Journal of Asia-Pacific Entomology 21 (2018) 58–68

Contents lists available at ScienceDirect

Journal of Asia-Pacific Entomology

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

Coding and long non-coding RNAs regulating adult migratory locust (*Locusta migratoria*) brain polyphenism revealed via whole transcriptome analyses

Gwan Seok Lee[a,1], Phuong Nguyen[b,1], Ah-Young Choi[c], A-Young Kim[d], Yeisoo Yu[c], Jin Kyo Jung[e,*], Young Ho Koh[b,d,**]

[a] Department of Plant Protection, National Academy of Agricultural Sciences, Jeollabuk-do, Republic of Korea
[b] Department of Bio-medical Gerontology, Graduate School, Hallym University, Chuncheon, Gangwon-do, Republic of Korea
[c] Phyzen Genomics Institute, Phyzen Inc., Seoul, Republic of Korea
[d] Ilsong Institute of Life Science, Hallym University, Anyang, Gyeonggi-do, Republic of Korea
[e] Department of Central Area, National Institute of Crop Science, Gyeonggi-do, Republic of Korea

ARTICLE INFO

Keywords:
Whole transcriptome analysis
Locusts
Polyphenisms
Long non-coding RNAs

ABSTRACT

The migratory locust, *Locust migratoria* is the most widely distributed and destructive grasshopper in the world. A draft whole genome of this pest were recently reported. However, there have been no reports of coding and long-non coding (LNC) RNAs, whose expression varies in adult locust brains depending on different habitats, phases, or sexes. Thus, an RNA−Seq analysis using 12 RNA seq data sets was performed in the present study. A total of 10.973 LNC and 22,845 coding RNAs were assembled based on 12 transcriptome datasets. Among these RNAs, the number of differentially expressed coding RNAs between phases and sexes ranged from 43 to 269 and 286–385, respectively. In addition, the number of differentially expressed LNC RNAs between phases and sexes were ranged from 19 to 57 and 109–173, respectively. Gene Ontology enrichment analysis and KEGG pathway analyses showed that metabolic pathways, such as purine, pyruvate, inositol phosphate and amino acids metabolism were significantly altered according to phases and sexes. Furthermore, the expression of 7 coding and 3 LNC RNAs was verified using real-time quantitative reverse transcription PCR. In this study, we identified habitat-, phase- and sex-specific coding and LNC RNAs. We further confirmed and analyzed 7 coding RNAs that have been suggested to be involved in inositol metabolisms, the regulation of micro RNA expression and juvenile hormone synthesis and the induction of caste development and adult behavioral changes. Further studies using LNC-RNAs would provide new insights into the molecular and cellular basis of polyphenisms in the migratory locusts.

Introduction

The most widely distributed grasshopper species in the world is the migratory locust, *Locusta migratoria*. The distribution area of the migratory locusts includes for most of the temperate and tropical regions of the eastern hemisphere, covering the distribution areas of the diverse feeding flora upon which it feeds (Chapuis et al., 2008; Ma et al., 2012). This broad distribution area of the migratory locusts suggests that it has the ability to adapt to different climates and vegetation in different regions, resulting in various subspecies. Recent phylo-genetics and -geographic analyses based on mitochondrial genomes and several genes from the migratory locusts collected from the main distribution regions in Africa, Europe, Asia and Oceania have shown that they can be divided into the southern and northern lineages, which were predicted to have separated 895,000 years ago from ancestral populations in Africa (Ma et al., 2012).

The ability of the migratory locust to adapt to various climatic conditions of the migratory locusts may be partially explained by its polyphenism, including changes in its morphology, behavior, physiology, and others phenotypes depending on developmental conditions, without altering genetic information (Ernst et al., 2015). When their habitat conditions are favorable, the migratory locust occurs in its solitary phage, but when the habitat environment is unfavorable, it switches to its gregarious phage and migrates over long distance (Ernst et al., 2015; Ma et al., 2012). One of the most important known factors in phase transition in the migratory locust is its population density.

* Corresponding author.
** Correspondence to: Y. H. Koh, Department of Bio-medical Gerontology, Graduate School, Hallym University, Chuncheon, Gangwon-do, Republic of Korea.
E-mail addresses: jungjk@korea.kr (J.K. Jung), kohyh@hallym.ac.kr (Y.H. Koh).
[1] Those authors equally contributed.

https://doi.org/10.1016/j.aspen.2017.11.007
Received 17 August 2017; Received in revised form 3 November 2017; Accepted 7 November 2017
Available online 07 November 2017

Gregarious locusts show morphological and physiological characteristics that are suitable for long-distance migration, such as an increased ratio between forewing length and hind femur length or an decreased ratio between hind femur length and head width, along with relatively small body sizes compared with those of solitary locusts (Yamagishi and Tanake, 2009). The migratory locust has been considered the most destructive pest worldwide, since it devours herbaceous plants while migrating long distances (Ma et al., 2012; Yamagishi and Tanake, 2009).

Recently, a draft of the whole-genome sequences of the migratory locusts was reported, providing genomic information that could be valuable for promoting various related studies (Wang et al., 2014). The draft whole genome size was 6.5 Gigabytes, with a total of 17,307 gene models being annotated. The larger genome size of the locust compared with that of other insects was due in part to the proliferation of diverse repetitive elements, including transposable elements, intron size expansion caused by transposable element (TE) insertions, and expansion of gene families related to metabolism and detoxification (Wang et al., 2014).

Polyphenism is known to be found in various insects and plays a pivotal role in adaptation to the environment (Ernst et al., 2015). Since polyphenism does not require an alteration of genetic information, it is highly likely to be controlled by epigenetics that regulate the expression of certain sets of genes (Ernst et al., 2015). Epigenetic regulation mechanisms include genomic DNA methylation, histone protein modification, nucleosome positioning and regulation by long non-coding (LNC) RNAs (Ernst et al., 2015). Especially. LNC-RNAs known to regulate expression of mRNAs by binding to its complementary mRNAs, forming complexes with protein translational machineries, or other unknown mechanisms in insects (Ernst et al., 2015).

Analyses of genomic DNA methylation associated with polyphenism have been reported in the migratory locust. The DNA methylation machinery is highly conserved in the migratory locusts, and significant expression differences have been reported between solitary- and gregarious-phage embryos (Robinson et al., 2016). In addition, analysis of the third-instar nymph methylome suggested that various signal transduction pathways, including microtubule dynamics-mediated synaptic plasticity, are involved in phase transitions (Wang et al., 2014). However, the involvement of LNC RNAs in phase transition in the migratory locust remained to be investigated. In this study, we performed whole-transcriptome analysis of migratory locusts in certain habitats, phases and sexes to elucidate differentially expressed coding and LNC RNAs in adult brains and found that many of coding and LNC RNAs were differentially expressed according to habitats, phases and sexes.

Material and methods

Rearing of migratory locust laboratory strains and capture methods

The laboratory-raised strains were established using *Locust migratoria* collected from Haenam, Jeonllanam-do in 2014 and were maintained at 30 ± 1 ˚C, under a 16-h light/8-h dark photoperiod, with 60% relative humidity. Laboratory-raised solitary and gregarious locust strains were maintained as previously reported (Kang et al., 2004). Briefly, after hatching, each locust was separated into an insect-rearing box (72 mm × 72 mm × 200 mm) containing a corn leaf, to produce solitary-phase adult locusts. To produce gregarious phase locusts, 30 locusts were reared in an insect-rearing box (72 mm × 72 mm × 200 mm) with a corn leaf. A total of 30 third-instar nymphs were subsequently transferred to an insect cage (200 mm × 200 mm × 200 mm) until they became adults, and group of 5 emerged adults were transferred into insect-rearing boxes for the gregarious-phase locusts. All locusts were reared for 10 days before extracting total RNAs from their heads.

Field migratory locusts were collected from Haenam and Muan, Jeonllanam-do in Korea. There were no gregarious-phase locusts in Korea in 2015 and 2016, and two different body colors of locusts were collected. Locusts with green bodies are considered to be in the solitary phase, while those with gray bodies are in an intermediate state between the solitary- and gregarious-phases. The collected locusts were anesthetized with CO_2, and their heads were immediately cut and frozen in LN_2 in the field. The frozen heads were maintained in liquid nitrogen (LN_2) until being processed for total RNA extraction.

Total RNA extration, RNASeq library construction and sequencing

Total RNA was extracted using the TRIzol reagent (Thermo Fischer Scientific, Waltham, MA, USA) from the frozen heads of L. *migratoria* maintained in LN_2, and the results of three independent extractions per sample were pooled for further analysis. The purity of the RNA (260/230 absorbance ratio of 2.0–2.2) was measured using a NanoDrop8000 spectrophotometer (Thermo Fischer Scientific), and RNA integrity was determined based on the RNA Integrity number (Rin) values (> 8.0) using an Agilent RNA Nano Chip with Bioanalyzer 2000 (Agilent Technologies, Santa Clara, CA, USA). Illumina RNA−Seq libraries were prepared following the manufacturer's recommendation with the TruSeq Stranded mRNA Library Prep kit (Illumina, San Diego, CA, USA). Briefly, mRNA was purified using poly-T magnetic beads with attached oligos and fragmented at 98 ˚C for 4 min with divalent cations. The first strand of cDNA was synthesized using reverse transcriptase in the presence of random hexamers and second strand cDNA synthesis was conducted using DNA polymerase with dUTP instead of dTTP. The 3′ ends of the cDNA molecules were subsequently adenylated and adapter ligation was carried out. Finally, PCR was performed to enrich the strand specific cDNA libraries. The quality of the amplified libraries was verified using a Bioanalyzer (Agilent USA), and each library was sequenced on the Illumina NextSeq 500 platform (Illumina, USA) platform with 2 × 150 bp.

Raw data quality control and L. migratoria transcriptome assembly

The raw data of the 12 RNA-seq. Datasets used in this study were deposited in the NCBI short read archive (SRR5580661 ~ SRR558072). FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used to examine the quality of the raw data, and Trimmomatic (Bolger et al., 2014) was then used to remove low quality and adapter sequences. Since the available reference genome of L. *migratoria* (AVCP000000000.1, 1,397,492 contigs, total, 5.76 Gigabytes) is highly fragmented, 369,771 contigs (4.24 Gigabyte) longer than 5 kb in length, was selected for mapping. HISAT (Kim et al., 2015) was employed to map the trimmed reads from each sample to L. *migratoria* draft genome assembly (Supplementary Table 1), and then StringTie (Pertea et al., 2015) was employed to assemble the mapped reads to transcripts. Twelve independent assemblies were merged into 33,818 transcripts using Cuffmerge (Trapnell et al., 2012) (Table 1), which was considered to represent the complete set of L. *migratoria* transcripts.

Identification of long non-coding RNAs

To define LNC and coding RNA sequences from the assembled transcripts, the assembled transcripts were processed according to the computational pipeline shown in Fig. 1. Among the 33,818 transcripts, 11,125 transcripts were assigned to a coding transcript bin because they were either translated into proteins larger than 100 amino acids or their lengths were shorter than 200 nucleotides. Another 10,276 and 216 transcripts were considered coding transcripts after performing BLASTx search against the SwissProt (cut off *E*-value < 0.001) and Pfam (cut off E-value < 0.001) databases, respectively. An additional 1189 transcripts were discarded from the LNC RNA set based on performing coding potential analysis (CPC < -1.0 & CPAT < 0.39). Finally, thirty-nine Housekeeping RNAs (tRNA, rRNA, snRNA, and snoRNAs) (cut off E-value $< 1.0 \times 10^{-10}$) were further removed and the