, including *L. decemlineata* [19–22]. This gives diapause a central role in insect pest management [23,24]. Developing a more comprehensive understanding of diapause and its influence on insect life-history traits will offer insights into other biological processes and help in planning new pest management strategies [25].

The first large-scale use of insecticides in agricultural crops was for the suppression of *L. decemlineata* [3]. Insecticides were very successful in controlling this serious pest until it developed widespread resistance to DDT in the mid-1950s [2]. In the laboratory conditions, the beetle has since developed resistance to at least 52 different compounds covering all the major pesticide classes, including pyrethroids, organophosphates, neonicotinoids [1], and, only in laboratory populations, also to *Bacillus thuringiensis* (Bt) [26,27]. Despite this, insecticide compounds remain the most used and only efficient means of managing beetle populations. At the same time, there is growing concern over the evolution of resistance and the environmental consequences of increased dosages of insecticides [1,28].

The beetle is important for both basic and applied biology – from invasive biology, through insect phenology to pest-species management. In fact, *L. decemlineata* has been included in the i5k insect genomes project (http://arthropodgenomes.org/wiki/Main_Page; http://www.ncbi.nlm.nih.gov/bioproject/PRJNA171749) in 2012 [29]. A first un–annotated draft of the genome has been made available while our manuscript was in its final preparation and therefore it could not be included in our analysis. Genetic investigations of *L. decemlineata* biology and its resistance to both chemical pesticides and Bt have relied on homology-based gene-by-gene cloning, on low throughput EST sequencing [18,30,31] and, more recently, the beetle larval midgut has been subject to 454 pyrosequencing [32].

Next-generation sequencing methods, such as 454 pyrosequencing, are cost-effective methods for the transcriptome characterization of insect species that lack a fully-sequenced genome [33–35]. The deeper sequencing coverage of the 454 method and an accurate base calling allow for *de-novo* transcriptome assembly and the characterization of genes without a reference genome. The massive number of expressed sequence tags obtained with this method facilitates the discovery and identification of new genes and the analysis of gene expression by providing a reference transcriptome for cDNA microarrays. It also facilitates the identification of such novel Type I genetic markers as microsatellites and SNPs for population genomic and quantitative traits locus (QTL) analyses [36,37].

We used 454 FLX Titanium-based pyrosequencing to generate a substantial dataset of transcripts reads of the *L. decemlineata* transcriptome. Together with the publicly available midgut larval reads [32], we obtained 121,912 contigs, of which 41.15% were similar to known protein or nucleotide sequences. We performed the *in silico* identification of Type I genetic markers and characterized genes of interest for diapause, detoxification pathways and insecticide target proteins. We annotated the combined assembly, including 8,993 transcripts available at NCBI (June 2012). All data and assembly are available at http://www.bio.unipd.it/grapputo/CPB-Webpage. Our results will provide the basis for performing future gene expression studies and functional analysis in *L. decemlineata* and improve our understanding of the biology of this invasive species at the molecular level.

Results and Discussion

Transcriptome assembly characteristics

Using the Roche 454 pyrosequencing method, we obtained 456,909 transcriptomic reads from adult beetles and 444,435 reads from beetle larvae for a total of 901,344 reads corresponding to 64.23 Mbp. After a cleaning step for removing adapters, low quality bases and contaminants (bacteria, viruses and potato sequences) we remained with 445,257 (97.45% of the original reads) and 442,791 (99.63%) reads for adults and larvae, respectively. These reads were combined with the publicly available 839,061 Roche 454 midgut larval reads [38], which, after cleaning as describe above, produced 804,056 (95.83%) reads. The total 1,702,104 reads from these three datasets were assembled into 117,848 contigs and 4,064 singletons for a total of 121,912 partial transcripts (**Table S1**) using the Mira 3.2 assembler [39]. Contig lengths varied from 41 to 7,034 bp with an average of 537 bp (**Figure S1 and Table S1**). Singletons ranged from 40 to 583 bp with an average length of 214 bp (**Tables 1 and S1**). Contigs and singletons had an average GC content of 36% and 35%, respectively (**Tables S1**). This was lower than the GC content (46%) of the red flour beetles (*Tribolium castaneum*) transcriptome suggesting that the genome of *L. decemlineata* could be very rich in AT as shown in other insects [40].

For simplicity we will not make further distinction between contigs and singletons and refer to both as contigs unless otherwise stated.

BLAST analyses

Of 121,912 *L. decemlineata* contigs, 41.15% showed significant similarity (E value$<$1e^{-3}) to proteins in the GenBank non-redundant (nr) database. Although we used an E-value threshold of $<$1e^{-3} for this analysis, the majority of matches were below this threshold, i.e. between 1e^{-4} and 1e^{-180} (**Figure 1A**). As we expected, the major fraction of sequences with hits in GenBank matched insect proteins (79.23%) (**Figure 1B**). Within insects, *T. castaneum* had the highest share of matches with 37.8% (**Figure 1B**). In contrast *L. decemlineata* had only 2.81% of the hits, confirming the low amount of annotated genomic information available on this species.

GO ontology assignment

Functional annotation is an essential requirement for understanding the transcriptomic data of non-model organisms. Gene Ontology (GO) facilitates the functional characterization of genes, transcripts and proteins of many organisms in terms of cellular components, biological processes and molecular functions in a species-independent fashion [41]. Currently, this approach is a standard method for corroborating overall annotation of 454-sequenced transcriptome data [42–46]. We have used this method for the functional annotation of *L. decemlineata* transcripts using the Blast2GO suite [41].

The derived *L. decemlineata* transcripts were assigned to three functional groups based on GO terminology: biological process, molecular function and cellular component (**Table S2**). We traced 42,208 contigs to biological process terms (**Figure 2**) with the following five top categories: catabolic process (5,453) signal transduction (2,919) carbohydrate metabolic process (2,869), protein modification process (2,696) and cell differentiation (2,603). Similarly, 18,738 contigs were assigned to cellular component terms (**Figure 2**) with the five top components being: protein complex (5,878), mitochondrion (2,174), ribosome (1,560), plasma membrane (1,453) and nucleoplasm (1,408). Finally, 20,438 contigs were linked to molecular function terms

Table 1. Summary of assembly statistics.

MIRA contigs (#)	117,848
MIRA contigs on assembly (%)	96.67
length (Mb)	63.359063
average_length (bp)	537.634
median_length (bp)	426
min_length (bp)	41
max_length (bp)	7,034
GC_conten (%)	36.1
average quality (phred)	41.5986
MIRA average_coverage (reads/contig)	11.9784
N25 stats	25% of total sequence length is contained in the 9245 sequences >= 1,081 bp
N50 stats	50% of total sequence length is contained in the 30645 sequences >= 570 bp
N75 stats	75% of total sequence length is contained in the 63854 sequences >= 409 bp
MIRA average_coverage (no singlets) (bp/position)	4.02896
MIRA median_coverage (no singlets) (bp/position)	2.17
MIRA standard_dev coverage (no singlets) (bp/position)	6.2343
MIRA Singletons (#)	4,064
MIRA Singletons (%)	3.33
length (Mb)	0.87
average_length (bp)	214.044
median_length (bp)	205
min_length (bp)	41
max_length (bp)	538
GC_conten (%)	34.89
average quality (phred)	26.4683
N25 stats	25% of total sequence length is contained in the 521 sequences >= 369 bp
N50 stats	50% of total sequence length is contained in the 1,179 sequences >= 295 bp
N75 stats	75% of total sequence length is contained in the 2,065 sequences >= 203 bp

doi:10.1371/journal.pone.0086012.t001

(**Figure 2**) with nucleotide binding (6,058), peptidase activity (2,037), DNA binding (1,881), structural molecule activity (1,593) and enzyme regulator activity (948), being the five top categories. All GO terms in these three categories are listed in **Table S3.** Further, we assigned 667 enzymatic codes (EC) to 16,656 contigs (**Table S3**) encompassing all six groups, EC1–EC6, with the highest transcript numbers assigned to EC2 and the lowest to EC5 (**Table S3**). A total of 539 transcripts were assigned more than one enzymatic code.

The annotated contigs are provided on our group website (http://www.bio.unipd.it/grapputo/CPB-Webpage/).

Enzymatic pathway analyses

We mapped 16,656 contigs to 663 enzymes of 134 different reference canonical pathways in the Kyoto Encyclopedia of Genes and Genomes (KEGG) (**Table S4**). The coverage of transcriptomic sequences per pathway ranged from 1 to 888; whereas the coverage of these sequences per enzyme ranged from 1 to 224. There were 17 pathways covered by more than 250 sequences (**Figure 3A**). The top five were: purine metabolism (888 sequences; KEGG map00230), starch and sucrose metabolism (596; map00500), oxidative phosphorylation (553; map00190), Nitrogen metabolism (525; map00910) and pyrimidine metabo-

lism (410; map00240). A further 23 KEGG pathways had a coverage range of between 150 and 250 putative enzymatic sequences (**Figure 3B**). The ratio of contig to singleton sequences in KEGG pathways was approximately 3.4:1. Similar results for KEGG pathways have been observed for other insect transcriptomes [43–46].

Identification of putative microRNAs

MicroRNAs (miRNAs) are small non-coding RNAs that have significant roles in the regulation of gene and protein expression in various biological processes [47,48]. In order to identify putative novel microRNAs in the transcriptome of *L. decemlineata*, we scanned all known metazoan microRNA sequences from a miRNAs database (miRBase - Release 20, June 2013) [49]. We identified 460 putative miRNA as summarized in **Table S5.**

Comparison of adult and larval transcriptomes

Since each read was labeled with the library of origin before the assembly, we were able to identify those contigs formed by reads from either the adult or the larval transcriptome and those from either one of the two larval transcriptomes. Out of 212,912 contigs obtained from the combined assembly of the three data sets, 43,050 contained reads from both adults and larvae while 19,811
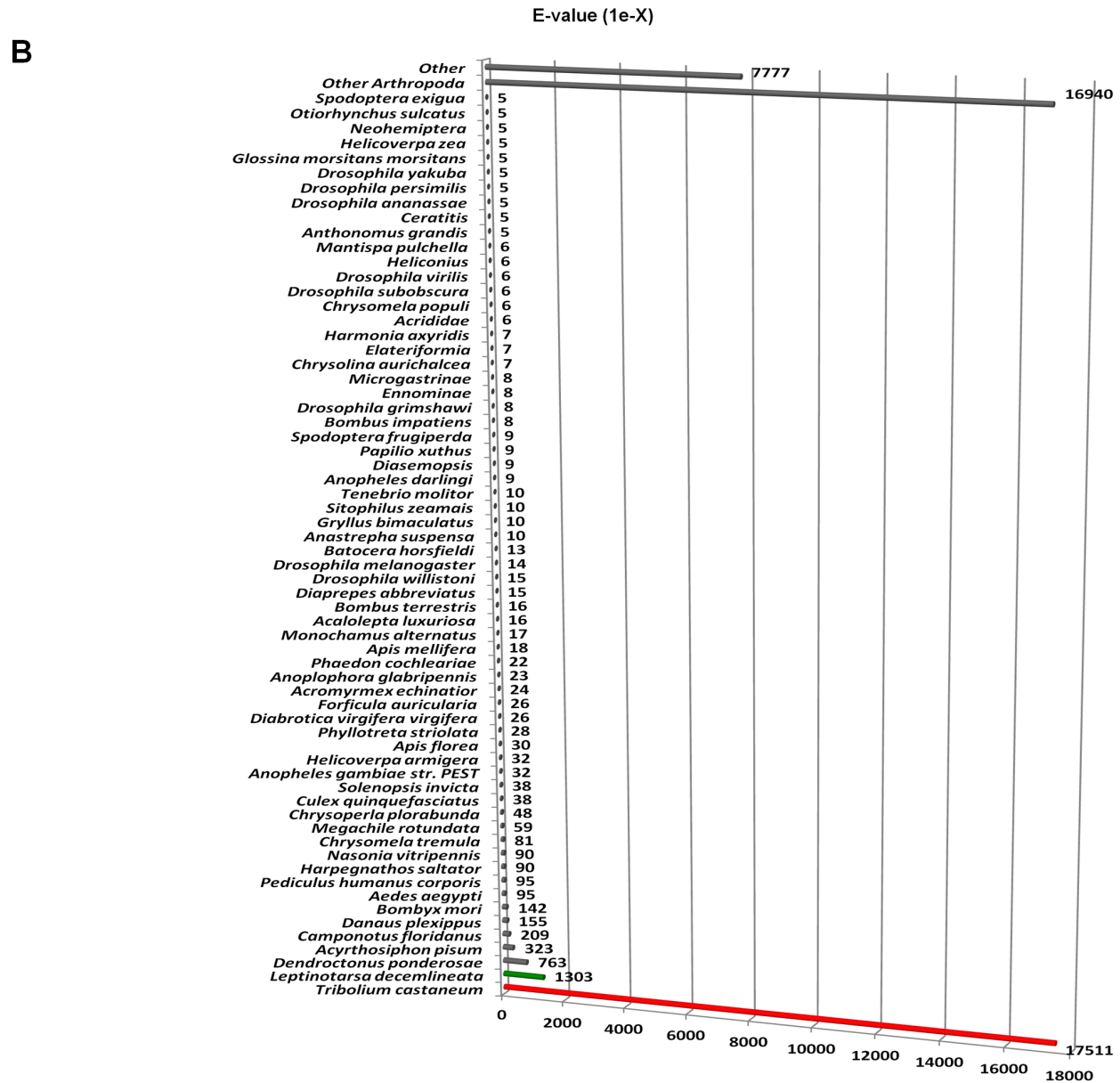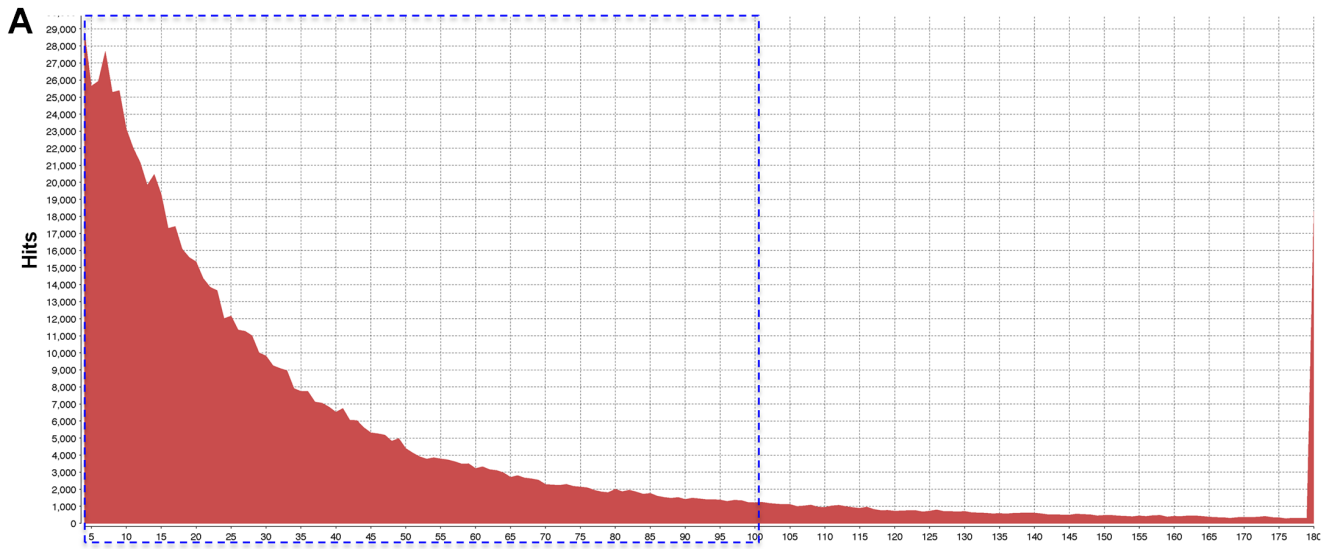
A

Hits

E-value (1e-X)

B

Taxon	Hits
Other	7777
Other Arthropoda	16940
Spodoptera exigua	5
Otiorhynchus sulcatus	5
Neohemiptera	5
Helicoverpa zea	5
Glossina morsitans morsitans	5
Drosophila yakuba	5
Drosophila persimilis	5
Drosophila ananassae	5
Ceratitis	5
Anthonomus grandis	5
Mantispa pulchella	6
Heliconius	6
Drosophila virilis	6
Drosophila subobscura	6
Chrysomela populi	6
Acrididae	6
Harmonia axyridis	7
Elateriformia	7
Chrysolina aurichalcea	7
Microgastrinae	8
Ennominae	8
Drosophila grimshawi	8
Bombus impatiens	8
Spodoptera frugiperda	9
Papilio xuthus	9
Diasemopsis	9
Anopheles darlingi	9
Tenebrio molitor	10
Sitophilus zeamais	10
Gryllus bimaculatus	10
Anastrepha suspensa	10
Batocera horsfieldi	13
Drosophila melanogaster	14
Drosophila willistoni	15
Diaprepes abbreviatus	15
Bombus terrestris	16
Acalolepta luxuriosa	16
Monochamus alternatus	17
Apis mellifera	18
Phaedon cochleariae	22
Anoplophora glabripennis	23
Acromyrmex echinatior	24
Forficula auricularia	26
Diabrotica virgifera virgifera	26
Phyllotreta striolata	28
Apis florea	30
Helicoverpa armigera	32
Anopheles gambiae str. PEST	32
Solenopsis invicta	38
Culex quinquefasciatus	38
Chrysoperla plorabunda	48
Megachile rotundata	59
Chrysomela tremula	81
Nasonia vitripennis	90
Harpegnathos saltator	90
Pediculus humanus corporis	95
Aedes aegypti	95
Bombyx mori	142
Danaus plexippus	155
Camponotus floridanus	209
Acyrthosiphon pisum	323
Dendroctonus ponderosae	763
Leptinotarsa decemlineata	1303
Tribolium castaneum	17511

Figure 1. Summary of BLAST results for transcriptomic sequences vs Genbank non-redundant (nr) database. A. E-value distribution suggested that majority of hits ranged from $1e^{-5}$ to $1e^{-100}$ (blue dashed square). **B**. Taxonomic distribution of the top BLAST hits of *L. decemlineata* contigs. Majority of fractions belongs to insects with the beetles *T. castaneum* (red bar) and *L. decemlineata* (green bar) as the two top species. The low number of hits to *L. decemlineata* was due to the low coverage of this beetle in the current nr database.
doi:10.1371/journal.pone.0086012.g001

and 59,051 contained reads from only adults and from only larvae (full larval [FL] data set + midgut larval [ML] data set). The comparison between the two larval transcriptomes allowed to identify contigs entirely composed by reads of the same library. These putative specific contigs were 18,470 for FL and 30,540 for ML.

Genes of Interest

We searched for homologies between *L. decemlineata* sequences and insect model genomes, such as *Drosophila* and *Tribolium*, focusing on genes putatively involved in diapause, detoxification and insecticide resistance.

Diapause-associated genes. We scanned for protein sequences encoded by eight transcripts previously reported to be down-regulated as beetles enter the diapause-maintenance phase of diapause development [18]. These transcripts are summarized in **Table S6**. To explore how many homologs are found in our transcriptome dataset in comparison to *Drosophila* and *Tribolium*, we used serpin as a test case. The serpin superfamily is characterized by a highly conserved domain of 35–50 kDa and these protease inhibitors play essential roles in the regulation of proteolytic cascades [50,51]. About 30 different serpins have been previously identified in other insects with some alternatively-spliced isoforms [52–54]. We report 41 contigs with homologous to serpins from the *L. decemlineata* transcriptome, which include most of the serpins known in insects as summarized by a Bayesian phylogenetic tree (**Figure 4**). Several serpins are *Drosophila*-specific including alternatively-spliced isoforms of Spn4/Spn42 (**Figure 4**) [52–54]. Similarly, as the phylogenetic tree shows, *L. decemlineata* presents several serpins arose likely by several duplication events (red in **Figure 4**), as previously reported in *T. castaneum* on chromosome 8 [53], suggesting that this duplication event could be common to all Cucujiformia or even all Coleoptera.

A recent study on *Drosophila montana* unraveled 108 genes that are instrumental in photoperiodic reproductive diapause [55]. Upon scanning these 108 genes, 101 were found in the beetle transcriptome spanning different functional categories (**Table 2**) with various homologs as listed in **Table S7**. The seven diapause candidate genes missing in this combined assembly were: disco, couch potato (cpo), CrebB17A, inaF, narrow abdomen (na), nnchung (nan) and timeless. **Table S7** also shows the results of the identification of dataset specific transcripts indicating that all these 101 transcripts involved in reproductive diapause in drosophila were expressed in both larvae and adult beetles. With our data we could not investigate differential expression of these transcripts in the different stages and therefore further studies are needed to assess if these genes have a role in the beetle diapause.

Putative transcripts involved in insecticide resistance. Since the first insecticide applications against the Colorado potato beetle in the 1950s, this pest has developed resistance to all the major classes of chemical insecticides (reviewed in [1,28]). Resistance mechanisms can be highly diverse even within a small geographic area [56], involving a variety of different genes including cytochrome P450 monooxygenases (CYPs), glutathione S-transferases (GSTs), superoxide dismutases (SODs), catalases (CATs), ADAM metalloproteases (ADAMs), cadherins (CADHs), Calmodulins (CALMs), glutathione peroxidases (GPXs), esterases and ascorbate peroxidases. Insect CYPs are also important in

metabolizing plant secondary metabolites [57] and Zhang et al. [58] found 38 up-regulated CYPs in Colorado potato beetles when the insects were fed on Solanaceae.

We have created a catalogue of genes putatively involved in insecticide resistance in *L. decemlineata* from the transcriptome assembled here using BLAST homology searches [59] at E-values lower than $1e^{-3}$ and listed in **Table 3**. The contigs encoding these genes are not necessarily complete ORFs and therefore it is likely that several contigs correspond to the same gene. From our dataset, it is difficult to discriminate between alleles of the same gene from recent duplication events as we sequenced a pool of many individuals. Therefore the number of genes per gene family could have been overestimated. However, the number of genes reported in **Table 4** is smaller than the number reported in the mosquito *Culex pipiens quinquefasciatus* [60].

GSTs are a diverse family of enzymes that play a central role in insecticide resistance in insects. GSTs can metabolize insecticides by facilitating their reductive dehydrochlorination or by conjugation reactions with reduced glutathione to produce water-soluble metabolites that are more readily excreted [61,62]. One or more GSTs have been implicated in resistance to organophosphates in *Musca domestica*, to organochlorine 1,1,1-trichloro-2,2-bis(*p*-chlorophenyl)-ethane (DDT) in *D. melanogaster* and to pyrethroids in *Nilaparvata lugens* [63]. We examined the status of GSTs in our assembled transcriptome and compared to *T. castaneum* and *D. melanogaster*. We found 45 contigs homologous to GSTs in *L. decemlineata* transcriptome sequences and their number is comparable to other insects with the expansion of epsilon and delta subgroups as illustrated in the Bayesian phylogenetic tree (**Figure 5**). GSTZs are found in many eukaryotic species, including insects [64]. They are implicated in the detoxification of xenobiotics containing chloride in the silkmoth, *Bombyx mori* [62,65]. We found one contig (Ld_c3961) homologous to the zeta-class glutathione S-transferases (GSTZs). This class of GST has not been observed in the red flour beetles and in the mosquito *C. p. quinquefasciatus* [60].

L. decemlineata laboratory populations have shown resistance to crystal (Cry) proteins derived from *Bacillus thuringiensis* (Bt) [26,27]. Transgenic crops producing toxins and Cry proteins of Bt continue to be widely used in insect pest management. Mechanisms of resistance are different in different species and resistance to Bt has been subject of several studies also in the beetle [27,66,67]. On *L. decemlineata*, it has been shown that ADAM metalloprotease serves as receptor for Cry3Aa toxin [68]. Cry3Aa toxin specifically binds to calmodulin (CalM) in a calcium-independent manner [69] and also to the toxin-binding fragments of cadherin [70]. We have taken in consideration these three genes in our analyses and using scientific literature, we built a catalog of those genes known to be involved in resistance to Bt, including several transcripts recently isolated from *Diabrotica virgifera* showing responses to the Bt toxin Cry3Bb1 [71]. We traced these transcripts in our assembly and found the majority of them present (**Table 3**) with contigs Ld_c74929, Ld_rep_c32791 and Ld_c240 the closest hits for ADAM, CalM and cadherin, respectively.

We selected actin, among the genes overexpressed in responsive *D. virgifera* to Bt (from **Table S8**), and examined how many actin homologs are expressed in the beetle transcriptome. We scanned actin transcripts in *L. decemlineata* and compared them with known
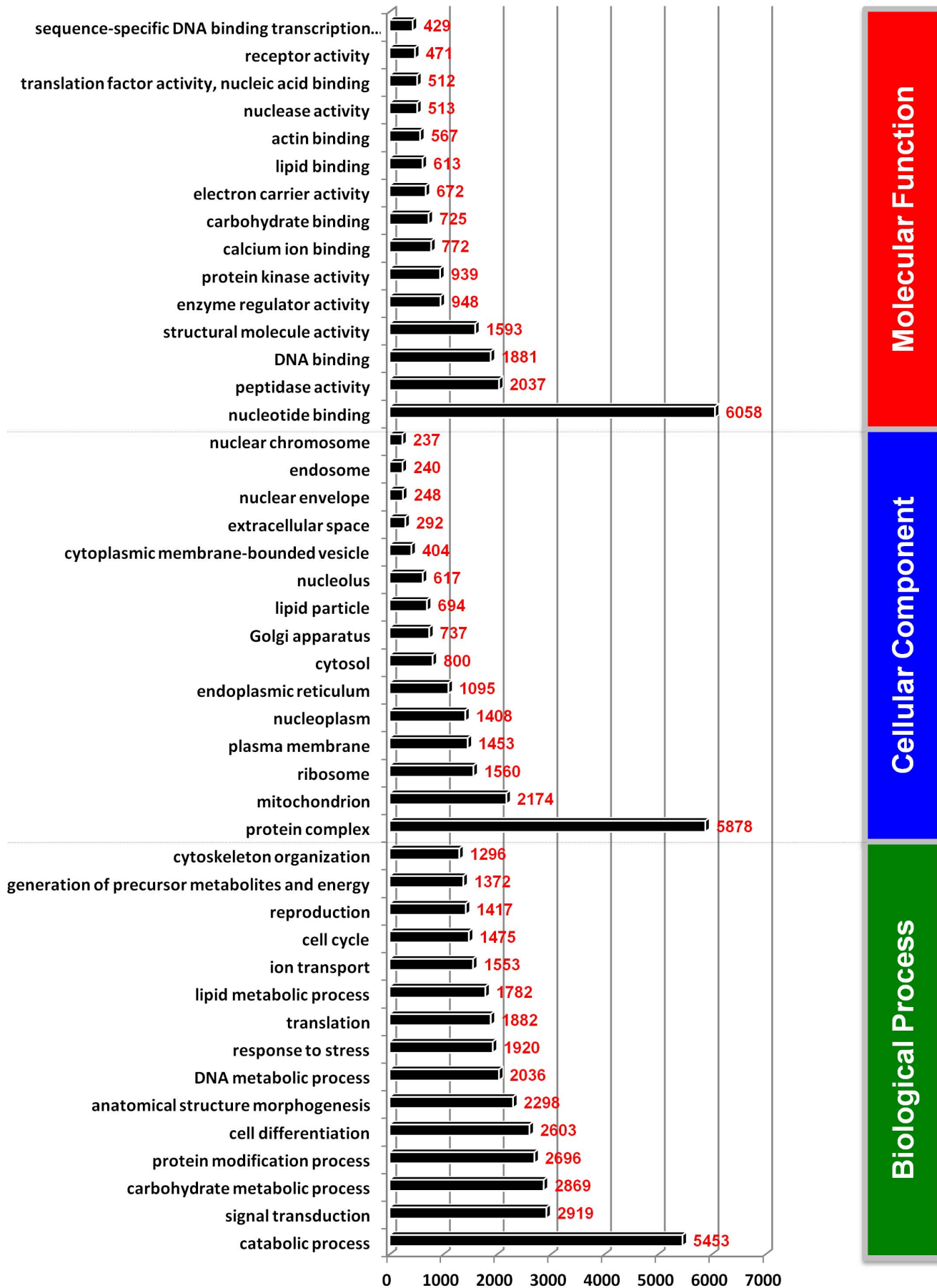
Figure 2. Description of the three categories of Gene Ontology (GO) terms for the transcriptomic sequences of *L. decemlineata.* **The** top 15 GO terms of each category are shown. A detailed summary is listed in **Table S4**.
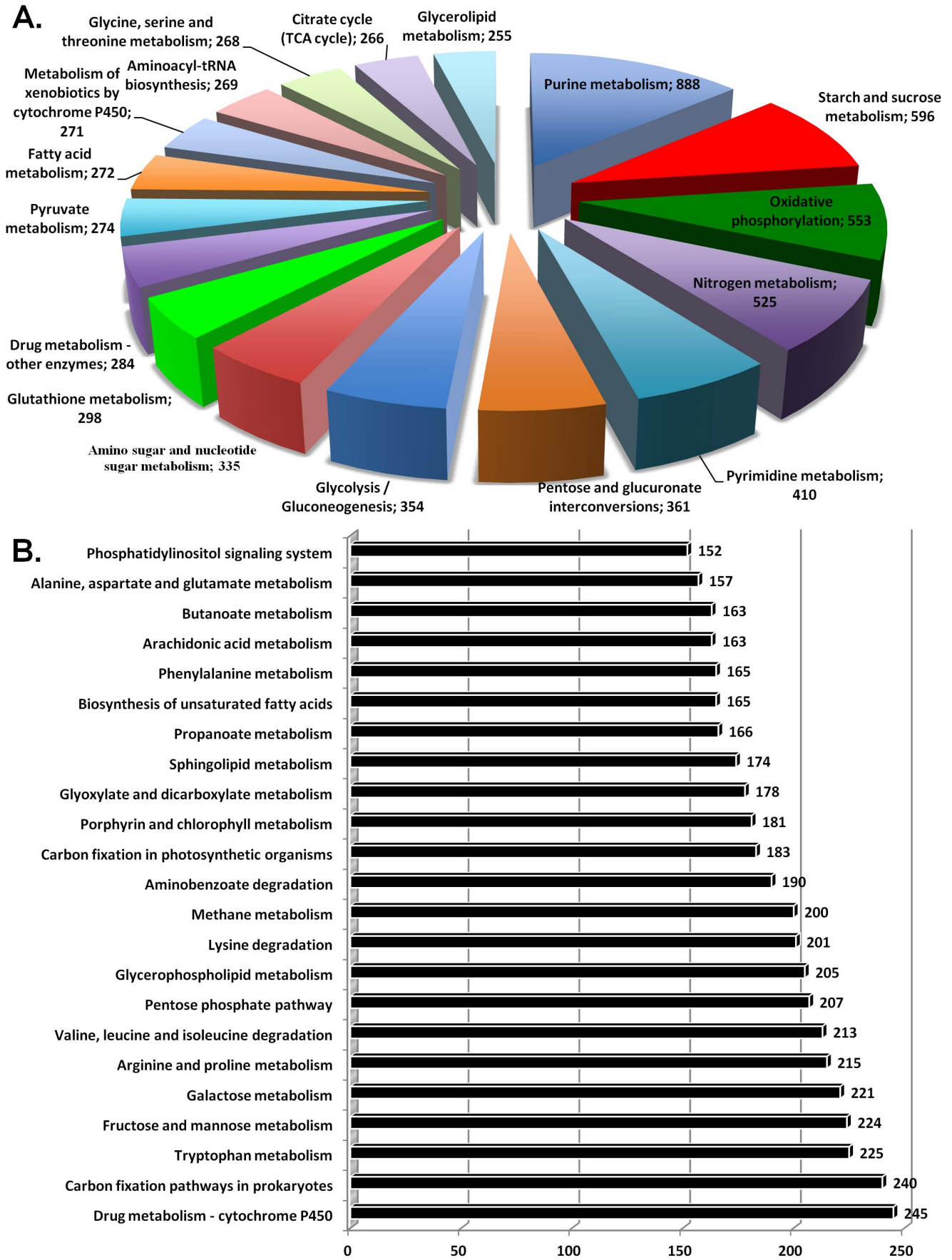
Figure 3. Summary of enzymatic KEGG pathways. A. KEGG pathways comprising more than 250 transcriptomic sequences. **B**. KEGG pathways comprising transcriptomic sequences between 150 and 250. A detailed summary is listed in **Table S5**.
doi:10.1371/journal.pone.0086012.g003

actins in *T. castaneum* and *D. melanogaster*. Actin is a major contractile protein found in all eukaryotic cells and constitutes 1–2% of the total cellular protein of eukaryotic genomes [72,73]. Several actin-like proteins, known as actin-related proteins (ARPs), are also present in various eukaryotic organisms. There are at least eight different ARP sub-families conserved in insects with different physiological roles such as actin polymerization (ARP2-3), chromatic remodeling (ARP4-6 and ARP8) and dynein mobility (ARP1 and ARP10) [72,73]. We identified several contigs homologous to the majority of conventional actins and the actin-related proteins (ARPs) of *T. castaneum* and *D. melanogaster* as depicted by the Bayesian phylogenetic tree (**Figure 6**).

In summary, we scanned the combined assembly of the adult and larval transcriptome of *L. decemlineata* for genes of interest. We found a comparable number of genes to, and which were conserved in, two other insects models, *T. castaneum* and *D. melanogaster*. We also carried out Bayesian phylogenetic tree for three representative genes from this dataset. The number of hits was generally higher for the larva than the adult transcriptome as it was higher for the ML than the FL dataset. This could be due to the higher number of reads sequenced for the midgut larval transcriptome which generated 1.65 times more contigs than ours datasets. An alternative hypothesis would be that most of Bt target receptors are specifically expressed in the midgut.
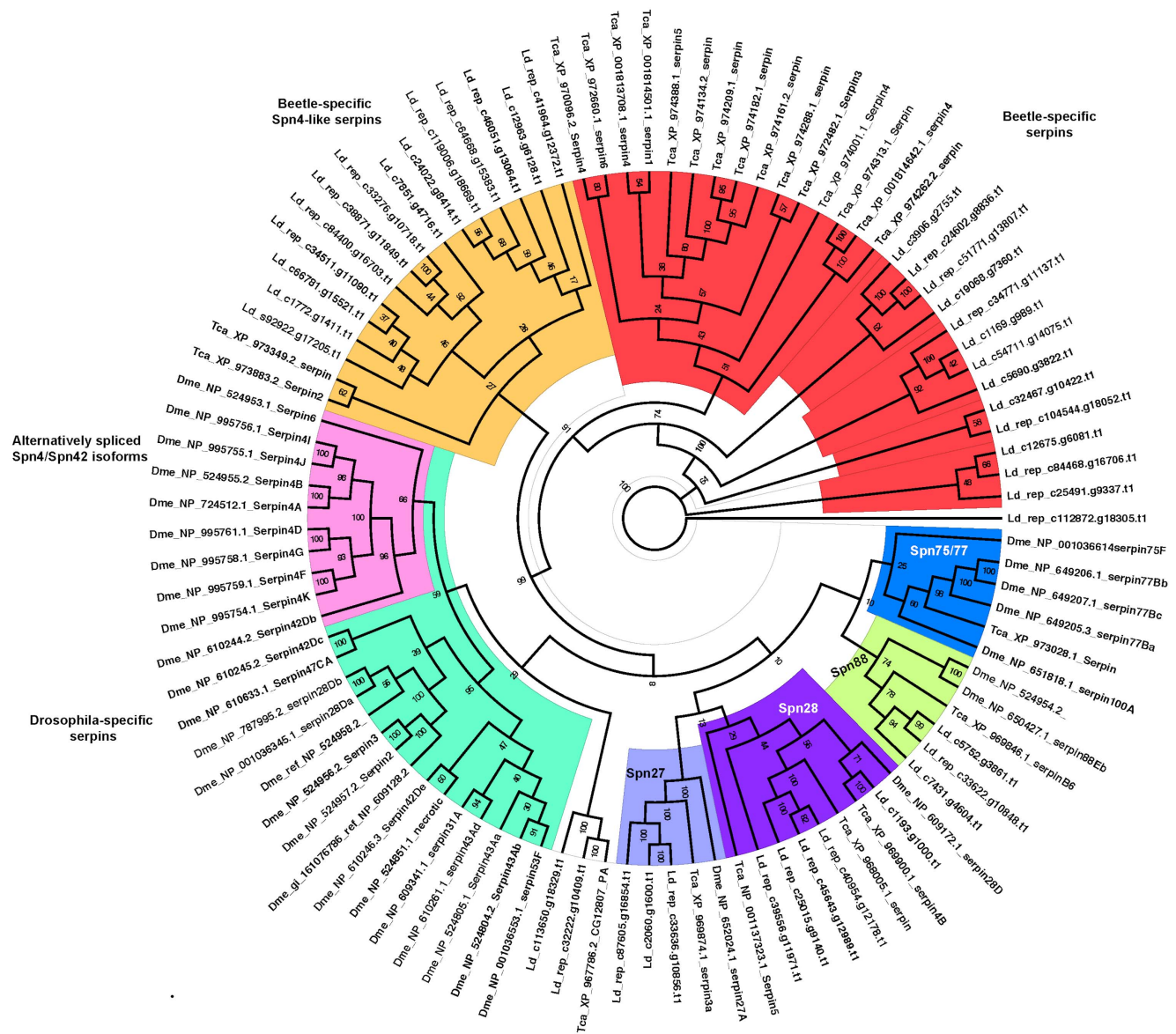

Figure 4. Bayesian phylogeny of serpins from *T. castaneum* and *D. melanogaster* genomic and *L. decemlineata* transcriptomic sequences. Several serpins are specific for *D. melanogaster* (green box) with many alternatively-spliced isoforms of Spn4 (synonym Spn42, pink box) whereas beetle specific Spn4-like serpins are marked in yellow. *L. decemlineata* has several tandemly-duplicated serpins (red box), as previously reported in *T. castaneum* on chromosome 8 [53]. The tree was based on the WAG +G+I protein model and run for 10,342,000 generations. Posterior probability values are indicated at the nodes. Sequences are named by a prefix of three letters indicating the species (Dme – *D. melanogaster*, Tca – *T. castaneum*, Ld – *L. decemlineata*) followed by either the accession ids and name in NCBI (for Dme and Tca) or contig name (for Ld).
doi:10.1371/journal.pone.0086012.g004

Table 2. Summary of putative genes involved in diapause in comparison to *D. montana*.

Function	D. montana	L. decemlineata
Circadian rhythm	26	23
Cold tolerance	02	02
Courtship behavior	14	13
Diapause	07	05
Heat tolerance	27	27
Housekeeping gene	07	07
Phototransduction	25	24
Total	**108**	**101**

doi:10.1371/journal.pone.0086012.t002

Status of frequently-occurring eukaryotic Pfam protein domains

Protein domains are the building blocks of proteins as well as their evolutionary conserved units. The Pfam database is a large collection of multiple sequence alignments covering approximately 13,000 protein families [74]. The curated protein domains in Pfam have been used extensively in the annotation of new genomes and transcriptomes. We traced full Pfam domains in the *L. decemlineata* transcriptome using the CLC Genomics Workbench 6.05 [75]. We found a total of 8,927 transcripts associated with all eukaryotic protein pfam domains with E-values below $1e^{-3}$. The zinc finger type C2H2 topped these top domains with a total of 878 hits (**Table 4**). Zinc finger domains are relatively small protein motifs which contain multiple finger-like protrusions that make tandem contacts with their target molecule such as DNA, RNA, protein or lipid and they regulate gene expression in different eukaryotes during various processes, such as photoreceptor cell specification and differentiation [76,77] Eye development [78] and larval-pupal metamorphosis [79] in beetles are governed by these zinc finger proteins. We detected 850 WD40 domains, involved in signal transduction, transcription regulation, cell cycle control, autophagy and apoptosis [80].

We found two tetratricopeptide-repeat-carrying domains, TPR_1 and TPT_2, in 419 and 516 transcripts, respectively. These motifs are involved in protein—protein interactions [81,82].

We also found 145 cytochrome P450 domains in *L. decemlineata* transcriptomic sequences at E-values below $1e^{-3}$. Cytochrome P450s play an important role in the metabolism of xenobiotics and there is a known correlation between induced levels of these P450 genes and resistance to synthetic insecticides [83].

Gag proteins mediate the telomer-specific transposition of retrotransposons for telomer maintenance in *Drosophila* [84,85] and a similar role of gag proteins was expected in *L. decemlineata*. We found 108 domains associated with reverse-transcriptase activity, which argues for transposable activities in this beetle. Furthermore, we found 159 transcripts coupled with the trypsin domain, a classical serine protease domain.

Molecular markers

We identified 29,205 putative single nucleotide polymorphisms (SNPs) in 8,181 sequences (**Table 5** and **Table S9**). This number is likely an underestimate of the number of SNPs in this species as we collected samples only in Europe where *L. decemlineata* was subject to a founder event during its invasion of the Eurasian continent with a substantial reduction in genetic diversity [10]. Among these SNPs, 18.882 were transitions (Ts) and 10,383 were transversions (Tv) with a Ts to Tv ratio of 1.8:1. We also predicted 17,284 single sequence repeats (SSRs or microsatellites). The majority of these microsatellites were di-nucleotide repeats (n = 13,473) while 3,294 were tri-nucleotide, 416 tetra-nucleotide, and 101 were penta-nucleotide repeats (**Table 6** and **Table S10**). These molecular markers, still to be verified by primer design and PCR amplification, could establish a platform for the research community to study the ecology, biology and genetics of *L. decemlineata* [86]. These markers, being EST-linked, may also provide a suitable tool to identify footprints of selective pressures due to natural or anthropogenic stress.

Conclusions

We have established a new genetic resource for *L. decemlineata*, a species of high importance in the field of invasion biology. The major results of this study are: (1) a general annotation of *L. decemlineata* expressed genes; (2) the identification of a significant number of enzymatic pathways from these transcripts; (3) a catalog

Table 3. Summary of genes that induce insecticide resistance in *L. decemlineata*.

Gene	Contigs in Ld_Assembly	Adult	Larva	FL only	ML only
Catalases (CAT)	13	0	9	5	1
Glutathione peroxidases (GPX)	22	3	8	3	2
Glutathione S-transferases (GSTs)	45	2	11	4	4
ADAM metalloprotease (ADAM)	3	1	0	0	0
Cadherin (Cadh)	10	1	5	1	4
Calmodulin (CalM)	98	13	30	8	14
Cytochrome P450 monooxygenases (CYPs)	221	24	77	20	44
Esterase	194	17	111	5	89
Superoxide dismutases (SOD)	28	4	4	2	2

Genes were scanned by BLAST [59] at E-value lower than $1e^{-3}$. The table also shows the number of contigs identified as library specific for each gene.
FL – full larva; ML – midgut-larva.
doi:10.1371/journal.pone.0086012.t003

Table 4. Summary of top Pfam protein domains in *L. decemlineata* transcriptome.

Accession id	Pfam domain name	Domain description	# Occurrence
PF00096	zf-C2H2	Zinc finger C2H2 type	878
PF00400	WD40	WD domain G-beta repeat	850
PF00560	LRR_1	Leucine Rich Repeat	556
PF07719	TPR_2	Tetratricopeptide repeat	516
PF00023	Ank	Ankyrin repeat	465
PF00515	TPR_1	Tetratricopeptide repeat	419
PF00036	efhand	EF hand	412
PF00069	Pkinase	Protein kinase domain	324
PF02985	HEAT	HEAT repeat	279
PF00076	RRM_1	RNA recognition motif. (a.k.a. RRM RBD or RNP domain)	267
PF07714	Pkinase_Tyr	Protein tyrosine kinase	251
PF07679	I-set	Immunoglobulin I-set domain	220
PF00112	Peptidase_C1	Papain family cysteine protease	177
PF00435	Spectrin	Spectrin repeat	175
PF00005	ABC_tran	ABC transporter	169
PF00379	Chitin_bind_4	Insect cuticle protein	164
PF01607	CBM_14	Chitin binding Peritrophin-A domain	157
PF08246	Inhibitor_I29	Cathepsin propeptide inhibitor domain (I29)	154
PF00135	COesterase	Carboxylesterase	148
PF00067	p450	Cytochrome P450	145
PF00153	Mito_carr	Mitochondrial carrier protein	143
PF00047	ig	Immunoglobulin domain	140
PF00232	Glyco_hydro_1	Glycosyl hydrolase family 1	134
PF00514	Arm	Armadillo/beta-catenin-like repeat	128
PF00106	adh_short	short chain dehydrogenase	126
PF00271	Helicase_C	Helicase conserved C-terminal domain	124
PF00004	AAA	ATPase family associated with various cellular activities (AAA)	114
PF00018	SH3_1	SH3 domain	110
PF00071	Ras	Ras family	108
PF07690	MFS_1	Major Facilitator Superfamily	104

Pfam protein domains were scanned at E-value$<$1e^{-3}.

of putative SNPs and microsatellite markers which, upon validation, could facilitate the identification of polymorphisms within and between *L. decemlineata* populations; and (4) the characterization of genes of interest: those involved in diapause, detoxification and insecticide resistance. *L. decemlineata* is an important pest beetle species and is commonly used for studying plant-herbivore interactions and resistance to insecticides. A transcriptome assembly is, therefore of great importance to this community. The new genetic resource and putative miRNA candidates established by our study provide new insights into the biology of *L. decemlineata*.

Materials and Methods

We have summarized our entire *L. decemlineata* transcriptome analysis approach in **Figure 7**.

Beetle samples within Europe

Beetles were collected from four field populations, two in north-eastern Italy (Camposampiero 45 33 39.72 – 11 56 52.08 and Montello 45 47 49.10 – 12 07 17.85) and two in Russia (one near Petroskoy 61 48 28.81 – 34 7 51 and one near Ufa 54 47 05.26 – 55 57 57.69). No official permits were required to collect the beetles. Permission was granted by land owners to access the fields, which were not in protected areas. No endangered or protected species were involved in the project. Permits were obtained to take the beetles into Finland (Evira DNr: 4140/0614/2008). Beetles from Russia and Camposampiero were reared in the Finnish laboratory (Evira permit DNr: 3861/541/2007) on Van Gogh potato plants and at a constant 23°C. Beetles from Montello were reared in the laboratory in Padova (no permits are required to rear beetles in Italy) on the Monalisa potato and at a constant 24°C. Beetles were mated in the laboratory and larvae were grown on potato plants until adults emerged. Half of the samples from St. Petersburg and Camposampiero were grown under short-day length conditions (12L:12D) to induce diapause and the other half under long-day length (18L:6D). Twenty-three adults, 12 females and 11 males (6 individuals per population except for Montello, 5 individuals), were collected at intervals – from adult emergence to diapause – and preserved either at -80°C or in RNA later at −
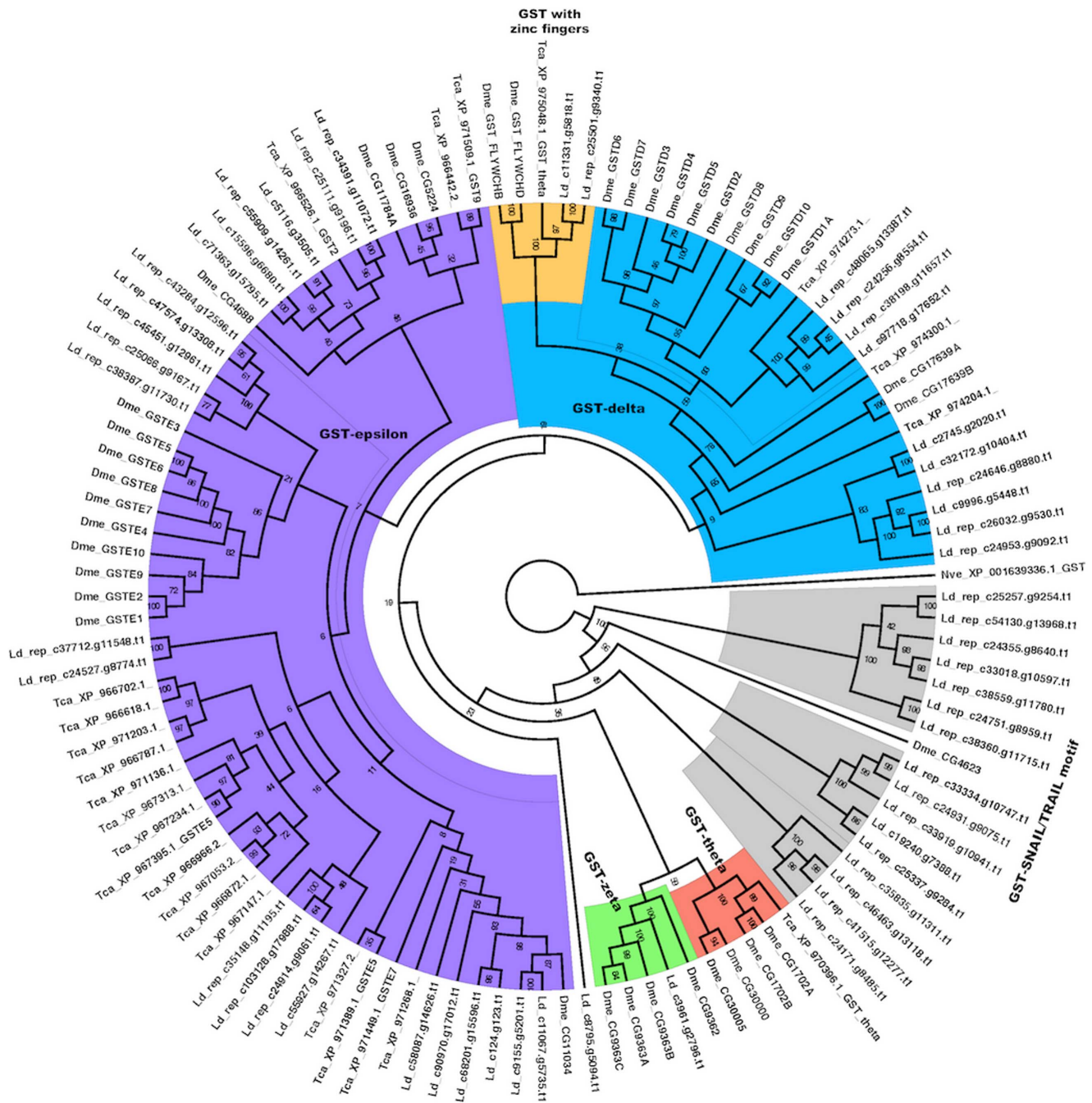
Figure 5. Bayesian phylogeny of GSTs from *T. castaneum*, and *D. melanogaster* genomic and *L. decemlineata* transcriptomic sequences. The tree was based on the WAG +G+I protein model and run for 9,248,000 generations. Posterior probability values are marked at the nodes. Sequences are named by a prefix of three letters indicating the species (Dme – *D. melanogaster*, Tca – *T. castaneum*, Lde – *L. decemlineata*) followed by either the accession ids and name in NCBI (for Dme and Tca) or contig name (for Ld). A GST from *N. vectensis* (Nve) was used as outgroup.
doi:10.1371/journal.pone.0086012.g005

20°C. A total of 142 larvae were also collected at different instar stages (1 to 4) from the four populations and preserved either at − 80°C or in RNA later at -20°C. Ten of these larvae (first instar) from Montello were exposed to potato leaf dipped in a solution of 3 mg/ml of *Bacillus thuringiensis* var *tenebrionis* strain NB 176 sierotype H 8a8b crystal proteins (Novodor FC; Serbios).

RNA isolation and cDNA library construction

RNA was extracted from 23 adults and 142 larvae of *L. decemlineata* using the RNeasy Mini Kit (Qiagen) following the manufacturer's protocol for animal tissues. The RNA was extracted from 20 mg of tissue from the head, thorax or abdomen of adult beetles, from the head of 3rd and 4th instar larvae and from pools of entire 1st and 2nd instar larvae. Extracted RNA was checked for integrity and size and then quantified using a Quant-IT RNA BR assay kit (Invitrogen). The single extracts were diluted
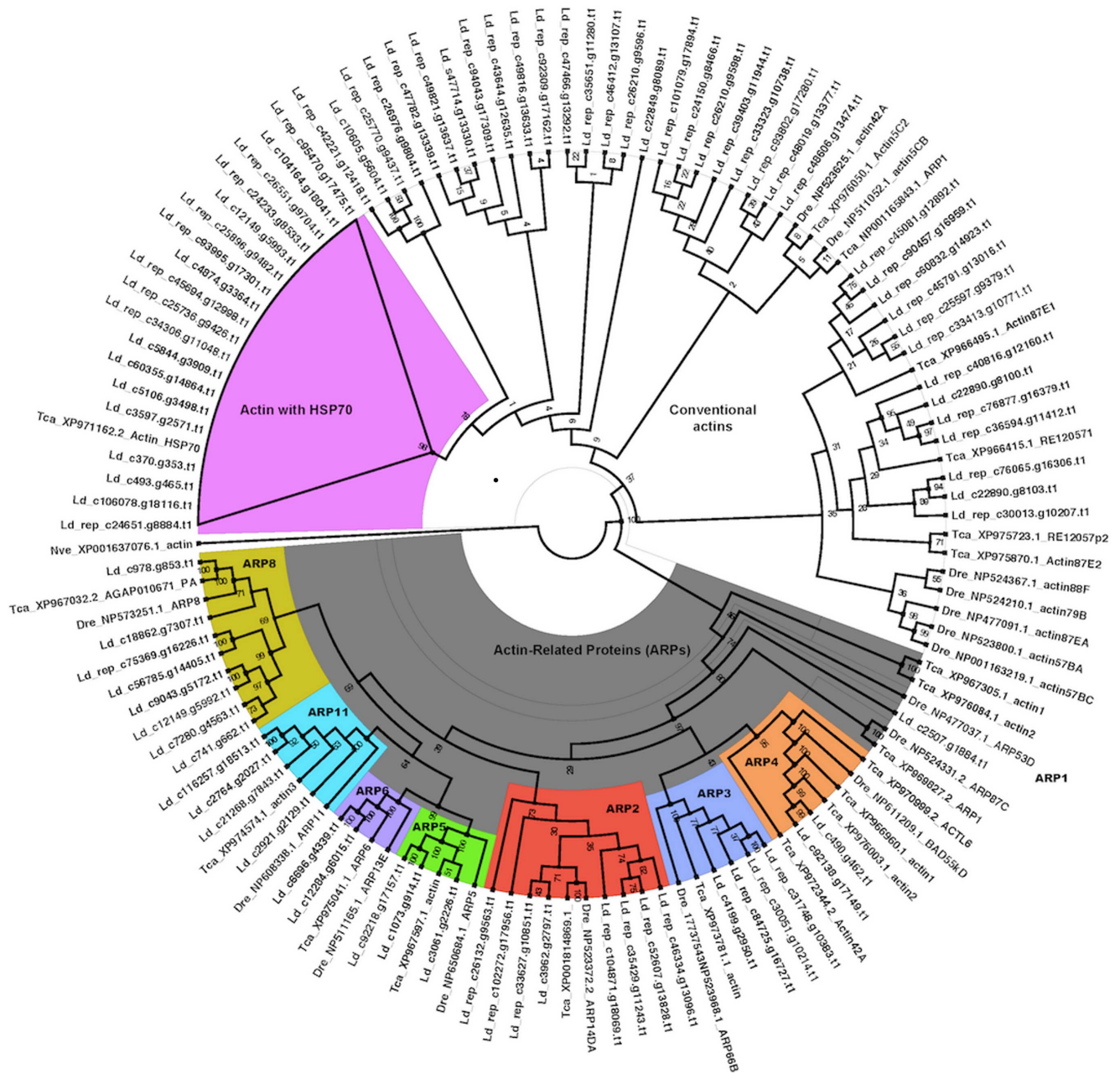
Figure 6. Bayesian phylogeny of actin and actin-related protein (ARP) from *T. castaneum*, and *D. melanogaster* genomic and *L. decemlineata* transcriptome sequences. The tree was based on the BLOSUM+G protein model and run for 10,927,000 generations. Posterior probability values are marked at the nodes. Sequences are named by a prefix of three letters indicating the species (Dme – *D. melanogaster*, Tca – *T. castaneum*, Ld – *L. decemlineata*) followed by either the accession ids and name in NCBI (for Dme and Tca) or contig name (for Ld). An actin from *N. vectensis* (Nve) was used as outgroup.
doi:10.1371/journal.pone.0086012.g006

to obtain a concentration of 100 ng/µl. Extraction was conducted in two different laboratories, and two pools of equal amounts of RNA were obtained, one for Russia and Camposampiero and one for Montello, for both adults and larvae. Each pool, consisting of a total of 5 µg of RNA, was stored in pure ethanol and shipped to Evrogen Labs Ltd., Moscow. The two adult and the two larval RNA pools were used for ds cDNA synthesis using the SMART approach [87]. SMART-prepared amplified cDNAs were pooled (ratio: ¾ of the Russia/Camposampiero pool and ¼ of the Montello pool) and then normalized using the DSN normalization

method [88] to reduce overabundant transcripts. Normalization included cDNA denaturation/reassociation, a treatment by duplex-specific nuclease (DSN [89]) and the amplification of the normalized fraction by PCR.

454 pyrosequencing

Approximately 20 µg of normalized cDNA were used for the sequencing libraries construction (one for adults and one for larvae) at BMR Genomics, Padova, Italy, according to the described protocol [90]. The sequencing was performed on a half

Table 5. Summary of putative SNPs in *L. decemlineata* transcriptomic sequences.

SNP types	Number
Transition	
A-G	9,515
C-T	9,306
Transversion	
A-C	2,394
A-T	4,049
C-G	1,567
G-T	2,368
T-R	5
G-R	1
Total	29,205

R – purine (G or A).
doi:10.1371/journal.pone.0086012.t005

plate for each data set in a 454 GS-FLX titanium series pyrosequencer (Roche Applied Science).

Preprocessing of 454 sequence reads

The raw reads from the two libraries were extracted from 454 SFF pyrograms through the open source alternative sff_extract 0.2.10. Sequence and qualities were tagged for the library of origin. We preprocessed the raw 454 sequence reads using the est_process module of the est2assembly 1.13 package [91], which performs sequencing adaptor removal, low complexity region masking, quality trimming, poly A/T detection and removal. We further cleaned the data sets from contaminants such as bacterial, viral, 18S RNA and *Solanum tuberosum* sequences using Deconseq [92] with the parameters: coverage ≥90% and identity ≥94%. These preprocessing steps with est2assembly and Deconseq were also performed on the publicly available midgut larval data set.

Assembly of 454 sequence reads

We assembled the transcriptomic sequence reads from the three data sets (adult, full larval [FL] and midgut larval [ML]) with one assembly round using the MIRA 3.2 assembler [39] on "EST" and "accurate" usage mode. Settings adopted for this *de novo* assembly round were those defined by the 454 pyrosequencing technology (mira -job = denovo, est, accurate, 454 –notraceinfo). A summary of parameters and quality of this assembly is provided in **File S1**.

Annotation of the transcriptomic dataset using homology searches

We annotated *L. decemlineata* transcriptome sequences by similarity search using BLASTX [59]. We used batch BLAST similarity searches for the entire transcriptome locally conducted against the non-redundant (nr) peptide database (downloaded in October 2010, including all non-redundant GenBank CDS translations + PDB + SwissProt + PIR+ PRF) with E-values< $1e^{-3}$. We utilized the Blast2GO suite [41] for functional annotation of transcripts, applying the function to map GO terms to transcripts with BLAST hits generated from BLAST searches against nr. Only ontologies obtained from hits of E-values< $1e^{-3}$, annotation cut-offs >25, and a GO weight >1 were used for the annotation.

Table 6. Summary of microsatellite loci predicted in the *L. decemlineata* transcriptome.

Number of repeats	Nucleotides in repeats			
	Di-	Tri-	Tetra-	Penta-
4	11,603	2,368	306	54
5	1,264	577	70	14
6	329	220	23	8
7	144	72	10	2
8	64	31	1	1
9	33	11	0	5
10	12	9	1	2
11	5	3	0	1
12	7	0	3	0
13	5	0	0	2
14	2	1	1	1
15	0	0	0	2
16	1	0	0	1
17	1	0	0	0
18	0	0	0	1
19	2	0	0	2
20	0	0	0	0
21	0	0	0	1
22	0	0	0	0
23	0	0	1	0
24	0	0	0	1
25	0	0	0	0
26	0	0	0	0
27	0	0	0	0
28	0	0	0	1
30	0	1	0	1
31	0	0	0	0
32	0	0	0	0
33	0	1	0	0
36	0	0	0	0
38	0	0	0	0
41	0	0	0	1
141	1	0	0	0
Total	**13,473**	**3,294**	**416**	**101**	**17,284**

doi:10.1371/journal.pone.0086012.t006

The taxonomic classifications of annotated contigs were computed using MEGAN 4 [93] based on the absolute best BLAST hits after excluding contigs with multiple best BLAST records.

Identification of stage specific transcripts

In order to identify transcripts that were uniquely expressed in either the larval or the adult stage, we used the same approach described by Vidotto *et al.* [94] to identify library specific contigs for transcriptome completeness estimation. First combined contigs were classified as being "adult", "larva" or "common" on the bases of the reads composition. Larval contigs were further divided into FL and ML contigs, if originated from full larval and midgut
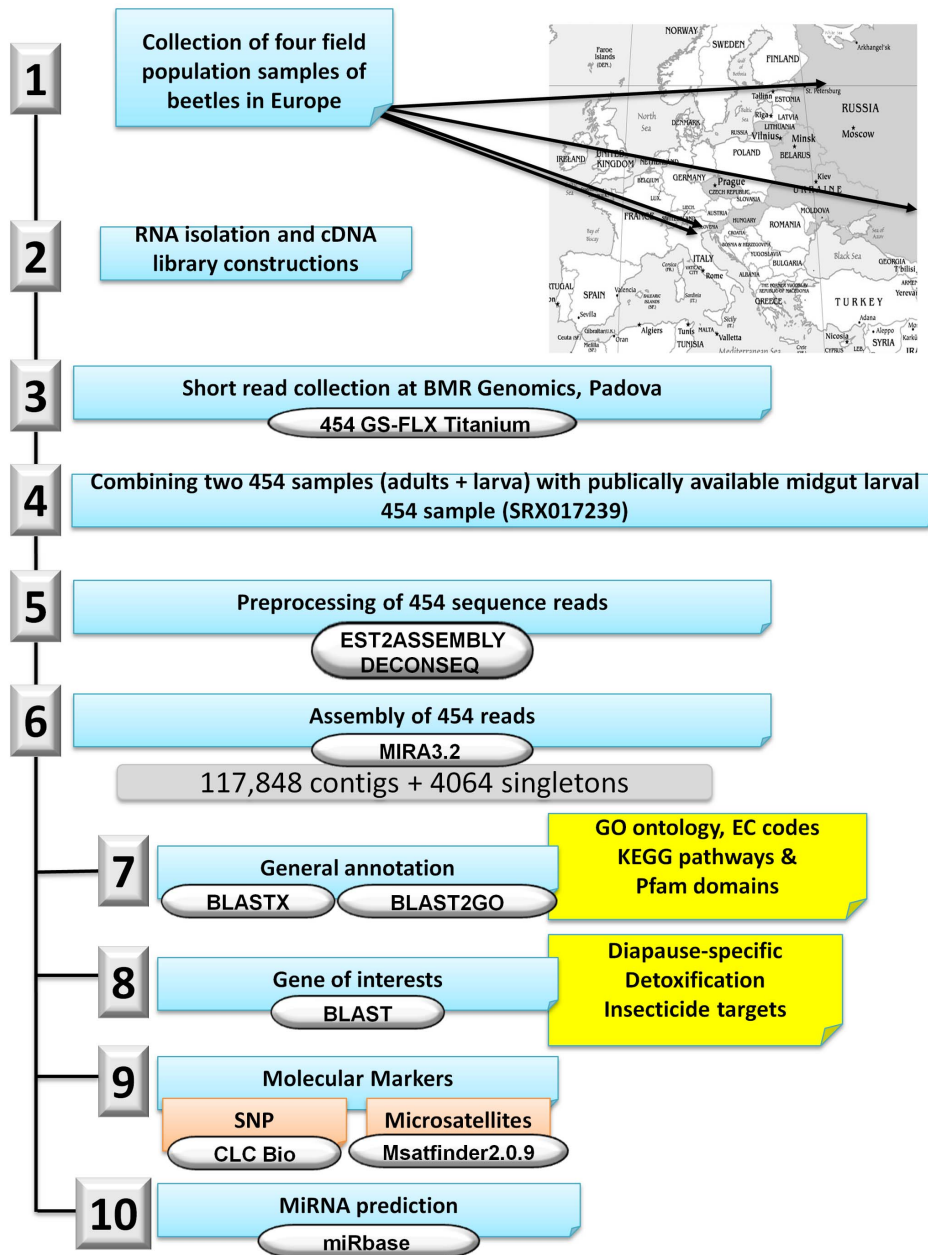
Figure 7. Overall approach to *L. decemlineata* 454 transcriptome analysis. Analysis steps are numbered and the software and tools used in each step are shown.
doi:10.1371/journal.pone.0086012.g007

larval datasets, respectively. Finally the common contigs were the fractions of those composed by reads from both libraries and then represented by transcripts considered to be expressed in the two developmental stages. We performed a bidirectional BLASTN [59] between the libraries specific contigs and those that align for more than 80% of their length, with e-values above $1e^{-50}$ were considered not library specific and then moved into the common fraction. The indirect subtraction was also performed to take into account contigs representing non overlapping fragments of the same transcript, which had not been assembled together. Library-specific contigs were searched for similarities, against the complete cDNA set of *T. castaneum*, stored in Ensembl Metazoa release-17 (Feb 2013), using TBLASTX [59]. The library-specific contigs with match against the same subjects, with e-values grater then $1e^{-6}$ and >80% query coverage, were considered belonging to the common fraction.

Detection of molecular markers

Microsatellites were identified using Msatfinder version 2.0.9 software [95]. SNPs were predicted using the CLC Genomics *Workbench* 6 [75] with a criterion of at least 4 reads supporting either the consensus or variant within a minimum of 10 reads.

miRNA detection

We scanned assembled sequences against all known miRNA sequences from miRbase Release 20 (June 2013) using BLAST suite [59] with E-values<$1e^{-3}$.

Search for genes of interest and phylogenetic analyses

We traced all genes of interest in the transcriptomic sequences using the BLAST suite [59] with E-values$<$1e^{-3}. We aligned protein sequences from *D. melanogaster*, *T. castaneum* and those recovered from the *L. decemlineata* transcriptome for the specific protein superfamily under consideration using the MUSCLE alignment tool [96] with default settings. All phylogenetic trees were constructed using a Bayesian approach with 2 runs, a number of generations and protein model as specified in figures 4–6, and 25% burn-in-period in MrBayes 3.2 [97]. The best protein model for each gene was estimated using TOPALi V2.5 [98].

Pfam Domain detection

We searched for all eukaryotic protein domains in the *L. decemlineata* transcriptome using the Pfam Domain Search function (which uses the HMMER software [99]) under the protein analysis module in the CLC Genomics Workbench 6.05 [70] with E-values$<$1e^{-3}.

Supporting Information

Figure S1 Length distributions of *L. decemlineata* transcriptomic sequences.
(PPTX)

Table S1 List of contigs with nucleotide frequencies, length and percentage of GC content from transcriptomic sequences of *L. decemlineata*. Contigs are named with suffix as per annotations provided by MIRA 3.2 assembler [39] as follow: c – "normal" contig; rep - these are contigs containing only repetitive areas; and s – singletons (contigs made up of a single read).
(XLS)

Table S2 Gene Ontology (GO) annotation results of *L. decemlineata* transcriptomic sequences. A. Biological Process. **B**. Molecular Function. **C**. cellular component.
(XLS)

Table S3 Complete list of GO terms in the three BLAST2GO categories with enzyme code assignment.
(XLSX)

Table S4 List of KEGG pathways encompassing *L. decemlineata* sequences.
(XLS)

Table S5 List of putative miRNA genes from *L. decemlineata* 454 transcriptomic sequences.
(XLS)

Table S6 List of *L. decemlineata* best hits with previously known diapause-specific genes obtained via suppressive subtractive hybridization.
(XLSX)

Table S7 List of putative genes involved in diapause in *D. montana* and their homologs in *L. decemlineata* transcriptome. The table also shows the number of contigs identified as library specific for each gene. FL – full larva; ML – midgut-larva Red - genes not found in *L. decemlineata* assembly.
(XLSX)

Table S8 List of transcripts with hits on genes known to be involved in resistance to Bt-Cry proteins. The table also shows the number of contigs identified as library specific for each gene. FL – full larva; ML – midgut-larva
(XLSX)

Table S9 Putative SNPs in *L. decemlineata* transcriptomic sequences.
(XLS)

Table S10 Putative microsatellite loci in *L. decemlineata* transcriptomic sequences. (A) Contigs wise microsatelites. (B) Repeat classes.
(XLSX)

File S1 A summary of parameters and quality of *L. decemlineata* transcriptome assembly.
(DOC)

Acknowledgments

Author Contributions

Conceived and designed the experiments: AG LL SP LC. Performed the experiments: AG SP LL. Analyzed the data: AK AG MV LC. Contributed reagents/materials/analysis tools: AG LL. Wrote the paper: AK AG LL SP LC MV.

References

1. Alyokhin A, Baker M, Mota-Sanchez D, Dively G, Grafius E (2008) Colorado potato beetle resistance to insecticides. Am J Pot Res 85: 395–413.
2. Casagrande RA (1987) The Colorado potato beetle: 125 years of mismanagement. Bulletin of the ESA 33: 142–150.
3. Gauthier NL, Hofmaster R, Semel M (1981) History of Colorado potato beetle control. In: Lashomb JH, Casagrande RA, editors. Advances in Potato Pest Management Stroudsburg, PA.: Hutchinson Ross Publishing Company.
4. Weber DC, Ferro DN (1994) Colorado potato beetle: diverse life history poses challenge to management. In: Zehnder GW, Powelson ML, Jansson RK, Raman KV, editors. Advances in potato pest biology and management.St. Paul, Minnesota.: APS Press. pp. 256–259.
5. Smith IM, Charles LMF (1998) Distribution maps of quarantine pests for Europe: CABI.
6. Hare JD (1990) Ecology and management of the Colorado potato beetle. Annu Rev Entomol 35: 81–100.
7. Tower WL (1906) An investigation of evolution in chrysomelid beetles of the genus *Leptinotarsa*. Stroudsburg, PA.: Carnegie Institution of Washington.
8. Alyokhin A, Chen YH, Udalov M, Benkovskaya G, Lindström L (2013) Chapter 19 - Evolutionary considerations in potato pest management. Insect Pests of Potato. San Diego: Academic Press.pp. 543–571.
9. Weber D (2003) Colorado beetle: pest on the move. Pestic Outlook 14: 256–259.
10. Boman S, Grapputo A, Lindstrom L, Lyytinen A, Mappes J (2008) Quantitative genetic approach for assessing invasiveness: geographic and genetic variation in life-history traits. Biol Inv 10: 1135–1145.
11. EPPO (2003) Distribution maps of quarantine pests of Europe, *Leptinotarsa decemlineta*. European and Mediterranean plant protection organization (http://www.eppo.org).
12. Grapputo A, Boman S, Lindstrom L, Lyytinen A, Mappes J (2005) The voyage of an invasive species across continents: genetic diversity of North American and European Colorado potato beetle populations. Mol Ecol 14: 4207–4219.
13. Forgash AG (1985) Insecticide resistance in the Colorado potato beetle. In: Ferro DN, Voss RH, editors. Proceedings of the Symposium on the Colorado potato beetle: XVII International Congress of Entomology.pp. 256–259.
14. de Wilde J, Hsiao TH (1981) Geographic diversity of the Colorado potato beetle and its infestation in Eurasia. In: Lashomb JH, Casagrande RA, editors. Advances in Potato Pest Management Stroudsburg, PA.: Hutchinson Ross Publishing Company.
15. Tauber MJ, Tauber CA, Obrycki JJ, Gollands B, Wright RJ (1988) Voltinism and the induction of estival diapause in the Colorado potato beetle, *Leptinotarsa decemlineata* (Coleoptera, Chrysomelidae). Ann Entomol Soc Am 81: 748–754.
16. Kort CAD (1990) Thirty-five years of diapause research with the Colorado potato beetle. Entomol Exp Appl 56: 1–13.

17. Noronha C, Cloutier C (2006) Effects of potato foliage age and temperature regime on prediapause Colorado potato beetle *Leptinotarsa decemlineata* (Coleoptera: Chrysomelidae). Environ Entomol 35: 590–599.

18. Yocum GD, Rinehart JP, Larson ML (2009) Down regulation of gene expression between the diapause initiation and maintenance phases of the Colorado potato beetle, *Leptinotarsa decemlineata* (Coleoptera: Chrysomelidae). Eur J Entomol 106: 471–476.

19. Baker MB, Porter AH (2008) Use of sperm precedence to infer the overwintering cost of insecticide resistance in the Colorado potato beetle. Agr Forest Entomol 10: 181–187.

20. Boivin T, Bouvier JC, Beslay D, Sauphanor B (2004) Variability in diapause propensity within populations of a temperate insect species: interactions between insecticide resistance genes and photoperiodism. Biol J Linn Soc 83: 341–351.

21. Carriere Y, Roff DA, Deland JP (1995) The joint evolution of diapause and insecticide resistance - a test of an optimality model. Ecology 76: 1497–1505.

22. Grewal PS, Power KT, Shetlar DJ (2001) Neonicotinoid insecticides alter diapause behavior and survival of overwintering white grubs (Coleoptera: Scarabaeidae). Pest Manag Sci 57: 852–857.

23. Tauber MJ, Tauber CA (1973) Quantitative response to daylength during diapause in insects. Nature 244: 296–297.

24. Tauber MJ, Tauber CA, Masaki S (1986) Seasonal adaptations of insects: Oxford University Press.

25. Denlinger DL (2008) Why study diapause? Entomol Res 38: 1–9.

26. Whalon ME, Miller DL, Hollingworth RM, Grafius EJ, Miller JC (1993) Selection of a Colorado potato beetle (Coleoptera, Chrysomelidae) strain resistant to *Bacillus thuringiensis*. J Econ Entomol 86: 226–233.

27. Rahardja U, Whalon ME (1995) Inheritance of resistance to *Bacillus thuringiensis* subsp. tenebrionis CryIIIA delta-endotoxin in Colorado potato beetle (Coleoptera: Chrysomelidae). J Econ Entomol 8: 21–26.

28. Alyokhin A (2009) Colorado potato beetle management on potatoes: current challenges and future prospects. In: Tennant P, Benkeblia N, editors. Potato II Fruit, Vegetable and Cereal Science and Biotechnology 3: Global Science Books Ltd.pp. 10–19.

29. Evans JD, Brown SJ, Consortium iK (2013) The i5K Initiative: advancing arthropod genomics for knowledge, human health, agriculture, and the environment. J Hered 104: 595–600.

30. Petek M, Turnsek N, Gasparic MB, Novak MP, Gruden K, et al. (2012) A complex of genes involved in adaptation of *Leptinotarsa decemlineata* larvae to induced potato defense. Arch Insect Biochem Physiol 79: 153–181.

31. Yocum GD, Rinehart JP, Chirumamilla-Chapara A, Larson ML (2009) Characterization of gene expression patterns during the initiation and maintenance phases of diapause in the Colorado potato beetle, *Leptinotarsa decemlineata*. J Insect Physiol 55: 32–39.

32. Pauchet Y, Wilkinson P, Chauhan R, Ffrench-Constant RH (2010) Diversity of beetle genes encoding novel plant cell wall degrading enzymes. PLoS One 5: e15635.

33. Choi J-H, Kijimoto T, Snell-Rood E, Tae H, Yang Y, et al. (2010) Gene discovery in the horned beetle *Onthophagus taurus*. BMC Genomics 11: 703.

34. 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.

35. Vera JC, Wheat CW, Fescemyer HW, Frilander MJ, Crawford DL, et al. (2008) Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. Mol Ecol 17: 1636–1647.

36. Morozova O, Hirst M, Marra MA (2009) Applications of new sequencing technologies for transcriptome analysis. Annu Rev Genomics Hum Genet 10: 135–151.

37. Schuster SC (2008) Next-generation sequencing transforms today's biology. Nat Meth 5: 16–18.

38. 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.

39. Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Muller WE, et al. (2004) Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome Res 14: 1147–1159.

40. 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.

41. 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.

42. Conesa A, Gotz S (2008) Blast2GO: A comprehensive suite for functional analysis in plant genomics. Int J Plant Genomics 2008: 619832.

43. Bai X, Mamidala P, Rajarapu SP, Jones SC, Mittapalli O (2011) Transcriptomics of the bed bug (*Cimex lectularius*). PLoS One 6: e16336.

44. Bai X, Rivera-Vega L, Mamidala P, Bonello P, Herms DA, et al. (2011) Transcriptomic signatures of ash (*Fraxinus* spp.) phloem. PLoS One 6: e16368.

45. Bai X, Zhang W, Orantes L, Jun TH, 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.

46. 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 (*Agrilus planipennis*). PLoS One 5: e13708.

47. Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. Cell 136: 215–233.

48. He L, Hannon GJ (2004) MicroRNAs: small RNAs with a big role in gene regulation. Nat Rev Genet 5: 522–531.

49. Kozomara A, Griffiths-Jones S (2011) miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res 39: D152–157.

50. Kumar A, Ragg H (2008) Ancestry and evolution of a secretory pathway serpin. BMC Evol Biol 8: 250.

51. Silverman G, Bird P, Carrell R, Church F, Coughlin P, et al. (2001) The serpins are an expanding superfamily of structurally similar but functionally diverse proteins. Evolution, mechanism of inhibition, novel functions, and a revised nomenclature. J Biol Chem 276: 33293–33296.

52. Reichhart JM (2005) Tip of another iceberg: *Drosophila* serpins. Trends Cell Biol 15: 659–665.

53. Zou Z, Evans JD, Lu Z, Zhao P, Williams M, et al. (2007) Comparative genomic analysis of the *Tribolium* immune system. Genome Biol 8: R177.

54. Kruger O, Ladewig J, Koster K, Ragg H (2002) Widespread occurrence of serpin genes with multiple reactive centre-containing exon cassettes in insects and nematodes. Gene 293: 97–105.

55. Kankare M, Salminen T, Laiho A, Vesala L, Hoikkala A (2010) Changes in gene expression linked with adult reproductive diapause in a northern malt fly species: a candidate gene microarray study. BMC Ecol 10: 3.

56. Ioannidis PM, Grafius E, Whalon ME (1991) Patterns of insecticide resistance to azinphosmethyl, carbofuran, and permethrin in the Colorado potato beetle (Coleoptera, Chrysomelidae). J Econ Entomol 84: 1417–1423.

57. Schuler MA (2011) P450s in plant-insect interactions. Biochim Biophys Acta 1814: 36–45.

58. Zhang S, Shukle R, Mittapalli O, Zhu YC, Reese JC, et al. (2010) The gut transcriptome of a gall midge, *Mayetiola destructor*. J Insect Physiol 56: 1198–1206.

59. 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.

60. Yan L, Yang P, Jiang F, Cui N, Ma E, et al. (2012) Transcriptomic and phylogenetic analysis of *Culex pipiens quinquefasciatus* for three detoxification gene families. BMC Genomics 13: 609.

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

62. Friedman R (2011) Genomic organization of the glutathione S-transferase family in insects. Mol Phylogenet Evol 61: 924–932.

63. Che-Mendoza A, Penilla RP, Rodriguez DA (2009) Insecticide resistance and glutathione S-transferases in mosquitoes: A review. Afr J Biotechnol 8: 1386–1397.

64. Board PG, Baker RT, Chelvanayagam G, Jermiin LS (1997) Zeta, a novel class of glutathione transferases in a range of species from plants to humans. Biochem J 328 (Pt 3): 929–935.

65. Yu Q, Lu C, Li B, Fang S, Zuo W, et al. (2008) Identification, genomic organization and expression pattern of glutathione S-transferase in the silkworm, *Bombyx mori*. Insect Biochem Mol Biol 38: 1158–1164.

66. Fabrick J, Oppert C, Lorenzen MD, Morris K, Oppert B, et al. (2009) A novel *Tenebrio molitor* cadherin is a functional receptor for *Bacillus thuringiensis* Cry3Aa toxin. J Biol Chem 284: 18401–18410.

67. Loseva O, Ibrahim M, Candas M, Koller CN, Bauer LS, et al. (2002) Changes in protease activity and Cry3Aa toxin binding in the Colorado potato beetle: implications for insect resistance to *Bacillus thuringiensis* toxins. Insect Biochem Mol Biol 32: 567–577.

68. Ochoa-Campuzano C, Real MD, Martínez-Ramírez AC, Bravo A, Rausell C (2007) An ADAM metalloprotease is a Cry3Aa *Bacillus thuringiensis* toxin receptor. Biochemical and Biophysical Research Communications 362: 437–442.

69. Ochoa-Campuzano C, Sánchez J-RI, Real MD, Rausell C, Sánchez J (2012) Identification of a calmodulin-binding site within the domain i of *Bacillus thuringiensis* Cry3Aa toxin. Archives of Insect Biochemistry and Physiology 81: 53–62.

70. Park Y, Abdullah MAF, Taylor MD, Rahman K, Adang MJ (2009) Enhancement of *Bacillus thuringiensis* Cry3Aa and Cry3Bb toxicities to coleopteran larvae by a toxin-binding fragment of an insect cadherin. Applied and Environmental Microbiology 75: 3086–3092.

71. Sayed A, Wiechman B, Struewing I, Smith M, French W, et al. (2010) Isolation of transcripts from *Diabrotica virgifera virgifera* LeConte responsive to the *Bacillus thuringiensis* toxin Cry3Bb1. Insect Mol Biol 19: 381–389.

72. Blessing CA, Ugrinova GT, Goodson HV (2004) Actin and ARPs: action in the nucleus. Trends Cell Biol 14: 435–442.

73. Goodson HV, Hawse WF (2002) Molecular evolution of the actin family. J Cell Sci 115: 2619–2622.

74. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, et al. (2012) The Pfam protein families database. Nucleic Acids Res 40: D290–D301.

75. Knudsen T, Knudsen B (2013) CLC Genomics Benchwork 6. Available: http://www.clcbio.com. Accessed on 2013 Sept 20.

76. Klug A (1999) Zinc finger peptides for the regulation of gene expression. J Mol Biol 293: 215–218.

77. Matthews JM, Sunde M (2002) Zinc fingers–folds for many occasions. IUBMB Life 54: 351–355.

78. Liu Z, Friedrich M (2004) The *Tribolium* homologue of glass and the evolution of insect larval eyes. Dev Biol 269: 36–54.

79. Parthasarathy R, Tan A, Bai H, Palli SR (2008) Transcription factor broad suppresses precocious development of adult structures during larval-pupal metamorphosis in the red flour beetle, *Tribolium castaneum*. Mech Dev 125: 299–313.

80. Stirnimann CU, Petsalaki E, Russell RB, Muller CW (2010) WD40 proteins propel cellular networks. Trends Biochem Sci 35: 565–574.

81. Blatch GL, Lassle M (1999) The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. Bioessays 21: 932–939.

82. Smith DF (2004) Tetratricopeptide repeat cochaperones in steroid receptor complexes. Cell Stress Chaperones 9: 109–121.

83. Feyereisen R (2006) Evolution of insect P450. Biochem Soc Trans 34: 1252–1255.

84. Rashkova S, Karam SE, Kellum R, Pardue ML (2002) Gag proteins of the two *Drosophila* telomeric retrotransposons are targeted to chromosome ends. J Cell Biol 159: 397–402.

85. Rashkova S, Karam SE, Pardue ML (2002) Element-specific localization of *Drosophila* retrotransposon Gag proteins occurs in both nucleus and cytoplasm. Proc Natl Acad Sci U S A 99: 3621–3626.

86. Behura SK (2006) Molecular marker systems in insects: current trends and future avenues. Mol Ecol 15: 3087–3113.

87. Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PD (2001) Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. Biotechniques 30: 892–897.

88. Zhulidov PA, Bogdanova EA, Shcheglov AS, Vagner LL, Khaspekov GL, et al. (2004) Simple cDNA normalization using kamchatka crab duplex-specific nuclease. Nucleic Acids Res 32: e37.

89. Shagin DA, Rebrikov DV, Kozhemyako VB, Altshuler IM, Shcheglov AS, et al. (2002) A novel method for SNP detection using a new duplex-specific nuclease from crab hepatopancreas. Genome Res 12: 1935–1942.

90. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437: 376–380.

91. Papanicolaou A, Stierli R, Ffrench-Constant RH, Heckel DG (2009) Next generation transcriptomes for next generation genomes using est2assembly. BMC Bioinformatics 10: 447.

92. Schmieder R, Edwards R (2011) Fast identification and removal of sequence contamination from genomic and metagenomic datasets. PloS one 6.

93. Huson DH, Mitra S, Ruscheweyh HJ, Weber N, Schuster SC (2011) Integrative analysis of environmental sequences using MEGAN4. Genome Res 21: 1552–1560.

94. Vidotto M, Grapputo A, Boscari E, Barbisan F, Coppe A, et al. (2013) Transcriptome sequencing and de novo annotation of the critically endangered *Adriatic sturgeon*. BMC Genomics 14: 407.

95. Thurston MI, Field D (2005) Msatfinder: detection and characterisation of microsatellites. Available: http://www.genomics.ceh.ac.uk/msatfinder/. Accessed on 2011 April 20.

96. Marden JH, Fescemyer HW, Saastamoinen M, MacFarland SP, Vera JC, et al. (2008) Weight and nutrition affect pre-mRNA splicing of a muscle gene associated with performance, energetics and life history. Journal of Experimental Biology 211: 3653–3660.

97. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572–1574.

98. Milne I, Lindner D, Bayer M, Husmeier D, McGuire G, et al. (2009) TOPALi v2: a rich graphical interface for evolutionary analyses of multiple alignments on HPC clusters and multi-core desktops. Bioinformatics 25: 126–127.

99. Finn RD, Clements J, Eddy SR (2011) HMMER web server: interactive sequence similarity searching. Nucleic Acids Res 39: W29–37.