

RESEARCH

Open Access



# Genome sequencing and identification of cellulase genes in *Bacillus paralicheniformis* strains from the Red Sea

Siham Fatani<sup>1†</sup>, Yoshimoto Saito<sup>1,2†</sup>, Mohammed Alarawi<sup>1</sup>, Takashi Gojobori<sup>1\*</sup> and Katsuhiko Mineta<sup>1\*</sup>

## Abstract

**Background:** Cellulolytic microorganisms are considered a key player in the degradation of plant biomass in various environments. These microorganisms can be isolated from various environments, such as soils, the insect gut, the mammalian rumen and oceans. The Red Sea exhibits a unique environment in terms of presenting a high seawater temperature, high salinity, low nutrient levels and high biodiversity. However, there is little information regarding cellulase genes in the Red Sea environment. This study aimed to examine whether the Red Sea can be a resource for the bioprospecting of microbial cellulases by isolating cellulase-producing microorganisms from the Red Sea environment and characterizing cellulase genes.

**Results:** Three bacterial strains were successfully isolated from the plankton fraction and the surface of seagrass. The isolated strains were identified as *Bacillus paralicheniformis* and showed strong cellulase activity. These results suggested that these three isolates secreted active cellulases. By whole genome sequencing, we found 10 cellulase genes from the three isolates. We compared the expression of these cellulase genes under cellulase-inducing and non-inducing conditions and found that most of the cellulase genes were generally upregulated during cellulolysis in the isolates. Our operon structure analysis also showed that cellulase genes form operons with genes involved in various kinds of cellular reactions, such as protein metabolism, which suggests the existence of crosstalk between cellulolysis and other metabolic pathways in the bacterial isolates. These results suggest that multiple cellulases are playing important roles in cellulolysis.

**Conclusions:** Our study reports the isolation and characterization of cellulase-producing bacteria from the Red Sea. Our whole-genome sequencing classified our three isolates as *Bacillus paralicheniformis*, and we revealed the presence of ten cellulase orthologues in each of three isolates' genomes. Our comparative expression analysis also identified that most of the cellulase genes were upregulated under the inducing conditions in general. Although cellulases have been roughly classified into three enzyme groups of beta-glucosidase, endo- $\beta$ -1,4-glucanase and exoglucanase, these findings suggest the importance to consider microbial cellulolysis as a more complex reaction with various kinds of cellulase enzymes.

**Keywords:** Cellulase, The Red Sea, Bioprospecting, Cellulolysis, Operon, Whole genome sequencing, Gene expression analysis

\* Correspondence: [takashi.gojobori@kaust.edu.sa](mailto:takashi.gojobori@kaust.edu.sa);  
[katsuhiko.mineta@kaust.edu.sa](mailto:katsuhiko.mineta@kaust.edu.sa)

<sup>†</sup>Siham Fatani and Yoshimoto Saito contributed equally to this work.

<sup>1</sup>Computational Bioscience Research Center (CBRC), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

Cellulose, which is the major component of plant biomass, is the most abundant organic compound on Earth and a sustainable source of energy [1]. It is composed of a linear homologous polymer chain consisting of D-glucose residues, containing up to 10,000 glucose residues linked by  $\beta$ -1,4-glycosidic bonds [2, 3]. The efficient conversion of cellulose into its glucose monomers by microbes as a source of high-energy molecules helps to meet future energy needs and serves as an alternative source of renewable energy [4]. The biodegradation of  $\beta$ -1,4-glycosidic bonds in cellulose biomass is carried out by free cellulases or a multienzyme complex referred to as the cellulosome, which can catalyze the hydrolysis of cellulose into sugars. These enzymes are produced by various microorganisms, such as bacteria and fungi [5].

Several cellulase-producing microorganisms with high cellulolytic activity have been isolated from the fungal genera *Aspergillus* and *Trichoderma* [6]. Cellulase activity has also been observed in bacterial genera including *Alteromonas*, *Acetivibrio*, *Bacillus*, *Bacteroides*, *Cellulomonas*, *Clostridium*, and *Ruminococcus* [7]. *Bacillus* species have been employed for production of cellulase [8, 9]. As represented by *B. subtilis*, a lot of studies regarding cellulase gene sequences, enzymatic activities, optimal condition for cellulolysis were published from *Bacillus* species. Cellulases from *Bacillus* are still now reported frequently [10, 11].

Cellulases are a group of three types of enzymes with different activities. The first type is endo- $\beta$ -1,4-glucanase (EC 3.2.1.4), which can perform cleavage on internal bonds in the cellulose fibers. The second is exoglucanase (EC 3.2.1.91), which binds at the reducing or non-reducing ends of cellulose fibers and cleaves them to produce short disaccharides [12]. The third is  $\beta$ -glucosidase (EC 3.2.1.21), which hydrolyzes cellobioses to produce glucose molecules [8]. Recent studies have classified the cellulases of all three types as glycoside-hydrolases (GHs). The glycoside-hydrolase is a group of hydrolases composed of a great number of enzymes. The Carbohydrate-Active Enzymes CAZy database [13] provides that the cellulases reported thus far are classified into of 16 widely ranging GH orthologous groups.

In contrast to the terrestrial environment, a small number of studies have been conducted to investigate cellulase-producing microorganisms in marine environments. However, the isolation and characterization of cellulases from marine bacteria are now more frequently reported [8]. Studies regarding marine cellulases have revealed characteristic of these enzymes with possible application of the enzyme and producing strains.

The Red Sea has unique marine environmental features compared to the other oceans. The Red Sea has been described as an oligotrophic environment

consisting of one of the warmest and saltiest water bodies in the world, with year-round high UV radiation [14]. These characteristics are thought to have given rise to and modulated the evolution and diversity of microbial forms in the Red Sea. To our knowledge, cellulase-producing microorganisms from the Red Sea environment remain unstudied. In particular, the diversity, abundance, and characteristics of cellulase genes or cellulolytic microorganisms are still far from fully explored.

In this study, cellulase-producing bacteria were isolated from the Red Sea environment. We also identified cellulase genes and revealed that they were expressed during cellulolysis and provided important information to understand the mechanism of microbial cellulolysis in detail.

## Results

### Isolation and Screening of Cellulase-Producing Microorganisms

Surface seawater samples were collected from a coastal region of the Red Sea at Thuwal in Saudi Arabia. The samples were diluted and spread on Nutrient Media (NM) plates and incubated at 30°C following a previously reported protocol [15]. Four hundred fifty-six colonies were isolated on NM plates and subsequently streaked on media containing Carboxymethyl cellulose (CMC) as the sole carbon source for the screening of cellulase-producing microorganisms. No isolates showed cellulase activity on the CMC medium plates (Table 1).

We then used the plankton fractions obtained from surface seawater as the source for the isolation of cellulase-producing microorganisms. Sixty strains were isolated on NM plates, and one isolate (PB1) formed a zone of clearance around the colonies on CMC medium plates after staining with Congo Red (Fig. 1). We also tried to isolate cellulase-producing bacteria from the surface of green seagrass. One hundred and twenty-six strains were obtained on NM plates, and two strains (SB2, SB3) showed a clear halo zone on CMC medium plates after staining (Table 1). We then characterized these three strains as cellulase-producing isolates from the Red Sea.

### Cellulase enzyme assay

The measurement of cellulase activities for the three strains was in broth media that containing a strip of cellulose filter paper as the sole carbon source. The complete degradation of the filter paper was observed in all the PB1, SB2 and SB3 cultures whereas the control (i.e., the same culture conditions without the bacterial inoculate) showed very slight degradation of the filter paper (Fig. S1). The measurement of the amount of reducing sugar formed in broth media after four days of cultivation revealed that all three strains showed

**Table 1** Screening of cellulase-producing bacteria from seawater, seagrass and plankton fraction samples

Sample	Seawater	Seagrass	Plankton
No. of picked colonies from NM plate	456	126	60
No. of clones showing cellulase activity on CMC plate	0	2	1

NM nutrient media; CMC carboxymethyl cellulose

maximum cellulase activity after 72 h of incubation (Fig. 2). The maximum activities were 0.75, 0.59 and 0.70 filter paper unit (FPU)/ml in the PB1, SB2 and SB3 isolates, respectively.

#### Growth test under high salinity conditions

We investigated whether our three isolates from the Red Sea have abilities to do cellulolysis under high salinity condition. First, we tested the growth of the isolates in NM broth under the salinity conditions ranging from 0.2% (34mM) to 10% (1.7 M) at 24 h after inoculation. Although the growth of the three strains gradually retarded along with the concentration of NaCl, the results clearly showed that OD<sub>600</sub> of the isolates were 0.5–0.9 at the concentration of 8% (1.4 M) and 0.12–0.36 at 10% (1.7 M) NaCl (Fig. S2a). Next, we tested the growth of the isolates on CMC broth where they are able to use CMC only as a carbon source under various salinity stress conditions at 48 h after inoculation. Compared with NM broth, the isolates showed lower OD<sub>600</sub> values in CMC broth at all the NaCl concentrations. However, the isolates still showed the growth (around 0.1 of

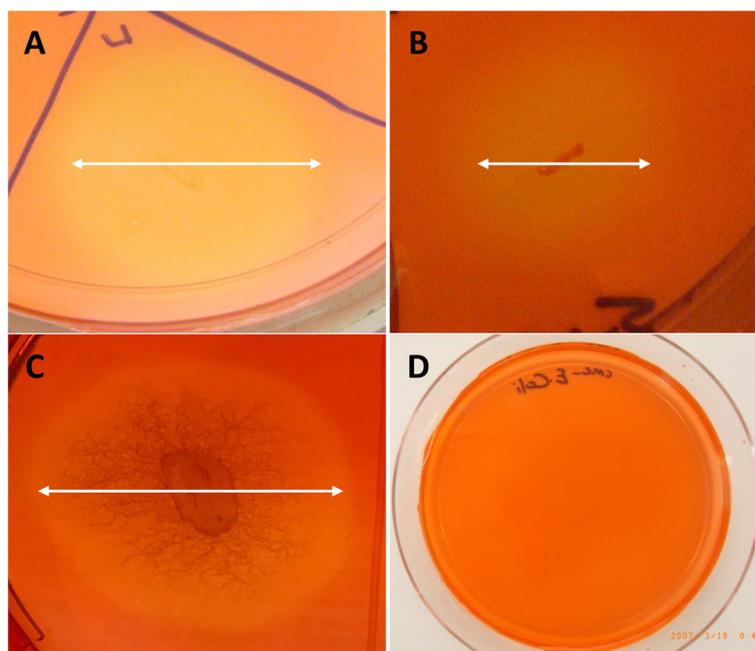
OD<sub>600</sub>) in CMC broth under 8% salinity (Fig. S2b, Fig. S2c).

#### Whole-Genome Sequencing of Bacterial Isolates

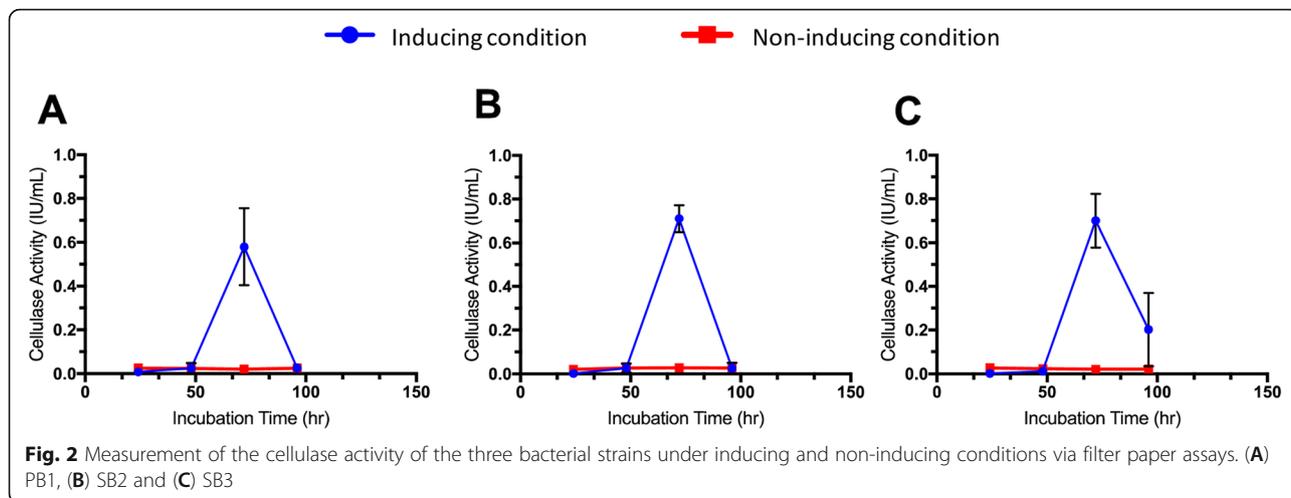
The genome sequences of the three strains were determined using the PacBio RSII platform. The average of sequencing coverage was 186x. The obtained sequences were *de novo* assembled as a single circular chromosome in each strain, and no plasmids were detected in their genomes. The total genome sizes of PB1, SB2 and SB3 were 4,318,221, 4,318,038, and 4,317,481 bp, respectively (Table S1). We then conducted gene prediction, resulting in the prediction of 4362, 4441 and 4675 genes in the genomes of the three isolates, respectively.

#### Multilocus Sequence Typing (MLST) Analysis

To determine the phylogenetic positions of the three strains within genus *Bacillus*, we conducted MLST by using thirteen housekeeping genes amino acid sequences (adk, ccpA, recF, rpoB, spo0A, sucC, glpF, ilvD, pta, purH, pycA, rpoD, tpiA and gmk). The three bacterial strains were all included in the cluster of *Bacillus*



**Fig. 1** Screening of cellulase-producing microorganisms by using Congo red on CMC agar plates. (A) Bacterial isolate from the plankton sample (PB1), (B and C) bacterial isolates from the seagrass samples (SB2 and SB3) and (D) *Escherichia coli* as a negative control

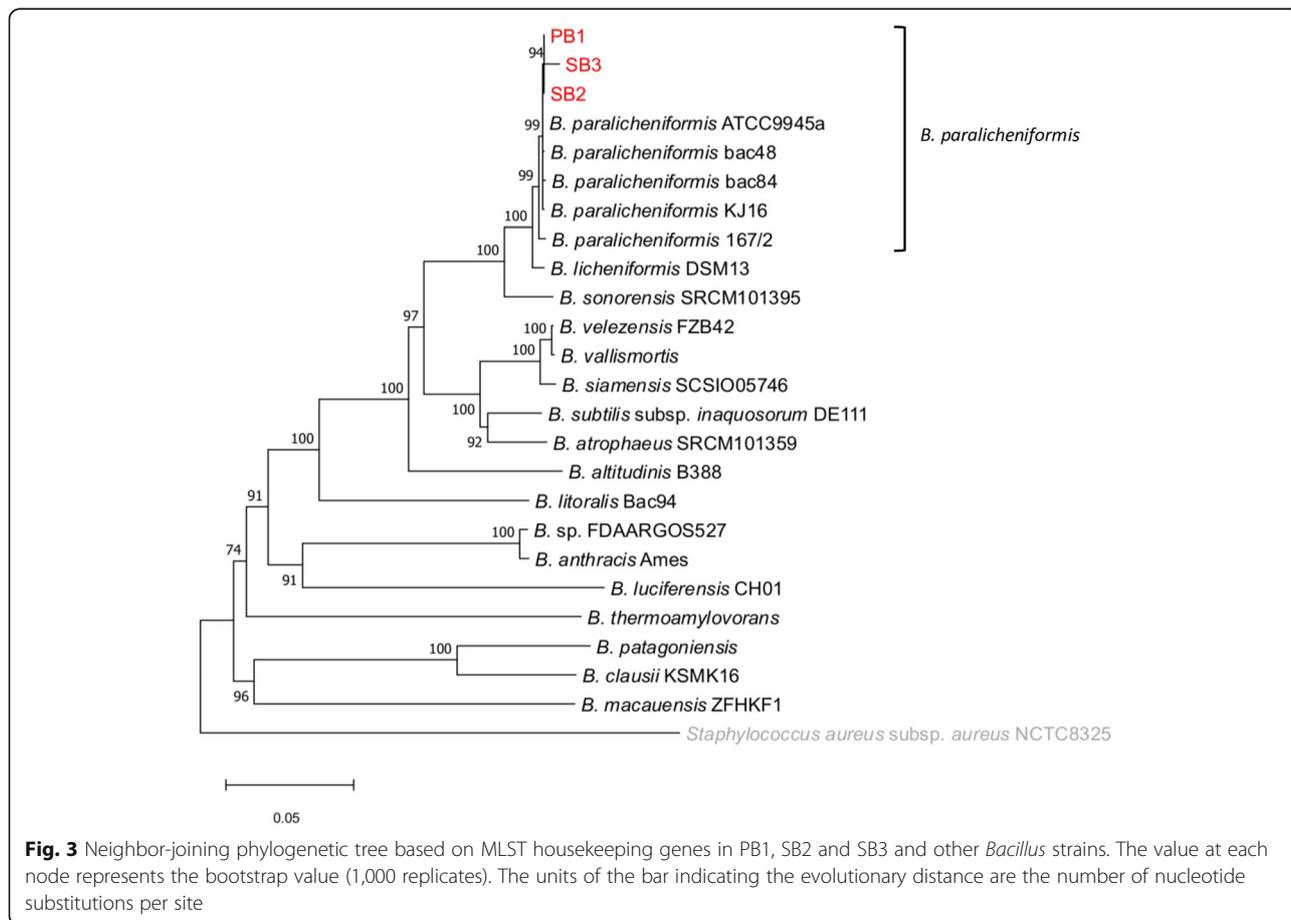


*paralicheniformis* strains with a 100 % bootstrap value. Furthermore, the three isolates formed a cluster with 94 % bootstrap support (Fig. 3).

**Identification of Cellulase Genes in the Isolates’ genomes**

We examined GH functional domains in the amino acid sequences of predicted genes and found 10, 10 and 11

cellulase genes from PB1, SB2 and SB3 strain genomes, respectively. Based on the similarity, these genes were classified into 10 orthologues and designated as Cellulase 1 to 10 (Cel-1 to Cel-10) respectively (Table 2). Amino acid sequences of each cellulase orthologue were 100 % identical among three isolates. The isolates possess each orthologue as a single copy in general. Only SB3 has



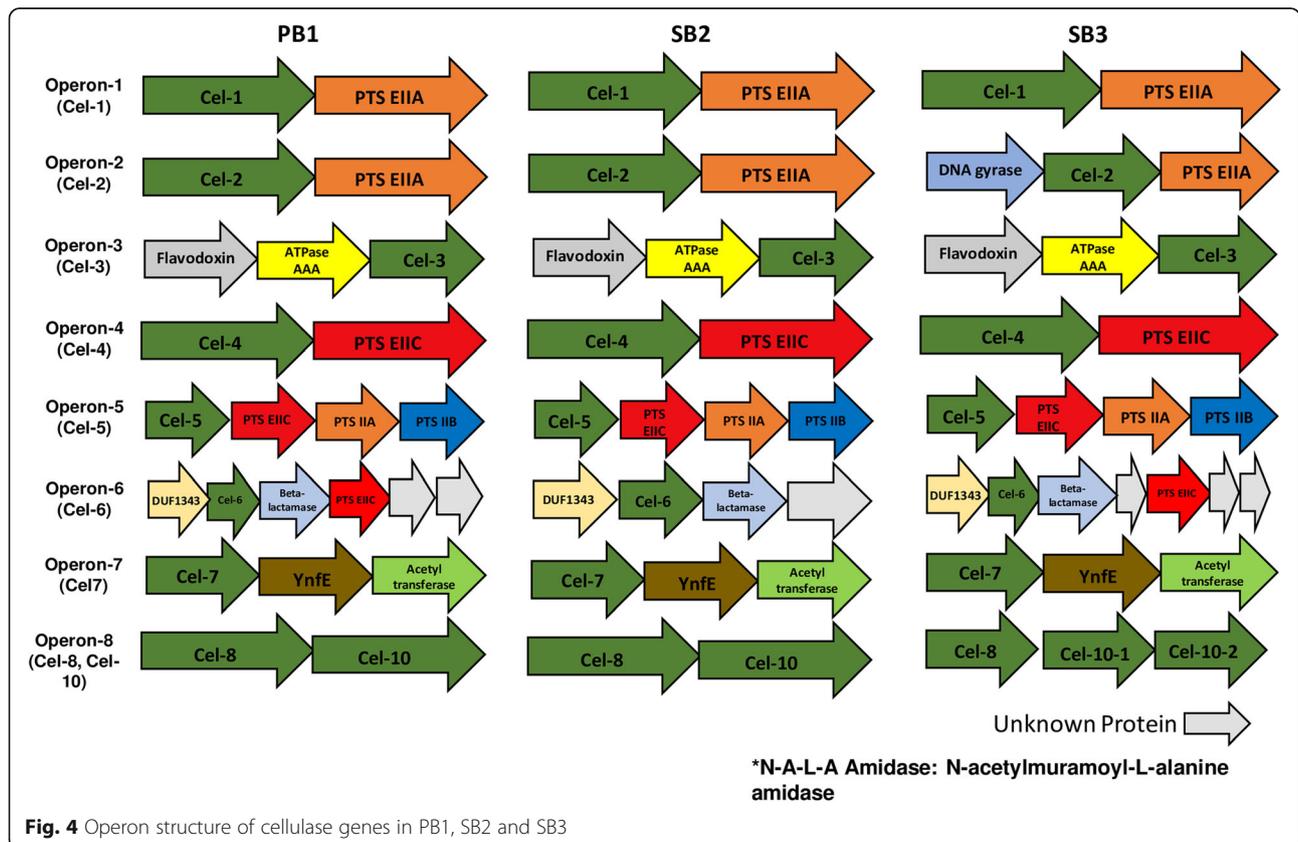
**Table 2** Cellulase genes in the three isolates

Cellulase orthologues	Genes in PB1	Genes in SB2	Genes in SB3	GH family	Cellulase activity
Cel-1	GENE_769	GENE_2609	GENE_4396	GH1	$\beta$ -glucosidase/exoglucanase
Cel-2	GENE_743	GENE_2582	GENE_4369	GH1	
Cel-3	GENE_1298	GENE_3145	GENE_288	GH1	
Cel-4	GENE_718	GENE_2557	GENE_4343	GH1	
Cel-5	GENE_3516	GENE_958	GENE_2665	GH1	
Cel-6	GENE_1100	GENE_2942	GENE_74	GH3	$\beta$ -glucosidase
Cel-7	GENE_3008	GENE_443	GENE_2102	GH5	Endo- $\beta$ -1,4-glucanase
Cel-8	GENE_2740	GENE_173	GENE_1821	GH9	Endo- $\beta$ -1,4-glucanase/ $\beta$ -glucosidase/ exoglucanase
Cel-9	GENE_1657	GENE_3511	GENE_679	GH26	Endo- $\beta$ -1,4-glucanase
Cel-10	GENE_2741	GENE_174	GENE_1822	GH48	Endo- $\beta$ -1,4-glucanase/ exoglucanase

exceptionally two Cel-10 orthologues, GENE\_1822 and GENE\_1823. It is noteworthy that the amino acid sequence of GENE\_1822 shows identity to the former part of Cel-10 orthologues in PB1 and SB2, while GENE\_1823 is identical to the latter part of Cel-10 (Fig. S3). Considering homologies to GH domain, orthologues of Cel-1, Cel-2, Cel-3, Cel-4 and Cel-5 were annotated as members of the GH1 family while Cel-6, Cel-7, Cel-8, Cel-9 and Cel-10 were predicted as GH3, GH5, GH9, GH26 and GH48 cellulases, respectively (Table 2).

**Identification of Operon Structure**

To understand what kind of genes were co-regulated with these cellulase genes, we surveyed structures of operons containing cellulase genes we found. As shown in Fig. 4, nine of the ten cellulase orthologues formed operons. We designated them as Operon-1 to Operon-8. Here, Cel-8 and Cel-10 appeared in the same operons, Operon-8. The structures of these eight operons were generally conserved in the three isolates.



**Fig. 4** Operon structure of cellulase genes in PB1, SB2 and SB3

Cel-1, Cel-2, Cel-4 and Cel-5 orthologues, all of which belonged to GH1 cellulase ( $\beta$ -glucosidase/exoglucanase), formed operons with the phosphoenolpyruvate-dependent sugar phosphotransferase system (PTS) components. However, inferred functions of those PTSs were different among each operon. PTS component in the Operon-1 showed 77 % identity in amino acid sequence with the PTS component of the sucrose transporting in *B. subtilis*, BglP while the sequence of the PTS component in Operon-2 was similar to the *B. subtilis* glucose-transporting PTS component, PtsG with 60.8 % identity [16].

Operon-5 also included three PTS components with the cellulase Cel-5. These three PTS proteins showed 76 %, 70 and 82 % sequence identities with cellobiose importing PTS components in *B. subtilis*, LicC (CelB), LicB (CelA) and LicA (CelC), respectively [17]. Cel-4 also formed with a PTS component in Operon-4, however, this protein did not show high homology with any known PTS components in *B. subtilis*. It only showed 32 % amino acid sequence identities with LicC of *B. subtilis*.

On the other hand, Operon-3, which also included the GH1-type  $\beta$ -glucosidase Cel-3, did not contain PTS components. Genes encoding flavodoxin and AAA-type ATPase were included in this operon, instead [15].

Cel-6, another type of  $\beta$ -glucosidase (GH3), formed operons with two genes in Operon-6. The one encoded a protein with the conserved unknown functional domain DUF1343 (Pfam) and the other is predicted to be a N-acetylmuramyl-L-alanine amidase gene [18].

Cel-7, which was classified as an endo- $\beta$ -1,4-glucanase (GH5), formed Operon-7 with a function-unknown protein YnfE and acetyltransferase genes. The two cellulase, Cel-8 and Cel-10, which were classified as endo- $\beta$ -1,4-glucanase of GH9 and GH48 respectively, formed Operon-8 together in all the three isolates. On the other hand, only Cel-9, which encoded GH26-type endo- $\beta$ -1,4-glucanase, was not included in any operons in all three isolates. It is also revealed that expression of genes included in the same operon were generally co-regulated in these three isolates (Table S2).

#### Comparative Expression Analysis of Cellulase Genes in the Isolates

To reveal whether the identified 10 cellulase genes are actually functional during cellulolysis, we conducted comparative expression analysis. We investigated expression of each gene under two different culture conditions: cellulase-inducing and non-inducing conditions. Enzymatic activity of cellulase was observed under cellulase-inducing conditions, i.e., after 72 h of incubation when it reached the maximum cellulase activity, while no cellulase activity was detected under the non-inducing

conditions in all strains (Fig. 2). Because of the low number of mapped reads at PB1 (Table S3), we only focused on SB2 and SB3. The results revealed that 9 of 10 cellulase genes were upregulated in both SB2 and SB3 under the cellulase-inducing conditions (Table 3). In addition, each strain exhibited particular genes that were particularly highly upregulated under the cellulase-inducing conditions (i.e., showing ten-fold higher expression than under non-inducing conditions) as below.

In the SB2 strain, the expression of Cel-5 (GH1) and Cel-6 (GH3) was highly upregulated. In the SB3 strain, four genes belonging to the GH1 family, Cel-1, Cel-2, Cel-3 and Cel-4, were highly expressed or expressed specifically in the inducing condition. Cel-6 (GH3) and Cel-8 (GH9) were also highly expressed in this isolate. In addition, one of the two Cel-10 (GH48) orthologues in SB3, GENE\_1822 showed the expression pattern specific to the inducing condition.

## Discussion

### Isolation of Cellulase-Producing Microorganisms

Cellulase-producing microorganisms are usually isolated from terrestrial environments [19], whereas there are only a few reports regarding the isolation of these microorganisms from a seawater environment [20]. Harshvardhan et al. (2013) reported the isolation of the cellulolytic marine species *Bacillus* sp. H1666 from seawater samples along the western Indian coast [21]. Additionally, Samira et al. (2011) reported the isolation of *Stenotrophomonas maltophilia* from surface seawater in the Persian Gulf [22]. In our study, we could not isolate any cellulase-producing microorganisms from the surface seawater of the Red Sea. On the other hand, we report the isolation of two bacterial strains from the surface of seagrass.

This result indicates that cellulase-producing microorganisms are usually present in association with cellulose-containing substrates such as seagrass [23], which might also explain the difficulty in isolating cellulase-producing microorganisms from shallow seawater. Previous studies supporting our results such as that of Trivedi et al. (2011) have also reported the isolation of *Bacillus aquimaris*, *B. flexus* NT and *Pseudoalteromonas* CSMCRI-5 strains with cellulolytic potential from green seagrass [4].

In addition, we isolated one cellulase-producing strain from the plankton fraction (i.e., samples collected from the surface seawater in which phytoplankton cells were condensed), suggesting that although the number of phytoplankton cells was very low, cellulase-producing bacteria probably live on the surface of phytoplankton cells in the Red Sea surface waters. To date, there have been no reports of the isolation of cellulase-producing microorganisms from plankton fractions; therefore, our

**Table 3** Comparative expression analysis of cellulase genes in three isolates

Orthologue	GH	SB2			SB3				
		Gene	Inducing	Non inducing	Ratio	Gene	Inducing	Non inducing	Ratio
Cel-1	GH1	GENE_2609	188.18	81.24	2.32	GENE_4396	3.23	0.18	18.22
Cel-2	GH1	GENE_2582	66.47	37.67	1.76	GENE_4369	347.44	17.03	20.40
Cel-3	GH1	GENE_3145	15.46	7.49	2.07	GENE_288	157.71	7.62	20.70
Cel-4	GH1	GENE_2557	7.90	2.82	2.80	GENE_4343	19.90	0.00	CMC
Cel-5	GH1	GENE_958	949.03	2.29	414.79	GENE_2665	21.05	6.57	3.21
Cel-6	GH3	GENE_2942	30.19	0.78	38.78	GENE_74	261.19	0.91	288.46
Cel-7	GH5	GENE_443	48.48	261.73	0.19	GENE_2102	4.23	2.26	1.87
Cel-8	GH9	GENE_173	1.23	0.35	3.48	GENE_1821	43.32	2.08	20.87
Cel-9	GH26	GENE_3511	21.97	6.73	3.26	GENE_679	13.55	7.48	1.81
Cel-10	GH48	GENE_174	4.51	0.66	6.87	GENE_1822	16.64	4.61	3.61

Values in inducing and non-inducing columns represent RPKM in each condition. Ratio columns represents ratio of RPKM values in these two conditions. "CMC" at the ratio column means the expression specific to the non-inducing condition

PB1 strain is the first isolate of a cellulase-producing microorganism associated with marine plankton.

Our growth test revealed that the three isolates are able to grow well at the concentration of 8 % NaCl and still slightly proliferated at 10 % NaCl in NM broth. The tolerances of the isolates against salinity are not at the same level with those of so-called halotolerant strains like *B. subtilis* strain FP-133 since they can grow at the concentration higher than 12.5 % NaCl [24]. However, Schroeter et al. reported that *B. lincheniformis* DSM 13 did not grow at the concentration of more than 8 % (1.4 M) NaCl [25]. Although the culturing conditions are different between Schroeter et al.'s study and ours, the isolates might be tolerant against salinity moderately. Our results also revealed that the isolates showed the growth in CMC broth at the concentration of 8 % NaCl, which allows us to expect that the isolates can degrade the cellulose under high salinity conditions. These might be a feature of cellulase producing bacteria in the Red Sea.

#### Taxonomic Prediction and cellulase activity measurement

*B. paralicheniformis* was recently described as a new species of genus *Bacillus* and isolated from various environments, including marine, freshwater, and food-related niches. In a previous study, the isolation of the bacteria *B. paralicheniformis* bac48 and bac84 from the Red Sea environment was reported by Othoum et al. (2018) [26]. Dass et al. (2018) isolated *B. paralicheniformis* F47 from a salty lake in Algeria [27]. Here, our MLST analysis showed that our three strains were closely related to *B. paralicheniformis* species. Although cellulases remain unexplored in *B. paralicheniformis* species, this species including our isolates might potentially be adaptive to high salinity environments under

the hot climates such as the Red Sea. Our strains showed cellulase activities of 0.75 FPU/ml, 0.70 FPU/ml and 0.59 FPU/ml, respectively. Although no common criteria (unit) for evaluating cellulase activity have been developed, Samira et al. (2011) measured the cellulase activities using the same method as ours except the type of the buffer (see Materials and Methods). They measured the activities of three marine bacterial isolates obtained from the Arabian/Persian Gulf and reported activities of 0.079, 0.074, and 0.072 FPU/ml for their strains. Our isolates showed more than ten times higher cellulase activity than their strains [18].

#### Identification of Cellulase Genes in the Isolates' genomes

We also identified cellulase genes expected to be responsible for cellulolysis in each strain. Recent genome sequencing projects in cellulase-producing microorganisms have revealed the presence of several cellulase genes in their genomes [28]. The genomic analysis of our strains revealed that these three isolates possess the gene set of ten cellulase orthologues in general, and the amino acid sequences of each cellulase orthologue were identical among these three isolates. Regarding Cel-10, PB1 and SB2 have only one orthologue while SB3 has two Cel-10 orthologues. Either of Cel-10 sequences in SB3 was short and almost identical to the former or the latter part of the other isolates' Cel-10 sequences. Therefore, we should take into account the possibility of misannotation and/or sequencing errors in the genomic region including two Cel-10 orthologues in SB2, although this study conducted genome sequencing with PacBio RS II platform at 186 x coverage, allowing us to expect the high accuracy of resultant genome sequences. As far as we surveyed, this is the first report on the active cellulases in *B. paralicheniformis*.

### Operon Structure Identification and Comparative Expression Analysis of Cellulase Genes in the Isolates

To determine cellulase genes expressed during cellulolysis, we conducted comparative expression analysis between two different conditions: cellulase-inducing and non-inducing conditions. Our comparative expression analysis showed that 9 of 10 cellulase genes were upregulated in SB2 and SB3 under the inducing condition. In this analysis, we omitted PB1 as the number of RNA-seq short reads mapped on open reading frame (ORF) regions in PB1 under the inducing condition was much smaller than those in the other strains (Table S3).

This study also revealed that particular types of cellulase genes were highly upregulated during cellulolysis. Most of these genes were predicted to encode  $\beta$ -glucosidases/exoglucanase classified as GH1 family. Our operon structure analysis identified that most of the GH1 cellulase genes formed operons with PTS system components involved in the import of various kind of sugars such as, sucrose, glucose and cellobiose [29–31]. The GH1 cellulase Cel-3 gene were also found to be included in the operon with flavodoxin and ATPases AAA protein, which is expected to be involved in diverse cellular activities [32]. Cel-6 gene, which encoded a  $\beta$ -glucosidase of GH3 family, were commonly included in the operon with a N-acetylmuramyl-L-alanine amidase gene showing 63.2% identity with AmiE of *B. subtilis*. AmiE is revealed to be involved in the pathway of peptidoglycan recycling and in cell wall biogenesis [33]. These results suggest that these highly upregulated cellulase genes were co-regulated together with various genes present in the same operon, which enhanced a wide range of cellular reactions during cellulolysis. It is also noteworthy that the gene sets of highly upregulated GH1  $\beta$ -glucosidases genes were slightly different from each other among the three orthologues, indicating that the strain-specific intracellular regulation might occur during the cellulolysis in each isolated.

On the other hand, no significant upregulation of endo-  $\beta$ -1,4-glucanases were observed during the cellulolysis in the expression analysis, although we confirmed the clear cellulolytic activity from the samples we used for the RNA-seq. It may be because the expression peak of endo-  $\beta$ -1,4-glucanase genes had come a little bit earlier than the peak of enzymatic activity. The cellulase activity was measured by the amount of reducing sugars generated as the final product of the filter paper degradation.  $\beta$ -glucosidases catalyze the division of disaccharides to monosaccharides. In particular, GH3  $\beta$ -glucosidases were predicted to be a secreted protein while the other  $\beta$ -glucosidases (i.e., GH1 proteins) were predicted to be localized inside the cell (Table S4). GH3  $\beta$ -glucosidases were highly upregulated in all three isolates under the inducing condition, and the gene

products were probably catalyzing the degradation of disaccharides outside the cell. High activity of GH3  $\beta$ -glucosidases is consistent with the observation of the highest enzymatic activities at the same timing. More detailed time course settings may be required to characterize the expression pattern of endo- $\beta$ -1,4-glucanases during microbial cellulolysis.

### Conclusions

Three cellulase-producing bacteria were obtained from the plankton fraction and seagrass surface in the Red Sea environment. We identified ten cellulase genes in their genomes and revealed that those genes are expressed during cellulolysis. The Red Sea exhibits high salinity (36–40 p.s.u.) and high surface temperatures (24 °C in spring and up to 35 °C in summer) [14]. The isolates obtained in this study are expected to produce cellulases that may be stable under such harsh conditions. Further analysis will provide valuable information on microbial cellulases in the Red Sea, which will contribute to industrial applications such as the development of plant biomass biorefineries.

### Methods

#### Collection of Samples from the Red Sea

Marine samples were collected from a coastal region of the Red Sea at Thuwal in Saudi Arabia on August 26 and September 30, 2015 for the isolation of cellulase-producing microorganisms. A seawater sample was obtained from the seawater surface at the site at 22°17.444'N, 39°03.183'E using a Niskin bottle. A seagrass (sargassum weed) sample was obtained from the KAUST coastal marina (22°18'16.7"N 39°06'12.1"E), and a plankton sample was collected from the sea surface by drawing a net with a mesh size of 0.63  $\mu$ m at 1 knot for 10 min. All samples were placed in sterile tubes and stored at 4 °C until use. The sample collection was followed by the institutional field research policy and procedure.

#### Isolation and Screening of Cellulase-Producing Microorganisms

The seawater and plankton samples were vortexed for 15 min in sterilized 50 ml tubes and then allowed to settle for 5 min. Ten-fold serial dilutions of each sample were prepared in sterilized distilled water, and 0.1 ml diluted samples were spread on the surface of NM plates containing 0.3 % beef extract, 0.5 % peptone, 0.5 % NaCl, and 1.7 % agar [pH 7.0]. The plates were incubated at 30 °C for 48 h. One gram of seagrass was also measured and added to 10 ml of sterilized water. One gram of glass beads (425–600  $\mu$ m) was placed in 1 M HCl for 1 h and then rinsed with distilled water. The resultant acid-washed glass beads were added to the seagrass

tubes, which were then vortexed for 10 min. The tubes were left for 5 min to allow the solids to settle, and the supernatant was collected and diluted to use it as an inoculation source and plated on NM plates [15].

To screen for cellulase-producing microorganisms, single colonies from NM were transferred to CMC agar composed of 0.2 % NaNO<sub>3</sub>, 0.1 % K<sub>2</sub>HPO<sub>4</sub>, 0.05 % MgSO<sub>4</sub>, 0.05 % KCl, 0.2 % CMC sodium salt, 0.02 % peptone, and 1.7 % agar. Following it, the plates were incubated at 30 °C for 48 h. Zones of hydrolysis were visualized by flooding the plates with 0.1 % Congo red for 20 min and then washing the plates with 1 M NaCl for 20 min [15].

#### Preparation of Extracellular Cellulase Enzymes

The obtained cellulase-producing strains were precultured in 100 ml of nutrient broth and incubated at 30 °C for 48 h at 200 rpm. Aliquot of 2 ml was used as the inoculum for enzyme production, and the broth culture system was composed of 0.2 % NaNO<sub>3</sub>, 0.1 % K<sub>2</sub>HPO<sub>4</sub>, 0.05 % MgSO<sub>4</sub>, 0.05 % KCl, 0.02 % peptone and a Whatman No. 1 Filter paper (1 × 6 cm strip, 0.05 g per 20 ml) [34]. The broth cultures were incubated for four days at 30 °C with shaking at 200 rpm. Cell growth was monitored every 24 h by determining the optical density at 600 nm, and cellulase activity was measured every 24 h during incubation. The cultures were centrifuged at 8,000 rpm for 10 min, and the supernatant was used as a source of crude enzyme for the determination of enzyme activity.

#### Growth test under increasing salinity condition

The growth of the three bacterial isolates under high salt condition (i.e., 0.2 %, 2 %, 4 %, 6 %, 8 and 10 % (w/v) of NaCl) was studied in Nutrient broth media (NM) and compared with CMC broth media (i.e., where the cellulose is a sole carbon source).

Six Nutrient broth media were prepared which composed of (0.3 % beef extract, 0.5 % peptone, and NaCl (i.e., 0.2 %, 2 %, 4 %, 6 %, 8 and 10 % (w/v))). Likewise, for the CMC broth media, six media were prepared that contain (0.2 % NaNO<sub>3</sub>, 0.1 % K<sub>2</sub>HPO<sub>4</sub>, 0.05 % MgSO<sub>4</sub>, 0.05 % KCl, 0.02 % peptone, CMC 0.2 % and six different concentrations of NaCl i.e., (0.2 %, 2 %, 4 %, 6 %, 8 and 10 % (w/v)). The broth media were incubated at 30 °C with shaking at 200 rpm. Cell growth was monitored every 24 h up to 5 days by determining the optical density at 600 nm.

#### Measurement of Cellulase Activity

The filter paper assay of Hankin and Anagnostakis was used to measure total cellulase activity in the culture [5]. Total cellulase activity was determined by measuring the amount of reducing sugar formed by the degradation of

filter paper strips. Then, 0.5 mL of the supernatant of the culture was incubated in 1.0 mL of 0.05 M sodium citrate buffer (pH 4.8) with a Whatman No. 1 filter paper strip, 1.0 × 6.0 cm (= 50 mg). After incubation for one hour at 50 °C, the reaction was stopped by adding 3 ml of dinitrosalicylic acid to the reaction mixture [35]. The amount of reducing sugars released was estimated spectrophotometrically at 540 nm using glucose as a standard. The enzymatic activity of total cellulases was defined in FPU/ml. One unit of cellulase activity is defined as the amount of enzyme releasing 1 μmol of reducing sugars (measured as glucose) from filter paper per mL per min [34].

#### Whole-Genome Sequencing

DNA samples for the whole-genome sequencing were prepared by culturing the isolates in nutrient broth overnight at 30 °C with shaking at 150 rpm. DNAs were extracted from the isolates using the Qiagen DNeasy Blood & Tissue Kit following the manufacturer's instructions [36]. The obtained DNA was quantified with a Qubit dsDNA BR assay kit (Thermo Fisher Scientific). Electrophoresis was also performed in a 1 % agarose gel to confirm that the length of the DNAs was longer than 40 Kb. Fifty micrograms of DNA from each strain were used for the library construction at the Bioscience Core Lab at KAUST following the manufacturer's instruction (Pacific Biosciences) [37, 38]. The sequencing was also performed at the Bioscience Core Lab using a PacBio RS II sequencing platform (Pacific Biosciences). The large-insert libraries were sequenced in single-molecule real-time (SMRT) sequencing cells using P6-C4 chemistry.

#### De novo Assembly of the Genome Sequencing Data

The raw reads of each isolate generated by the PacBio RSII platform were *de novo* assembled and polished with HGAP3/Quiver [39]. The overlapping ends were visually checked by using Gepard v1.40, which would help indicates the circular genomes [40]. Circular closure was performed by using Minimus2 (<http://amos.sourceforge.net/wiki/index.php/Minimus2>) to trim the ends and permute the genome to begin at the DnaA gene (identified by BLAST), followed by Quiver-based error correction for a final closed genome. We used default parameters for Minimus2 [41].

#### Genome Annotation

The FGENESB\_annotator was used to predict the presence of likely genes in the genomes of the bacterial isolates (<http://www.softberry.com/berry.phtml?topic=fgenesb&group=programs&subgroup=gfindb>) [42]. Recently, all known cellulases were classified based on sequence comparison into 16 glycoside-hydrolase (GH) orthologous groups [43]. Since each GH group in the

CAZy database is known to have a corresponding Pfam domain (Table S5) [44], the Pfam annotation was used for the identification of cellulase genes in each strain obtained in this study. The Pfam annotation of the deduced amino acid sequences of predicted genes was conducted at the Pfam-A database using the hmmscan program in HMMER (v3.0). The annotations of GH families shown in Table S5 were used to extract candidate cellulase genes, with an E-value cutoff of  $<1.0e-60$  [45].

### Multilocus Sequence Typing (MLST) Analysis

The phylogenetic relationships between the isolates and other *Bacillus* species were determined by multilocus sequence typing (MLST). The amino acid sequences of thirteen housekeeping genes from *Bacillus licheniformis* Table 4 (adk, ccpA, recF, rpoB and sucC) [46], *Bacillus subtilis* WB800N (glpF, ilvD, pta, purH, pycA, rpoD and tpiA) [47] and *Bacillus anthracis* CZC5 (gmk) [48] were obtained from PubMLST (<http://pubmlst.org/>). These thirteen genes are all the gene sets which linked to *Bacillus* in PubMLST, except for apo0A gene which didn't find in the outgroup genome (*Staphylococcus aureus* subsp. *aureus* NCTC8325).

To perform the MLST analysis, the protein sequences of twenty-two different *Bacillus* species (Table S6) were obtained from the assembly database at the National Center for Biotechnology Information (<https://www.ncbi.nlm.nih.gov/>) and converted to the database format with the makeblastdb program of the Blast+ package version 2.2.31 [49]. For three isolates obtained in this study, amino acid sequences of predicted genes were converted to the BLASTp database format. The identification of housekeeping genes for MLST analysis was conducted via BLASTp searches using thirteen housekeeping gene sequences obtained from PubMLST as a query against the protein databases of each *Bacillus* species. The sequences of the top-scoring hit with an E-value lower than  $1.0e-80$  and 100% query coverage were selected from the *Bacillus* genomes for each of thirteen genes. The selected genes were then used as query for the reciprocal Blastp search against proteomes of *B. licheniformis* DSM 13 (GenBank assembly accession: GCF\_000011645.1), *B. anthracis* CZC5 (GCF\_000534935.1) and *B. subtilis* WB800N (GCF\_003610955.1) and confirmed their orthology. The sequence alignment of each gene was conducted using MAFFT version 7.394 with default parameters, and the obtained alignments were concatenated to a single alignment manually [50]. The phylogenetic tree was constructed by the neighbor-joining method with MEGA7 [51]. Branching quality was evaluated by using a bootstrapping confidence value with 1,000 replicates [52].

### RNA Extraction

The isolates were cultured under two different types of culture conditions: cellulase-inducing and non-inducing conditions. Under inducing conditions, the isolates were cultured in media composed of 0.02% peptone, 0.2%  $K_2HPO_4$ , 0.05%  $MgSO_4 \cdot 7H_2O$ , and 0.2%  $NaNO_3$  with Whatman No.1 filter paper. In the non-inducing conditions, the filter paper was excluded from the media. Cellulase activities were examined in both conditions with the method described in the Measurement of Cellulase Activity section.

Total RNA was extracted from each of these conditions after 72 h of incubation when it reached the maximum cellulase activity by using a QIAGEN RNeasy mini Kit (Qiagen, Valencia, CA) according to the manufacturer's protocol. Total RNA quality and concentrations were determined using the Agilent RNA 6000 Pico kit (Agilent, Santa Clara, CA) in a 2100 Bioanalyzer (Agilent). Paired-end libraries with approximate average insert lengths of 200 base pairs were synthesized using the Genomic Sample Prep kit (Illumina, San Diego, CA) according to the manufacturer's instructions. Libraries were sequenced on the Illumina HiSeq 4000 platform (Illumina, San Diego, CA) with support from the KAUST Bioscience Core laboratory [53].

### Expression Analysis in the Isolates

The nucleotide sequences of ORF region of genes predicted with FGENESB program from the genome sequences were employed to build an index using the bowtie2-build program in the bowtie2 package [54]. Only one side of the paired-end reads generated for each isolate in the RNA-seq experiments described above were aligned to the sequence index by using the Bowtie2 alignment program. The gene expression rate was determined in Reads Per Kilobase of transcript per Million mapped reads (RPKM) units with the following RPKM equation, where the number of short reads mapped onto each ORF region of predicted gene ( $r_g$ ) was normalized and divided by the feature length ( $fl_g$ ) multiplied by the total number of mapped reads from the sequencing run (R) [55]:

$$RPKM = \frac{r_g \times 10^9}{fl_g \times R}$$

### Operon Structure Identification

The FGENESB\_annotator web server was used under the default setting to predict the operon structure using the extracted genes as the input. The expression rate of the genes in the operons was determined by the same method described in the expression analysis in the isolates section.

## Abbreviations

NM: Nutrient Media; CMC: Carboxymethyl Cellulose; ORF: Open Reading Frame; RPM: Revolutions Per Minute; GH: Glycoside hydrolysis; FPU: Filter paper unit; RPKM: Reads Per Kilobase of transcript per Million mapped reads; MLST: Multilocus Sequence Typing

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12866-021-02316-w>.

**Additional file 1.** Supplemental Materials. Description of data: Figures S1 to S3; Tables S1 to S6

## Acknowledgements

We acknowledge KAUST Core laboratories of Coastal and Marine Resources and Bioscience for their support on the field sampling and sequencing.

## Authors' contributions

TG, KM conceived the study; YS, KM designed the study and collected samples; SF, YS, MA conducted experiments; SF prepared the first draft with YS; TG, KM supervised research; All authors read, edited, and approved the manuscript.

## Funding

The research reported in this publication was supported through funding from King Abdullah University of Science and Technology (KAUST), under award numbers BAS/1/1059-01-01, URF/1/1976-03-01, URF/1/1976-17-01, URF/1/1976-20-01 and FCS/1/3326-01-01. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

## Availability of data and materials

PacBio sequencing data and Illumina RNA-seq data have been deposited to the DDBJ Sequence Read Archive under accession numbers DRR228801-DRR228806, and nucleotide sequences of genome assemblies and predicted genes are available in the DDBJ/ENA/GenBank nucleotide database under accession numbers AP023088-AP023090. PB1, SB2 and SB3 are also aliased as RSC (Red Sea Cellulase-producing bacterial strain)-1, RSC-2 and RSC-3 respectively in these databases. CAZy (<http://www.cazy.org/>), PubMLST (<http://pubmlst.org/>), Pfam (<https://pfam.xfam.org/>) and NCBI (<https://www.ncbi.nlm.nih.gov/>) were used for obtaining reference data in this study, and these databases can be accessed openly.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Computational Bioscience Research Center (CBRC), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia. <sup>2</sup>Marine Open Innovation Institute (MaOI), Shizuoka, Japan.

Received: 1 March 2021 Accepted: 8 September 2021

Published online: 22 September 2021

## References

- Zverlov WV, Schwarz WH. Bacterial cellulose hydrolysis in anaerobic environmental subsystems - *Clostridium thermocellum* and *Clostridium stercorarium*, thermophilic plant-fiber degraders. *Ann N Y Acad Sci*. 2008; 1125(1):298–307.
- Teeri TT. Crystalline cellulose degradation: new insight into the function of cellobiohydrolases. *Trends Biotechnol*. 1997;15(5):160–7.
- Fukuoka A, Dhepe PL. Catalytic conversion of cellulose into sugar alcohols. *Angew Chemie Int Ed Engl*. 2006;45(31):5161–3.
- Trivedi N, Gupta V, Kumar M, Kumari P, Reddy CRK, Jha B. An alkali-halotolerant cellulase from *Bacillus flexus* isolated from green seaweed *Ulva lactuca*. *Carbohydr Polym*. 2011;83(2):891–7.
- Behera BC, Sethi BK, Mishra RR, Dutta SK, Thatoi HN. Microbial cellulases – Diversity & biotechnology with reference to mangrove environment: A review. *J Genet Eng Biotechnol*. 2017;15(1):197–210.
- Galante YM, De Conti A, Monteverdi R. Application of Trichoderma enzymes in the textile industry. *Trichoderma & Gliocladium*. 2014;2:311–25.
- Robson LM, Chambliss GH. Cellulases of bacterial origin. *Enzyme Microb Technol*. 1989;11(10):626–44.
- Sukharnikov LO, Cantwell BJ, Podar M, Zhulin IB. Cellulases: ambiguous nonhomologous enzymes in a genomic perspective. *Trends Biotechnol*. 2011;29(10):473–9.
- Yang G, Yang D, Wang X, Cao W. A novel thermostable cellulase-producing *Bacillus licheniformis* A5 acts synergistically with *Bacillus subtilis* B2 to improve degradation of Chinese distillers' grains. *Bioresour Technol*. 2021; 325:124729.
- Thapa S, Mishra J, Arora N, Mishra P, Li H, O'Hair J, et al. Microbial cellulolytic enzymes: diversity and biotechnology with reference to lignocellulosic biomass degradation. *Rev Environ Sci Biotechnol*. 2020;19(3):621–48.
- Malik WA, Khan HM, Javed S. Bioprocess Optimization for Enhanced Production of Bacterial Cellulase and Hydrolysis of Sugarcane Bagasse. *BioEnergy Res*. 2021.
- Davies G, Henrissat B. Structures and mechanisms of glycosyl hydrolases. *Structure*. 1995;3(9):853–9.
- Cantarel BI, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. *Nucleic Acids Res*. 2009;37(SUPPL. 1):233–8.
- Behzad H, Ibarra MA, Mineta K, Gojobori T. Metagenomic studies of the Red Sea. *Gene*. 2016;576(2):717–23.
- Kasana RC, Salwan R, Dhar H, Dutt S, Gulati A. A rapid and easy method for the detection of microbial cellulases on agar plates using Gram's iodine. *Curr Microbiol*. 2008;57(5):503–7.
- Deutscher J, Galinier A, Martin-verstraete I. Carbohydrate Uptake and Metabolism. In: Sonenshein AL, Hoch JA, Losick R, editors. *Bacillus subtilis Its Closest Relatives: From Genes to Cells*. ASM press; 2001. p. 129–50.
- Snider J, Thibault G, Houry WA. The AAA+ superfamily of functionally diverse proteins. *Genome Biol*. 2008;9(4):1–8.
- Litzinger S, Duckworth A, Nitzsche K, Risinger C, Wittmann V, Mayer C. Muropeptide rescue in *Bacillus subtilis* involves sequential hydrolysis by  $\beta$ -N-acetylglucosaminidase and N-acetylmuramyl-L-alanine amidase. *J Bacteriol*. 2010;192(12):3132–43.
- Niyonzima FN. Detergent-compatible bacterial cellulases. *J Basic Microbiol*. 2019;59(2):134–47.
- Lynd LR, Weimer PJ, Zyl WH Van, Pretorius IS. Microbial cellulose utilization: Fundamentals and Biotechnology. *Bioresour Technol*. 2002;66(3):506–77.
- Harshvardhan K, Mishra A, Jha B. Purification and characterization of cellulase from a marine *Bacillus* sp. H1666: A potential agent for single step saccharification of seaweed biomass. *J Mol Catal B Enzym*. 2013;93:51–6.
- Samira M, Mohammad R, Gholamreza G. Carboxymethyl-cellulase and filter-paperase activity of new strains isolated from Persian Gulf. *Microbiol J*. 2011; 1(1):8–16.
- Joint I, Mühling M, Querellou J. Culturing marine bacteria - An essential prerequisite for biodiscovery: Minireview. *Microb Biotechnol*. 2010;3(5):564–75.
- Setyorini E, Kim Y, Takenaka S, Murakami S, Aoki K. Purification and characterization of a halotolerant intracellular protease from *Bacillus subtilis* strain FP-133. *J Basic Microbiol*. 2006;46(4):294–304.
- Schroeter R, Hoffmann T, Voigt B, Meyer H, Bleisteiner M, Muntel J, et al. Stress responses of the industrial workhorse *Bacillus licheniformis* to osmotic challenges. *PLoS One*. 2013;8(11):e80956.
- Othoum G, Bougouffa S, Razali R, Bokhari A, Alamoudi S, Antunes A, et al. In silico exploration of Red Sea *Bacillus* genomes for natural product biosynthetic gene clusters. *BMC Genomics*. 2018;19(1):1–11.
- Daas MS, Rosana ARR, Acedo JZ, Douzane M, Nateche F, Kebbouche-Gana S, et al. Draft Genome Sequence of *Bacillus paralicheniformis* F47, Isolated from an Algerian Salty Lake. *Genome Announc*. 2018;6(13):4–6.
- Feng Y, Duan C-J, Pang H, Mo X-C, Wu C-F, Yu Y, et al. Cloning and identification of novel cellulase genes from uncultured microorganisms in

- rabbit cecum and characterization of the expressed cellulases. *Appl Microbiol Biotechnol.* 2007;75(2):319–28.
29. Bourassa S, Vadeboncoeur C. Expression of an inducible enzyme II fructose and activation of a cryptic enzyme II glucose in glucose-grown cells of spontaneous mutants of *Streptococcus salivarius* lacking the low-molecular-mass form of III<sup>man</sup>, a component of the phosphoenolpyruvate:mannose phosphotransferase system. *J Gen Microbiol.* 1992;138(4):769–77.
  30. Arnaud M, Vary P, Zagorec M, Klier A, Debarbouille M, Postma P, et al. Regulation of the sacPA operon of *Bacillus subtilis*: Identification of phosphotransferase system components involved in SacT activity. *J Bacteriol.* 1992;174(10):3161–70.
  31. Kleerebezemab M, Hols P, Hugenholtz J. Lactic acid bacteria as a cell factory: Rerouting of carbon metabolism in *Lactococcus lactis* by metabolic engineering. *Enzyme Microb Technol.* 2000;26(9–10):840–8.
  32. Sancho J. Flavodoxins: Sequence, folding, binding, function and beyond. *Cell Mol Life Sci.* 2006;63(7–8):855–64.
  33. Zoll S, Pätzold B, Schlag M, Götz F, Kalbacher H, Stehle T. Structural basis of cell wall cleavage by a staphylococcal autolysin. *PLoS Pathog.* 2010;6(3):e1000807.
  34. Gupta P, Samant K, Sahu A. Isolation of cellulose-degrading bacteria and determination of their cellulolytic potential. *Int J Microbiol.* 2012;5:78925.
  35. Shareef I, Sathesh M, Christopher SX. Isolation and Identification of Cellulose Degrading Microbes. *Int J Innov Res Sci Eng Technol.* 2015;8(4):6788–93.
  36. Troedsson C, Simonelli P, Nägele V, Nejstgaard JC, Frischer ME. Quantification of copepod gut content by differential length amplification quantitative PCR (dla-qPCR). *Mar Biol.* 2009;156(3):253–9.
  37. Woo HL, Ballor NR, Hazen TC, Fortney JL, Simmons B, Davenport KW, et al. Complete genome sequence of the lignin-degrading bacterium *Klebsiella* sp. strain BRL6-2. *Stand Genomic Sci.* 2014;9(1):1–9.
  38. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics.* 2012;13(1):341.
  39. Qiu J, Yang Y, Zhang J, Wang H, Ma Y, He J, et al. The complete genome sequence of the nicotine-degrading bacterium *Shinella* sp. HZN7. *Front Microbiol.* 2016;7:1348.
  40. Krumsiek J, Arnold R, Rattei T. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics.* 2007;23(8):1026–8.
  41. Sommer DD, Delcher AL, Salzberg SL, Pop M. Minimus: a fast, lightweight genome assembler. *BMC Bioinformatics.* 2007;8(1):64.
  42. Solovyev VV, Salamov A, Seledtsov I, Vorobyev D, Bachinsky A. Automatic Annotation of Bacterial Community Sequences and Application To Infections Diagnostic. In: *Proceedings of the International Conference on Bioinformatics Models, Methods and Algorithms.* 2011. p. 346–53.
  43. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* 2008;37(suppl\_1):D233–8.
  44. Park BH, Karpinetz T V., Syed MH, Leuze MR, Uberbacher EC. CAZymes Analysis Toolkit (cat): Web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZy database. *Glycobiology.* 2010;20(12):1574–84.
  45. Punta M, Coghill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, et al. The Pfam protein families database. *Nucleic Acids Res.* 2012;40(D1):290–301.
  46. Jeong DW, Lee B, Lee H, Jeong K, Jang M, Lee JH. Urease characteristics and phylogenetic status of *Bacillus paralicheniformis*. *J Microbiol Biotechnol.* 2018;28(12):1992–8.
  47. Bóka B, Manczinger L, Kocsusbé S, Shine K, Alharbi NS, Khaled JM, et al. Genome analysis of a *Bacillus subtilis* strain reveals genetic mutations determining biocontrol properties. *World J Microbiol Biotechnol.* 2019;35(3):1–14.
  48. Bartoszewicz M, Marjańska PS. Milk-originated *Bacillus cereus sensu lato* strains harbouring *Bacillus anthracis*-like plasmids are genetically and phenotypically diverse. *Food Microbiol.* 2017;67:23–30.
  49. Tsimidis M, Bachoumis G, Mimouli K, Kyriakopoulou Z, Robertson DL, Markoulatos P, et al. T-RECS: Rapid and large-scale detection of recombination events among different evolutionary lineages of viral genomes. *BMC Bioinformatics.* 2017;18(1):1–8.
  50. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013;30(4):772–80.
  51. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol.* 2016;33(7):1870–4.
  52. Akamatsu R, Suzuki M, Okinaka K, Sasahara T, Yamane K, Suzuki S, et al. Novel Sequence Type in *Bacillus cereus* Strains Associated with Nosocomial Infections and Bacteremia, Japan. *Emerg Infect Dis.* 2019;25(5):883–90.
  53. Mizrahi E, Hefer CA, Ranik M, Joubert F, Myburg AA. De novo assembled expressed gene catalog of a fast-growing Eucalyptus tree produced by Illumina mRNA-Seq. *BMC Genomics.* 2010;11(1).
  54. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9(4):357.
  55. Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci.* 2012;131(4):281–5.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://www.biomedcentral.com/submissions)

