JB Journals.ASM.org

GENOME ANNOUNCEMENT

Insights from the Genome Sequence of a *Salmonella enterica* Serovar Typhi Strain Associated with a Sporadic Case of Typhoid Fever in Malaysia

Kien-Pong Yap,[a,b] Cindy Shuan Ju Teh,[a,b] Ramani Baddam,[c] Lay-Ching Chai,[a,b] Narender Kumar,[c] Tiruvayipati Suma Avasthi,[a,c] Niyaz Ahmed,[a,c] 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]; and Pathogen Biology Laboratory, Department of Biotechnology, School of Life Sciences, University of Hyderabad, Hyderabad, India[c]

***Salmonella enterica* serovar Typhi is the causative agent of typhoid fever, which causes nearly 21.7 million illnesses and 217,000 deaths globally. Herein, we describe the whole-genome sequence of the *Salmonella* Typhi strain ST0208, isolated from a sporadic case of typhoid fever in Kuala Lumpur, Malaysia. The whole-genome sequence and comparative genomics allow an in-depth understanding of the genetic diversity, and its link to pathogenicity and evolutionary dynamics, of this highly clonal pathogen that is endemic to Malaysia.**

Salmonella enterica serovar Typhi (*S.* Typhi) is a human intracellular pathogen of global importance, infecting 21.7 million people and causing 217,000 deaths annually (4). The challenges of higher incidence of typhoid fever in developing countries have led to an increased health burden (4).

In 2008, 201 cases of sporadic outbreaks were reported in Malaysia (18). The genome sequence of this sporadically associated strain will provide insights into possible genetic events that would confer a fitness advantage.

S. Typhi strain ST0208 was isolated from the stool sample of a typhoid fever patient admitted to University Malaya Medical Centre (UMMC), Kuala Lumpur, Malaysia, in 2008. The strain was characterized by pulsed-field gel electrophoresis (PFGE), repetitive extragenic palindromic (REP)-PCR, and antimicrobial susceptibility profiling (25). The genome sequence of *S.* Typhi ST0208 was determined using an Illumina genome analyzer (GA2x, pipeline version 1.60) with an insert size of 300 bp, which generated 1.83 gigabytes of data with an average coverage of 165× and yielded 1,499,986 paired-end reads with a 100-bp read length. Genome assembly was constructed *de novo* using Velvet (26), which generated 222 contigs. The resultant contigs were uploaded into the RAST server (2, 17) to predict the open reading frames (ORFs) by using Glimmer3 (5) and validated with the ISGA integrated system (8). In brief, the predicted ORFs were annotated by searching against clusters of orthologous group (21) and SEED (7) databases, whereas tRNA and rRNA genes were identified by using tRNAscan-SE (15) and RNAmmer (12), respectively. The draft genome size is approximately 4,798,272 bp in length, with an average GC content of 52.0%, and is composed of 4,890 predicted coding sequences with an average length of 810 bp. A mean percentage of 83.7% of nucleotides of the genome are predicted to encode proteins. The genome reveals 71 tRNA and 22 rRNA predicted genes.

The genome contains several monosaccharide and polysaccharide metabolism-related genes, which were not reported in *S.* Typhi strains Ty2 and CT18 (6, 11, 18), such as D-galactarate permease, gluconate permease, tagatose-6-phosphate kinase, trehalase, and arabinose-proton transporter, that could be associated with host persistence (20). The genome sequence revealed four multidrug resistance clusters, *mdtABCR* and *marABCR* proteins (1, 3, 19), DNA gyrase subunit A and B, and topoisomerase subunit (IV) A and B (9), which were also identified in *S.* Typhi Ty2 and *S.* Typhi CT18 (6, 11, 19). Hypothetical proteins and pathogenicity islands were annotated.

High genetic diversity of *S.* Typhi was detected among human strains in Malaysia and Southeast Asia (10, 14, 22, 23, 24). Variation in the genome sequence revealed by the strain ST0208 is consistent with the proposed key theory of persistent adaptation and optimization of function (13, 16, 19). Genomic information from locality-specific strains associated with clinical manifestation allows genome evolution and endemicity to be studied.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited in DDBJ/EMBL/GenBank under the accession number AJXA00000000. The version described in this paper is the first version, AJXA01000000. The BioProject designation for this project is PRJNA160181.

ACKNOWLEDGMENTS

This research is supported by the University of Malaya High Impact Research Grant (reference UM.C/625/1HIR/MOHE/-02 [A000002-50001]), Molecular Genetics, in which K.-L. Thong is the grant holder for the high-impact research project under the title of Pathogenomics and Phenomics of Food-borne Bacterial Diseases.

We are also grateful to M/s Genotypic Technology Pvt. Ltd., Bengaluru, India, for their assistance with Illumina sequencing. We acknowledge Safwan Jusoh from the ICT department, University of Malaya, for assisting us with computing solutions and allowing us to use the servers and computational facilities and S. T. Ngoi from LBSMM, IPS, for technical assistance with DNA preparation.

REFERENCES

1. **Alekshun MN, Levy SB.** 1999. The mar regulon: multiple resistance to antibiotics and other toxic chemicals. Trends Microbiol. 7:410–413.

Received 15 June 2012 **Accepted** 29 June 2012

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

doi:10.1128/JB.01062-12

2. **Aziz RK, et al.** 2008. The RAST server: rapid annotations using subsystems technology. BMC Genomics **9**:75.
3. **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.
4. **Crump J, Mintz E.** 2010. Global trends in typhoid and paratyphoid fever. Clin. Infect. Dis. **50**:241–246.
5. **Delcher AL, Bratke KA, Powers EC, Salzberg SL.** 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics **23**:673–679.
6. **Deng W, et al.** 2003. Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2 and CT18. J. Bacteriol. **185**:2330–2337.
7. **Disz T, et al.** 2010. Accessing the SEED genome database via web services API: tools for programmers. BMC Bioinformatics **11**:319.
8. **Hemmerich C, Buechlein A, Podicheti R, Revanna KV, Dong Q.** 2010. An Ergatis-based prokaryotic genome annotation web server. Bioinformatics **26**:1122–1124.
9. **Hirose K, et al.** 2002. DNA sequence analysis of DNA gyrase and DNA topoisomerase IV quinolone resistance-determining regions of *Salmonella enterica* serovar Typhi and serovar Paratyphi A. Antimicrob. Agents Chemother. **46**:3249–3252.
10. **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.
11. **Holt KE, et al.** 2008. High throughput sequencing provides insights into genome variation and evolution in *Salmonella* Typhi. Nat. Genet. **40**:987–993.
12. **Lagesen K, et al.** 2007. RNAmmer: consistent annotation of rRNA genes in genomic sequences. Nucleic Acids Res. **35**:3100–3108.
13. **Lapierre P, Gogarten JP.** 2009. Estimating the size of the bacterial pangenome. Trends Genet. **25**:107–110.
14. **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.
15. **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.
16. **Matthews TD, Rabsch W, Maloy S.** 2011. Chromosomal rearrangements in *Salmonella enterica* serovar Typhi strains isolated from asymptomatic human carriers. mBio **2**:e00060-11. doi:10.1128/mBio.00060-11.
17. **Meyer F, et al.** 2008. The metagenomic RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics **9**:386.
18. **Ministry of Health Malaysia.** 2008. Epidemiology of foodborne diseases in Malaysia, p 89. Director General Ministry of Health Malaysia technical report. Ministry of Health Malaysia, Putrajaya, Malaysia.
19. **Parkhill J, et al.** 2001. Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. Nature **413**:848–852.
20. **Shelburne SA, et al.** 2008. A direct link between carbohydrate utilization and virulence in the major human pathogen group A Streptococcus. Proc. Natl. Acad. Sci. U. S. A. **105**:1698–1703.
21. **Tatusov RL, Koonin EV, Lipman DJ.** 1997. A genomic perspective on protein families. Science **278**:631–637.
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. **Tiong V, et al.** 2010. Macrorestriction analysis and antimicrobial susceptibility profiling of *Salmonella enterica* at a university teaching hospital, Kuala Lumpur. Jpn. J. Infect. Dis. **63**:317–322.
26. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. **18**:821–829.