


Complete genome sequence of an indigo producing strain *Yangia* sp. TSBP01, isolated from oil-contaminated sediment


Hae-Seon Kim¹, Sun Ho Cha², Ho Young Suk³, Nyun-Ho Park¹, and Jung-Hee Woo^{1*} 

¹Gyeongbuk Institute for Marine Bio-Industry (GIMB), Uljin 36315, Republic of Korea

²GenoTech Corporation, Daejeon 34113, Republic of Korea

³Department of Life Sciences, Yeungnam University, Gyeongsan 38541, Republic of Korea

인디고 생산능이 있는 *Yangia* sp. TSBP01의 유전체 분석

김해선¹ · 차선호² · 석호영³ · 박년호¹ · 우정희^{1*} 

¹(재)경북해양바이오산업연구원, ²(주)제노텍, ³영남대학교 생명과학과

(Received May 30, 2018; Revised July 13, 2018; Accepted July 16, 2018)

Yangia sp. TSBP01, isolated from tidal flat sediment contaminated by the oil spill, is known to convert indole to indigo via an intermediate called indoxyl. Our analysis revealed that *Yangia* sp. TSBP01 contained the genome of 5,165,974 bp (G + C content: 66.5%) being composed of two chromosomes and five plasmids. This strain had genes encoding several oxygenases such as indole oxygenase directly involved in the conversion of indole to indoxyl.

Keywords: *Yangia*, complete genome, indigo, indole oxygenase

Indigo, one of the most widely used blue dyes, has traditionally been extracted from plants such as *Indigofera*. Although indigo has dominantly been produced by chemical synthesis to minimize the production costs in recent years, production methods using microorganisms have recently begun to be proposed with the possibility of minimizing environmental pollution (O'Connor *et al.*, 1997; Qu *et al.*, 2010). *Yangia* sp. TSBP01 was isolated from a sample of tidal flat sediment contaminated by the oil spill collected on 4 May in 2011 at Taean, a small-sized city located on the west coast in Republic

of Korea. This strain is known to have activities of degrading polyaromatic hydrocarbons and converting indole to indigo. We hereby provided its genomic information to confirm the presence of the genes encoding the target proteins in indigo production.

The genomic DNA extracted from the TSBP01 strain was used to construct the library using SMRTbell™ template prep kit for sequencing on a Pacific Biosciences RS II platform (Pearson *et al.*, 2006). A total of 67,096 long reads of 528,776,319 bp were *de novo* assembled using Hierarchical Genome Assembly Process (HGAP, Version 2.3) workflow (Chin *et al.*, 2013) following subreads filtering, and the draft genome was polished using Quiver (Chin *et al.*, 2013). The estimated total genome size and average coverage were 5,313,796 bp and 82X, respectively. We performed error correction in longest reads to achieve the target of 30X seed bases (150,003,936 bp) with the remaining shorter reads and assembly was done for those reads that survived the correction. Because typical bacterial genomes and plasmids are circular, each of contigs was checked using MUMmer 3.5 (Kurtz *et al.*, 2004) and one of the self-similar ends was trimmed for manual genome closure. Genome annotation was performed using NCBI's Prokaryotic Genome Annotation Pipeline (PGAP).

*For correspondence. E-mail: wjh1991@naver.com;
Tel.: +82-54-780-3454; Fax: +82-54-780-3469

Table 1. Quantitative genomic characterization of *Yangia* sp. TSBP01

Features	Genome
Genome size (bp)	5,165,974
G + C content (%)	66.5
rRNA genes	15
tRNA genes	55
Protein coding genes	4,944
Protein coding genes with function prediction	4,170

Table 2. Chromosomal and plasmid composition of *Yangia* sp. TSBP01

Features	Size (bp)	GC content (%)	Accession No.
Chromosome 1	2,693,385	66.68	CP022189
Chromosome 2	1,350,393	66.09	CP022190
Plasmid 1	497,024	66.46	CP022191
Plasmid 2	205,291	68.11	CP022192
Plasmid 3	182,129	66.29	CP022193
Plasmid 4	120,216	65.12	CP022194
Plasmid 5	117,530	66.93	CP022195

Yangia sp. TSBP01 genome consists of two chromosomes (2,693,385 bp and 1,350,393 bp) and five plasmids (497,024 bp, 205,291 bp, 182,129 bp, 120,216 bp, and 117,530 bp). The genome was predicted to comprise 15 rRNAs, 55 tRNAs and 4,944 protein coding genes.

Indigo is known to be produced from indole via an intermediate called indoxyl by the action of oxygenases (Ensley *et al.*, 1983; Kim *et al.*, 2003). The TSBP01 strain was found to contain an indole oxygenase probably responsible for the production of indigo from indole (Kang and Lee, 2009).

Nucleotide sequence accession numbers

The complete genome sequence of *Yangia* sp. TSBP01 was deposited in GenBank under accession number of CP022189 to CP022195. This strain is available from the Korean Culture Center of Microorganisms under the accession number of KCCM12029P.

적 요

유류 누출 사고로 오염된 해양 갯벌에서 분리된 *Yangia* sp. TSB01은 인돌을 인디고로 생물전환 할 수 있는 특성이 있다. 미

생물에서 인디고는 산화효소에 의해 인돌로부터 인독실을 거쳐 생성된다. *Yangia* sp. TSBP01의 유전체는 2개의 chromosome 과 5개의 plasmid로 구성되어 있으며, 총 5,165,974 bp의 크기에 G + C 함량은 66.5%이다. TSBP01는 인디고를 생성하는데 관여하는 indole oxygenase를 보유하고 있다.

Acknowledgements

This work (Grants No. C0563841) was supported by Business for Cooperative R&D between Industry, Academy, and Research Institute funded Korea Small and Medium Business Administration in 2017 and Gyeongsangbuk-do R&D Program.

References

- Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, *et al.* 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* **10**, 563–569.
- Ensley BD, Ratzkin BJ, Osslund TD, Simon MJ, Wackett LP, and Gibson DT. 1983. Expression of naphthalene oxidation genes in *Escherichia coli* results in the biosynthesis of indigo. *Science* **222**, 167–169.
- Kang MS and Lee JH. 2009. Cloning of indole oxygenase from *Rhodococcus*. *Korean J. Microbiol. Biotechnol.* **37**, 197–203.
- O'Connor KE, Dobson AD, and Hartmans S. 1997. Indigo formation by microorganisms expressing styrene monooxygenase. *Appl. Environ. Microbiol.* **63**, 4287–4291.
- Kim JY, Lee K, Kim Y, Kim CK, and Lee K. 2003. Production of dyestuffs from indole derivatives by naphthalene dioxygenase and toluene dioxygenase. *Lett. Appl. Microbiol.* **36**, 343–348.
- Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, and Salzberg SL. 2004. Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12.
- Pearson GA, Serrao EA, Procaccini G, Duarte C, and Marba N. 2006. Genomic DNA isolation from green and brown algae (caulerpales and fucales) for microsatellite library construction. *J. Phycol.* **42**, 741–745.
- Qu Y, Pi W, Ma F, Zhou J, and Zhang X. 2010. Influence and optimization of growth substrates on indigo formation by novel isolate *Acinetobacter* sp. PP-2. *Bioresour. Technol.* **101**, 4527–4532.