De Novo Transcriptome Analysis of Wing Development-Related Signaling Pathways in *Locusta migratoria* Manilensis and *Ostrinia furnacalis* (Guenée)

Suning Liu[9], **Wei Wei**[9], **Yuan Chu, Long Zhang, Jie Shen***, **Chunju An***

Department of Entomology, College of Agriculture and Biotechnology, China Agricultural University, Beijing, China

Abstract

Background: Orthopteran migratory locust, *Locusta migratoria*, and lepidopteran Asian corn borer, *Ostrinia furnacalis*, are two types of insects undergoing incomplete and complete metamorphosis, respectively. Identification of candidate genes regulating wing development in these two insects would provide insights into the further study about the molecular mechanisms controlling metamorphosis development. We have sequenced the transcriptome of *O. furnacalis* larvae previously. Here we sequenced and characterized the transcriptome of *L. migratoria* wing discs with special emphasis on wing development-related signaling pathways.

Methodology/Principal Findings: Illumina Hiseq2000 was used to sequence 8.38 Gb of the transcriptome from dissected nymphal wing discs. *De novo* assembly generated 91,907 unigenes with mean length of 610 nt. All unigenes were searched against five databases including Nt, Nr, Swiss-Prot, COG, and KEGG for annotations using blastn or blastx algorithm with an cut-off E-value of 10^{-5}. A total of 23,359 (25.4%) unigenes have homologs within at least one database. Based on sequence similarity to homologs known to regulate *Drosophila melanogaster* wing development, we identified 50 and 46 potential wing development-related unigenes from *L. migratoria* and *O. furnacalis* transcriptome, respectively. The identified unigenes encode putative orthologs for nearly all components of the Hedgehog (Hh), Decapentaplegic (Dpp), Notch (N), and Wingless (Wg) signaling pathways, which are essential for growth and pattern formation during wing development. We investigated the expression profiles of the component genes involved in these signaling pathways in forewings and hind wings of *L. migratoria* and *O. furnacalis*. The results revealed the tested genes had different expression patterns in two insects.

Conclusions/Significance: This study provides the comprehensive sequence resource of the wing development-related signaling pathways of *L. migratoria*. The obtained data gives an insight into better understanding the molecular mechanisms involved in the wing development in *L. migratoria* and *O. furnacalis*, two insect species with different metamorphosis types.

Introduction

Insects are the only group of invertebrates that have evolved flight [1]. Their wings serve not only as organs of flight, but also may be adapted variously as protective covers [2], thermal collectors [3], gyroscopic stabilizers [4], sound producers [5], or visual cues for species recognition and sexual contact [6]. For those insects with wings during the adult stage, complete wings are not always visible throughout the life cycle. The insects undergoing incomplete metamorphosis have represented functional wings during the stage of nymph [7], while the insects going through complete metamorphosis only have wing discs inside the body in the larval stage [7]. Therefore, the comparative studies on how the insect wings are developed will be helpful to understand the insect metamorphosis. Additionally, given the high evolutionary conservation of the proteins involved in the wing development, the understanding of the molecular mechanisms involved in wing development also sheds light on organogenesis, tissue homeostasis, human disease and so on [8–11].

Insect wing development is controlled with amazing precision and complication. Current understanding of insect wing development mechanisms is mainly from the fruit fly *Drosophila melanogaster* [12–14]. The adult wing of fruit fly is derived from the wing imaginal disc, formed at the end of embryonic
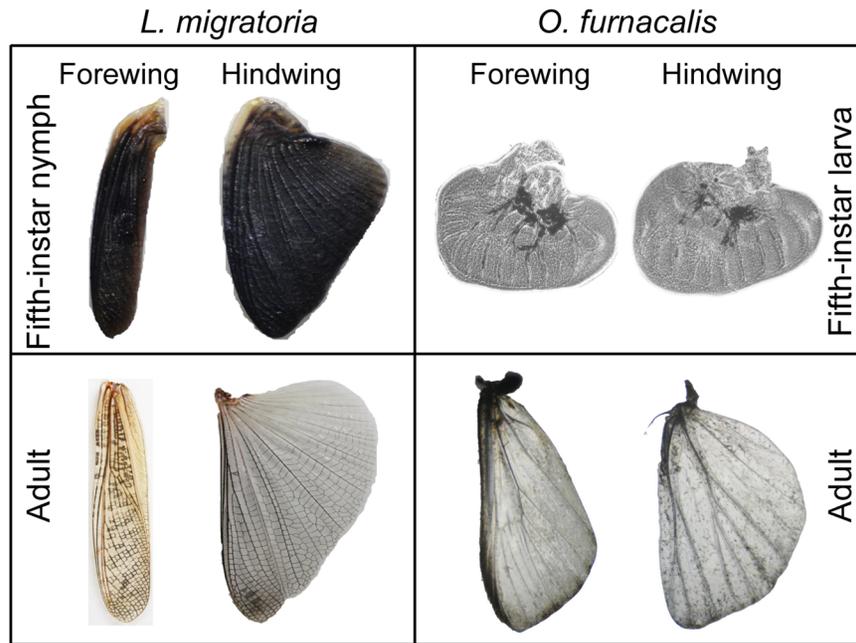
Figure 1. Observation and comparison of dissected wing discs and adult wings from *L. migratoria* and *O. furnacalis*. Note that *O. furnacalis* adult wings were manually removed scales.
doi:10.1371/journal.pone.0106770.g001

development [15]. In *Drosophila*, the wing imaginal disc is subdivided into anterior (A) and posterior (P) compartments at very early stage and then further subdivided into dorsal (D) and ventral (V) compartments at second instar stage [16–18]. Organizers located in the A/P and D/V boundaries coordinate the patterning of the wing disc by secreting signal molecules including the long-range morphogens Decapentaplegic (Dpp; the vertebrate homolog of which is TGFβ) and Wingless (Wg; the vertebrate homolog of which is Wnt) [11,19,20], and short-range morphogen Hedgehog (Hh) [21]. The produced morphogens form gradients to regulate the expression of target genes and control all aspects of wing development at cell level via the specific signaling pathways [19,22–27]. These signaling pathways are highly evolutionarily conserved in many different animals. Some key genes for the wing disc development have also been identified in a few limited insect species, such as *Tribolium castaneum* [28]. However, knowledge about the identification and involvement of wing development-related genes in various insects, especially in non-model insects, is still unclear and incomplete.

Orthopteran migratory locust, *Locusta migratoria* manilensis, and lepidopteran Asian corn borer, *Ostrinia furnacalis* (Guenée), are two different types of insects undergoing incomplete and complete metamorphosis, respectively [29,30]. Identification of candidate genes regulating wing development would provide insights into the further study about the molecular mechanisms controlling metamorphosis development. Traditionally, such gene identification in "non-model" insects relied on degenerate PCR, which is labor-intensive, expensive, prone to failure, and only produces incomplete fragments [31]. The introduction of novel high throughput sequencing technologies greatly facilitates the global analysis of the blueprint of development-related genes [32]. This technology has been used, for example, to characterize the wing development-related genes in the milkweed bug *Oncopeltus fasciatus* [33], the oriental fruit fly *Bactrocera dorsalis* [34], and the salt marsh beetle *Pogonus chalceus* [35] etc.

In this study, we combined the Illumina sequencing and *de novo* assembly to obtain and characterize the transcriptome of the wing discs of *L. migratoria* nymph. 91,907 unigenes were assembled and 23,359 ones were annotated to known databases. We also recharacterized the previous transcriptome of *O. furnacalis* larvae [36], with special emphasis on wing development-related genes. Overall, we identified 50 potential wing development-related unigenes from *L. migratoria* transcriptome, and 46 unigenes from *O. furnacalis* transcriptome, respectively. Additionally, we performed qRT-PCR analysis to investigate the gene expression profiles of several key wing development genes during the stage of rapid growth in *L. migratoria* and *O. furnacalis*. All these results provide valuable information for studying the molecular mechanisms involved in the insect wing development, and are useful resources for further exploring the mechanism how signaling pathways control wing development.

Materials and Methods

Insect rearing

Migratory locust *L. migratoria manilensis* was reared on fresh wheat seedlings at 28–30°C, 60% relative humidity with an 18L/6D photoperiod. Asian corn borer (*O. furnacalis* (Guenée)) larvae were reared on an artificial diet at 28°C under a relative humidity of 70–90% and a photoperiod of 18L/6D [36].

Dissection and total RNA extraction of wing discs

The wing discs of *L. migratoria* fifth instar nymph were dissected along the wing root with small scissors under microscope. Thirty pairs of dissected wing discs were combined, and total RNA samples were prepared using TRizol Reagent (TIANGEN, Beijing, China) following the manufacturer's instructions. Total RNA was dissolved in H_2O and RNA quantity was determined on a Nanodrop ND-2000 spectrophotometer (NanoDrop products, Wilmington, DE, USA). RNA integrity was checked on Agilent 2100 BioAnalyzer (Agilent Technologies, Englewood, CO, USA).
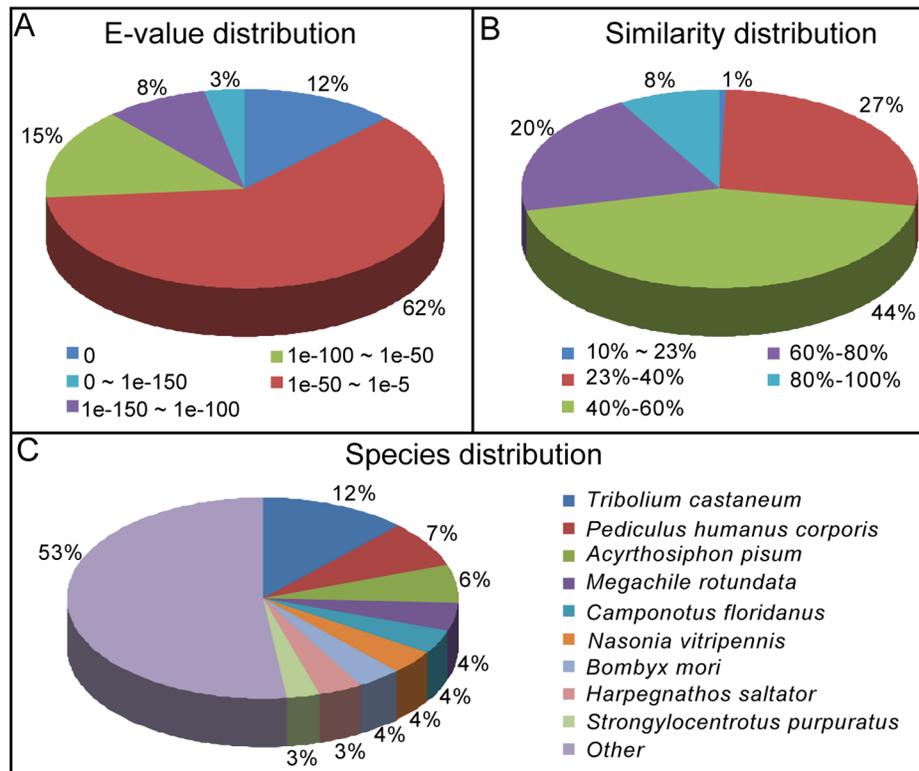
Figure 2. Homology analysis of *L. migratoria* unigenes. (A) E-value distribution. (B) Similarity distribution. (C) Species distribution. All unigenes that had BLASTX annotations within the NCBI Nr database with a cut-off E-value of 10–5 were analyzed. The first hit of each sequence was used for analysis.

doi:10.1371/journal.pone.0106770.g002

Library construction and Illumina sequencing

Ten μg of total RNA was used to isolate mRNA using oligo(dT) magnetic beads. The cDNA library was constructed using NEBNext mRNA Library Prep Reagent Set (NEB, Ipswich, MA, USA) following the manufacturer's protocols. Briefly, enriched poly(A) RNA of each sample was fragmented into 200–700 nt pieces with RNA Fragmentation Reagents. The cleaved RNA fragments were transcribed into the first-strand cDNA using random hexamer-primers, followed by second-strand cDNA synthesis. The resulting double-stranded cDNA (dsDNA) was purified with QiaQuick PCR extraction kit (Qiagen, Hilden, Germany) and dissolved in EB buffer. The purified dsDNA was treated with T4 DNA Polymerase and T4 Polynucleotide Kinase for end-repairing and dA-tailing. After that, they were ligated to sequencing adaptors with barcode using T4 DNA ligase. Finally, fragments with around 200 bp-length were purified with Qia-Quick GelPurify Kit (Qiagen, Hilden, Germany), and used as templates for PCR amplification to create the cDNA library. The library was sequenced on Illumina HiSeq 2000 (Illumina, San Diego, CA, USA) in Baimaike company (Beijing, China).

Assembly and annotation of transcriptomes

Raw reads were filtered to remove low quality reads with Q20 less than 20 and the sequence reads containing adapters and poly-A/T tails. The resulting clean reads were assembled to produce unigenes using the short reads assembling program – Trinity [37]. For functional annotations, we first searched all unigene sequences against various protein databases such as Nr, Swiss-Prot, COG, and KEGG using BLASTX, and then searched nucleotide database Nt using BLASTN, with an E-value cut-off of 10^{-5}

[38]. The BLAST results were used to extract coding region sequences (CDS) from the unigene sequences, and translate them into peptide sequences. When a unigene happened to have no BLAST hits, ESTScan software [39] would be used to determine the sequence direction. In addition, we performed the Gene Ontology (GO) annotations for each unigene with Blast2GO program according to the GO association done by a BLASTX against the Nr database [40,41].

Identification and sequence analysis of wing development-related genes from *L. migratoria* and *O. furnacalis* transcriptome

The available wing development-related gene sequences from *Drosophila* were used as references to screen *L. migratoria* transcriptome database obtained above and *O. furnacalis* transcriptome obtained previously [36]. The potential candidates of *L. migratoria* and *O. furnacalis* wing development-related genes were confirmed by searching the BLASTX algorithm against the non-redundant (nr) NCBI nucleotide database using a cut-off E-value of 10^{-5}.

For the sequence analysis of putative wing development-related genes identified above, the deduced protein domains were determined by using Pfam (http://www.sanger.ac.uk/Software/Pfam/) and SMART (http://smart.embl.de/). Analysis of deduced amino acid sequences, including prediction of signal peptide, molecular weight and isoelectric point, was carried out in the EXPASY (Expert Protein Analysis System) proteomics server (http://www.expasy.org). Sequence comparisons and phylogenetic analysis were performed by MEGA5 software [42]. Phylogenetic trees were constructed by the neighbor-joining
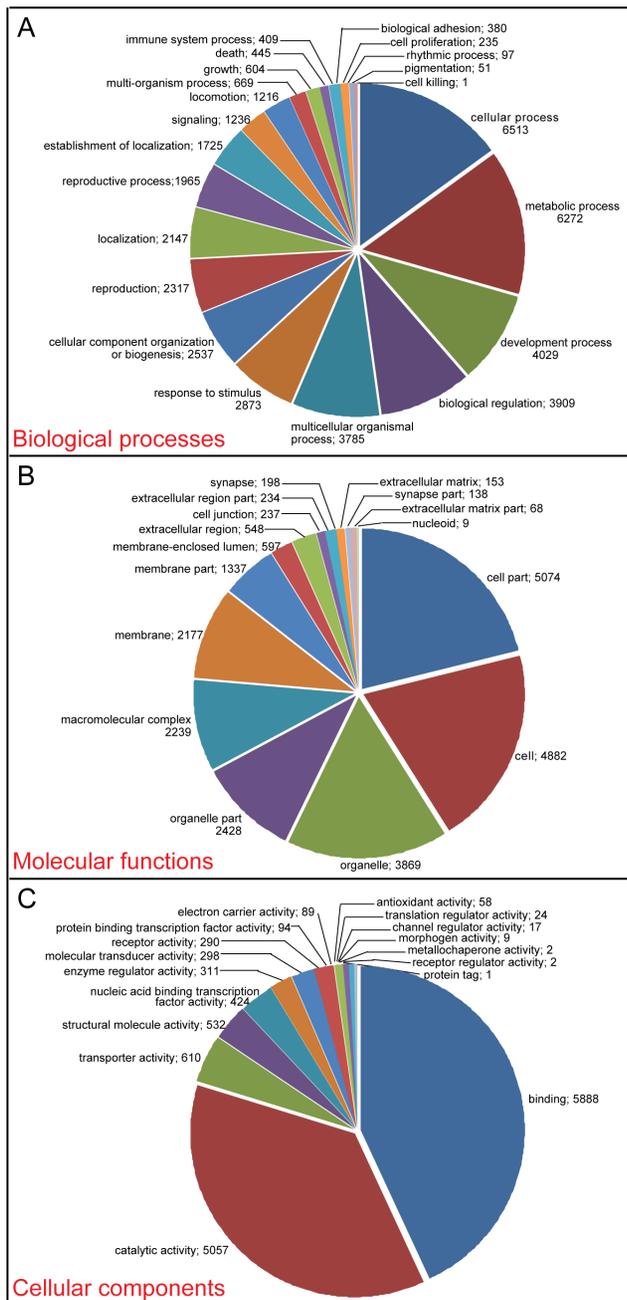
Figure 3. Gene ontology (GO) assignment for the *L. migratoria* transcriptome. GO assignments (level 2) as predicted for their involvement in (A) biological processes, (B) cellular components, and (C) molecular functions. The number of unigenes assigned to each GO term is shown behind semicolon.
doi:10.1371/journal.pone.0106770.g003

method, with statistical analysis by the bootstrap method, using 1000 repetitions.

Expression assay of several identified wing development-related genes

To investigate the expression profiles of several key wing development-related genes in forewings and hind wings, we dissected the wing discs as described above, and extracted total RNA independently from 3 biological replicates. DNase I-treated

RNA (1µg) was converted into first-strand cDNA using TIAN-Script RT Kit (TIANGEN, Beijing, China). The cDNA products were diluted 2 fold for use as template. Specific primers for each gene were designed and listed in Table S1. *L. migratoria* Actin and *O. furnacalis* ribosomal protein L8 (*rpL8*) was used as an internal standard to adjust the template amounts in a preliminary PCR experiment. The qRT-PCR was performed on a Applied Biosystems 7500 Real-time PCR system (Life Technologies, Grand Island, NY, USA) using the GoTaq qPCR Master (Promega, Madison, WI, USA), according to the manufacture's instructions. The thermal cycling conditions for qRT-PCR and calculation methods were same as described previously [36].
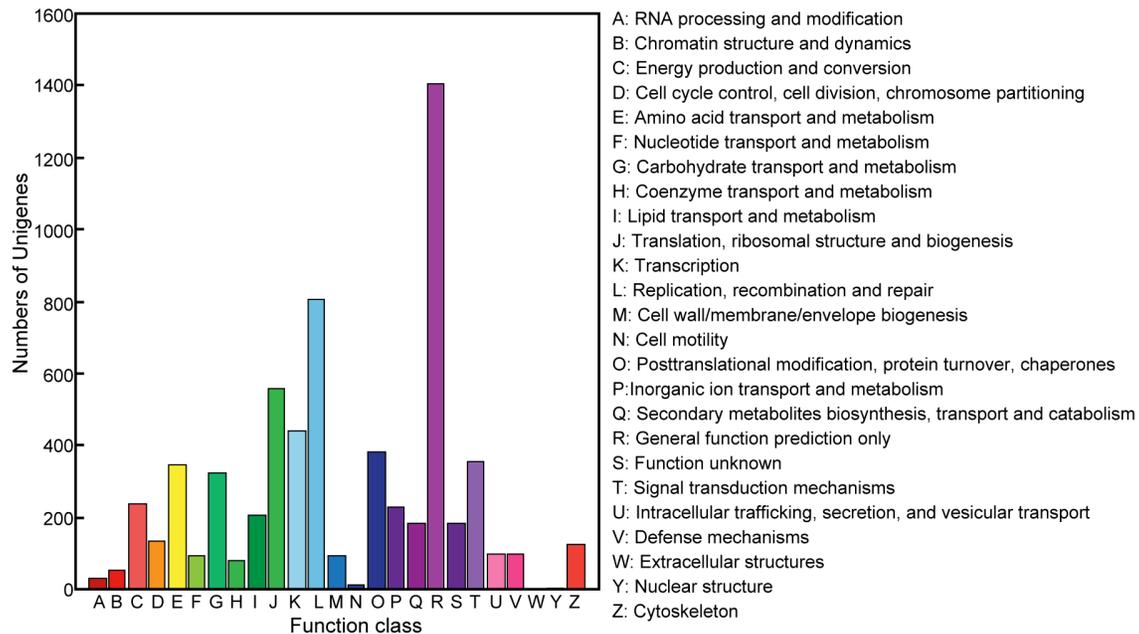
Results and Discussion

Dissection and observation of wing discs in *L. migratoria* and *O. furnacalis*

All insects in the Pterygota undergo metamorphosis from immature to adult. Among them, insects with incomplete metamorphosis, such as migratory locust *L. migratoria*, have young nymph resembling the adult with visible forewings and hind wings. Meanwhile, insects with complete metamorphosis, such as corn borer *O. furnacalis*, go through four stage processes from egg, larva, pupa, and adult in which only adult has visible wings. Although there are some minor differences, the overall appearances of forewings and hind wings in *L. migratoria* nymph and adult are similar (Fig. 1). However, actual wings are only visible in *O. furnacalis* adult. In fifth-instar *O. furnacalis* larva, a pair of pea-like wing discs composed of large amount of cells was dissected from the position where the forewing and hind wing will be derived. There is no similarity on the appearance between the wing discs and adult wings (Fig. 1). The huge difference about the wing development in *L. migratoria* and *O. furnacalis* suggests that the molecular mechanisms controlling the wing development in insects with incomplete or complete metamorphosis might be largely different. As a first step to understand these molecular mechanisms, it is important and necessary to identify as many as possible genes functioning in the wing development. Previously, we have combined the Illumina sequencing and *de novo* assembly to obtain the high quality transcriptome from *O. furnacalis* larvae [36]. In this study, we also obtained the data for *L. migratoria* transcriptome, and we can comparatively characterize both transcriptomes and identify wing development-related genes. This work will provide useful information for studying the molecular basis involved in the wing development in *L. migratoria* and *O. furnacalis*, two insect species with different type of metamorphosis.

Sequencing and unigene assembly of *L. migratoria* transcriptome

In order to obtain detailed information about *L. migratoria* transcriptome, we prepared cDNA from the wing discs of fifth-instar nymph, and subjected it to Hiseq 2000 sequencing. After cleaning of dirty reads and quality checks, a total of 83,000,540 high-quality clean reads (SRA accession number SRX491784) with a cumulative length of 8,382,624,748 nucleotides (8.38 Gb) were generated from *L. migratoria* wing disc library. The GC percentage of the reads is 42.65%, which is comparable with genome sequence of other insects. Using Trinity software by the manner of paired-end joining and gap-filling, these reads were assembled into 91,907 unigenes longer than 200 nt (mean length of 610 nt and N50 of 1024 nt). The size distribution indicated that the ratio of unigenes with a length of 200–1000 bp was 86.49%, while the length of 12,420 (13.51%) unigenes was more then 1000 bp (Fig.S1). It was significantly larger than that in previous

Legend:

A: RNA processing and modification
B: Chromatin structure and dynamics
C: Energy production and conversion
D: Cell cycle control, cell division, chromosome partitioning
E: Amino acid transport and metabolism
F: Nucleotide transport and metabolism
G: Carbohydrate transport and metabolism
H: Coenzyme transport and metabolism
I: Lipid transport and metabolism
J: Translation, ribosomal structure and biogenesis
K: Transcription
L: Replication, recombination and repair
M: Cell wall/membrane/envelope biogenesis
N: Cell motility
O: Posttranslational modification, protein turnover, chaperones
P: Inorganic ion transport and metabolism
Q: Secondary metabolites biosynthesis, transport and catabolism
R: General function prediction only
S: Function unknown
T: Signal transduction mechanisms
U: Intracellular trafficking, secretion, and vesicular transport
V: Defense mechanisms
W: Extracellular structures
Y: Nuclear structure
Z: Cytoskeleton

Figure 4. Clusters of orthologous groups (COG) classification of *L. migratoria* **unigenes.** A total of 6,508 produced functional annotations were among the 25 categories. The Y-axis shows the number of unigene in each COG term.
doi:10.1371/journal.pone.0106770.g004

Functional annotation and classification

In order to annotate the unigenes, the obtained sequences were first aligned by BLASTX to various protein databases of nr, Swiss-Prot, KEGG, COG, and GO (E-value$<10^{-5}$), and then aligned by BLASTN to nucleotide database nt (E-value$<10^{-5}$). Among 91,907 unigenes, 20,746 (22.6%), 8,737 (9.5%), 11,450 (12.5%), 5,312 (5.8%), 4,986 (5.4%), and 10,644 (11.6%) ones were annotated in nr, nt, Swiss-Prot, KEGG, COG, and GO, respectively (Table S2). A total of 23,359 unigenes (25.4%) were annotated to at least one database. The remaining 68,548 unigenes (74.6%) were not annotated to any referred databases. The low annotated percentage might be due to the transcripts derived from the cDNA of untranslated regions, assemblage errors, and nonconserved areas of proteins where homology is not detected [43]. Another possibility was that a large part of the genes in *L. migratoria* transcriptome database were with unknown functions, and the un-annotated unigenes were the potential sources of novel genes.

The functional annotations of unigenes were performed mainly based on the BLASTX results against the nr database. Among the 20,746 unigenes annotated to nr database, 7,883 (38%) showed strong homology (E-value smaller than 1e-50) (Fig. 2A). The identity comparison showed 5,809 (28%) unigenes have more than 60% identity with other insects (Fig. 2B). For species distribution of the top BLAST hits against the Nr databse, *L. migratoria* unigenes had the greatest number of matches with *Tribolium castaneum* (2,490 unigenes, 12%), followed by *Pediculus humanus corporis* (1,452 unigenes, 7%), and *Acyrthosiphon pisum* (1,245 unigenes, 6%) (Fig. 2C).

Gene Ontology (GO) assignment programs were used to classify the functions of all unigenes. GO contains three categories: biological process, cellular component, and molecular function.

Using the Blast2GO and WEGO software, 43,415, 24,188, and 13,706 unigenes were associated with biological process (22 subcategories), cellular component (16 sub-categories), and molecular function (17 sub-categories), respectively (Fig. 3). Among the biological process assignments, cellular processes (15.0%) and metabolic processes (14.4%) represented the most abundant subcategories. It indicated the importance of cell cycle, generation as well as metabolic activities in wing development stage. Under the category of cellular component, the top 3 sub-categories were cell part (21.0%), cell (20.2%) and organelle (16.0%). In the molecular function category, binding (43.0%) and catalytic activities (36.9%) were the most abundant (Fig. 3). The subcategories taking up the largest two proportions in each category were consistent with that in transcriptomic studies of other insects [45,46].

The unigenes were further compared against COG for the analysis of the putative protein functions. A total of 6,508 unigenes were functionally classified into 25 COG categories (Fig. 4). The cluster for 'General function prediction only' (1,407 unigenes) constituted the largest group, followed by 'Replication, recombination and repair' (809 unigenes), 'Translation, ribosomal structure and biogenesis' (560 unigenes), and 'Transcription' (441 unigenes) (Fig. 4). The group of 'signal transduction mechanisms' contains 354 unigenes, the fifth largest one within the COG classification of the *L. migratoria* transcriptome. It suggests the possible importance of signaling transduction pathways in wing development in *L. migratoria*.

Identification and comparison of wing development-related genes in *L. migratoria* and *O. furnacalis* transcriptome

Wing development is a complicated biological process comprising multiple integrated signaling pathways. In particular, four signaling pathways – Notch, Hedgehog (Hh), Decapentaplegic (Dpp), Wingless (Wg) – play important roles in growth control and cell fate determination during the wing development

Table 1. Summary of the wing development-related unigenes identified in *L. migratoria* and *O. furnacalis* transcriptome.

	Function[a]	*Locusta migratoria*						*Ostrinia furnacalis*					
		Unigene ID	Nucleotide Length (nt)	Full length[b]	Protein length (aa)[c]	Identity to Best hit (%)[d]	P value[e]	Unigene ID	Nucleotide length (nt)	Full length[b]	Protein Length (aa)[c]	Identity to Best hit (%)[d]	P value[e]
Hedgehog pathway													
hedgehog (hh)	L	a37032	4051	3'	383	66	e-101	Unigene16362	590	M	196	73	9e-064
patched (ptc)	R	a57898	1327	5'	372	75	e-135	Unigene153	3111	Y	775	62	0.0
smoothened (smo)	R	a59513	2125	3'	651	62	6e-073	CL2764.Contig1	2937	Y	751	67	0.0
cubitus interruptus (ci)	F	a69651	650	M	216	79	4e-096	Unigene15615	910	M	303	77	5e-098
costal-2 (co-2)	B	a8431	3970	Y	843	52	e-102	?					
Fused	K	a32509	2482	3'	650	74	7e-071	Unigene11591	1346	M	448	64	1e-071
suppressor of fused (sufu)	B	a11681	2472	3'	488	55	3e-086	Unigene16575	637	M	212	59	3e-042
Decapentaplegic pathway													
decapentaplegic (dpp)	L	a136595	815	3'	209	69	1e-068	Unigene7326	990	3'	306	63	1e-078
glass bottom boat (gbb)	L	a2659	1653	3'	332	64	5e-077	CL9574.Contig1	1573	Y	448	61	8e-097
saxophone (sax)	R	a14241	2261	Y	568	76	0.0	Unigene21017	796	M	256	77	1e-098
punt	R	a17842	5910	Y	503	69	e-144	CL628.Contig1	2051	Y	547	49	1e-051
thickveins (tkv)	R	a4697	2436	Y	512	72	e-174	Unigene5852	2162	3'	335	82	e-137
mothers against Dpp (mad)	F	a11173	2860	Y	475	84	0.0	Unigene16360	1775	Y	423	86	0.0
		a12619	3021	Y	436	59	e-152	Unigene6852	1575	Y	447	58	e-147
medea (med)	A	a26380	1235	3'	356	67	1e-095	Unigene4133	440	M	146	89	3e-045
		a11651	3401	M	397	66	e-131	Unigene17363	589	M	196	89	3e-080
optomotor-blind (omb)	F	a4984	803	M	267	96	e-121	CL105.Contig5	1131	M	376	89	9e-142
spalt major (salm)	F	a6922	1073	5'	325	55	2e-070	CL2350.Contig2	2278	M	759	54	2e-097
spalt related (salr)	F	a12971	2087	5'	651	47	2e-089	?					
Notch signaling Pathway													
Notch (n)	R	a1269	12296	Y	2484	72	0.0	Unigene6905	2340	M	780	83	0.0
delta (dl)	L	a123129	926	M	308	69	e-114	Unigene3068	987	M	328	69	e-120
serrate (ser)	L	a1910	7128	Y	1324	49	0.0	CL7020.Contig1	1927	M	409	52	e-164
Hairless (h)	R	a35245	1975	5'	538	61	5e-032	Unigene3346	243	M	80	68	4e-015

Table 1. Cont.

	Function[a]	Locusta migratoria						Ostrinia furnacalis					
		Unigene ID	Nucleotide Length (nt)	Full length[b]	Protein length (aa)[c]	Identity to Best hit (%)[d]	P value[e]	Unigene ID	Nucleotide length (nt)	Full length[b]	Protein Length (aa)[c]	Identity to Best hit (%)[d]	P value[e]
suppressor of hairless (su(H))	A	a4077	3351	Y	496	95	0.0	Unigene16575	637	M	211	59	3e-042
mastermind (mam)	A	a14121	1140	5'	153	54	1e-025	Unigene19131	1464	5'	487	54	3e-036
notchless	B	a6959	1804	Y	474	78	0.0	CL2896.Contig1	1054	5'	281	80	e-116
fringe	K	a17301	2466	3'	259	82	e-111	Unigene9365	1276	Y	327	73	e-100
hairy	B	a4291	3634	3'	407	52	6e-070	Unigene22517	673	5'	100	72	3e-32
Wingless pathway													
wingless (wg)	L	a40139	948	3'	227	85	7e-050	Unigene1219	292	M	97	65	4e-020
frizzled (fz)	R	a1577	5642	Y	571	68	0.0	CL4000.Contig1	1705	Y	497	71	e-157
frizzled2 (fz2)	R	?						Unigene2716	1952	5'	516	57	e-113
arrow (arr)	R	a39913	4189	5'	1381	65	0.0	Unigene11432	3712	5'	1167	67	0.0
disheveled (dsh)	F	a6911	3503	Y	721	60	e-154	Unigene24227	256	M	85	100	3e-044
armadillo (arm)	F	a1041	3398	Y	820	87	0.0	Unigene21327	2918	Y	737	84	0.0
axin	A	a4962	3922	3'	676	64	2e-031	Unigene17645	493	M	164	53	8e-023
adenomatous polyposis coli (APC)	A	a20732	4723	5'	1541	47	e-160	Unigene11483	2788	5'	880	60	e-156
vestigial (vg)	F	a302963	243	M	81	62	4e-013	Unigene4539	380	M	126	55	2e-021
scalloped (sd)	F	a1575	4661	Y	459	69	0.0	CL11169.Contig4	2469	Y	413	61	2e-167
shaggy	K	a2857	10994	Y	1913	83	0.0	Unigene16270	3819	Y	428	86	e-176
Others													
ultrabithorax (Ubx)	F	a12834	2418	3'	174	56	6e-053	Unigene20941	1207	Y	253	46	6e-068
apterous (ap)	F	a188016	391	M	130	59	5e-044	Unigene4961	1440	3'	348	53	1e-098
engrailed (en)	F	a1647	2530	3'	138	80	3e-052	CL5152.Contig1	2062	Y	517	70	5e-057
homothorax (Hth)	F	a22527	721	M	240	75	e-102	CL5887.Contig2	1800	3'	235	77	2e-099
teashirt (Tsh)	F	a866117	347	M	115	55	8e-027	?					
epidermal growth factor receptor (EGFR)	R	a1492	4431	Y	1394	53	0.0	CL625.Contig2	6301	Y	1458	60	0.0
rhomboid (rho)	K	a36342	2527	Y	294	58	1e-072	?					
nubbin (Nub)	F	a3913	1608	3'	460	45	1e-080	Unigene16843	2180	Y	355	65	2e-052
pannier (Pnr)	F	a455482	478	M	159	71	7e-053	Unigene16014I	2513	3'	341	46	3e-065
notum	K	a538614	267	M	88	62	9e-029	Unigene26847	643	M	214	47	1e-046
fat (ft)	R	a9170	5093	M	1697	51	0.0	Unigene18347	1879	M	626	55	0.0
four-jointed (fj)	K	a37684	381	M	127	66	9e-036	Unigene6909	1600	3'	530	41	e-107

Table 1. Cont.

	Locusta migratoria							Ostrinia furnacalis					
	Function^a	Unigene ID	Nucleotide Length (nt)	Full length^b	Protein length (aa)^c	Identity to Best hit (%)^d	P value^e	Unigene ID	Nucleotide length (nt)	Full length^b	Protein Length (aa)^c	Identity to Best hit (%)^d	P value^e
dally-like (dlp)	B	a6329	2624	Y	629	51	e-104	?					

a The mark of L, R, A, F, K, and B means that the unigens potentially function as ligand, receptor, transcription factor, transcription activator, protein kinase, or binding protein.
b The mark of Y, 5', 3', and M means that the fragment of the unigene consists of complete open reading frame, 5'-end containing start codon, 3'-end containing stop codon, and the middle part without start and stop codons, respectively.
c The number indicates the length of amino acid sequences deduced from available coding region, which is complete or not.
d Identity indicates the percentage of identical amino acid residues in unigenes to the sequence best hit in BLASTX.
e P-value is determined by FDR (false discovery rate).
doi:10.1371/journal.pone.0106770.t001

[20,22,25,47–49]. In order to obtain the global perspective on the wing development-related transcripts in *L. migratoria*, we searched the assembled transcriptome for orthologous genes known to be involved in the above four pathways in *Drosophila*. In total, we have identified 50 unigenes with significant similarity to *Drosophila* wing development-related genes, including 7 ones in Notch pathway, 12 in Hh pathway, 9 in Dpp pathway, and 10 in Wg pathway (Table 1). Meanwhile we performed the similar analysis for *O. furnacalis* transcriptome reported previously [36], and identified 46 unigenes with high potential to be related to the wing development (Table 1).

Genes involved in the Hh signaling pathway. The Hh pathway is an important signaling cascade in *Drosophila* to pattern the embryonic cuticles and adult appendages [50]. It is also vital for diverse aspects of animal development and essential in humans to pattern their limbs and internal organs [51,52]. The pathway takes its name from its polypeptide ligand, an intercellular signaling molecule called as Hh [53]. *Hh* was initially discovered as a segment polarity gene in *Drosophila* [54]. Mammals have three Hh orthologues, Sonic (Shh), Indian (Ihh), and Desert (Dhh) [55]. We have identified the potential *hh* genes, a37032 from *L. migratoria* and Unigene16362 from *O. furnacalis* transcriptome, respectively (Table 1). Although these two unigenes were incomplete, they have 66% and 73% amino acid sequence identity to *Drosophila* Hh (Fig. 5). On the other hand, *Drosophila* Hh is synthesized as inactive precursors with an N-terminal signaling region linked to a C-terminal autoprocessing region which begins with a cysteine for an acyl rearrangement analogous to step 1 of the protein splicing pathway [56]. This Cys residue is completely conserved in both *L. migratoria* and *O. furnacalis* Hh homologues (Fig. 5). The conserved motif A, B, F, J, K, and L in *Drosophila* Hh [57] were found in *L. migratoria*, but not in *O. furnacalis* Hh because its C-terminus was presently unknown (Fig. 5). The phylogenetic analysis of Hh from different insect species and other animals reveals that almost all nematode Hh-related proteins form a distinct clade while the other eukaryote Hh proteins, including *L. migratoria* a37032 and *O. furnacalis* Unigene16362, cluster into another Hh clade (Fig. 6). *L. migratoria* a37032 is grouped with other Orthopteran Hhs, and *O. furnacalis* Unigene16362 is clustered with other Lepidopteran Hhs. We performed qRT-PCR analysis to analyze the transcriptional expression of *hh* during different developmental stages. The mRNA abundance of *hh* in *L. migratoria* remained unchanged from the first instar through the fifth instar stage. The transcript abundance of *O. furnacalis hh* was consistent before the fourth instar stage, but increased significantly in the fifth instar larvae (Fig. 7). This result also suggests that the quality of the transcriptome assembly is high enough to be used for primer design.

Hh signaling is mediated by a multi-component receptor complex in the cell membrane. This receptor complex consists of a 12-span transmembrane protein, Patched (Ptc) as the receptor and a 7-span transmembrane protein, Smoothened (Smo) as the obligatory signal transducer across the plasma membrane [51,58]. When extracellular Hh binds to and is inhibited by Ptc, Smo starts to accumulate, and inhibit the proteolytic cleavage of zinc-finger transcription factor Cubitus interruptus (Ci) which is normally bound by the kinesin-like protein Costal-2 (Cos2). The intact Ci protein then translocate into the nucleus, allowing the transcription of some genes such as *dpp*. In the absence of Hh, Ptc blocks Smo activity, and full-length Ci protein is degraded into Ci fragment (CiR) that functions as a transcriptional repressor to block the transcription of target genes (Fig. S3A) [50]. By searching *L. migratoria* and *O. furnacalis* transcriptome using *Drosophila* corresponding genes as reference, we have identified

```
Of_Unigene16362   ------------------------------------------------------------
Lm_a37032         ------------------------------------------------------------
Dm_Hh             MDNHSSVPWASAASVTCLSLDAKCHSSSSSSSSSKSAASSISAIPQEETQTMRHIAHTQRC   60


Of_Unigene16362   --------------------AAACGPGRGFNPRQQARRSYPLVFRQHEPNYSEINKSAS     39
Lm_a37032         ------------------------GP--------SRKRLTPLVFKQHVPNVHEVSLAAA     27
Dm_Hh             LSRLTSLVALLLIVLPMVFSPAHSCGPGRGLG-RHRARNLYPLVLKQTIPNLSEYTNSAS    119
                                          *     :.  ***::*   **   * . :*:


Of_Unigene16362   GPAEGRITRHDEKFKDLVPNYNPDIDFKDDEGTGADRLMTQRCKEKLNTLAISVMNQWPG     99
Lm_a37032         GKPERRVTRNDDRFRDLVPNYNADIIFKDEEGTGADRLMTQRCKERLNTLAISVMNQWPG     87
Dm_Hh             GPLEGVIRRDSPKFKDLVPNYNRDILFRDEEGTGADRLMSKRCKEKLNVLAYSVMNEWPG    179
                  *   *   : *.. :*:****** **  *:*:*:*******:****:**.** ****:***


Of_Unigene16362   VHLRVNEGWDEENSHGENSLHYEGRAVDITTSDRDTSKIGMLARLAYQAGFDWVYYRSRS    159
Lm_a37032         VRLRVTEGWDEEGHHSADSLHYEGRAVDVTTSDRDRAKYGMLARLAVEAGFDWVYYESRA    147
Dm_Hh             IRLLVTESWDEDYHHGQESLHYEGRAVTIATSDRDQSKYGMLARLAVEAGFDWVSYVSRR    239
                  ::*  *.*.***:   *. :********* :*:***** :* ******* :****** *  **


Of_Unigene16362   YIHCSVKTESSVGTG-AG□FPSGAVVHTES------GPIDIAAL----------------    196
Lm_a37032         HIHCSVKSESSQAARSGG□FPGDATVLLSDGRRLPLRRLRVGQRVQAVDLSSGQTLFSEV    207
Dm_Hh             HIYCSVKSDSSISSHVHG□FTPESTALLESGVRKPLGELSIGDRVLSMTAN-GQAVYSEV    298
                  :*:*****:**  .:   **.  :..  ..        : : .
                                      Motif A


Of_Unigene16362   ------------------------------------------------------------
Lm_a37032         ILFLDRSPREERPFLRLSLASGGALLVTPAHLLPVDR-GSGGARQVFAGSVRPGDRLFVS    266
Dm_Hh             ILFMDRNLEQMQNFVQLHTDGGAVLTVTPAHLVSVWQPESQKLTFVFADRIEEKNQVLVR    358
                                           Motif B


Of_Unigene16362   ------------------------------------------------------------
Lm_a37032         DGNATSVHSERVLAVTEERRPGVFAPLTRQGTVVIDGIAASCYAVVSSHSLAHWSYAPFR    326
Dm_Hh             DVETGELRPQRVVKVGSVRSKGVVAPLTREGTIVVNSVAASCYAVINSQSLAHWGLAPMR    418
                                           Motif F        Motif J        Motif K


Of_Unigene16362   ---------------------------------------------------------
Lm_a37032         FGSGVRDFFRKLSSTVPEDGVESVSPQWATPGSGVHWYARALYAIARFVLPSGMLYE    383
Dm_Hh             LLSTLEAWLPAKEQLHSSPKVVSS----AQQQNGIHWYANALYKVKDYVLPQSWRHD    471
                                               Motif L
```

Figure 5. Alignments of potential Hhs in *L. migratoria*, *O. furnacalis* and *Drosophila*. The deduced amino acid sequences of *L. migratoria* a37032 and *O. furnacalis* Unigene16362 were compared with *Drosophila* Hh (Dm_Hh). Completely conserved amino acids are indicated by "*", and conservative substitutions by ":" and "." below the sequences. The conserved cysteine residue for an acyl rearrangement analogous to step 1 of the protein splicing pathway is boxed. The conserved motifs in *Drosophila* Hedgehog are underlined.
doi:10.1371/journal.pone.0106770.g005

the potential 1:1 orthologs of Ptc, Smo, Ci from both transcriptome databases (Table 1). It suggests that Hh signaling pathway is conserved in both *L. migratoria* and *O. furnacalis*.

Genes involved in Dpp signaling pathway. Dpp, directed by the Hh signaling, is a member of the bone morphogenetic protein (BMP) growth factor family [59]. Dpp forms a long-range dynamic and precise gradients to control cell survival, cell morphogenesis, cell proliferation, cell differentiation during the wing development [22,25,26,60]. When a cell receives a Dpp signal, its heteromeric receptor complex composed of type I receptor Thickveins (Tkv) and type II receptor Punt is activated [61,62]. Once activated, the receptors are able to phosphorylate an intracellular protein called mothers against Dpp (Mad) [63]. The phosphorylated Mad then associate with the Medea (Med), and the complex translocates to the nucleus where it binds to DNA and activates or suppresses the expression of the target genes in conjunction with other transcription factors [64]. Within this signaling pathway, the *Drosophila* Hox gene *Ultrabithorax* (*Ubx*)

restricts both the transcription and the mobility of Dpp [65]. Genes activated by Dpp signaling pathway include transcription factor *optomotor-blind* (*omb*) and two members of Spalt (sal) family *spalt major* (*salm*), *spalt related* (*salr*) and so on (Fig. S3B) [66,67].

In this study, we identified almost all transcripts of the above key components in Dpp signaling pathway (Table 1). Among them, one unigene with sequence similarity to *Drosophila* Dpp was identified from the transcriptome of *L. migratoria* (a136595) and *O. furnacalis* (Unigene7326), respectively. The 5′-end of the cDNA sequences of both potential *Dpp* is incomplete. In order to establish the orthology of these two fragments, we performed a phylogenetic analysis incorporating a selection of BMP growth factor members from *Drosophila*, other arthropods and a variety of deuterostome taxa. The resulting phylogenetic tree distinguishes two groups of proteins. One group consists of the arthropod *dpp* genes and their deuterostome homologs, the BMP2/4 genes. The second group comprises the *Drosophila* TGF-beta genes *screw* (*scw*) and *glass bottom boat* (*gbb*), and the remaining deuterostome

Figure 6. Phylogenetic analysis of Hhs. The amino acid sequences of *L. migratoria* a37032 and *O. furnacalis* Unigene16362, together with 68 potential Hhs from other organisms were used to build the neighbor joining tree. *L. migratoria* a37032 and *O. furnacalis* Unigene16362 are boxed. Nematode Hh-related proteins and other eukaryote Hh proteins are indicted with brackets. The arrows at nodes denote bootstrap value greater than 700 from 1000 trials.
doi:10.1371/journal.pone.0106770.g006

BMP genes with TGF-beta homology (Fig. 8). The identified fragments, *L. migratoria* a136595 and *O. furnacalis* Unigene7326, reside in the *dpp* group with statistical support (reliability value = 98 and 99 for a136595 and Unigene7326, respectively). Therefore, they are postulated as potential *dpp* genes. We conducted qRT-PCR analysis to investigate the expression profile of *dpp*. *L. migratoria* *dpp* was expressed at high level in the first instar nymphs, then decreased significantly in the second, third, and fourth instar nymphs, but significantly increased in the hind wings of the fifth instar nymph. The mRNA level of *O. furnacalis* *dpp* remained unchanged in all tested stages (Fig. 7). The different expression patterns of *dpp* in two insects suggest that *dpp* might be related to the different metamorphosis.

In addition, it is interesting that only one *mad* gene was present in *Drosophila* but two transcripts encoding for potential *mad* were identified from both *L. migratoria* and *O. furnacalis* transcriptomes (Table 1). Two Mads in *L. migratoria* (a11173 and a12619) shared 60% identity while two *O. furnacalis* Mads (Unigene16360 and Unigene6852) had 61% identity in amino acid sequences. A similar situation happened to the identification of *med*. There is only one *med* gene in *Drosophila* but two transcript fragments were identified for both *L. migratoria* and *O. furnacalis* *med*, including a26380 and a11651 in *L. migratoria*, and unigene4133 and unigene17363 in *O. furnacalis*, respectively. The encoded amino

acid sequences of a26380 and unigene4133 were highly similar to the N-terminus of *Drosophila* Med while a11651 and unigene17363 were highly similar to the C-terminus of *Drosophila* Med (Fig. S2). We doubted that the two unigenes from each transcriptome were just two partial fragments within *med* gene. However, no overlapped sequences were observed in a26380 and a11651, or in unigene4133 and unigene17363 (Fig. S2). One possible reason might be due to the sequencing errors because the predicted overlapping part was just located at the terminus of sequenced fragment. Further experiments are required to determine the full length of *med* in these two insect species.

Genes involved in the Notch signaling pathway. The Notch signaling pathway is highly conserved throughout the animal kingdom. It regulates cell-fate determination during development and maintains adult tissue homeostasis [47]. The key component of this signaling pathway is the Notch receptor. Notch is a 300 kDa single-pass transmembrane protein, composed of a large extracellular domain, a single transmembrane portion, and a small intracellular region [68–70]. After binding to its ligands, Delta and Serrate (known as Jagged in mammals), inactive Notch precursor undergoes two proteolytic cleavage events: the first cleavage is catalyzed by ADAM-family metalloproteases; the second cleavage is mediated by γ-secretase which is an enzyme complex containing presenilin, nicastrin, PEN2 and APH1 [71].
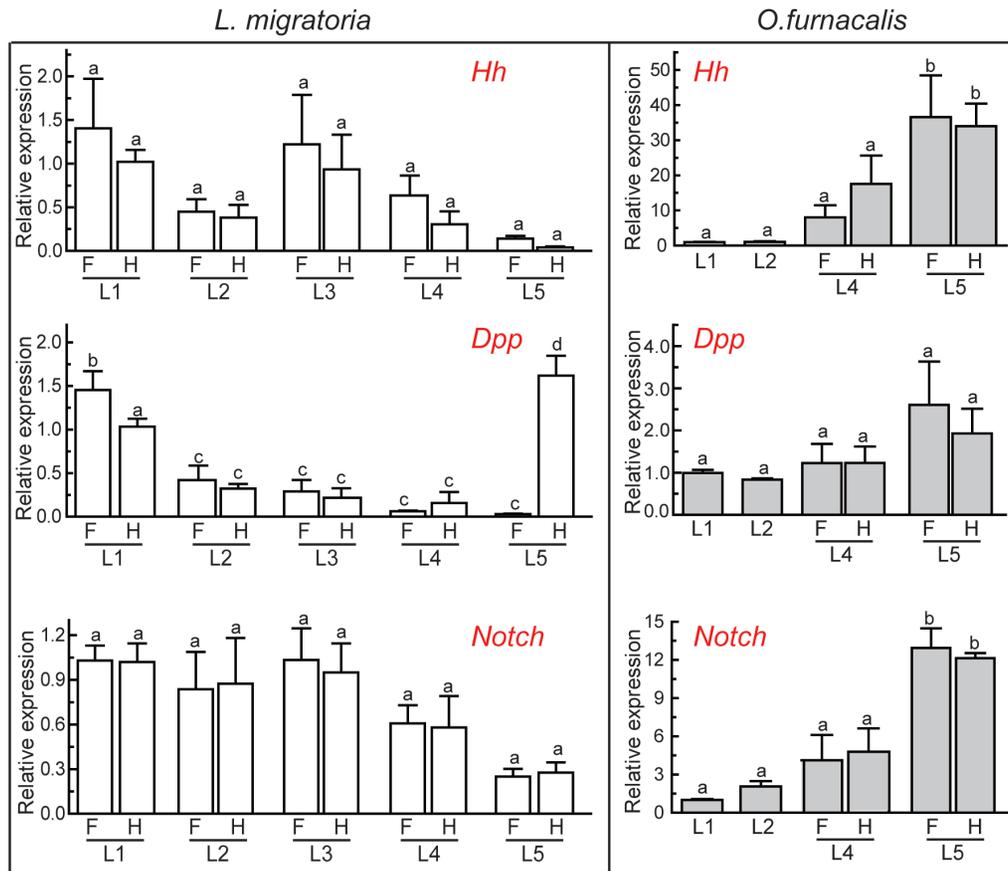
Figure 7. qRT-PCR analysis of transcriptional levels of *hh*, *dpp*, and *Notch* in *L. migratoria* and *O. furnacalis*. *Actin* and ribosomal protein L8 (*rpL8*) were used as an internal standard to normalize the templates in *L. migratoria* and *O. furnacalis*, respectively. RNA was extracted from wing discs of forewings (F) or hind wings (H) collected from the first-instar (L1), second-instar (L2), third-instar (L3), fourth-instar (L4) and fifth-instar (L5) *L. migratoria* nymph or *O. furnacalis* larvae. The bars represent the mean ± S. D. ($n = 3$). Bars labeled with different letters are significantly different (one-way ANOVA, followed by the Newman-Keuls test, $P < 0.05$).
doi:10.1371/journal.pone.0106770.g007

The second cleavage liberates the Notch intracellular domain (NICD), which then migrates into the nucleus and cooperates with the DNA-binding protein CSL (also named as CBF1, Suppressor of hairless (Su(H)), LAG-1, or RBP) and its co-activator Mastermind (Mam) to promote the transcription of downstream target genes, such as *wg* (Fig. S3C) [72]. Using the corresponding components in *Drosophila* as referred sequences, we have identified 9 transcripts for potential components in Notch signaling pathway from *L. migratoria* and *O. furnacalis* transcriptomes, respectively (Table 1). Four out of nine unigenes in *L. migratoria* are complete while only one unigene is complete in *O. furnacalis*. The possible reason was that the sequencing quality of *L. migratoria* transcription was better than that of *O. furnacalis* transcriptome (8.38 Gb vs. 4.72 Gb). Given the importance of Notch in this signaling pathway, we selected it for further analysis. Similar to the case in other invertebrates, we only identified one transcript for Notch gene, a1269 from *L. migratoria* transcriptome, and Unigene6905 from *O. furnacalis* transcriptome (Table 1). The canonical Notch protein consists of three repeated sequence motifs: the extracellular domains contain 10~36 copies of an ~40-amino acid epidermal growth factor-like (EGFL) sequence motif and 3 copies of an ~40-amino acid *lin/Notch/glp* (LNG) sequence motif; the intracellular domains contain 6–7 copies of a *cdc10/SW16/*ankyrin (CDC/ANK) sequence motif flanked by stretches of nonrepetitive sequences [73]. *L. migratoria*

a1269 encodes a 2,484-amino acids full-length protein which has 63% similarity to *Drosophila* Notch. *L. migratoria* Notch contains 36 EGFL tandem repeats, 3 LNG repeats, and 7 tandem ANK repeats (Fig. 9A), suggesting it is a canonical Notch. *O. furnacalis* Unigene6905 encodes a 780-amino acids polypeptide with 70% similarity to *Drosophila* Notch. Only 20 tandem EGFL repeats are predicted in the current identified Unigene6905 fragment (Fig. 9A). It is unknown whether *O. furnacalis* Notch also contain classic LNG and ankyrin repeats because the current transcript is incomplete. We performed the phylogenetic analysis for *L. migratoria* and *O. furnacalis* Notch and 30 Notch sequences from other species to investigate the evolutionary relationship of the Notch protein family. As shown in Fig. 9B, the Notch from insects forms a separate group which includes *L. migratoria* and *O. furnacalis* Notch, and the Notch from vertebrates is clustered into another group. Notch from *Ciona*, sea urchin, and amphioxus is grouped with vertebrate Notch, however, with a low bootstrap value of 82. *Caenorhabditis elegans* Notch is out of any group, showing a great differentiation from the other taxa. We analyzed the expression profiles of *Notch* using qRT-PCR methods. As shown in Fig. 7, the expression profile of *Notch* was similar to that of *Hh* in two insects. It kept unchanged from the first instar through the fifth instar stage in *L. migratoria*, while it increased significantly in *O. furnacalis* fifth instar larvae (Fig. 7).
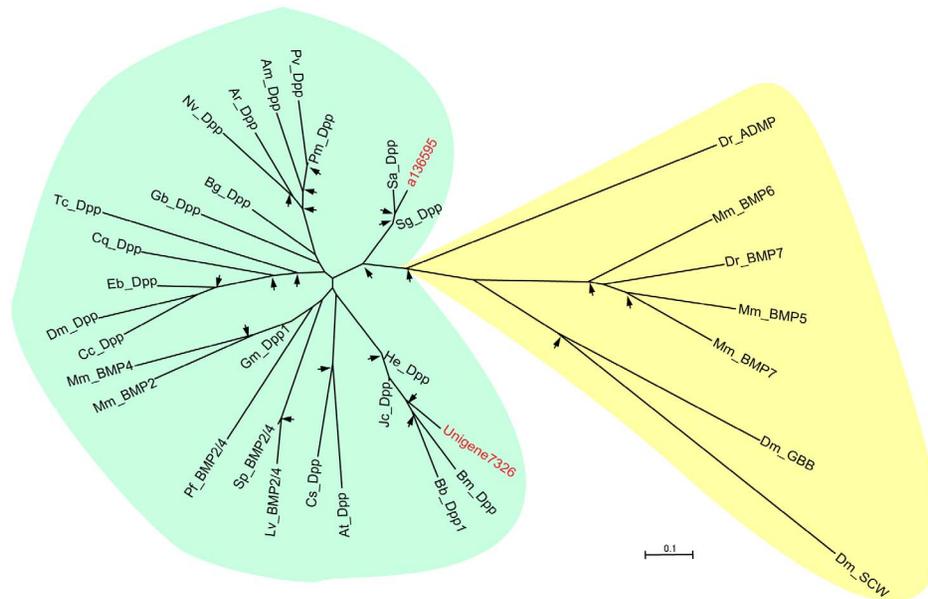
Figure 8. Phylogenetic analysis of Dpps. Except for *L. migratoria* a136595 and *O. furnacalis* Unigene7326 (marked in red), the used amino acid sequences of other 33 Dpps are from mouse (Mm), zebrafish (Dr), lancet (*Branchiostoma floridae*; Bf), acorn worm (*Ptychodera flava*; Pf), the sea urchins *Strongylocentrotus purpuratus* (Sp) and *Lytechinus variegatus* (Lv), fruit fly (Dm), flour beetle (*Tribolium castaneum*; Tc), sawfly (*Athalia rosae*; Ar), buckeye butterfly (*Junonia coenia*; Jc), the grasshoppers *Schistocerca americana* (Sa) and *S. gregaria* (Sg), cricket (*Gryllus bimaculatus*; Gb), *Cupiennius salei* (Cs), *Ceratitis capitata* (Cc), *Blattella germanica* (Bg), *Polyrhachis vicina* (Pv), *Episyrphus balteatus* (Eb), *Apis mellifera* (Am), *Pheidole morrisi* (Pm), *Biston betularia* (Bb), *Bombyx mori* (Bm), *Culex quinquefasciatus* (Cq), *Nasonia vitripennis* (Nv). The branches specific for arthropod Dpps and TGF-beta homologs are shaded in blur and yellow, respectively. For explanation of the arrows see Fig. 6.
doi:10.1371/journal.pone.0106770.g008

Genes involved in the wg signaling pathway. Wg (the vertebrate homolog of which is Wnt) is transcriptionally activated by Notch signaling, and Wg signaling pathway plays critical roles in axis patterning, cell fate specification, cell proliferation, and cell migration etc [74]. During the development of *Drosophila* wing, canonical Wg signaling pathway specifies pattern formation along the dorsal/ventral (D/V) axis while Dpp signaling pathway play this role along the A/P axis [75]. The ligand Wg is the founding member of the Wnt family, and is a secreted lipid-modified signaling glycoprotein that has 350–400 amino acids in length [76]. Wg is expressed at the D/V boundary and forms a stable and long-range gradient by symmetrically diffusing at both sides of the boundary. When Wg binds to a receptor complex consisting of the seven-transmembrane protein Frizzled (Fz) and the single-pass transmembrane protein Arrow (Arr, homologous to murine and human low density lipoprotein (LDL) receptor-related protein 5 or 6 (LRP5/6)), the downstream cytoplasmic protein disheveled (Dsh in *Drosophila* and Dvl in vertebrates) is activated [77]. Dsh in turn inhibits glycogen synthase kinase (GSK)-3β in the β-catenin destruction complex, which mainly consists of Axin, GSK-3β, adenomatous polyposis coli (APC) and β-catenin (armadillo (arm) in *Drosophila*). Consequently, β-catenin accumulates in the cytoplasm, and stabilized β-catenin then translocates into the nucleus and acts together with the transcription factor Pangolin to regulate the transcription of Wg target genes (Fig. S3D) [77]. In this study, we identified 10 and 11 unigenes for the known components in the Wg signaling pathway from *L. migratoria* and *O. furnacalis* transcriptome, respectively (Table 1). No putative *Frizzled 2* (*fz2*) ortholog was identified in *L. migratoria* transcriptome. A possible reason is that *fz2* gene is missing in *L. migratoria* because of the evolutionary event. The other reason with higher possibility is that the transcript level of *fz2* is low and it

is not captured in the RNA-seq. We attempted to perform further analysis for identified *wg* gene, a40139 in *L. migratoria* and Unigene1219 in *O. furnacalis*. However, unigene1219 only encodes a 97-amino acid peptide which is too short to have no common sequences with other Wg during the alignment. Therefore, we failed to conduct phylogenetic analysis to reveal the evolutionary relationship of Wg. Additionally, we identified another 12 and 9 unigenes from *L. migratoria* and *O. furnacalis* transcriptome, respectively (Table 1), which were potentially involved in the wing development, including homologs to *Drosophila apterous*, *engrailed*, *homothorax*, *teashirt*, *epidermal growth factor receptor*, *rhomboid*, *nubbin*, *pannier*, *notum*, *fat*, *four-jointed*, *dally-like*.

Conclusions

In summary, we sequenced and characterized the transcriptome from the wing discs of *L. migratoria* nymph. The assembled sequence data comprising 91,907 unique transcripts provides a comprehensive sequence source for future *L. migratoria* study. We identified a large set of genes relevant to wing development with high significance, especially the genes involved in four signaling pathways – Notch, Hh, Dpp, and Wg signaling pathways, from *L. migratoria* transcriptome and another *O. furnacalis* transcriptome obtained previously. The explored wing development-related genes constitute an integrated picture of the development network, which provides the valuable clues for a better understanding of the wing development in *L. migratoria* and *O. furnacalis*. These development repertoire genes appear to be evolutionarily conserved to different extent. Functional analyses are necessary to verify our predictions. Nevertheless, the framework of information presented in this study should help to further

Figure 9. (A) Schematic representation of the *L. migratoria* and *O. furnacalis* Notch. The domain organization was predicted using the SMART program (http://smart.embl.de/). Question mark means the end is incomplete. (B) Phylogenetic analysis of the Notch from insects, hydrobios and vertebrates. The unrooted tree was generated using the distance-based method Neighbor-Joing, with *Caenorhabditis elegans* protein as outgroup. The branch lengths reflect evolutionary divergence. The circled bootstrap value indicates that *L. migratoria* a1269 and *O. furnacalis* Unigene6905 (marked in red) belong to insect Notch group.
doi:10.1371/journal.pone.0106770.g009

understand the complex molecular mechanisms involved in wing development in *L. migratoria* and *O. furnacalis*, two insect species with different type of metamorphosis.

Supporting Information

Figure S1 Assembled unigene length distribution of *L. migratoria* transcriptome. The *x*-axis indicates unigene size and the *y*-axis indicates the number (*left*) or percentage (*right*) of unigenes of each size.
(EPS)

Figure S2 Alignments of potential Meds in *L. migratoria*, *O. furnacalis* and *Drosophila*. Note: *L. migratoria* a26380 (green) and *O. furnacalis* Unigene4133 (purple) match well with the N-terminus of *Drosophila* Med (blue), while *L. migratoria* a11651 (black) and *O. furnacalis* Unigene17363 (red) are highly similar to

the C-terminus of *Drosophila* Med (blue). The part without identity is shaded.
(EPS)

Figure S3 Schematic drawing of the Hh (A), Dpp (B), Notch (C), and Wg (D) signaling pathways. **(A)** In the presence of Hh, Hh binds to its receptor Ptc and Ptc relieves Smo repression. Smo accumulates and is activated by phosphorylation. This promotes its association with the complex including Costal 2 (Cos2), Fused (Fu), Suppressor of Fu (Su(Fu)) and Cubitus interruptus (Ci). Uncleaved Ci is released from the complex and acts as transcriptional activator in the nucleus and induces expression of target genes. **(B)** Dpp binds to its receptor complex including the type I receptor Tkv and type II receptor Punt which induces the phosphorylation of Tkv and in turn phosphorylates the signal transducer Mad. Phosphorylated Mad forms complex with Med and then translocates to the nucleus and induces expression of

target genes. (**C**) Notch receptor is a transmembrane protein, with a large extracellular domain (NECD), a transmembrane domain and an intracellular domain (NICD). When its ligands Dl or Ser bind to the Notch, two proteolytic cleavage events are activated: the first cleavage is catalyzed by ADAM-family TACE metalloproteases; the second cleavage is mediated by γ-secretase which is an enzyme complex containing presenilin, nicastrin, PEN2 and APH1. The second cleavage liberates the NICD, which then migrates into the nucleus and cooperates with the DNA-binding protein CSL and its co-activator Mam to promote the transcription of downstream target genes. (**D**) Wg binds to the receptors Fz and Arr. The Wg-Arr-Fz complex binds to and activates the protein Dsh which then disrupts the complex of APC/Axin/Sgg and results in an increase of free cytosolic Arm. Arm then translocates to the nucleus where it binds to transcription factor Pangolin (dTCF) and activates the expression of target genes. (EPS)

Table S1 Primers for RT-PCR analysis. (XLS)

Table S2 The statistics of annotated unigenes. (DOC)

Author Contributions

Conceived and designed the experiments: SL LZ JS CA. Performed the experiments: SL WW YC. Analyzed the data: SL JS CA. Contributed reagents/materials/analysis tools: SL WW YC. Contributed to the writing of the manuscript: SL LZ JS CA.

References

1. Dudley R (2002) The biomechanics of insect flight: form, function, evolution. Princeton University Press.
2. Wagner GP (2007) The developmental genetics of homology. Nat Rev Genet 8: 473–479.
3. Miaoulis IN, Heilman BD (1998) Butterfly thin films serve as solar collectors. Ann Entomol Soc Am 91: 122–127.
4. Sudo S, Tsuyuki K, Kanno K (2005) Wing characteristics and flapping behavior of flying insects. Exp Mech 45: 550–555.
5. Simmons RB, Conner WE (1996) Ultrasonic signals in the defense and courtship of *Euchaetes egle* Drury and *E. bolteri* Stretch (Lepidoptera: Arctiidae). J Insect Behav 9: 909–919.
6. Gibson G, Warren B, Russell IJ (2010) Humming in tune: sex and species recognition by mosquitoes on the wing. J Assoc Res Otolaryngol 11: 527–540.
7. Truman JW, Riddiford LM (1999) The origins of insect metamorphosis. Nature 401: 447–452.
8. Alberts B, Johnson A, Lewis J, Walter P, Raff M, et al. (2002) Organogenesis and the Patterning of Appendages. In: Alberts B, Johnson A, Lewis J, Walter P, Raff M, et al, editors. Molecular Biology of the Cell 5th Edition. Routledge. pp. 1347–1363.
9. Herrera SC, Martin R and Morata G (2013) Tissue homeostasis in the wing disc of *Drosophila melanogaster*: immediate response to massive damage during development. PLoS Genet 9: e1003446.
10. Fortini ME, Bonini NM (2000) Modeling human neurodegenerative diseases in *Drosophila*: on a wing and a prayer. Trends Genet 16: 161–167.
11. Neto-Silva RM, Wells BS, Johnston LA (2009) Mechanisms of growth and homeostasis in the *Drosophila* wing. Annu Rev Cell Dev Biol 25: 197–220.
12. Baena-Lopez LA, Nojima H, Vincent JP (2012) Integration of morphogen signalling within the growth regulatory network. Curr Opin Cell Biol 24: 166–172.
13. Waldron JA, Newbury SF (2012) The roles of miRNAs in wing imaginal disc development in *Drosophila*. Biochem Soc Trans 40: 891–895.
14. Worley MI, Setiawan L, Hariharan IK (2012) Regeneration and transdetermination in *Drosophila* imaginal discs. Annu Rev Genet 46: 289–310.
15. Bate M, Arias AM (1991) The embryonic origin of imaginal discs in *Drosophila*. Development 112: 755–761.
16. Morata G, Lawrence PA (1975) Control of compartment development by the engrailed gene in *Drosophila*. Nature 255: 614–617.
17. Garcia-Bellido A, Ripoll P, Morata G (1976) Developmental compartmentalization in the dorsal mesothoracic disc of *Drosophila*. Dev Biol 48: 132–147.
18. Zecca M, Struhl G (2002) Subdivision of the *Drosophila* wing imaginal disc by EGFR-mediated signaling. Development 129: 1357–1368.
19. Podos SD, Ferguson EL (1999) Morphogen gradients: new insights from DPP. Trends Genet 15: 396–402.
20. Tabata T, Takei Y (2004) Morphogens, their identification and regulation. Development 131: 703–712.
21. Strigini M, Cohen SM (1997) A Hedgehog activity gradient contributes to AP axial patterning of the *Drosophila* wing. Development 124: 4697–4705.
22. Teleman AA, Cohen SM (2000) Dpp Gradient Formation in the *Drosophila* Wing Imaginal Disc. Cell 103: 971–980.
23. Strigini M, Cohen SM (2000) Wingless gradient formation in the *Drosophila* wing. Curr Biol 10: 293–300.
24. Zecca M, Basler K, Struhl G (1996) Direct and long-range action of a wingless morphogen gradient. Cell 87: 833–844.
25. Zhang X, Luo D, Pflugfelder GO, Shen J (2013) Dpp signaling inhibits proliferation in the Drosophila wing by Omb-dependent regional control of bantam. Development 140: 2917–2922.
26. Shen J, Dahmann C (2005) Extrusion of cells with inappropriate Dpp signaling from *Drosophila* wing disc epithelia. Science 307: 1789–1790.
27. Shen J, Dahmann C, Pflugfelder GO (2010) Spatial discontinuity of optomotor-blind expression in the *Drosophila* wing imaginal disc disrupts epithelial architecture and promotes cell sorting. BMC Dev Biol 10: 23.
28. 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.
29. Uvarov BP (1936) The oriental migratory locust (*Locusta migratoria manilensis*, Meyen 1835). Bull Entomol Res 27: 91–104.
30. Nafus DM, Schreiner IH (1991) Review of the biology and control of the Asian corn borer, *Ostrinia furnacalis* (Lep: Pyralidae). International Journal of Pest Management 37: 41–56.
31. Preston G (1997) Cloning Gene Family Members Using PCR with Degenerate Oligonucleotide Primers. PCR Cloning Protocols. Humana Press. pp. 433–449.
32. Wang X, Fang X, Yang P, Jiang X, Jiang F, et al. (2014) The locust genome provides insight into swarm formation and long-distance flight. Nat Commun 5: 2957.
33. Ewen-Campen B, Shaner N, Panfilio KA, Suzuki Y, Roth S, et al. (2011) The maternal and early embryonic transcriptome of the milkweed bug *Oncopeltus fasciatus*. BMC Genomics 12: 61.
34. Zheng W, Peng T, He W, Zhang H (2012) High-throughput sequencing to reveal genes involved in reproduction and development in *Bactrocera dorsalis* (Diptera: Tephritidae). PLoS One 7: e36463.
35. Van Belleghem SM, Roelofs D, Van Houdt J, Hendrickx F (2012) *De novo* transcriptome assembly and SNP discovery in the wing polymorphic salt marsh beetle *Pogonus chalceus* (Coleoptera, Carabidae). PLoS One 7: e42605.
36. Liu Y, Shen D, Zhou F, Wang G, An C (2014) Identification of Immunity-Related Genes in *Ostrinia furnacalis* against Entomopathogenic Fungi by RNA-Seq Analysis. PLoS One 9: e86436.
37. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29: 644–652.
38. 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.
39. Iseli C, Jongeneel CV, Bucher P (1999) ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. Proc Int Conf Intell Syst Mol Biol: 138–148.
40. 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.
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. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28: 2731–2739.
43. Zhu JY, Yang P, Zhang Z, Wu GX, Yang B (2013) Transcriptomic immune response of *Tenebrio molitor* pupae to parasitization by *Scleroderma guani*. PLoS One 8: e54411.
44. Karatolos N, Pauchet Y, Wilkinson P, Chauhan R, Denholm I, et al. (2011) Pyrosequencing the transcriptome of the greenhouse whitefly, *Trialeurodes vaporariorum* reveals multiple transcripts encoding insecticide targets and detoxifying enzymes. BMC Genomics 12: 56.
45. Bai X, Mamidala P, Rajarapu SP, Jones SC, Mittapalli O (2011) Transcriptomics of the bed bug (*Cimex lectularius*). PLoS One 6: e16336.
46. Badisco L, Huybrechts J, Simonet G, Verlinden H, Marchal E, et al. (2011) Transcriptome analysis of the desert locust central nervous system: production and annotation of a *Schistocerca gregaria* EST database. PLoS One 6: e17274.

47. Rafel N, Milan M (2008) Notch signalling coordinates tissue growth and wing fate specification in *Drosophila*. Development 135: 3995–4001.

48. Brook WJ (2000) Hedgehog signaling and the axial patterning of *Drosophila* wings. Biochem Cell Biol 78: 585–591.

49. Neumann CJ, Cohen SM (1997) Long-range action of Wingless organizes the dorsal-ventral axis of the *Drosophila* wing. Development 124: 871–880.

50. Jiang J, Hui CC (2008) Hedgehog signaling in development and cancer. Dev Cell 15: 801–812.

51. Ingham PW, McMahon AP (2001) Hedgehog signaling in animal development: paradigms and principles. Genes Dev 15: 3059–3087.

52. Weedon MN, Lango H, Lindgren CM, Wallace C, Evans DM, et al. (2008) Genome-wide association analysis identifies 20 loci that influence adult height. Nat Genet 40: 575–583.

53. Tabata T, Kornberg TB (1994) Hedgehog is a signaling protein with a key role in patterning *Drosophila* imaginal discs. Cell 76: 89–102.

54. Perrimon N, Mahowald AP (1987) Multiple functions of segment polarity genes in *Drosophila*. Dev Biol 119: 587–600.

55. Varjosalo M, Taipale J (2008) Hedgehog: functions and mechanisms. Genes Dev 22: 2454–2472.

56. Perler FB (1998) Protein splicing of inteins and hedgehog autoproteolysis: structure, function, and evolution. Cell 92: 1–4.

57. Burglin TR (2008) Evolution of *hedgehog* and *hedgehog-related* genes, their origin from Hog proteins in ancestral eukaryotes and discovery of a novel Hint motif. BMC Genomics 9: 127.

58. Carpenter D, Stone DM, Brush J, Ryan A, Armanini M, et al. (1998) Characterization of two patched receptors for the vertebrate hedgehog protein family. Proc Natl Acad Sci U S A 95: 13630–13634.

59. Spencer FA, Hoffmann FM, Gelbart WM (1982) Decapentaplegic: a gene complex affecting morphogenesis in *Drosophila melanogaster*. Cell 28: 451–461.

60. Akiyama T, Kamimura K, Firkus C, Takeo S, Shimmi O, et al. (2008) Dally regulates Dpp morphogen gradient formation by stabilizing Dpp on the cell surface. Dev Biol 313: 408–419.

61. Affolter M, Basler K (2007) The Decapentaplegic morphogen gradient: from pattern formation to growth regulation. Nat Rev Genet 8: 663–674.

62. Bangi E, Wharton K (2006) Dual function of the *Drosophila* Alk1/Alk2 ortholog Saxophone shapes the Bmp activity gradient in the wing imaginal disc. Development 133: 3295–3303.

63. Newfeld SJ, Chartoff EH, Graff JM, Melton DA, Gelbart WM (1996) Mothers against dpp encodes a conserved cytoplasmic protein required in DPP/TGF-beta responsive cells. Development 122: 2099–2108.

64. Wisotzkey RG, Mehra A, Sutherland DJ, Dobens LL, Liu X, et al. (1998) Medea is a *Drosophila* Smad4 homolog that is differentially required to potentiate DPP responses. Development 125: 1433–1445.

65. Crickmore MA, Mann RS (2006) Hox control of organ size by regulation of morphogen production and mobility. Science 313: 63–68.

66. Grimm S, Pflugfelder GO (1996) Control of the gene *optomotor-blind* in *Drosophila* wing development by decapentaplegic and wingless. Science 271: 1601–1604.

67. de Celis JF, Barrio R, Kafatos FC (1996) A gene complex acting downstream of dpp in *Drosophila* wing morphogenesis. Nature 381: 421–424.

68. Blaumueller CM, Qi H, Zagouras P, Artavanis-Tsakonas S (1997) Intracellular cleavage of Notch leads to a heterodimeric receptor on the plasma membrane. Cell 90: 281–291.

69. Logeat F, Bessia C, Brou C, LeBail O, Jarriault S, et al. (1998) The Notch1 receptor is cleaved constitutively by a furin-like convertase. Proc Natl Acad Sci U S A 95: 8108–8112.

70. Chan YM, Jan YN (1998) Roles for proteolysis and trafficking in notch maturation and signal transduction. Cell 94: 423–426.

71. Bray SJ (2006) Notch signalling: a simple pathway becomes complex. Nat Rev Mol Cell Biol 7: 678–689.

72. Diaz-Benjumea FJ, Cohen SM (1995) Serrate signals through Notch to establish a Wingless-dependent organizer at the dorsal/ventral compartment boundary of the *Drosophila* wing. Development 121: 4215–4225.

73. Maine EM, Lissemore JL, Starmer WT (1995) A phylogenetic analysis of vertebrate and invertebrate Notch-related genes. Mol Phylogenet Evol 4: 139–149.

74. Swarup S, Verheyen EM (2012) Wnt/Wingless signaling in *Drosophila*. Cold Spring Harb Perspect Biol 4.

75. Zhang XB, Lu J, Shen J (2010) How to pattern a tissue? –Knowledge from *Drosophila* wing model system. Chinese Bulletin of Entomology 1: 009.

76. Cadigan KM, Nusse R (1997) Wnt signaling: a common theme in animal development. Genes Dev 11: 3286–3305.

77. Stadeli R, Hoffmans R, Basler K (2006) Transcription under the control of nuclear Arm/beta-catenin. Curr Biol 16: R378–385.