Bacteroidetes contribute to the carbon and nutrient cycling of deep sea through breaking down diverse glycans

Rikuan Zheng^{1,2,3,4}, Ruining Cai^{1,2,3,4}, Rui Liu^{1,2,4}, Ge Liu^{1,2,4}, Chaomin Sun^{1,2,4*}

¹CAS Key Laboratory of Experimental Marine Biology & Center of Deep Sea Research, Institute of Oceanology, Chinese Academy of Sciences, Qingdao, China

²Laboratory for Marine Biology and Biotechnology, Qingdao National Laboratory for Marine Science and Technology, Qingdao, China

³College of Earth Science, University of Chinese Academy of Sciences, Beijing, China

⁴Center of Ocean Mega-Science, Chinese Academy of Sciences, Qingdao, China

* Corresponding author

Chaomin Sun Tel.: +86 532 82898857; fax: +86 532 82898857.

E-mail address: sunchaomin@qdio.ac.cn

22 **Abstract**

23 Bacteroidetes are thought to be specialized for the degradation of algae-derived ocean
 24 polysaccharides and are a major contributor to the marine carbon and nutrient cycling.
 25 Here, we first show Bacteroidetes are the second most abundant phylum bacteria in
 26 deep-sea cold seep and possess more genes associated with polysaccharides
 27 degradation than other bacteria through metagenomics methods. We further isolate a
 28 novel Bacteroidetes species, *Maribellus comscasis* WC007^T, which can efficiently
 29 degrade numerous different polysaccharides including: cellulose, pectin, fucoidan,
 30 mannan, xylan and starch. These results are verified by transcriptomic analyses and
 31 growth assays. Notably, we find cellulose promotes abundant bacterial growth, and
 32 using transcriptomics and metabolomics we finally report on the underlying
 33 mechanisms of cellulose degradation and utilization, as well as potential contributions
 34 to the carbon cycling. Overall, our results suggest Bacteroidetes play key roles in the
 35 deep-sea carbon and nutrient cycling, likely due to their high abundance and
 36 prominent polysaccharide degradation capabilities.

37

38

39 **One Sentence Summary**

40 Bacteroidetes contribute to ocean carbon and nutrient cycle.

41

42

43 **Introduction**

44 Roughly half of the global net primary biomass production is oceanic and carried out
 45 predominantly by small marine phytoplankton. As the primary biomass production,
 46 complex carbohydrates are a ubiquitous energy source for microorganisms in both
 47 terrestrial and marine ecosystems (**Grondin et al., 2017**). Complex carbohydrates,
 48 mostly in the form of polysaccharides, are the largest repository of organic carbon in
 49 the biosphere (**Malhi, 2002**). In the ocean, polysaccharides constitute a substantial
 50 fraction of biomass and marine algae are a dominant source of polysaccharides
 51 (**Benner et al., 1992; Engel et al., 2004**). Marine algae are likely composed of greater
 52 than 50% polysaccharides, which are found as structural components of cell walls and
 53 intracellular energy storage compounds (**KloaregQuatrano, 1988**). The
 54 polysaccharides produced at the ocean surface can sink to a lower water column and
 55 even the deep sediments. With this, large amounts of mostly algal plants and animal
 56 debris are deposited from the upper ocean, providing an important organic carbon
 57 source for polysaccharide-degrading bacteria in deep-sea sediments (**Gao et al., 2017**).
 58 Algal debris is rich in difficult-to-degrade polysaccharides, such as pectin, cellulose
 59 and hemicellulose (including fucoidan, mannan, xylan, glucan, arabinogalactan and
 60 others) (**Snajdr et al., 2011**). These polysaccharides are the main components of the
 61 cell wall and intercellular layer. Breakdown of these polysaccharides by heterotrophic
 62 microorganisms plays a significant role in organic biomass degradation and greatly
 63 contributes to the marine carbon and nutrient cycling (**Azam, 1998; Moran et al.,**

2016). Likewise, polysaccharides are a major nutrient source for degradation microorganisms (**AzamMalfatti, 2007**), providing these organisms with a significant growth advantage and allowing them to efficiently occupy niches in harsh deep-sea environments.

Notably, many members of the bacterial phylum Bacteroidetes are specialized in polysaccharide degradation and are thought to be the most abundant group of ocean bacteria after Proteobacteria and Cyanobacteria (**Kirchman, 2002**). Bacteroidetes are globally distributed in coastal areas (**Pommier et al., 2007**), sediment regions, hydrothermal vents and numerous other marine environments (**Alonso et al., 2007; Teeling et al., 2012**). Marine Bacteroidetes are commonly thought to play a key role in degrading phytoplankton polysaccharides (**Fernandez-Gomez et al., 2013**), likely owing to the number and diversity of carbohydrate-active enzymes (CAZymes) in their genomes (**Hahnke et al., 2016**). Bacteroidetes CAZymes are categorized into families of glycoside hydrolases (GHs), glycoside transferases (GTs), carbohydrate-binding modules (CBMs), carbohydrate esterases (CEs), polysaccharide lyases (PLs), sulfatases and a broad range of auxiliary enzymes (**Bauer et al., 2006; Cantarel et al., 2009; Fernandez-Gomez et al., 2013**). As such, Bacteroidetes have evolved unique and sophisticated degradation systems, typically called polysaccharide utilization loci (PULs). These loci are tightly regulated and are comprised of colocalized gene clusters that encode CAZymes and the protein assemblies required for the degradation of complex carbohydrates. In addition to substrate-specific

85 CAZymes genes, Bacteroidetes PULs also contain genes encoding TonB-dependent
86 transporters (TBDTs), which are SusCD-like extracellular lipoproteins with integral
87 membrane beta-barrels (**Kappelmann et al., 2019**). Polysaccharides initially bind to
88 outer membrane proteins and are cleaved into oligosaccharides by endo-active
89 enzymes (**Kappelmann et al., 2019**). Oligosaccharides are transported from the outer
90 membrane into the periplasm in a TonB-dependent manner. In the periplasm,
91 oligosaccharides are protected from competition by other bacteria and are further
92 degraded into monosaccharides, which are then moved by dedicated transporters
93 across the cytoplasmic membrane into the cytoplasm for utilization (**Glenwright et**
94 **al., 2017**).

95 Various mechanisms of Bacteroides polysaccharide degradation have been
96 investigated in the human gut, including the degradation of pectin and hemicellulose
97 (**Larsbrink et al., 2014; Rogowski et al., 2015**). Intriguingly, a study suggests
98 human gut bacteria could acquire CAZymes encoding genes from marine bacteria,
99 which may be a general factor of CAZyme diversity in human gut microbes
100 (**Hehemann et al., 2010**). However, as compared to human gut Bacteroides, few
101 studies have investigated polysaccharide degradation by marine Bacteroides,
102 especially the deep-sea variety.

103 Here we first investigate the abundance of Bacteroidetes in the deep-sea cold
104 seep by metagenomics methods. We successfully isolate a novel Bacteroidetes strain
105 *Maribellus comscasis* WC007^T by a goal-directed method from the deep-sea

106 sediments. Strain WC007^T efficiently degrades various polysaccharides. Using
 107 genomics, transcriptomics and metabolomics, we deeply characterize the degradation
 108 and utilization of different polysaccharides by *M. comscasis* WC007^T. Lastly, we
 109 discuss the contribution of Bacteroidetes to the deep ocean carbon and nutrient
 110 cycling.

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

Results

Bacteroides are high abundance bacteria in deep-sea cold seep. To gain insights into the bacterial community composition of the deep sea, we performed OTUs sequencing analysis of bacteria present in cold seep sediments. Our results indicated that, at the phylum level, Proteobacteria was the most abundant domain (68.8%), followed by Bacteroidetes (13.2%), Chloroflexi (4.5%), Planctomycetes (4.0%), Firmicutes (1.9%), Acidobacteria (1.7%), Lateseibacteria (1.4%) and Spirochaetes (1.1%). Together, these constituted roughly 96.6% of all bacteria in the sampled cold seep. The remaining 3.4% was other phyla bacteria (Figure 1A). Notably, Bacteroidetes were the second most abundant phylum, suggesting Bacteroidetes are

Figure. 1 Quantification and polysaccharide degradation predictions of Bacteroides derived from deep-sea cold seep sediments. (A) 16S rRNA amplicon-based assessment of bacterial abundance in the deep-sea cold seep sediments. (B) Calculated number of genes associated with polysaccharide degradation from metagenome-assembled genomes of Bacteroides, Chloroflexi

142 and Proteobacteria derived from deep-sea cold seep. (C) Calculated number of genes encoding
143 SusC, SusD, CAZymes and sulfatases identified from 8 metagenome-assembled genomes of
144 Bacteroides derived from deep-sea cold seep. X-axis indicates the names of Bacteroides
145 metagenome-assembled genomes.

146 dominant in these deep-sea regions. To understand the ability of deep-sea
147 Bacteroidetes to degrade polysaccharides, we used metagenome sequencing to predict
148 and enumerate genes encoding CAZymes. Fifty-one high quality
149 metagenome-assembled genomes (MAGs) for the three most abundant phyla
150 (Proteobacteria, Bacteroidetes and Chloroflexi) were obtained by using a hybrid
151 binning strategy combined with manual inspection and data curation (Supplementary
152 file 2). Through the analysis of 51 MAGs, we found that Bacteroidetes had the highest
153 abundances of CAZymes, followed by Chloroflexi and Proteobacteria (Figure 1B).
154 We also analyzed the number of genes encoding CAZymes, SusC, SusD and sulfatase
155 proteins in our Bacteroidetes MAGs. As expected, all 8 MAGs were rich in CAZymes,
156 SusC, SusD and sulfatase genes (Figure 1C, Supplementary file 3), as these genes
157 were necessary components for polysaccharide degradation.

158 To further understand the types of polysaccharides degraded by deep-sea
159 Bacteroidetes, we conducted in-depth analysis of GHs obtained from above 8 MAGs.
160 The results showed that the major GH families of Bacteroidetes were GH2 (a family
161 of diverse functions), GH3 (also a family of diverse functions), GH29 (a family of
162 alpha-L-fucosidases) and GH92 (a family of mostly alpha-mannosidases) (Figure

1-figure supplement 1). This suggests Bacteroidetes may be important participants of
diverse polysaccharides degradation in the deep-sea cold seep environment.

Isolation and characterization of a new deep-sea Bacteroides species, *Maribellus comscasis* WC007^T, that has prominent polysaccharide degradation functions.

Given their high abundance and potential polysaccharide degradation functions, we
isolated deep-sea derived Bacteroides to investigate novel metabolic pathways of
polysaccharide degradation. Considering the superior polysaccharide degradation
capability of Bacteroides, the deep-sea sediment sample was anaerobically enriched at

Figure 2. Goal-directed isolation of a novel deep-sea Bacteroides species possessing strong
polysaccharide degradation capabilities. (A) Diagrammatic scheme for enrichment and isolation of
deep-sea Bacteroides. (B) TEM observation of *Maribellus comscasis* WC007^T, bar = 2 μm. (C)

175 Genomic distribution of SusC, SusD, CAZymes, sulfatases genes in *M. comscasis* WC007^T. (D)

176 Quantification of SusC, SusD, CAZymes, sulfatases and peptidases genes in the *M. comscasis*

177 WC007^T genome.

178 28 °C for one month in an inorganic medium supplemented with various

179 polysaccharides (Figure 2A). Enriched samples were then plated on solid medium in

180 Hungate tubes and individual colonies with distinct morphology were picked and

181 cultured (Figure 2A). As expected, most cultured colonies were identified as

182 Bacteroides. Among them, strain WC007^T grew significantly faster than other

183 Bacteroides strains and was chosen for further study.

184 Cells of strain WC007^T were curved, approximately $2.0-4.0 \times 0.5-0.8 \mu\text{m}$ in size,

185 and had no flagellum, as indicated by TEM (Figure 2B). The whole genome size of

186 strain WC007^T was 7,811,310 bp with a DNA G+C content of 38.38%. A sequence

187 similarity calculation using the NCBI server indicated the closest relatives of strain

188 WC007^T were *Maribellus luteus* XSD2^T (95.70%), *Mariniphaga sediminis* SY21^T

189 (93.44%) and *Draconibacterium sediminis* MCCC 1A00734^T (92.99%). Phylogenetic

190 analyses based on the 16S rRNA and genome sequences showed strain WC007^T

191 belonged to the genus *Maribellus* and formed an independent phyletic line with the

192 type strain *Maribellus luteus* XSD2^T (Figure 2-figure supplement 2 and 3). Therefore,

193 we propose strain WC007^T to be classified as the type strain of a novel species in the

194 genus *Maribellus*, for which the name *Maribellus comscasis* sp. nov. is proposed.

Details of the phenotypic, chemotaxonomic, and genotypic characterizations of WC007^T were shown in Supplementary files 4-6. When compared to its closest relative *Maribellus luteus* XSD2^T, strain WC007^T showed a distinct capability for degrading cellulose, pectin and xylan (Supplementary file 5), indicating its potential for polysaccharide degradation. Consistently, genes encoding SusC, SusD, CAZymes and sulfatase were ubiquitously distributed in the WC007^T genome (Figure 2C). Moreover, detailed analysis with the Pfam, TIGRFAM, dbCAN, MEROPS and CAZyme databases showed that strain WC007^T had 634 CAZymes (including 374 GHs, 88 CEs, 21 PLs, 77 CBMs and 74 GTs), 120 SusC-like proteins, 82 SusD-like proteins, 422 sulfatases and 475 peptidases (Figure 2D). The numbers of genes encoding CAZymes, SusC/SusD pairs and sulfatases in WC007^T were significantly higher than those of other *Bacteroides* from different habitats including human gut, lake, shallow ocean and deep sea (Figure 2-figure supplement 4 and Supplementary file 7). Overall, these data suggest WC007^T is a bacterium with a notable polysaccharide degradation capability. Studies focusing on this strain may provide a more comprehensive understanding of carbon cycling in the deep sea.

Prediction and validation of polysaccharides degradation by *M. comscasis* WC007^T. To gain further insights into the potential mechanisms of polysaccharide degradation by *M. comscasis* WC007^T, we conducted an in-depth analysis of the CAZymes, Sus transport genes and sulfatases present in the WC007^T genome. In previous studies (Kabisch et al., 2014; Kappelmann et al., 2019; Lapebie et al.,

216 **2019**), polysaccharide utilization loci (PULs) are defined as the presence of CAZymes
217 with predicted polysaccharide degradation function and a gene pair encoding
218 SusC/D-like proteins (**Sonnenburg et al., 2010**). We identified 82 PUL-containing
219 gene clusters

Figure 3. Prediction and verification of polysaccharide degradation and utilization by *M. comscasis* WC007^T. (A) Selected putative PULs responsible for degradation of cellulose, pectin, fucoidan, mannan, xylan and starch polysaccharides. Genes encoding CAZymes, SusC/SusD,

224 sigma factor, anti-sigma factor, sulfatase and peptidase are depicted in different colors.
 225 Hypothetical proteins and genes involved in other metabolic functions are depicted in white. **(B)**
 226 Growth assays of *M. comscasis* WC007^T in inorganic medium supplemented with cellulose, pectin,
 227 fucoidan, mannan, xylan or starch polysaccharides (all at a final concentration of 1 g/L).
 228 with a total of 82 unique *susC/D* gene pairs (Figure 3A, Figure 3-figure supplement
 229 1-6, Supplementary file 8). Every PUL carried one *susC/D* pair and one or more
 230 CAZyme encoding gene(s). Of these 82 PULs, 44 were linked to either dedicated
 231 polysaccharide or polysaccharide classes according to in-depth annotations. The
 232 remaining 38 PULs could not be definitively inferred. Surprisingly, 58 of these PULs
 233 comprised sigma-70/anti-sigma factors, a rare observation in Bacteroidetes PULs,
 234 suggesting potentially novel regulation pathways exist in deep-sea Bacteroidetes.
 235 Substrate predictions suggested these PULs likely degraded six types of
 236 polysaccharides including: cellulose, pectin, fucoidan, mannan, xylan and starch
 237 (Figure 3A and Supplementary file 9). These were further described in the
 238 Supplementary results.

239 To verify the predicted polysaccharide substrate spectra of *M. comscasis*
 240 WC007^T, we grew WC007^T in medium containing various polysaccharides as the sole
 241 organic carbon source. Unsurprisingly, all six polysaccharides (cellulose, pectin,
 242 fucoidan, mannan, xylan and starch) promoted growth of WC007^T (Figure 3B).
 243 However, cellulose had an almost 7-fold increased growth compared to the control

group (Figure 3B). From these findings we conclude that *M. comscasis* WC007^T possesses versatile polysaccharide degradation capabilities.

Transcriptomic profiling of polysaccharide degradation and utilization by *M. comscasis* WC007^T. To further validate the roles of predicted PULs and polysaccharides degradation, we performed transcriptomic analysis of *M. comscasis* WC007^T during growth in medium containing different polysaccharides. For consistency, we focused on significantly up-regulated PULs and expression changes of genes encoding CAZymes, SusC/SusD, sigma-70 factor and anti-sigma factor within the corresponding PULs. We found, respectively, there were 2, 4, 5, 1, 4, 3 up-regulated PULs in cellulose (Figure 4-figure supplement 1), pectin (Figure 4-figure supplement 2), fucoidan (Figure 4-figure supplement 3), mannan (Figure 4-figure supplement 4), xylan (Figure 4-figure supplement 5) and starch (Figure 4-figure supplement 6) treatment groups. PUL 29 was expressed in three polysaccharides treatment groups (pectin, fucoidan and starch), PUL 54 was expressed in both pectin and starch treatment groups, PUL 65 was expressed in both pectin and xylan treatment groups, and the remaining PULs were only presented in a single polysaccharide treatment group. It is worth noting that only four PULs were consistent with predicted and detected activities. These included PUL 73 (pectin degradation), PUL 26 (mannan degradation), PUL 27 and PUL 76 (xylan degradation). These results suggest a single PUL might function in degrading different substrates.

In future studies, it will be important to clarify the specific activity of each PUL in the genome of WC007^T.

Figure 4. Transcriptomic analysis of *M. comscasis* WC007^T cultured in inorganic medium supplemented with different polysaccharides. Heat map showing differentially expressed genes encoding glycoside hydrolase genes (A) and SusC, SusD, sigma-70 factor and anti-sigma factor (B) in *M. comscasis* WC007^T supplemented with 1 g/L of different polysaccharides. Genes whose expression was up-regulated in three or more polysaccharides treatment groups are shown. The heat map is generated by the Heml 1.0.3.3 software.

To further characterize the dynamic changes of individual genes within WC007^T PULs under different treatment conditions, we analyzed the glycoside hydrolases, which were up-regulated simultaneously in three or more polysaccharides treatment groups. Genes encoding GH2, GH3, GH20, GH27, GH28, GH43, GH92 and GH95 family enzymes were selected (Figure 4A), as they are proposed to function in

different polysaccharides degradation functions. We also analyzed some annotated glycoside hydrolases outside predicted PULs. We found that expression of the same GH could be induced by different polysaccharides, such as alpha-L-fucosidase (GM_418 30400), GH43 (GM_418 28935) and glycoside hydrolase (GM_418 21515). Additionally, glycoside hydrolase (GM_418 21515), whose expression was significantly up-regulated by all six polysaccharides, may play a key role in mediating polysaccharides degradation of WC007^T. Future studies should help determine the exact function of this hydrolase. We also observed upregulation of genes encoding glycoside hydrolases outside predicted PULs, suggesting cryptic PULs may exist in WC007^T and may aid in the breakdown of glycans. Upon analysis of up-regulated genes encoding SusC/SusD and sigma-70 factor/anti-sigma factor located in the same PUL, we found 8 PULs were simultaneously induced in 3 or more polysaccharides treated groups (Figure 4B). The genes encoding sigma-70 factor (GM418_00730) and SusC (GM418_09255) were consistently up-regulated by 6 different polysaccharides, indicating their universal functions in polysaccharide degradation.

Metabolomics analysis of cellulose degradation by *M. comscasis* WC007^T. Of the 6 degradation polysaccharides, cellulose promoted the most growth of *M. comscasis* WC007^T (Figure 3B). From this, we further investigated the degradation and utilization capabilities of cellulose by *M. comscasis* WC007^T. We first observed degradation of cellulose by *M. comscasis* WC007^T by SEM. In the control group the

298
 299 **Figure 5.** Characterization of cellulose degradation by *M. comscasis* WC007^T. (A) SEM
 300 observation of cellulose. (B-D) SEM observation of cellulose (1 g/L) degradation by *M. comscasis*
 301 WC007^T. Scale bar is 10 μ m. (E) PUL 32 genes are significantly up-regulated when culturing *M.*
 302 *comscasis* WC007^T in the presence of 1 g/L cellulose. (F) qRT-PCR validation of gene expression
 303 changes observed in panel e.
 304 morphological surface of cellulose was uniform and intact (Figure 5A); however, in
 305 the *M. comscasis* WC007^T group, cellulose was clearly degraded into small pieces,

306 and a large amount of bacterial cells were attached to and even embedded in cellulose
307 particles (Figure 5B-D). Transcriptomic analysis also showed PUL 32 was
308 significantly up-regulated (up to 256-fold) when the bacterium was incubated in the
309 medium supplemented with cellulose (Figure 5E). Additionally, all of genes within
310 PUL 32 were significantly up-regulated, as detected by qRT-PCR (Figure 5F).
311 Overall, our above results suggest that *M. comscasis* WC007^T can efficiently degrade
312 cellulose for cellular energy and growth.

313 To further understand the mechanisms of cellulose degradation by *M. comscasis*
314 WC007^T and the effect of cellulose degradation on bacterial growth, we performed
315 metabolomic analysis of *M. comscasis* WC007^T cultured with or without cellulose.
316 Using the criteria of an OPLS-DA model VIP > 1 and a *P*-value < 0.05, 65 different
317 metabolites were identified when comparing control and cellulose treatment groups
318 (Supplementary file 10). To better understand the candidate metabolites and the
319 relationship between metabolites and differential expression patterns, we conducted a
320 hierarchical cluster analysis based on different metabolites. A heat map of the 65
321 metabolites shows the production of most detected metabolites (55 vs 65) were
322 up-regulated and a few metabolites (10 vs 65) were down-regulated (Figure 6). The
323 up-regulated metabolites were closely associated with amino acid metabolism,
324 carotenoid biosynthesis, fatty acid biosynthesis, and other cellular processes. The top
325 20 enriched pathways were further investigated by KEGG analysis (Figure 6-figure
326 supplement 1), which showed aminoacyl-tRNA biosynthesis, biosynthesis of amino

327

328 **Figure 6.** Metabonomic analysis of *M. comscasis* WC007^T cultured in inorganic medium

329 supplemented with cellulose. Differential metabolites heatmap showing accumulation of different

330 metabolites in *M. comscasis* WC007^T cultured with 1 g/L cellulose. “Up” indicates metabolites are

331 up-regulated in *M. comscasis* WC007^T cultured with cellulose. “Down” indicates metabolites that

are down-regulated in *M. comscasis* WC007^T cultured with cellulose. “Control 1-6” and “Cellulose 1-6” represent six different biological replicates. The color key scale indicates the abundance of metabolites.

acids and 2-Oxocarboxylic acid metabolism pathways were enriched with statistical significance (P -value < 0.05). This metabolomic and growth observation data (Figure 3B) as well as transcriptomic analysis (Figure 5) suggest degradation of cellulose promotes the metabolism of different amino acids and other key factors related to energy production.

Combined analysis of transcriptomic and metabolomic data during cellulose

degradation by *M. comscasis* WC007^T. To better understand the correlation between our transcriptomic and metabolomic data of *M. comscasis* WC007^T cultured with cellulose, we selected common metabolic pathways for further analysis, based on gene expression differences and metabolite production (Supplementary file 11). A total of 18 pathways were enriched, of which 4 pathways were enriched significantly. These 4 pathways were biosynthesis of amino acids, carbon metabolism, arginine biosynthesis and purine metabolism (Figure 7-figure supplement 1). These metabolic pathways are mainly involved in energy production and likely contribute to the growth of *M. comscasis* WC007^T. Major metabolites involved in saccharide metabolism, amino acids metabolism and tricarboxylic acid (TCA) cycle were increased, and most of the key related enzymes were also up-regulated (Figure 7).

Figure 7. Combined analysis of the *M. comscasis* WC007^T transcriptome and metabolome while cultured with cellulose. Metabolite and gene expression changes in three biosynthesis pathways (metabolism of saccharides and amino acids, and TCA cycle) from the combined analysis of transcriptomic and metabolomics data. Enzymes involved in the above pathways are marked in red. The relative content of each metabolite identified in *M. comscasis* WC007^T is shown as a heatmap from low (yellow) to high (red) as represented by the color scale. Similarly, gene expression levels for each enzyme associated with the above pathways are shown as a heatmap from low (blue) to high (red) as represented by the color scale. Two columns for each metabolite or gene represent the control (left) and the cellulose treatment group (right), respectively.

Investigating saccharide metabolism, we found endoglucanase, beta-glucosidase and xylose isomerase gene expression was increased. These enzymes were associated with degradation of cellulose to D-glucose, which is then converted to D-fructose. It is worth noting that crucial amino acids are also increased, including: L-arginine,

366 L-glutamate, L-threonine, homogentisic acid and guanidoacetic acid. These amino
367 acids could produce intermediates required for the TCA cycle (Figure 7).
368 Correspondingly, aconitate hydratase, isocitrate dehydrogenase (NADP⁺),
369 oxoglutarate dehydrogenase and succinate dehydrogenase gene expression was
370 up-regulated. This may be coordinated to enhance the biosynthesis and degradation of
371 isocitric acid, oxoglutaric acid, succinic acid and fumaric acid, and eventually
372 facilitate the TCA cycle and energy production. Overall, the observed changes in
373 metabolite concentrations were consistent with enzyme gene expression changes,
374 further supporting our transcriptomic and metabolomic results and the contribution of
375 cellulose to bacterial growth.

376 **Discussion**

377 Carbohydrates are a valuable natural resource and account for roughly 75% of the
378 biomass on Earth. Different microorganisms have developed unique strategies to
379 utilize this chemical energy for survival (Thomas et al., 2011). Bacteroidetes,
380 globally distributed in human gut, coastal ocean, marine sediments and other
381 environments, play key roles in the degradation of high molecular weight
382 carbohydrates (CottrellKirchman, 2000; Fernandez-Gomez et al., 2013; Thomas
383 et al., 2011). As such, Bacteroidetes are recognized as one of the dominant phyla of
384 ocean bacterioplankton (Diez-Vives et al., 2019) and are the most important
385 decomposer for algae derived carbohydrates, actively driving the ocean carbon and
386 nutrient cycling (Kruger et al., 2019; Unfried et al., 2018). As compared to studies

387 of Bacteroidetes derived from the shallow ocean, very little research has been done on
388 the deep-sea counterparts, likely due to the difficulties of sampling and cultivation.

389 In our current study, we first investigate the abundance of Bacteroides in the
390 deep-sea cold seep through metagenomic analysis. We find that similar to other
391 environments, Bacteroides are the second most abundant bacteria in the deep-sea
392 biotope (Figure 1A). We also find the number of genes encoding CAZymes in
393 Bacteroides is significantly higher than that in other deep-sea bacteria (Figure 1B and
394 1C), suggesting this phylum plays an essential role in carbohydrate degradation and
395 even the carbon cycle of this habitat. Many Bacteroides strains have been isolated
396 from common environments, however, there are few pure cultures obtained from
397 deep-sea habitats. To enrich for deep-sea Bacteroides, we perform an anaerobic
398 enrichment step, supplementing different polysaccharides in the inorganic medium.
399 As expected, we enrich and isolate a novel species of Bacteroides (*M. comscasis*
400 WC007^T) capable of degrading various polysaccharides. It is worth noting that this
401 novel species prefers the anaerobic growth conditions of its native deep-sea habitat
402 and duplicates very slowly in aerobic conditions. Most importantly, we develop an
403 efficient goal-directed method to enrich and isolate Bacteroides from deep-sea
404 environments (Figure 2A), a method which will likely be useful for isolating
405 Bacteroides from other environments in the future.

406 *M. comscasis* WC007^T contains numerous glycoside hydrolases (374 GHs) and
407 82 pairs SusC/D-like protein genes (Figure 2C and 2D). This gene abundance is far

408 more than the number identified in *Bacteroides* isolated from other environments
 409 (Figure 2 – figure supplement 3). We predict *M. comscasis* WC007^T has a total of 82
 410 PULs and propose these PULs function to degrade six different polysaccharides
 411 (cellulose, pectin, fucoidan, mannan, xylan and starch) (Figure 3A). We validate these
 412 findings using growth assays (Figure 3B). The biggest difference between PULs
 413 identified in other *Bacteroides* and *M. comscasis* WC007^T is the large number of
 414 sigma/anti-sigma factors in the PULs of *M. comscasis* WC007^T (Figure 3 – figure
 415 supplement 1-6). We find that the expression of these sigma/anti-sigma factors
 416 changes with different polysaccharide supplements (Figure 4B), suggesting they
 417 function for polysaccharides degradation by *M. comscasis* WC007^T. Transport and
 418 degradation of polysaccharides by *M. comscasis* WC007^T depend on typical
 419 TonB-dependent regulatory systems (**Koebnik, 2005**). These systems consist of a
 420 specialized outer membrane-localized TonB-dependent receptor (such as SusC) that
 421 interacts with its energizing TonB-ExbBD protein complex, a cytoplasmic
 422 membrane-localized anti-sigma factor and an extracytoplasmic function
 423 (ECF)-subfamily sigma factor. This sophisticated complex senses signals from
 424 outside the bacterial cell and transmits them via two membranes into the cytoplasm,
 425 leading to transcriptional activation of target genes (**Koebnik, 2005**). In Bacteroidetes,
 426 predictions have been made for the existence of sigma/anti-sigma factors in the PULs,
 427 but few PULs have been verified for their polysaccharide degradation activity
 428 (**Koebnik, 2005**). Our transcriptomic results clearly suggest polysaccharide

429 degradation and utilization in *M. comscasis* WC007^T is strictly regulated by
430 sigma/anti-sigma factors and is mediated by the TonB-dependent regulatory systems.
431 This may be a universal mechanism for deep-sea Bacteroidetes and warrants further
432 investigation.

433 From our metagenomic, transcriptomic and metabolomic results, we propose *M.*
434 *comscasis* WC007^T and other Bacteroidetes are major contributors to polysaccharide
435 degradation in the deep-sea environment. Polysaccharide degradation by
436 heterotrophic microbes is known to be a key process in Earth's carbon and nutrient
437 cycling (Unfried et al., 2018). Notably, during cellulose degradation, *M. comscasis*
438 WC007^T contributes to the carbon cycle not only by promoting the saccharide and
439 amino acid metabolism but also by facilitating the urea cycle and methane metabolism,
440 the most common carbon source for cold seep microorganisms (Figure 7). Therefore,
441 we propose Bacteroidetes play an essential role in the carbon and nutrient cycling of
442 the deep sea, owing to their polysaccharide degradation functions and biological
443 abundance.

444 Algal polysaccharides are an important bacterial nutrient source and a central
445 component of marine food webs. They are also key contributors to the carbon cycling
446 at the ocean surface and the deep ocean (Kabisch et al., 2014; Kappelmann et al.,
447 2019; Kruger et al., 2019). Given the carbon cycling associated with algal
448 polysaccharides in the whole ocean ecosystem, it is reasonable to imagine the most
449 easily degraded polysaccharides are utilized by aerobic microorganisms at the ocean

450 surface (Figure 8). The more difficult-to-degrade polysaccharides aggregate to form
 451 particulate detritus and these sinking particles become hotspots for organic carbon,
 452 which subsequently play a key role in the export of matter from the euphotic

453
 454 **Figure 8.** Proposed model of polysaccharide degradation by Bacteroides and contributions to the
 455 ocean carbon and nutrient cycling. SusC, starch utilization system subunit C (TonB-dependent
 456 receptor/transporter); SusD, starch utilization system subunit D (outer membrane sugar-binding
 457 protein); CAZymes, carbohydrate-active enzymes; GH, glycoside hydrolases; GT, glycosyl

458 transferases; PL, polysaccharide lyases; CBM, carbohydrate-binding modules; anti- σ , anti-sigma
 459 factor; ECF σ , ECF-subfamily sigma factor; RNAPol, RNA polymerase; TonB, a kind of typical
 460 periplasmic protein; ExbB and ExbD, plasma membrane proteins providing energy to TonB; CM,
 461 cytoplasmic membrane; OM, outer membrane.

462 surface to the deep ocean sediment (**AzamMalfatti, 2007**). Once the
 463 carbohydrate-derived particles reach the deep sea, efficient degraders like *M.*
 464 *comscasis* WC007^T initially recognize the extracellular polysaccharides or
 465 oligosaccharides via TonB-dependent outer membrane transporters (SusC and SusD
 466 complex). SusC then interacts with TonB through the TonB box, providing energy
 467 through the TonB-ExbB-ExbD complex located on the cytoplasmic membrane,
 468 allowing these degraders to uptake the oligosaccharides into the periplasm.
 469 Meanwhile, TonB-dependent transducers possess a unique N-terminal extension that
 470 interacts with the anti-sigma factor and promotes ECF sigma factor binding to the
 471 RNA polymerase core enzyme. CAZymes are then activated and released to cleave
 472 oligosaccharides into monosaccharide and disaccharides, which are then transported
 473 into the cytoplasm for metabolism and energy production to promote bacterial growth.
 474 Notably, TonB-dependent uptake systems are sometimes semi-selfish (**Cuskin et al.,**
 475 **2015**), with some cleavage products becoming available to other bacteria, which
 476 could greatly promote the carbon and nutrient cycling in the deep biosphere.

477 **Materials and Methods**

478 **Sampling and operational taxonomic units (OTUs) analysis.** The deep-sea
 479 sediment sample was collected by *RV KEXUE* from a typical cold seep in the South
 480 China Sea (E 119°17'07.322", N 22°06'58.598") at a depth of approximately 1,146 m
 481 in July of 2018 as described previously (**Zhang et al., 2020**). In order to understand
 482 the abundance of Bacteroidetes in deep-sea cold seep, operational taxonomic units
 483 (OTUs) analysis was performed. Briefly, total genomic DNA extracted from the
 484 sample was diluted to 1 ng/μL using sterile water and used for PCR template.
 485 Bacterial 16S rRNA genes of distinct regions (16S V4/V5) were amplified using
 486 specific primers (515F-GTGCCAGCMGCCGCGG and
 487 907R-CCGTCAATTCMTTTRAGTTT) with the barcode, and the PCR products were
 488 purified with Qiagen Gel Extraction Kit (Qiagen, Germany) for library construction.
 489 Sequencing library was generated using TruSeq® DNA PCR-Free Sample
 490 Preparation Kit (Illumina, USA) following manufacturer's instructions. And
 491 sequences analyses were performed by Uparse software (**Edgar, 2013**). Sequences
 492 with ≥97% similarity were assigned to the same OTU. Representative sequence for
 493 each OTU was screened for further annotation. For each representative sequence, the
 494 Silva Database (**Quast et al., 2013**) was used based on Mothur algorithm to annotate
 495 taxonomic information.

496 **Metagenomic sequencing, assembly, binning and annotation.** For metagenomic
 497 sequencing, total DNA from the sample (20 g) was extracted using the Qiagen
 498 DNeasy® PowerSoil® Pro Kit (Qiagen, Hilden, Germany) and the integrity of DNA

499 was evaluated by gel electrophoresis. Then 0.5 µg DNA of each sample was used to
500 prepare libraries. Libraries were prepared with an amplification step for each sample.
501 DNA was cleaved into 50 ~ 800 bp fragments using Covaris E220 ultrasonicator
502 (Covaris, Brighton, UK) and some fragments between 150 ~ 250 bp were selected
503 using AMPure XP beads (Agencourt, Beverly, MA, USA) and repaired using T4
504 DNA polymerase (Enzymatics, Beverly, MA, USA). All NGS sequencing was
505 performed on the BGISEQ-500 platform (BGI, Qingdao, China), generating 100 bp
506 paired-end raw reads. Quality control was performed by SOAPnuke (v1.5.6) (setting:
507 -l 20 -q 0.2 -n 0.05 -Q 2 -d -c 0 -5 0 -7 1) (**Chen et al., 2017**) and the clean data were
508 assembled using MEGAHIT (v1.1.3) (setting:--min-count 2 --k-min 33 --k-max 83
509 --k-step 10) (**Li et al., 2015**). Maxbin2 (**Wu et al., 2016**), metaBAT2 (**Kang et al.,**
510 **2019**) and Concoct (**Alneberg et al., 2014**) were used to automatically bin from
511 assemblies of these samples. MetaWRAP (**Uritskiy et al., 2018**) was used to purify
512 and organize data to generate the final bins. Finally, completeness and contamination
513 of metagenome-assembled genomes (MAGs) were assessed using checkM (v1.0.18)
514 (**Parks et al., 2015**). These obtained MAGs were subsequently annotated by
515 searching the predicted genes against NR (20180814), KEGG (Release 87.0), COG
516 (update-2018_08) and Swissprot (release-2017_07). Additionally, in order to search
517 for CAZymes genes, CAZymes were annotated based on HMMer, DIAMOND and
518 Hotpep searches of the dbCAN2 meta server (release-2019_07), a web server for

519 automated carbohydrate-active enzyme annotation. CAZymes were annotated and
520 selected only as such when at least two of the searches were positive.

521 **Isolation and cultivation of deep-sea bacteria.** To enrich the bacteria degrading
522 polysaccharide, the sediment sample was cultured at 28 °C for one month in an
523 anaerobic enrichment medium containing (per litre of seawater): 1.0 g NH₄Cl, 1.0 g
524 NaHCO₃, 1.0 g CH₃COONa, 0.5 g KH₂PO₄, 0.2 g MgSO₄·7H₂O, 1.0 g polysaccharide
525 (cellulose, pectin or xylan), 0.7 g cysteine hydrochloride, 500 µL 0.1 % (w/v)
526 resazurin and pH 7.0. This medium was prepared anaerobically as previously
527 described (Fardeau et al., 1997). A 100 µL enrichment was spread on Hungate tube
528 with modified medium (1 g yeast extract, 1 g peptone, 1 g NH₄Cl, 1 g NaHCO₃, 1 g
529 CH₃COONa, 0.5 g KH₂PO₄, 0.2 g MgSO₄·7H₂O, 0.7 g cysteine hydrochloride, 500
530 µL 0.1 % (w/v) resazurin, 1,000 mL seawater, 15 g agar, pH 7.0) after 10,000 times
531 dilution, and this medium was named ORG in this study. The Hungate tubes were
532 anaerobically incubated at 30 °C for 7 days. Individual colonies with distinct
533 morphology were picked using sterilized bamboo sticks and then cultured in the ORG
534 broth. *Maribellus comscasis* WC007^T was isolated and purified with ORG medium by
535 repeated use of the Hungate roll-tube methods for several rounds until it was
536 considered to be axenic. The purity of strain WC007^T was confirmed routinely by
537 transmission electron microscopy (TEM) and repeated partial sequencing of the 16S
538 rRNA gene. Strain WC007^T is preserved at -80 °C in ORG broth supplemented with
539 20% (v/v) glycerol.

540 **Electron microscopy observation.** To observe the morphological characteristics of
 541 *M. comscasis* WC007^T, cells were examined using transmission electronic
 542 microscopy (TEM) (HT7700; Hitachi, Japan) with a JEOL JEM 12000 EX (equipped
 543 with a field emission gun) at 100 kV. The cells suspension of *M. comscasis* WC007^T
 544 was washed with Milli-Q water and centrifuged at 5,000 g for 5 min. Subsequently,
 545 the sample was taken by immersing copper grids coated with a carbon film for 20 min
 546 in the bacterial suspensions and washed for 10 min in distilled water and dried for 20
 547 min at room temperature(Han et al., 2014).

548 To detect the degradation of cellulose by *M. comscasis* WC007^T, 10 µL sample
 549 was dripped on cover slips and soaked in gelatin and dried for 30 min to allow the
 550 sample to adhere to the surface of copper grid. These samples were fixed in 2.5%
 551 glutaraldehyde for 30 min. Samples were then washed three times with PBS and
 552 dehydrated in ethanol solutions of 30%, 50%, 70%, 90% and 100% for 10 min each
 553 time. All samples were observed with scanning electronic microscopy (SEM)
 554 (S-3400N, Hitachi, Japan) at 5 kV.

555 **Phylogenetic analysis.** Phylogenetic trees were constructed with the full-length 16S
 556 rRNA sequences by the neighbor-joining algorithm (SaitouNei, 1987), maximum
 557 Likelihood (Felsenstein, 1981) and minimum-evolution methods (RzhetskyNei,
 558 1992). 16S rRNA sequences of *M. comscasis* WC007^T (accession number MT460676)
 559 and other related taxa used for phylogenetic analysis were obtained from NCBI
 560 GenBank. Phylogenetic analysis was performed using the software MEGA version

561 6.0 (Tamura et al., 2013). And a genome-based phylogenetic tree was reconstructed
562 using the TYGS algorithm (<https://tygs.dsmz.de/>) (Meier-KolthoffGoker, 2019).

563 **Genomic characterizations of *M. comscasis* WC007^T.** Genome relatedness values
564 were calculated by multiple approaches: Average Nucleotide Identity (ANI) based on
565 the MUMMER ultra-rapid aligning tool (ANIm), ANI based on the BLASTN
566 algorithm (ANIB), tetranucleotide signatures (TETRA), and *in silico* DNA-DNA
567 similarity. ANIm, ANIB, and TETRA frequencies were calculated using JSpecies WS
568 (<http://jspecies.ribohost.com/jspeciesws/>) (Richter et al., 2016). Recommended
569 species criteria cut-offs were used: 95% for ANIB and ANIm and 0.99 for TETRA
570 signature (RichterRossello-Mora, 2009). Amino acid identity (AAI) values were
571 calculated by AAI-profiler (<http://ekhidna2.biocenter.helsinki.fi/AAI/>) (Medlar et al.,
572 2018). *In silico* DNA-DNA similarity values were calculated by the
573 Genome-to-Genome Distance Calculator (GGDC) (<http://ggdc.dsmz.de/>)
574 (Meier-Kolthoff et al., 2013). The *isDDH* results were based on the recommended
575 formula 2, which is independent of genome size and is robust when using
576 whole-genome sequences.

577 **PULs prediction and annotation.** PULs of *M. comscasis* WC007^T were detected
578 based on the presence of *susC*-like and *susD*-like gene pairs, which in most cases was
579 also accompanied by the presence of CAZyme clusters, as previously suggested
580 (Bjursell et al., 2006). Additional consideration required a PUL to have at least one
581 *susC*- and *susD*-like gene pair and encode at least one degradative CAZymes from the

GH families. PUL genes were predicted using a combination of HMMer searches against the Pfam (Finn et al., 2014), TIGRFAM (Selengut et al., 2007), dbCAN (Yin et al., 2012), MEROPS (Rawlings et al., 2012) and CAZy (Lombard et al., 2014) databases. CAZymes were annotated based on HMMer searches against the Pfam v25 and dbCAN 3.0 databases and BLASTp searches (Altschul et al., 1990) against the CAZy database. SusC-like protein (TonB-linked outer membrane protein) and SusD-like protein (starch-binding associating with outer membrane) were annotated by the DOE-JGI Microbial Annotation Pipeline (MGAP) (Huntemann et al., 2016), which used the TIGRFam model (TIGR04056) to detect SusC-like proteins and the Pfam models (PF12741, PF12771, and PF14322) to detect SusD-like proteins (Terrapon et al., 2015). Sulfatases were annotated by the SulfAtlas database v1.0 (Barbeyron et al., 2016). Peptidases were annotated based on the BLASTp searches against the MEROPS 9.13 database using the default settings of $E \leq 10^{-4}$.

Growth assay of *M.comscasis* WC007^T. Growth assays were performed at atmospheric pressure. Briefly, 100 μ L *M.comscasis* WC007^T culture was inoculated in Hungate tubes containing 10 mL inorganic medium (1.0 g/L NH_4Cl , 1.0 g/L NaHCO_3 , 1.0 g/L CH_3COONa , 0.5 g/L KH_2PO_4 , 0.2 g/L $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 0.7 g/L cysteine hydrochloride, 500 μ L/L of 0.1 % (w/v) resazurin and pH 7.0) with different polysaccharides (1.0 g/L cellulose, 1.0 g/L pectin, 0.1 g/L fucoidan, 0.1 g/L mannan, 1.0 g/L xylan or 1.0 g/L starch), respectively. The Hungate tubes were anaerobically incubated at 30 °C for 10 days. Since *M.comscasis* WC007^T grew very slowly in the

603 inorganic culture medium, we adopted the colony forming units (CFU) to measure the
604 growth curve of *M.comscasis* WC007^T in an anaerobic cabinet.

605 **Transcriptomics analysis.** Transcriptomics analysis was performed by Novogene
606 (Tianjin, China). Cells suspension of *M. comscasis* WC007^T cultured in inorganic
607 medium for 7 days was used as a control, while cell suspensions cultured in inorganic
608 medium supplemented with different polysaccharides (1.0 g/L cellulose, 1.0 g/L
609 pectin, 0.1 g/L fucoidan, 0.1 g/L mannan, 1.0 g/L xylan or 1.0 g/L starch) for 7 days
610 were the experimental groups. All cultures were incubated in 2 L anaerobic bottles.
611 For transcriptomics analyses, total *M. comscasis* WC007^T RNA was extracted using
612 TRIzol reagent (Invitrogen, USA) and DNA contamination was removed using the
613 MEGA clear™ Kit (Life technologies, USA). Detailed protocols of the following
614 procedures including library preparation, clustering and sequencing and data analyses
615 were described in the Supplementary information.

616 **Metabolomics analysis.** Metabolomics analysis was performed by Genedenovo
617 Biotechnology (Guangzhou, China). The samples for metabolomics analysis were
618 obtained from the culture supernatant used for transcriptomic analysis, where *M.*
619 *comscasis* WC007^T was cultured in inorganic medium supplemented with 1.0 g/L
620 cellulose for 7 days. Briefly, 100 mg supernatant was taken and placed in containers,
621 extraction liquid containing an internal target was added and the mixture was
622 homogenized in ball mill kept in ice water for 4 min at 45 Hz with ultrasound
623 treatment. After 3 rounds of homogenization, the samples were incubated for 1 h at

624 -20 °C to precipitate proteins. Samples were then centrifuged at 12,000 g for 15 min
625 at 4 °C and the supernatant was transferred into fresh tubes. Extracts were dried in a
626 vacuum concentrator without heating and extraction liquid was added for
627 reconstitution. Finally, the extracts were shaken for 30 s and centrifuged at 12,000 g
628 for 15 min at 4 °C. Supernatants were transferred into 2 mL LC/MS glass vials for the
629 LC-MS/MS analysis by using an UHPLC system (Agilent Technologies 1290 with a
630 UPLC BEH Amide column) coupled to TripleTOF 6600 (Q-TOF, AB Sciex).

631 Metabolite identification was performed based on public metabolite databases,
632 such as HMDB (<https://www.hmdb.ca/>) (Wishart et al., 2013), METLIN
633 (<https://metlin.scripps.edu/index.php>) (Zhu et al., 2013), MassBank
634 (<https://www.massbank.jp/>), LipidMaps (<http://www.lipidmaps.org>), mzCloud
635 (<https://www.mzcloud.org>). Principle component analysis (PCA), partial least squares
636 discriminant analysis (PLS-DA) and hierarchical clustering analysis (HCA) were used
637 to integrate the obtained metabolite mass spectral peaks.

638 The threshold of variables determined to be important in the projection (VIP)
639 scores ≥ 1.0 together with a *P*-value of *T*-test < 0.05 were adopted to assess
640 significant different metabolites. All differentially expressed genes and metabolites
641 were mapped to the Kyoto Encyclopedia of Genes and Genomes (KEGG) KEGG
642 database for pathway and enrichment analysis (Okuda et al., 2008).

643 **Data availability.** The BioProject and BioSample accession numbers of
644 metagenome-assembled genomes (MAGs) are shown in the Supplementary file 1. The

complete genome sequence of *M. comscasis* WC007^T has been deposited at GenBank under the accession number CP046401. The raw sequencing reads for transcriptomic analysis have been deposited to NCBI Short Read Archive (accession number: PRJNA664312).

Description of *Maribellus comscasis* sp. nov.

Maribellus comscasis (com.sca'sis. L. gen. pl. n. *comscasis* from the Center for Ocean Mega-Science, Chinese Academy of Sciences).

Cells are Gram-stain-negative curve-shaped, 2.0-6.0 µm in length and 0.5-0.8 µm in width. Facultative anaerobic and oxidase-positive. The temperature range for growth is 28-37 °C with an optimum at 30 °C. Growing at pH values of 6.0-8.0 (optimum, pH 7.0). Growth occurs at NaCl concentrations between 0.0-5.0% with optimum growth at 1.0% NaCl. By analyzing the hydrolysis of polysaccharides, the growth is promoted significantly by cellulose, pectin and xylan. From the sole carbon source utilization test, growth is stimulated by acetate, maltose, fructose, lactate, sorbitol and D-mannose. Weak growth occurs with ethanol and formate. The major polar lipids are phosphatidylethanolamine, unidentified phospholipid, unidentified aminolipid, unidentified lipid. Containing significant proportions (>10%) of the cellular fatty acids iso-C_{15:0}, C_{16:0}, summed feature 3 (containing C_{16:1}ω7c and/or C_{16:1}ω6c) and summed feature 8 (containing C_{18:1}ω7c and/or C_{18:1}ω6c).

664 The type strain, WC007^T (=MCCC 1K04777^T = KCTC 25169^T), was isolated from
665 the sediment of deep-sea cold seep, P.R. China. The DNA G+C content of the type
666 strain is 38.4%.

667 **Acknowledgements**

668 This work was funded by the Strategic Priority Research Program of the Chinese
669 Academy of Sciences (Grant No. XDA22050301), China Ocean Mineral Resources
670 R&D Association Grant (Grant No. DY135-B2-14), National Key R and D Program
671 of China (Grant No. 2018YFC0310800), the Taishan Young Scholar Program of
672 Shandong Province (tsqn20161051), and Qingdao Innovation Leadership Program
673 (Grant No. 18-1-2-7-zhc) for Chaomin Sun. This study is also funded by the Open
674 Research Project of National Major Science & Technology Infrastructure (*RV*
675 *KEXUE*) (Grant No. NMSTI-KEXUE2017K01).

676 **Author contributions**

677 RZ and CS conceived and designed the study; RZ conducted most of the experiments;
678 RL and GL collected the samples from the deep-sea cold seep; RC helped to analyze
679 the metagenomes; RZ and CS lead the writing of the manuscript; all authors
680 contributed to and reviewed the manuscript.

681 **Conflict of interest**

682 The authors declare that there are no any competing financial interests in relation to
683 the work described.

References

- 684 **References**
- 685 **Aineberg J**, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson
- 686 AF, Quince C. 2014. Binning Metagenomic Contigs by Coverage and Composition. *Nat Methods*
- 687 **11**: 1144-1146.
- 688 **Alonso C**, Warnecke F, Amann R, Pernthaler J. 2007. High Local and Global Diversity of
- 689 *Flavobacteria* in Marine Plankton. *Environ Microbiol* **9**: 1253-1266.
- 690 **Altschul SF**, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic Local Alignment Search Tool. *J*
- 691 *Mol Biol* **215**: 403-410.
- 692 **Azam F**. 1998. Microbial Control of Oceanic Carbon Flux: The Plot Thickens. *Science* **280**: 694-696.
- 693 **Azam F**, Malfatti F. 2007. Microbial Structuring of Marine Ecosystems. *Nat Rev Microbiol* **5**:
- 694 782-791.
- 695 **Barbeyron T**, Brillet-Gueguen L, Carre W, Carriere C, Caron C, Czjzek M, Hoebeke M, Michel G.
- 696 2016. Matching the Diversity of Sulfated Biomolecules: Creation of a Classification Database for
- 697 Sulfatases Reflecting Their Substrate Specificity. *Plos One* **11**: e0164846.
- 698 **Bauer M**, Kube M, Teeling H, Richter M, Lombardot T, Allers E, Wurdemann CA, Quast C, Kuhl H,
- 699 Knaust F, Woebken D, Bischof K, Musmann M, Choudhuri JV, Meyer F, Reinhardt R, Amann
- 700 RI, Glockner FO. 2006. Whole Genome Analysis of the Marine Bacteroidetes '*Gramella Forsetii*'
- 701 Reveals Adaptations to Degradation of Polymeric Organic Matter. *Environ Microbiol* **8**:
- 702 2201-2213.
- 703 **Benner R**, Pakulski JD, Mccarthy M, Hedges JI, Hatcher PG. 1992. Bulk Chemical Characteristics of
- 704 Dissolved Organic-Matter in the Ocean. *Science* **255**: 1561-1564.
- 705 **Bjursell MK**, Martens EC, Gordon JI. 2006. Functional Genomic and Metabolic Studies of the
- 706 Adaptations of a Prominent Adult Human Gut Symbiont, *Bacteroides Thetaiotaomicron*, to the
- 707 Suckling Period. *J Biol Chem* **281**: 36269-36279.
- 708 **Cantarel BL**, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. 2009. The
- 709 Carbohydrate-Active Enzymes Database (Cazy): An Expert Resource for Glycogenomics. *Nucleic*
- 710 *Acids Res* **37**: D233-D238.

711 **Chen YX**, Chen YS, Shi CM, Huang ZB, Zhang Y, Li SK, Li Y, Ye J, Yu C, Li Z, Zhang XQ, Wang J,
712 Yang HM, Fang L, Chen Q. 2017. Soapnuke: A Mapreduce Acceleration-Supported Software for
713 Integrated Quality Control and Preprocessing of High-Throughput Sequencing Data. *Gigascience*
714 **7**.

715 **Cottrell MT**, Kirchman DL. 2000. Natural Assemblages of Marine Proteobacteria and Members of the
716 *Cytophaga-Flavobacter* Cluster Consuming Low- and High-Molecular-Weight Dissolved Organic
717 Matter. *Appl Environ Microb* **66**: 1692-1697.

718 **Cuskin F**, Lowe EC, Temple MJ, Zhu YP, Cameron EA, Pudlo NA, Porter NT, Urs K, Thompson AJ,
719 Cartmell A, Rogowski A, Hamilton BS, Chen R, Tolbert TJ, Piens K, Bracke D, Vervecken W,
720 Hakki Z, Speciale G, Munoz-Munoz JL, Day A, Pena MJ, McLean R, Suits MD, Boraston AB,
721 Atherly T, Ziemer CJ, Williams SJ, Davies GJ, Abbott DW, Martens EC, Gilbert HJ. 2015.
722 Human Gut Bacteroidetes Can Utilize Yeast Mannan through a Selfish Mechanism. *Nature* **517**:
723 165-169.

724 **Diez-Vives C**, Nielsen S, Sanchez P, Palenzuela O, Ferrera I, Sebastian M, Pedros-Alio C, Gasol JM,
725 Acinas SG. 2019. Delineation of Ecologically Distinct Units of Marine Bacteroidetes in the
726 Northwestern Mediterranean Sea. *Mol Ecol* **28**: 2846-2859.

727 **Edgar RC**. 2013. Uparse: Highly Accurate Otu Sequences from Microbial Amplicon Reads. *Nat*
728 *Methods* **10**: 996-998.

729 **Engel A**, Thoms S, Riebesell U, Rochelle-Newall E, Zondervan I. 2004. Polysaccharide Aggregation
730 as a Potential Sink of Marine Dissolved Organic Carbon. *Nature* **428**: 929-932.

731 **Fardeau ML**, Ollivier B, Patel BKC, Magot M, Thomas P, Rimbault A, Rocchiccioli F, Garcia JL.
732 1997. *Thermotoga Hypogea* Sp. Nov., a Xylanolytic, Thermophilic Bacterium from an
733 Oil-Producing Well. *Int J Syst Bacteriol* **47**: 1013-1019.

734 **Felsenstein J**. 1981. Evolutionary Trees from DNA-Sequences - a Maximum-Likelihood Approach. *J*
735 *Mol Evol* **17**: 368-376.

736 **Fernandez-Gomez B**, Richter M, Schuler M, Pinhassi J, Acinas SG, Gonzalez JM, Pedros-Alio C.
737 2013. Ecology of Marine Bacteroidetes: A Comparative Genomics Approach. *Isme J* **7**:
738 1026-1037.

739 **Field CB**, Behrenfeld MJ, Randerson JT, Falkowski P. 1998. Primary Production of the Biosphere:
740 Integrating Terrestrial and Oceanic Components. *Science* **281**: 237-240.

741 **Finn RD**, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K,
742 Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M. 2014. Pfam: The Protein Families
743 Database. *Nucleic Acids Res* **42**: D222-D230.

744 **Gao BL**, Jin M, Li L, Qu W, Zeng RY. 2017. Genome Sequencing Reveals the Complex
745 Polysaccharide-Degrading Ability of Novel Deep-Sea Bacterium *Flammeovirga Pacifica* Wpaga1.
746 *Front Microbiol* **8**: 38248.

747 **Glenwright AJ**, Pothula KR, Bhamidimarri SP, Chorev DS, Basle A, Firbank SJ, Zheng HJ, Robinson
748 CV, Winterhalter M, Kleinekathofer U, Bolam DN, van den Berg B. 2017. Structural Basis for
749 Nutrient Acquisition by Dominant Members of the Human Gut Microbiota. *Nature* **541**: 407-411.

750 **Grondin JM**, Tamura K, Dejean G, Abbott DW, Brumer H. 2017. Polysaccharide Utilization Loci:
751 Fueling Microbial Communities. *J Bacteriol* **199**: e00860-00816.

752 **Hahnke RL**, Meier-Kolthoff JP, Garcia-Lopez M, Mukherjee S, Huntemann M, Ivanova NN, Woyke
753 T, Kyrpides NC, Klenk HP, Goker M. 2016. Genome-Based Taxonomic Classification of
754 Bacteroidetes. *Front Microbiol* **7**: 2003.

755 **Han ZZ**, Yan HX, Zhao H, Zhou SX, Han M, Meng XQ, Zhang Y, Zhao YY, Sun B, Yao CK, Wang
756 YF, Wang CL, Li FB, Tian CC, Xu LL. 2014. Bio-Precipitation of Calcite with Preferential
757 Orientation Induced by *Synechocystis* Sp. Pcc6803. *Geomicrobiol J* **31**: 884-899.

758 **Hehemann JH**, Correia G, Barbeyron T, Helbert W, Czjzek M, Michel G. 2010. Transfer of
759 Carbohydrate-Active Enzymes from Marine Bacteria to Japanese Gut Microbiota. *Nature* **464**:
760 908-912.

761 **Huntemann M**, Ivanova NN, Mavromatis K, Tripp HJ, Paez-Espino D, Tennessen K, Palaniappan K,
762 Szeto E, Pillay M, Chen IMA, Pati A, Nielsen T, Markowitz VM, Kyrpides NC. 2016. The
763 Standard Operating Procedure of the Dore-Jgi Metagenome Annotation Pipeline (Map V.4). *Stand*
764 *Genomic Sci* **11**: 17.

765 **Kabisch A**, Otto A, Konig S, Becher D, Albrecht D, Schuler M, Teeling H, Amann RI, Schweder T.
766 2014. Functional Characterization of Polysaccharide Utilization Loci in the Marine Bacteroidetes
767 '*Gramella Forsetii*' Kt0803. *Isme J* **8**: 1492-1502.

768 **Kang DWD**, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. 2019. Metabat 2: An Adaptive
769 Binning Algorithm for Robust and Efficient Genome Reconstruction from Metagenome
770 Assemblies. *Peerj* **7**.

771 **Kappelmann L**, Kruger K, Hehemann JH, Harder J, Markert S, Unfried F, Becher D, Shapiro N,
772 Schweder T, Amann RI, Teeling H. 2019. Polysaccharide Utilization Loci of North Sea
773 *Flavobacteriia* as Basis for Using SusC/D-Protein Expression for Predicting Major Phytoplankton
774 Glycans. *Isme J* **13**: 76-91.

775 **Kirchman DL**. 2002. The Ecology of *Cytophaga-Flavobacteria* in Aquatic Environments. *Fems*
776 *Microbiol Ecol* **39**: 91-100.

777 **Kloareg B**, Quatrano RS. 1988. Structure of the Cell-Walls of Marine-Algae and Ecophysiological
778 Functions of the Matrix Polysaccharides. *Oceanogr Mar Biol* **26**: 259-315.

779 **Koebnik R**. 2005. Tonb-Dependent Trans-Envelope Signalling: The Exception or the Rule? *Trends*
780 *Microbiol* **13**: 343-347.

781 **Kruger K**, Chafee M, Ben Francis T, del Rio TG, Becher D, Schweder T, Amann RI, Teeling H. 2019.
782 In Marine Bacteroidetes the Bulk of Glycan Degradation During Algae Blooms Is Mediated by
783 Few Clades Using a Restricted Set of Genes. *Isme J* **13**: 2800-2816.

784 **Lapebie P**, Lombard V, Drula E, Terrapon N, Henrissat B. 2019. Bacteroidetes Use Thousands of
785 Enzyme Combinations to Break Down Glycans. *Nat Commun* **10**.

786 **Larsbrink J**, Rogers TE, Hemsworth GR, McKee LS, Tauzin AS, Spadiut O, Klintner S, Pudlo NA,
787 Urs K, Koropatkin NM, Creagh AL, Haynes CA, Kelly AG, Cederholm SN, Davies GJ, Martens
788 EC, Brumer H. 2014. A Discrete Genetic Locus Confers Xyloglucan Metabolism in Select Human
789 Gut Bacteroidetes. *Nature* **506**: 498-502.

790 **Li DH**, Liu CM, Luo RB, Sadakane K, Lam TW. 2015. Megahit: An Ultra-Fast Single-Node Solution
791 for Large and Complex Metagenomics Assembly Via Succinct De Bruijn Graph. *Bioinformatics*
792 **31**: 1674-1676.

793 **Lombard V**, Ramulu HG, Drula E, Coutinho PM, Henrissat B. 2014. The Carbohydrate-Active
794 Enzymes Database (Cazy) in 2013. *Nucleic Acids Res* **42**: D490-D495.

795 **Malhi Y**. 2002. Carbon in the Atmosphere and Terrestrial Biosphere in the 21st Century. *Philos T Roy*
796 *Soc A* **360**: 2925-2945.

797 **Medlar AJ**, Toronen P, Holm L. 2018. Aai-Profiler: Fast Proteome-Wide Exploratory Analysis
798 Reveals Taxonomic Identity, Misclassification and Contamination. *Nucleic Acids Res* **46**:
799 W479-W485.

800 **Meier-Kolthoff JP**, Auch AF, Klenk HP, Goker M. 2013. Genome Sequence-Based Species
801 Delimitation with Confidence Intervals and Improved Distance Functions. *Bmc Bioinformatics* **14**:
802 60.

803 **Meier-Kolthoff JP**, Goker M. 2019. Tygs Is an Automated High-Throughput Platform for
804 State-of-the-Art Genome-Based Taxonomy. *Nat Commun* **10**: 2182.

805 **Moran MA**, Kujawinski EB, Stubbins A, Fatland R, Aluwihare LI, Buchan A, Crump BC, Dorrestein
806 PC, Dyhrman ST, Hess NJ, Howe B, Longnecker K, Medeiros PM, Niggemann J, Obernosterer I,
807 Repeta DJ, Waldbauer JR. 2016. Deciphering Ocean Carbon in a Changing World. *P Natl Acad*
808 *Sci USA* **113**: 3143-3151.

809 **Okuda S**, Yamada T, Hamajima M, Itoh M, Katayama T, Bork P, Goto S, Kanehisa M. 2008. Kegg
810 Atlas Mapping for Global Analysis of Metabolic Pathways. *Nucleic Acids Res* **36**: W423-W426.

811 **Parks DH**, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. Checkm: Assessing the
812 Quality of Microbial Genomes Recovered from Isolates, Single Cells, and Metagenomes. *Genome*
813 *Res* **25**: 1043-1055.

814 **Pommier T**, Canback B, Riemann L, Bostrom KH, Simu K, Lundberg P, Tunlid A, Hagstrom A. 2007.
815 Global Patterns of Diversity and Community Structure in Marine Bacterioplankton. *Mol Ecol* **16**:
816 867-880.

817 **Quast C**, Priesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO. 2013. The Silva
818 Ribosomal Rna Gene Database Project: Improved Data Processing and Web-Based Tools. *Nucleic*
819 *Acids Res* **41**: D590-D596.

820 **Rawlings ND**, Barrett AJ, Bateman A. 2012. Merops: The Database of Proteolytic Enzymes, Their
821 Substrates and Inhibitors. *Nucleic Acids Res* **40**: D343-D350.

822 **Richter M**, Rossello-Mora R. 2009. Shifting the Genomic Gold Standard for the Prokaryotic Species
823 Definition. *P Natl Acad Sci USA* **106**: 19126-19131.

824 **Richter M**, Rossello-Mora R, Glockner FO, Peplies J. 2016. Jspeciesws: A Web Server for
825 Prokaryotic Species Circumscription Based on Pairwise Genome Comparison. *Bioinformatics* **32**:
826 929-931.

827 **Rogowski A**, Briggs JA, Mortimer JC, Tryfona T, Terrapon N, Lowe EC, Basle A, Morland C, Day
828 AM, Zheng HJ, Rogers TE, Thompson P, Hawkins AR, Yadav MP, Henrissat B, Martens EC,
829 Dupree P, Gilbert HJ, Bolam DN. 2015. Glycan Complexity Dictates Microbial Resource
830 Allocation in the Large Intestine. *Nat Commun* **6**: 7481.

831 **Rzhetsky A**, Nei M. 1992. A Simple Method for Estimating and Testing Minimum-Evolution Trees.
832 *Mol Biol Evol* **9**: 945-967.

833 **Saitou N**, Nei M. 1987. The Neighbor-Joining Method - a New Method for Reconstructing
834 Phylogenetic Trees. *Mol Biol Evol* **4**: 406-425.

835 **Selengut JD**, Haft DH, Davidsen T, Ganapathy A, Gwinn-Giglio M, Nelson WC, Richter AR, White
836 O. 2007. Tigrfams and Genome Properties: Tools for the Assignment of Molecular Function and
837 Biological Process in Prokaryotic Genomes. *Nucleic Acids Res* **35**: D260-D264.

838 **Snajdr J**, Cajthaml T, Valaskova V, Merhautova V, Petrankova M, Spetz P, Leppanen K, Baldrian P.
839 2011. Transformation of Quercus Petraea Litter: Successive Changes in Litter Chemistry Are
840 Reflected in Differential Enzyme Activity and Changes in the Microbial Community Composition.
841 *Fems Microbiol Ecol* **75**: 291-303.

842 **Sonnenburg ED**, Zheng HJ, Joglekar P, Higginbottom SK, Firbank SJ, Bolam DN, Sonnenburg JL.
843 2010. Specificity of Polysaccharide Use in Intestinal Bacteroides Species Determines Diet-Induced
844 Microbiota Alterations. *Cell* **141**: 1241-1252.

845 **Tamura K**, Stecher G, Peterson D, Filipski A, Kumar S. 2013. Mega6: Molecular Evolutionary
846 Genetics Analysis Version 6.0. *Mol Biol Evol* **30**: 2725-2729.

847 **Teeling H**, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennke CM, Kassabgy M, Huang SX,
848 Mann AJ, Waldmann J, Weber M, Klindworth A, Otto A, Lange J, Bernhardt J, Reinsch C, Hecker
849 M, Peplies J, Bockelmann FD, Callies U, Gerdt G, Wichels A, Wiltshire KH, Glockner FO,
850 Schweder T, Amann R. 2012. Substrate-Controlled Succession of Marine Bacterioplankton
851 Populations Induced by a Phytoplankton Bloom. *Science* **336**: 608-611.

852 **Terrapon N**, Lombard V, Gilbert HJ, Henrissat B. 2015. Automatic Prediction of Polysaccharide
853 Utilization Loci in Bacteroidetes Species. *Bioinformatics* **31**: 647-655.

854 **Thomas F**, Hehemann JH, Rebuffet E, Czejek M, Michel G. 2011. Environmental and Gut
855 Bacteroidetes: The Food Connection. *Front Microbiol* **2**: 93.

856 **Unfried F**, Becker S, Robb CS, Hehemann JH, Markert S, Heiden SE, Hinzke T, Becher D, Reintjes
857 G, Kruger K, Avci B, Kappelmann L, Hahnke RL, Fischer T, Harder J, Teeling H, Fuchs B,
858 Barbeyron T, Amann RI, Schweder T. 2018. Adaptive Mechanisms That Provide Competitive
859 Advantages to Marine Bacteroidetes During Microalgal Blooms. *Isme J* **12**: 2894-2906.

860 **Uritskiy GV**, DiRuggiero J, Taylor J. 2018. Metawrap-a Flexible Pipeline for Genome-Resolved
861 Metagenomic Data Analysis. *Microbiome* **6**.

862 **Wishart DS**, Jewison T, Guo AC, Wilson M, Knox C, Liu YF, Djoumbou Y, Mandal R, Aziat F, Dong
863 E, Bouatra S, Sinelnikov I, Arndt D, Xia JG, Liu P, Yallou F, Bjorn Dahl T, Perez-Pineiro R, Eisner
864 R, Allen F, Neveu V, Greiner R, Scalbert A. 2013. Hmdb 3.0-the Human Metabolome Database in
865 2013. *Nucleic Acids Res* **41**: D801-D807.

866 **Wu YW**, Simmons BA, Singer SW. 2016. Maxbin 2.0: An Automated Binning Algorithm to Recover
867 Genomes from Multiple Metagenomic Datasets. *Bioinformatics* **32**: 605-607.

868 **Yin YB**, Mao XZ, Yang JC, Chen X, Mao FL, Xu Y. 2012. DbcAn: A Web Resource for Automated
869 Carbohydrate-Active Enzyme Annotation. *Nucleic Acids Res* **40**: W445-W451.

870 **Zhang J**, Liu R, Xi SC, Cai RN, Zhang X, Sun CM. 2020. A Novel Bacterial Thiosulfate Oxidation
871 Pathway Provides a New Clue About the Formation of Zero-Valent Sulfur in Deep Sea. *Isme J* **14**:
872 2261-2274.

873 **Zhu ZJ**, Schultz AW, Wang JH, Johnson CH, Yannone SM, Patti GJ, Siuzdak G. 2013. Liquid
 874 Chromatography Quadrupole Time-of-Flight Mass Spectrometry Characterization of Metabolites
 875 Guided by the Metlin Database. *Nat Protoc* **8**: 451-460.

876

877

878

879

880

881

882

883

884

885

886

887

888

889

890

891

892

Figures Supplements

Figure 1 – figure supplement 1. Calculation of the numbers of GH families distributed in the metagenome-assembled genomes (MAGs) of Bacteroides derived from the deep-sea cold seep. The number above the column indicates the number of GH families identified in corresponding Bacteroides MAG. The X-axis indicates the names of different Bacteroides MAGs. The Y-axis indicates the ratio of each GH family number to the total number of GH families identified in each Bacteroides MAG.

Figure 2 – figure supplement 1. Neighbor-joining tree based on 16S rRNA sequences showing the position of the novel species *Maribellus comscasis* WC007^T. The filled circles indicate branches of the tree that were also formed using the maximum-likelihood and minimum-evolution methods. The numbers above or below the branches are bootstrap values based on 1,000 replicates. The access number of each 16S rRNA is indicated after the strain's name. The sequence of *Agarivorans albus* LMG 21761^T is used as an outgroup. Bar, 0.05 substitutions per nucleotide position.

Figure 2 – figure supplement 2. Phylogenomic tree analysis of *M. comscasis* WC007^T based on the TYGS algorithm (<https://tygs.dsmz.de/>).

934

935 **Figure 2 – figure supplement 3.** Comparison of the number of genes encoding SusC/SusD pairs,
 936 CAZymes and sulfatases identified in the genomes of *Bacteroides* from different environments
 937 (including gut, lake, shallow ocean, deep-sea etc.). *M. comscasis* WC007^T is framed by a dotted
 938 line and indicated with two pentacle stars. The detailed information related to this figure is listed
 939 in the Supplementary file 7.

940

Figure 3 – figure supplement 3. PULs predicted to target fucoidan (mainly fucose-containing sulfated polysaccharides and L-fucoidan) in the genome of *M. comscasis* WC007^T. Color coding of genes indicates gene types and corresponding numbers indicate CAZymes family associations.

Figure 3 – figure supplement 4. PULs predicted to target mannan (sulfated α -glucuronomannan, α -mannan and β -mannan) in the genome of *M. comscasis* WC007^T. Color coding of genes indicates gene types and corresponding numbers indicate CAZymes family associations.

Figure 3 – figure supplement 5. PULs predicted to target xylan (mainly sulfated xylan, α -xylose and β -xylose), α -mannan and β -mannan in the genome of *M. comscasis* WC007^T. Color coding of genes indicate gene types and corresponding numbers indicate CAZymes family associations.

Figure 3 – figure supplement 6. PULs predicted to target glucan (mainly starch and laminarin) in the genome of *M. comscasis* WC007^T. Color coding of genes indicates gene types and corresponding numbers indicate CAZymes family associations.

1017

1018 **Figure 4 – figure supplement 1.** Transcriptomics based heat map showing the PULs with all
 1019 genes up-regulated when cultured *M. comscasis* WC007^T in the medium supplemented with 1 g/L
 1020 cellulose.

1021

1022

1023

1024

1025

1026

Figure 4 – figure supplement 2. Transcriptomics based heat map showing the PULs with all genes up-regulated when cultured *M. comscasis* WC007^T in the medium supplemented with 1 g/L pectin.

Figure 4 – figure supplement 3. Transcriptomics based heat map showing the PULs with all genes up-regulated when cultured *M. comscasis* WC007^T in the medium supplemented with 1 g/L fucoidan.

1043

1044 **Figure 4 – figure supplement 4.** Transcriptomics based heat map showing the PULs with all

1045 genes up-regulated when cultured *M. comscasis* WC007^T in the medium supplemented with 1 g/L

1046 mannan.

1047

1048

Figure 4 – figure supplement 5. Transcriptomics based heat map showing the PULs with all genes up-regulated when cultured *M. comscasis* WC007^T in the medium supplemented with 1 g/L xylan.

Figure 4 – figure supplement 6. Transcriptomics based heat map showing the PULs with all genes up-regulated when cultured *M. comscasis* WC007^T in the medium supplemented with 1 g/L starch.

1060

1061 **Figure 6 – figure supplement 1.** KEGG pathway enrichment of differential metabolites when
 1062 cultured *M. comscasis* WC007^T in the medium supplemented with 1 g/L cellulose. Only Top 20
 1063 pathway enrichments were shown.

1064

1065

1066

1067

1068

1069

1070

1071

Figure 7 – figure supplement 1. Combination analysis of enriched pathways associated genes and metabolites respectively identified in the transcriptomic and metabonomic analyses of *M. comscasis* WC007^T treated with 1 g/L cellulose. Blue column represents the number of differentially expressed genes identified in transcriptomic analysis. Red column represents the number of differentially abundant metabolites identified in metabonomic analysis.