

# Complete genome sequence of *Salmonella* Thompson strain MFDS1004024 isolated from crab-stick

Sewook Park, Woojung Lee, 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

## 게맛살에서 분리된 *Salmonella* Thompson MFDS1004024의 유전체 염기서열 분석

박세욱 · 이우정 · 유란희 · 주인선 · 곽효선 · 김순한\*

식품의약품안전처 식품의약품안전평가원 미생물과

(Received February 28, 2018; Revised April 23, 2018; Accepted May 3, 2018)

*Salmonella enterica* subsp. *enterica* serovar Thompson strain MFDS1004024 was isolated from crab-stick in Korean food-borne outbreak in 2014. Here, we present the complete genome sequence of strain MFDS1004024 with a size of 4,742,942 bp and a mean G + C content of 52%. The genome included 4,373 coding sequences, and 22 ribosomal RNA and 84 transfer RNA genes. Also, we found that strain MFDS1004024 has some genes for *Salmonella* infection and beta-lactam resistance in its genome based on the result of genome analysis.

**Keywords:** *Salmonella enterica*, food pathogen, genome

*Salmonella enterica* subsp. *enterica* is a major cause of food-borne illnesses associated with a wide variety of foods, including meat, eggs, fruits, vegetables, nuts, and spices (Edwards and Galton, 1966). Here, we present the complete genome sequence of *Salmonella enterica* subsp. *enterica* serovar Thompson strain MFDS1004024, isolated from a commercial crab-stick linked to food-borne illness in the Republic of Korea in 2014. The isolate was purified on tryptic soy agar and subsequently treated with lysozyme. High-quality genomic DNA for use in complete genome sequencing was extracted using phenol-

chloroform isoamyl alcohol DNA extraction protocols (Wilson, 1987), and then sheared into ~10–20 kb fragments for PacBio library preparation (Pacific Biosciences). The library was prepared and sequenced using P6/C4 chemistry on a single-molecule real-time (SMRT) cell.

The 96,479 reads were assembled using the HGAP/Quiver software package in SMRT Analysis version 2.3.0., resulting in a single contig. The coding sequences (CDSs) and tRNAs were predicted using Prodigal and tRNAscan-SE, respectively (Lowe and Eddy, 1997; Schattner *et al.*, 2005; Hyatt *et al.*, 2010). The clustered regularly interspaced short palindromic repeat (CRISPR) sequences were detected using PILER-CR and CRISPR Recognition Tool (Bland *et al.*, 2007; Edgar, 2007). rRNAs and other non-coding RNAs were searched by covariance model search with inference of Rfam 12.0 (Nawrocki *et al.*, 2014). Gene calling and annotation of each CDSs were performed with homology searches against Swiss-prot, EggNOG 4.5, SEED, and KEGG databases (Bairoch and Apweiler, 2000; Kanehisa and Goto, 2000; Aziz *et al.*, 2012; Overbeek *et al.*, 2014; Huerta-Cepas *et al.*, 2015).

As a result of the final assembly, a single contig of 4,742,945 bp was generated with a mean coverage of 239× and a G + C content of 52%. The genome did not contain any plasmids. A

\*For correspondence. E-mail: [lambndog@korea.kr](mailto:lambndog@korea.kr);  
Tel.: +82-43-719-4302; Fax: +82-43-719-4300

**Table 1. General features and properties of complete genome sequence of *Salmonella enterica* strain MFDS1004024**

Property	Value
Genome assembly	
Assembly method	SMRT Analysis version 2.3.0
Genome coverage	239.17×
Sequencing technology	PacBio P6/C4 chemistry
Genome features	
Size (bp)	4,742,945
G + C content (%)	52.24
Coding sequences	4,373
rRNA (23S, 16S, 5S)	22 (7, 7, 8)
tRNA	84
CRISPR	2

total of 4,479 genes were identified and annotated including the features of 4,373 CDSs, 22 rRNAs, and 84 tRNAs, and then 2 CRISPR regions were also annotated (Table 1).

Whole-genome sequence based multilocus sequence typing (wgMLST) showed that strain MFDS1004024 belongs to sequence type 26 (ST26) (Larsen *et al.*, 2017). *In silico* serotyping predicted an antigenic profile of 7:k:1,5 (O antigen: 7, H antigen phase1: k, H antigen phase2: 1,5) Thompson using the Kauffmann-White scheme (Zhang *et al.*, 2015). Functional categories based on COG and KEGG metabolic pathway analysis revealed that the genome of MFDS1004024 contains the 25 genes associated with *Salmonella* infection including various invasion proteins. The genome also contains the 19 genes related to beta-lactam resistance such as beta-lactamase and penicillin-binding proteins.

#### Nucleotide sequence accession numbers

The annotated genome sequence of the chromosome of *S. enterica* strain MFDS1004024 has been deposited at NCBI GenBank under accession number CP025745.

## 적 요

*Salmonella enterica* subsp. *enterica* serovar Thompson strain MFDS1004024 는 2014 년 한국에서 발생한 식중독 사고의 계 맛살에서 분리되었다. 본 연구에서는 4,742,942 bp 의 크기와 약 52%의 G + C 함량을 가진 MFDS1004024 균주의 완전한

유전체 염기서열을 분석하였다. 이 유전체에는 4,373 개의 코딩 서열, 22 개의 리보솜 RNA 유전자 및 84 개의 전사 RNA 유전자가 존재한다. 또한 유전체 분석 결과를 통해 살모넬라 감염과 베타락탐계 항생제 내성에 관련이 있는 유전자를 발견하였다.

## Acknowledgements

The mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the Ministry of Food and Drug Safety, Republic of Korea.

This work was funded by Ministry of Food and Drug Safety, Republic of Korea (17161MFDS033).

## References

- Aziz, R.K., Devoid, S., Disz, T., Edwards, R.A., Henry, C.S., Olsen, G.J., Olson, R., Overbeek, R., Parrello, B., Pusch, G.D., *et al.* 2012. SEED servers: high-performance access to the SEED genomes, annotations, and metabolic models. *PLoS One* **7**, e48053.
- Bairoch, A. and Apweiler, R. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28**, 45–48.
- Bland, C., Ramsey, T.L., Sabree, F., Lowe, M., Brown, K., Kyrpides, N.C., and Hugenholtz, P. 2007. CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* **8**, 209.
- Edgar, R.C. 2007. PILER-CR: fast and accurate identification of CRISPR repeats. *BMC Bioinformatics* **8**, 18.
- Edwards, P. and Galton, M. 1966. Salmonellosis. *Adv. Vet. Sci.* **11**, 1–63.
- 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.
- Larsen, M.V., Joensen, K.G., Zankari, E., Ahrenfeldt, J., Lukjancenko, O., Kaas, R.S., Roer, L., Leekitcharoenphon, P., Saputra, D.,

- Cosentino, S., et al.** 2017. The CGE tool box, pp. 65–90. *In* Applied genomics of foodborne pathogens. Springer.
- 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.
- 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.
- Schattner, P., Brooks, A.N., and Lowe, T.M.** 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* **33**, W686–W689.
- Wilson, K.** 1987. Preparation of genomic DNA from bacteria. *In* Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (eds.), Current Protocols in Molecular Biology, Wiley & Sons, New York, 2.4.1–2.4.5.
- 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.