

Complete genome sequence of *Salmonella* Enteritidis MFDS1004839 isolated from food

Woojung Lee[†], Sewook Park[†], Ran Hee Yoo, In-Sun Joo, Hyo Sun Kwak, and Soon Han Kim*

Food Microbiology Division, Ministry of Food and Drug Safety, Cheongju 28159, Republic of Korea

식품에서 분리된 *Salmonella* Enteritidis MFDS1004839의 유전체 서열 분석

이우정[†] · 박세욱[†] · 유란희 · 주인선 · 곽효선 · 김순한*

식품의약품안전처 식품의약품안전평가원 미생물과

(Received February 27, 2018; Revised May 8, 2018; Accepted May 8, 2018)

Salmonella enterica subsp. *enterica* is a foodborne pathogen that has been detected throughout the world. Here, we present the complete genome sequence of *Salmonella* Enteritidis isolated from a commercial kimchap that caused foodborne illness in the Republic of Korea in 2014. Complete genome sequence analysis of *Salmonella* Enteritidis MFDS1004839 revealed a 4,679,649 bp chromosome and a 96,994 bp plasmid, with G + C contents of 52.2% and 49.3%, respectively. The chromosome and plasmid genome included 4,482 predicted protein-coding sequences, 84 tRNAs and 22 rRNAs genes.

Keywords: *Salmonella* sp., foodborne illness, genome, kimchap

Salmonella spp. is a major foodborne and waterborne pathogen in the Republic of Korea causing disease in humans and animals (Lee *et al.*, 2001; Kim, 2010). *Salmonella* pathogenicity has been associated with the presence of virulence factors, such as adhesion, invasion, and toxin genes (Ochman *et al.*, 2000; Pallen and Wren, 2007). Those genes play important roles in invasion, survival, and proliferation inside the host cells (Das *et al.*, 2012). Whole genome sequencing is employed for the identifi-

cation and characterization of pathogenic bacteria. *Salmonella enterica* subsp. *enterica* Serovar Enteritidis (also known as *Salmonella* Enteritidis) MFDS1004839 was isolated from a commercial kimchap that caused foodborne disease in the Republic of Korea in 2014. The isolate has been deposited at Korean Culture Collection for Foodborne Pathogens under the strain number MFDS1004839. Here, we report the complete genome sequence of *Salmonella* Enteritidis MFDS1004839.

The genomic DNA was extracted from bacterial cells grown on tryptic soy agar using the HiGeneTM Genomic DNA Prep Kit (Biofact) and the DNA was then sheared into 10 kb to 20 kb fragments using g-Tubes (Covaris, Inc.). The sheared DNA was used to generate a Pacific Biosciences (PacBio) library using a template preparation kit according to the manufacturer's instructions. Sequencing was performed using the single-molecule real-time (SMRT) sequencing system (PacBio RS II) with P6/C4 chemistry, using a library constructed by Chunlab Inc.

The assembly process was performed using the HGAP/Quiver software package in the SMRT analysis program version 2.3.0 (Pacific Biosciences). The complete alignment showed approximately 161.32 × coverage and two contigs were generated. Coding DNA sequences (CDSs) and tRNAs were predicted using prodigal (Hyatt *et al.*, 2010) and tRNAscan-SE (Lowe

[†]These authors contributed equally to this work.

*For correspondence. E-mail: lambdog@korea.kr;
Tel.: +82-43-719-4302; Fax: +82-43-719-4300

Table 1. Genome features of *Salmonella* Enteritidis MFDS1004839

Features	Chromosome	Plasmid
Molecular shape	Circular	Circular
Size (bp)	4,679,649	96,994
G + C content (%)	52.2	49.3
Total number of		
Genes	4,465	123
CDSs	4,359	123
rRNA (23S, 16S, 5S)	22 (7, 7, 8)	-
tRNA	84	-
GenBank Accession No.	CP026569	CP026570

and Eddy, 1997), respectively. rRNAs and other non-coding RNAs were detected using Rfam 12.0 (Nawrocki *et al.*, 2014). The functions of the predicted proteins were annotated by homology searches against the Swiss-prot (Bairoch and Apweiler, 2000), EggNOG 4.5 (Huerta-Cepas *et al.*, 2015), SEED (Overbeek *et al.*, 2014) and KEGG databases (Kanehisa and Goto, 2000).

The complete genome sequence of *Salmonella* Enteritidis MFDS1004839 consisted of a 4,679,649 bp chromosome and a circular 96,994 bp plasmid (pSE1004839), with G + C contents of 52.2% and 49.3%, respectively. As a result of gene prediction, this strain was found to possess 4,359 CDSs, 84 tRNAs and 22 rRNAs in the chromosome and 123 CDSs in pSE1004839. The serovar of *Salmonella* Enteritidis MFDS1004839 was predicted to be 9:g,m:-(O antigen: 9, H antigen phase1: g,m, H antigen phase2: not detected) Enteritidis using SeqSero software (Zhang *et al.*, 2015). This strain includes 25 genes associated with *Salmonella* infection and the 20 genes related to beta-lactam resistance.

Nucleotide sequence accession numbers

The complete genome sequence of *S. Enteritidis* MFDS1004839 has been deposited at the NCBI GenBank database under the accession numbers CP026569 (chromosome) and CP026570 (plasmid, pSE1004839).

적 요

본 연구에서는 2014년 국내에서 식중독 원인식품인 김밥 으로부터 분리된 *Salmonella* Enteritidis의 유전체 분석을 수

행하였다. *Salmonella* Enteritidis MFDS1004839는 한 개의 chromosome (4,679,649 bp)과 plasmid (96,994 bp)로 구성되어 있고, 각각의 G + C contents는 52.2%와 49.3%로 확인되었다. chromosome와 plasmid DNA에 예측된 유전자의 총 수는 4,482개의 단백질 코딩유전자와 84개 tRNA, 그리고 22개의 rRNA였다.

Acknowledgements

This research was financially supported by the Ministry of Food and Drug Safety, Republic of Korea (17161MFDS033).

References

- Bairoch, A. and Apweiler, R.** 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28**, 45-48.
- Das, A., Hari, S.S., Shalini, U., Ganeshkumar, A., and Karthikeyan, M.** 2012. Molecular screening of virulence genes from *Salmonella enterica* isolated from commercial food stuffs. *Biosci. Biotechnol. Res. Asia* **9**, 363-369.
- Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M.C., Rattei, T., Mende, D.R., Sunagawa, S., Kuhn, M., *et al.*** 2015. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* **44**, D286-D293.
- Hyatt, D., Chen, G.L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J.** 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119.
- Kanehisa, M. and Goto, S.** 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27-30.
- Kim, S.** 2010. *Salmonella* serovars from foodborne and waterborne diseases in Korea, 1998-2007: total isolates decreasing versus rare serovars emerging. *J. Korean Med. Sci.* **25**, 1693-1699.
- Lee, W.C., Lee, M.J., Kim, J.S., and Park, S.Y.** 2001. Foodborne illness outbreaks in Korea and Japan studied retrospectively. *J. Food Prot.* **64**, 899-902.
- Lowe, T.M. and Eddy, S.R.** 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955-964.
- Nawrocki, E.P., Burge, S.W., Bateman, A., Daub, J., Eberhardt, R.Y., Eddy, S.R., Floden, E.W., Gardner, P.P., Jones, T.A., Tate, J., *et al.*** 2014. Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res.* **43**, D130-D137.
- Ochman, H., Lawrence, J.G., and Groisman, E.A.** 2000. Lateral gene

- transfer and the nature of bacterial innovation. *Nature* **405**, 299.
- Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Parrello, B., Shukla, M., et al.** 2014. The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Res.* **42**, D206–D214.
- Pallen, M.J. and Wren, B.W.** 2007. Bacterial pathogenomics. *Nature* **449**, 835–842.
- Zhang, S., Yin, Y., Jones, M.B., Zhang, Z., Kaiser, B.L.D., Dinsmore, B.A., Fitzgerald, C., Fields, P.I., and Deng, X.** 2015. *Salmonella* serotype determination utilizing high-throughput genome sequencing data. *J. Clin. Microbiol.* **53**, 1685–1692.