Marine Genomics xxx (xxxx) xxxx

Contents lists available at ScienceDirect

Marine Genomics

journal homepage: www.elsevier.com/locate/margen

Complete genome sequence of *Echinicola rosea* JL3085, a xylan and pectin decomposer

Peiwen Zhan[a], Jianing Ye[a], Xiaopei Lin[a], Fan Zhang[b], Dan Lin[a], Yao Zhang[a], Kai Tang[a,*]

[a] State Key Laboratory of Marine Environmental Science, Fujian Key Laboratory of Marine Carbon Sequestration, College of Ocean and Earth Sciences, Xiamen University, Xiamen, China

[b] Department of Molecular Virology & Microbiology, Center for Metagenomics and Microbiome Research, Baylor College of Medicine, Houston, TX, USA

ARTICLE INFO

Keywords:
Bacteroidetes
Polysaccharides
Enzyme
Degradation

ABSTRACT

Marine *Bacteroidetes* are well known for their functional specialization on the decomposition of polysaccharides which results from a great number of carbohydrate-active enzymes. Here we represent the complete genome of a *Bacteroitedes* member *Echinicola rosea* JL3085^T that was isolated from surface seawater of the South China Sea. The genome is 6.06 Mbp in size with a GC content of 44.1% and comprises 4613 protein coding genes. A remarkable genomic feature is that the number of glycoside hydrolase genes in the genome of *E. rosea* JL3085^T is high in comparison with most of the sequenced members of marine *Bacteroitedes*. *E. rosea* JL3085^T genome harbored multi-gene polysaccharide utilization loci (PUL) systems involved in the degradation of pectin, xylan and arabinogalactan. The large diversity of hydrolytic enzymes supports the use of *E. rosea* JL3085^T as a candidate for biotechnological applications in enzymatic conversion of plant polysaccharides.

1. Introduction

Plant biomass represents the largest renewable carbon source on Earth. Microbes are generally thought to play critical roles in the process of the biomass degradation to ensure a global carbon cycle. To date, microbial enzymatic degradation of plant polysaccharides has many industrial and agricultural applications in the production of bio-based products such as food, feed, chemicals and biofuels. For instance, pectinases are one of the upcoming enzymes of fruit and textile industries (Kashyap et al., 2001). Xylan, the second most abundant polysaccharide and the principal type of hemicellulose, can be hydrolyzed for manufacture of food, animal feed, pulp bleaching and subsequent fermentation to biofuels (Beg et al., 2001; Polizeli et al., 2005). Arabinogalactan and its degradation products have been reported to be beneficial to body health through regulating blood cholesterol (Saeed et al., 2011). Marine *Bacteroidetes* are commonly assumed to be specialized in degrading phytoplankton polysaccharides (Fernández-Gomez et al., 2013) due to a great number and diversity of carbohydrate-active enzymes (CAZymes) (Cantarel et al., 2009) in their genomes (Hahnke et al., 2016), including glycoside hydrolases (GHs), glycosyl transferases (GTs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), auxiliary activities (AAs) and carbohydrate-binding modules (CBMs). Marine bacteria may provide the most common CAZymes resources for polysaccharides degradation, however, they are yet poorly explored biotechnological potential in comparison to Fungi or bacteria isolated from other environments.

Here we report the complete genome sequence of a *Bacteroidetes* member *Echinicola rosea* JL3085^T, a Gram-negative, light-pink-pigmented strain isolated from surface seawater around Yongxing Island in the South China Sea (Table 1). The genus *Echinicola* was firstly proposed by Nedashkovskaya et al. (2006) with the description of *E. pacifica* isolated from sea urchin *Strongylocentrotus intermedius*. So far, members of this genus have been isolated from marine sources like seawater (Nedashkovskaya et al., 2007), solar saltern (Kim et al., 2011),

* Corresponding author.
E-mail address: tangkai@xmu.edu.cn (K. Tang).

https://doi.org/10.1016/j.margen.2019.100722
Received 20 August 2019; Received in revised form 5 October 2019; Accepted 17 October 2019

Table 1
General features of *E. rosea* JL3085^T based on MIGS recommendation and general information on its genome (Yilmaz et al., 2011).

Items	Description
General features	
Classification	Domain *Bacteria* Order *Cytophagales* Family *Cyclobacteriaceae*
Particle shape[a]	Rod
Gram stain	Negative
Pigmentation	Light-pink-pigmented
Temperature[a]	4.0—50.0 °C
Salinity[a]	0—15.0 %
pH range[a]	4.0—11.0
MIxS data	
Submitted_to_insdc	CP040106 (GenBank)
Investigation_type	Bacteria
Project_name	*Echinicola rosea* JL3085^T genome sequencing and assembly (PRJNA541188)
Geo_loc_name	China: Yongxing Island
Lat_lon	16° 49′ 4″ N, 112° 20′ 24″ E
Depth	Surface
Collection_date	2013—05
Env_biome	Marine biome (EVO: 00000447)
Env_feature	Coastal water body (ENVO: 02000049)
Env_material	Sea water (ENVO: 00002149)
Source_mat_id	NBRC 111782; CGMCC 1.15407
Biotic_relationship	Free living
Trophic_level	Chemoorganotroph
Rel_to_oxygen	Aerobic
Isol_growth_condt	Marine agar 2216 (MA; Difco) medium
Seq_meth	Illumina Miseq; PacBio
Assembly method	A5-miseq; SPAdes; HGAP4; CANU
Finishing_strategy	Complete
Genome features	
Genome size (Mbp)	6.06
Contig numbers	1
GC content (%)	44.1
Number of genes (coding)	4731
tRNAs	42
rRNAs	12
Summary of CAZymes	
Glycoside hydrolases (GHs)	171
Glycosyl transferases (GTs)	58
Polysaccharide lyases (PLs)	9
Carbohydrate esterases (CEs)	23
Auxiliary activities (AAs)	5
Carbohydrate-binding modules (CBMs)	13

[a] Data taken from Liang et al. (2016).

brackish water (Srinivas et al., 2012), sea urchin (Jung et al., 2017) and coastal sediment (Lee et al., 2017). Furthermore, several strains showed capacity on hydrolyzing polymeric substances. For example, *E. pacifica* KMM 6172^T and *E. strongylocentrotus* MEBiC08714^T could hydrolyze agar and starch (Nedashkovskaya et al., 2006; Jung et al., 2017). Here we describe the complete genome of *E. rosea* JL3085^T, together with the examination of its polysaccharides degradation ability.

2. Data description

The strain was grown at 28 °C on marine agar 2216 (MA; Difco) medium, and genomic DNA was extracted using a TIANamp Bacteria DNA Kit. The whole genome of *E. rosea* JL3085^T was sequenced using the Illumina Miseq platform accompanied with the SMRT platform to build the Illumina PE library and PacBio library, resulting in 1933 Mb data (6,494,140 reads with 400 bp average insert size and 299-fold coverage depth) and 762 Mb data (97,177 reads with 7838 bp average length and the N_{50} length of the reads is 9992 bp) separately. The genome was assembled using A5-miseq (Tritt et al., 2012), SPAdes genome assembler (Bankevich et al., 2012), HGAP4 (Chin et al., 2016) and CANU (Koren et al., 2017). Contigs obtained from second and third-generation sequencing were analyzed collinearly using MUMmer (Delcher et al., 1999). Finally, the results were corrected using pilon (Walker et al., 2014), after which gene prediction was accomplished using GeneMarkS (Besemer et al., 2001). tRNA-encoding genes and rRNA operons were found by tRNAscan-SE (Lowe and Eddy, 1997) and Barrnap (https://github.com/tseemann/barrnap), respectively. CAZymes were searched using hmmscan program (Krogh et al., 1994).

JL3085^T has a circular chromosome of 6.06 Mbp with a GC content of 44.1% and devoid of plasmids (Fig. 1). The genome contains a total of 4613 protein coding genes, 42 tRNA genes and 12 rRNA operons (Table 1). Moreover, its genome harbors 279 CAZymes (Table 1), including 171 GH which was high in comparison with most of the sequenced strains of marine *Bacteroidetes* (Tang et al., 2017). The putative genes encoding GHs belonged to 44 different families with gene counts ranging between 1 and 32, in which GH43 genes was highest (Fig. 2A). GH43 are classified based on their mode of action and substrate preference into xylanases, xylosidases, arabinofuranosidases, arabinosidase and others in JL3085^T genome, indicating that JL3085^T have the capacity of xylan utilization (Charnock et al., 2000; Sainz-Polo et al., 2014). In addition, numerous genes assigned to other GH families involved in the degradation of xylan were detected including three xylanases from GH10 (FDR09_RS20905, _20975, _20990), one xylanase from GH30 (FDR09_RS14855), two xylosidases from GH31(FDP09_RS07600, _RS13310) and two arabinofuranosidases from GH51 (FDP09_RS02130, _RS07525). The CAZyme annotation of *E. rosea* JL3085^T identified five polysaccharide lyases (CAZyme superfamily, PL1, PL10 and PL11, FDR09_RS03610, _RS03750, _RS14670, _RS14685, _RS07500) responsible for the degradation of pectin which possess pectate lyase activities. Moreover, its genome harbors nine glycoside hydrolases belonging to GH28 (FDP09_RS03755, _RS05505, _RS07425, _RS07450, _RS14640, _RS14680, _RS16090, _RS16095, _RS17050), two pectin methylesterases (CE8) (FDP09_RS14630, _RS14650) and three pectin acetylesterases from CE12 (FDP09_RS07395, _RS07520, _RS14645), which could contribute to release galacturonic acid from pectin (Benoit et al., 2012; Tang et al., 2017). Genes encoding polygalacturonic acid-binding proteins (CBM32, FDP09_RS09745) were also found in the genome which have been demonstrated in *Yersinia enterolitica* before (Abbott et al., 2007).

Some xylanolytic enzymes were located on the multi-gene polysaccharide utilization loci (PUL) of *E. rosea* JL3085^T, including genes encoding xylanases (GH10, FDP09_RS20905, _RS20975, _RS20990) and xylosidase (GH43, FDP09_RS20965, _RS20970), and two carbohydrate-binding modules (CBM6, FDP09_RS20930, _RS20935) (Fig. 2B). It was

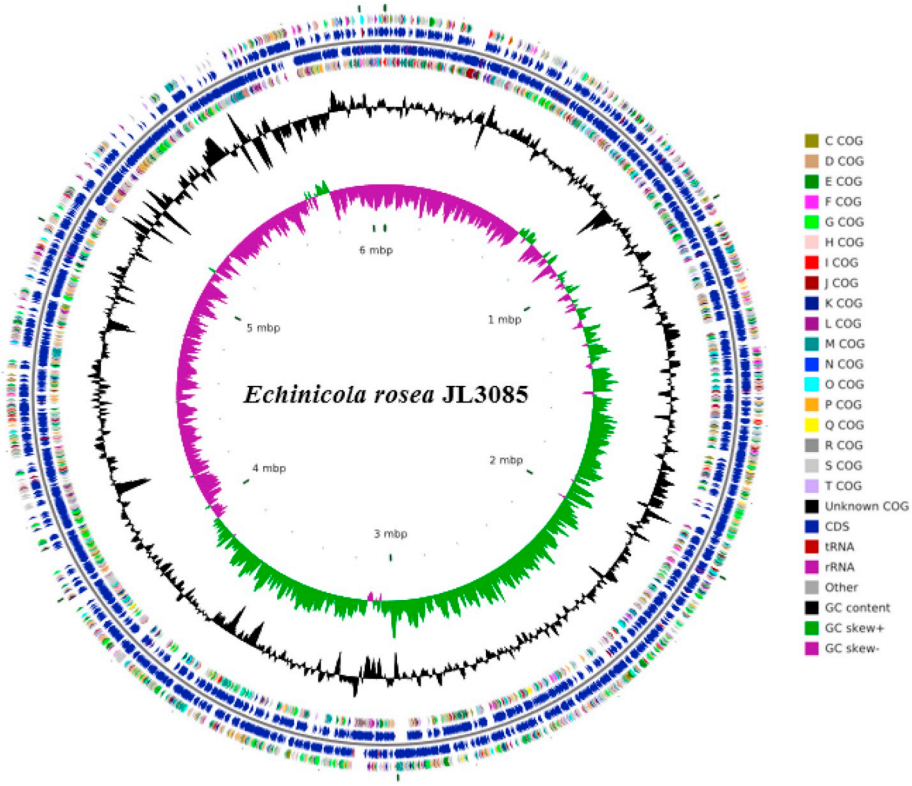

Fig. 1. Complete genome map of *E. rosea* JL3085^T. From the innermost circle to the outermost circle, circle 1 shows the scale line in Mbps. Circle 2 denotes GC skew and circle 3 illustrates the GC content. Circle (4,7) indicates the CDSs, colored according to COG function categories. Circle (5,6) shows CDS/tRNA/rRNA genes.

found that *E. rosea* JL3085^T contained a PUL predicted to code for enzymes known to be involved in the degradation of pectin, including genes encoding two pectate lyases (PL1, FDP09_RS14670, _RS14685), two pectin esterases (CE8, FDP09_RS14630, _RS14650), and two polygalacturonases (GH28, FDP09_RS14640, _RS14680), rhamnogalacturonides (FDP09_RS14635), degradation protein (GH88, FDP09_RS14625), rhamnogalacturonan acetylesterases (FDP09_RS14645) and a set of genes associated with hexuronate metabolism (Fig. 2B). *E. rosea* JL3085^T harbored a arabinan PUL containing gene-encoding enzymes (FDP09_RS02145, _RS02150) that target 1,3- or 1,5- linkages of α-L-arabinofuranosides in arabinans and their derivate such as arabinoxylans and arabinogalactans (Fig. 2B). In addition, the presence of four α-galactosidases (GH57 and GH97), 24 β-galactosidases (GH35 and GH2) in the genome indicates bacteria have the ability of degrading arabinogalactan. Three PULs contain the transporter systems consisting of the oligosaccharide-binding protein susD and an outer-membrane TonB-dependent receptor porin susC.

Pectin, xylan, sodium alginate, arabinogalactan, chitin, fucoidan, laminarin and mannan were used to test the strain's degradation ability according to previous methods (Zhan et al., 2017). Growth experiments showed that strain JL3085^T could utilize pectin, xylan and arabinogalactan (Fig. 2C), which were in consistent with genomic prediction of polysaccharide utilization.

In summary, the great number and diversity of CAZymes possessed by *E. rosea* JL3085^T may provide a genomic content for biotechnology applications in xylan and pectin degradation.

Strain and nucleotide sequence accession numbers

The strain has been deposited in the NITE Biological Resource Center and China General Microbiological Culture Collection Center (=NBRC 111782^T = CGMCC 1.15407^T). The complete genome sequence is available in the NCBI database (accession number CP040106).

Declaration of Competing Interest

The authors declare that they have no conflicts of interest

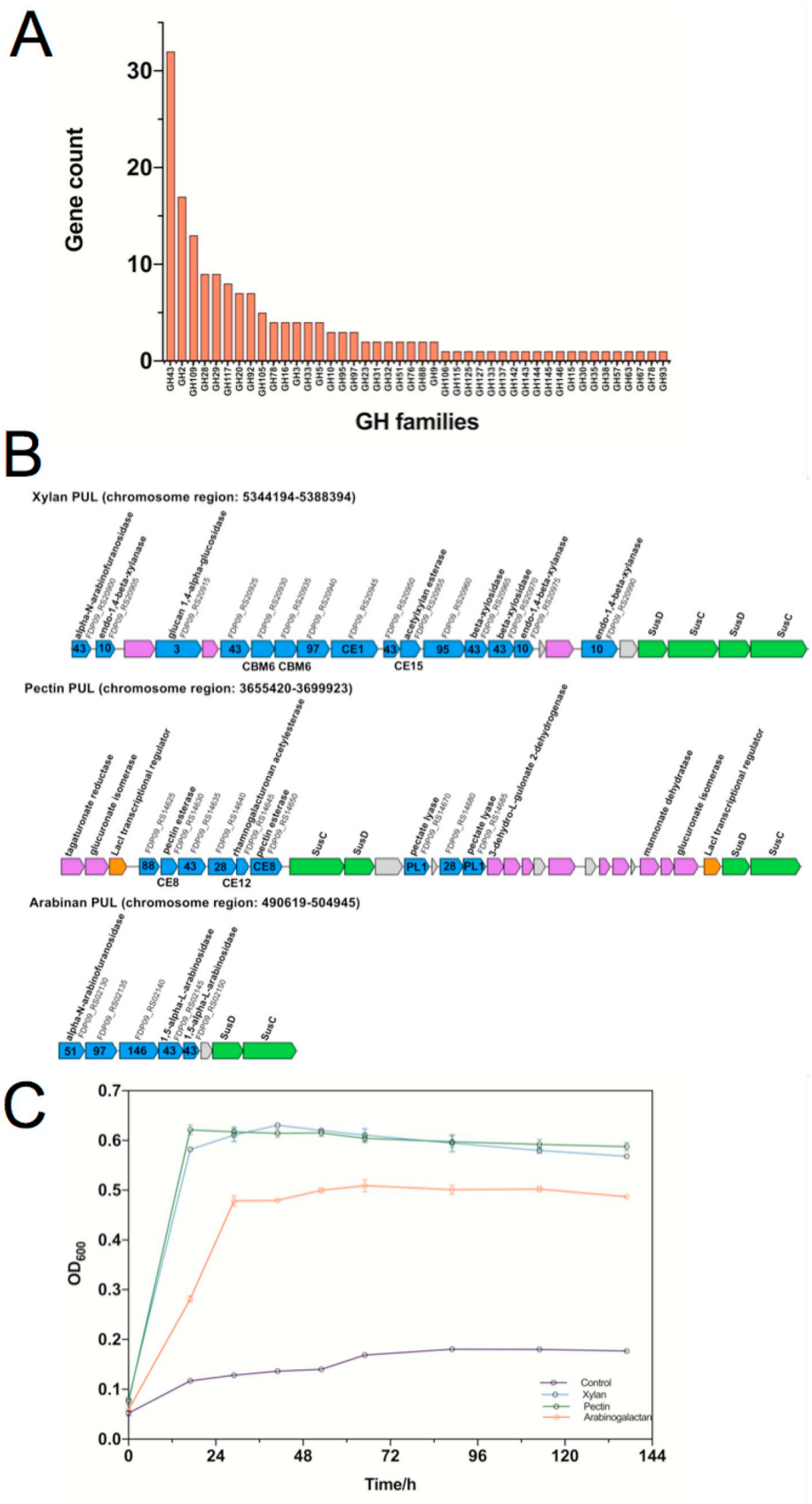

Fig. 2. A) Distribution of GH families in *E. rosea* JL3085^{T} genome; **B)** genetic organization of the predicted PULs of *E. rosea* JL3085^{T}. The position of the gene cluster in the genome is shown in a bracket. The functions of the proteins are colour-coded: blue, CAZymes; green, membrane proteins involved in binding/transport; purple, other enzymes; orange, regulation factor; grey, unknown function; **C)** growth curves of strain JL3085^{T}. Bacteria were cultured in marine minimal media (2.3% (*w/v*) sea salts, 0.05% (w/v) yeast extract, 0.05% (w/v) NH_4Cl and 50 mM Tris-HCl, pH 7.6–7.8), with a final concentration of 0.2% of one of the following carbon sources: pectin, xylan and arabinogalactan. Cultures using only minimal media were treated as controls.

Acknowledgements

This study was supported by the National Key Research and Development Program of China (2016YFA0601100, 2016YFA0601400), the National Natural Science Foundation of China project (41776167, 41676070, U1805242).

References

Abbott, D.W., Hrynuik, S., Boraston, A.B., 2007. Identification and characterization of a novel periplasmic polygalacturonic acid binding protein from *Yersinia enterolitica*. J. Mol. Biol. 367, 1023–1033.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–477.

Beg, Q., Kapoor, M., Mahajan, L., Hoondal, G.S., 2001. Microbial xylanases and their industrial applications: a review. Appl. Microbiol. Biotechnol. 56, 326–338.

Benoit, I., Coutinho, P.M., Schols, H.A., Gerlach, J.P., Henrissat, B., de Vries, R.P., 2012. Degradation of different pectins by fungi: correlations and contrasts between the pectinolytic enzyme sets identified in genomes and the growth on pectins of different origin. BMC Genomics 13, 321.

Besemer, J., Lomsadze, A., Borodovsky, M., 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res. 29, 2607–2618.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., Henrissat, B., 2009. The Carbohydrate-Active enzymes database (CAZy): an expert resource for glycogenomics. Nucleic Acids Res. 37, D233–D238.

Charnock, S.J., Bolam, D.N., Turkenburg, J.P., Gilbert, H.J., Ferreira, L.M., Davies, G.J., Fontes, C.M., 2000. The X6 "thermostabilizing" domains of xylanases are carbohydrate-binding modules: structure and biochemistry of the *Clostridium thermocellum* X6b domain. Biochemistry 39, 5013–5021.

Chin, C.S., Peluso, P., Sedlazeck, F.J., Nattestad, M., Concepcion, G.T., Clum, A., Dunn, C., O'Malley, R., Figueroa-Balderas, R., Morales-Cruz, A., Cramer, G.R., Delledonne, M., Luo, C., Ecker, J.R., Cantu, D., Rank, D.R., Schatz, M.C., 2016. Phased diploid genome assembly with single-molecule real-time sequencing. Nat. Methods 13, 1050.

Delcher, A.L., Kasif, S., Fleischmann, R.D., Peterson, J., White, O., Salzberg, S.L., 1999. Alignment of whole genomes. Nucleic Acids Res. 27, 2369–2376.

Fernández-Gomez, B., Richter, M., Schüler, M., Pinhassi, J., Acinas, S.G., González, J.M., Pedros-Alio, C., 2013. Ecology of marine Bacteroidetes: a comparative genomics approach. ISME J. 7, 1026.

Hahnke, R.L., Meier-Kolthoff, J.P., García-López, M., Mukherjee, S., Huntemann, M., Ivanova, N.N., Woyke, T., Kyrpides, N.C., Klenk, H.-P., Göker, M., 2016. Genome-based taxonomic classification of Bacteroidetes. Front. Microbiol. 7, 2003.

Jung, Y.J., Yang, S.H., Kwon, K.K., Bae, S.S., 2017. *Echinicola strongylocentroti* sp. nov., isolated from a sea urchin *Strongylocentrotus intermedius*. Int. J. Syst. Evol. Microbiol. 67, 670–675.

Kashyap, D.R., Vohra, P.K., Chopra, S., Tewari, R., 2001. Applications of pectinases in the commercial sector: a review. Bioresour. Technol. 77, 215–227.

Kim, H., Joung, Y., Ahn, T.S., Joh, K., 2011. Echinicola jeungdonensis sp. nov., isolated from a solar saltern. Int. J. Syst. Evol. Microbiol. 61, 2065–2068.

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., Phillippy, A.M., 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 27, 722–736.

Krogh, A., Brown, M., Mian, I.S., Sjölander, K., Haussler, D., 1994. Hidden Markov models in computational biology: applications to protein modeling. J. Mol. Biol. 235, 1501–1531.

Lee, D.W., Lee, A.H., Lee, H., Kim, J.J., Khim, J.S., Yim, U.H., Kim, B.S., 2017. Echinicola sediminis sp. nov., a marine bacterium isolated from coastal sediment. Int. J. Syst. Evol. Microbiol. 67, 3351–3357.

Liang, P., Sun, J., Li, H., Liu, M., Xue, Z., Zhang, Y., 2016. Echinicola rosea sp. nov., a marine bacterium isolated from surface seawater. Int. J. Syst. Evol. Microbiol. 66, 3299–3304.

Lowe, T.M., Eddy, S.R., 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25, 955–964.

Nedashkovskaya, O.I., Kim, S.B., Vancanneyt, M., Lysenko, A.M., Shin, D.S., Park, M.S., Lee, K.H., Jung, W.J., Kalinovskaya, N.I., Mikhailov, V.V., Bae, K.S., Swings, J., 2006. *Echinicola pacifica* gen. nov., sp. nov., a novel flexibacterium isolated from the sea urchin *Strongylocentrotus intermedius*. Int. J. Syst. Evol. Microbiol. 56, 953–958.

Nedashkovskaya, O.I., Kim, S.B., Hoste, B., Shin, D.S., Beleneva, I.A., Vancanneyt, M., Mikhailov, V.V., 2007. Echinicola vietnamensis sp. nov., a member of the phylum Bacteroidetes isolated from seawater. Int. J. Syst. Evol. Microbiol. 57, 761–763.

Polizeli, M.L.T.M., Rizzatti, A.C.S., Monti, R., Terenzi, H.F., Jorge, J.A., Amorim, D.S., 2005. Xylanases from fungi: properties and industrial applications. Appl. Microbiol. Biotechnol. 67, 577–591.

Saeed, F., Pasha, I., Anjum, F.M., Sultan, M.T., 2011. Arabinoxylans and arabinogalactans: a comprehensive treatise. Crit. Rev. Food Sci. Nutr. 51, 467–476.

Sainz-Polo, M.A., Valenzuela, S.V., González, B., Pastor, F.J., Sanz-Aparicio, J., 2014. Structural analysis of glucuronoxylan-specific Xyn30D and its attached CBM35 domain gives insights into the role of modularity in specificity. J. Biol. Chem. 289, 31088–31101.

Srinivas, T.N.R., Tryambak, B.K., Kumar, P.A., 2012. Echinicola shivajiensis sp. nov., a novel bacterium of the family "Cyclobacteriaceae" isolated from brackish water pond. Antonie Van Leeuwenhoek 101, 641–647.

Tang, K., Lin, Y., Han, Y., Jiao, N., 2017. Characterization of potential polysaccharide utilization systems in the marine bacteroidetes Gramella Flava JLT2011 using a multi-omics approach. Front. Microbiol. 8, 402–413.

Tritt, A., Eisen, J.A., Facciotti, M.T., Darling, A.E., 2012. An integrated pipeline for de novo assembly of microbial genomes. PLoS One 7, e42304.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., Earl, A.M., 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9, e112963.

Yilmaz, P., Kottmann, R., Field, D., Knight, R., Cole, J.R., Amaral-Zettler, L., et al., 2011. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. Nat. Biotechnol. 29, 415–420.

Zhan, P., Tang, K., Chen, X., Yu, L., 2017. Complete genome sequence of *Maribacter* sp. T28, a polysaccharide-degrading marine flavobacteria. J. Biotechnol. 259, 1–5.