

Draft genome sequence of *Streptococcus* sp. strain NM isolated from head and neck cancer patients

Young Suk Kim¹, Kyoung-Tag Do², and Soo-Je Park^{3*} 

¹Department of Radiation Oncology, Jeju National University Hospital, School of Medicine, Jeju National University, Jeju 63243, Republic of Korea

²Department of Animal Biotechnology, Faculty of Biotechnology, Jeju National University, Jeju 63243, Republic of Korea

³Department of Biology, Jeju National University, Jeju 63243, Republic of Korea

두경부암 환자로 부터 분리된 *Streptococcus* sp. strain NM의 유전체 분석

김영석¹ · 도경탁² · 박수제^{3*} 

¹제주대학교의학전문대학원 제주대학교병원, ²제주대학교생명공학부, ³제주대학교생물학과

(Received November 23, 2018; Revised December 17, 2018; Accepted December 17, 2018)

Streptococcus sp. strain NM belonging to *Firmicutes* was isolated from head and neck cancer patients. Here, we report the draft genome sequence of strain NM with a size of approximately 1.90 Mbp and a mean G+C content of 39.3%. The draft genome included 1,845 coding sequences, and 12 ribosomal RNA and 58 transfer RNA genes. In the draft genome, genes involved in the antimicrobial resistance, hemolysis and defense system have been identified.

Keywords: *Streptococcus*, genome, oral

Many members of the genus *Streptococcus* classified into a member of the *Firmicutes* have been known as a major colonizer in humans or animals. To date, there are validated 129 species isolated from various environments including human specimens (<http://www.bacterio.net/streptococcus.html>). Normally, the streptococcal members are found as an opportunistic pathogen of the oral cavity and upper respiratory tract (Mitchell, 2003; Doern and Burnham, 2010). They are Gram-staining positive, aerobic or facultative anaerobic, coccus-shaped. Here, we describe the draft genome sequence and annotation of *Streptococcus* sp.

strain NM isolated from oral microflora of head and neck cancer patient.

To isolate, the sample was swabbed using sterilized cotton, rinsed into phosphate buffered saline (PBS, pH 7.0), immediately. Then, the PBS was serially diluted to five-folds with fresh PBS. Then 100 µl of the aliquot from the diluted sample was spread on Blood agar plates (BAPs, Hanil KOMED) and incubated at 37°C for one week under a microaerobic condition generated by BD GasPak EZ CampyPouch system (BD). Under naked eyes, we selected some strains with hemolysis activity, and repetitively transferred to new BAPs in order to get purified colony, which designed as strain NM. The purified strain NM has been deposited at Korean Culture Center for Microorganisms (KCCM) as KCCM 43307. To extract genomic DNA (gDNA) of strain NM, we used a commercial DNA extraction kit (GeneAll Biotechnology, Co. Ltd.) according to the manufacturer's instructions. Before whole-genome sequencing, the phylogenetic relationship for the strain NM was determined using the sequence of the 16S ribosomal RNA gene (Koh *et al.*, 2015) and used EzBioCloud server (<https://www.ezbiocloud.net/>). Then, we confirmed that strain NM was most closely related to *Streptococcus pseudopneumoniae* ATCC BAA-960^T (99.7% 16S

*For correspondence. E-mail: sjpark@jejunu.ac.kr;
Tel.: +82-64-754-3524; Fax: +82-64-756-3541

rRNA gene sequence similarity) isolated from lower respiratory tract (Arbique *et al.*, 2004). Whole-genome sequencing was performed on the PacBio RS II sequencing platform (Pacific Biosciences of California) at MacroGen. A reads passed filtering a total of about 1.54 Gb in sequenced reads used into contig assembly. The trimming of the resulting nucleotide sequences and assembling *de novo* were accomplished by Falcon (v.0.2.1) (Chin *et al.*, 2016). Finally, 10 contigs (a mean of 260 depth) were obtained in this study. To estimate genome completeness and quality, we used checkM (Parks *et al.*, 2015). The resulting assembled sequences were annotated by NCBI Prokaryotic Genome Annotation Pipeline with GeneMarkS + version 4.5, using the best-placed reference protein method (Angiuoli *et al.*, 2008). Finally, the draft genome size of the strain NM is *ca.* 1.90 Mb with 39.9% G+C content. The results of CheckM estimation indicated that genome completeness at 93.8% with 0.76% contamination and 75% strain heterogeneity.

The genome includes 1,918 predicted genes, and 12 ribosomal RNA and 58 transfer RNA genes (Table 1). Among coding sequences ($n = 1,776$), only 938 CDSs were matched in KEGG database (52.8% of total CDSs), in which most of them were affiliated into protein families: genetic information processing category. Despite the similarity for 16S rRNA gene sequence was high, the result (92.0%) of average nucleotide identify calculation indicated that strain NM might be new species of the genus *Streptococcus*, against *Streptococcus pseudopneumoniae*

ATCC BAA-960^T (Figueras *et al.*, 2014).

Genome analysis revealed that there were hyaluronidase/collagenases and hemolysin genes, both of which have been reported to play a damaging toxin for membrane or extracellular matrix in staphylococci. Also, we identified some genes involved in two-component regulatory systems (TCRS, e.g. *ciaRH* and *vicRK*) in *Streptococcus pneumoniae* (AlonsoDeVelasco *et al.*, 1995). These TCRS have been known as an essential for growth (Wagner *et al.*, 2002). The genome also contained the antimicrobial resistance genes such as aminoglycoside resistance gene (*aadK*), accessory gene for vancomycin resistance (*vanWY*), and cationic antimicrobial peptide resistance operon (*dltABCD*).

Accession number

This Whole Genome Shotgun project of the strain NM (= KCCM 43307) has been deposited at DDBJ/ENA/GenBank under the accession QUOS00000000. The version described in this paper is version QUOS01000000.

적 요

*Firmicutes*에 속하는 *Streptococcus* sp. strain NM을 두경부 암 환자로부터 분리하였다. 본 연구에서는 약 1.90 Mb의 크기와 39.3%의 평균 G+C 함량을 가진 NM 균주의 비완전한 유전체를 보고한다. 유전체는 1,845개의 코딩서열, 12개의 리보솜 RNA 및 58개의 전사 RNA 유전자를 포함하였다. 본 유전체로부터, 항생제내성, 용혈 및 방어시스템과 관련된 유전자들이 확인되었다.

Table 1. *Streptococcus* sp. strain NM genome assembly and its general features

Item	Description
Genome Assembly Data	
Assembly Method	FALCON v. 0.2.1
Genome Coverage	260X
Sequencing Technology	PacBio RSII
No. of contigs	10
Genome features	
Size (Mbp)	1.90
GC content (%)	39.9
No. of total predicted genes	1,918
No. of total coding sequences	1,845
No. of hypothetical proteins	244
rRNA (23S, 16S, 5S)	12 (4, 4, 4)
tRNA	58

Acknowledgements

This work was supported by a research grant from Jeju National University Hospital in 2018.

References

- AlonsoDeVelasco E, Verheul AF, Verhoef J, and Snippe H. 1995. *Streptococcus pneumoniae*: virulence factors, pathogenesis, and vaccines. *Microbiol. Rev.* **59**, 591–603.
- Angiuoli SV, Gussman A, Klimke W, Cochrane G, Field D, Garrity G,

- Kodira CD, Kypides N, Madupu R, Markowitz V, et al.** 2008. Toward an online repository of Standard Operating Procedures (SOPs) for (meta)genomic annotation. *OMICS* **12**, 137-141.
- Arbique JC, Poyart C, Trieu-Cuot P, Quesne G, Carvalho Mda G, Steigerwalt AG, Morey RE, Jackson D, Davidson RJ, and Facklam RR.** 2004. Accuracy of phenotypic and genotypic testing for identification of *Streptococcus pneumoniae* and description of *Streptococcus pseudopneumoniae* sp. nov. *J. Clin. Microbiol.* **42**, 4686-4696.
- Chin CS, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, Dunn C, O'Malley R, Figueroa-Balderas R, Morales-Cruz A, et al.** 2016. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* **13**, 1050-1054.
- Doern CD and Burnham CA.** 2010. It's not easy being green: the viridans group streptococci, with a focus on pediatric clinical manifestations. *J. Clin. Microbiol.* **48**, 3829-3835.
- Figueras MJ, Beaz-Hidalgo R, Hossain MJ, and Liles MR.** 2014. Taxonomic affiliation of new genomes should be verified using average nucleotide identity and multilocus phylogenetic analysis. *Genome Announc.* **2**, e00927-14.
- Koh HW, Hong H, Min UG, Kang MS, Kim SG, Na JG, Rhee SK, and Park SJ.** 2015. *Rhodanobacter aciditrophus* sp. nov., an acidophilic bacterium isolated from mine wastewater. *Int. J. Syst. Evol. Microbiol.* **65**, 4574-4579.
- Mitchell TJ.** 2003. The pathogenesis of streptococcal infections: from tooth decay to meningitis. *Nat. Rev. Microbiol.* **1**, 219-230.
- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, and Tyson GW.** 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043-1055.
- Wagner C, de Saizieu A, Schonfeld HJ, Kamber M, Lange R, Thompson CJ, and Page MG.** 2002. Genetic analysis and functional characterization of the *Streptococcus pneumoniae* vic operon. *Infect. Immun.* **70**, 6121-6128.