Journal of Bacteriology

Genome Sequence and Comparative Pathogenomics Analysis of a Salmonella enterica Serovar Typhi Strain Associated with a Typhoid Carrier in Malaysia

Kien-Pong Yap, Han Ming Gan, Cindy Shuan Ju Teh, Ramani Baddam, Lay-Ching Chai, Narender Kumar, Suma Avasthi Tiruvayipati, Niyaz Ahmed and Kwai-Lin Thong
J. Bacteriol. **2012, 194(21):5970. DOI: 10.1128/JB.01416-12.**

JB Journals.ASM.org

GENOME ANNOUNCEMENT

Genome Sequence and Comparative Pathogenomics Analysis of a *Salmonella enterica* Serovar Typhi Strain Associated with a Typhoid Carrier in Malaysia

Kien-Pong Yap,[a,b] Han Ming Gan,[c] Cindy Shuan Ju Teh,[a,b] Ramani Baddam,[d] Lay-Ching Chai,[a,b] Narender Kumar,[d] Suma Avasthi Tiruvayipati,[a,d] Niyaz Ahmed,[a,d] and Kwai-Lin Thong[a,b]

Institute of Biological Sciences, Faculty of Science, University of Malaya, Kuala Lumpur, Malaysia[a]; Laboratory of Biomedical Science and Molecular Microbiology, UMBIO Research Cluster, University of Malaya, Kuala Lumpur, Malaysia[b]; ScienceVision SB, Setia Alam, SeksyenU13, Shah Alam, Selangor, Malaysia[c]; and Pathogen Biology Laboratory, Department of Biotechnology, School of Life Sciences, University of Hyderabad, Hyderabad, India[d]

***Salmonella enterica* serovar Typhi is a human pathogen that causes typhoid fever predominantly in developing countries. In this article, we describe the whole genome sequence of the *S.* Typhi strain CR0044 isolated from a typhoid fever carrier in Kelantan, Malaysia. These data will further enhance the understanding of its host persistence and adaptive mechanism.**

Typhoid fever caused by human-specific *Salmonella enterica* serovar Typhi (*S.* Typhi) remains a major health problem that affects 21.7 million people, with 217,000 deaths worldwide annually (6). *S.* Typhi is transmitted through the oral-fecal route and sometimes persists in the body, establishing an asymptomatic chronic carrier (10, 12). The risk of developing gallbladder diseases, including carcinoma, is also higher among typhoid carriers (5, 10, 21).

Although typhoid fever is endemic in many countries, including Malaysia, little is known about the mechanism of survival and persistence of *S.* Typhi in the host. Therefore, the genome sequence and comparative pathogenomics analysis of carrier strain will provide in-depth understanding of its persistence and adaptive mechanism within its host.

S. Typhi CR0044 was isolated from stool sample of a typhoid carrier in Kelantan, Malaysia, in 2007. This strain was subtyped as ST1 by multilocus sequence typing (14) and was highly similar to the outbreak strain in 2005 by pulsed field gel electrophoresis (PFGE) (2). Genome sequencing of *S.* Typhi strain CR0044 was performed using the Ilumina Genome Analyzer (GA2x, pipeline version 1.60, insert size 300), which generated 1.0 gigabyte of data with a 90× depth coverage and a 73-bp read length. Genome assembly was constructed with Velvet (26) using the *de novo* approach, which generated 201 contigs with a minimum contig length of more than 200 bp and an average size of 23,367 bp. The open reading frames (ORFs) of the resultant contigs were predicted using RAST (1) and Prodigal (13) and subsequently annotated using Blast2GO (4), whereas tRNA and rRNA genes were identified with tRNAscan-SE (17) and RNAmmer (15), respectively. The predicted genome size is approximately 4,769,054 bp, with an average GC content of 52.1% and coding percentage of 85.8. The genome revealed approximately 4,884 coding sequences (CDS) with an average length of 825 bp. The genome also contains predicted 69 tRNA and 22 rRNA genes.

The genome revealed a type III secretion system and flagellum subsystem as reported in *S.* Typhi strains Ty2 and CT18 (7, 12). The genome contains genes reported in Ty2 and CT18, such as the gene coding for type 4 fimbrial assembly protein, the *yjbEFGH* locus, *yhjD* conserved clusters, and *wca* genes, which are related to cell wall and biofilm formation and host persistence (3, 8, 7, 12, 18, 25). It is noteworthy that the genome sequence also revealed the presence of the GGDEF family protein YeaJ, which is associated with cell surface adhesion and biofilm formation, which was not identified in Ty2 and CT18 (9, 19). The gene encoding the rhamnogalacturonide transporter RhiT for rhamnose utilization was also found adjacent to a transposase gene in CR0044 (20). Interestingly, the genome also revealed a zonular occludens toxin family protein that was not previously reported in *Salmonella* spp.

S. Typhi in Southeast Asia is genetically diverse, with genome variations and clonal expansion reported (2, 11, 15, 16, 21, 22, 23, 24). The dynamic nature of the *S.* Typhi chromosome greatly enhances its persistence and adaptation within the host, which allows the pathogen to survive and thrive in typhoid carriers. The genomic information obtained here could unveil the genome evolution and mechanism involved in carrier-state transformation.

Nucleotide sequence accession numbers. This Whole Genome Shotgun project has been deposited in GenBank under accession no. AKZO00000000. The version described in this paper is the first version, AKZO01000000. The Bioproject designation for this project is PRJNA160187.

ACKNOWLEDGMENTS

This research is supported by a University of Malaya High Impact Research Grant—Molecular Genetics (reference no. UM.C/625/1HIR/MOHE/-02 [A000002-5000 1]), for which K.-L. Thong is the grant holder for the high-impact research project under the title "Pathogenomic and Phenomic of Food-Borne Disease." We also acknowledge Indo-German International Research Training Group—Internationales Graduiertenkolleg (GRK1673)—Functional Molecular Infection Epidemiology, an initiative of the German Research Foundation (DFG) with N.A. from the University of Hyderabad (India). We are also grateful to M/s Genotypic Technology Pvt., Ltd., Bengaluru, India, for support with Illumina sequencing.

We acknowledge Safwan Jusoh from the ICT department, University of Malaya, for assisting us with computing solutions and allowing us to

Received 6 August 2012 **Accepted** 17 August 2012

Address correspondence to Kwai-Lin Thong, thongkl@um.edu.my.

doi:10.1128/JB.01416-12

use their servers and computational facilities and Soo Tein Ngoi from LBSMM, IPS, for technical assistance with DNA preparation.

REFERENCES

1. **Aziz RK, et al.** 2008. The RAST server: rapid annotations using subsystems technology. BMC Genomics **9**:75. doi:10.1186/1471-2164-9-75.
2. **Baddam R, et al.** 2012. Genetic fine structure of a *Salmonella enterica* serovar Typhi strain associated with the 2005 outbreak of typhoid fever in Kelantan, Malaysia. J. Bacteriol. **194**:3565–3566.
3. **Cano DA, Bernal GD, Tierrez A, Portillo FG, Casadesus J.** 2002. Regulation of capsule synthesis and cell motility in *Salmonella enterica* by the essential gene igaA. Genetics **162**:1513–1523.
4. **Conesa A, et al.** 2005. Blast2Go: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics **21**: 3674–3676.
5. **Crawford RW, et al.** 2010. Gallstones play a significant role in Salmonella spp. gallbladder colonization and carriage. Proc. Natl. Acad. Sci. U. S. A. **107**:4353–4358.
6. **Crump J, Mintz E.** 2010. Global trends in typhoid and paratyphoid fever. Clin. Infect. Dis. **50**:241–246.
7. **Deng W, et al.** 2003. Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2 and CT18. J. Bacteriol. **185**:2330–2337.
8. **Ferrieres L, Aslam SN, Cooper RM, Clarke DJ.** 2007. The yjbEFGH locus in Escherichia coli K-12 is an operon encoding proteins involved in exopolysaccharide production. J. Microbiol. **153**:1070–1080.
9. **Garcia B, et al.** 2004. Role of the GGDEF protein family in Salmonella cellulose biosynthesis and biofilm formation. J. Mol. Microbiol. **54**:264–277.
10. **Gonzalez-Escobedo G, Marshall JM, Gunn JS.** 2011. Chronic and acute infection of the gall bladder by *Salmonella* Typhi: understanding the carrier state. Nat. Rev. Microbiol. **9**:9–14.
11. **Holt KE, et al.** 2011. Temporal fluctuation of multidrug resistant *Salmonella* Typhi haplotypes in the Mekong River delta region of Vietnam. PLoS Negl. Trop. Dis. **5**:e929. doi:10.1371/journal.pntd.0000929.
12. **Holt KE, et al.** 2008. High throughput sequencing provides insights into genome variation and evolution in *Salmonella* Typhi. Nat. Genet. **40**:987–993.
13. **Hyatt D, et al.** 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11**:119. doi: 10.1186/1471-2105-11-119.
14. **Kidgell C, et al.** 2002. Salmonella typhi, the causative agent of typhoid fever, is approximately 50,000 years old. Infect. Genet. Evol. **2**:39–45.
15. **Lagesen K, et al.** 2007. RNammer: consistent annotation of rRNA genes in genomic sequences. Nucleic Acids Res. **35**:3100–3108.
16. **Le TA, et al.** 2007. Clonal expansion and microevolution of quinolone-resistant *Salmonella enterica* serotype Typhi in Vietnam from 1996 to 2004. J. Clin. Microbiol. **45**:3485–3492.
17. **Lowe TM, Eddy SR.** 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25**: 955–964.
18. **Mamat U, et al.** 2008. Single amino acid substitutions in either YhjD or MsbA confer viability to 3-deoxy-D-manno-oct-2-ulosonic acid-depleted Escherichia coli. J. Mol. Microbiol. **67**:633–648.
19. **Parkhill J, et al.** 2001. Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. Nature **413**:848–852.
20. **Rodionov DA, Gelfand MS, Pattat NHC.** 2004. Comparative genomics of the KdgR regulon in Erwinia chrysanthemi 3937 and other gammaproteobacteria. J. Microbiol. **150**:3571–3590.
21. **Shukla VK, Singh H, Pandey M, Upadhyay SK, Nath G.** 2000. Carcinoma of the gallbladder—is it a sequel of typhoid? Dig. Dis. Sci. **45**:900–903.
22. **Thong KL, Cheong YM, Puthucheary SD, Koh CL, Pang T.** 1994. Epidemiologic analysis of sporadic *Salmonella* Typhi isolates and those from outbreaks by pulsed-field gel electrophoresis, J. Clin. Microbiol. **32**: 1135–1141.
23. **Thong KL, Puthucheary SD, Pang T.** 1996. Genome size variation among recent human isolates of *Salmonella* Typhi. Res. Microbiol. **148**:229–235.
24. **Thong KL, et al.** 1995. Analysis of *Salmonella* Typhi from Southeast Asia by pulsed-field gel electrophoresis. J. Clin. Microbiol. **33**:1938–1941.
25. **Townsend SM, et al.** 2001. *Salmonella enterica* serovar Typhi possesses a unique repertoire of fimbrial gene sequences. J. Infect. Immun. **69**:2894–2901.
26. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. **18**:821–829.