



Data in Brief

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



Hsiao-Mei Liao^a, Chien-Chung Chao^{b,c}, Haiyan Lei^a, Bingjie Li^a, Shien Tsai^a, Guo-Chuan Hung^a, Wei-Mei Ching^{b,c}, Shyh-Ching Lo^{a,*}

^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

ARTICLE INFO

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

ABSTRACT

We recently reported the genome of *Orientia tsutsugamushi* (OT) strain Karp (GenBank Accession #: NZ_LYMA00000000.2, https://www.ncbi.nlm.nih.gov/nucore/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/nucore/1035784408>) and AFSC7 (GenBank Accession #: NZ_LYMB00000000.1, <https://www.ncbi.nlm.nih.gov/nucore/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/>).

Specifications

Organism/cell line/tissue	<i>Orientia tsutsugamushi</i> strain Karp, AFSC4 and AFSC7
Sex	N/A
Sequencer or array type	Sanger, Illumina MiSeq V2 250 × 2 PE
Data format	Analyzed
Experimental factors	Bacteria isolated from different geographic areas with distinguished sensitivities to antibiotics
Experimental features	The genome sequences of three <i>Orientia tsutsugamushi</i> strains were annotated and compared.
Consent	N/A
Sample source location	The strains were originally collected from Papua New Guinea and Thailand.

1. Direct link to deposited data

Orientia tsutsugamushi, Karp https://www.ncbi.nlm.nih.gov/genome/710?genome_assembly_id=290025.

* Corresponding author at: WO52-72 RM3128 HFM-735, 10903 New Hampshire Ave, Silver Spring, MD-20993-0002, United States.

E-mail address: shyhching.lo@fda.hhs.gov (S.-C. Lo).

Orientia tsutsugamushi, AFSC4 https://www.ncbi.nlm.nih.gov/genome/710?genome_assembly_id=276762.

Orientia tsutsugamushi, AFSC7 http://www.ncbi.nlm.nih.gov/genome/710?genome_assembly_id=276760.

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.

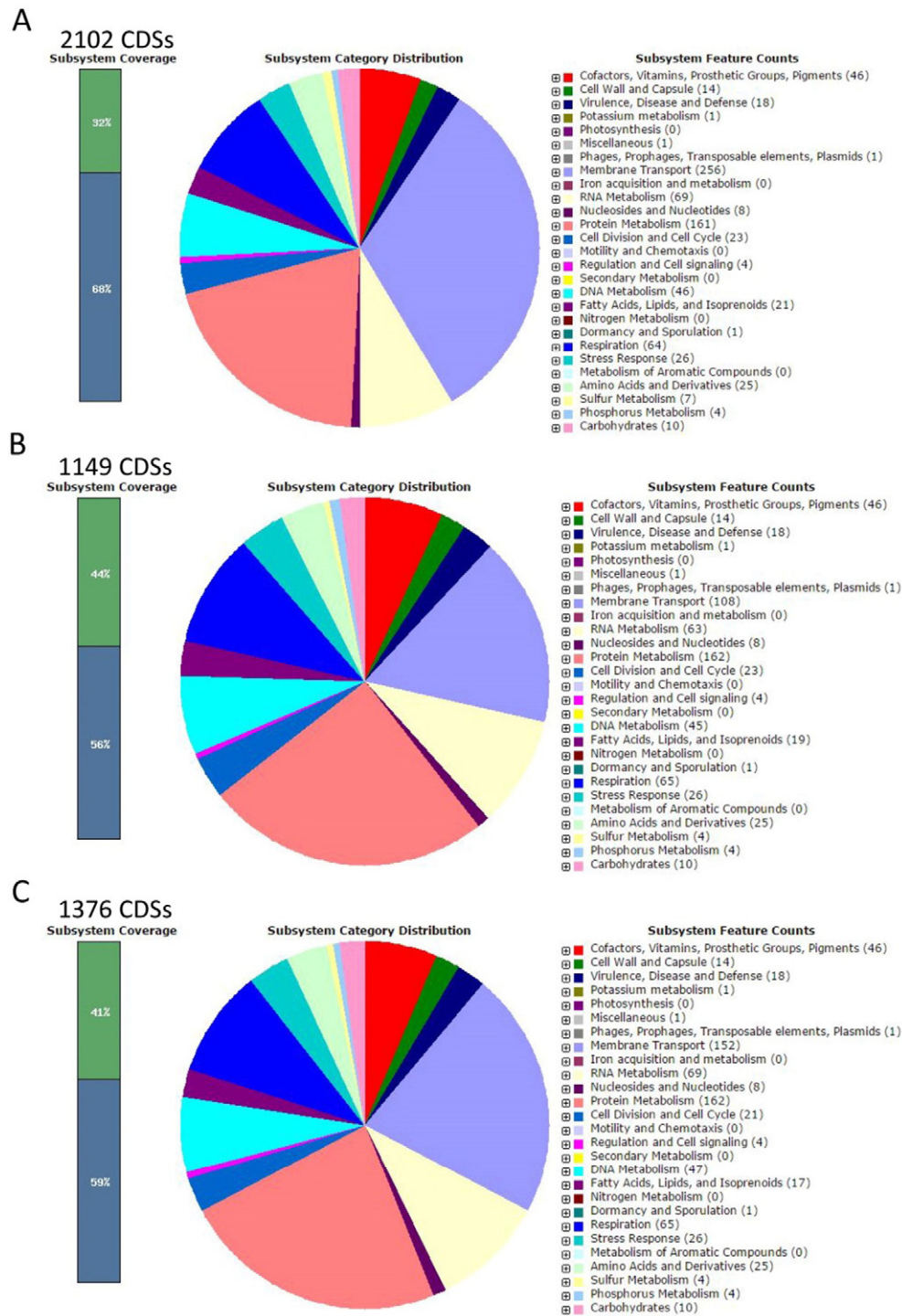


Fig. 1. The gene subsystem distribution of OT strain KARP, AFSC4 and AFSC7. The analysis was performed using RAST server Version 2.0. The pie charts show the distribution of OT strain (A) KARP (B) AFSC4 and (C) AFSC7.

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 = 5, 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

(GenBank Accession#: LYMA00000000), 2) strain Boryong (NC_009488.1) and 3) strain Ikeda (NC_010793.1). The original contigs were also blasted in parallel against nt/nr database to remove those with sequence homology to mouse and other environmental organisms as well as those with only very low coverage. The reads from the contigs, which passed the filtering, were retrieved and subjected to a second round of *de novo* assembly.

All retained WGS contigs were able to be mapped to at least one of the three OT genome references, the final contig numbers of AFSC4 and AFSC7 were 452 and 485, respectively. The draft genome of AFSC4 has 1,295,323 bp (29.9% G + C) and AFSC7 has 1,437,569 bp (30.0% G + C). 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, pairwise, 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).

Result

[change organism selection](#)

You chose to compute data for the following organisms:

Reference: Orientia tsutsugamushi str. Ikeda (334300.3)
 Comparison Organism 1: Orientia tsutsugamushi Karp-v2_109 contigs (784.7) | [BlastDotPlot](#)
 Comparison Organism 2: Orientia tsutsugamushi str. AFSC4_Ikeda Aligned (1359185.13) | [BlastDotPlot](#)
 Comparison Organism 3: Orientia tsutsugamushi str. AFSC7_Ikeda Aligned (1359185.14) | [BlastDotPlot](#)

Bidirectional best hit: 100 99.9 99.8 99.5 99.5 98 98 95 90 80 70 60 50 40 30 20 10
 Unidirectional best hit: 100 99.9 99.8 99.5 99.5 98 98 95 90 80 70 60 50 40 30 20 10
[export table](#) | [clear all filters](#)

display 30 items per page
 displaying 2 - 31 of 2191

percent identity 784.7		percent identity 1359185.13									
334300.3		1359185.13									
Contig	Gene	Length	Hit	Contig	Gene	Hit	Contig	Gene	Hit	Contig	Gene
1	2	818	bi	46	762	bi	1	2	bi	1	2
1	3	1229	bi	46	764	bi	1	3	bi	1	3
1	4	4007	bi	46	760	bi	1	4	bi	1	4
1	5	98	bi	46	758	bi	1	5	bi	1	5
1	6	434	bi	46	758	bi	1	6	bi	1	6
1	7	418	bi	46	757	bi	2	7	bi	1	6
1	8	190	bi	46	758	bi	2	8	bi	1	5
1	9	262	bi	46	758	bi	2	9	bi	1	7
1	10	234	bi	46	758	bi	2	10	bi	1	8
1	11	207	uni	5	27	uni	95	118	uni	246	1183
1	12	39	*	*	*	bi	3	32	*	*	*
1	13	354	bi	46	740	bi	3	33	bi	2	12
1	14	340	bi	46	740	bi	3	34	bi	2	14
1	15	307	bi	46	742	bi	3	35	bi	2	15
1	16	42	bi	46	746	bi	3	36	*	*	*
1	17	123	bi	46	748	bi	3	37	bi	2	16
1	18	69	*	*	*	bi	3	38	bi	2	17
1	19	158	bi	46	744	bi	3	39	bi	2	18
1	20	216	bi	46	748	bi	3	40	bi	2	19
1	21	284	bi	46	742	bi	3	41	bi	2	20
1	22	164	bi	46	743	bi	3	42	bi	2	21
1	23	1140	bi	46	740	bi	3	43	bi	2	22
1	24	404	bi	46	738	bi	3	44	bi	2	23
1	25	383	bi	46	738	bi	3	45	bi	2	24
1	26	299	bi	46	737	bi	3	46	bi	2	25
1	27	103	bi	46	736	bi	3	47	bi	2	27
1	28	292	bi	46	738	bi	3	48	bi	2	28
1	29	274	bi	46	734	bi	3	49	bi	2	29
1	30	399	bi	46	733	bi	3	50	bi	2	30
1	31	354	bi	46	732	bi	3	51	bi	2	31

displaying 2 - 31 of 2191

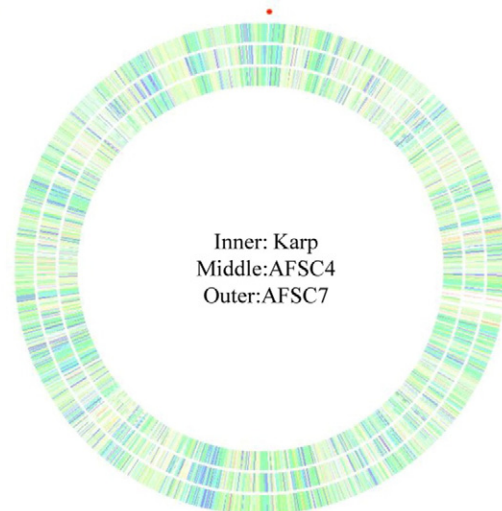


Fig. 2. The results of the SEED Viewer sequence compare.

Table 1
The summary of the two coding sequences present in AFSC4 and AFSC7 but absent in Karp.

Reference <i>Orientia tsutsugamushi</i> str. Ikeda				Blast against Karp, AFSC4 and AFSC7 draft genome	
Gene ID	Function	Length (bp)	Sequence	Identity	Hit notes
1436	Hypothetical protein	1032	atgcaacaagaagaatatataaaattcttttttaggagaagatagctacattaataaaaaacttttttagcatt aaatcaactacagcatgatatactgtctgtactatagctatataatagcaagcctcaaaaagcacttttga gctatcttaaaaatcaacaattgagtcagtttttagaaaggtttgtgaaaataaaggatctcaagcaagttataaa aaagtccaataaaaaaacaagacttagcaagtggttaaggactcttgaatataaaaataagcagcatcaaca aaaatctcaaaagttgaagatgataaaatcgttactgcaagctattaatagcaaaaactacaagctcttatgc tagatcttgatgctcaggatgatctctgaaggatggacaggttgaatagcctctcggatcaattctcagagtta ttatcaagataaaaacgcttattgtttatggcagatgaagaatggggcaacgtattataattgtataccagatgg ctctatattacctaatttagcatatttcttttgataggaattgggtcgtttgataaattagcaaggaaactttat ctaagttactcaaaaaggatctataaaggtagatgctctttttattcattgcatccagatttattggtataatt caggccatagttcaagaaaacatccaagtcgcaaaaaattcagatttagcactataatcagcaaggatgagcattta actgcttcaacattgattgtaaacattagtttaagcgtcaacattgaggtaggagatgatgaagattgcaaaaa aaatctgaaaatttccaaaataataccactaaacggttaatttgcgctatgaattctatgatgaaccaatttgata gtaatggcagatgactaagttattaggaatagaagaagctctttgattgctgctctatttcttag ttgaagcaatgtaccagtaataagaagaatagatgtacaaaaatagtaagttataactagatcacaataatc aaaatattcagttggtacattgcaaaaattaccgattactaaagtaaaataaggaaataaaaaaagtaatttttctt aacctataagctactcaaaagtccattaa	24/26 (92%)	One full length hit was found in AFSC4 and AFSC7, but not in Karp. (a 24 nt fragment was blast matched)
2073	Hypothetical protein	192	ttgaagcaatgtaccagtaataagaagaatagatgtacaaaaatagtaagttataactagatcacaataatc aaaatattcagttggtacattgcaaaaattaccgattactaaagtaaaataaggaaataaaaaaagtaatttttctt aacctataagctactcaaaagtccattaa	75/77 (97%), 14/14 (100%), 14/14 (100%)	One full length was found in AFSC4 and AFSC7 genomes, but not found in Karp. (three fragments found with 75, 14, and 14 nt blast match, respectively)

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 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_contig0000097 has partial sequence matched to the CDS#2073.

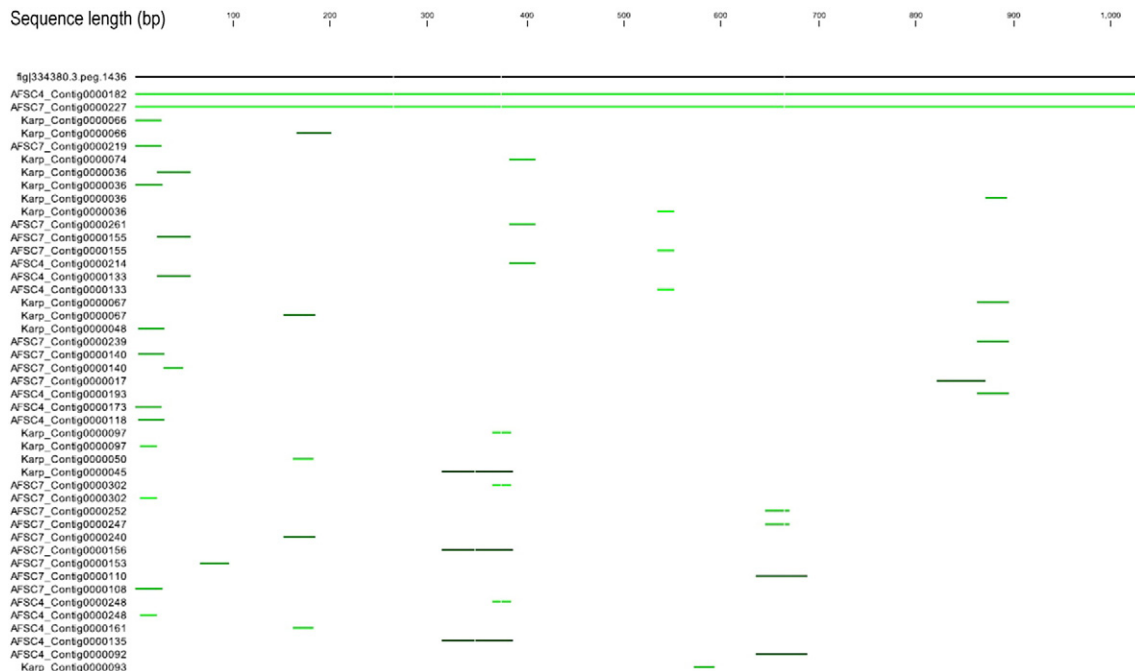


Fig. 3. The blast hits of the CDS#1436 sequence against the draft genomes of OT strain Karp, AFSC4, AFSC7.

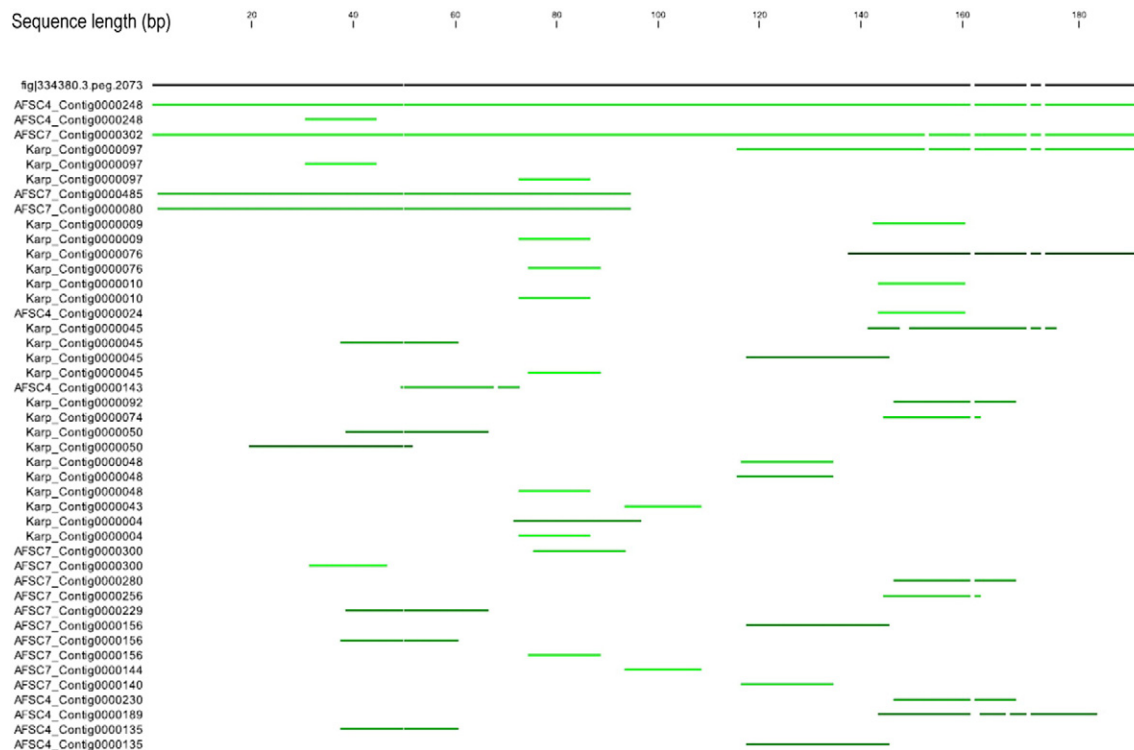


Fig. 4. The blast hits of the CDS# 2073 sequence against the draft genomes of OT strain Karp, AFSC4, AFSC7.

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.J.A0310 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 Wei-Mei Ching are employees of the U. S. Government. This work was prepared 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

- [1] D.J. Kelly, P.A. Fuerst, W.M. Ching, A.L. Richards, Scrub typhus: the geographic distribution of phenotypic and genotypic variants of *Orientia tsutsugamushi*. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 48 (Suppl. 3) (2009) S203–S230, <http://dx.doi.org/10.1086/596576> PubMed PMID: 19220144.
- [2] M.E. Balcells, R. Rabagliati, P. Garcia, H. Poggi, D. Oddo, M. Concha, et al., Endemic scrub typhus-like illness, Chile. *Emerg. Infect. Dis.* 17 (9) (2011) 1659–1663, <http://dx.doi.org/10.3201/eid1709.100960> PubMed PMID: 21888791; PubMed Central PMCID: PMC3322051.
- [3] L. Izzard, A. Fuller, S.D. Blacksell, D.H. Paris, A.L. Richards, N. Aukkanit, et al., Isolation of a novel *Orientia* species (*O. chuto* sp. nov.) from a patient infected in Dubai. *J. Clin. Microbiol.* 48 (12) (2010) 4404–4409, <http://dx.doi.org/10.1128/JCM.01526-10> PubMed PMID: 20926708; PubMed Central PMCID: PMC3008486.
- [4] E. Mathai, J.M. Rolain, G.M. Verghese, O.C. Abraham, D. Mathai, M. Mathai, et al., Outbreak of scrub typhus in southern India during the cooler months. *Ann. N. Y. Acad. Sci.* 990 (2003) 359–364 PubMed PMID: 12860654.
- [5] G. Watt, C. Chouriyagune, R. Ruangveerayud, P. Watcharapichat, D. Phuulsuksombati, K. Jongsakul, et al., Scrub typhus infections poorly responsive to antibiotics in northern Thailand. *Lancet* 348 (9020) (1996) 86–89 PubMed PMID: 8676722.
- [6] G. Watt, P. Kantipong, K. Jongsakul, P. Watcharapichat, D. Phuulsuksombati, Azithromycin activities against *Orientia tsutsugamushi* strains isolated in cases of scrub typhus in Northern Thailand. *Antimicrob. Agents Chemother.* 43 (11) (1999) 2817–2818 PubMed PMID: 10543774; PubMed Central PMCID: PMC389570.
- [7] G. Watt, P. Kantipong, K. Jongsakul, P. Watcharapichat, D. Phuulsuksombati, D. Strickman, Doxycycline and rifampicin for mild scrub-typhus infections in northern Thailand: a randomised trial. *Lancet* 356 (9235) (2000) 1057–1061, [http://dx.doi.org/10.1016/S0140-6736\(00\)02728-8](http://dx.doi.org/10.1016/S0140-6736(00)02728-8) PubMed PMID: 11009140.
- [8] D. Strickman, T. Sheer, K. Salata, J. Hershey, G. Dasch, D. Kelly, et al., In vitro effectiveness of azithromycin against doxycycline-resistant and -susceptible strains of *Rickettsia tsutsugamushi*, etiologic agent of scrub typhus. *Antimicrob. Agents Chemother.* 39 (11) (1995) 2406–2410 PubMed PMID: 8585717; PubMed Central PMCID: PMC3162956.
- [9] H.M. Liao, C.C. Chao, H. Lei, B. Li, S. Tsai, G.C. Hung, et al., Genomic sequencing of *Orientia tsutsugamushi* strain Karp, an assembly comparable to the genome size of the strain Ikeda. *Genome Announc.* 4 (4) (2016) <http://dx.doi.org/10.1128/genomeA.00702-16> PubMed PMID: 27540052; PubMed Central PMCID: PMC4991697.
- [10] G.A. Dasch, S. Halle, A.L. Bourgeois, Sensitive microplate enzyme-linked immunosorbent assay for detection of antibodies against the scrub typhus rickettsia, *Rickettsia tsutsugamushi*. *J. Clin. Microbiol.* 9 (1) (1979) 38–48 PubMed PMID: 107185; PubMed Central PMCID: PMC3272954.
- [11] M. Galardini, E.G. Biondi, M. Bazzicalupo, A. Mengoni, CONTIGuator: a bacterial genomes finishing tool for structural insights on draft genomes. *Source Code Biol Med.* 6 (2011) 11, <http://dx.doi.org/10.1186/1751-0473-6-11> PubMed PMID: 21693004; PubMed Central PMCID: PMC3133546.
- [12] R. Overbeek, R. Olson, G.D. Pusch, G.J. Olsen, J.J. Davis, T. Disz, et al., The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* 42 (Database issue) (2014) D206–D214, <http://dx.doi.org/10.1093/nar/gkt1226> PubMed PMID: 24293654; PubMed Central PMCID: PMC3965101.
- [13] R. Overbeek, T. Begley, R.M. Butler, J.V. Choudhuri, H.Y. Chuang, M. Cohoon, et al., The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.* 33 (17) (2005) 5691–5702, <http://dx.doi.org/10.1093/nar/gki866> PubMed PMID: 16214803; PubMed Central PMCID: PMC1251668.