Complete Genome Sequence of *Salmonella enterica* subsp. *enterica* Serovar Typhi P-stx-12

Su Yean Ong,[a] Chandra Bhan Pratap,[b] Xuehua Wan,[c] Shaobin Hou,[c] Ahmad Yamin Abdul Rahman,[a] Jennifer A. Saito,[a] Gopal Nath,[b] and Maqsudul Alam[a]

Centre for Chemical Biology, Universiti Sains Malaysia, Bayan Lepas, Penang, Malaysia[a]; Department of Microbiology, Institute of Medical Sciences, Banaras Hindu University, India[b]; and Advanced Studies in Genomics, Proteomics and Bioinformatics, University of Hawaii, Honolulu, Hawaii, USA[c]

We report here the complete genome sequence of *Salmonella enterica* subsp. *enterica* serovar Typhi P-stx-12, a clinical isolate obtained from a typhoid carrier in India.

Salmonella enterica serovar Typhi is the causative agent of typhoid fever. *S.* Typhi does not have an animal reservoir and can be transmitted from a typhoid carrier only through contaminated water or food (11). It was estimated that the global incidence of typhoid is 16,000,000 cases, with 500,000 deaths per year (9). In this study, we isolated and sequenced the *S.* Typhi strain of a chronic carrier from a region in India where the disease is highly endemic.

Whole-genome sequencing was performed with both Roche 454 and Illumina paired-end sequencing technologies. A 4-kb genomic library was constructed and 177,021 paired-end and 65,478 single-end reads were generated using the GS FLX Titanium system, giving ~18-fold coverage of the genome. A total of 97.09% of the reads were assembled into 11 scaffolds using Newbler (Roche). A total of ~500 Mbp of 3-kb mate pair (MP) sequencing data (100-fold coverage) were generated with an Illumina Solexa GA IIx. These sequences were mapped to the scaffolds by using the Burrows-Wheeler Alignment (BWA) tool (7). Gaps were closed by sequencing PCR products. Coding sequences were predicted using the ISGA (Integrative Services for Genomic Analysis) pipeline (5) and DIYA (Do-It-Yourself Annotator) pipeline (12), which comprises Glimmer (3), tRNAscan-SE (8), RNAmmer (6), BLAST (1), and Asgard (2). Annotation results were improved and checked using CLC Genomics Workbench.

The complete genome of *S.* Typhi P-stx-12 consists of a single circular chromosome of 4,768,352 bp with a GC content of 52.1% and a 181,431-bp plasmid with a GC content of 46.4%. The chromosome consists of 4,691 predicted coding sequences (CDS), 22 rRNA genes, and 76 tRNA genes, while the plasmid consists of 234 protein-coding genes. Over 75% of the genes were assigned to specific clusters of orthologous groups (COG), and approximately 25% were assigned an enzyme classification number and were involved in 268 predicted metabolic pathways. A clustered regularly interspaced short palindromic repeat (CRISPR) element was detected in the chromosome at position 2900675 to 2901069.

The genome of *S.* Typhi P-stx-12 was significantly different from the other two published genomes of *S.* Typhi strains, CT18, which was isolated in Vietnam (GenBank accession number AL513382), and Ty2, which was isolated in Russia (GenBank accession number AE014613). Comparison between these three genomes revealed that the coding genes of *S.* Typhi P-stx-12 were 84% similar to those of CT18 (10) and Ty2 (4).

The 17 pathogenic islands which were found in the previous two genomes were also identified in *S.* Typhi P-stx-12. This strain has one plasmid which shares 169 orthologous CDS with pHCM1, the plasmid belonging to CT18 (GenBank accession number AL513383). It is worth noting that the plasmid of P-stx-12 carries genes encoding the tetracycline resistance protein and tetracycline repressor protein TetR, possibly conferring drug resistance to this strain. Interestingly, this genome has fewer pseudogenes than CT18 and Ty2 but a higher number of hypothetical proteins.

Nucleotide sequence accession numbers. The genome sequences of *S.* Typhi P-stx-12 have been deposited in GenBank under accession numbers CP003278 (chromosome) and CP003279 (plasmid).

REFERENCES

1. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J. Mol. Biol. **215**:403–410.
2. **Alves JM, Buck GA.** 2007. Automated system for gene annotation and metabolic pathway reconstruction using general sequence databases. Chem. Biodivers. **4**:2593–2602.
3. **Delcher AL, Bratke KA, Powers EC, Salzberg SL.** 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics **23**:673–679.
4. **Deng W, et al.** 2003. Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2 and CT18. J. Bacteriol. **185**:2330–2337.
5. **Hemmerich C, Buechlein A, Podicheti R, Revanna KV, Dong Q.** 2010. An Ergatis-based prokaryotic genome annotation web server. Bioinformatics **26**:1122–1124.
6. **Lagesen K, et al.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. **35**:3100–3108.
7. **Li H, Durbin R.** 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics **26**:589–595.
8. **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.

9. **Pang T, Bhutta ZA, Finlay BB, Altwegg M.** 1995. Typhoid fever and other salmonellosis: a continuing challenge. Trends Microbiol. **3:**253–255.
10. **Parkhill J, et al.** 2001. Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. Nature **413:**848–852.
11. **Raffatellu M, Tukel C, Chessa D, Wilson RP, Baumler AJ.** 2007. The intestinal phase of salmonella infections, p 30–51. *In* Rhen M, Maskell D, Mastroeni P, Threlfall EJ (ed), Salmonella: molecular biology and pathogenesis. Horizon Bioscience, Norfolk, United Kingdom.
12. **Stewart AC, Osborne B, Read TD.** 2009. DIYA: a bacterial annotation pipeline for any genomics lab. Bioinformatics **25:**962–963.