3 Biotech (2023) 13:50
https://doi.org/10.1007/s13205-023-03472-8

GENOME REPORTS

Draft genome sequence of *Joostella atrarenae* M1-2^T with cellulolytic and hemicellulolytic ability

Kok Jun Liew[1] · Muhammad Ramziuddin Zakaria[1] · Clarine Wan Ling Hong[1] · Melvin Chun Yun Tan[1] · Chun Shiong Chong[1]

Received: 12 July 2022 / Accepted: 6 January 2023 / Published online: 17 January 2023

Abstract

The halophilic genus *Joostella* is one of the least-studied genera in the family of *Flavobacteriaceae*. So far, only two species were taxonomically identified with limited genomic analysis in the aspect of application has been reported. *Joostella atrarenae* M1-2^T was previously isolated from a seashore sample and it is the second discovered species of the genus *Joostella*. In this project, the genome of *J. atrarenae* M1-2^T was sequenced using NovaSeq 6000. The final assembled genome is comprised of 71 contigs, a total of 3,983,942 bp, a GC ratio of 33.2%, and encoded for 3,416 genes. The 16S rRNA gene sequence of *J. atrarenae* M1-2^T shows 97.3% similarity against *J. marina* DSM 19592^T. Genome-genome comparison between the two strains by ANI, dDDH, AAI, and POCP shows values of 80.8%, 23.3%, 83.4%, and 74.1% respectively. Pan-genome analysis shows that strain M1-2^T and *J. marina* DSM 19592^T shared a total of 248 core genes. Taken together, strain M-2^T and *J. marina* DSM 19592^T belong to the same genus but are two different species. CAZymes analysis revealed that strain M1-2^T harbors 109 GHs, 40 GTs, 5 PLs, 9 CEs, and 6 AAs. Among these CAZymes, while 5 genes are related to cellulose degradation, 12 and 24 genes are found to encode for xylanolytic enzymes and other hemicellulases that involve majorly in the side chain removal of the lignocellulose structure, respectively. Furthermore, both the intracellular and extracellular crude extracts of strain M1-2^T exhibited enzymatic activities against CMC, xylan, *p*NPG, and *p*NPX substrates, which corresponding to endoglucanase, xylanase, β-glucosidase, and β-xylosidase, respectively. Collectively, description of genome coupled with the enzyme assay results demonstrated that *J. atrarenae* M1-2^T has a role in lignocellulosic biomass degradation, and the strain could be useful for lignocellulosic biorefining.

Keywords Halophile · *Joostella* · *Flavobacteriaceae* · Cellulase · Hemicellulase

Marine microorganisms remain a great source for the development of different biotechnological applications. Due to the huge pool of untapped species in the sea, it has attracted the attention of many scientists to discover novel antibiotics, drugs, biosurfactant, carotenoids and enzymes from the marine microorganisms (Selvaratnam et al. 2016; Thevarajoo et al. 2016; Wang et al. 2016). Discovery and studies of cellulases, hemicellulases, and lignin-degrading enzymes had long been focused on those terrestrially isolated microorganisms (Sethupathy et al. 2021). For example, fungi like *Aspergillus* spp., and *Trichoderma* spp., or bacteria such as *Caldicellulosiruptor* spp., and *Cellulomonas* spp. are commonly used in lignocellulolytic enzymes cocktail production (Rai et al. 2019; Chukwuma et al. 2020; Sethupathy et al. 2021). Mining of the lignocellulolytic genes from marine microorganisms seems to be a good breakthrough for industrial enzymes development. This is because special environmental conditions in the sea, such as low temperature, low light, high salinity, and high pressure, had caused the marine residents to evolve, and this corresponding to the production of enzymes with unique and diverse characteristics (Wang et al. 2016).

Genus *Joostella* was proposed in 2008 and it is classified under the family *Flavobacteriaceae* (Quan et al. 2008). Its members are Gram-negative, non-spore-forming, non-motile, yellow-pigmented, and are mostly isolated from marine samples (Quan et al. 2008; Kim et al. 2011). This genus only consists of two identified species up-to-date, which are *J. marina* and *J. atrarenae* (Quan et al. 2008; Kim

✉ Chun Shiong Chong
cschong@utm.my

[1] Department of Biosciences, Faculty of Science, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia

et al. 2011). Based on NCBI taxonomy database search, up to 21 strains of the genus are currently unclassified. Except for the two identified species, which had been subjected to a polyphasic taxonomy (phenotypic and chemotaxonomic) study, other members are only verified via 16S rRNA phylogenetic analyses and are not deposited in any culture collection yet. At the time of writing, only genome of *J. marina* DSM 19592^T was reported without detailed analysis on its application aspects (Stackebrandt et al. 2013). Collectively, due to this lack of study and limited information, *Joostella* is considered an underexplored genus. Current project aimed to sequence *J. atrarenae* M1-2^T genome and explore the potential of this strain in lignocellulose degradation.

J. atrarenae M1-2^T (KCTC 23194^T) was purchased from the Korean Collection of Type Cultures. The bacterium was grown in 10 mL Marine Broth (BD, USA) for 3 days at 30 °C with 200 rpm shaking. Total DNA was extracted from the culture using ZR Quick-DNA Miniprep plus kit (Zymo Research, USA). A NEBNext® Ultra™ DNA Library Prep Kit for Illumina® (New England Biolabs, USA) was used to construct the total DNA into sequencing library. Whole genome sequencing was conducted in an Illumina® NovaSeq 6000 with running mode of PE150 (paired-end 150 bp). A total of 2.81 billion bases in 9,374,372 raw paired-end reads were obtained from the genome sequencing. The raw sequence data from the sequencing was deposited in NCBI Sequence Read Archive (SRA) database with accession number SRR15341968.

The raw paired-end reads were trimmed and quality filtered using soapNuke v2.1.6 (Chen et al. 2018). De novo assembly of the clean paired-end reads was carried out using SPAdes v3.13.0 (Prjibelski et al. 2020), which resulted in total genome size of 3,983,942 bp, with 350× coverage, 71 contigs, N50 of 267,267 bp, and GC content of 33.2%. Gene prediction and annotation of the assembled genome was conducted using NCBI Prokaryotic Genome Annotation Pipeline (PGAP) v4.12 (Tatusova et al. 2016). A total of 3,416 genes were encoded by the genome, including 3,343 protein-coding sequences, 26 pseudogenes, 3 rRNAs, 40 tRNAs, and 4 ncRNAs. Table 1 illustrates the minimum information of genome sequence (MIGS) of strain M1-2^T.

For identification, 16S rRNA gene sequence was extracted from the draft genome and compared with the reported partial sequence (accession number: GQ872420), which achieved 100% similarity, and this confirmed that this draft genome belongs to strain M1-2^T. Then, the 16S rRNA gene sequence was also searched against EzTaxon database (Yoon et al. 2017), which resulted in 97.3%

Table 1 Minimum information of genome sequence (MIGS) of *J. atrarenae* M1-2^T

Investigation type	Bacteria
BioProject accession	PRJNA694488
BioSample accession	SAMN17525809
Taxonomy ID of DNA sample	679257
Geographical location	Black sand from seashore of Jeju Island, South Korea
Longitude and latitude	33°14′39″N, 126°34′29″E
Collection date	2009
Environment	Sea sand [ENVO: 00002118]
Sequencing method	Illumina® NovaSeq 6000
Number of replicons	1
Isolation and growth condition	Marine agar or broth 2216, optimum growth conditions at 30 °C, pH 8.0–8.5, and 1–4% (*w/v*) NaCl
Reference for biomaterial	PMID: 20820783; https://doi.org/10.1007/s00284-010-9750-y
Assembly software	SPAdes; v3.13.0; default parameters
Assembly quality	High-quality draft genome
Number of contigs	71
Coverage	350×
Other genome features	
Genome size	3,983,942 bp
GC content	33.2%
Protein-coding sequences	3,343
Pseudogenes	26
tRNAs	40
rRNAs	3
ncRNAs	4
Genome accession number	JAETXX000000000

similarity against *Joostella marina* DSM 19592^T (16S rRNA accession: JH651379). Besides, Average Nucleotide Identity (ANI) was conducted using FastANI v1.33 (Jain et al. 2018), and a 80.8% ANI value was found between strain M1-2^T and *J. marina* DSM 19592^T (genome accession: GCA_000260115.1). Digital DNA-DNA hybridization (dDDH) by GGDC 3.0 shows 23.3% between genomes of strain M1-2^T and *J. marina* DSM 19592^T (Meier-Kolthoff et al. 2022). In term of amino acid similarities, Average Amino Acid Identities (AAI) by EzAAI was found to be 83.4% (Kim et al. 2021), and Percentage of Conserved Proteins (POCP) was found to be 74.1% between strain M1-2^T and *J. marina* DSM 19592^T (Qin et al. 2014). Pan-genome analysis via Roary v3.13.0 shows that both strains share a total of 248 core genes (Page et al. 2015). Collectively, the findings showed that strain M1-2^T and *J. marina* DSM 19592^T are two different species in genus *Joostella*.

EggNOG-mapper v2.1.4 was used for Cluster of Orthologous Groups of protein (COGs) analysis (Cantalapiedra et al. 2021). It successfully annotated a total of 3,106 protein-coding sequences into different COGs (Table 2). Due to the low similarities of strain M1-2^T protein sequences against the existing COGs database, many of the proteins (23.0%) were classified as Function Unknown, which indicating that the member of the genus *Joostella* is underexplored, causing their encoding genes or proteins are not well-documented. Nevertheless, around 7.1% of the total proteins are involved in carbohydrate transport and metabolism, this is an important information for the subsequent analysis of Carbohydrate-Active Enzymes (CAZymes).

To analyze carbohydrate-active enzymes (CAZymes), the genes encoding CAZymes in the genome were classified via run_dbCAN v2.0.11 (Zhang et al. 2018), and the annotated CAZymes genes were cross-checked using Diamond v2.0.11.149 Blastp search against NCBI nonreduntant and Uniprot Swiss-Prot databases (Buchfink et al. 2021); sequences from genus *Joostella* were excluded from the search. CAZymes of interest were further analyzed using InterProScan v5.52–86.0, SignalP

Table 2 Cluster of Orthologous Group of proteins (COGs) of *J. atrarenae* M1-2^T

COG	Functional description	Count	Proportion (%)
Cellular processing and signalling			
D	Cell cycle control, cell division, chromosome partitioning	29	0.9
M	Cell wall/membrane/envelope biogenesis	289	9.3
N	Cell motility	27	0.9
O	Post-translational modification, protein turnover, and chaperones	122	3.9
T	Signal transduction mechanisms	120	3.9
U	Intracellular trafficking, secretion, and vesicular transport	76	2.4
V	Defense mechanisms	57	1.8
W	Extracellular structures	4	0.1
Y	Nuclear structure	0	0
Z	Cytoskeleton	5	0.2
Information storage and processing			
A	RNA processing and modification	0	0.0
B	Chromatin structure and dynamics	0	0.0
J	Translation, ribosomal structure and biogenesis	166	5.3
K	Transcription	186	6.0
L	Replication, recombination and repair	137	4.4
Metabolism			
C	Energy production and conversion	151	4.9
E	Amino acid transport and metabolism	238	7.7
F	Nucleotide transport and metabolism	72	2.3
G	Carbohydrate transport and metabolism	220	7.1
H	Coenzyme transport and metabolism	137	4.4
I	Lipid transport and metabolism	103	3.3
P	Inorganic ion transport and metabolism	205	6.6
Q	Secondary metabolites biosynthesis, transport, and catabolism	48	1.5
Poorly characterized			
R	General function prediction only	0	0.0
S	Function unknown	714	23.0

v5.0, and TMHMM v2.0 to predict their domain, motifs, and other structures (Möller et al. 2001; Jones et al. 2014; Armenteros et al. 2019).

As a result, a total of 169 proteins were confirmed as CAZymes encoded by strain M1-2^T genome, which include 109 glycoside hydrolases (GHs), 40 glycosyltransferases (GTs), 5 polysaccharide lyases (PLs), 9 carbohydrate esterases (CEs), and 6 auxiliary activities (AAs) enzymes. Moreover, 26 carbohydrate-binding modules (CBMs), which have no catalytic function but are responsible for enzyme binding or cell adhesion to carbohydrate polymers, were also found in the genome. To identify the genes that are involved in lignocellulose degradation, these genes of interest were further refined from the CAZymes as shown above. The genes encode for lignocellulolytic enzymes by strain M1-2^T were summarized in Table 3. Out of the 40 lignocellulolytic enzymes identified in strain M1-2^T, 30 were found to be homologous to the sequences in other reported *Joostella* spp., with percentage of identity ranging from 24.8%–90.6%. Although these sequences can be found in *Joostella* spp., none of them was further studied. Thus, this shows the importance and novelty of studying these lignocellulolytic enzymes. As compared to other species, the lignocellulolytic enzymes in strain M1-2^T had percentage of identity ranging from 32.7%–95.8%.

The bacterium harbors 5 genes that could be responsible for cellulose degradation, including JM658_02870, JM658_07570, JM658_07655, JM658_10395, and JM658_10405, and these genes encode for either β-glucosidases or cellulases that classified in GH3, 5, and 144. A total of 12 genes involves in xylan degradation (the major composition of hemicellulose), which are JM658_00590, JM658_00820, JM658_02860, JM658_07595, JM658_07630, JM658_07640, JM658_07650, JM658_07705, JM658_07745, JM658_09135, JM658_09145 and JM658_09160. These genes can be translated into various enzymes such as bifunctional xylanase/deacetylase, β-xylanase, β-xylosidase, xylosidase, and bifunctional xylosidase/arabinosidase, that are classified under CE4, GH3, 5, 39, 43, 120, and some are non-classified glycoside hydrolases (GH0). Moreover, there are 24 genes that are responsible for other hemicellulose degradation including side chain removal. For instance, JM658_11595, JM658_12360, and JM658_15590, JM658_15640 (all encode for GH2 β-galactosidases) are responsible for galactan degradation. Interestingly, strain M1-2^T consists of 13 α-L-fucosidase encoded genes, of which 2 of them belong to GH95 and 11 are GH29. These α-L-fucosidases are responsible in the side chain removal of hemicellulose that is made up of L-fucose sugar monomer, which is commonly found in xyloglucan polysaccharide (Zhou et al. 2017; Cardoso et al. 2020). Collectively, strain M1-2^T is most probably effective in degrading cellulose and hemicellulose (xylan and xyloglucan) (Fig. 1). The strain M1-2^T could degrade cellulose and xylan completely as it encodes for full set of enzymes for degradation. Whereas for xyloglucan, the bacterium is missing an important enzyme that cleaves the α-1,6-glycosidic bond that is formed between D-glucose and D-xylose. Thus, the bacterium can only degrade xyloglucan partially, especially in the side chain removal of L-fucose and D-galactose (Fig. 1).

Strain M1-2^T was previously isolated from a black sand sample located at a seashore in Jeju Island, South Korea (Kim et al. 2011). The region is rich in seaweeds and microalgae growth, thanks to the large submarine groundwater discharge at the coastal regions that with a lot of organic matters and nutrients (Hwang et al. 2005; Song et al. 2018). Compared to the terrestrial plants, the cell walls of most seaweed and microalgae are lacking in lignin structure, and are composed mainly of carbohydrate polymers like cellulose, hemicellulose, laminarin, fucoidan, ulvan, etc. (Khan et al. 2018). Therefore, this in turn leads to the growth of cellulolytic and hemicellulolytic microorganisms. In general, seaweed and microalgae involve in carbon sequestration process, where they capture and store atmospheric carbon dioxides through photosynthesis. Then, the microorganisms carry out respiration and degradation of the cellulolytic compounds, and these release carbon dioxides back into the atmosphere. Collectively, this provides a possible way to involve in marine carbon recycling. Through the analysis of genome earlier, strain M1-2^T encodes for different types of cellulolytic and hemicellulolytic enzymes, this enables the strain to contribute to the carbon cycling as discussed above.

The ability of *J. atrarenae* M1-2^T to utilize different types of polysaccharides was screened via hydrolytic plate assays. As a result, strain M1-2^T was able to utilize carboxylmethyl cellulose (CMC), beechwood xylan, locust bean gum (LBG), and arabinan, forming halo zone in the agar plate supplemented with the respective substrates after 5 days of incubation (Fig. 2a). Besides, the releases of reducing sugars in suspension cultures supplemented with different substrates (1% *w/v*) were also inspected (Fig. 2b). As a negative control, culture without addition of substrates was performed and no reducing sugar detected throughout the 5 days of incubation. CMC and Avicel, which represent the cellulose, were hydrolysed into glucose and its oligomers by strain M1-2^T, with the highest reducing sugar concentration (0.200 ± 0.062 mM and 0.160 ± 0.056 mM) detected at day 1. Similarly, strain M1-2^T was able to hydrolyse the xylan into xylose and its derivatives with the highest concentration (0.673 ± 0.004 mM) found at day 1 of incubation. For LBG, which composed mainly of mannan, was hydrolysed into mannose monomer and the highest concentration (0.240 ± 0.053 mM) can be found at day 5. Lastly, the culture supplemented with arabinan had the least amount of reducing sugar detected throughout the 5 days incubation.

Table 3 List of cellulolytic and hemicellulolytic enzymes encode by *J. atrarenae* M1-2 ^{T}

No	Locus_Tag	Family	Annotation	Closest Sequence to *Joostella* sp. % Identity	Closest Sequence to other species / % Identity	CBM[a]	SP[b]	Other Secretion[c]	TMHMM[d]	Loc.[e]
Cellulose-degrading enzymes										
1	JM658_02870	GH3	Beta-glucosidase	WP_008613904.1 [*Joostella marina*] 34.0%	RWW99618.1 [*Flavobacterium cerinum*] 75.1%	O	P	O	O	Extra
2	JM658_07570	GH5	Cellulase	–	MBN2351537.1 [Spirochaetales bacterium] 39.9%	O	P	O	O	Extra
3	JM658_07655	GH5	Cellulase	–	WP_068216333.1 [*Lewinella* sp. 4G2] 42.3%	O	P	O	O	Extra
4	JM658_10395	GH144	Beta-glucosidase	WP_008613900.1 [*Joostella marina*] 84.1%	WP_202026894.1 [*Galbibacter mesophilus*] 81.1%	O	P	O	O	Extra
5	JM658_10405	GH3	Beta-glucosidase	WP_008613904.1 [*Joostella marina*] 88.7%	WP_178952300.1 [*Galbibacter* sp. BG1] 86.4%	P	P	O	O	Extra
Xylan-degrading enzymes										
6	JM658_00590	GH43	Beta-xylosidase	WP_008611881.1 [*Joostella marina*] 90.6%	WP_202027825.1 [*Galbibacter mesophilus*] 85.3%	P	P	O	O	Extra
7	JM658_00820	CE4	Xylanase/deacetylase	WP_008611816.1 [*Joostella marina*] 70.3%	WP_178950007.1 [*Galbibacter* sp. BG1] 64.2%	O	O	O	O	Intra
8	JM658_02860	GH5	Beta-xylosidase	–	WP_202029153.1 [*Galbibacter mesophilus*] 78.3%	P	P	O	O	Extra
9	JM658_07595	GH5	Beta-xylanase	–	WP_008658681.1 [*Rhodopirellula europaea*] 44.2%	P	O	P	O	Extra
10	JM658_07630	GH120	Beta-xylosidase	–	WP_117880452.1 [*Lutibacter* sp. SM1352] 61.2%	P	P	O	O	Extra
11	JM658_07640	GH3	Beta-xylosidase	WP_008611324.1 [*Joostella marina*] 39.9%	MBN1996334.1 [candidate division KSB1 bacterium] 47.3%	P	P	O	O	Extra
12	JM658_07650	GH39	Beta-xylosidase	–	WP_077845495.1 [*Clostridium puniceum*] 52.8%	O	P	O	P	Ecto

Table 3 (continued)

No	Locus_Tag	Family	Annotation	Closest Sequence to *Joostella* sp. % Identity	Closest Sequence to other species / % Identity	CBM[a]	SP[b]	Other Secretion[c]	TMHMM[d]	Loc.[e]
13	JM658_07705	GH5	Beta-xylanase	–	WP_202244991.1 [*Fulvivirga* sp. 2943] 35.1%	P	O	P	O	Extra
14	JM658_07745	GH3	Beta-xylosidase	WP_008613873.1 [*Joostella marina*] 36.6%	HIJ70784.1 [Planctomycetes bacterium] 54.9%	P	P	O	O	Extra
15	JM658_09135	GH43	Xylosidase	–	WP_089380719.1 [*Lutibacter agarilyticus*] 74.4%	O	P	O	O	Extra
16	JM658_09145	GH0	Xylosidase	–	WP_089381103.1 [*Lutibacter agarilyticus*] 80.1%	O	P	O	O	Extra
17	JM658_09160	GH0	Xylosidase/arabinosidase	–	WP_010420365.1 [*Anaerophaga thermohalophila*] 52.1%	P	P	P	O	Extra
Other hemicellulose-degrading enzymes										
18	JM658_00480	GH16	Beta-glucanase	WP_008611030.1 [*Joostella marina*] 76.7%	WP_108372171.1 [*Flavobacterium magnum*] 70.4%	O	P	O	O	Extra
19	JM658_02175	GH95	Alpha-L-fucosidase	WP_050989689.1 [*Joostella marina*] 36.5%	WP_154853733.1 [*Cyclobacterium xiamenense*] 57.9%	P	P	O	O	Extra
20	JM658_03210	GH29	Alpha-L-fucosidase	WP_008614653.1 [*Joostella marina*] 29.8%	WP_143398590.1 [*Fulvivirga* sp. M361] 57.8%	O	P	O	O	Extra
21	JM658_03835	GH97	Alpha-glucosidase	WP_008610695.1 [*Joostella marina*] 88.9%	WP_202027890.1 [*Galbibacter mesophilus*] 82.8%	O	P	O	O	Extra
22	JM658_06890	GH31	Alpha-glucosidase	WP_008614297.1 [*Joostella marina*] 88.4%	WP_178949212.1 [*Galbibacter* sp. BG1] 87.0%	O	O	O	O	Intra
23	JM658_07720	GH31	Alpha-xylosidase	WP_008614297.1 [*Joostella marina*] 24.8%	MBN2211898.1 [Sedimentisphaerales bacterium] 32.7%	P	P	O	O	Extra
24	JM658_07730	GH29	Alpha-L-fucosidase	WP_008615295.1 [*Joostella marina*] 32.4%	WP_139958298.1 [*Flavicella sediminum*] 64.3%	O	P	O	O	Extra

Table 3 (continued)

No	Locus_Tag	Family	Annotation	Closest Sequence to *Joostella* sp. % Identity	Closest Sequence to other species / % Identity	CBM[a]	SP[b]	Other Secretion[c]	TMHMM[d]	Loc.[e]
25	JM658_07755	GH29	Alpha-L-fucosidase	WP_008613150.1 [*Joostella marina*] 30.4%	HGT31229.1 [Verrucomicrobia subdivision 3 bacterium] 63.6%	P	P	O	O	Extra
26	JM658_10600	GH97	Alpha-glucosidase	WP_008613871.1 [*Joostella marina*] 50.1%	WP_072402860.1 [*Flaviramulus basaltis*] 75.8%	O	P	O	O	Extra
27	JM658_11205	GH29	Alpha-L-fucosidase	WP_008611906.1 [*Joostella marina*] 26.8%	WP_136465879.1 [*Flagellimonas* sp. XY-359] 71.6%	O	P	O	O	Extra
28	JM658_11580	GH43	Beta-galactosidase	WP_008613251.1 [*Joostella marina*] 30.0%	WP_202029452.1 [*Galbibacter mesophilus*] 95.3%	O	P	O	O	Extra
29	JM658_11595	GH2	Beta-galactosidase	WP_040880371.1 [*Joostella marina*] 34.9%	WP_202029455.1 [*Galbibacter mesophilus*] 95.8%	P	O	O	O	Intra
30	JM658_12270	GH29	Alpha-L-fucosidase	WP_008613150.1 [*Joostella marina*] 73.9%	WP_149276527.1 [*Pareuzebyella sediminis*] 63.9%	P	P	O	O	Extra
31	JM658_12280	GH95	Alpha-L-fucosidase	WP_050989689.1 [*Joostella marina*] 64.7%	EKF55753.1 [*Galbibacter marinus*] 62.5%	P	P	O	O	Extra
32	JM658_12285	GH29	Alpha-L-fucosidase	WP_008614653.1 [*Joostella marina*] 30.9%	WP_167615279.1 [*Maribellus* sp. GM1-28] 59.9%	O	P	O	O	Extra
33	JM658_12290	GH29	Alpha-L-fucosidase	WP_008614653.1 [*Joostella marina*] 35.3%	SFW57402.1 [*Sinomicrobium oceani*] 65.4%	O	P	O	O	Extra
34	JM658_12310	GH29	Alpha-L-fucosidase	WP_008615295.1 [*Joostella marina*] 34.5%	WP_074977762.1 [*Zhouia amylolytica*] 72.3%	O	P	O	O	Extra
35	JM658_12360	GH2	Beta-galactosidase	WP_008611892.1 [*Joostella marina*] 67.5%	WP_038266115.1 [*Zhouia amylolytica*] 72.3%	P	P	O	O	Extra
36	JM658_12995	GH29	Alpha-L-fucosidase	WP_008613150.1 [*Joostella marina*] 33.4%	WP_046150553.1 [*Parabacteroides* sp. HGS0025] 42.5%	P	P	O	O	Extra
37	JM658_15500	GH29	Alpha-L-fucosidase	WP_008615295.1 [*Joostella marina*] 85.1%	WP_141384486.1 [*Flavobacterium flevense*] 77.0%	O	P	O	O	Extra

Table 3 (continued)

No	Locus_Tag	Family	Annotation	Closest Sequence to *Joostella* sp. % Identity	Closest Sequence to other species / % Identity	CBM[a]	SP[b]	Other Secretion[c]	TMHMM[d]	Loc.[e]
38	JM658_15510	GH29	Alpha-L-fucosidase	WP_008615291.1 [*Joostella marina*] 78.6%	WP_139694957.1 [*Tamlana fucoidanivorans*] 74.1%	O	P	O	O	Extra
39	JM658_15590	GH2	Beta-galactosidase	WP_008615267.1 [*Joostella marina*] 71.2%	WP_203395008.1 [*Polaribacter* sp. Q13] 65.1%	O	O	O	P	Intra
40	JM658_15640	GH2	Beta-galactosidase	WP_008615235.1 [*Joostella marina*] 89.4%	WP_076456730.1 [*Zobellia uliginosa*] 82.6%	O	P	O	O	Extra

Note: Listed proteins were subjected to domain search and other structure predictions (*P* present, *O* absent,— Not applicable), including [a]carbohydrate binding modules, [b]signal-peptide, [c]other secretion system, and [d]transmembrane helices structure. [e]Protein localization (intra-, extracellular, or ectoenzyme) is predicted (*Intra* Intracellular enzyme, *Extra* Extracellular enzyme, *Ecto* Ectoenzyme)

To further verify the cellulolytic and xylanolytic abilities of strain M1-2^T, quantitative enzyme assays of endoglucanase, xylanase, β-glucosidase and β-xylosidase were performed. *J. atrarenae* M1-2^T was grown in Marine Broth supplemented with 0.1% (*w/v*) carboxylmethyl cellulose (CMC) and 0.1% (*w/v*) beechwood xylan, to induce the production of its cellulolytic and xylanolytic enzymes. The cultures with different incubation period (1–5 days) were harvested by centrifugation (8000× g, 5 min, 4 °C). The supernatant was treated as the extracellular crude enzymes of the bacterium, whereas the intracellular crude enzymes were extracted from the pelleted cells by B-PER™ Bacterial Protein Extraction Reagent (Thermo Fisher Scientific, USA). Prior for enzyme assay, both extra- and intracellular enzymes were dialyzed against 100 mM Tris–HCl buffer (pH 7). Unless specified, enzyme assays were conducted by mixing 0.5 mL enzymes solution with 0.5 mL substrates and incubating at 30 °C, pH 7 for 30 min. Endoglucanase and xylanase activities were tested using 1% (*w/v*) carboxymethyl cellulose (CMC) and 1% (*w/v*) beechwood xylan as substrates, respectively, and coupled with 3,5-dinitrosalicyclic acid (DNS) assay to measure the release of reducing sugars. Absorbance changes for DNS assay were detected at 540 nm in a spectrophotometer SP-UV 300 (Perkin Elmer, USA). One unit (U) of endoglucanase or xylanase activity is defined as the release of 1 nmol reducing sugar per minute per mL under the assay condition. β-glucosidase and β-xylosidase activities were measured using substrates *p*-nitrophenyl-β-D-glucopyranoside (*p*NPG, 5 mM) and *p*-nitrophenyl-β-D-xylopyranoside (*p*NPX, 5 mM), respectively. The release of *p*-nitrophenol (*p*NP) compound from these substrates was detected at 405 nm in a spectrophotometer. One unit (U) of β-glucosidase or β-xylosidase activity is defined as the release of 1 nmol *p*NP per minute per mL under the assay condition. The protein concentration of both extra- and intracellular extracts of *Joostella atrarenae* M1-2^T were determined using the Pierce™ BCA Protein Assay kit (Thermo Fisher Scientific, USA), following the manufacturer's protocol with bovine serum albumin (BSA) as the standard. Absorbance changes of the protein assays were measured at 562 nm in a spectrophotometer SP-UV 300 (Perkin Elmer, USA). All assays were conducted in triplicates.

Based on the enzymatic assay results represented in Fig. 3, both the intracellular and extracellular extracts of *J. atrarenae* M1-2^T were found to exhibit all four tested enzymes. In overall, the extracellular extracts of the bacterium had higher accumulative specific enzyme activities as compared to the intracellular portions throughout the five days incubation. This finding is in complement to the description of genome (Table 1), where most lignocellulolytic related enzymes in strain M1-2^T are expressed extracellularly. Nonetheless, as compared among the four enzymes, xylanase had higher

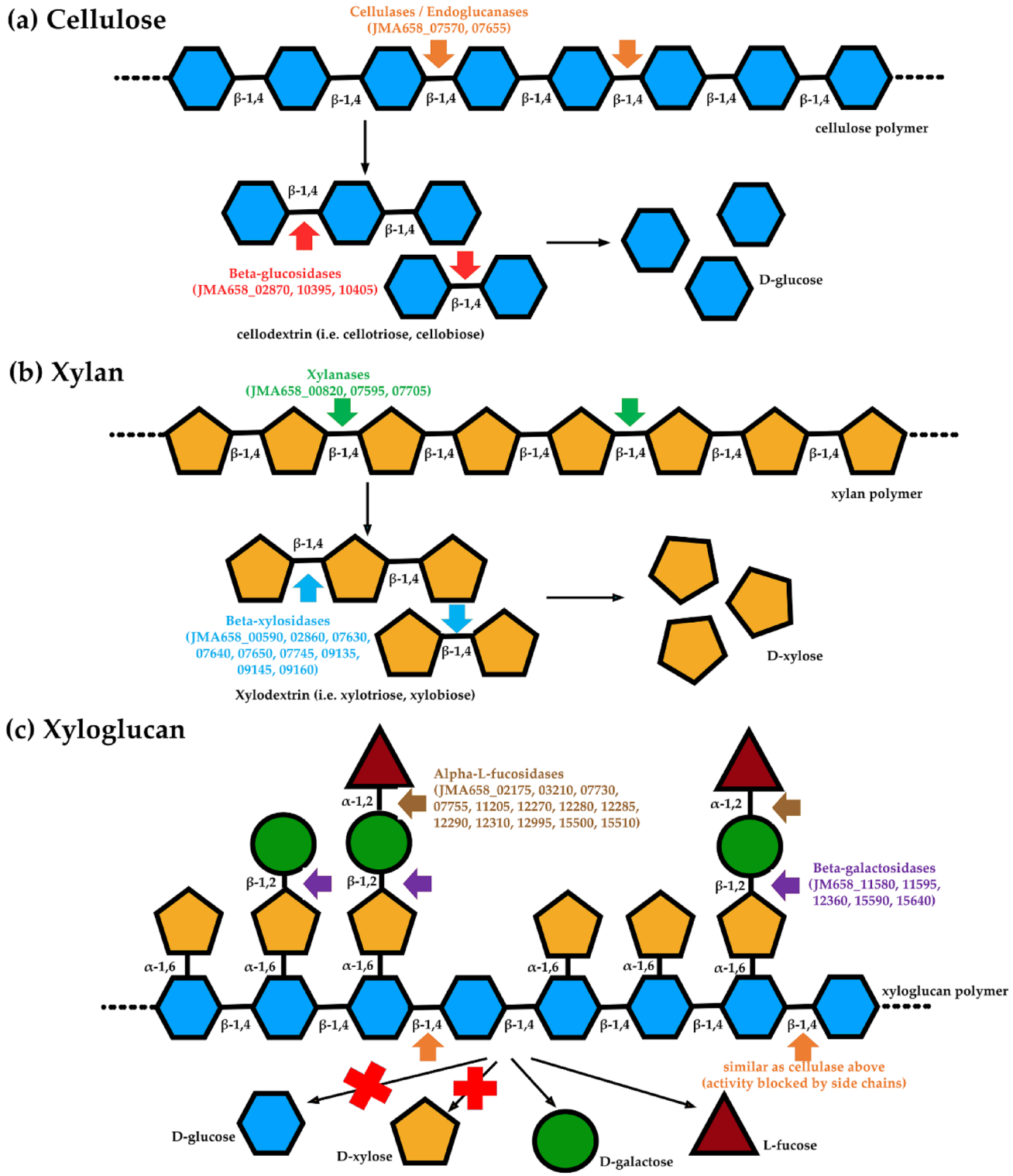
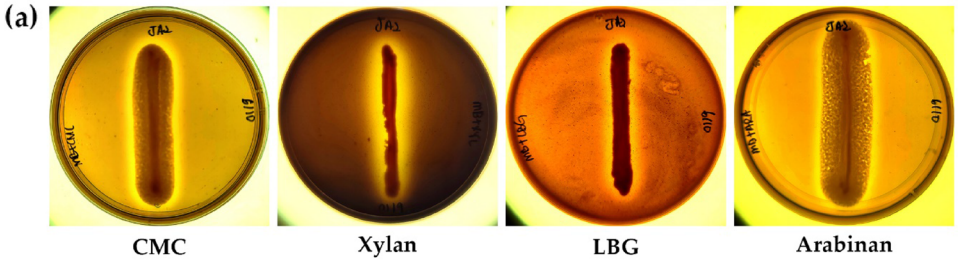


Fig. 1 Prediction of cellulose, xylan, and xyloglucan degradation by *J. atrarenae* M1-2^T based on lignocellulolytic gene mining

specific enzyme activities, with maximum achieved at day 4 incubation (69.56 ± 1.30 nmol·mg^{-1}·min^{-1}) as tested in the extracellular portion. This indicated that the bacterium had higher preference against xylan substrate rather than the CMC, when both substrates were provided at the same time during the bacterial growth. In other words, strain M1-2^T most probably has better degrading ability against the hemicellulose portion (represented by xylan in this study) than



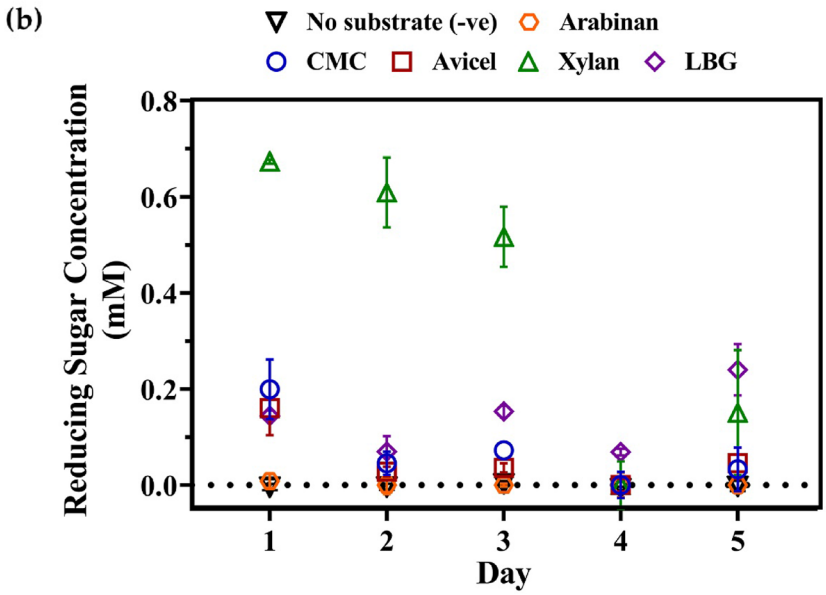


Fig. 2 The ability to utilize different polysaccharides by *J. atrarenae* M1-2[T]. **a** Hydrolytic plate assays and **b** measurement of reducing sugar released from suspension culture supplemented with different substrates from Day 1 – 5

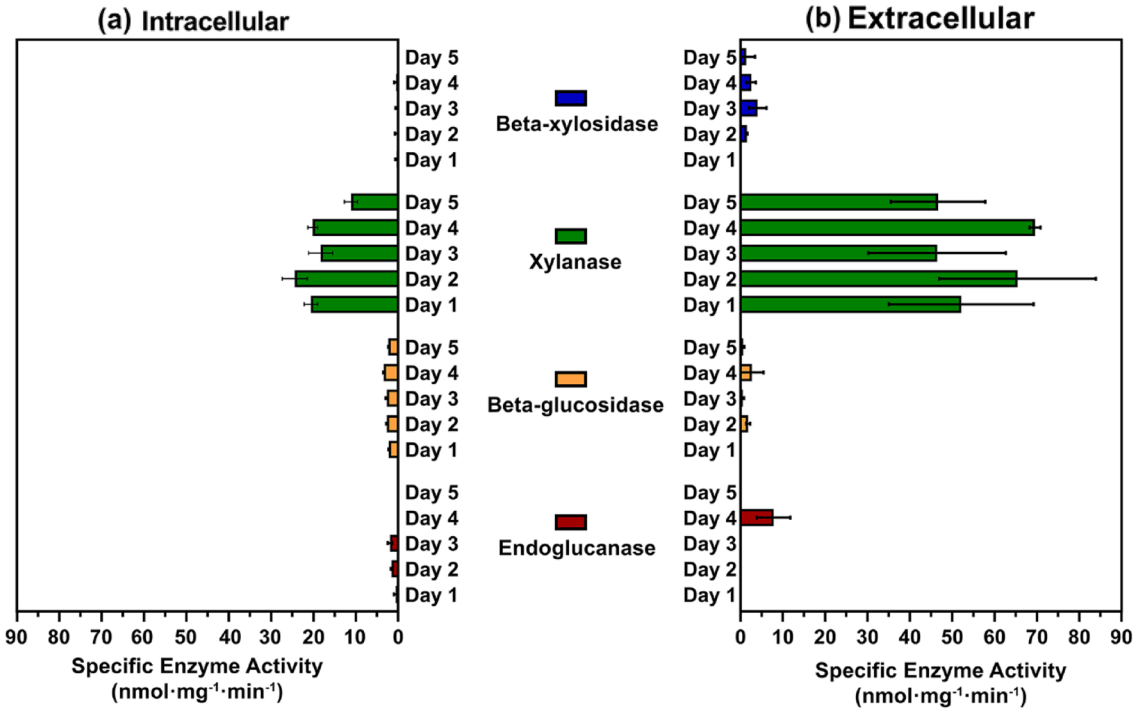


Fig. 3 Specific enzyme activities of endoglucanase, beta-glucosidase, xylanase, and beta-xylosidase as tested in the intracellular and extracellular crude extracts of *J. atrarenae* M1-2.[T], which were harvested from Day 1 – 5

the cellulose portion (represented by CMC) of seaweed biomass in nature. Other marine or halophilic bacteria that can produce lignocellulolytic enzymes had been studied elsewhere (Liew et al. 2018; Jin et al. 2019; Lam et al. 2020; Zhu et al. 2020). For instance, a halophile called *Roseithermus sacchariphilus* strain RA was found to have 11 lignocellulolytic encoded genes (Liew et al. 2018). As further studies, two xylanases (XynRA1 and XynRA2) of *R. sacchariphilus* strain RA were cloned, expressed, purified, and characterized, and both xylanases exhibited different characteristics (XynRA1 is non-salt tolerant; XynRA2 is salt tolerance up to 5 M NaCl with higher thermostability) (Liew et al. 2019; Teo et al. 2019). So far, none of the enzymes from *Joostella* spp. has been heterologous expressed and characterized. Since *J. atrarenae* M1-2^T harbors multiple copies of genes that encode for enzyme with same catalytic function (i.e. 6 β-xylosidases) (Table 1), the characterization of these enzymes in future will permit the identification of their differences and unique features.

In conclusion, *J. atrarenae* M1-2^T genome encoded 5 and 12 enzymes for degradation of cellulose and xylan, respectively. Meanwhile, a total of 24 genes related to side-chain removal of the hemicellulose are also identified. Furthermore, enzyme assays verified the presence of endoglucanase, xylanase, β-glucosidase, and β-xylosidase activities. To the best of our knowledge, this is the first genome analysis describing the lignocellulose degradation by *J. atrarenae* M1-2^T. The findings could provide new enzymes for lignocellulosic biomass processing.

Acknowledgements K.J.L. is a post-doctoral researcher of Universiti Teknologi Malaysia funded by NFIS 2020-2021. M.R.Z. and C.W.L.H. would like to thank Zamalah Program of Universiti Teknologi Malaysia for scholarship support.

Author contributions K.J.L. performed the genome analysis and wrote the manuscript. M.R.Z. and C.W.L.H. collected the sequencing data and carried out the enzyme assay experiments. M.C.Y.T. reviewed and edited the manuscript. C.S.C. conceived the project. All authors reviewed and approved the final manuscript.

Funding This work was supported by the Ministry of Higher Education Malaysia under Fundamental Research Grant Scheme (FRGS/1/2019/WAB13/UTM/02/1) as main sponsor. This work was also supported by Newton Fund Impact Scheme (NFIS) 2020–2021 from Malaysian Industry-Government Group for High Technology with project ID 536423090 (RADIS cost centre number:4B568).

Data availability All data generated or analyzed during this study are included in this published article. Accession Numbers: The whole-genome sequence of *J. atrarenae* M1-2^T can be accessed via National Center for Biotechnology Information (NCBI) database, either with BioProject accession number PRJNA694488, BioSample accession number SAMN17525809, or Genbank accession number JAETXX000000000. The raw sequence data from the sequencing was also deposited in NCBI Sequence Read Archive (SRA) database with accession number SRR15341968.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Armenteros JJA, Tsirigos KD, Sønderby CK et al (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol 37:420–423. https://doi.org/10.1038/s41587-019-0036-z

Buchfink B, Reuter K, Drost H-G (2021) Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods 18:366–368. https://doi.org/10.1038/s41592-021-01101-x

Cantalapiedra CP, Hernández-Plaza A, Letunic I, et al (2021) eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. bioRxiv. https://doi.org/10.1093/molbev/msab293

Cardoso J, Fonseca L, Abrantes I (2020) α-l-fucosidases from Bursaphelenchus xylophilus secretome—Molecular characterization and their possible role in breaking down plant cell walls. Forests 11:265. https://doi.org/10.3390/f11030265

Chen Y, Chen Y, Shi C et al (2018) SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. Gigascience. https://doi.org/10.1093/gigascience/gix120

Chukwuma OB, Rafatullah M, Tajarudin HA, Ismail N (2020) Lignocellulolytic enzymes in biotechnological and industrial processes: a review. Sustainability 12:7282. https://doi.org/10.3390/su12187282

Hwang D, Lee Y, Kim G (2005) Large submarine groundwater discharge and benthic eutrophication in Bangdu Bay on volcanic Jeju Island, Korea. Limnol Oceanogr 50:1393–1403

Jain C, Rodriguez-R LM, Phillippy AM et al (2018) High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9:1–8. https://doi.org/10.1038/s41467-018-07641-9

Jin M, Gai Y, Guo X et al (2019) Properties and applications of extremozymes from deep-sea extremophilic microorganisms: A mini review. Mar Drugs 17:656. https://doi.org/10.3390/md17120656

Jones P, Binns D, Chang H-Y et al (2014) InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Khan MI, Shin JH, Kim JD (2018) The promising future of microalgae: current status, challenges, and optimization of a sustainable and renewable industry for biofuels, feed, and other products. Microb Cell Fact 17:1–21

Kim B-C, Lee KH, Kim MN et al (2011) Joostella atrarenae sp. nov., a novel member of the Flavobacteriaceae originating from the black sea sand of Jeju Island. Curr Microbiol 62:606–611. https://doi.org/10.1007/s00284-010-9750-y

Kim D, Park S, Chun J (2021) Introducing EzAAI: a pipeline for high throughput calculations of prokaryotic average amino acid identity. J Microbiol 59:476–480. https://doi.org/10.1007/s12275-021-1154-0

Lam MQ, Oates NC, Thevarajoo S et al (2020) Genomic analysis of a lignocellulose degrading strain from the underexplored genus Meridianimaribacter. Genomics 112:952–960. https://doi.org/10.1016/j.ygeno.2019.06.011

Liew KJ, Ngooi CY, Shamsir MS et al (2019) Heterologous expression, purification and biochemical characterization of a new endo-1,4-β-xylanase from Rhodothermaceae bacterium RA. Protein Express Purific. https://doi.org/10.1016/j.pep.2019.105464

Liew KJ, Teo SC, Shamsir MS, et al (2018) Complete genome sequence of Rhodothermaceae bacterium RA with cellulolytic and xylanolytic activities. 3 Biotech https://doi.org/10.1007/s13205-018-1391-z

Meier-Kolthoff JP, Carbasse JS, Peinado-Olarte RL, Göker M (2022) TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. Nucleic Acids Res 50:D801–D807. https://doi.org/10.1093/nar/gkab902

Möller S, Croning MDR, Apweiler R (2001) Evaluation of methods for the prediction of membrane spanning regions. Bioinformatics 17:646–653. https://doi.org/10.1093/bioinformatics/17.7.646

Page AJ, Cummins CA, Hunt M et al (2015) Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31:3691–3693. https://doi.org/10.1093/bioinformatics

Prjibelski A, Antipov D, Meleshko D et al (2020) Using SPAdes de novo assembler. Curr Protocols Bioinform. https://doi.org/10.1002/cpbi.102

Qin Q-L, Xie B-B, Zhang X-Y et al (2014) A proposed genus boundary for the prokaryotes based on genomic insights. J Bacteriol 196:2210–2215. https://doi.org/10.1128/JB.01688-14

Quan Z-X, Xiao Y-P, Roh SW et al (2008) Joostella marina gen. nov., sp. nov., a novel member of the family Flavobacteriaceae isolated from the East Sea. Int J Syst Evol Microbiol 58:1388–1392. https://doi.org/10.1099/ijs.0.65611-0

Rai M, Ingle AP, Pandit R et al (2019) Emerging role of nanobiocatalysts in hydrolysis of lignocellulosic biomass leading to sustainable bioethanol production. Catalysis Rev 61:1–26. https://doi.org/10.1080/01614940.2018.1479503

Selvaratnam C, Thevarajoo S, Ee R et al (2016) Genome sequence of Roseivirga sp. strain D-25 and its potential applications from the genomic aspect. Mar Genom 28:29–31. https://doi.org/10.1016/j.margen.2016.04.004

Sethupathy S, Morales GM, Li Y et al (2021) Harnessing microbial wealth for lignocellulose biomass valorization through secretomics: a review. Biotechnol Biofuels 14:1–31. https://doi.org/10.1186/s13068-021-02006-9

Song J-W, Kim J, Kim T-H (2018) Distributions of dissolved organic matter in submarine groundwater discharge (SGD) in Jeju Island. Ocean Polar Res 40:77–85

Stackebrandt E, Chertkov O, Lapidus A et al (2013) High-quality-draft genome sequence of the yellow-pigmented flavobacterium Joostella marina type strain (En5 T). Stand Genom Sci 8:37–46. https://doi.org/10.4056/sigs.3537045

Tatusova T, DiCuccio M, Badretdin A et al (2016) NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569

Teo SC, Liew KJ, Shamsir MS et al (2019) Characterizing a halotolerant GH10 xylanase from Roseithermus sacchariphilus strain RA and its CBM-truncated variant. Int J Mol Sci. https://doi.org/10.3390/ijms20092284

Thevarajoo S, Selvaratnam C, Goh KM et al (2016) Vitellibacter aquimaris sp. nov., a marine bacterium isolated from seawater. Int J Syst Evol Microbiol 66:3662–3668. https://doi.org/10.1099/ijsem.0.001248

Wang Y, Song Q, Zhang X-H (2016) Marine microbiological enzymes: studies with multiple strategies and prospects. Mar Drugs 14:171. https://doi.org/10.3390/md14100171

Yoon S-H, Ha S-M, Kwon S et al (2017) Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. Int J Syst Evol Microbiol 67:1613. https://doi.org/10.1099/ijsem.0.001755

Zhang H, Yohe T, Huang L et al (2018) dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 46:W95–W101. https://doi.org/10.1093/nar/gky418

Zhou X, Li W, Mabon R, Broadbelt LJ (2017) A critical review on hemicellulose pyrolysis. Energ Technol 5:52–79. https://doi.org/10.1002/ente.201600327

Zhu D, Adebisi WA, Ahmad F et al (2020) Recent development of extremophilic bacteria and their application in biorefinery. Front Bioeng Biotechnol 8:483. https://doi.org/10.3389/fbioe.2020.00483