Intraspecies comparative genomics of three strains of *Orientia tsutsugamushi* with different antibiotic sensitivity

Hsiao-Mei Liao  
*Center for Biologics Evaluation and Research, Food and Drug Administration*

Chien-Chung Chao  
*Naval Medical Research Center*

Haiyan Lei  
*Center for Biologics Evaluation and Research, Food and Drug Administration*

Bingjie Li  
*Center for Biologics Evaluation and Research, Food and Drug Administration*

Shien Tsai  
*Center for Biologics Evaluation and Research, Food and Drug Administration*

*See next page for additional authors*

Follow this and additional works at: [http://digitalcommons.unl.edu/usuhs](http://digitalcommons.unl.edu/usuhs)
Authors
Hsiao-Mei Liao, Chien-Chung Chao, Haiyan Lei, Bingjie Li, Shien Tsai, Guo-Chiuan Hung, Wei-Mei Ching, and Shyh-Ching Lo
Data in Brief

Intraspecies comparative genomics of three strains of Orientia tsutsugamushi with different antibiotic sensitivity

Hsiao-Mei Liaoa, Chien-Chung Chaob,c, Haiyan Leia, Bingjie Li a, Shien Tsai a, Guo-Chiuan Hung a, Wei-Mei Chingb,c, Shyh-Ching Loab,⁎

a Tissue Microbiology Laboratory, Division of Cellular and Gene Therapies, Office of Tissues and Advanced Therapies, Center for Biologics Evaluation and Research, Food and Drug Administration, Silver Spring, MD, USA
b Naval Medical Research Center, Silver Spring, MD, USA
c Uniformed Services University of the Health Sciences, Bethesda, MD, USA

Abstract

Article history:
Received 24 January 2017
Received in revised form 20 March 2017
Accepted 26 March 2017
Available online 28 March 2017

Keywords:
Orientia tsutsugamushi
Comparative genomic analysis
Antibiotic sensitivity
RAST genome annotation
Gene function comparison
Genome sequence comparison

We recently reported the genome of Orientia tsutsugamushi (OT) strain Karp (GenBank Accession #: NZ_LYMA00000000.2, https://www.ncbi.nlm.nih.gov/nuccore/NZ_LYMA00000000.2) with >2 Mb in size through clone-based sequencing and high throughput genomic shotgun sequencing (HTS). The genomes of OT strains AFSC4 and AFSC7 were similarly sequenced by HTS Since strains AFSC4 (GenBank Accession #: NZ_LYMT00000000.1, https://www.ncbi.nlm.nih.gov/nuccore/1035784408) and AFSC7 (GenBank Accession #: NZ_LYMB00000000.1, https://www.ncbi.nlm.nih.gov/nuccore/1035854767) were more resistant to antibiotics than strain Karp, we conducted comparative analysis of the three draft genomes annotated by RAST server aimed to identify possible genetic bases of difference in microbial antibiotic sensitivity. Intraspecies comparative genomics analysis of the three OT strains revealed that two ORFs encoding hypothetical proteins in both strains AFSC4 and AFSC7 are absent in strain Karp.

© 2017 Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Direct link to deposited data


2. Experimental design, materials and methods

Orientia tsutsugamushi (OT), the etiological agent of scrub typhus infects human via chigger bite, is an obligate intracellular organism. The disease is endemic in the scrub typhus triangle or so called “tsutsugamushi triangle” which extends from northern Japan and far-eastern Russia in the north, to northern Australia in the south, and to Pakistan in the west with a million cases reported yearly and a billion people under the risk [1]. Many OT isolates have been obtained from endemic areas but also from the areas outside of the scrub typhus triangle [2,3]. Clinical antibiotic resistant infections have been reported in Southern India [4] and Southern Thailand [5–7]. In vitro studies and animal models showed that Karp strain was sensitive to doxycycline and azithromycin whereas AFSC4 and AFSC7 were much less sensitive [8]. We recently reported the 2 Mb genome of strain Karp using both clone-based sequencing and high-throughput genomic shotgun sequencing [9]. In this study, we conducted genomic sequencing of strains AFSC4 and AFSC7 isolated from Thailand and made genomic comparison among the three strains of OT.
All Orientia tsutsugamushi strains were cultured in irradiated L929 cells. Purified bacteria with minimum amount of host material was prepared by serious ultracentrifugation in renografin density gradient [10]. The DNAs extracted from the purified bacteria of strain AFSC4 and AFSC7 were subjected to whole genome shotgun sequencing (WGS) using Illumina MiSeq platform with the 2 × 250 bp pair-end mode. Briefly, 50 ng of purified DNA was subjected to DNA library construction using the Nextera DNA Sample Prep Kit (Illumina), and the generated fastq files were used in the downstream sequencing data analysis, which was performed by CLC Genomics Workbench Version 9.0.1. The raw reads were trimmed and filtered under the parameters: reads phred score >Q30, base error probability score >0.05, base ambiguities <2 and read-length <15 bp. The trimmed raw reads of strain AFSC4 and AFSC7 were de novo assembled into original 1201 and 1803 contigs, with the average coverage 120 and 179, respectively. The parameters of de novo assembly in CLC Genomics Workbench v.8.5 were minimum contig length 300 bp, no scaffolding, mismatch cost = 2, insertion cost = 3, deletion cost = 3, length fraction = 3, similarity fraction = 0.8. The original contigs of two strains obtained from WGS were first mapped to genomes of three reference strains of OT: 1) strain Karp

---

Fig. 1. The gene subsystem distribution of OT strain KARP, AFSC4 and AFSC7. The analysis was performed using RAST server Version 2.0. The pie charts show the distribution of OT strain (A) KARP (B) AFSC4 and (C) AFSC7.

---
Ikeda reference genome using CONTIGuator genome finishing tool. The contigs of AFSC4 and AFSC7 were aligned to the Ikeda reference genome using CONTIGuator genome finishing tool to obtain the oriented and arranged contig series as the draft genomes [11]. In comparison with the Karp draft genome 2,026,724 bp in length with 30.41% G + C content [9], both of AFSC 4 and AFSC 7 draft genome contributed totally from the contigs are significantly shorter. The smaller sizes of draft genome with higher numbers of contigs were evidently contributed by the high content (about 40%) of repetitive sequence in genomes of OT. The ambiguity and abundance of the repeats prevented generation of long contigs in de novo assembly. Many junctions between contigs of AFSC4 and AFSC7 could only be possibly closed with the long reads generated by other sequencing approaches or methods in the future.

All three genomes were annotated using RAST server Version 2.0 [12]. The numbers of the contigs would not influence the gene annotation. The strain Karp was annotated with 2089 functional elements, including 2052 coding sequences (CDSs) and 37 RNAs. Total 1090 CDSs transcribed from the positive strand and 999 transcribed from negative strand. 32% (656) of the CDSs were categorized into 185 subsystems (Fig. 1A). The strain AFSC4 has 1149 annotated CDS and 36 RNAs, 560 CDSs were transcribed from the positive strand and 625 were transcribed from negative strand. 44% (502) of the CDSs were categorized into 186 subsystems (Fig. 1B). In the study, the strain AFSC7 was annotated with 1376 CDSs and 37 RNAs. In that, 652 CDSs were transcribed from the positive strand and 761 were transcribed from negative strand: 41% (557) of the CDSs were categorized into 185 subsystems (Fig. 1C). AFSC4 and AFSC7 have similar subsystem distribution, while Karp has more CDSs categorized into the membrane transporter, the most abundant subsystems. Protein metabolism and RNA metabolism are the second and third abundant subsystems of these three strains, respectively.

The SEED Viewer functional comparison [13] compared the genes encompassed in functional groups or subsystems in genomes, pairwisely, which identified 682 CDSs of AFSC4 and 677 CDSs of AFSC7 with known functions and are comparable to those of CDS identified in Karp genome. The Seed Viewer functional-based comparative analysis revealed five protein-coding genes that are present in the genomes of AFSC4 and AFSC7, but absent in the genome of Karp. The sequence length of these five genes ranged from 135 bp to 543 bp. However, all of these five genes could be found in the Karp genome, when the gene sequences were blasted against the draft genome. The findings indicated that the functional comparison between genomes by the computational tool may not be optimal for some of the shorter genes, and shown as false positive findings.

To identify genomic differences or possible genes that might contribute to the differential of the antibiotic sensitivity among the closely related strains of OT, we further analyzed the CDSs of the genomes using the SEED Viewer computational tool for sequence comparison. The genome sequences of the compared strains were aligned to a reference genome. The output listed the genes on the reference genome following the chromosomal order and displayed the gene hits on the compared strains without concerning the functions of the listed genes. The genome of OT strain Ikeda was used as the reference strain for the detail sequence comparison (Fig. 2). The comparative analysis revealed 28 CDSs that are present in AFSC4 and AFSC7 but are absent in Karp. However, specific Blast verification for gene sequences of each identified CDS showed only two CDSs, number 1436 and 2073, were truly missing in the Karp genome (Table 1). Figs. 3 and 4 show the sequences of these two genes were blasted against the draft genomes of the three OT strains. The proteins encoded by these two genes are hypothetical without known biological function. Although the two genes identified are coding for hypothetical proteins without known biological function, the close genomic comparisons of three OT strains with evident differences in antibiotic tolerance has provided the possible targets for investigation of OT drug-resistance mechanism(s).
Four genome sequences were chosen as the inputs for sequence compare. The database available complete genome of OT strain Ikeda was used as reference for comparison. The circles represent the blast identity of the compared strains against the reference strain, the red dot indicates the first based of the reference sequence, and the continuous bases were extended clockwise. The inner circle indicates the blast hit in Karp, the middle indicates the blast hit in AFSC4 and the outer circle indicates the blast hit in AFSC7. The rainbow spectrum of color was used to indicate the identity (purple: high to red: low) of the sequence hit on the corresponding position of the reference genome. Most of the blast hits on the three strains have high identity comparing to the reference sequences.

On the top, the black line represents the sequence of the coding sequence Fig|334380.3.peg.1436 (CDS#1436 in short), the numbers above indicate the length of the sequence. The other lines (green) below the top line indicate the sequences of the contigs of the three draft genomes that can be matched to the CDS#1436. The brighter green indicates the higher identity. Only the FASC4_contig0000162 and AFSC7_contig000227 encompass the full-length sequence of CDS#1436.

On the top, the black line represents the sequence of the coding sequence Fig|334380.3.peg.2073 (CDS#2073 in short), the numbers above indicate the length of the sequence. The other lines (green) below the top line indicate the sequences of the contigs of the three draft genomes that can be matched to the CDS#2073. The brighter green indicates the higher identity. Only the AFSC4_contig0000248 and AFSC7_contig000302 encompass the full-length sequence of CDS#2073. Karp_contig000097 has partial sequence matched to the CDS#2073.

---

**Table 1**

The summary of the two coding sequences present in AFSC4 and AFSC7 but absent in Karp.

<table>
<thead>
<tr>
<th>Reference Orientia tsutsugamushi str. Ikeda</th>
<th>Blast against Karp, AFSC4 and AFSC7 draft genome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene ID</td>
<td>Function</td>
</tr>
</tbody>
</table>
| 1436 | Hypothetical protein | 1032 | atgcaacaagagatatataaatatatgattttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt

---

**Fig. 3.** The blast hits of the CDS#1436 sequence against the draft genomes of OT strain Karp, AFSC4, AFSC7.
3. Nucleotide sequence accession number

The first versions of draft genome sequences of Orientia tsutsugamushi strain AFSC4 and AFSC7 were deposited in GenBank under accession number LYMT00000000 and LYMB00000000, respectively.

Acknowledgement

The authors wish to thank Dr. Gregory Dasch for purifying Karp, AFSC4 and AFSC7 strains of Orientia tsutsugamushi from L929 cells and Ms. Zhiwen Zhang and Ms. Tatyana Belinskaya for subsequent DNA extraction for WGS. This work was supported in part by Work Unit Number (WUN) 6000.RAD1.JA0310 and an FDA Modernizing Science grant. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication. The opinions and assertions contained herein are the private ones of the authors and are not to be construed as official or as reflecting the views of the Department of the Navy, the Naval service at large, the Department of Defense, or the U. S. Government. Authors Chien-Chung Chao and Hai-Ming Liao are military employees of the United States Government as part of official duties. Title 17 U.S.C. §105 provides that ‘Copyright protection under this title is not available for any work of the United States Government’. Title 17 U.S.C. §101 defines a U.S. Government work as a work prepared by employee of the U.S. Government as part of that person’s official duties.

References


