Comparative Biochemistry and Physiology, Part D 16 (2015) 20–27

Contents lists available at ScienceDirect

Comparative Biochemistry and Physiology, Part D

journal homepage: www.elsevier.com/locate/cbpd

De novo sequencing and characterization of the *Bradysia odoriphaga* (Diptera: Sciaridae) larval transcriptome

Haoliang Chen, Lulu Lin, Minghui Xie, Guangling Zhang, Weihua Su *

Institute of Plant Protection and Agro-Products Safety, Anhui Academy of Agricultural Sciences, Hefei 230031, China

ARTICLE INFO

Article history:
Received 14 February 2015
Received in revised form 11 July 2015
Accepted 11 July 2015
Available online 17 July 2015

Key words:
Bradysia odoriphaga
De novo assembly
Functional annotation
Next generation sequencing
Transcriptome

ABSTRACT

The most serious pestilent threat to the Chinese chive, *Allium tuberosum* Rottle ex Spreng (Liliaceae) is the *Bradysia odoriphaga* Yang and Zhang. There is limited genetic research focused on *B. odoriphaga*, partially due to the lack of genomic resources. The advent of high-throughput sequencing technologies has enabled generation of genomic resources in a short time frame and at minimal costs. In this study, we performed, for the first time, *de novo* transcriptome sequencing of the *B. odoriphaga*. Here, 16,829 unigenes were assembled from the total reads, 12,024 of these unigenes were annotated in the NCBI NR protein database, and 9784 were annotated in the Swiss-Prot database. Of these annotated unigenes, 7903 and 5060 unigenes have been assigned to gene ontology categories and clusters of orthologous groups, respectively. Furthermore, 8647 unigenes were mapped to 257 pathways using the Kyoto Encyclopedia of Genes and Genomes Pathway database. We found that 408 unigenes were related to insecticide resistance and metabolism. In addition, 23,122 simple sequence repeats (SSRs) were identified in 11,009 unigenes, and 100 PCR primers of SSR loci were used to validate the assembly quality and polymorphisms. These results provide a good platform for further investigations into the insecticide resistance of *B. odoriphaga*. Finally, the SSRs identified in *B. odoriphaga* may be a useful genomic resource.

1. Introduction

The Chinese chive maggot, *Bradysia odoriphaga* Yang & Zhang (Diptera: Sciaridae), is the most harmful pest to the Chinese chive, *Allium tuberosum* Rottle ex Spreng (Liliaceae) crop (Li et al., 2007). Other plants, such as shallot (*Allium ascalonicum*), garlic (*Allium sativum*), cabbage (*Brassica oleracea*), radish (*Raphanus sativus*), celery (*Apium graveolens*), and mushrooms (*Pleurotus ostreatus*) are also negatively impacted by this pest (Mei et al., 2003). It can cause > 50% crop loss in Chinese chive and in the absence of insecticide treatments it can destroy the entire crop (Dang et al., 2001). It is difficult to control this pest due to its cryptic life style and its ability to rapidly evolve, developing resistance to insecticides (Gao et al., 2000). Aggressive application of chemical insecticides (e.g., organophosphates, carbamates, neonicotinoids) is the most prevalent *B. odoriphaga* prevention method used on Chinese chive crops in China. This has led to an excessive use of insecticides, resulting in increased environmental pollution and high residues of insecticides in commercial chives crops (Wang et al., 2006; Li et al., 2007).

Determination of the genetic pathways and specific genes involved in the *B. odoriphaga* detoxification pathways could be beneficial for controlling the organism. There is limited genomic information for the *B. odoriphaga* and sparse research has been conducted in this area, however. Until now, minimal genomic data have existed on the *Bradysia* genus. Only two expressed sequence tags (EST) from *Bradysia coprophila* has been deposited in the NCBI EST database (Lee et al., 2010; Van der Nest et al., 2011). Moreover, a limited number of DNA and RNA sequences have been published in the NCBI database: 404 nucleotides from 31 species of the *Bradysia* genus. However, there is no information regarding *B. odoriphaga*.

Today, next generation sequencing (NGS) technologies, including the Illumina Solexa, Roche 454, and ABI SOLiD platforms, can provide genomic and transcriptomic data cheaply and rapidly (Schuster, 2008; Ansorge, 2009; Metzker, 2010). Because of these advantages, NGS has been used in many areas of research, such as resequencing, small RNA expression, DNA methylation, and *de novo* transcriptome (RNA-Seq) of non-model organisms (Chen et al., 2010; Crawford et al., 2010; Huang et al., 2010; Mikheyev et al., 2010; Nobuta et al., 2010; Wang et al., 2010; Feldmeyer et al., 2011; Zeng et al., 2013). Recently, Illumina RNA-seqs have been used for efficient assembly of *de novo* transcriptomes for many insects, such as *Bemisia tabaci* (Wang et al., 2010), *Anopheles funestus* (Crawford et al., 2010), and *Locusta migratoria* (Chen et al., 2010). Transcriptomic information is also being used in gene discovery and SSR mining (Wang et al., 2010; Li et al., 2012; Zhang et al., 2012; Chen et al., 2014). These studies indicated that Illumina deep sequencing technology can produce exceptional coverage of expressed sequences.

In this study, we performed *de novo* transcriptome sequencing using the Illumina NGS platform, to determine the *B. odoriphaga* larval transcriptome. The unigenes obtained here were annotated by BLASTX, using public databases. Next, the putative functions of the unigenes

* Corresponding author. Tel.: +86 551 6516 0850.
E-mail address: suwh850@163.com (W. Su).

http://dx.doi.org/10.1016/j.cbd.2015.07.001
1744-117X/

Table 1
Summary of the *Bradysia odoriphaga* transcriptome.

Total raw reads	71,682,542
Total clean reads	65,875,094
Total clean nucleotides (nt)	5,928,758,460
Q20 percentage	97.84%
N percentage	0.00%
GC percentage	43.09%
Total number of contigs	40,922
Mean length of contigs	294
N50 of contigs	446
Total number of unigenes	16,829
Mean length of unigenes	573
N50 of unigenes	762
Distinct clusters	2124
Distinct singletons	14,705

were categorized using Gene Ontology (GO), Cluster of Orthologous Groups (COG), and grouped into pathways using the Kyoto Encyclopedia of Genes and Genomes (KEGG). In addition, unigenes related to insecticide resistance and metabolisms were identified. Finally, SSRs in unigenes were identified and 100 primers were designed to validate the assembly quality and predict polymorphisms.

2. Material and methods

2.1. Insect rearing

The culture of Chinese chive maggots (*B. odoriphaga*) used in this study was obtained from Bozhou, Anhui Province, China (115.90E, 34.00N), and maintained in Petri dishes (9 cm), on Chinese chive stems. In order to keep moisture, a 2% agar gel was used to make a bottom layer in the Petri dishes. After the gel solidified, one piece of fitted filter paper was placed on top of the gel. Chive stems were cut into short pieces, placed in a plastic bowl (9.0 cm diameter at the bottom, 14.0 cm diameter at the top, and 8.0 cm high), and several pairs of adults maggots were introduced into the bowl using an aspirator with a tube. The bowl was then covered with a lid. The adults laid eggs on the chive stems, and the stems were replaced every 48 h. The chive stems with eggs were introduced into Petri dishes containing agar gel and filter paper, and incubated for 7–10 days. After incubation, a moist brush was used to transfer the larvae into a new Petri dish with agar gel, filter paper, and chive stems. The larvae were transferred to a new Petri dish, with agar gel, filter paper, and chive stems, every 3–4 days until they pupated. Petri dishes were placed in a walk-in environmental chamber at 25 ± 1 °C, with a photoperiod of 14 h light:10 h darkness, and 70 ± 10% relative humidity (Mei et al., 2004). Ten 3rd instar larvae were collected and here considered one sample. They were subjected to RNA extraction after removal of the gut to prevent contamination from any ingested plants or gut microorganisms.

Table 2
Number of unigenes annotated using different databases.

Database	NR	NT	Swiss-Prot	KEGG	COG	GO	ALL
Number of annotated unigenes	12,024	4171	9784	8647	5060	7903	12,480

2.2. RNA isolation and cDNA library preparation

An RNA isolation system (SV Total RNA Isolation System Kit, Promega Corporation, Madison, WI, USA) was used for total RNA isolation, following the manufacturer's protocol. Next, samples were treated with deoxyribonuclease (Dnase I: Fermentas Inc., Burlington, ON, Canada) to remove possible residual genomic DNA. A 2100 Bioanalyzer (Agilent Technologies) was used to determine RNA quality. A sample RIN (RNA integrated number) value of 7.0 was chosen for transcriptome analysis, using the Illumina's standard procedure. Briefly, oligo (dT) magnetic beads were used with mRNA purified from 6 μg of total RNA. In a thermomixer, set at 95 °C, divalent cations were used to fragment purified mRNA into small pieces. First strand cDNA synthesis was performed using the small RNA fragments, reverse transcriptase, and random primers. Next, DNA polymerase I and RNaseH were used for second strand cDNA synthesis. The cDNA fragments first underwent end repair and then adapters were ligated to the cDNA fragments. Finally, PCR was performed on the products, approximately 200 bp in size, to create a cDNA library.

2.3. Illumina sequencing and de novo assembly

An Illumina HiSeq™ 2000 instrument was used to sequence the cDNA. The cDNA library sample size was approximately 200 bp, and both ends were sequenced. Adaptor sequences, reads containing more than 5% unknown nucleotides, reads containing more than 50% bases with Q-value ≤ 20, and empty reads were removed to obtain clean reads. Next, *de novo* assembly was performed using Trinity v2.0.2 software to generate unigenes (Grabherr et al., 2011). The K-mer of 25 bp was chosen. This K-mer was chosen with two things in mind. First, large K-mers can result in low sequencing depth at poor overlap regions. Second, de Bruijn graphs are highly complex when too small a K-mer is used. The raw data from Illumina deep-sequencing were

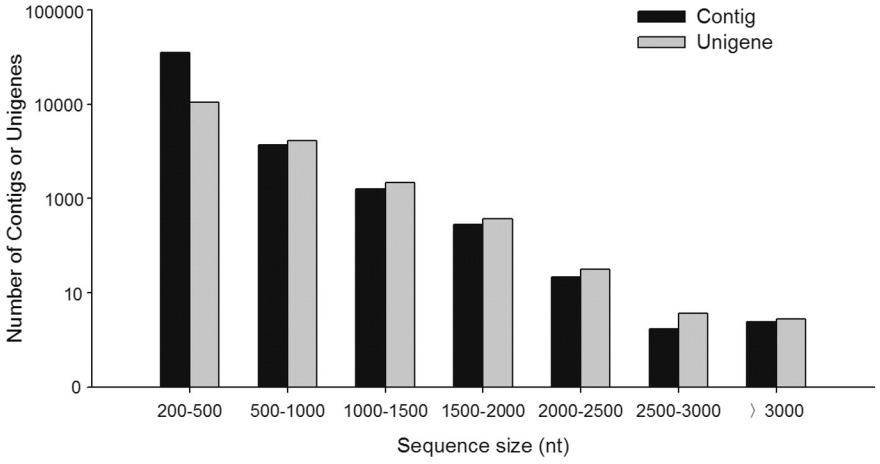

Fig. 1. Contig and unigene length distributions. The x-axis indicates the length of the contig or unigene and the y-axis indicates the number of contigs or unigenes identified.