

RESEARCH ARTICLE

Open Access

# Identification of differential expression genes associated with host selection and adaptation between two sibling insect species by transcriptional profile analysis

Haichao Li<sup>1,2</sup>, Hao Zhang<sup>1,2</sup>, Ruobing Guan<sup>1,2</sup> and Xuexia Miao<sup>1\*</sup>

## Abstract

**Background:** Cotton bollworm (*Helicoverpa armigera*) and oriental tobacco budworm (*Helicoverpa assulta*) are noctuid sibling species. Under artificial manipulation, they can mate and produce fertile offspring. As serious agricultural insect pests, cotton bollworms are euryphagous insects, but oriental tobacco budworms are oligophagous insects. To identify the differentially expressed genes that affect host recognition and host adaptation between the two species, we constructed digital gene expression tag profiles for four developmental stages of the two species. High-throughput sequencing results indicated that we have got more than 23 million 17nt clean tags from both species, respectively. The number of unique clean tags was nearly same in both species (approximately 357,000).

**Results:** According to the gene annotation results, we identified 83 and 68 olfaction related transcripts from *H. armigera* and *H. assulta*, respectively. At the same time, 1137 and 1138 transcripts of digestion enzymes were identified from the two species. Among the olfaction related transcripts, more odorant binding protein and G protein-coupled receptor were identified in *H. armigera* than in *H. assulta*. Among the digestion enzymes, there are more detoxification enzyme, e.g. P450, carboxypeptidase and ATPase in *H. assulta* than in *H. armigera*. These differences partially explain that because of the narrow host plant range of *H. assulta*, more detoxification enzymes would help them increase the food detoxification and utilization efficiency.

**Conclusions:** This study supplied some differentially expressed genes affecting host selection and adaptation between the two sibling species. These genes will be useful information for studying on the evolution of host plant selection. It also provides some important target genes for insect species-specific control by RNAi technology.

**Keywords:** Development, Host plant range, Transcripts, Digital gene expression tag profile (DGE-Tag), Sibling species, Differential expression gene, *Helicoverpa armigera*, *Helicoverpa assulta*

## Background

Cotton bollworm (*Helicoverpa armigera*, Hübner) and oriental tobacco budworm (*Helicoverpa assulta*, Guenée) are two sibling noctuid species of Lepidoptera. They are distributed in almost same region from 50°S to 50°N and from 45°S to 45°N, respectively. *H. armigera* are slightly broader than oriental tobacco budworms [1,2]. In the field, similar external morphology makes them easily confused.

Interestingly, these two species can mate under artificial manipulation and produce offspring. However, when female *H. armigera* mated with male *H. assulta*, the first filial generations are all males [3]. This result further confirms that they are two distinct species [4]. Under natural conditions, because of differences in sex pheromone composition, the two species seldom mate. Their sex pheromones comprise *cis*-11-hexadecenal (Z11-16: Ald) and *cis*-9-hexadecenal (Z9-16: Ald), but the compositions are reversed in the two species. The ratio of these two components is 97:3 in *H. armigera*, but it is 7:93 in *H. assulta* [5-7].

\* Correspondence: xxm@sippe.ac.cn

<sup>1</sup>Key Laboratory of Insect Developmental and Evolutionary Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China

Full list of author information is available at the end of the article

In addition, the host ranges of these two species are significantly different. *H. armigera* is a euryphagous insect whose host range includes 40 families of over 200 different plants. However, *H. assulta* is an oligophagous insect, they are mainly feeding on the plants of the Solanaceae, for example, tobacco and hot pepper [1,8,9]. Although each species have their own preferred host plants, both of them love feed on tobacco and hot pepper [4]. These similarity and difference may be depend on the host plant selection by adult, or depend on the food digestion or detoxification enzymes from larvae. Host plant selection is a complicated and continuous process. The color, odor and shape of the plants will affect the insects choice on host plants, among which odors are a critically important element for the lifestyle and reproduction of an insect species [10]. Accordingly, insects with different feeding habits possess their own specific odor identification and odor-binding proteins [11]. At the same time, enzymes for food digestion and detoxification are also important factors for insect growth and development. These enzymes probably effect on the survival of insects, consequently affecting on the host range of an insect species. Therefore, identifying enzymes related to insect development and feeding habits will benefit to research on insect host range and on insect pest control.

The two sibling species are non-model insects. Their genome sequences are not available till now. To identify differentially expressed genes from *H. armigera* and *H. assulta*, digital gene expression tag (DGE-tag) profile libraries were constructed and sequenced using high throughput second-generation sequencing technology [12,13]. A DGE-tag profile is according to the theory and method of SAGE (serial analysis of gene expression) combined with high-throughput sequencing technology [14]. In this study, eight DGE-tag libraries were constructed and sequenced for four developmental stages (embryo, larva, pupa and adult) of the two species. Differentially expressed transcripts or genes between *H. armigera* and *H. assulta* were analyzed by bioinformatics. Most growth and development related genes have similar expression modes. The differentially expressed genes are mainly focus on olfactory-related genes and enzymes for food detoxification or digestion. Therefore, the two sibling species represent a good model for host plant selection and adaptability. These results also provide valuable data for insect pest control.

## Results and discussion

### The main identifying characteristics of *H. armigera* and *H. assulta*

The two insect species in genus *Helicoverpa*, Cotton bollworm (*H. armigera*) and oriental tobacco budworm (*H. assulta*), are important insect pests of crops in China. They have similar external morphology. Figure 1 shows

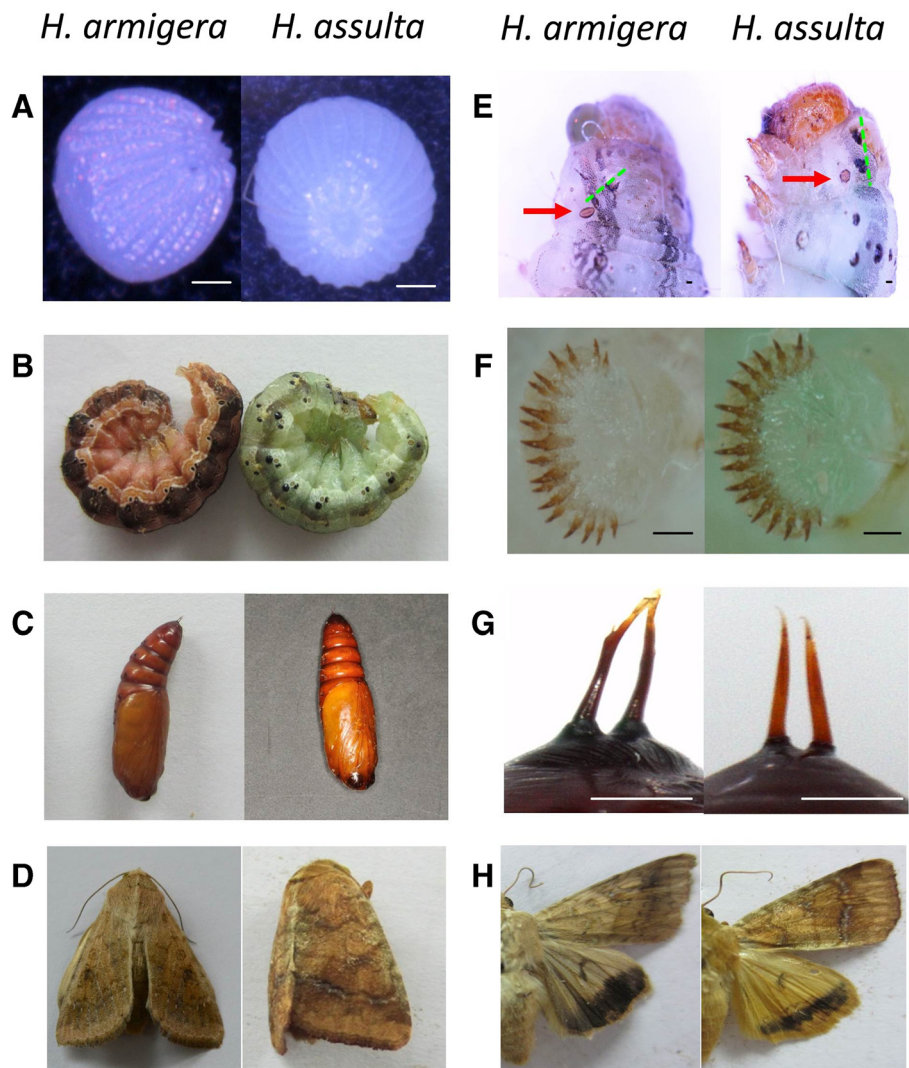
some taxonomic characteristics to distinguish these two species. Their eggs, larvae and pupae look like nearly same (Figure 1A-C). Only under the microscope, according to some taxonomic characteristics they can be distinguished (Figure 1E-G). The adult is the easiest stage to be distinguished by entomologist, because there are some special speckles and stripes on the wings (Figure 1D, H).

To show the relationship of *H. armigera* and *H. assulta*, multiple sequence alignments of spanning the 18S rRNA across 26 species from 23 orders (Additional file 1, Additional file 2: Figure S1A) and the expansion segment of the *COI* gene across 20 species of lepidopteran moths (Additional file 3, Additional file 2: Figure S1B) were constructed and supplied as Additional files. The results provide clues about the evolutionary origin of the phytophagous Noctuidae. The sister group of *H. assulta* and *H. armigera* is clustered on a clade. Some hypotheses on the sister group relatedness based on morphology are concordant with our molecular results.

### Sequencing of DGE-tag libraries and unique tag annotation

DGE-tag profile libraries were constructed from total RNA of *H. armigera* and *H. assulta* for four development stages (embryo, larva, pupa and adult). The summary sequence results are shown in Table 1. Low frequency tags were discounted under the assumption that many could have arisen through sequencing errors such as base substitution, deletion or addition at a single position [14]. Therefore, after eliminating low quality tags (containing Ns), copy numbers less than two and adaptor sequences, the remaining reads were called clean tags, of which more than 50% were singletons (tags with count equal to 1), which is typically observed in SAGE experiments [15]. We obtained approximately 23 million 17nt clean tags from both insect species. Their total unique clean tag (Uni-tag) numbers were also similar at approximately 357,000 (Table 1). Unique tag-to-gene assignments were conducted for the four development stages of *H. armigera* and *H. assulta* using SOAPdenovo program just permitting 1 bp mismatch [16]. On average, more than 75% of the uni-tags of *H. armigera* were mapped on transcripts; however, only 64.5% uni-tags of *H. assulta* mapped on transcripts. The total numbers of transcripts or genes were 268,145 and 230,591 for *H. armigera* and *H. assulta*, respectively, among which the annotated transcripts or genes were 88,857 and 75,157, respectively (Table 1). The Illumina short-reads sequence of *H. armigera* and *H. assulta* were submitted to NCBI Sequence Read Archive under the accession number of SRR628282 and SRR620569, respectively.

Although we obtained similar amounts of total unique clean tags from *H. armigera* and *H. assulta*, the numbers of uni-tags obtained from the different developmental stages in each insect were quite different (Table 1, Figure 2A).



**Figure 1** Main taxonomic characteristics of *H. armigera* and *H. assulta*. (A) (B) (C) and (D) show the phenotypes of the egg, larva, pupa and adult of *H. armigera* (left panel) and *H. assulta* (right panel). (E) The green dashed line shows that the two black hair-base and stigma are in a straight line on the prothorax of *H. armigera* (left panel); the green dashes show that the two black hair-base and stigma are not in a straight line on the prothorax of *H. assulta* (right panel). The red arrows show the stigma of the prothorax. (F) The proleg crochet is in double order for *H. armigera* (left panel) but in single order for *H. assulta* (right panel). (G) Two anal spines of the abdominal end are born on a large black protuberance for the pupae of *H. armigera* (left panel); the two anal spines of the abdominal end do not share the black protuberance in *H. assulta* (right panel). (H) Forewing without white stripes, and the outer edge of the hindwing with a wide brown belt of *H. armigera* (left panel); forewing with a white stripe, and the outer edge of the hindwing with a narrow brown belt of *H. assulta* (right panel). Scale bars = 0.1 mm.

In theory, each uni-tag should be derived from one transcript [14]. According to this theory, the embryo stage has the highest number of transcripts compared with any other stages in the two species (97,646 and 93,672 in *H. armigera* and *H. assulta*, respectively). Unexpectedly, the larval stage of *H. armigera* has the lowest number of transcripts, 80,673 (Table 1, Figure 2A). The gene annotation result indicated that 75.14% and 64.52% uni-tags of *H. armigera* and *H. assulta* corresponded to EST sequences in the *H. armigera* transcriptome library. The number of

identified genes in *H. armigera* is higher than in *H. assulta*; however, the number of identified genes has no significant difference at each developmental stage of the same species (Table 1, Figure 2B).

#### Global analysis of differentially expressed genes between the two species

The unique clean tags provide transcripts information for one species. Using a Venn diagram, developmental stage-specific transcripts and coexpressed transcripts

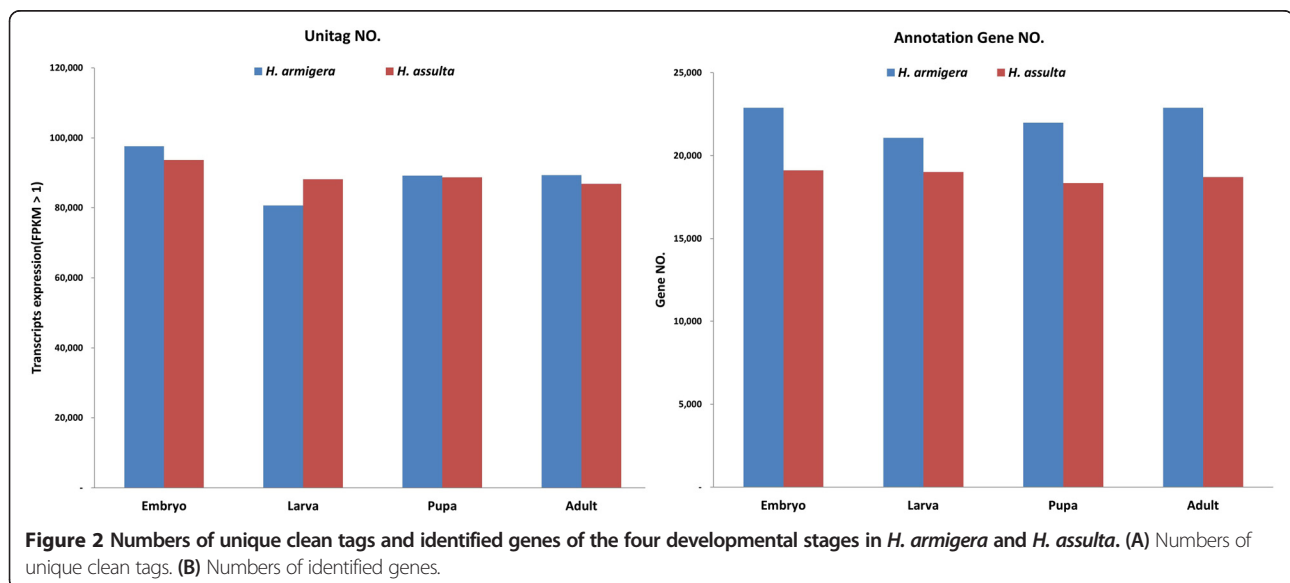
**Table 1 DGE-tag unique clean tags and tags percentage that map to genes in the four developmental stages of *H. armigera* and *H. assulta***

		Embryos	Larva	Pupa	Adult	Total
Raw data	<i>H. armigera</i>	6,202,369	5,766,391	6,126,237	6,021,919	24,116,916
	<i>H. assulta</i>	6,072,324	6,161,560	6,195,186	5,876,709	24,305,779
Clean tags	<i>H. armigera</i>	6,058,338	5,646,953	5,998,625	5,891,808	23,595,724
	<i>H. assulta</i>	5,955,877	6,048,552	6,084,239	5,759,082	23,847,750
Unique clean tags (Uni-tags)	<i>H. armigera</i>	97,646	80,673	89,193	89,330	356,842
	<i>H. assulta</i>	93,672	88,171	88,711	86,860	357,414
Uni-tag mapping to gene	<i>H. armigera</i>	71,623	64,679	64,458	67,385	268,145
	<i>H. assulta</i>	61,518	60,287	55,503	53,283	230,591
Mapping tag ratio (%)	<i>H. armigera</i>	73.35%	80.17%	72.27%	75.43%	75.14%
	<i>H. assulta</i>	65.67%	68.37%	62.57%	61.34%	64.52%
Gene numbers of tag mapping	<i>H. armigera</i>	22,890	21,080	21,987	22,901	88,857
	<i>H. assulta</i>	19,106	19,006	18,340	18,705	75,157
Mapping gene ratio (%)	<i>H. armigera</i>	34.27%	31.56%	32.92%	34.29%	
	<i>H. assulta</i>	28.61%	28.46%	27.46%	28.01%	

between two to four developmental stages were shown between of *H. armigera* and *H. assulta* (Figure 3A, B). The analysis results revealed that the minimum numbers of coexpressed transcripts existed between the larval and adult stage in both *H. armigera* and *H. assulta* (3155 and 3630, respectively). This indicated that the biggest differences exist between these two stages among the four developmental stages. However, in *H. armigera*, the embryo and adult stage are probably “the closest neighbors” and have the most amount of 6703 coexpressed transcripts or genes. In *H. assulta*, the largest number of coexpressed transcripts was 10052 between the embryo and larvae stage. The uni-tag annotation

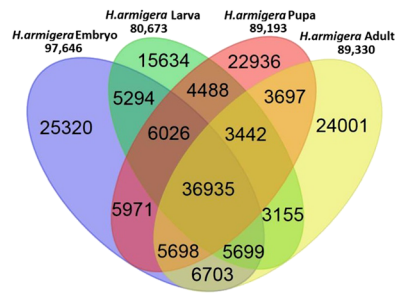
results were also analyzed for differential expression and coexpressed transcripts or genes using a Venn diagram (Figure 3C, D).

The copy number of each unique tag provides quantitative information for the abundance of the transcripts or genes detected by the tags. Using the tag copy number, we can roughly estimate the expression level of each transcripts or gene. The dynamics of gene expression can be reflected by up- or down-regulation among the four developmental stages by pairwise comparisons (Figure 3E, F). Overall, the changes in gene expression levels between two developmental stages in *H. assulta* are more extreme than in *H. armigera*.

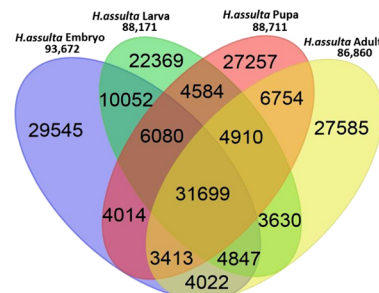




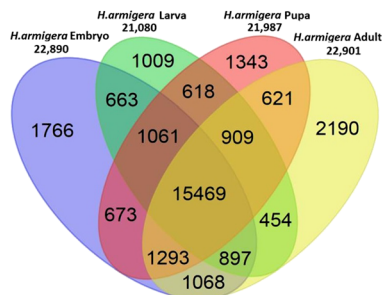
**A** unique clean tags of *H. armigera*



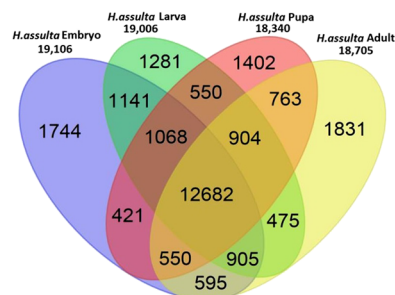
**B** unique clean tags of *H. assulta*



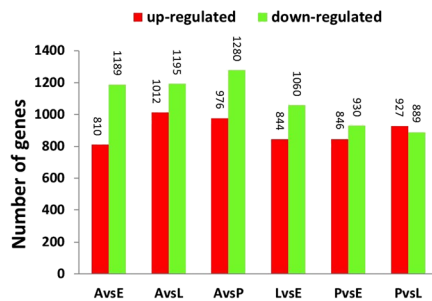
**C** known genes of *H. armigera*



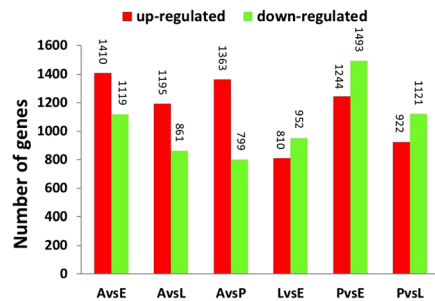
**D** known genes of *H. assulta*



**E** *H. armigera*



**F** *H. assulta*



**Figure 3 Gene expression analysis of the *H. armigera* and *H. assulta*.** (A) and (B): Venn diagram of unique clean tags showing the differential expression or coexpression between pairs developmental stages of *H. armigera* and *H. assulta*. (C) and (D): Venn diagram of known genes showing the differential expression or coexpression between two to four developmental stages of *H. armigera* and *H. assulta*. (E): Up- or downregulated expression of genes in *H. armigera*. (F): Up- or downregulated expression genes in *H. assulta*. E: Embryo stage; L: Larval stage; P: Pupal stage; A: Adult stage.

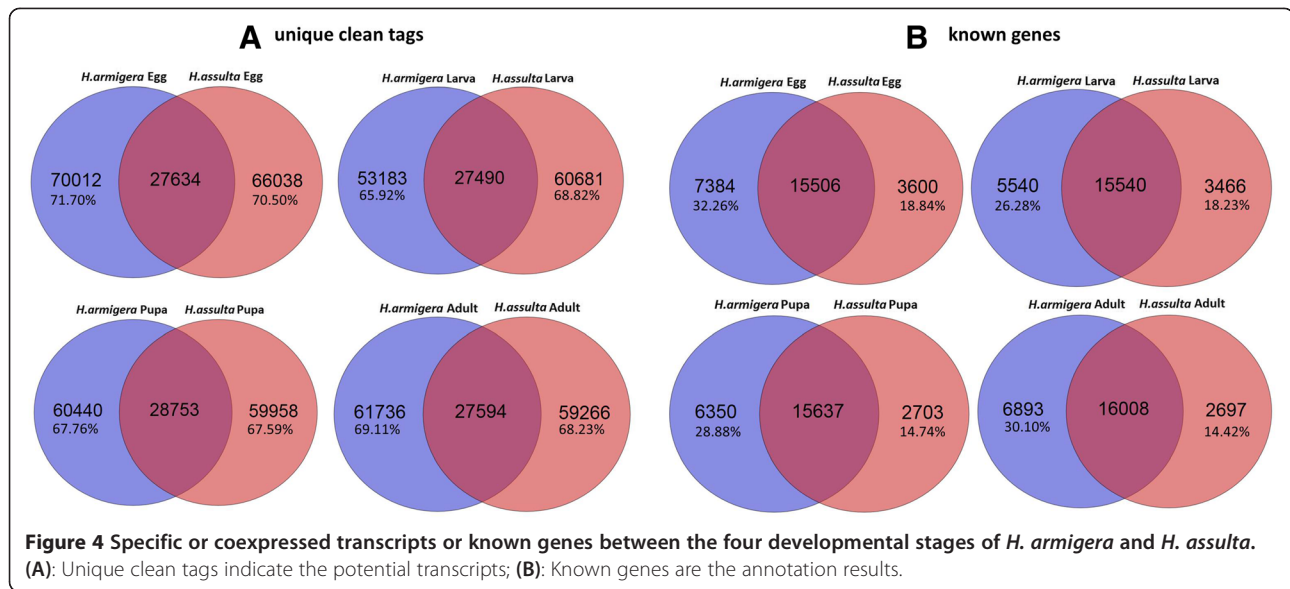
**Coexpressed transcripts or genes between two species**

Comparing the unique clean tags between *H. armigera* and *H. assulta*, approximately 30% transcripts or genes are coexpressed in each developmental stage (Figure 4A, red region of overlap). This reflects the real situation of coexpression transcripts in these two insects. Because there are no whole genome annotation information for the two insect species, most of the differentially or coexpressed genes are unknown proteins, hypothetical proteins, enzymes or cytoskeletal proteins, which are annotated according to the *H. armigera* transcriptome results. In terms of annotated genes, about 67% to 85% genes in *H. armigera* and *H. assulta* are coexpression

during the four developmental stages (Figure 4B, red region of overlap). GO analysis also confirmed that only in the larval stage there were more functional genes in *H. assulta* than in *H. armigera* (Additional file 4).

**Development and host range related transcripts or genes between two insect species**

To explain why the two species have these similarities and differences, we focused on comparing the transcripts or genes that are related to growth and development, food digestion or detoxification enzymes, and host plant recognition. We identified 246 and 240 growth and development related transcripts, 1137 and 1138 transcripts



for food digestion or detoxification related enzymes, and 83 and 68 olfaction related transcripts from *H. armigera* and *H. assulta*, respectively. The relative expression levels (by tag copy numbers) for these transcripts in each developmental stage of the two species and the annotation results are listed in Additional files 5, 6 and 7. In summary, the amounts of each type of transcript are similar between the two species (Figure 5A-C). The biggest difference in the number is odorant binding proteins (OBPs). There are 42 OBP transcripts in *H. armigera*, but only 31 in *H. assulta* (Figure 5B).

#### Expression patterns of possible host-range-related transcripts or genes

The expression patterns of transcripts can be divided into two categories by tag copy number among the four developmental stages between two species (Table 2). The first type is the similar expression pattern. For example, in Table 2, the Trypsin-1 gene just express at the embryo stage in the two species (Table 2, line 1). The Cytochrome C oxidase polypeptide III gene has nearly same high expression levels at all four development stages (Table 2, line 2). More information is shown in Additional files 5, 6, and 7. Genes and transcripts with similar expression patterns were not further analyzed in this study. The other expression pattern is showing a significant difference between two species. For example, the OBP3 gene is a carrier of odor molecules, which can protect odor molecules from enzymatic degradation [17,18]. Thus, OBP3 is likely to be an important gene in host plant selection. In this study, the expression pattern of OBP3 was significantly different among the four development stages between the two insect species (Table 2, line 5 D3). These kinds of transcripts or genes are probably the main reasons

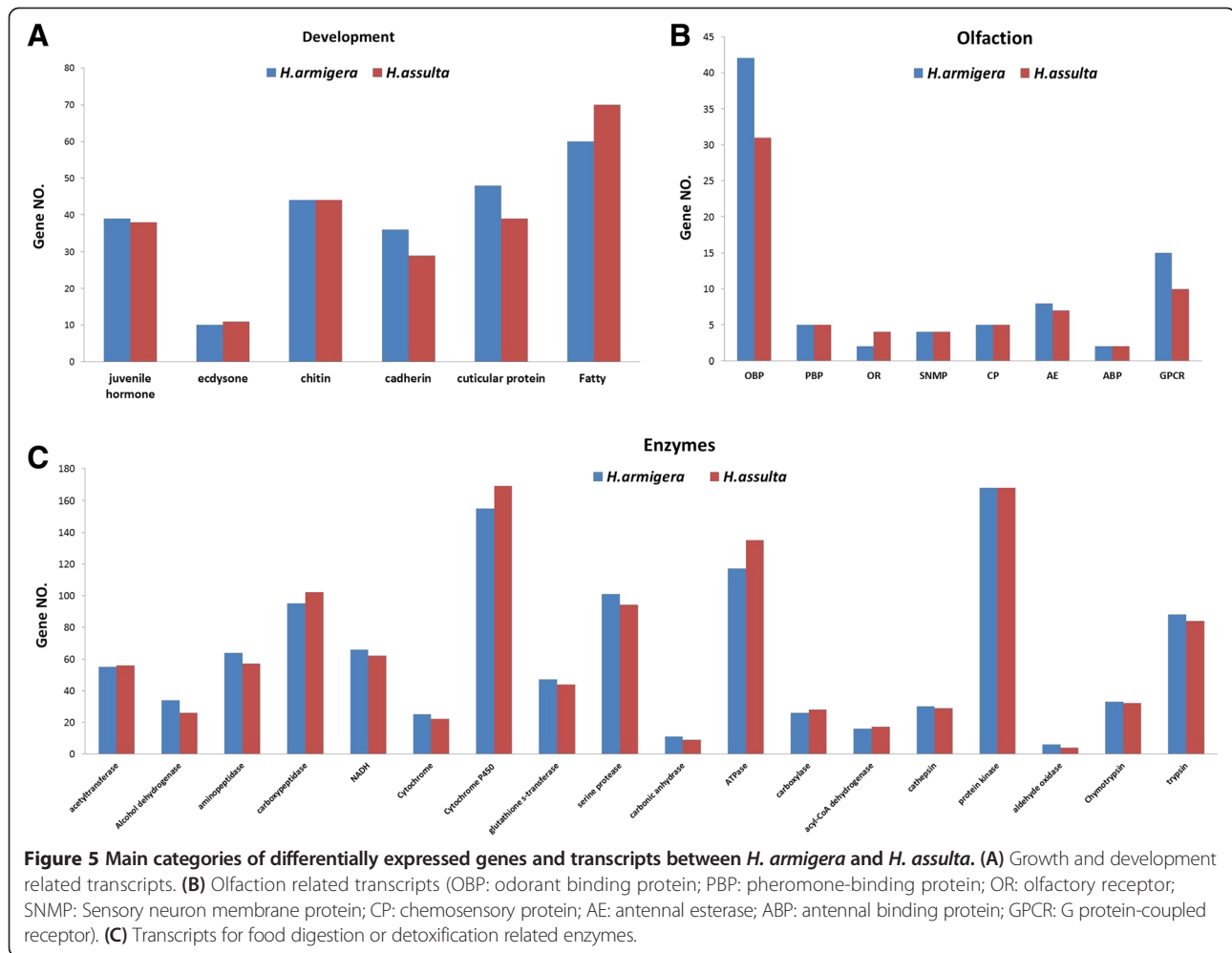
underlying the differences between the two species. The expressions of these types of transcripts or genes were confirmed by reverse transcription polymerase chain reaction (RT-PCR) (Figure 6). Most of expression patterns are nearly consistent with the digital expression tag copy number. These transcripts or genes should be further studied in host selection and adaptation.

#### Conclusions

In this paper, we systematically analyzed the differences and similarities between the two sibling insect species, *H. armigera* and *H. assulta*. These characteristics make the two sibling species fitting for research on host range, evolution and pest control. By comparing the tag copy number of each developmental stage, we identified some differentially expressed transcripts or genes that are probably associated with host plant recognition and food digestion or detoxification. These genes provide important clues for further study.

SAGE is a method of large-scale gene expression analysis [19]. It is an 'open' system that permits the relative expression levels of almost all transcripts in an organism. DGE-tag profiles are a development of this technology using second-generation high throughput sequencing technology [20,21]. The DGE-tag profile results indicated that even between two sibling species, only 30% of transcripts are coexpressed in each developmental stage. The annotation results indicated that 67–85% genes are coexpressed in each developmental stage between the two species (Figure 4). These results further confirm that they are distinct two species at the genome level.

The unique clean tags number can provide quantitative information for the number of transcripts detected by tags. Using traditional SAGE technology, 50,000 to



100,000 tags could be collected, which represent 20,000 to 40,000 unique tags [15,19]. In this study, the high-throughput approach was adopted to implement the tag sequencing protocol on the Illumina platform [13]. Using this technology, we obtained more than 23 million clean tags from *H. armigera* and *H. assulta*, respectively. These data far more exceed the saturation requirements of sequencing [15,22]. The total unique clean tags were 356,842 and 357,414 for each species, which probably represent the total transcripts in the whole life cycle of the two insects. A total of 80,673 to 97,646 transcripts were identified for each developmental stage (Table 1, unique clean tags). These results indicated that the data sets were suitable for analyzing the differential expression of transcripts or genes.

In this study, we found that embryo stage expressed the most transcripts or genes in both species. Unexpectedly, the larval stage expressed the lowest numbers of transcripts in *H. armigera* (Figure 3A). Because the total clean tag number far more exceeded the saturation requirements, the difference is unlikely to be caused by sequencing

bias. The most significant difference between the two insects is in their host ranges, which should be reflected at the larval stage.

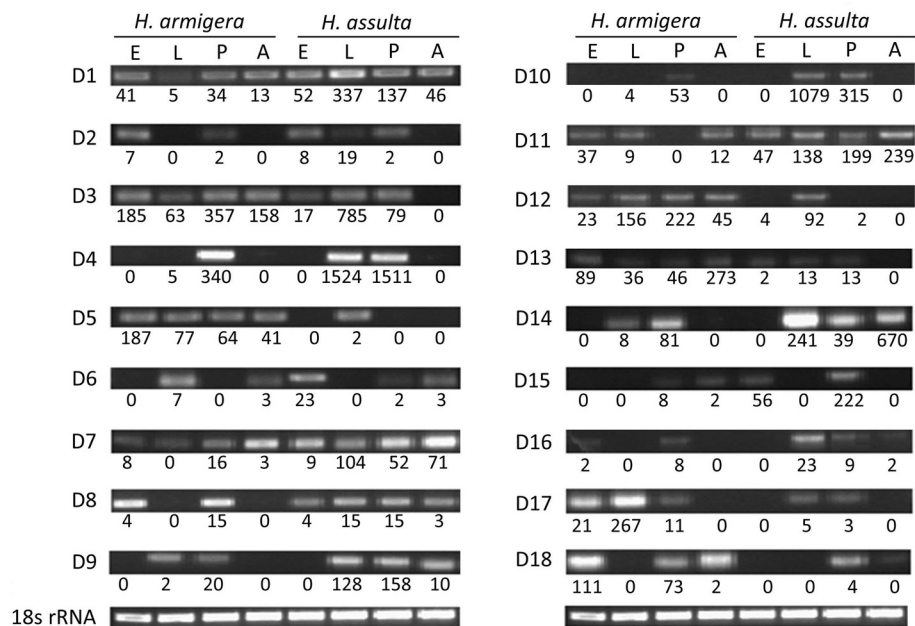
Actually, the selection and adaptation of the host plant is decided by the insect's internal factors and external environmental stimuli. When insects choose a plant as a host, they will be spawning and feeding, and then growth and reproduction on the plant. During this process, as external environmental factors, plant volatile odors are the most important cues for host plant selection. At the same time, as internal factors, insect OBPs not only can selectively bind certain types of odor molecules, but also can remove toxic substances and protect the odor molecules from enzymatic degradation [17,18]. Then, the downstream ORs (odorant receptors) will be activated, the chemical odor molecule information will be converted into an electrical signal, spread to the central nervous system, triggering an insect behavioral response [23-26]. In this research, we identified 42 and 31 OBP-related transcripts or genes from *H. armigera* and *H. assulta*, respectively (Figure 5B, Additional file 6: Table S5). This

**Table 2 Expression patterns of food digestion or detoxification-related enzyme genes or transcripts according to tag copy numbers**

Pattern	Gene ID	seq_id	Embryo		Larva		Pupa		Adult		Annotation results
			<i>H. armigera</i>	<i>H. assulta</i>	<i>H. armigera</i>	<i>H. assulta</i>	<i>H. armigera</i>	<i>H. assulta</i>	<i>H. armigera</i>	<i>H. assulta</i>	
*S1	P35035	HARM050658	74	71	0	0	0	0	0	0	Trypsin-1
S2	Q35826	HARM066614	234233	174746	185711	167262	159977	163095	132374	81008	Cytochrome c oxidase polypeptide III
D1	Q9VYY4	HARM012503	41	52	5	337	34	137	13	46	Cytochrome P450 4 g15
D2	P54191	HARM016746	7	8	0	19	2	2	0	0	Pheromone-binding protein-related protein 1
D3	AEB54582	HARM004409	185	17	63	785	357	79	158	0	OBP3
D4	Q00871	HARM066585	0	0	5	1524	340	1511	0	0	Chymotrypsin BI
D5	P46441	HARM051295	187	0	77	2	64	0	41	0	Putative ATPase N2B
D6	Q9VGG8	HARM002903	0	23	7	0	0	2	3	3	Probable G-protein coupled receptor Mth-like 5
D7	Q11001	HARM001041	8	9	0	104	16	52	3	71	Membrane alanyl aminopeptidase
D8	Q964T2	HARM001432	4	4	0	15	15	15	0	3	Cytochrome P450 9e2
D9	Q9QYZ9	HARM001863	0	0	2	128	20	158	0	10	Serine protease 30
D10	Q0II73	HARM066611	0	0	4	1079	53	315	0	0	Carboxypeptidase O
D11	P62333	HARM025783	37	47	9	138	0	199	12	239	26S protease regulatory subunit 10B
D12	O18598	HARM027201	23	4	156	92	222	2	45	0	Glutathione S-transferase
D13	Q9V675	HARM053132	89	2	36	13	46	13	273	0	Probable cytochrome P450 6 g2
D14	Q9V7G5	HARM039905	0	0	8	241	81	39	0	670	Probable cytochrome P450 4aa1
D15	P46430	HARM034436	0	56	0	0	8	222	2	0	Glutathione S-transferase 1
D16	ACV60230	HARM024407	2	0	0	23	8	9	0	2	antennal esterase CXE3
D17	Q27377	HARM028696	21	0	267	5	11	3	0	0	Putative odorant-binding protein A10
D18	AEX07273	HARM051397	111	3	0	0	73	4	2	0	odorant-binding protein

\* S = similarity; D = difference.





**Figure 6** Reverse transcription PCR results of different expression transcripts or genes. E: Embryo stage; L: Larval stage; P: Pupal stage; A: Adult stage.

is the biggest difference among all the transcripts types. This is probably the main reason why these two sibling species have different host ranges. We also identified two and four OR-related transcripts or genes from *H. armigera* and *H. assulta*, respectively. These transcripts should be further studied to increase our understanding of the host range difference between the two species.

In addition to host selection, the other important aspect is host adaptation. During feeding, insect will inevitably swallow some poisonous secondary metabolites from plants. Therefore, insects have to develop an adaptation mechanism involving a series of detoxification enzymes [27,28]. These detoxification enzymes include cytochrome P450-dependent monooxygenases (P450s), glutathione-S-transferases and carboxylesterases (COEs) [27,29-32]. In this study, we found that there are more transcripts or genes for P450s, COEs and ATPases in *H. assulta* than in *H. armigera* (Figure 5C). GO analysis also confirmed that only in the larval stage there are more functional genes in *H. assulta* than in *H. armigera* (Additional file 4). Therefore, we suspected that because oriental tobacco budworm has a narrower host range, more detoxification enzymes would help them increase food detoxification and utilization efficiency. This should be further investigated in a future study.

The two species are also important agricultural insect pests. Considering all the similarities and differences between the two sibling species, we think they are a good model insect-pair for developing species-selective RNAi technology. Many studies have shown that RNAi is feasible

technology in insect pest control [33-37]. The appeal of RNAi technology in pest control is that it is possible to design the pesticide to target only a single species or a group of related species, with minimal threat to other organisms [38]. To this end, it is necessary to identify species-specific target genes. The present study represents an effective strategy for identifying differentially expressed genes from related species that do not have genome sequences. Using DGE-tag profile technology, we identified many differentially expression transcripts or genes from the two sibling insect species. These genes not only provide clues for host range difference studies, but may also represent important targets for species-specific control by RNAi technology.

## Methods

### Insect culture and sample collection

*H. armigera* and *H. assulta* were originally obtained from Henan Agricultural University and maintained for several generations in our laboratory. They were fed an artificial diet at  $25 \pm 1^\circ\text{C}$  under a light-dark cycle of 14:10 h. Moths were provided with 10% honey solution as food; larvae were fed on a modified artificial diet (wheat germ: 84.0 g; casein: 64.0 g; sucrose: 64.0 g; cellulose: 10.0 g; vitamin C: 10.0 g; Wechsler salt: 10.0 g; choline chloride: 3.0 g; sorbic acid: 3.0 g; nipagin: 3.0 g; agar: 30.0 g; vitamin complex: 8.0 g; vitamin C: 8.0 g; and water: 1000 mL). Thirty-four samples were collected from embryo to adult stages during the whole life cycle of *H. armigera* (1d, 2d, 3d embryos; one to six instar larvae; 1-7d pupae; and 1-9d

adults (male and female separately)). The samples were immediately frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  before RNA extraction.

#### RNA isolation

Total RNA was isolated using a Qiagen RNA Extraction kit according to the manufacturer's instructions. The RNA was treated with RNase-free DNase I for 30 min at  $37^{\circ}\text{C}$  (New England BioLabs) to remove residual DNA. Equivalent amounts of the 34 samples were merged into four pools of embryo, larva, pupa and adult. mRNA was isolated from DNA-free total RNA using a Dynabeads mRNA Purification Kit (Invitrogen).

#### cDNA synthesis

Before cDNA synthesis, 5  $\mu\text{g}$  total RNA was treated with RQ1 RNase-free DNase (Promega), according to the manufacturer's instructions, to ensure no DNA contamination. cDNA synthesis was then carried out with the purified RNA using the SuperScript III First-Strand Synthesis System (Invitrogen), following the manufacturer's instructions. The RT reaction was performed using Mastercycler Gradient (Eppendorf). Briefly, 1  $\mu\text{g}$  RNA, 50  $\mu\text{M}$  oligo dT(20) and 10 mM dNTP mix were added together and incubated at  $65^{\circ}\text{C}$  for 5 min. The samples were then placed on ice for at least 1 min. After that, 2  $\mu\text{l}$  10  $\times$  RT buffer, 1  $\mu\text{l}$  25 mM  $\text{MgCl}_2$ , 2  $\mu\text{l}$  0.1 M DTT, 40 U RNaseOUT and 200 U SuperScript III were added and incubation carried at  $50^{\circ}\text{C}$  for 50 min. The RT reaction was terminated by incubating at  $85^{\circ}\text{C}$  for 5 min and the residual RNA was removed by incubating at  $37^{\circ}\text{C}$  for 20 min with the addition of 1  $\mu\text{l}$  RNaseH. The cDNA was stored at  $-20^{\circ}\text{C}$ .

#### Sequence tag preparation, sequencing and DGE-tag annotation

Sequence tags were prepared with Illumina's Digital Gene Expression Tag Profiling Kit, according to the manufacturer's protocol. A schematic overview of the procedure can be found in reference [39]. We extracted 6  $\mu\text{g}$  total RNA, use Oligo(dT) magnetic beads adsorption to purify mRNA, and then use Oligo(dT) as primer to synthesize the first and second-strand cDNA. The 5' ends of tags can be generated by two types of Endonuclease: NlaIII or DpnII. Usually, the bead-bound cDNA is subsequently digested with restriction enzyme NlaIII, which recognizes and cuts off the CATG sites. The fragments apart from the 3' cDNA fragments connected to Oligo(dT) beads are washed away and the Illumina adaptor 1 is ligated to the sticky 5' end of the digested bead-bound cDNA fragments. The junction of Illumina adaptor 1 and CATG site is the recognition site of MmeI, which is a type of Endonuclease with separated recognition sites and digestion sites. It cuts at 17 bp downstream of the CATG

site, producing tags with adaptor 1. After removing 3' fragments with magnetic beads precipitation, Illumina adaptor 2 is ligated to the 3' ends of tags, acquiring tags with different adaptors of both ends to form a tag library. After 15 cycles of linear PCR amplification, 95 bp fragments are purified by 6% TBE PAGE Gel electrophoresis.

After denaturation, the single-chain molecules are fixed onto the Illumina Sequencing Chip (flowcell). Each molecule grows into a single-molecule cluster sequencing template through Situ amplification. Then add in four types of nucleotides which are labeled by four colors, and perform sequencing with the method of sequencing by synthesis (SBS). Each tunnel will generate millions of raw reads with sequencing length of 35 bp. Image analysis and basecalling were performed using the Illumina Pipeline, where sequence tags were obtained after purity filtering. This was followed by sorting and counting the unique tags.

We filtered out low quality tags (containing Ns), copy number below 2 and adaptor sequences. Ultimately,  $\approx 6$  million clean sequence DGE-tags for each developmental stages of embryo, larva, pupa and adult were obtained. The DGE tags, which consist of the CATG restriction enzyme digested site and an additional 17 bp from each transcript, were *de novo* assembled using SOAPdenovo program just permitting 1 bp mismatch [16]. All of the tags were compared with the reference database of *H. armigera* cDNA library [40-42] and other insect nucleotide sequences (*Bombyx mori*, *Heliothis virescens*, *Spodoptera exigua*, *Prodenia litura* and *Manduca sexta*) from NCBI. The number of tags mapped on a transcript was used as a measure of the abundance of this transcript. The DGE-tag expression level was calculated by the RPKM (Reads Per kb per Million reads) method [43]. Functional annotation by Gene Orthology (GO, <http://www.geneontology.org/>) was run on Blast2go (<http://www.blast2go.org/>) [44]. The COG and KEGG pathway annotation were performed using Blastall online program against the Cluster of Orthologous Groups of proteins (COG, [www.ncbi.nlm.gov/COG](http://www.ncbi.nlm.gov/COG)) and Kyoto Encyclopedia of Genes and Genomes (KEGG, [www.genome.jp/kegg](http://www.genome.jp/kegg)) databases, respectively. Briefly, sequences were searched against GenBank non-redundant database (Nr) with BLASTx algorithm [44]. The blast results were mapped to gene ontology terms and annotation was carried out using default annotation parameters in the Blast2Go software suit [44-46]. For further functional annotation, the KEGG mapping was carried out in Blast2Go.

#### Semi-quantitative RT-PCR

The cDNA was amplified in a 50- $\mu\text{l}$  reaction mixture containing 50 mM KCl, 10 mM Tris-HCl, 2 mM  $\text{MgCl}_2$ , 0.2 mM of each deoxyribonucleoside triphosphates, 0.4  $\mu\text{M}$  primers (Additional file 8), and 1 U *Ex Taq* polymerase

(TaKaRa Biotech) using hot-start PCR. The PCR reaction conditions are shown in Additional file 8. To confirm that the products were fragments of the target genes, the Beijing Genomics Institute sequenced the PCR products using the same set of primers. Furthermore, cDNA samples with the strongest amplification were serially diluted, and a close correlation between the amount of product and initial cDNA was seen after PCR analysis. Ten microliters of each PCR product was visualized by electrophoresis on a 1% agarose gel stained with ethidium bromide.

## Additional files

- Additional file 1: Table S1.** Sequence name of the 18S Ribosomal RNA Genes.
- Additional file 2: Figure S1.** Phylogenetic tree of insects from 18S rRNA and the Col gene to show the relationship of *H. armigera* and *H. assulta*.
- Additional file 3: Table S2.** COI gene sequences name and gene ID.
- Additional file 4: Table S3.** Growth and development related genes and transcripts.
- Additional file 5: Table S4.** Olfaction related genes and transcripts.
- Additional file 6: Table S5.** Food digestion or detoxification related enzymes.
- Additional file 7: Table S6.** GO analysis results for all of the transcripts between the two species.
- Additional file 8: Table S7.** Primer pairs for RT-PCR.

## Abbreviations

DGE-Tag: Digital gene expression tag profile; SAGE: Serial analysis of gene expression; Uni-tag: Unique clean tag; EST: Expressed sequence tag; NCBI: National center for biotechnology information; GO: Gene ontology; SQ-RT-PCR: Semi-quantitative reverse transcriptase polymerase chain reaction; ORs: Odorant receptors; OBPs: Odorant-binding proteins; P450s: Cytochrome P450-dependent monooxygenases; GSTs: Glutathione-S-transferases; COEs: Carboxylesterases.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

HL analyzed data; HZ and RG collected the samples, designed the primers and performed RT-PCR test; XM and HL designed the experiments and wrote the manuscript. All authors read and approved the final manuscript.

## Acknowledgments

This work was supported by the National Transgenic Great Subject from the Ministry of Agriculture of China (2011ZX08009-003-001), the grants from NSFC (31172152) and CAS (KZSD-EW-Z-021-2-1). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Author details

<sup>1</sup>Key Laboratory of Insect Developmental and Evolutionary Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China. <sup>2</sup>University of Chinese Academy of Sciences, Beijing 100049, China.

Received: 11 March 2013 Accepted: 14 August 2013

Published: 28 August 2013

## References

1. Fitt GP: The ecology of heliothis species in relation to agroecosystems. *Annu Rev Entomol* 1989, **34**:17–52.

2. Teng ZQ, He Q, Li HT, Zhang QW: Interspecific and intraspecific comparisons of ejaculates in the cotton bollworm *Helicoverpa armigera* and the tobacco budworm *H. assulta*. *J Ethol* 2009, **27**:19–24.
3. Wang CZ, Dong JF: Interspecific hybridization of *Helicoverpa armigera* and *H. assulta* (Lepidoptera: Noctuidae). *Chinese Sci Bull* 2001, **46**:489–491.
4. Tang QB, Jiang JW, Yan YH, van Loon JJA, Wang CZ: Genetic analysis of larval host-plant preference in two sibling species of *Helicoverpa*. *Entomol Exp Appl* 2006, **118**:221–228.
5. Wang HL, Zhao CH, Wang CZ: Comparative study of sex pheromone composition and biosynthesis in *Helicoverpa armigera*, *H. assulta* and their hybrid. *Insect Biochem Mol Biol* 2005, **35**:575–583.
6. Wu H, Hou C, Huang LQ, Yan FS, Wang CZ: Peripheral coding of sex pheromone blends with reverse ratios in two *helicoverpa* species. *Plos One* 2013, **8**(7):e70078.
7. Zhao XC, Yan YH, Wang CZ: Behavioral and electrophysiological responses of *Helicoverpa assulta*, *H. Armigera* (Lepidoptera: Noctuidae), their F1 hybrids and backcross progenies to sex pheromone component blends. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 2006, **192**:1037–1047.
8. Anonymous: *Heliothis assulta*. *Distribution Maps of Plant Pests*. Map. No. 262. Wallingford, UK: CAB International; 1969.
9. Liu ZD, Li DM, Gong PY, Wu KJ: Life table studies of the cotton bollworm, *Helicoverpa armigera* (Hübner) (Lepidoptera: Noctuidae), on different host plants. *Environ Entomol* 2004, **33**:1570–1576.
10. Krieger J, Raming K, Dewar YME, Bette S, Conzelmann S, Breer H: A divergent gene family encoding candidate olfactory receptors of the moth *Heliothis virescens*. *Eur J Neurosci* 2002, **16**:619–628.
11. Xu P, Atkinson R, Jones DN, Smith DP: *Drosophila* OBP LUSH is required for activity of pheromone-sensitive neurons. *Neuron* 2005, **45**:193–200.
12. Ghosh T, Soni K, Scaria V, Halimani M, Bhattacharjee C, Pillai B: MicroRNA-mediated up-regulation of an alternatively polyadenylated variant of the mouse cytoplasmic  $\beta$ -actin gene. *Nucleic Acids Res* 2008, **36**:6318–6332.
13. Morrissy AS, Morin RD, Delaney A, Zeng T, McDonald H, Jones S, Zhao Y, Hirst M, Marra MA: Next-generation tag sequencing for cancer gene expression profiling. *Genome Res* 2009, **19**:1825–1835.
14. Popesco MC, Lin SL, Wang ZL, Ma ZXJ, Friedman L, Frostholm A, Rotter A: Serial analysis of gene expression profiles of adult and aged mouse cerebellum. *Neurobiol Aging* 2008, **29**:774–788.
15. Wang SM: Understanding SAGE data. *Trends Genet* 2007, **23**:42–50.
16. Li R, Li Y, Kristiansen K, Wang J: SOAP: short oligonucleotide alignment program. *Bioinformatics* 2008, **24**:713–714.
17. Kaissling KE, Rospars JP: Dose-response relationships in an olfactory flux detector model revisited. *Chem Senses* 2004, **29**:529–531.
18. Steinbrecht RA: Odorant-binding proteins: expression and function. *Ann N Y Acad Sci* 1998, **855**:323–332.
19. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW: Serial analysis of gene-expression. *Science* 1995, **270**:484–487.
20. Asmann YW, Klee EW, Thompson EA, Perez EA, Middha S, Oberg AL, Therneau TM, Smith DI, Poland GA, Wieben ED, Kocher JPA: 3' tag digital gene expression profiling of human brain and universal reference RNA using illumina genome analyzer. *Bmc Genomics* 2009, **10**:531.
21. Wang Z, Dong D, Ru B, Young RL, Han N, Guo T, Zhang S: Digital gene expression tag profiling of bat digits provides robust candidates contributing to wing formation. *Bmc Genomics* 2010, **11**:619.
22. Pleasance ED, Marra MA, Jones SJ: Assessment of SAGE in transcript identification. *Genome Res* 2003, **13**:1203–1215.
23. Pophof B: Pheromone-binding proteins contribute to the activation of olfactory receptor neurons in the silkmoths *antheraea polyphemus* and *Bombyx mori*. *Chem Senses* 2004, **29**:117–125.
24. Jacquín-Joly E, Merlin C: Insect olfactory receptors: contributions of molecular biology to chemical ecology. *J Chem Ecol* 2004, **30**:2359–2397.
25. Wicher D, Schafer R, Bauernfeind R, Stensmyr MC, Heller R, Heinemann SH, Hansson BS: *Drosophila* odorant receptors are both ligand-gated and cyclic-nucleotide-activated cation channels. *Nature* 2008, **452**:1007–1011.
26. Grosse-Wilde E, Kuebler LS, Bucks S, Vogel H, Wicher D, Hansson BS: Antennal transcriptome of *Manduca sexta*. *Proc Natl Acad Sci U S A* 2011, **108**:7449–7454.
27. Tao XY, Xue XY, Huang YP, Chen XY, Mao YB: Gossypol-enhanced P450 gene pool contributes to cotton bollworm tolerance to a pyrethroid insecticide. *Mol Ecol* 2012, **21**:4371–4385.
28. Klot A, Ghanim M: Fitness costs associated with insecticide resistance. *Pest Manag Sci* 2012, **68**:1431–1437.

29. Grogan G: **Cytochromes P450: exploiting diversity and enabling application as biocatalysts.** *Curr Opin Chem Biol* 2011, **15**:241–248.
30. Zhang H, Tian W, Zhao J, Jin L, Yang J, Liu C, Yang Y, Wu S, Wu K, Cui J, et al: **Diverse genetic basis of field-evolved resistance to *Bt* cotton in cotton bollworm from China.** *Proc Natl Acad Sci U S A* 2012, **109**:10275–10280.
31. Li X, Schuler MA, Berenbaum MR: **Molecular mechanisms of metabolic resistance to synthetic and natural xenobiotics.** *Annu Rev Entomol* 2007, **52**:231–253.
32. Oakeshott P, Hunt GM: **Long-term outcome in open spina bifida.** *Br J Gen Pract* 2003, **53**:632–636.
33. Zhang H, Li HC, Miao XX: **Feasibility, limitation and possible solutions of RNAi-based technology for insect pest control.** *Insect Sci* 2012, **00**:1–16.
34. Wang YB, Zhang H, Li HC, Miao XX: **Second-generation sequencing supply an effective Way to screen RNAi targets in large scale for potential application in pest insect control.** *PLoS One* 2011, **6**:e18644.
35. Price DR, Gatehouse JA: **RNAi-mediated crop protection against insects.** *Trends Biotechnol* 2008, **26**:393–400.
36. Mao YB, Cai WJ, Wang JW, Hong GJ, Tao XY, Wang LJ, Huang YP, Chen XY: **Silencing a cotton bollworm P450 monooxygenase gene by plant-mediated RNAi impairs larval tolerance of gossypol.** *Nat Biotechnol* 2007, **25**:1307–1313.
37. Baum JA, Bogaert T, Clinton W, Heck GR, Feldmann P, Ilagan O, Johnson S, Plaetinck G, Munyikwa T, Pleau M, et al: **Control of coleopteran insect pests through RNA interference.** *Nat Biotechnol* 2007, **25**:1322–1326.
38. Whyard S, Singh AD, Wong S: **Ingested double-stranded RNAs can act as species-specific insecticides.** *Insect Biochem Mol Biol* 2009, **39**:824–832.
39. 't Hoen PAC, Ariyurek Y, Thygesen HH, Vreugdenhil E, Vossen RHAM, de Menezes RX, Boer JM, van Ommen GJB, den Dunnen JT: **Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms.** *Nucleic Acids Res* 2008, **36**(21):e141.
40. Zhang Q, Lu YX, Xu WH: **Integrated proteomic and metabolomic analysis of larval brain associated with diapause induction and preparation in the cotton bollworm, *Helicoverpa armigera*.** *J Proteome Res* 2012, **11**:1042–1053.
41. Liu Y, Gu S, Zhang Y, Guo Y, Wang G: **Candidate olfaction genes identified within the *Helicoverpa armigera* antennal transcriptome.** *PLoS One* 2012, **7**:e48260.
42. Celorio-Mancera Mde L, Courtiade J, Muck A, Heckel DG, Musser RO, Vogel H: **Sialome of a generalist lepidopteran herbivore: identification of transcripts and proteins from *Helicoverpa armigera* labial salivary glands.** *PLoS One* 2011, **6**:e26676.
43. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B: **Mapping and quantifying mammalian transcriptomes by RNA-Seq.** *Nat Methods* 2008, **5**:621–628.
44. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M: **Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research.** *Bioinformatics* 2005, **21**:3674–3676.
45. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**:403–410.
46. Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M, Dopazo J, Conesa A: **High-throughput functional annotation and data mining with the Blast2GO suite.** *Nucleic Acids Res* 2008, **36**:3420–3435.

doi:10.1186/1471-2164-14-582

**Cite this article as:** Li et al.: Identification of differential expression genes associated with host selection and adaptation between two sibling insect species by transcriptional profile analysis. *BMC Genomics* 2013 **14**:582.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

