

RESEARCH ARTICLE

Long-read whole genome sequencing and comparative analysis of six strains of the human pathogen *Orientia tsutsugamushi*

Elizabeth M. Batty^{1,2,3}, Suwittra Chaemchuen², Stuart Blacksell^{2,3}, Allen L. Richards⁴, Daniel Paris^{2,3,5,6}, Rory Bowden¹, Caroline Chan⁷, Ramkumar Lachumanan⁷, Nicholas Day^{2,3}, Peter Donnelly^{1,8}, Swaine Chen^{9,10}, Jeanne Salje^{2,3,11*}

1 Wellcome Centre for Human Genetics, University of Oxford, Oxford, United Kingdom, **2** Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand, **3** Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom, **4** US Naval Medicine Research Center, Silver Spring, Maryland, United States of America, **5** Swiss Tropical and Public Health Institute, Basel, Switzerland, **6** Faculty of Medicine, University of Basel, Basel, Switzerland, **7** Pacific Biosciences, 1305 O'Brien Drive, Menlo Park, California, United States of America, **8** Department of Statistics, University of Oxford, Oxford, United Kingdom, **9** Department of Medicine, Division of Infectious Diseases, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, **10** Genome Institute of Singapore, A*STAR, Singapore, **11** Public Health Research Institute, Rutgers Biomedical and Health Science, Newark, New Jersey, United States of America

* jeanne.salje@ndm.ox.ac.uk



OPEN ACCESS

Citation: Batty EM, Chaemchuen S, Blacksell S, Richards AL, Paris D, Bowden R, et al. (2018) Long-read whole genome sequencing and comparative analysis of six strains of the human pathogen *Orientia tsutsugamushi*. PLoS Negl Trop Dis 12(6): e0006566. <https://doi.org/10.1371/journal.pntd.0006566>

Editor: José Reck, Jr., Instituto de Pesquisas Veterinarias Desiderio Finamor, BRAZIL

Received: April 13, 2018

Accepted: May 28, 2018

Published: June 6, 2018

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: The sequence data and assemblies generated in this study have been uploaded to the EBI under project PRJEB24834. They can be accessed through the following link: <https://www.ebi.ac.uk/ena/data/view/PRJEB24834>.

Funding: JS was funded by a Royal Society Dorothy Hodgkin Research Fellowship (DH140154, <https://royalsociety.org/>). PD is supported by a Wellcome Trust Core Award (090532/Z/09/Z, <https://wellcome.ac.uk/>). This work was partially

Abstract

Background

Orientia tsutsugamushi is a clinically important but neglected obligate intracellular bacterial pathogen of the Rickettsiaceae family that causes the potentially life-threatening human disease scrub typhus. In contrast to the genome reduction seen in many obligate intracellular bacteria, early genetic studies of *Orientia* have revealed one of the most repetitive bacterial genomes sequenced to date. The dramatic expansion of mobile elements has hampered efforts to generate complete genome sequences using short read sequencing methodologies, and consequently there have been few studies of the comparative genomics of this neglected species.

Results

We report new high-quality genomes of *O. tsutsugamushi*, generated using PacBio single molecule long read sequencing, for six strains: Karp, Kato, Gilliam, TA686, UT76 and UT176. In comparative genomics analyses of these strains together with existing reference genomes from Ikeda and Boryong strains, we identify a relatively small core genome of 657 genes, grouped into core gene islands and separated by repeat regions, and use the core genes to infer the first whole-genome phylogeny of *Orientia*.

Conclusions

Complete assemblies of multiple *Orientia* genomes verify initial suggestions that these are remarkable organisms. They have larger genomes compared with most other

supported (ALR) by the US Military Infectious Diseases Research Program, work unit # A1230. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: PD is Founder, Director, and Executive Officer of Genomics plc and a Partner of Peptide Groove LLP. CC and RL are employees of Pacific Biosciences.

Rickettsiaceae, with widespread amplification of repeat elements and massive chromosomal rearrangements between strains. At the gene level, *Orientia* has a relatively small set of universally conserved genes, similar to other obligate intracellular bacteria, and the relative expansion in genome size can be accounted for by gene duplication and repeat amplification. Our study demonstrates the utility of long read sequencing to investigate complex bacterial genomes and characterise genomic variation.

Author summary

Orientia tsutsugamushi is an obligate intracellular bacterial pathogen, and the causative agent of the human disease scrub typhus. This vector-borne bacterial infection can be life-threatening if not treated rapidly with appropriate antibiotics. *Orientia* is endemic across large parts of Asia and is estimated to infect at least one million people per year. *O. tsutsugamushi* has a highly unusual genome. At 2.1 Mbp in length, it is almost double the size of most other obligate intracellular bacterial genomes, which have undergone extensive genome reduction in their adaptation to the intracellular lifestyle. The comparatively expanded genome of *Orientia* is accounted for by an abundance of repeated DNA elements. Indeed, the genome of *Orientia* is the most highly repeated bacterial genome reported to date. As a consequence of this, it has been difficult to generate complete genome assemblies of *Orientia* using short-read sequencing technologies. In the current report, the authors use Pacific Biosciences long-read sequencing to generate six complete genomes of diverse strains of *O. tsutsugamushi*. By comparing these genomes, the authors show that there is very little colinearity in the ordering of genes within the genome, even between closely related strains. This lack of synteny suggests that there have been frequent events of massive chromosomal rearrangement in this species, although the underlying mechanisms for this have not been shown. The genome sequences and analysis reported here will benefit those working on various aspects of the biology of this clinically important but neglected bacterial species.

Introduction

O. tsutsugamushi is an obligate intracellular bacterial pathogen of the order Rickettsiales, family Rickettsiaceae which causes the life-threatening human disease scrub typhus. *Orientia* is transmitted by *Leptotrombidium* mites that occasionally feed on humans during the larval stage of development (“chiggers”), inoculate bacteria into the skin, and initiate infection. *Orientia* is maintained in mite populations by transovarial transmission. The mites normally feed only once on a vertebrate host, and cannot transmit bacteria directly from one host to another [1]. Bacteria propagate within endothelial cells, dendritic cells and monocytes at the site of inoculation, sometimes resulting in a visible red skin feature called an eschar [2]. Bacteria subsequently spread through the endothelial and lymphatic systems to cause a systemic infection characterised by lymphadenopathy, headache, fever, rash and myalgia, which typically begin 7–10 days after inoculation. The non-specificity of these symptoms makes scrub typhus difficult to diagnose based purely on clinical observations, and this is an important reason why the prevalence of scrub typhus has been historically under-recognised. Scrub typhus has now been shown to be a leading cause of severe fever and sepsis in studies in Thailand, India, China, Laos and Myanmar [3] and untreated or severe cases are associated with CNS

infection, morbidity and death [4,5]. Locally acquired cases of scrub typhus have been reported in South America and the Middle East [6,7], suggesting that the global burden of this disease may stretch beyond the traditionally known endemic areas of Asia and Northern Australia [3].

O. tsutsugamushi is distinct from other members of the Rickettsiaceae. The genus *Orientia* currently includes two known species, *O. tsutsugamushi* and *O. chuto*, the latter represented to date by a single strain inoculated from a patient with a febrile illness contracted in Dubai [6]. High antigenic diversity among strains of *O. tsutsugamushi* is reflected in the poor immunological protection that recovered patients exhibit towards strains different from their original infection and this has hampered efforts towards vaccine development.

Despite its importance as a pathogen, few genomic analyses of *O. tsutsugamushi* have been published. The first whole genome sequence, Boryong, [8] reported a proliferation of type IV secretion systems in a repeat-dense genome of which 37.1% comprised identical repeats. A comparison of Boryong and the second complete genome, Ikeda [9], revealed similar repeats present in each genome, dominated by an integrative and conjugative element named the *O. tsutsugamushi* amplified genetic element (OtAGE), and identified a core genome of 520 genes shared between the two *O. tsutsugamushi* strains and the 5 available sequences of other *Rickettsia* [9]. Extensive genomic reshuffling was thought to have been mediated by amplification of repetitive sequences.

In comparison to other *Rickettsiae*, many of which have small and extremely stable genomes, *O. tsutsugamushi* has a large genome with an extraordinary proliferation of repeat sequences and conjugative elements. Some of the conjugative elements present in multiple copies across the genome are homologues of a gene cluster found in a single copy in *Rickettsia bellii*. Many of the genes in these elements are fragmented, suggesting they are non-functional [10]. Other intracellular pathogens also contain repetitive elements, such as the mobile genetic elements in *Wolbachia* [11] and the tandem intergenic repeats in *Ehrlichia ruminantium* [12]. These mechanisms may evolve to increase genetic variability and aid immune evasion in bacteria which cannot easily take up novel DNA.

Larger collections of *O. tsutsugamushi* strains have been extensively studied using MLST and sequence typing of the *groES* and *groEL* [13] genes, and the outer membrane proteins 47kDa (also called HtrA or TSA47) [14] and 56kDa (also called OmpA or TSA56) [15] genes. The 56kDa and 47kDa genes are highly immunogenic in human patients and animal models and have long been investigated as candidates for vaccine design, but high levels of diversity between strains, especially in the 56kDa gene, have limited the potential of developing a universal vaccine based on these epitopes.

Multiple studies in South East Asia have looked at the diversity of strains by MLST and 56kDa typing, and shown a high level of diversity, with many MLSTs unique to an individual strain [16–19]. Work in Thailand and Laos has shown recombination between MLSTs, as well as evidence for multiple infections in individual patients, implying that multiple strains may co-exist in mites [16]. Comparisons of 56kDa typing with MLST [16] and 47kDa [14] also show low congruence between methods, suggesting that single gene typing of *Orientia* may not represent the true relationships between strains; by extension, a 7-gene MLST scheme may not capture the full set of genomic relationships among strains.

Attempts to generate complete *O. tsutsugamushi* genomes by whole genome sequencing have been limited by the difficulties of sequencing and assembling a repeat-dense genome, and no further genomes have been completed since the Boryong and Ikeda genomes in 2008. Current draft assemblies are fragmented with over 50 contigs per genome, and vary in size—the two assemblies of the genome of *O. tsutsugamushi* str. Karp available on Genbank are 1,459,958bp (<https://www.ebi.ac.uk/ena/data/view/LANM01000000>) and 2,022,909bp (<https://www.ncbi.nlm.nih.gov/nuccore/LYMA00000000>) [20] in length, suggesting that assemblies are

either incomplete, or have problems caused by the misassembly of repeats or the inclusion of contaminating sequences.

In this work, we have used Pacific Biosciences long-read sequencing to assemble six complete genomes of *O. tsutsugamushi* strains representing a range of geographical origins and serotypes. This expanded genomic perspective will contribute to our understanding of the phylogeography and epidemiology of this species, as well as contribute to more detailed studies of virulence mechanisms.

Methods

Bacterial propagation

All experiments were performed using *O. tsutsugamushi* grown in the mouse fibroblast cell line L929 (European Collection of Authenticated Cell Cultures 85103115). Uninfected L929 cells were grown in 25 cm² and 75 cm² plastic flasks at 37 °C and 5% CO₂, using DMEM or RPMI 1640 (Thermo Fisher Scientific, USA) media supplemented with 10% FBS (Sigma) as described previously (Giengkam et al 2015). Infected L929 cells were grown in the same way, but at 35 °C. Frozen stocks of bacteria were grown for 5 days, then the bacterial content was calculated using qPCR against the bacterial gene TSA47 [21]. Bacteria were isolated onto fresh L929 cells in 75 cm² flasks at an Multiplicity of Infection of 10:1 and then grown for an additional 7 days. At this point bacteria were isolated from host cells and prepared for DNA extraction.

DNA extraction

The supernatant were removed from infected flasks and replaced with 6–8 ml pre-warmed media. Infected cells were harvested by mechanical scraping and then lysed using a bullet blender (BBX24B, Bullet Blender Blue, Nextadvance, USA) operated at power 8 for 1 min. Host cell debris was removed by centrifugation at 300xg for 3 minutes, and the supernatant was filtered through a 2.0 µm filter unit (Puradisc, GE Whatman, USA). 10 µl of 1.4 µg/µl DNase (Deoxyribonuclease I from bovine pancreas, Merck, UK) was added per 1 ml of bacterial solution, then incubated at room temperature for 30 minutes. This procedure removed excess host cell DNA. The bacterial sample was then isolated by centrifugation at 14,000xg for 10 min at 4 °C, and washed two times with 0.3M sucrose (Merck). After the washing steps were completed DNA was extracted using a QIAGEN DNeasy Blood & Tissue Kit (QIAGEN, UK) following the manufacturer's instructions.

Purified DNA samples were analysed by gel electrophoresis using 0.8% agarose gel, in order to assess the DNA integrity. The yield of genomic DNA was quantified using a nanodrop (Nanodrop 2000, Thermo Scientific, UK) and Qubit Fluorometric Quantitation (Qubit 3.0 Fluorometer, Thermo Scientific).

Sequencing

SMRTBell templates were prepared from purified *Orientia* genomic DNA according to PacBio's recommended protocols (PacBio, USA). Briefly, 20kb libraries were targeted; enrichment for large fragments was done using BluePippin (Sage Science, USA) size selection method or successive Ampure (Beckman Coulter, USA) clean-ups, depending on the original DNA size distribution and quantity, as recommended by PacBio. SMRTBell templates were sequenced on a Pacific Biosciences RSII Sequencer using P6 chemistry with a 240min run time. Genomes were assembled using the RS_HGAP_Assembly.3 protocol from the PacBio SMRTPortal (version 2.3.0), with initial polishing performed on trimmed initial assemblies using the same raw sequencing data with the RS_Resequencing.1 protocol. Each assembly was further polished

using paired-end reads sequenced on an Illumina Miseq machine. Sequencing information and data availability for each sample can be found in [S1 Supporting Information](#). For each assembly, the corresponding Illumina reads were aligned to the PacBio assembly using Stampy v1.0.23 [22]. Pilon v1.16 [23] was then used to generate a final genome, and corrected 2 to 265 errors in the assemblies, with the majority of the errors being single base deletions at the end of A or T homopolymer runs. All genomes were rotated and reverse complemented as needed so that the predicted start codon for the *dnaA* gene formed the first nucleotide in the genome sequence. Sequencing and assembly statistics can be found in [Table 1](#) and [S1 Supporting Information](#). The Boryong, Ikeda, and non-*Orientia* Rickettsial genomes used in this study were obtained from NCBI ([S1 Supporting Information](#)).

The finished assemblies were annotated using Prokka v1.11 [24], using a custom database created from the Boryong and Ikeda strains, which were previously annotated using the NCBI prokaryotic annotation pathway. The Boryong and Ikeda strains were re-annotated using Prokka for consistency with the other samples. Short gene names for all non-hypothetical gene products were checked manually (607 products). Where genes names were present for Boryong and/or Ikeda a consensus name based on these was selected. Where no short name was available, the long gene name was searched for in *E. coli* using the UniProt database, and where a single and unambiguous match was selected this was used. In cases of ambiguity the protein sequence from *Orientia* was used in a BLAST search against *E. coli*, *R. rickettsii* and *H. sapiens* and the short name of the closest match was selected. The key *Orientia* genes *TSA56*, *TSA47*, *TSA22*, *ScaA*, *ScaC*, *ScaD*, and *ScaE* were also manually annotated by taking known protein sequences from the UT76 strain and using BLAST to find the homologous genes in the other strains and give them the correct names.

Repetitive regions of the genome were defined as regions of at least 1000bp in length which had a match with another 1000bp region with up to 100 differences (mismatches, insertions, and deletions) allowed. The repetitive regions were identified with Vmatch [25].

Table 1. Overview of *O. tsutsugamushi* strains.

Strain	Serotype	Original Source	Source in this study	Reference	Genome length (bp)	Contigs	GC percentage	Errors corrected by Illumina sequencing
Boryong*	Boryong	South Korea, human patient, 1980s	-	[50]	2,127,051	1	31	-
Ikeda*	Ikeda	Japan, human patient, 1979	-	[51]	2,008,987	1	31	-
Karp	Karp	New Guinea, human patient, 1943	Naval Medical Research Centre (NMRC)	[52]	2,469,803	1	31	48
Kato	Kato	Niigata, Japan, human patient, 1955	NMRC	[52]	2,319,449	1	31	5
Gilliam	Gilliam	Indian-Burmese border, human patient, 1943	NMRC	[53]	2,465,012	1	31	7
TA686	TA686	Thailand, animal (<i>Tupaia glis</i>), 1963	NMRC	[52]	2,254,485	1	31	28
TA763	TA763	Thailand, animal <i>Rattus rajah</i>), 1963	NMRC	[52]	2,089,396	8	31	88
FPW1038	TA716	Thailand-Burmese border, human patient (pregnant), 2010	Mahidol-Oxford Research Centre (MORU)	[54]	2,035,338	25	31	265
UT76	Karp	Udon Thani, Thailand, human patient, 2003	MORU	[55]	2,078,193	1	30	2
UT176	Karp	Udon Thani, Thailand, human patient, 2004	MORU	[56]	1,932,116	1	30	13

Bacterial strains used in this study and assembly statistics for the 10 assemblies used in this analysis. Genomes marked with * are previously-assembled reference strains.

<https://doi.org/10.1371/journal.pntd.0006566.t001>

The core and accessory genome was identified using Roary [26] with a threshold of 80% sequence identity required to consider two sequences part of the same gene group. Core genes were defined as genes present in every sample and as a single copy in every sample. The accessory genes identified using Roary were re-clustered using CD-Hit [27,28] with a cutoff of 80% identity across 95% of the length of the shortest protein to identify accessory genes which were truncated copies of other proteins. The correlation between gene order in each pair of samples was calculated by taking the order of the genes relative to the Karp strain and calculating the Spearman's rank coefficient between each pair. COG categories were assigned using RPS-BLAST to find matches in the NCBI Conserved Domain Database [29] and assigning a COG category to these using cdd2cog [30]. Repeat genes were identified using protein clusters generated by CD-Hit to find gene groups which were present at more than 1 copy. CD-Hit was used to identify protein clusters on the proteins predicted by Prokka with a cutoff of 80% identity across 90% of the length of the shortest protein. Pseudogenes were identified from CD-Hit protein clusters where at least one protein was a truncated version of the longest protein in the group. As pseudogenes which are truncated at the 5' end will not be annotated by Prokka, tBLASTn [31] was used to screen for any additional pseudogenes in non-genic regions by searching for BLAST hits with protein identity $\geq 80\%$ and an E-value $< 10^{-15}$. This method found a further 26–37 pseudogenes per strain.

Further analysis used BioPython [32] and the GenomeDiagram package [33]. Fig 1 was created with Circos [34]. Statistical tests were carried out in R [35] and the Python SciPy library [36].

Phylogenies were inferred using Maximum Likelihood methods in RaxML [37] under the GAMMA model of rate heterogeneity and bootstrap values calculated using the rapid bootstrap method. The input sequences were aligned with Clustal Omega [38] (for the 56kDa/46kDa/MLST trees) or using the MAFFT alignments produced by Roary (for the core gene tree). Phylogenetic trees were drawn using the ape [39] and phytools [40] R packages, and Robinson-Foulds distances were calculated using the phangorn [41] R package.

Results

Sequencing, assembly, and annotation

Eight genomes (Table 1) were assembled using the PacBio reads to perform initial genome assembly and Illumina sequencing reads to polish the genomes and reduce errors. Six of the eight genomes could be assembled into a single finished contig, while two genomes remain in multiple contigs (Table 1). In addition, two previously assembled reference genomes, *O. tsutsugamushi* str. Boryong and *O. tsutsugamushi* str. Ikeda, were incorporated into our analysis. The genome size ranges from 1.93Mb to 2.47Mb, and the GC content for all strains is consistent at 30–31%. We assessed the genomes to identify core genes shared between all genomes, and look for repetitive regions and repeat genes in each strain. Fig 1 plots the genetic elements of each complete genome.

The number of predicted genes in each strain ranges from 2086 (UT176) to 2709 (Gilliam) (Table 2) and is highly correlated with genome size (Spearman's correlation coefficient 0.94, $p < 2.2 \times 10^{-16}$). A function could not be assigned, by similarity to reported sequences, to 325–547 genes (16 to 22% of the identified coding regions) in each strain.

Core genome analysis

The set of 8 complete, single-contig genomes was used to identify core genes (present in all genomes) and accessory genes (present in a subset of genomes), using the criterion that all members of a group of putative orthologues should be at least 80% identical to all other members of the group. The number of predicted genes and total genome size for the unfinished

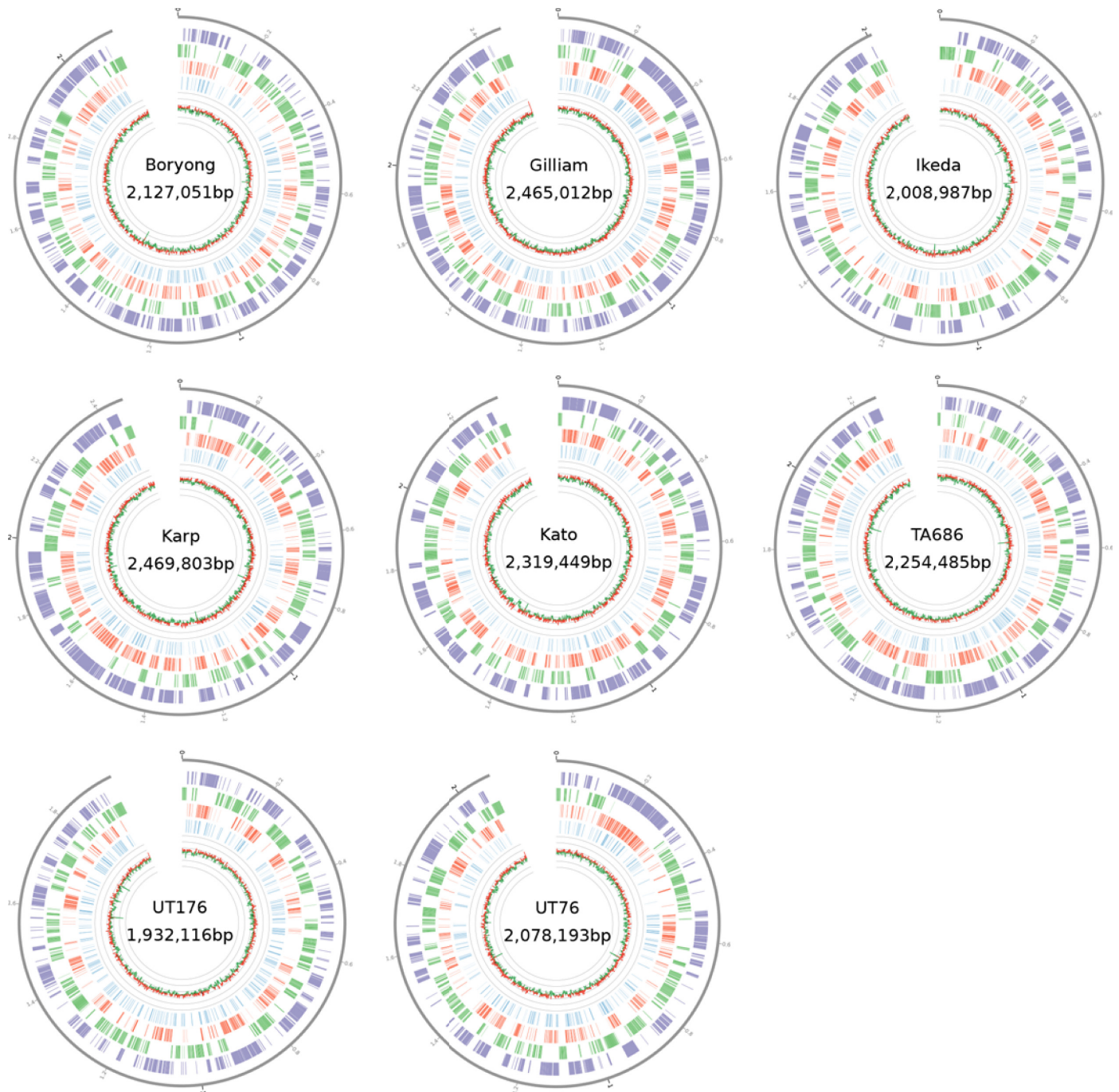


Fig 1. Ring diagrams for all single-contig strains. From outermost feature in each genome, moving inwards: repetitive regions are shown in purple, core genes in green, repeat genes in red and pseudogenes in blue. The track shows the GC percentage in windows of 1000bp. Values above the median GC are in green, and values below the median GC are in red.

<https://doi.org/10.1371/journal.pntd.0006566.g001>

genomes was comparable with those in the finished genomes, which suggests that while the assemblies are not assembled into a single contig they are not missing large pieces of the genome sequence of these strains. However, since we are unable to verify that they contain the complete genome, we have excluded them from the core gene analysis to avoid concluding

Table 2. Predicted gene numbers in different strains of *O. tsutsugamushi*.

Strain	Genes	Annotated as hypothetical
Boryong	2443	547
Ikeda	2186	417
FPW1038	2198	369
Gilliam	2709	463
Karp	2578	470
Kato	2406	465
TA686	2546	474
TA763	2212	396
UT76	2247	420
UT176	2086	325

Number of genes predicted in each strain after annotation with Prokka, and the number of genes which were annotated as hypothetical. The Boryong and Ikeda strains were reannotated with Prokka for consistency between strains.

<https://doi.org/10.1371/journal.pntd.0006566.t002>

that genes are not in the core genome based on their absence from these assemblies. A total of 657 genes were present in all 8 strains and therefore form a putative core genome, while 2812 gene clusters (which may contain multiple genes from a single strain) were present in 2–7 of the 8 strains, and a further 4687 gene clusters were found in a single strain. The 657 core genes make up 28–35% of the genome of each strain (S1 Supporting Information). The number of core genes does not continue to decrease as more genomes are added to the analysis, suggesting that the core genome of *Orientia* can be defined with 8 representative genomes (Fig 2). In the initial analysis with Roary, the total number of gene clusters continued to grow, suggesting an open pan-genome. However, while the total number of distinct gene clusters continues to grow, many of these clusters contain genes which are annotated with the same function as previously observed gene clusters. Of the 6050 clusters where a function can be assigned to one or more gene, there are only 122 different predicted gene products, many of them conjugal transfer proteins, transposases, DNA helicases, and other functions shared by genes known to be part of the *O. tsutsugamushi* amplified genetic element identified in the Ikeda strain [42]. Re-clustering these accessory genes using a more lenient length cut-off to determine clusters allows genes which are only a match to part of a gene sequence to cluster together to include more truncated and fragmented copies of genes. A comparison between the standard Roary clusters and this lenient clustering shows that the number of accessory gene clusters continues to increase, but at a slower rate (S1 Supporting Information). The number of gene products remains constant at 122 no matter how many strains are included in the analysis. This suggests that the increase in non-core gene clusters is mainly due to further duplication and truncation of existing genes, rather than by the import of novel genes.

Genome synteny

With the completed genomes produced by long read sequencing, the synteny of the genomes can be investigated. Previous work on the Boryong and Ikeda genomes showed extensive genome shuffling between the two strains. Analysis of the order and grouping of the core genes which are conserved in each genome shows that the genome has undergone massive rearrangement, with the core genes found in ‘core gene islands’ with repeat regions interspersed between these islands. The 657 core genes are present in 145–157 distinct islands, of which only 51 are conserved (defined as the same genes present in the same order, including

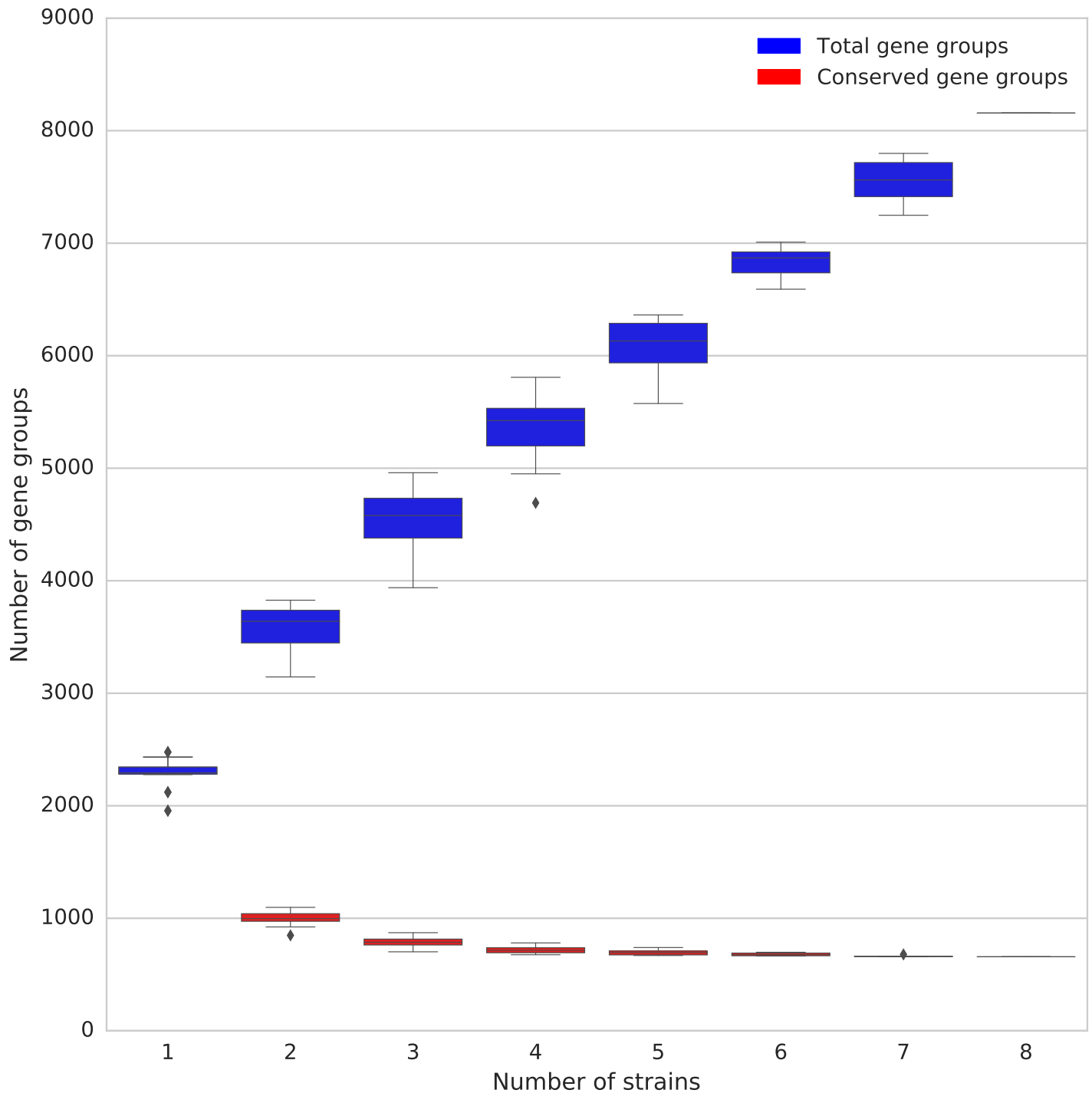


Fig 2. The number of core gene groups and the total number of gene groups (including the core gene groups) as more strains are added to the analysis. Boxplots represent all possible combinations of the number of strains given on the x-axis.

<https://doi.org/10.1371/journal.pntd.0006566.g002>

single member islands) in all genomes. Fig 3 shows the position and ordering of these conserved core gene islands which are maintained in all samples relative to the position and ordering in the Karp strain. The correlation between gene order in each pair of samples is shown in

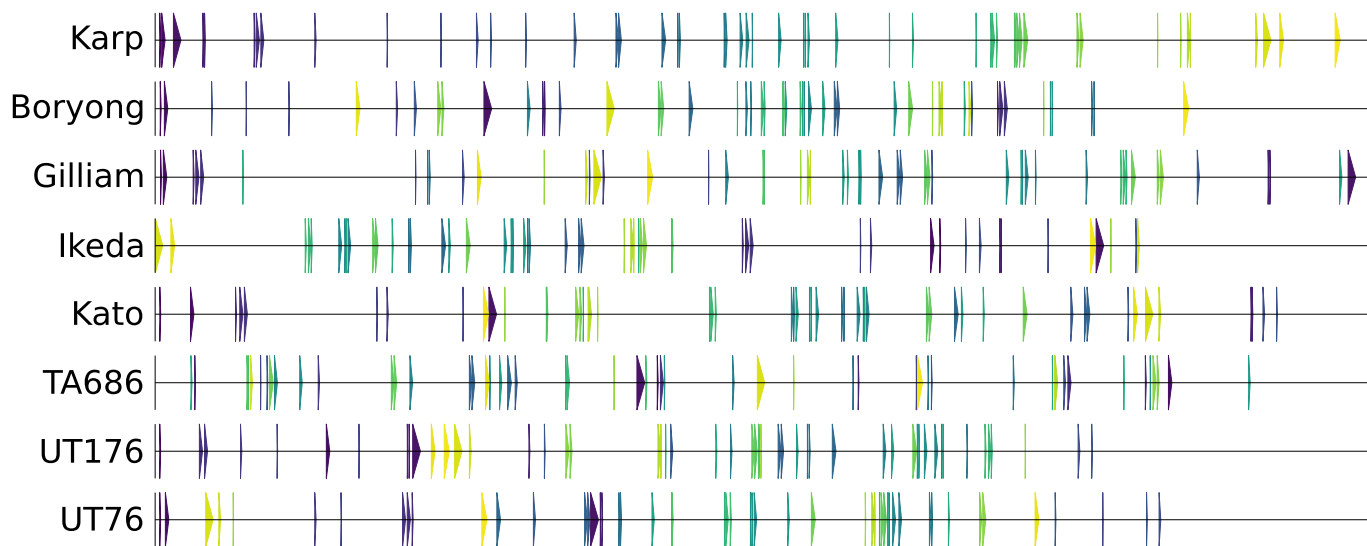


Fig 3. Each arrow represents the location of a core gene island containing one or more core genes which are conserved in the same order within an island across all strains. The arrows are coloured relative to their order in the Karp genome.

<https://doi.org/10.1371/journal.pntd.0006566.g003>

S1 Supporting Information. A value close to 0 shows low correlation in gene order, while values closer to 1 show higher correlation in gene order. As there are differences in the correlation of gene order between strains, this suggests that the process of genome rearrangement is happening in multiple steps and not as a single event.

The identities of genes present on conserved islands is shown in **S1 Supporting Information**. Conserved islands range from 1–13 genes in size, with larger islands often containing genes linked by plausible biological functions. For example, groups 3 and 6 include a number of cell division and peptidoglycan biosynthesis genes (including *mraY*, *murF*, *murE*, *pbp*, *ftsL*, *dnaJ* and *dnaK* in group 3 and *murC*, *murB*, *ddl* and *ftsQ* in group 6) and groups 31 and 32 include a number of 30S and 50S ribosomal proteins. Analysis of the number of conserved islands shared between samples shows that the number of conserved islands continues to decrease as more genomes are included (Figure S3 in **S1 Supporting Information**), and suggests that gene order and clustering is not always constrained in *O. tsutsugamushi*. There is no difference seen in the size of the islands between conserved and non-conserved islands (Figure S4 in **S1 Supporting Information**) (two-sample Kolmogorov-Smirnov test $D = 0.085$, p -value = 0.86), the nucleotide diversity between genes in the two categories of islands (two-sample Kolmogorov-Smirnov test $D = 0.052$, p -value = 0.86) (Figure S5 in **S1 Supporting Information**), or the Clusters of Orthologous Groups (COG) categories assigned to genes in the two island categories (Chi-squared test $\chi^2 = 15.03$, $p = 0.82$) (Figure S6 in **S1 Supporting Information**).

Repeats and pseudogenes

The genomes of *O. tsutsugamushi* are known to be highly repetitive, including a highly amplified genetic element known as the *O. tsutsugamushi* amplified genetic element (OtAGE), as well as other transposable elements.

Our results emphasise the large number of repeated genes and regions, including many genes related to the Type IV secretion system. The total proportion of the genome which is repetitive (see **Methods** for our definition of repetitive) differs markedly from 33% in UT176 to 51% in Gilliam (**S1 Supporting Information**). In contrast, the extremely compact (and

therefore non-repetitive) *Rickettsia typhi* genome is 0% repetitive by our measure and even, intriguingly, the *Rickettsia* endosymbiont of *Ixodes scapularis*, known to encode multiple copies of the same repetitive element found in *Orientia* [43], is 20% repetitive in our analysis, despite our methods giving more conservative figures than previously determined for the Ikeda strain [42].

We identified 530 groups of repeat genes containing 12043 genes present in multiple copies in at least one strain. Of the 530 groups, 427 represent genes found in multiple strains, while 103 are found only in a single strain. Despite clustering in 530 groups, the genes have only 66 different functional products, as is expected from the earlier results looking at all the non-core genes. The repeat genes are mainly transposase and conjugal transfer genes, similar to those previously reported in the OtAGE (S1 Supporting Information), and cluster into genetic elements which are interspersed between the core genes. Many of these genes are present in high copy number, with all strains carrying over 200 transposases and 300 conjugal transfer genes and gene fragments. These 530 repeat elements occupy 35–47% of the *O. tsutsugamushi* genome and represent 57–67% of the genes in these genomes (S1 Supporting Information).

O. tsutsugamushi genes are known to exhibit high levels of pseudogenisation and gene decay. We searched for pseudogenes in each genome, and identified up to 484 pseudogenes per strain (S1 Supporting Information). This is lower than previously reported in Ikeda, but due to methodological differences the figures cannot be directly compared. We also assessed whether the pseudogene had been caused by truncation at the 5' or 3' end of the sequencing, or by frameshift. We did not see a larger number of pseudogenes caused by frameshifts in the genomes new to this study compared to the Boryong and Ikeda strains, suggesting that we do not have a large number of frameshift errors caused by PacBio sequencing.

Phylogenetics

A phylogenetic tree was constructed using the core genes from each strain. This can be compared to trees built using the 56kDa (Fig 4) and 47kDa (Figure S7 in S1 Supporting Information) genes, which are often used for phylogenetic analysis of *O. tsutsugamushi*, or to trees built using the 7 MLST genes from the MLST scheme developed in Sonthayanon et al., 2010 (Figure S8 in S1 Supporting Information). *Orientia* strains are commonly based on their similarity to reference strains, either from phylogenetics or serology. Compared to the 56kDa tree, the core gene tree suggests the Kato and Ikeda strains are more closely related to the Karp, UT176, and UT76 strains than the TA686 and Gilliam strains (Fig 4). Robinson-Foulds distances between trees are shown in S1 Supporting Information; for this small number of strains, the distance is lowest between the 47kDa tree and the core genome tree.

Discussion

We present the first large-scale genomic comparison of *O. tsutsugamushi*, a bacterium which is important both for the study of human disease and for its unique insights into genome evolution.

Previous studies of *O. tsutsugamushi* genomes have used BAC cloning and Sanger sequencing to produce complete genomes [8,42], or have used next-generation sequencing strategies which have produced only incomplete and fragmented genomes [20]. We demonstrate that a combination of PacBio and Illumina sequencing is sufficient to produce a single-contig genome, allowing us to study the gene content and synteny of this organism. For the two genomes which could not be assembled into single contigs in our study (FPW1038 and TA763), we found that the sequencing produced fewer reads at the high end of the length distribution. This suggests that given the highly repetitive nature of the *O. tsutsugamushi* genome, the DNA preparation and sequencing methods must be carefully chosen to produce very long

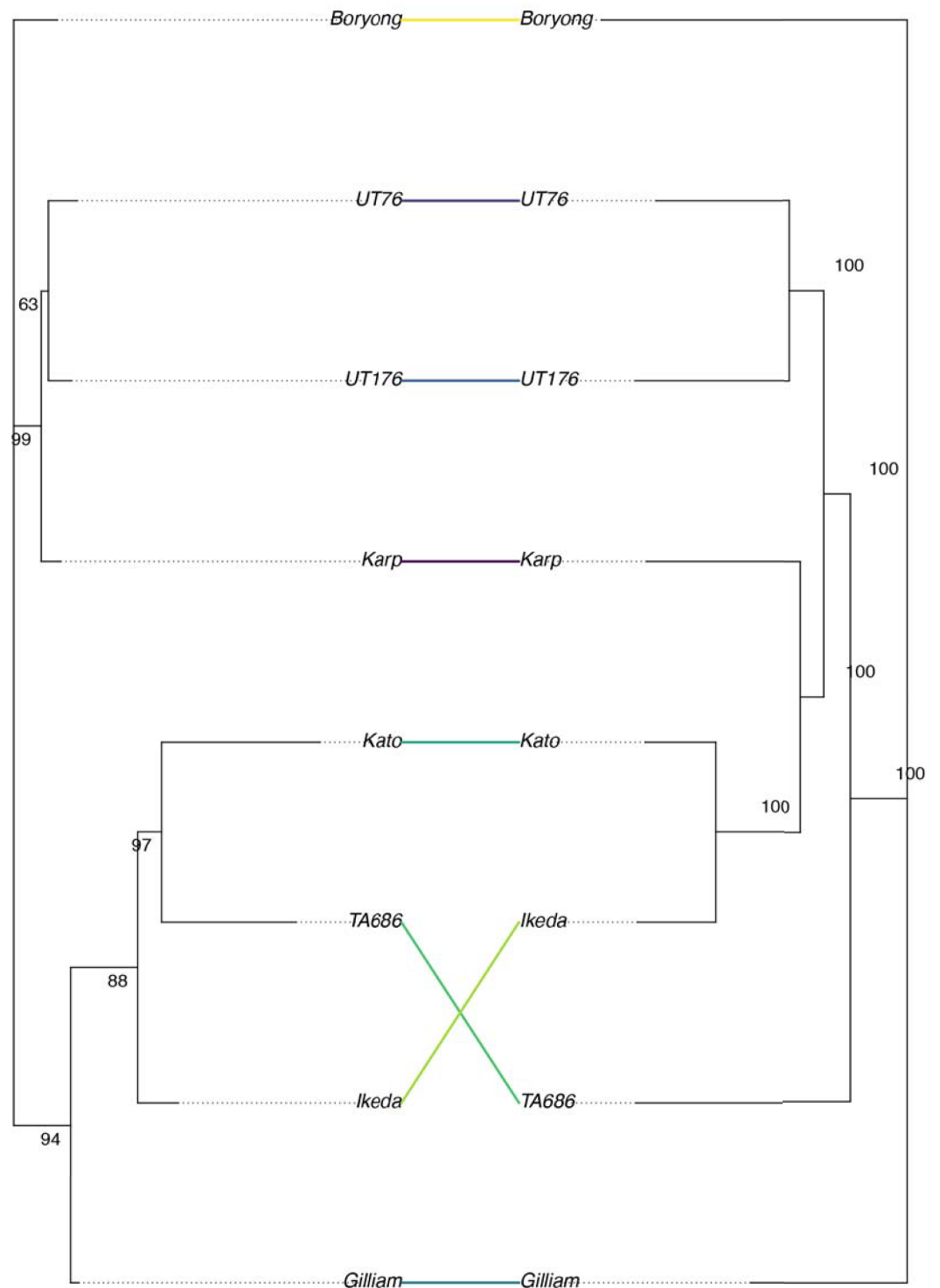


Fig 4. Phylogenetic trees generated from the 56kDa antigen sequence (left) and the sequence of the 657 core genes (right). The tree was inferred using the maximum likelihood method implemented in RaxML, and bootstrap values were calculated with the RaxML rapid bootstrap method.

<https://doi.org/10.1371/journal.pntd.0006566.g004>

reads in order to produce complete assemblies. We used Illumina sequencing to correct errors in our genomes, which was vital to reduce the number of homopolymer errors, which could otherwise suggest frameshift errors and affect gene annotation. While the fewest errors we corrected in a strain was two, this is likely an underestimate as no Illumina reads map with high confidence to 5–15% of the genome due to the repeats, and regions where Illumina reads cannot map will be impossible to correct using this approach. While our analysis shows small differences when quantifying the extent of the repeat regions and repeat gene families in *Orientia* compared to previous work, a direct comparison is difficult due to differences in methodology between analyses.

Owing to the difficulties of producing complete genomes, most previous work has relied on single gene or MLST studies to investigate the genetic diversity of *O. tsutsugamushi*. We demonstrate that phylogenies generated from limited data are substantially different from those produced from the whole core genome. The common practice of grouping *Orientia* strains into ‘Karp-like’ or ‘Gilliam-like’ groups based on the genotype of the 56kDa antigen may not give an accurate view of the relatedness of these strains, especially when recombination is taken into account, although this may still be important when considering immune response. The core gene set we have defined can be sequenced using Illumina reads alone and will allow future studies to perform large-scale phylogenetic analysis of *Orientia* without needing long-read sequencing. We have not investigated the effect of recombination in this study due to the difficulties of analyzing recombination in a highly fragmented and rearranged genome, and it is possible that the core gene tree we present does not represent the true phylogeny of these strains.

Previous work has demonstrated limited synteny between the two reference strains of *O. tsutsugamushi*, but we extend this to demonstrate that there is minimal synteny between any known *O. tsutsugamushi* genome. The pattern of core gene islands separated by transposable elements and repeats suggests a repeat-mediated system of chromosome rearrangement. It is unclear whether this is a gradual process of genome rearrangement, or whether the genome is being broken apart and rearranged rapidly, similar to chromothripsis or the chromosome repair of *Deinococcus radiodurans* after exposure to ionizing radiation. In *Deinococcus*, it is thought that RecFOR pathway is particularly important for DNA repair, and it has no homologues to RecB or RecC [44]. Similarly, in *Orientia*, the core genome does not contain RecB or RecC, but does contain the RecFOR pathway genes, indicating this alternative DNA repair pathway may also be important. Longitudinal studies of *O. tsutsugamushi* genomes during passage or infection may be needed to determine the speed and processes of genome rearrangement in *Orientia*.

We report a core genome of only 657 genes, compared to the 519 previously reported as the core genome shared between *Orientia* and five other sequenced *Rickettsia*, but similar to the 665 core genes shared between the Ikeda and Boryong strains in a previous study [9]. Differences in methodology may lead to the reporting of different core gene sets, but more interesting is the pattern of core genome islands separated by amplified repeat regions, and the lack of conservation in the ordering and clustering of the core genes.

All of the *Orientia* genomes show high repetitiveness, which we measured as both non-unique regions of the genome, and genes which are present in multiple copies (some of which may be truncated). The genomes of intracellular bacteria tend towards genome reduction and gene loss [10,45], but maintain degraded genes and accumulate non-coding DNA. The transition to intracellularity leads to smaller effective population sizes, as the bacteria are sequestered inside cells [46]. As the majority of mutations have a neutral or slightly deleterious effect, they will be removed by purifying selection; however, purifying selection is less effective in smaller bottlenecked populations such as intracellular bacteria, which will lead to an accumulation of slightly deleterious mutations and an increased rate of sequence evolution [47]. The expansion

of the OtAGE (and other mobile elements) throughout the *Orientia* lineage appears to be another consequence of the decreased role of selection on *Orientia* in its intracellular niche, again leading to accelerated sequence evolution of the genome through rearrangement and gene loss. This is supported by the finding that the diversity of gene repertoire between strains of *O. tsutsugamushi* is partly due to the duplication and truncation of existing genes. The amplification of a transposable element has been seen in Rickettsial [43] and non-Rickettsial [48] species, but it is not known whether this is associated with rearrangement of the genome in other species. High levels of genome plasticity and recombination have also been seen in fungal endobacteria, and are thought to be a defence against Muller's ratchet [49]

In conclusion, we report the generation of six complete and a further two draft genomes from a diverse set of strains of the important but neglected human pathogen *O. tsutsugamushi*. This set includes the major reference strains Karp, Kato and Gilliam, and will serve as a valuable resource for scientists and clinicians studying this pathogen, in particular supporting future work on *Orientia* genomics, vaccine development, and cell biology. The new genomes reported here confirm the status of *Orientia* as one of the most fragmented and highly repeated bacterial genomes known, and exciting questions remain regarding the mechanisms and time-frames driving the evolution of these extraordinary genomes.

Supporting information

S1 Supporting Information. File containing all supporting figures and table.
(PDF)

Acknowledgments

Disclaimers

The views expressed in this presentation are those of the authors and do not necessarily represent the official policy or position of the Department of the Navy, Department of Defense, nor the U.S. Government.

Dr. Richards is an employee of the U.S. Government. This work was prepared as a part of his 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 a military service member or employee of the U.S. Government as part of that person's official duties.

Author Contributions

Conceptualization: Elizabeth M. Batty, Rory Bowden, Ramkumar Lachumanan, Nicholas Day, Swaine Chen, Jeanne Salje.

Data curation: Swaine Chen.

Formal analysis: Elizabeth M. Batty.

Funding acquisition: Jeanne Salje.

Investigation: Elizabeth M. Batty, Suwittra Chaemchuen, Swaine Chen.

Methodology: Suwittra Chaemchuen, Caroline Chan.

Project administration: Jeanne Salje.

Resources: Stuart Blacksell, Allen L. Richards, Daniel Paris, Rory Bowden, Ramkumar Lachumanan, Peter Donnelly.

Supervision: Jeanne Salje.

Validation: Swaine Chen.

Writing – original draft: Elizabeth M. Batty, Jeanne Salje.

Writing – review & editing: Stuart Blacksell, Daniel Paris, Rory Bowden, Nicholas Day, Peter Donnelly, Swaine Chen, Jeanne Salje.

References

1. Coleman RE, Monkanna T, Linthicum KJ, Strickman DA, Frances SP, Tanskul P, et al. Occurrence of *Orientia tsutsugamushi* in small mammals from Thailand. *Am J Trop Med Hyg*. 2003; 69: 519–24. PMID: 14695089
2. Paris DH, Phetsouvanh R, Tanganuchitcharnchai A, Jones M, Jenjaroen K, Vongsouvath M, et al. *Orientia tsutsugamushi* in human scrub typhus eschars shows tropism for dendritic cells and monocytes rather than endothelium. *PLoS Negl Trop Dis*. 2012; 6: e1466. <https://doi.org/10.1371/journal.pntd.0001466> PMID: 22253938
3. Luce-Fedrow A, Lehman M, Kelly D, Mullins K, Maina A, Stewart R, et al. A Review of Scrub Typhus (*Orientia tsutsugamushi* and Related Organisms): Then, Now, and Tomorrow. *Trop Med Infect Dis*. 2018; 3: 8. <https://doi.org/10.3390/tropicalmed3010008>
4. Bonell A, Lubell Y, Newton PN, Crump JA, Paris DH. Estimating the burden of scrub typhus: A systematic review. *PLoS Negl Trop Dis*. 2017; 11: e0005838. <https://doi.org/10.1371/journal.pntd.0005838> PMID: 28945755
5. Dittrich S, Rattanavong S, Lee SJ, Panyanivong P, Craig SB, Tulsiani SM, et al. *Orientia*, rickettsia, and leptospira pathogens as causes of CNS infections in Laos: a prospective study. *Lancet Glob Heal*. 2015; 3: e104–12. [https://doi.org/10.1016/S2214-109X\(14\)70289-X](https://doi.org/10.1016/S2214-109X(14)70289-X) PMID: 25617190
6. Izzard L, Fuller A, Blacksell SD, Paris DH, Richards AL, Aukkanit N, et al. Isolation of a Novel *Orientia* Species (*O. chuto* sp. nov.) from a Patient Infected in Dubai. *J Clin Microbiol*. 2010; 48: 4404–4409. <https://doi.org/10.1128/JCM.01526-10> PMID: 20926708
7. Weitzel T, Dittrich S, López J, Phuklia W, Martinez-Valdebenito C, Velásquez K, et al. Endemic Scrub Typhus in South America. *N Engl J Med*. 2016; 375: 954–961. <https://doi.org/10.1056/NEJMoa1603657> PMID: 27602667
8. Cho N-H, Kim H-R, Lee J-H, Kim S-Y, Kim J, Cha S, et al. The *Orientia tsutsugamushi* genome reveals massive proliferation of conjugative type IV secretion system and host-cell interaction genes. *Proc Natl Acad Sci U S A*. 2007; 104: 7981–6. <https://doi.org/10.1073/pnas.0611553104> PMID: 17483455
9. Nakayama K, Kurokawa K, Fukuhara M, Urakami H, Yamamoto S, Yamazaki K, et al. Genome comparison and phylogenetic analysis of *Orientia tsutsugamushi* strains. *DNA Res*. 2010; 17: 281–91. <https://doi.org/10.1093/dnares/dsq018> PMID: 20682628
10. Darby AC, Cho N-H, Fuxelius H-H, Westberg J, Andersson SGE. Intracellular pathogens go extreme: genome evolution in the Rickettsiales. *Trends Genet*. 2007; 23: 511–20. <https://doi.org/10.1016/j.tig.2007.08.002> PMID: 17822801
11. Wu M, Sun L V, Vamathevan J, Riegler M, Deboy R, Brownlie JC, et al. Phylogenomics of the Reproductive Parasite *Wolbachia pipientis* wMel: A Streamlined Genome Overrun by Mobile Genetic Elements. *PLoS Biol*. 2004; 2: e69. <https://doi.org/10.1371/journal.pbio.0020069> PMID: 15024419
12. Frutos R, Viari A, Ferraz C, Morgat A, Eychenié S, Kandassamy Y, et al. Comparative genomic analysis of three strains of *Ehrlichia ruminantium* reveals an active process of genome size plasticity. *J Bacteriol*. 2006; 188: 2533–42. <https://doi.org/10.1128/JB.188.7.2533-2542.2006> PMID: 16547041
13. Arai S, Tabara K, Yamamoto N, Fujita H, Itagaki A, Kon M, et al. Molecular phylogenetic analysis of *Orientia tsutsugamushi* based on the groES and groEL genes. *Vector Borne Zoonotic Dis*. 2013; 13: 825–9. <https://doi.org/10.1089/vbz.2012.1155> PMID: 24107204
14. Jiang J, Paris DH, Blacksell SD, Aukkanit N, Newton PN, Phetsouvanh R, et al. Diversity of the 47-kD HtrA nucleic acid and translated amino acid sequences from 17 recent human isolates of *Orientia*. *Vector Borne Zoonotic Dis*. 2013; 13: 367–75. <https://doi.org/10.1089/vbz.2012.1112> PMID: 23590326
15. Lu H-Y, Tsai K-H, Yu S-K, Cheng C-H, Yang J-S, Su C-L, et al. Phylogenetic analysis of 56-kDa type-specific antigen gene of *Orientia tsutsugamushi* isolates in Taiwan. *Am J Trop Med Hyg*. 2010; 83: 658–63. <https://doi.org/10.4269/ajtmh.2010.09-0608> PMID: 20810835
16. Sonthayanon P, Peacock SJ, Chierakul W, Wuthiekanun V, Blacksell SD, Holden MTG, et al. High rates of homologous recombination in the mite endosymbiont and opportunistic human pathogen

- Orientia tsutsugamushi*. PLoS Negl Trop Dis. 2010; 4: e752. <https://doi.org/10.1371/journal.pntd.0000752> PMID: 20651929
17. Duong V, Blassdell K, May TTX, Sreyrath L, Gavotte L, Morand S, et al. Diversity of *Orientia tsutsugamushi* clinical isolates in Cambodia reveals active selection and recombination process. *Infect Genet Evol.* 2013; 15: 25–34. <https://doi.org/10.1016/j.meegid.2010.08.015> PMID: 20854933
 18. Wongprompitak P, Duong V, Anukool W, Sreyrath L, Mai TTX, Gavotte L, et al. *Orientia tsutsugamushi*, agent of scrub typhus, displays a single metapopulation with maintenance of ancestral haplotypes throughout continental South East Asia. *Infect Genet Evol.* 2015; 31: 1–8. <https://doi.org/10.1016/j.meegid.2015.01.005> PMID: 25577986
 19. Phetsouvanh R, Sonthayanon P, Pukrittayakamee S, Paris DH, Newton PN, Feil EJ, et al. The Diversity and Geographical Structure of *Orientia tsutsugamushi* Strains from Scrub Typhus Patients in Laos. Yang R, editor. *PLoS Negl Trop Dis.* 2015; 9: e0004024. <https://doi.org/10.1371/journal.pntd.0004024> PMID: 26317624
 20. Liao H-M, Chao C-C, Lei H, Li B, Tsai S, Hung G-C, et al. Intraspecies comparative genomics of three strains of *Orientia tsutsugamushi* with different antibiotic sensitivity. *Genomics Data.* 2017. pp. 84–88. <https://doi.org/10.1016/j.gdata.2017.03.012> PMID: 28393016
 21. Giengkam S, Blakes A, Utsahajit P, Chaemchuen S, Atwal S, Blacksell SD, et al. Improved Quantification, Propagation, Purification and Storage of the Obligate Intracellular Human Pathogen *Orientia tsutsugamushi*. *PLoS Negl Trop Dis.* 2015; 9: e0004009. <https://doi.org/10.1371/journal.pntd.0004009> PMID: 26317517
 22. Lunter G, Goodson M. Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res.* 2010; <https://doi.org/10.1101/gr.111120.110> PMID: 20980556
 23. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS One.* 2014; 9: e112963. <https://doi.org/10.1371/journal.pone.0112963> PMID: 25409509
 24. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics.* 2014; 30: 2068–2069. <https://doi.org/10.1093/bioinformatics/btu153> PMID: 24642063
 25. Abouelhoda MI, Kurtz S, Ohlebusch E. Replacing suffix trees with enhanced suffix arrays. *J Discret Algorithms.* 2004; 2: 53–86. [https://doi.org/10.1016/S1570-8667\(03\)00065-0](https://doi.org/10.1016/S1570-8667(03)00065-0)
 26. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics.* 2015; 31: 3691–3. <https://doi.org/10.1093/bioinformatics/btv421> PMID: 26198102
 27. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics.* 2006; 22: 1658–1659. <https://doi.org/10.1093/bioinformatics/btl158> PMID: 16731699
 28. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics.* 2012; 28: 3150–3152. <https://doi.org/10.1093/bioinformatics/bts565> PMID: 23060610
 29. Marchler-Bauer A, Panchenko AR, Shoemaker BA, Thiessen PA, Geer LY, Bryant SH. CDD: a database of conserved domain alignments with links to domain three-dimensional structure. *Nucleic Acids Res.* 2002; 30: 281–3. PMID: 11752315
 30. Leimbach A. bac-genomics-scripts: Bovine *E. coli* mastitis comparative genomics edition. 2016; <https://doi.org/10.5281/ZENODO.215824>
 31. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997; 25: 3389–3402. PMID: 9254694
 32. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics.* 2009; 25: 1422–1423. <https://doi.org/10.1093/bioinformatics/btp163> PMID: 19304878
 33. Pritchard L, White JA, Birch PRJ, Toth IK. GenomeDiagram: a python package for the visualization of large-scale genomic data. *Bioinformatics.* 2006; 22: 616–617. <https://doi.org/10.1093/bioinformatics/btk021> PMID: 16377612
 34. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: An information aesthetic for comparative genomics. *Genome Res.* 2009; 19: 1639–1645. <https://doi.org/10.1101/gr.092759.109> PMID: 19541911
 35. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2014.
 36. Jones E, Oliphant T, Peterson P, others. SciPy: Open source scientific tools for Python Available: <http://www.scipy.org/>

37. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014; 30: 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033> PMID: 24451623
38. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. 2011;7. <https://doi.org/10.1038/msb.2011.75> PMID: 21988835
39. Paradis E, Claude J, Strimmer K. A{PE}: analyses of phylogenetics and evolution in R language. *Bioinformatics*. 2004; 20: 289–290. PMID: 14734327
40. Revell LJ. phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol*. 2012; 3: 217–223.
41. Schliep KP. phangorn: phylogenetic analysis in R. *Bioinformatics*. 2011; 27: 592–593. <https://doi.org/10.1093/bioinformatics/btq706> PMID: 21169378
42. Nakayama K, Yamashita A, Kurokawa K, Morimoto T, Ogawa M, Fukuhara M, et al. The Whole-genome sequencing of the obligate intracellular bacterium *Orientia tsutsugamushi* revealed massive gene amplification during reductive genome evolution. *DNA Res*. 2008; 15: 185–99. <https://doi.org/10.1093/dnares/dsn011> PMID: 18508905
43. Gillespie JJ, Joardar V, Williams KP, Driscoll T, Hostetler JB, Nordberg E, et al. A Rickettsia genome overrun by mobile genetic elements provides insight into the acquisition of genes characteristic of an obligate intracellular lifestyle. *J Bacteriol*. 2012; 194: 376–94. <https://doi.org/10.1128/JB.06244-11> PMID: 22056929
44. Cox MM, Keck JL, Battista JR. Rising from the Ashes: DNA Repair in *Deinococcus radiodurans*. *PLoS Genet*. 2010; 6: e1000815. <https://doi.org/10.1371/journal.pgen.1000815> PMID: 20084105
45. Merhej V, Raoult D. Rickettsial evolution in the light of comparative genomics. *Biol Rev* 2011; 86: 379–405. <https://doi.org/10.1111/j.1469-185X.2010.00151.x> PMID: 20716256
46. Moran NA. Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proc Natl Acad Sci U S A*. 1996; 93: 2873–8. PMID: 8610134
47. McCutcheon JP, Moran NA. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol*. 2012; 10: 13–26. <https://doi.org/10.1038/nrmicro2670> PMID: 22064560
48. Wiens GD, Rockey DD, Wu Z, Chang J, Levy R, Crane S, et al. Genome sequence of the fish pathogen *Renibacterium salmoninarum* suggests reductive evolution away from an environmental *Arthrobacter* ancestor. *J Bacteriol*. 2008; 190: 6970–82. <https://doi.org/10.1128/JB.00721-08> PMID: 18723615
49. Naito M, Pawlowska TE. Defying Muller's Ratchet: Ancient Heritable Endobacteria Escape Extinction through Retention of Recombination and Genome Plasticity. *MBio*. 2016; 7: e02057–15. <https://doi.org/10.1128/mBio.02057-15> PMID: 27329757
50. Chang WH, Kang JS, Lee WK, Choi MS, Lee JH. Serological classification by monoclonal antibodies of *Rickettsia tsutsugamushi* isolated in Korea. *J Clin Microbiol*. 1990; 28: 685–8. PMID: 2110179
51. Tamura A, Takahashi K, Tsuruhara T, Urakami H, Miyamura S, Sekikawa H, et al. Isolation of *Rickettsia tsutsugamushi* antigenically different from Kato, Karp, and Gilliam strains from patients. *Microbiol Immunol*. 1984; 28: 873–82. PMID: 6438448
52. Enatsu T, Urakami H, Tamura A. Phylogenetic analysis of *Orientia tsutsugamushi* strains based on the sequence homologies of 56-kDa type-specific antigen genes. *FEMS Microbiol Lett*. 1999; 180: 163–169. <https://doi.org/10.1111/j.1574-6968.1999.tb08791.x> PMID: 10556707
53. Rights FL, Smadel JE. Studies on scrub typhus; tsutsugamushi disease; heterogeneity of strains of *R. tsutsugamushi* as demonstrated by cross-vaccination studies. *J Exp Med*. 1948; 87: 339–51. PMID: 18904219
54. McGready R, Blacksell SD, Luksameetanasan R, Wuthiekanun V, Jedsadapanpong W, Day NPJ, et al. First report of an *Orientia tsutsugamushi* type TA716-related scrub typhus infection in Thailand. *Vector Borne Zoonotic Dis*. 2010; 10: 191–3. <https://doi.org/10.1089/vbz.2008.0199> PMID: 19492945
55. Blacksell SD, Luksameetanasan R, Kalambaheti T, Aukkanit N, Paris DH, McGready R, et al. Genetic typing of the 56-kDa type-specific antigen gene of contemporary *Orientia tsutsugamushi* isolates causing human scrub typhus at two sites in north-eastern and western Thailand. *FEMS Immunol Med Microbiol*. 2008; 52: 335–342. <https://doi.org/10.1111/j.1574-695X.2007.00375.x> PMID: 18312580
56. Paris DH, Aukkanit N, Jenjaroen K, Blacksell SD, Day NPJ. A highly sensitive quantitative real-time PCR assay based on the groEL gene of contemporary Thai strains of *Orientia tsutsugamushi*. *Clin Microbiol Infect*. 2009; 15: 488–95. <https://doi.org/10.1111/j.1469-0691.2008.02671.x> PMID: 19416296