



Complete genome sequence of *Rhodothermaceae* bacterium RA with cellulolytic and xylanolytic activities

Kok Jun Liew¹ · Seng Chong Teo¹ · Mohd Shahir Shamsir¹ · Rajesh Kumar Sani² · Chun Shiong Chong¹ · Kok-Gan Chan^{3,4} · Kian Mau Goh¹

Received: 13 June 2018 / Accepted: 6 August 2018 / Published online: 13 August 2018
© The Author(s) 2018

Abstract

Rhodothermaceae bacterium RA is a halo-thermophile isolated from a saline hot spring. Previously, the genome of this bacterium was sequenced using a HiSeq 2500 platform culminating in 91 contigs. In this report, we report on the resequencing of its complete genome using a PacBio RSII platform. The genome has a GC content of 68.3%, is 4,653,222 bp in size, and encodes 3711 genes. We are interested in understanding the carbohydrate metabolic pathway, in particular the lignocellulosic biomass degradation pathway. Strain RA harbors 57 glycosyl hydrolase (GH) genes that are affiliated with 30 families. The bacterium consists of cellulose-acting (GH 3, 5, 9, and 44) and hemicellulose-acting enzymes (GH 3, 10, and 43). A crude cell-free extract of the bacterium exhibited endoglucanase, xylanase, β -glucosidase, and β -xylosidase activities. The complete genome information coupled with biochemical assays confirms that strain RA is able to degrade cellulose and xylan. Therefore, strain RA is another excellent member of family *Rhodothermaceae* as a repository of novel and thermostable cellulolytic and hemicellulolytic enzymes.

Keywords Cellulase · Xylanase · Halophile · *Rhodothermaceae* · *Rhodothermus*

Introduction

Lignocellulosic biomass is mainly composed of cellulose, hemicellulose, and lignin. The degradation of these complex structures involves a series of enzymes that work synergistically. Enzymes responsible for cellulose degradation are cellulase, including endoglucanase (EC 3.2.1.4), β -glucosidase (EC 3.2.1.21), and exoglucanase (EC 3.2.1.91). These

enzymes are classified into glycosyl hydrolase (GH) families GH1, GH3, GH5, GH6, GH7, GH8, GH9, GH12, GH45, and GH48 (Bohra et al. 2018). Hemicellulose is enzymatically hydrolyzed by a mixture of enzymes, including xylanase (EC 3.2.1.8), β -galactosidase (EC 3.2.1.23), β -mannosidase (EC 3.2.1.25), β -glucuronidase (EC 3.2.1.31), β -xylosidase (EC 3.2.1.37), β -D-fucosidase (EC 3.2.1.38), and α -L-arabinofuranosidase (EC 3.2.1.55). These hemicellulases are mainly found in the GH families GH2, GH10, GH11, GH16, GH26, GH30, GH31, GH36, GH43, GH51, GH74, and GH95 (Bohra et al. 2018). Enzymes such as laccase, lignin peroxidase, and manganese peroxidase are also crucial in lignocellulosic biomass degradation, in particular, the lignin moiety. The majority of these enzymes can be found in the Auxiliary Activities (AA) families listed in the CAZY database (Lombard et al. 2013).

Family *Rhodothermaceae* has not been studied extensively for lignocellulose degradation. Members of this family are rod- or cocci-shaped, stain gram negative, non-sporulating, chemoorganotrophic aerobes, and are known to produce pigments (Park et al. 2014). Currently members of family *Rhodothermaceae* consists of six genera: *Rhodothermus* (Alfredsson et al. 1988; Marteinson et al.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s13205-018-1391-z>) contains supplementary material, which is available to authorized users.

✉ Kian Mau Goh
gohkianmau@utm.my

- ¹ Faculty of Science, Universiti Teknologi Malaysia, 81300 Skudai, Johor, Malaysia
- ² Department of Chemical and Biological Engineering, South Dakota School of Mines and Technology, Rapid City, USA
- ³ Institute of Biological Sciences, Faculty of Science, University of Malaya, 50603 Kuala Lumpur, Malaysia
- ⁴ International Genome Centre, Jiangsu University, Zhenjiang 212013, People's Republic of China

2010), *Salinibacter* (Antón et al. 2002; Makhdoumi-Kakhki et al. 2012), *Salisaeta* (Vaisman and Oren 2009), *Longimonas* (Xia et al. 2015), *Longibacter* (Xia et al. 2016), and *Natronotalea* (Sorokin et al. 2017). Members of the genera *Rubricoccus* and *Rubrivirga* were previously affiliated to family *Rhodothermaceae* (Park et al. 2011, 2013; Goh et al. 2016), but have recently been reassigned and classified as members of a new family, family *Rubricoccaceae* (Munoz et al. 2016). Genome sequences are available from representatives of hall genera of family *Rhodothermaceae* except for any representative of genus *Natronotalea*. To date, complete genome sequences of two representatives of this family, namely, *Rhodothermus marinus* and *Salinibacter ruber*, are available. Strain RA is a halo-thermophile (optimum growth at 2% w/v NaCl, 50 °C) which was isolated from a saline hot spring located on Langkawi Island, Malaysia (6°25'22"N, 99°48'49"E) (Goh et al. 2016; Chan et al. 2017). Due to the low DNA–DNA similarity as measured by Genome-to-Genome Distance Calculator (GGDC), 16S rRNA gene similarity, and the housekeeping genes to the other members of the family *Rhodothermaceae*, strain RA has been assigned as an unclassified taxon of the family *Rhodothermaceae*, order *Bacteroidetes* Order II. *Incertae sedis*, and phylum *Bacteroidetes*. Detailed the low 16S rRNA gene sequence similarity of strain RA (89%) compared to *Rhodothermus spp.*, *Salisaeta longa*, *Longibacter salinarum*, *Longimonas halophila*, *Salinibacter spp.*, and *Natronotalea spp.* suggests that it should be included as a new species in a newly created genus status. The genome of strain RA was initially sequenced in 2015 using a HiSeq 2500 platform and assembled into 91 contigs (Goh et al. 2016). Here, we report on the full genome using a PacBio single-molecule sequencing platform and in the presence of lignocellulose biomass-degrading enzymes.

Materials and methods

Genome sequencing, assembly, and annotation

Strain RA (KCTC 62031) was originally isolated from a hot spring located in Langkawi, a Malaysian island (Goh et al. 2016). The cells were resuscitated from 20% (v/v) glycerol stock, grown on marine agar plates (pH 7.5), and incubated at 50 °C for 48 h. Colonies on the agar plates were scraped and DNA extraction using a Quick-DNA™ Miniprep Plus kit (Zymo Research, Irvine, USA). The extracted genomic DNA was analyzed using a NanoDrop 1000 spectrophotometer and Qubit® 3.0 fluorometer (Thermo Scientific, Waltham, USA) to check its purity ($A_{260/280}$ ratio) and concentration. The genomic DNA was then constructed into a 20-kb SMRTbell™ template library and sequenced using a PacBio RSII sequencing platform (Pacific Biosciences, CA,

USA). The resulting sequence was assembled using a PacBio Hierarchical Genome Assembly Process (HGAP) algorithm version 2 (Chin et al. 2013). The final assembled genome was analyzed and annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) version 2.10 (Tatusova et al. 2016). (Tatusova et al. 2016). A cluster of orthologous genes (COG) (Tatusov et al. 2003) was carried out for gene function analysis. Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto 2000; Kanehisa et al. 2017) was utilized for pathway analysis. The GH proteins from strain RA were further classified using dbCAN HMMs 5.0 (Yin et al. 2012), and the results were validated with the annotations available online in the Carbohydrate-Active Enzymes (CAZy) database (Lombard et al. 2013).

Enzymatic assay of bacterial whole cell lysate

To induce production of both cellulolytic and xylanolytic enzymes, strain RA was grown in marine broth supplemented with both 0.1% (w/v) carboxymethyl cellulose (CMC) and 0.1% (w/v) beechwood xylan. After 72-h incubation, crude enzymes were extracted from the cells and dialyzed against 20-mM sodium phosphate buffer (pH 8) using a 10K MWCO SnakeSkin™ dialysis tubing (Thermo Scientific, Waltham, USA). Unless specified, all enzyme assays were carried out at 50 °C, pH 8 for 15 min by incubating 0.1 mL of crude enzymes with 1 mL of substrate, and subsequently measured using a 7300 Vis spectrophotometer (Jenway, Staffordshire, UK) with the wavelength adjusted to 540 nm (for reducing sugar detection by DNS assay), or at 405 nm (for detection of p-nitrophenol released from the artificial substrates). Substrates tested included Avicel®, CMC, beechwood xylan, p-nitrophenyl-β-D-glucopyranoside (pNPG), p-nitrophenyl-β-D-xylopyranoside (pNPX), cello-dextrins (cellobiose to celloheptaose, C2–C7), and xylo-dextrins (xylobiose to xylohexaose, X2–X6). The post-reaction products were determined using an Agilent 1260 Infinity High-Performance Liquid Chromatography, coupled with an Agilent 385-Evaporative Light Scattering Detector (Agilent Technologies, Santa Clara, USA) and a Rezex RSO-Oligosaccharide Ag + column (Phenomenex Inc, Torrance, USA).

Results and discussion

Genome features of strain RA

As strain RA is most likely a new genus of the family *Rhodothermaceae*, we resequenced the genome to fill in the gaps, as well as to confirm the orientation or order of contigs present in the draft genome. The PacBio RSII sequencer was able to close the gaps found in the earlier draft genome. An additional 71 CDS were also identified

in the newly assembled genome. The complete genome of this bacterium has been deposited in GenBank under accession number CP020382.1. The circular chromosome of 4,653,222 bp (132x coverage) had a GC content of 68.3%, and based on NCBI PGAP (Fig. 1), the genome encoded 3,711 genes, which included 3,506 protein-coding sequences (CDS), 155 pseudogenes, 3 rRNAs, 44 tRNAs, and 3 ncRNAs. Moreover, a total of 1730 genes (46.6% of the total genes) from strain RA are annotated as hypothetical protein or uncharacterized protein due to their low sequence similarities to the existing database. A total of 3417 genes are annotated into different functional categories according to COG analysis (Table 1). Based on KEGG

(Entry number T04780), strain RA possessed all genes for most of the carbohydrate metabolism pathways.

Figure 2 illustrates the distribution of glycosyl hydrolases (GHs) in the genome of strain RA and other genera affiliated with family *Rhodothermaceae* which includes *Rhodothermus marinus* DSM 4252 (CP001807.1), *R. marinus* SG0.5JP17-171 (GCA_000565305.1), and *R. marinus* SG0.5JP17-172 (CP003029.1) that exhibit average 54 GH sequences placed in 31 GH families. Strain RA has a distribution of GH sequences similar to *R. marinus*, with 57 GHs that are affiliated to 30 GH families. The total number of GHs annotated in the genome of strain RA is higher than other genera of the family *Rhodothermaceae*. For instance, *Rhodothermus profundus* (GCA_900142415.1),

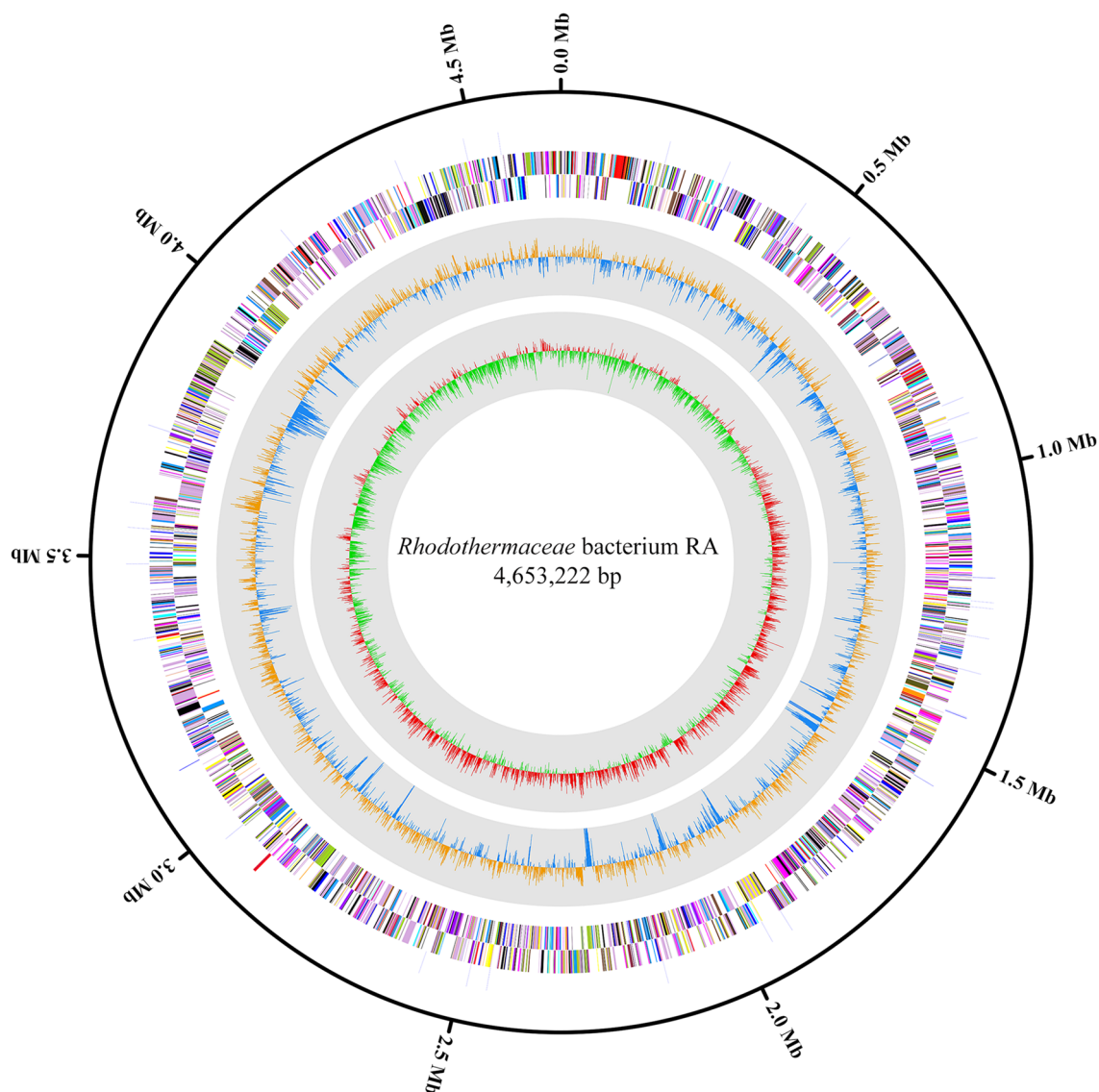


Fig. 1 Circular genome map of *Rhodothermaceae* bacterium RA. From the outermost circle to the center: RNA genes (rRNA, tRNA, and ncRNA), Reverse CDS, Forward CDS, GC skew, and GC ratio

Table 1 COG functional categories of *Rhodothermaceae* bacterium RA

COG functional categories	Count	Proportion (%)
Information storage and processing		
<i>J</i> —translation, ribosomal structure, and biogenesis	147	4.30
<i>A</i> —RNA processing and modification	1	0.03
<i>K</i> —transcription	124	3.63
<i>L</i> —replication, recombination, and repair	142	4.16
<i>B</i> —chromatin structure and dynamics	3	0.09
Cellular processes and signaling		
<i>D</i> —cell cycle control, cell division, and chromosome partitioning	25	0.73
<i>Y</i> —nuclear structure	0	0.00
<i>V</i> —defence mechanisms	43	1.26
<i>T</i> —signal transduction mechanisms	180	5.27
<i>M</i> —cell wall/membrane/envelope biogenesis	209	6.12
<i>N</i> —cell motility	38	1.11
<i>Z</i> —cytoskeleton	1	0.03
<i>W</i> —extracellular structures	1	0.03
<i>U</i> —intracellular trafficking, secretion, and vesicular transport	44	1.29
<i>O</i> —posttranslational modification, protein turnover, and chaperones	122	3.57
Metabolism		
<i>C</i> —energy production and conversion	148	4.33
<i>G</i> —carbohydrate transport and metabolism	183	5.36
<i>E</i> —amino acid transport and metabolism	237	6.94
<i>F</i> —nucleotide transport and metabolism	70	2.05
<i>H</i> —coenzyme transport and metabolism	92	2.69
<i>I</i> —lipid transport and metabolism	77	2.25
<i>P</i> —inorganic ion transport and metabolism	195	5.71
<i>Q</i> —secondary metabolites biosynthesis, transport, and catabolism	58	1.70
Poorly characterized		
<i>R</i> —general function prediction only	0	0.00
<i>S</i> —function unknown	1277	37.37

Longibacter salinarum (GCA_002554795.1), *Longimonas halophila* (GCA_002554705.1), and *Salisaeta longa* (GCA_000419585.1) have 29–35 GHs (Fig. 2). Most *Salinibacter* spp. (CP000159.1/FP565814.1/GCA_002894605.1/GCA_002894625.1/GCA_002894645.1) have around 21 sequences grouped into 17 different GH families, except for *Salinibacter* sp. 10B (GCA_002954405.1), which has 50 GHs across 22 families.

Several genes present in strain RA are annotated as GH enzymes related to cellulose and hemicellulose degradation (Table 2). These sequences include a GH2 β -galactosidase (NCBI locus tag: AWN76_014570), GH3 β -glucosidase (AWN76_006445), GH5 endoglucanase (AWN76_009395), GH9 endoglucanase (AWN76_010685), GH10 xylanase (AWN76_003690 and AWN76_008205), GH43 β -xylosidase (AWN76_012335), GH53 endo- β -1,4-galactanase (AWN76_017855), and GH92 α -mannosidase (AWN76_002955). Interestingly, these enzymes have low identities to other counterpart sequences available in the NCBI database (57–73% identity). In addition, these

sequences exhibited low similarity to *Rhodothermus* spp. counterparts, a clear indication of the novelty of enzymes from strain RA. A putative sequence (AWN76_009940) was annotated as glycoside hydrolase. The AWN76_009940 protein sequence consists of a typical GH16 domain as determined using InterProScan, and it is 75% identical to laminarinase (endo-1,3(4)- β -glucanase; PDB id: 3ILN_A) which originates from *R. marinus* (Bleicher et al. 2011). Another sequence (AWN76_008195) is annotated as a hypothetical protein but putatively functions as an endoglucanase associated with GH44. The protein sequence of AWN76_008195 is 56% identical to endoglucanase J of *Ruminiclostridium thermocellum* (Ahsan et al. 1996). Other than GHs, some of the genes from strain RA are also assigned to other CAZy families, including 60 glycosyl transferases (GTs), 4 polysaccharide lyases (PLs), 8 carbohydrate esterases (CEs), 16 carbohydrate-binding modules (CBMs), and 7 auxiliary activities (AA) affiliated enzymes. According to CAZy, AA consists of two groups of enzymes (ligninolytic enzymes and lytic polysaccharide mono-oxygenases), that are responsible

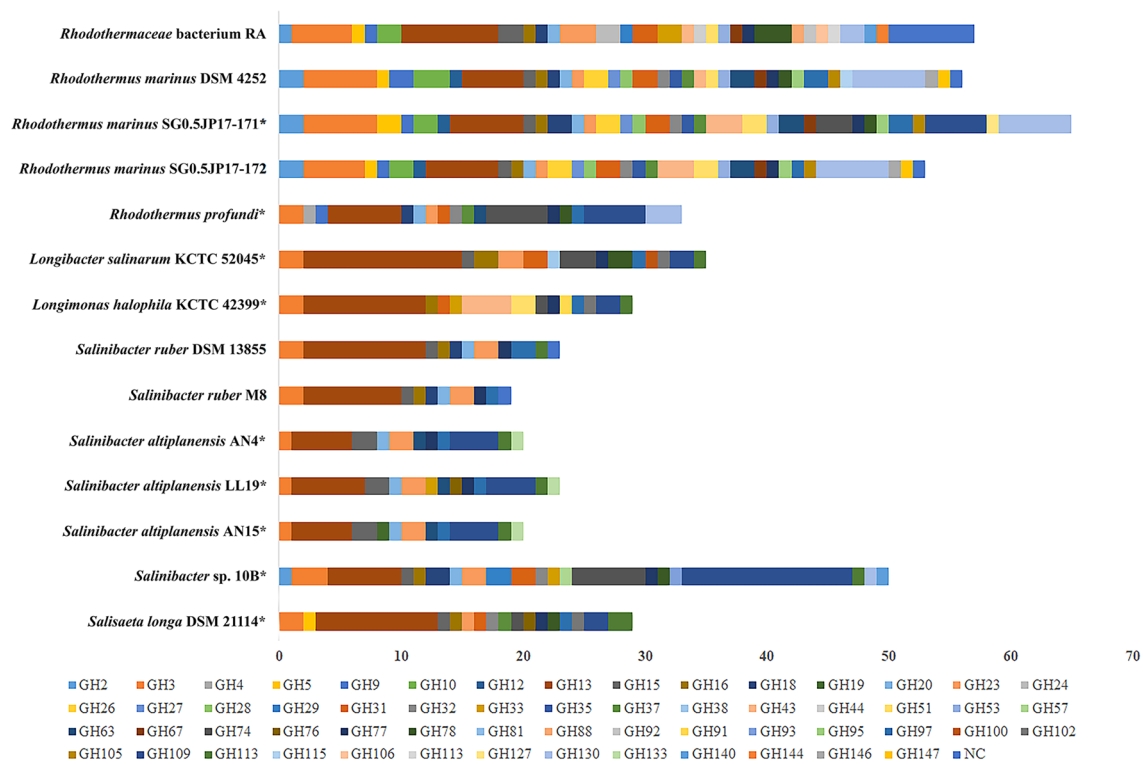


Fig. 2 Distribution and predicted numbers of GH in the genome *Rhodothermaceae* bacterium RA and other bacteria strains of the same genus. *indicates draft genome sequences

for lignin breakdown, as well as the hydrolysis of polysaccharide. Therefore, it is likely that these enzymes in strain RA may work cooperatively with GHs to efficiently degrade the lignocellulosic biomass.

Cellulolytic and xylanolytic potential of strain RA

Table 3 summarizes the results of both colorimetric assays and HPLC analysis. In brief, the cell-free crude enzymes of strain RA are active on the following substrates: CMC, beechwood xylan, pNPG, pNPX, cellodextrins (C2–C7), and xylodextrins (X2–X6). Under the current experimental setup, the crude enzymes of strain RA exhibit 0.41-U/mL endoglucanase, 0.02-U/mL β -glucosidase, 1.43-U/mL xylanase, and 0.17-U/mL β -xylosidase activities. The crude enzymes were not active against Avicel[®] suggesting the absence of exoglucanase activity, confirming the absence of such an enzyme from genome annotation. Similarly, exoglucanase gene is also absent from members of genus *Rhodothermus*, the closest relative of strain RA. Many of the distant thermophilic bacteria (*Thermotoga maritima*, *Dictyoglomus turgidum*, and *Thermomonospora curvata*) (Chertkov et al. 2011; Singh et al. 2015; Brumm et al. 2016), also lack of exoglucanase gene. Other thermophiles (*Ruminiclostridium thermocellum*, *Caldicellulosiruptor* spp., and

Thermobifida fusca) produce exoglucanase (Caspi et al. 2008; Blumer-Schuette et al. 2010; Sheng et al. 2016).

Crude enzymes were assayed with CMC and xylan and the end-products were analyzed using HPLC. The crude enzyme was found to hydrolyze CMC and C2–C7 to glucose, cellobiose, cellotriose, and cellotetraose, whereas the major degradation products from xylan and X2–X6 were xylose and xylobiose. In general, more total sugars were released from xylan than CMC (Table 3). Besides, cell-free crude enzymes efficiently hydrolyzed longer chains substrates (C5–C7, X3–X6), but the activities gradually dropped for shorter chains substrates (C2–C4, X2). It is hypothesized that under current experiment condition, supplementation of xylan and CMC to the growth culture favoured the expression of endoglucanase and xylanase activities, which are specific to longer chains substrates. Although the activities were detected for β -glucosidase and β -xylosidase, both enzymes were likely not over-expressed by the induction of xylan and CMC.

Comparison of strain RA enzyme activities with a list of other lignocellulolytic bacteria is summarized in Appendix A. *Cellvibrio mixtus*, *Jonesia denitrificans*, and *Gracilibacillus* sp. TSCPVG are outstanding xylanase producers (Giridhar and Chandra 2010; Nawel et al. 2011; Wu and He 2015). *Clostridium thermocellum* is an excellent bacterium for endoglucanase production (Mori 1992). At the current

Table 2 List of potential lignocellulolytic enzymes from *Rhodothermaceae* bacterium RA

CAZyme families	Annotation	locus_tag	RefSeq accession number	Closest Sequence	Identity (%)
GH2	β -glucosidase	AWN76_014570	ARA94254.1	glycoside hydrolase of <i>Gemmatimonadetes</i> bacterium	65
GH3	β -glucosidase	AWN76_006445	ARA95045.1	β -glucosidase BglX of <i>Rhodothermus marinus</i>	69
GH5	Endoglucanase	AWN76_009395	ARA93352.1	glycoside hydrolase of <i>Pedobacter</i> sp. V48	57
GH9	Endoglucanase	AWN76_010685	ARA95103.1	glycoside hydrolase family 9 of <i>Gemmatimonas</i> sp	73
GH10	Xylanase	AWN76_003690	ARA92359.1	glycoside hydrolase of <i>Rhodothermus marinus</i>	73
GH10	Xylanase	AWN76_008205	ARA95075.1	endo-1,4-beta-xylanase of <i>Candidatus Solibacter usitatus</i>	62
GH43	β -xylosidase	AWN76_012335	ARA93868.1	glycoside hydrolase of <i>Parapedobacter composti</i>	60
GH53	endo- β -1,4-galactanase	AWN76_017855	ARA94834.1	arabinogalactan endo-1,4- β -galactosidase of <i>Rhodothermus marinus</i>	60
GH92	α -mannosidase	AWN76_002955	ARA92237.1	α -mannosidase of <i>Spirosoma</i> sp. 209	58
GH16	Glycosyl hydrolase	AWN76_009940	ARA93444.1	laminarinase of <i>Rhodothermus marinus</i>	75
GH44	Hypothetical protein	AWN76_008195	ARA93141.1	endoglucanase J of <i>Ruminiclostridium thermocellum</i>	56
AA2	Catalase/peroxidase HPI	AWN76_014060	ARA94166.1	Catalase/peroxidase HPI of <i>alpha Proteobacterium</i>	74
AA3	GMC family oxidoreductase	AWN76_001955	ARA92052.1	GMC family oxidoreductase of <i>Rhodothermus marinus</i>	70
AA3	GMC family oxidoreductase	AWN76_003120	ARA92263.1	GMC family oxidoreductase of <i>Rhodothermus marinus</i>	70
AA3	Patatin	AWN76_007050	ARA92944.1	Patatin-like phospholipase family protein of <i>Catalinimonas alkaloidigena</i>	61
AA3	GMC family oxidoreductase	AWN76_011750	ARA93772.1	GMC family oxidoreductase of <i>Rhodothermus marinus</i>	67
AA12	Sorbosone dehydrogenase	AWN76_005825	ARA92731.1	Sorbosone dehydrogenase of <i>Rhodothermus marinus</i>	64
AA12	Sorbosone dehydrogenase	AWN76_011490	ARA95111.1	Sorbosone dehydrogenase of <i>Phormidesmis priestleyi</i>	64

time, data on lignocellulose saccharification studies using members of the family *Rhodothermaceae* are limited. In addition, currently, the ability to degrade cellulose and hemicellulose is known to be restricted to members of the genus *Rhodothermus*. According to a study by Dahlberg et al. (1993), *R. marinus* DSM 4252 exhibited <0.03-U/mL endoglucanase, 1.98-U/mL β -glucosidase, 1.14-U/mL xylanase, and 4.08-U/mL β -xylosidase activities after 24-h growth in modified M162 medium supplemented with 0.5% (w/v) xylan in a 2.5-L bioreactor (note: unit conversion from nkat/mL to U/mL). Moreover, other reports related to *Rhodothermus* spp. enzymes such as endoglucanase, xylanase, endo-1,4- β -mannosidase, and

α -L-arabinofuranosidase have also been reported (Karlsson et al. 1997; Halldórsdóttir et al. 1998; Politz et al. 2000; Gomes et al. 2000). All these enzymes have been purified, characterized, and were reported to be active and thermostable. Altogether, *Rhodothermus* is known as an interesting genus for lignocellulose degradation. In conclusion, the current work reports the improved genome sequence of strain RA. In addition to members of genus *Rhodothermus*, strain RA is yet another excellent candidate in the family *Rhodothermaceae* which possess a repertoire of novel and thermostable cellulolytic and hemicellulolytic enzymes.

Table 3 Hydrolysis of various substrates by *Rhodothermaceae* bacterium RA crude enzymes

Substrate	Spectrophotometric analysis ^{a,b} (Unit/mL)	HPLC analysis						
		Substrate depletion ^c (%)	Product formation (µg/mL)					
			Glucose	C2	C3	C4	Xylose	X2
Avicel	0.000 ± 0.000 ^a	–	0	0	0	0	–	–
CMC	0.411 ± 0.011 ^a	–	82	39	18	0	–	–
Xylan	1.428 ± 0.007 ^a	–	–	–	–	–	1397	41
PNPG	0.019 ± 0.001 ^b	–	–	–	–	–	–	–
PNPX	0.173 ± 0.001 ^b	–	–	–	–	–	–	–
C2	–	16.86	318	–	–	–	–	–
C3	–	30.06	314	810	–	–	–	–
C4	–	99.28	405	4928	464	–	–	–
C5	–	100.00	430	3042	3909	0	–	–
C6	–	100.00	359	3354	2185	229	–	–
C7	–	100.00	306	1230	494	58	–	–
X2	–	95.06	–	–	–	–	4233	–
X3	–	100.00	–	–	–	–	3958	419
X4	–	100.00	–	–	–	–	3786	368
X5	–	100.00	–	–	–	–	3392	240
X6	–	100.00	–	–	–	–	3276	193

– indicates not available. C2–C7 indicate cellobiose to celloheptaose, respectively. X2–X6 indicate xylobiose to xylohexaose, respectively

^aReading taken at wavelength 540 nm (DNS assay). One unit (U) of enzyme activity was defined as the enzyme amount that can liberate 1 µmol of reducing sugar per min per mL under assay condition

^bReading taken at wavelength 405 nm. One unit (U) of enzyme activity was defined as the enzyme amount that can liberate 1 µmol of p-nitrophenol per min per mL under assay condition

^cCalculated using formula: $\frac{\text{Initial amount of substrate} - \text{amount of substrate}}{\text{Initial amount of substrate}} \times 100\%$

Acknowledgements This work was supported by Universiti Teknologi Malaysia Research Grants (Grant Nos. 16H89 & 14H67) awarded to MS Shamsir and KM Goh. CS Chong is thankful to Newton Ungku Omar Fund 4B297. KJ Liew thanks the Zamalah Program of Universiti Teknologi Malaysia for scholarship support. K-G Chan gratefully acknowledges the financial support provided by grants from JBK, NRE, and PPP (GA001-2016, GA002-2016, PG226-2016).

Compliance with ethical standards

Conflict of interest The authors declare that there is no conflict of interest regarding the publication of this article.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

Ahsan MM, Kimura T, Karita S, Sakka K, Ohmiya K (1996) Cloning, DNA sequencing, and expression of the gene encoding *Clostridium thermocellum* cellulase CelJ, the largest catalytic component of the cellulosome. J Bacteriol 178:5732–5740

- Alfredsson GA, Kristjansson JK, Hjörleifsdóttir S, Stetter KO (1988) *Rhodothermus marinus*, gen. nov., sp. nov., a thermophilic, halophilic bacterium from submarine hot springs in Iceland. Microbiology 134:299–306
- Antón J, Oren A, Benlloch S, Rodríguez-Valera F, Amann R, Rosselló-Mora R (2002) *Salinibacter ruber* gen. nov., sp. nov., a novel, extremely halophilic member of the Bacteria from saltern crystallizer ponds. Int J Syst Evol Microbiol 52:485–491
- Bleicher L, Prates ET, Gomes TC, Silveira RL, Nascimento AS, Rojas AL, Golubev A, Martínez L, Skaf MS, Polikarpov I (2011) Molecular basis of the thermostability and thermophilicity of laminarinases: X-ray structure of the hyperthermostable laminarinase from *Rhodothermus marinus* and molecular dynamics simulations. J Phys Chem B 115:7940–7949
- Blumer-Schuette SE, Lewis DL, Kelly RM (2010) Phylogenetic, microbiological, and glycoside hydrolase diversities within the extremely thermophilic, plant biomass-degrading genus *Caldicellulosiruptor*. Appl Environ Microbiol 76:8084–8092
- Bohra V, Dafale NA, Purohit HJ (2018) *Paenibacillus polymyxa* ND25: candidate genome for lignocellulosic biomass utilization. 3 Biotech 8:248
- Brumm PJ, Gowda K, Robb FT, Mead DA (2016) The complete genome sequence of hyperthermophile *Dictyoglomus turgidum* DSM 6724TM reveals a specialized carbohydrate fermentor. Front Microbiol 7:1979
- Caspi J, Irwin D, Lamed R, Li Y, Fierobe H-P, Wilson DB, Bayer EA (2008) Conversion of *Thermobifida fusca* free exoglucanases into cellulosomal components: comparative impact on cellulose-degrading activity. J Biotechnol 135:351–357

- Chan CS, Chan K-G, Ee R, Hong K-W, Urbieta MS, Donati ER, Shamsir MS, Goh KM (2017) Effects of physiochemical factors on prokaryotic biodiversity in Malaysian circumneutral hot springs. *Front Microbiol* 8:1252
- Chertkov O, Sikorski J, Nolan M, Lapidus A, Lucas S, Del Rio TG, Tice H, Cheng J-F, Goodwin L, Pitluck S (2011) Complete genome sequence of *Thermomonospora curvata* type strain (B9 T). *Stand Genomic Sci* 4:13–22
- Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE (2013) Non-hybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10:563–569
- Dahlberg L, Holst O, Kristjansson JK (1993) Thermostable xylanolytic enzymes from *Rhodothermus marinus* grown on xylan. *Appl Microbiol Biotechnol* 40:63–68
- Giridhar PV, Chandra T (2010) Production of novel halo-alkali-thermostable xylanase by a newly isolated moderately halophilic and alkali-tolerant *Gracilicoccus* sp. *TSCPVG Process Biochem* 45:1730–1737
- Goh KM, Chan K-G, Lim SW, Liew KJ, Chan CS, Shamsir MS, Ee R, Adrian T-G-S (2016) Genome analysis of a new *Rhodothermaceae* strain isolated from a hot spring. *Front Microbiol* 7:1109
- Gomes J, Gomes I, Terler K, Gubala N, Ditzelmüller G, Steiner W (2000) Optimisation of culture medium and conditions for α -L-arabinofuranosidase production by the extreme thermophilic eubacterium *Rhodothermus marinus*. *Enzyme Microb Technol* 27:414–422
- Halldórsdóttir S, Thóroflsdóttir E, Spilliaert R, Johansson M, Thorbjarnardóttir S, Palsdóttir A, Hreggvidsson G, Kristjansson J, Holst O, Eggertsson G (1998) Cloning, sequencing and overexpression of a *Rhodothermus marinus* gene encoding a thermostable cellulase of glycosyl hydrolase family 12. *Appl Microbiol Biotechnol* 49:277–284
- Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28:27–30
- Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* 45:D353–D361
- Karlsson EN, Bartonek-Roxå E, Holst O (1997) Cloning and sequence of a thermostable multidomain xylanase from the bacterium *Rhodothermus marinus*. *Biochim Biophys Acta, Gene Struct Expr* 1353:118–124
- Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2013) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* 42:D490–D495
- Makhdoumi-Kakhki A, Amoozegar MA, Ventosa A (2012) *Salinibacter iranicus* sp. nov. and *Salinibacter luteus* sp. nov., isolated from a salt lake, and emended descriptions of the genus *Salinibacter* and of *Salinibacter ruber*. *Int J Syst Evol Microbiol* 62:1521–1527
- Marteinsson VT, Bjornsdóttir SH, Bienvenu N, Kristjansson JK, Birrien J-L (2010) *Rhodothermus profundus* sp. nov., a thermophilic bacterium isolated from a deep-sea hydrothermal vent in the Pacific Ocean. *Int J Syst Evol Microbiol* 60:2729–2734
- Mori Y (1992) Comparison of the cellulolytic systems of *Clostridium thermocellum* YM4 and JW20. *Biotechnol Lett* 14:131–136
- Munoz R, Rosselló-Móra R, Amann R (2016) Revised phylogeny of Bacteroidetes and proposal of sixteen new taxa and two new combinations including *Rhodothermaeota* phyl. nov. *Syst Appl Microbiol* 39:281–296
- Nawel B, Said B, Estelle C, Hakim H, Duchiron F (2011) Production and partial characterization of xylanase produced by *Jonesia denitrificans* isolated in Algerian soil. *Process Biochem* 46:519–525
- Park S, Yoshizawa S, Kogure K, Yokota A (2011) *Rubricoccus marinus* gen. nov., sp. nov., of the family ‘*Rhodothermaceae*’, isolated from seawater. *Int J Syst Evol Microbiol* 61:2069–2072
- Park S, Song J, Yoshizawa S, Choi A, Cho J-C, Kogure K (2013) *Rubrivirga marina* gen. nov., sp. nov., a member of the family *Rhodothermaceae* isolated from deep seawater. *Int J Syst Evol Microbiol* 63:2229–2233
- Park S, Akira Y, Kogure K (2014) The family *Rhodothermaceae*. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F (eds) *The prokaryotes*. Springer, Berlin, pp 849–856
- Politz O, Krah M, Thomsen K, Borriss R (2000) A highly thermostable endo-(1, 4)- β -mannanase from the marine bacterium *Rhodothermus marinus*. *Appl Microbiol Biotechnol* 53:715–721
- Sheng T, Zhao L, Gao L-F, Liu W-Z, Cui M-H, Guo Z-C, Ma X-D, Ho S-H, Wang A-J (2016) Lignocellulosic saccharification by a newly isolated bacterium, *Ruminiclostridium thermocellum* M3 and cellular cellulase activities for high ratio of glucose to cellobiose. *Biotechnol Biofuels* 9:172
- Singh R, Gradnigo J, White D, Lipzen A, Martin J, Schackwitz W, Moriyama E, Blum P (2015) Complete genome sequence of an evolved *Thermotoga maritima* isolate. *Genome Announc* 3:e00557–e00515
- Sorokin DY, Khijniak TV, Galinski EA, Kublanov IV (2017) *Natronotalea proteiniolytica* gen. nov., sp. nov. and *Longimonas haloalkaliphila* sp. nov., extremely haloalkaliphilic members of the phylum *Rhodothermaeota* from hypersaline alkaline lakes. *Int J Syst Evol Microbiol* 67:4161–4167
- Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN (2003) The COG database: an updated version includes eukaryotes. *BMC Bioinform* 4:41
- Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J (2016) NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res* 44:6614–6624
- Vaisman N, Oren A (2009) *Salisaeta longa* gen. nov., sp. nov., a red, halophilic member of the *Bacteroidetes*. *Int J Syst Evol Microbiol* 59:2571–2574
- Wu Y-R, He J (2015) Characterization of a xylanase-producing *Cellvibrio mixtus* strain J3-8 and its genome analysis. *Sci Rep* 5:10521
- Xia J, Zhou Y-X, Zhao L-H, Chen G-J, Du Z-J (2015) *Longimonas halophila* gen. nov., sp. nov., isolated from a marine solar saltern. *Int J Syst Evol Microbiol* 65:2272–2276
- Xia J, Dunlap CA, Flor-Weiler L, Rooney AP, Chen G-J, Du Z-J (2016) *Longibacter salinarum* gen. nov., sp. nov., isolated from a marine solar saltern. *Int J Syst Evol Microbiol* 66:3287–3292
- Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y (2012) dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* 40:W445–W451