Journal of Asia-Pacific Entomology 18 (2015) 377–382

Contents lists available at ScienceDirect

Journal of Asia-Pacific Entomology

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

Comparison of the genome profiles between head and body lice

Jae Soon Kang [a], Yong-Jun Cho [b], Ju Hyeon Kim [c], Sang Hyeon Kim [c], Seungil Yoo [d], Seung-Jae Noh [d], Junhyung Park [d], Kyong Sup Yoon [e], J. Marshall Clark [f], Barry R. Pittendrigh [g], Jongsik Chun [b], Si Hyeock Lee [c,h,*]

[a] Gyeongnam Department of Environmental Toxicology Chemistry, Korea Institute of Toxicology, Jin-Ju, Gyeongnam, Republic of Korea
[b] School of Biological Sciences, Seoul National University, Seoul, Republic of Korea
[c] Department of Agricultural Biotechnology, Seoul National University, Seoul, Republic of Korea
[d] Department of Research, Codes Division, Insilicogen, Inc., Suwon, 441-813, Republic of Korea
[e] Department of Biological Sciences and Environmental Sciences Program, Southern Illinois University-Edwardsville, Edwardsville, IL, 62026, USA
[f] Department of Veterinary & Animal Science, University of Massachusetts, Amherst, MA 01003, USA
[g] Department of Entomology, University of Illinois Urbana-Champaign, Urbana, IL, 61801, USA
[h] Research Institute for Agriculture Life Science, Seoul National University, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Received 26 February 2015
Revised 14 April 2015
Accepted 16 April 2015
Available online 24 April 2015

Keywords:
Pediculus humanus
Louse
Genome sequencing
Gene homologue analysis
SNP analysis
Endosymbiont

ABSTRACT

The body louse (*Pediculus humanus humanus*) is known to have diverged from the head louse (*P. humanus capitis*) but genomic differences between these two subspecies still remain unexplored. To compare genomic profiles between head and body lice, whole genome sequences of head lice were determined by next generation sequencing methods based on both Illumina Genome analyzer and Roche GS FLX pyrosequencing and compared with the reference genome sequences of the body louse. Total consensuses generated by mapping to the body louse genome in conjunction with *de novo* assembly of head louse genome sequences revealed a head louse genome size of 110 Mbp with a 96% coverage of the body louse genome sequences. A total of 12,651 genes were predicted from the head louse genome sequences although more precise assembly and functional annotation of the genome is required for a more accurate gene count. Among the 873 genes that were putatively specific to the head louse, 15 genes were confirmed to be transcribed in both head and body lice, suggesting the previously estimated gene number of the body louse was likely underestimated. The single nucleotide polymorphism analysis showed that the nucleotide diversity of genome between head and body lice was 2.2%, which was larger than that of the transcriptome between head and body lice. An endosymbiont genome analysis showed that the composition of endosymbionts in head lice was similar to that of body lice and *Candidatus* Riesia pediculicola was the primary endosymbiont in both head and body lice.

Introduction

Both the head louse (*Pediculus humanus capitis*) and the body louse (*P. humanus humanus*) are obligatory human ectoparasites feeding exclusively human blood. It has been suggested that they adapted to human when the common ancestor of human and chimpanzee diverged 5–6 million years ago (Pennisi, 2004). The body louse is speculated to have diverged from the head louse when human began to wear clothing (Kittler et al., 2003). However, the taxonomic status of the head and body lice is still disputable because fertile F1 hybrid can be generated between head and body lice under laboratory conditions although their interbreeding has not been observed in the wild (Mullen and Durden, 2009). Comparison of several molecular markers, such as mitochondrial DNA, nuclear ribosomal DNA and microsatellite DNA (Leo et al., 2002, 2005; Kittler et al., 2003; Reed et al., 2004; Leo and Barker, 2005; Light et al., 2008), suggested that head and body lice are conspecific except for the study using microsatellite DNA, in which head and body lice were proposed to be separate species (Leo et al., 2005).

Despite a similar genetic background, head and body lice have several differences in their biological features, such as niche, body size and vector competence. Head lice live only on the human scalp throughout its entire lifespan whereas body lice live primarily on clothes as well as on body hair except for hair on the scalp (Kittler et al., 2003). Body lice can transmit pathogenic bacteria to human, such as *Rickettsia prowazekii* (epidemic typhus), *Bartonella quintana* (trench fever), and *Borrelia recurrentis* (relapsing fever) (Brouqui et al., 1999; Raoult and Roux, 1999; Rydkina et al., 1999). In contrast, the head louse is not known to transmit pathogen to humans. Because of such differences, these two species are regarded as the appropriate models for studies of species differentiation and differential vector competence (Kim et al., 2011).

* Corresponding author at: Department of Agricultural Biotechnology, Seoul National University, 151-921, Seoul, Republic of Korea. Tel.: +82 2 880 4704; fax: +82 2 873 2319.
E-mail address: shlee22@snu.ac.kr (S.H. Lee).

http://dx.doi.org/10.1016/j.aspen.2015.04.010
1226-8615/

Recently, whole genome sequencing of the body louse was completed (Kirkness et al., 2010). The body louse has a 108 Mb genome, which is the smallest among insect genomes sequenced, and includes 10,773 protein-coding genes. This reduced number of genes in the body louse has been attributed to its simple life style, which includes feeding only on fresh human blood and having humans as a sole host. Similarly, it was reported that the number of immune-related genes in the body louse was less than that of other insects, such as *Drosophila melanogaster*, *Bombyx mori*, *Anopheles gambiae* and *Tribolium castaneum* (Kim et al., 2011). Since the body louse diverged from the head louse relatively recently, the genetic background of the head louse was assumed to be similar to the body louse, thus likely having almost identical genome size, gene number and genome structure.

The transcriptional profiles between body and head lice were recently compared (Olds et al., 2012). Among the 10,775 protein-coding genes predicted from the body louse genome, almost the same number of genes was annotated both in the head louse (10,770 genes) and the body louse (10,771 genes) transcriptomes. Among the 544 genes in the genome of *Candidatus* Riesia pediculicola, a primary endosymbiont of both head and body lice (Sasaki-Fukatsu et al., 2006), 539 genes were observed from the head louse transcriptome, which were similar to the 538 genes observed in the body louse transcriptome. These results suggested that the phenotypic differences between head and body lice were not likely due to different gene components but rather due to differential gene regulation of similar gene sets. To confirm this assumption, it is necessary to determine the whole genome sequences of the head louse and compare them with those of the body louse.

In this study, the whole genome sequencing of the head louse was performed using two next generation sequencing (NGS) methods, Genome analyzer IIx-platform sequencing and GS FLX Titanium-platform sequencing. The sequences generated from NGS were mapped to the genome of the body louse or *de novo* assembled. From this data set, features of the head louse genome were examined, single nucleotide polymorphisms (SNP) analyzed and putative genes predicted. These predicted genes were homologue-compared to the genes of the body louse, from which the head louse-specific genes were identified. In addition, the bacterial endosymbiont community of the head louse was analyzed and the genome of *Ca.* R. pediculicola in the head louse was sequenced and compared to that determined in the body louse.

Materials and methods

Head and body lice rearing

A highly inbred BR-HL strain of head lice was used for the whole genome sequence analysis. The BR-HL strain was originally collected in Bristol, UK, and has been reared on the *in vitro* rearing system (Yoon et al., 2006). For the reference genomic DNA extraction and cDNA preparation, another head louse strain (CA-HL, originally collected from Cambodia) and two body louse strains (SF-BL, collected from San Francisco; CP-BL, Culpepper strain that was used for the body louse genome analysis) were also used. The louse colonies were maintained under conditions of 30 °C, 70–80% RH and 16L:8D in a rearing chamber.

Genomic DNA extraction

Genomic DNA was extracted from approximately 500 newly hatched first instar nymphs before their first blood meal using DNeasy blood & tissue kit (Qiagen, Hilden, Germany). Both quality and quantity of genomic DNA were analyzed by gel electrophoresis and Quant-iT™ PicoGreen® dsDNA Quanititation reagent (Invitrogen, Carlsbad, CA, USA).

Whole genome sequencing

Genome Analyzer IIx (Illumina, San Diego, CA, USA) with a mean length of 101 bp paired-end and GS FLX Titanium (Roche, Indianapolis, CA, USA) using a 3 kb library were used for genome sequencing according to the manufacturer's recommendations at the National Instrumentation Center for Environmental Management (NICEM, Seoul, Korea) and DNA Link, Inc. (Seoul, Korea), respectively. The resulting sequence data were mapped to 8588 contigs of the body louse genome. Unmapped reads were used for creating *de novo* assembly. All analyses and statistics for *de novo* assembly and reference mapping were performed using CLC Genomics Workbench (CLC bio, Aarhus, Denmark). The gene coding regions were predicted by GeneMark-ES version 2.3a (Ter-Hovhannisyan et al., 2008) in the *de novo* assembled contigs. Predicted protein-encoding genes were analyzed by BLASTP and the ones, which showed significant BLAST similarity (e-value $< 10^{-5}$) to proteins from other organisms in the non-redundant (NR) database at the National Center for Biotechnology Information (NCBI), were annotated. Comparison of the genes between head and body lice were conducted by BLAST-searching of all genes of each species against opposite species genome database on the condition of $< 10^{-4}$ e-value.

Endosymbiont genomes were annotated using RAST server (Aziz et al., 2008). The reads that unmapped to the genome of the body louse were assembled and the assembled contigs containing 16S ribosomal RNA (16S rRNA) fragment were selected. Selected contigs were used for analyzing the bacterial community by CLcommunity™ (Ver. 2.04, CLC bio). In addition, the genome of *Ca.* R. pediculicola in the head louse was sequenced and mapped to the genome of *Ca.* R. pediculicola in the body louse.

Verification of newly identified putative genes

Among the putative head louse-specific genes, 30 genes, which were not identified from the body louse genome, were further analyzed by PCR to confirm their presence in the genomes of head and body lice. Primer pairs were designed from the putative exon regions of each gene (Supplementary Table 1) and used for PCR using either genomic DNA or cDNA. Genomic DNA was extracted from the BR-HL, CA-HL, SF-BL, and CP-BL. cDNA was synthesized from the total RNA extracted from the same strains of head and body lice. PCR was conducted with Ex Taq polymerase (Takara Korea Biomedical Inc., Seoul, Korea) and 5 pmol primers under the following thermal program: an initial denaturation at 95 °C for 2 min and a total of 34 cycles of 95 °C for 20 s, 52 °C for 10 s, and 72 °C for 1 min.

Calculation of Ka/Ks ratios

Through reciprocal blast-searching of protein sequences between head and body lice, 9015 pairs of orthologous gene sequences were extracted. Each of these pairs was aligned using ClustalW2. The ratios (*Ka*/*Ks*) of the non-synonymous substitutions per site (*Ka*) to synonymous substitutions per site (*Ks*) were estimated for each orthologue pair and averaged over the entire alignment using *Ka*/*Ks* Calculator v.2 (Zhang et al, 2006). Except for 2303 pairs that failed because of large gap opening in the sequence alignment, 6721 orthologue pairs were successfully analyzed using the *Ka*/*Ks* Calculator. The *Ka*/*Ks* Calculator adopts different models for codon substitutions, such as approximate methods (NG, LPB, MYN, etc.) and maximum likelihood methods (GY, MS, MA, etc.), among which the NG method was used to estimate *Ka*/*Ks* ratios.

Results

Head louse genome features

A total of 11.5 Gb nucleotide sequences (ca. 114 million reads having 101 bp) were obtained from Illumina GA platform (Table 1). These