Marine Genomics xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Marine Genomics

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

Genomic analysis of *Microbulbifer* sp. Q7 exhibiting degradation activity toward seaweed polysaccharides

Min Yang, Yuan Yu, Tianyi Jin, Haijin Mou *, Li Li *

College of Food Science and Engineering, Ocean University of China, Qingdao 266003, China

ARTICLE INFO

Article history:
Received 12 April 2017
Received in revised form 12 July 2017
Accepted 13 July 2017
Available online xxxx

Keywords:
Microbulbifer
Alginate lyase
Agarase
Seaweed polysaccharide

ABSTRACT

Microbulbifer sp. Q7 (=CGMCC 14061), which has the ability to degrade seaweed polysaccharides, was isolated from the gut of sea cucumber. The genome of *Microbulbifer* sp. Q7 was sequenced, and the organism was found to contain 3,953,171 bp with a G + C content of 58.36%. A total of 3131 protein-coding sequences were predicted, including seven agarases (EC 3.2.1.81) and five alginate lyases (EC 4.2.2.3). These results reveal that *Microbulbifer* sp. Q7 have potential to degrade seaweed polysaccharides. Finally, the genome sequence of *Microbulbifer* sp. Q7 provides the fundamental genomic information for future studies on novel depolymerase enzymes of seaweed polysaccharides.

1. Introduction

Microbulbifer is a genus which is Gram-negative, rod-shaped and strictly aerobic, first proposed by González in 1997 (González et al., 1997). So far, 20 *Microbulbifer* species have been proposed and published with valid names (http://www.bacterio.net/). Six complete and seven draft genomes of the *Microbulbifer* strains have been deposited into the NCBI genome database and released up till February 2017. The genomic features of these *Microbulbifer* species showed that they were capable of producing seaweed polysaccharide-degrading enzymes, indicating that *Microbulbifer* strains can be potentially applied in industry to produce oligosaccharide (Sun et al., 2014; Wakabayashi et al., 2012).

The current researches of *Microbulbifer* strains mainly isolated from sea water, sediment and seaweed (Vashist et al., 2012; Zhu et al., 2016; Lee and Choi, 2016). *Microbulbifer* sp. Q7 was isolated from sea cucumber gut, a commendable material to obtain seaweed polysaccharide-degrading strains (Purcell et al., 2016). Here, we present the genome of Q7 with the description of annotation, highlighting its potential for seaweed polysaccharides degradation.

2. Data description

Microbulbifer sp. Q7 was isolated from the gut of sea cucumber, the cells were cultivated in Marine Broth 2216 (BD Difco™, Sparks, USA), pH 7.2 at 28 °C for 24 h. Isolation of genomic DNA was carried using Rapid Bacterial Genomic DNA Isolation Kit (Sangon Biotech, China). Total DNA obtained was subjected to quality control by agarose gel electrophoresis and quantified by Qubit. The genome of Q7 was sequenced with MPS (massively parallel sequencing) Illumina technology. Two DNA libraries were constructed: a paired-end library with an insert size of 500 bp and a mate-pair library with an insert size of 6 kb. The 500 bp library was sequenced using an Illumina Miseq by PE300 strategy and the 6 kb library was sequenced using an Illumina Hiseq2000 by PE100 strategy. Library construction and sequencing was performed at the Beijing Novogene Bioinformatics Technology Co., Ltd. Quality control of both paired-end and mate-pair reads were performed using in-house program. After this step, Illumina PCR adapter reads and low quality reads were filtered. The filtered reads were assembled by SOAPdenovo (Li et al., 2010) (http://soap.genomics.org.cn/soapdenovo.html) to generate scaffolds. All reads were used for further gap closure. Transfer RNA (tRNA) genes were predicted with tRNAscan-SE (Lowe and Eddy, 1997). Ribosome RNAs (rRNA) were predicted with rRNAmmer (Lagesen et al., 2007) and sRNAs were predicted by BLAST against Rfam (Gardner et al., 2008).

Gene prediction was performed on the Q7 genome assembly by GeneMarKS (Besemer et al., 2001) (http://topaz.gatech.edu/) with integrated model which combine the GeneMarKS generated

* Corresponding authors at: College of Food Science and Engineering, Ocean University of China, Qingdao, Shandong 266003, China.
E-mail addresses: mousun@ouc.edu.cn (H. Mou), ll@ouc.edu.cn (L. Li).

http://dx.doi.org/10.1016/j.margen.2017.07.003
1874-7787/

(native) and Heuristic model parameters. A whole Blast search (*E*-value < = 1e − 5, minimal alignment length percentage > = 40%) was performed against 5 databases. They are KEGG (Kanehisa et al., 2016) (Kyoto Encyclopedia of Genes and Genomes), COG (Tatusov et al., 2003) (Clusters of Orthologous Groups), NR (Non-Redundant Protein Database databases), Swiss-Prot (Magrane, 2011), and GO.

The genomic features of Q7 were shown in Tables 1 and 2. The genome consisted of two contigs with a G + C content of 58.36% and a genome size of 3,953,171 bp. A total of 3268 coding sequences (CDSs) were identified, making up 87.66% of the genome. The genome also encoded 48 tRNAs, eight rRNAs, and one sRNA, making up 0.368% of the genome. Approximately 8% of the predicted protein-encoding genes had no homologs in the NR protein databases. Among the genes having homologs with other organisms, 1825 genes were mapped to the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al., 2016) database. According to the GOC (Galperin et al., 2015) database, 1789 predicted genes could be classified into 22 functional categories, and approximately 6% of the predicted genes belonged to the carbohydrate transport and metabolism category (Fig. 1).

The main purpose of this work was to identify seaweed polysaccharide depolymerases from Q7. Interestingly, seven agarase-encoding sequences and five alginate lyase-encoding sequences were found in the genome of Q7. All of these agarases and alginate lyases showed high similarity (>70%) with the amino acid of agarases and alginate lyases in NCBI, respectively (Supplementary Table 1 and Table 2). According to the KEGG database, nos. 0068 (Locus_tag = AU182_RS00335), 2553 (Locus_tag = AU182_RS12585), 2562 (Locus_tag = AU182_RS12630), 2563 (Locus_tag = AU182_RS12635), 2568 (Locus_tag = AU182_RS12660) and 2569 (Locus_tag = AU182_RS12665) could be classified as EC 3.2.1.81, a β-agarase that mainly hydrolyzes (1 → 4)-β-D-galactosidic linkages in agar, yielding a tetramer as the predominant product. Additionally, agarase nos. 0068, 2553, and 2562 belonged to glycoside hydrolase family 12 (GH12), and agarase no. 2563 belonged to GH16. According to their substrate specificities, alginate lyases can be classified as polyM lyases (EC 4.2.2.3) and polyG lyases (EC 4.2.2.11), which preferentially break up polyM and polyG blocks, respectively (Kim et al., 2011). All of the five alginate lyases in Q7 were polyM lyases (EC 4.2.2.3) according to the sequence analysis. Three-dimensional (3D) models of alginate lyase and agarase were obtained by homologous modeling using Swiss Model (https://swissmodel.expasy.org/interactive). The five alginate lyases were homology-modeled with alginate lyase, chondroitinase B, putative alginate lyase, and alginate lyase, respectively. The 3D structure prediction of agarase showed that the proteins were homology-modeled with agarase, β-porphyranase, and endo-1,4-β-glucanase.

In summary, because of the existence of a large repertoire of genes encoding key enzymes involved in the degradation of seaweed polysaccharides, we expect that Q7 may have the potential for production of efficient seaweed polysaccharides depolymerases when heterologous expression is realized.

Table 1
General features of Q7 and MIGS information.

Item	Description
General features	
Gram stain	Negative
Cell shape	Rod
Motility	Non-motile
Colour of colonies	light yellow
Temperature range; Optimum (°C)	24–30 °C; 28 °C
NaCl concentration range; Optimum (%)	1–5%; 1%
pH range; Optimum	6.0–9.0; 7.0
MIGS data	
Investigation_type	Bacteria_archaea
Project _name	Genome sequence of *Microbulbifer* sp. Q7
Country	China
Collection_date	2014-3
Env_biome	marine neritic zone (ENVO:00000206)
Env_feature	Intestine environment (ENVO:2100002)
Env_material	Alga (ENVO:02500019)
Biotic_relationship	Free living
Rel_to_oxygen	Aerobic
Num_replicons	1
Source_mat_id	CGMCC 14061
Submitted to INSDC	Accession number NZ_LROY01000000
Sequencing method	Illumina Miseq; Illumina Hiseq2000
Assembly data	
Assemble method	SOAPdenovo
Assembly name	LoxAfr_3.0
Genome coverage	200×
Finishing strategy	Prime design, PCR and massively parallel sequencing (MPS)

Table 2
General features of Q7 genome.

General features	Q7
Genome size (bp)	3,953,171
Contig	2
GC content (%)	58.36
Gene number	3268
tRNA number	48
rRNA number	8
sRNA number	1
Properties of predicted gene models	No. of genes
NR	3006
Swiss-Prot	1394
KEGG	1825
COG	1789
GO	1064
TrEMBL	2661

Nucleotide sequence accession number

The genome sequences of *Microbulbifer* sp. Q7 (=CGMCC 14061) have been deposited at DDBJ/EMBL/GenBank under the accession number of LROY00000000.

Acknowledgement

This work was supported by Public Science and Technology Research Funds Projects of Ocean (201505022), Shandong Science and Technology Development Project (2014GHY115037) and Fundamental Research Funds for the Central Universities (201762034).

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.margen.2017.07.003.

Please cite this article as: Yang, M., et al., Genomic analysis of *Microbulbifer* sp. Q7 exhibiting degradation activity toward seaweed polysaccharides, Mar. Genomics (2017), http://dx.doi.org/10.1016/j.margen.2017.07.003