



Genome-Wide Identification and Functional Characterization of β -Agarases in *Vibrio astriarenae* Strain HN897

Yupeng Liu^{††}, Xingkun Jin^{††}, Chao Wu¹, Xinyuan Zhu¹, Min Liu¹, Douglas R. Call² and Zhe Zhao^{1*}

¹ Department of Marine Biology, College of Oceanography, Hohai University, Nanjing, China, ² Paul G Allen School for Global Animal Health, Washington State University, Pullman, WA, United States

OPEN ACCESS

Edited by:

Frank T. Robb,
University of Maryland, Baltimore,
United States

Reviewed by:

Chaomin Sun,
Institute of Oceanology, CAS, China
Daniel Lundin,
Linnaeus University, Sweden

*Correspondence:

Zhe Zhao
zhezhaoh@hhu.edu.cn

[†] These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Evolutionary and Genomic
Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 12 February 2020

Accepted: 29 May 2020

Published: 24 June 2020

Citation:

Liu Y, Jin X, Wu C, Zhu X, Liu M,
Call DR and Zhao Z (2020)
Genome-Wide Identification
and Functional Characterization
of β -Agarases in *Vibrio astriarenae*
Strain HN897.
Front. Microbiol. 11:1404.
doi: 10.3389/fmicb.2020.01404

The genus *Vibrio* is a genetically and metabolically versatile group of heterotrophic bacteria that are important contributors to carbon cycling within marine and estuarine ecosystems. HN897, a *Vibrio* strain isolated from the coastal seawater of South China, was shown to be agarolytic and capable of catabolizing D-galactose. Herein, we used Illumina and PacBio sequencing to assemble the whole genome sequence for the strain HN897, which was comprised of two circular chromosomes (Vas1 and Vas2). Genome-wide phylogenetic analysis with 140 other *Vibrio* sequences firmly placed the strain HN897 into the Marisflavi clade, with *Vibrio astriarenae* strain C7 being the closest relative. Of all types of carbohydrate-active enzyme classes, glycoside hydrolases (GH) were the most common in the HN897 genome. These included eight GHs identified as putative β -agarases belonging to GH16 and GH50 families in equal proportions. Synteny analysis showed that GH16 and GH50 genes were tandemly arrayed on two different chromosomes consistent with gene duplication. Gene knockout and complementation studies and phenotypic assays confirmed that *Vas1_1339*, a GH16_16 subfamily gene, exhibits an agarolytic phenotype of the strain. Collectively, these findings explained the agar-decomposing of strain HN897, but also provided valuable resources to gain more detailed insights into the evolution and physiological capability of the strain HN897, which was a presumptive member of the species *V. astriarenae*.

Keywords: agarolytic activity, β -agarase, gene knockout, glycoside hydrolase (GH), *Vibrio astriarenae*, whole genome sequencing

INTRODUCTION

Agar, the major ingredient of the cell wall in many red algae, is a complex water-soluble polysaccharide that is composed of agarose and agaropectin (Renn, 1997). Agarose is a linear macromolecule composed of repeated D-galactose and 3,6-anhydro-L-galactose units joined by α -1,3- and β -1,4-glycosidic bonds, while agaropectin is a derivative of agarose with repeated units of

methylated, pyruvated, sulfonated, or glycosylated galactose moieties (Lee et al., 2017). Agarase is an important agarolytic enzyme involved in producing oligosaccharides or monomeric sugars from agarose, functions that are of considerable interest to the food, cosmetics and medical industries (Fu and Kim, 2010; Yun et al., 2017).

β -agarases (EC 3.2.1.81) are the most common class of agarase and to date, more than 30 have been functionally verified and documented in the carbohydrate-active enzymes (CAZymes) database (Chi et al., 2012). Based on amino acid sequence similarity, all β -agarases can be classified into the glycoside hydrolase (GH) subfamilies, including GH16, GH50, GH86, and GH118 (Lee et al., 2012). The mechanism of these enzymes is well established (Ekborg et al., 2006; Fu and Kim, 2010; Pluvinae et al., 2013; Yun et al., 2017). Initially, the endo-type β -agarase, including GH16 and GH86, hydrolyze internal β -1,4 glycosidic linkages of agarose and release a series of neoagarooligosaccharides (NAOS) (Ohta et al., 2004; Hehemann et al., 2012; Chi et al., 2014; Naretto et al., 2019). Other endo-type β -agarases, including GH50 and GH118, act in cleaving the terminal β -1,4 linkages at the non-reducing end of NAOS, and ultimately release hydrolysis products such as neoagarobiose (NeODP2) (Xie et al., 2013; Yu et al., 2020).

The majority of agarase-producing bacteria are from marine environments including taxonomically distinct genus such as *Pseudomonas* (Kang et al., 2003; Hsu et al., 2015), *Alteromonas* (Chi et al., 2014), *Microbulbifer* (Ohta et al., 2004; Jonnadula and Ghadi, 2011), and *Agarivorans* (Fu et al., 2009; Lin et al., 2012). Although agar utilization is not particularly common in *Vibrio* species, several β -agarases have been identified (Lee et al., 2014; Yun et al., 2015; Yu et al., 2020). For example, agarases have been purified and characterized from different agarolytic *Vibrio* strains (Aoki et al., 1990; Sugano et al., 1993; Araki et al., 1998), and a set of β -agarase-encoding genes have been described and functionally validated as recombinant proteins (Sugano et al., 1993; Dong et al., 2006; Zhang and Sun, 2007; Liao et al., 2011). The agarolytic system of the *Vibrio* sp. strain EJY3 has been studied extensively and this strain is the only agarolytic *Vibrio* that harbors four GH50 β -agarases (Lee et al., 2014; Yun et al., 2015; Yu et al., 2020).

Vibrios are Gram-negative, obligate heterotrophic bacteria that inhabit marine and estuarine environments as free-living populations or in association with a variety of aquatic organisms (Thompson et al., 2004). This is a diverse genus with more than 120 validated species (Sawabe et al., 2013; Parte, 2018). Some *Vibrio* species are pathogenic to marine animals and people, while the majority are mostly non-pathogenic and are found as the normal part of marine microbial communities (Thompson and Polz, 2006). *Vibrio* species are known to use complex organic carbohydrates, such as polysaccharides (cellulose, chitin, and agar etc.), as carbon and energy sources and play a significant role in environmental carbon cycling (Hunt et al., 2008; Zhang X. H. et al., 2018).

A marine *Vibrio* (strain HN897) was recently isolated from the coastal seawater of South China and shown to be agarolytic and able to catabolize D-galactose (Zhang J. Y. et al., 2018). In this study, we have generated a high-quality genome assembly

of the strain HN897 and described new agarase genes, including functional validation. Results from this work further our understanding of agar-decomposing organisms and provide a valuable resource to gain more detailed insights into the evolution and physiological capability of agar-degrading organisms that play a vital role in carbon cycling.

MATERIALS AND METHODS

Bacterial Strains, Plasmids and Growth Conditions

The bacterial strains and plasmids used in this study are listed in the **Supplementary Table S1**. The strain HN897 was isolated from seawater in the South China coastal areas. In brief, thiosulfate citrate bile salts sucrose (TCBS) plates were used to isolate 150 putative *Vibrio* strains, of which three were phenotypically tested for agarose-degrading activity. The most active strain was designated HN897. All the *Vibrio* strains were routinely grown in trypticase soy broth (TSB; BD) supplemented with 1% (w/v) NaCl or on TSB agar plates (TSA) supplemented with 1% (w/v) NaCl at 30°C. *Escherichia coli* S17 λ pir was used in gene deletion experiments and was cultured in Luria-Bertani (LB; BD) medium. Suicide plasmid pDM4 was used to generate gene knockouts and expression vector pMMB207 was used for complementation experiment (Morales et al., 1991; Milton et al., 1996). Unless otherwise indicated, chloramphenicol was used at the concentration of 34 μ g/mL in the media.

Genomic DNA Isolation and Whole Genome Sequencing

Genomic DNA was extracted from the strain HN897 using the CTAB method with minor modification (Liu et al., 2017). The concentration, quality and integrity of the DNA samples were evaluated using a NanoDrop 2000 Spectrophotometer (Thermo Scientific, United States) and a Bioanalyzer 2100 (Agilent Technologies, United States), respectively. Insertion libraries [400 bp pair-end (PE250)] were constructed with Illumina TruSeq sample preparation reagents according to manufacturer instructions and sequenced using an Illumina MiSeq (Illumina, United States) instrument. Larger 20k bp insertion (S20K) libraries were prepared by SMRTbell Template Prep Kit 1.0 according to the PacBio protocol, and sequenced using a PacBio RS II platform (Pacific Biosciences, United States).

Genome Assembly and Annotation

For the Illumina short reads, the adaptor sequences were removed by Adapter Removal v2.1.7 (Mikkil et al., 2016), and the k-mers for the paired-end libraries were tallied using SOAPec v2.0 (Luo et al., 2012) and verified using SPAdes v3.9.0 (Bankevich et al., 2012) followed by assembly with A5-miseq v20150522 software (Tritt et al., 2012). The PacBio sequences were assembled using HGAP4 (Chin et al., 2016) and CANU v1.6 (Koren et al., 2017). Subsequently, the assembled contigs from both Illumina and PacBio reads were merged, and gaps were filled using MUMmer v3 (Delcher et al., 1999). Finally, the full genome

assembly was generated with Pilon v1.22 (Walker et al., 2014) software. Microbial genes were predicted with GeneMark.hmm v4.32 using the Heuristic models option (Besemer et al., 2001). The coding sequences (CDS) of the putative genes were aligned to NCBI-NR database using BLASTP to retrieve annotations with an E-value cutoff of $1e-6$. Clusters of orthologous genes (COGs) were annotated with eggNOG-mapper (evolutionary genealogy of genes: non-supervised orthologous groups) with eggNOG database v4.5 and an E-value cutoff of $1e-6$ (Huerta-Cepas et al., 2017). The final assembled circular chromosomes (CGs) were compared to closely related species by sequence alignment using tBLASTx with the following thresholds: E-value: $1e-10$; identity: 90%; alignment length: 100. CG maps were constructed and visualized based on the sequence alignment blocks using GView Server (Stothard et al., 2019). KAAS (KEGG Automatic Annotation Server)¹ was used to assign the KEGG (Kyoto Encyclopedia of Genes and Genomes) orthology annotations² for the predicted proteins. Proteins were functionally classified via subfamily domain modules that had been annotated using PfamScan v1.6, with Pfam32.0 database (Finn et al., 2016). Putative CAZymes profiles were identified using Hmmscan v3.1b2 (Krogh et al., 1994).

Genome-Based Taxonomic Classification

Using the representative genomes in Genome Taxonomy Database (GTDB Release 89) (Parks et al., 2020) as references, the genome-based taxonomy of HN897 was classified using GTDB-tk 'classify_wf' (Chaumeil et al., 2019). Maximum-likelihood ML (WAG + gamma model) trees were constructed using FastTree2 (Price et al., 2010) based on multiple sequences alignment of 120 concatenated bacterial single-copy marker protein sequences according to the GTDB genome phylogeny (Parks et al., 2018). Further, the quality of HN897 genome assembly was assessed based on lineage-specific phylogenetic markers using CheckM (Parks et al., 2015). A precomputed and visualized genome-wide phylogeny with taxonomic annotations showing the members in γ -proteobacterial order was downloaded from AnnoTree (Mendler et al., 2019). As a complementary analysis, a ML phylogenetic tree, based on 16S rDNA sequences, was constructed using FastTree2 with a GTR + gamma model and MAFFT-aligned sequences from Al-Saari et al. (2015). This latter analysis included other *Vibrio* data having a high sequence similarity (>98%) to HN897. All the ML trees were constructed and visualized by iTOL (Letunic and Bork, 2019).

Sequence Clustering and Synteny Analysis of β -Agarase-Coding Genes

Based on the CAZymes classifications, the predicted full-length proteins of GH subfamilies were abstracted and characterized for conserved enzymatic domains (excluding carbohydrate-binding CBMs and lectins) were identified using InterProScan (Jones et al., 2014). Identified GH domain sequences were aligned

using COBALT (Constraint-based Multiple Alignment Tool) (Papadopoulos and Agarwala, 2007) and an unrooted ML tree was constructed using FastTree2 with the WAG + Gamma models. The protein domain architectures for putative β -agarases were further inspected and visualized according to SMART database (Letunic and Bork, 2018). The structure-based sequence similarity comparisons were based on alignment by MUSCLE v3.8.31, and the secondary structure information was parsed and annotated using ESPript 3.0 (Robert and Gouet, 2014).

A total of 45 subject genome assemblies were used for synteny comparisons with the strain HN897. These assemblies were selected because they (i) included genes that have high sequence similarities to the β -agarase from HN897 (>60% amino acid identity); (ii) appeared phylogenetically related to HN897 according to genome-based taxonomic classification from the present study and the phylogeny inference of the strain C7 (Al-Saari et al., 2015). A total of 46 assemblies were used and the genomic loci spanning eight β -agarases from HN897 were inspected and compared using MultiGeneBLAST v1.1.14 (Medema et al., 2013).

Construction of Deletion Mutants and Complementation

All deletion mutants were constructed by allelic exchange as described previously (Zhao et al., 2018). Briefly, deletion cassettes for chromosomal in-frame deletions were generated through the splice-overlap-extension (SOE) method (primers listed in **Supplementary Table S2**), which joins two approximately 400-bp amplicons corresponding to genomic regions flanking β -agarase genes. The deletion cassettes were then cloned into a suicide plasmid (pDM4) using standard cloning procedures. Each resulting construct was transformed into chemically competent cells of *E. coli* S17-1 λ pir, and then transferred into HN897 by conjugation. Mutant strains were selected on TCBS plates containing chloramphenicol (6.8 μ g/mL) followed by a 10% (w/v) sucrose selection process. Each positive deletion mutant was experimentally confirmed using PCR with gene-specific primers spanning the deleted region.

For complementation tests, the complete ORF of each indicated target gene was amplified using gene-specific primers (**Supplementary Table S2**) and cloned into the vector pMMB207. The insertion success of constructed plasmid was checked by Sanger sequencing, and only the positive ones were introduced into the mutant strain by conjugation.

Assays of Agarolytic and Agarase Activity

An agar-spot assay was used to evaluate the agarolytic activity of each HN897 strain. Briefly, each indicated strain was incubated overnight and spotted (5 μ l) onto a fresh TSA plate supplemented with 1% (w/v) NaCl and 0.5 mM IPTG. For the complementation strain, overnight culture was diluted 1:100 into TSB medium supplemented with chloramphenicol and 1% (w/v) NaCl and was cultured with shaking at 30°C. When the OD₆₀₀ reached \sim 0.6, IPTG (1 mM final) was added to induce protein expression and then continued to culture for 5 h. The induced culture

¹<https://www.genome.jp/tools/kaas/>

²<http://www.kegg.jp>

was subjected to agar-spot assay as described above. The plates were incubated at 30°C for 12 h, and the colonies that formed transparent lysis zones on agar plates were classified as agarolytic.

Agarase activity of each indicated strain was measured using the 3, 5-dinitrosalicylic acid (DNS) method as described previously (Miller, 1959). Briefly, after 24 h culture bacterial cells were pelleted by centrifugation (14000 g, 5 min) at 4°C and resulting supernatant was collected and kept for the agarase activity assay. For the complementation strain, the bacterial cells were cultured with chloramphenicol and IPTG as aforementioned. Culture supernatant (1 mL) was mixed with 1 mL 10 mM sodium citrate buffer (pH = 5) containing 0.2% agarose that was pre-solubilized using a conventional microwave oven. After incubation at 30°C for 60 min, 750 μ L of DNS solution (36.4 g of potassium sodium tartrate, 1.2 g of DNS, and 83.2 mL 5% NaOH in 200 mL distilled water) was added to the cultures and boiled at 100°C for 10 min, followed by ice-cold water-based cooling for 2 min. The absorbance at 520 nm for each sample was measured using spectrophotometer. Distilled water (1 mL) was included as a blank control. For the agarase activity assays, the extent of agarose reduction was estimated using a standard curve of D-galactose (Sangon, Shanghai, China).

RESULTS

High-Quality Genome Sequence of Marine *Vibrio* Strain HN897

Contig assembly were generated using 4,713,440 (96.86% of total) Illumina paired-end reads with 233 \times coverage. Scaffolds were generated using 152 \times coverage of PacBio sequences (11,636-bp N₅₀). The contigs generated by both short- and long-reads were scaffolded to fill gaps and increased the accuracy of sequence consensus of assembly. The complete genome for HN897 included two circular chromosomes spanning 3,120,326 bp with 46.06% GC, and 1,679,007 bp with 44.82% GC, respectively (**Supplementary Figures S1, S2**). The larger chromosome ("Vas1") encompasses 2,801 coding sequences (CDSs), 110 tRNAs and 34 rRNAs while the smaller chromosome ("Vas2") consists of 1,487 CDSs, 6 tRNA genes, and 16 miscellaneous ncRNAs (**Supplementary Table S3**).

The BLAST-NR based annotations allocated 2,736 and 1,420 gene hits for Vas1 and Vas2, respectively, and approximately half of the hits were derived from a single genome (*Vibrio* sp. C7) with >90% of identities and alignment coverage on average (**Supplementary Figure S3**). For the remaining top 10 hit species, the genus *Vibrio* was predominant although *Photobacterium gaebulicola* was also identified. *P. gaebulicola* is a member of *Vibrionaceae* (**Supplementary Figure S3**). The CheckM assessment and classification showed the genome assembly of HN897 was 98.21% complete and 3.24% contaminated, ensuring a critical foundation for further biological inference of this particular strain. It is worth mentioning that the genome contaminations reported by CheckM was based on occurrence of duplications of single-copy marker genes, which might be true expanded paralogs in HN897 genome. The full standard metrics for HN897 assembly according to the Genomic

Standards Consortium (GSC) (Bowers et al., 2017) are shown in **Supplementary Table S3**.

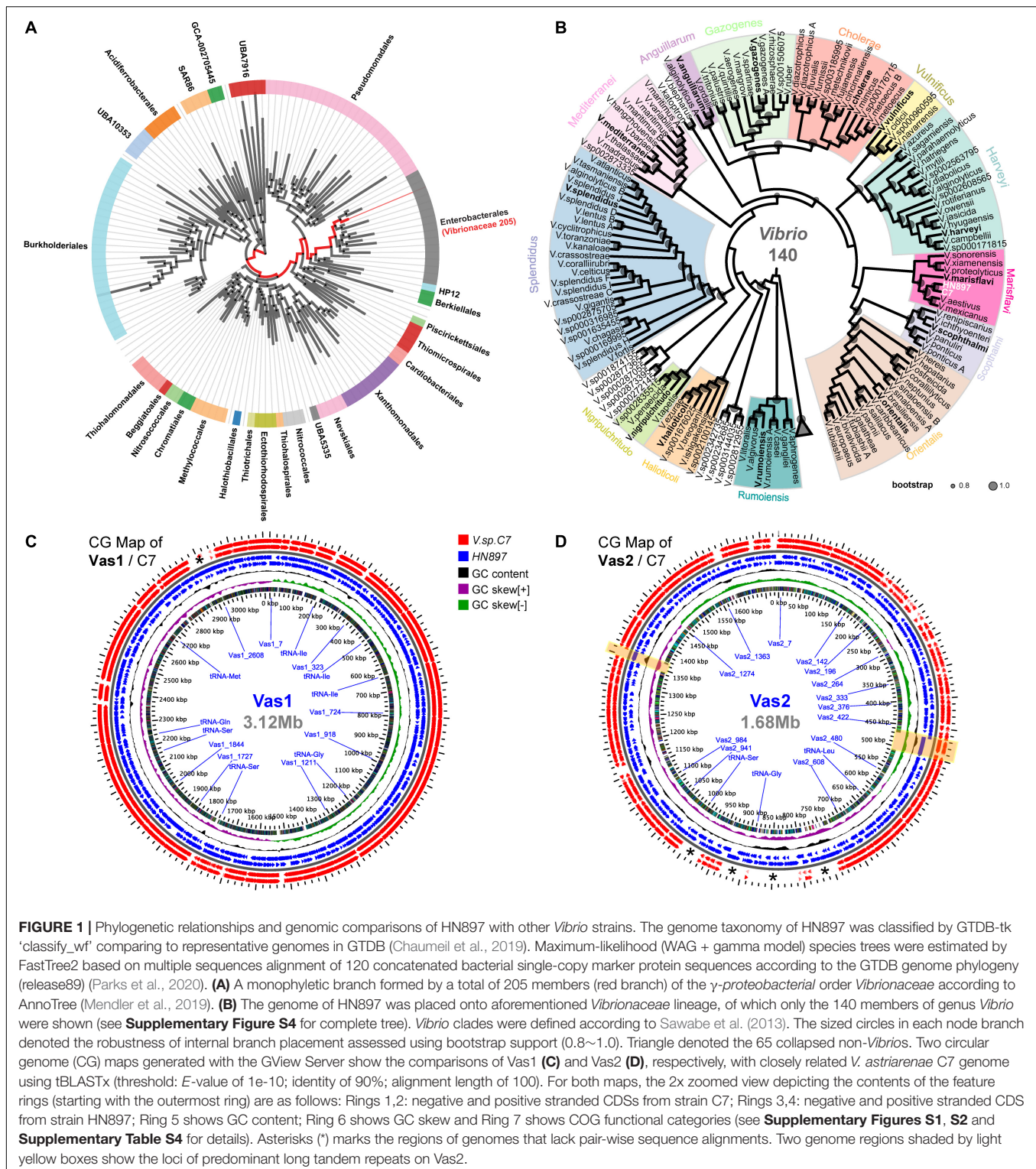
Taxonomic Assignment of HN897 Genome

According to GTDB, all 205 representative species of *Vibrionaceae* formed a monophyletic family level group (red clade in **Figure 1A**). Notably, the parent taxon of *Vibrionaceae* was *Vibrionales* under the NCBI taxonomy, but reclassified as *Enterobacterales* by GTDB taxonomy after taxonomic-rank normalization (Parks et al., 2018). Further, the genome of HN897 was firmly placed into the aforementioned *Vibrionaceae* subtree, together with 140 *Vibrio* members, and more precisely within a subclade shared with *V. astriarenae* strain C7 (**Figure 1B** and **Supplementary Figure S4**). Alignment-based classification showed that the average nucleotide identity (ANI) between HN897 and C7 was 96.6% (**Supplementary Table S3**), further indicating the close relatedness between these two strains. We further constructed two CG maps comparing the sequences of Vas1 and Vas2 with that of contig-level genome assembly of C7 respectively. Using BLAST comparisons conducted between CDS translations from both strains, the detected regions with high sequence similarity (90% identity threshold) included most of the coding regions in each chromosome (95.5% CDSs for Vas1; 77.2% for Vas2). The positions of contig boundaries and potential fragmentations for C7 assembly were revealed as marked with asterisks and shaded boxes shown in **Figures 1C,D**.

Interpretation of the HN897 Genome via Functional Annotations

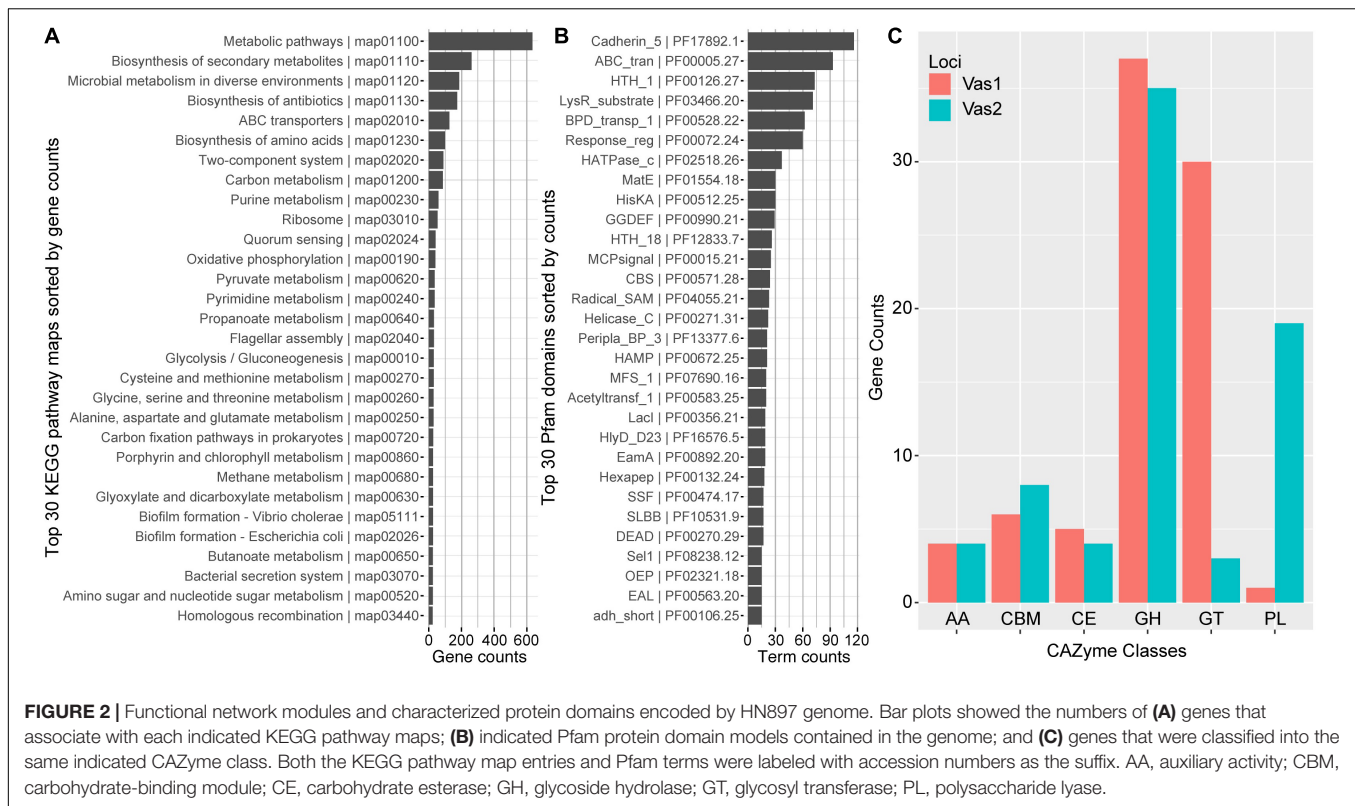
In total, 2,344 (54.66%) predicted CDSs were placed into ortholog groups according to the KEGG Orthology (KO) database. Most of the KO-annotated genes were related to metabolic pathways (634), especially anabolism such as biosynthesis of secondary metabolites (262), antibiotics (174) and amino acids (99). In addition, many other KO-annotated genes were related to different cellular signaling pathways, including ABC transporters pathway (126), two-component systems (99) and quorum sensing (41). Importantly, a set of gene components that take part in carbon (86) and pyruvate (36) metabolism, and glycolysis/gluconeogenesis (30) appeared relatively complete (**Figure 2A** and **Supplementary Table S4**).

The majority of CDSs for Vas1 (87.9%) and Vas2 (81.4%) had at least one Pfam model hit. The most abundant domain was cadherin_5 (PF17892) with 116 hits (**Figure 2B**), amongst which 111 were tandemly arranged within the longest CDS, Vas2_433 (12,014 aa, WP_164651137.1) (The larger shaded box in **Figure 1D**). Notably, the second longest CDS, Vas_1198 (5,779 aa, WP_164650763.1) also contained a single cadherin_5 domain with an additional 19 N-terminal tandem repeats of t1ss_rpt_143 domain (tigr:TIGR03660), and a C-terminal domain cluster including one each of tandem-95 repeat (CDD: NF012211), hemolysinCabin (PF00353) and peptidase_M10_C (PF08548) (the smaller shaded box in **Figure 1D**). Consistent with the KEGG annotation results, the ABC_transporter (PF00005) domains were abundant, which was not uncommon



as this group forms the largest protein family in many other sequenced bacterial genomes (Saurin et al., 1999). Many signal transduction related domains were found with high frequency, such as methyl-accepting chemotaxis protein (MCP) domain (MCPsignal, PF00015), HAMP (PF00672), response regulator

receiver domain (Response_reg, PF00072), and his kinase A (HisKA, PF00512). Notably, the latter two domains are critical components of two-component system regulators. The other abundant domains included evolutionarily conserved helix-turn-helix DNA binding motifs, such as HTH1 (PF00126),



HTH18 (PF12833), and LacI (PF00356) that regulates gene expression (Figure 2B).

Most of the agarolytic bacteria encoded a large repertoire of putative CAZymes and carbohydrate-binding proteins (Barbeyron et al., 2016; Osterholz et al., 2016). HN897 had a total of 72 glycoside hydrolases (GHs), 33 glycosyl transferases (GTs), 20 polysaccharide lyases (PLs), 9 carbohydrate esterases (CEs), and 8 auxiliary activities (AAs). Carbohydrate-binding modules (CBMs) were also identified (Figure 2C and Supplementary Table S4).

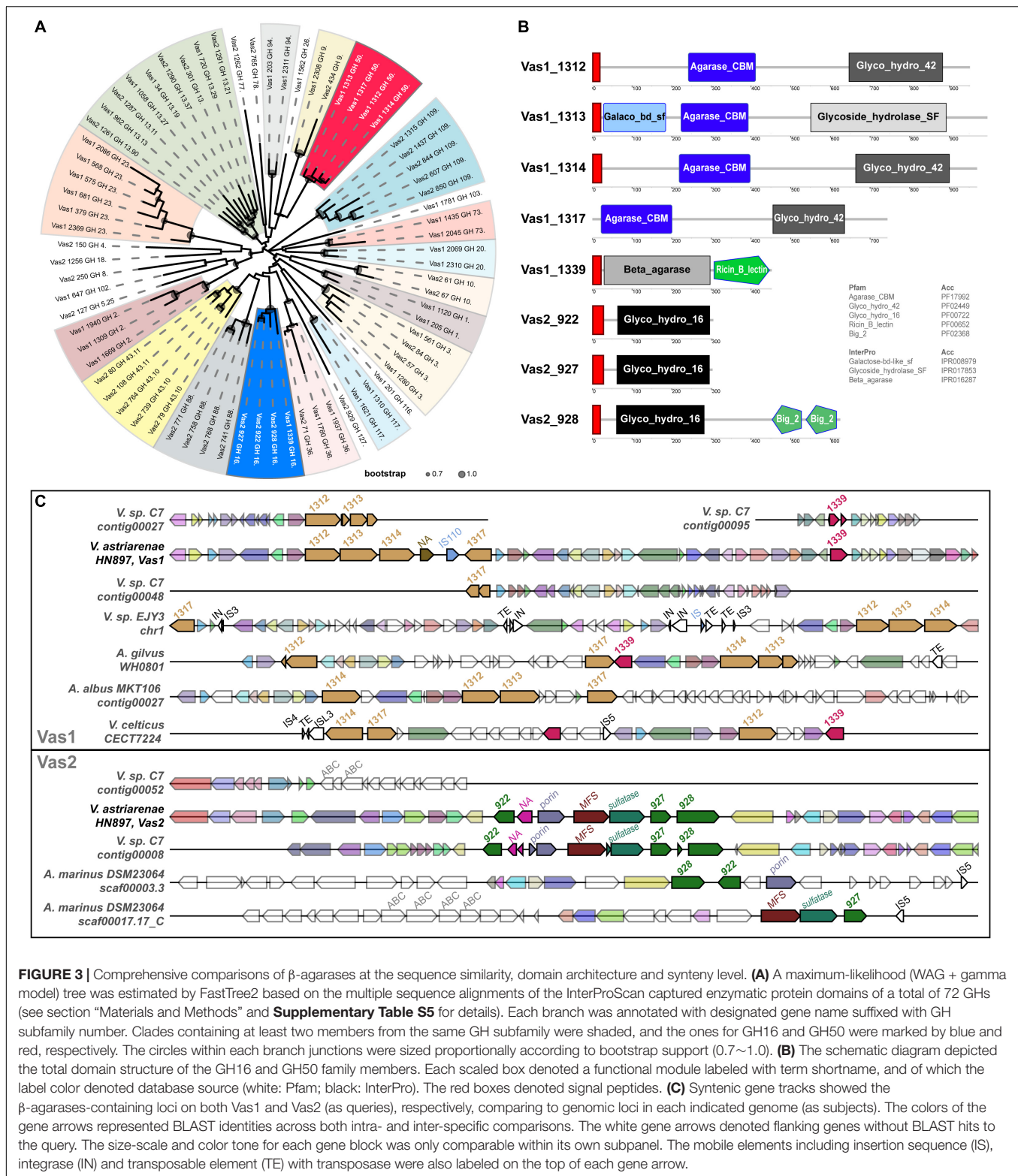
The β -Agarase Gene Repertoire of Strain HN897

Eight GH subfamily members were similar to β -agarases, including all the members of GH16 and GH50 in HN897 (Supplementary Table S5). As expected, a ML tree based on conserved enzymatic domains from 72 GH protein sequences revealed congruence at the GH subfamily level for all the GHs. The eight β -agarases from HN897 were included in subclade of GH50 and GH16 (Figure 3A).

The identified N-terminal hydrophobic signal peptides showed that members of both GH50s and GH16s were likely to be secretory, with the only exception of Vas1_1317 (WP_164648142.1) that lacked such structure (Figure 3B). As expected, the domains of agarase_CBM (PF17992) and glyco_hydro_42 (PF02449) were only shared by the GH50 members while the glyco_hydro_16 domain (PF00722) was specific to GH16 members. Vas1_1313 (a

GH50 member, WP_164648138.1) harbored an alternative glycoside_hydrolase_SF domain (IPR017853) in its C-terminus, and an extra galactose-bd-like_sf domain (IPR008979) in its N-terminus. Vas1_1339 (a GH16 member, WP_164648163.1) possessed an alternative domain, namely beta_agarase (IPR016287) in N-terminal region, and an extra ricin_B_lectin domain (PF00652) in its C-terminal region. Vas2_928 (a GH16 member, WP_164650524.1) had two extra tandemly repeated domains namely bacterial_Ig-like_2 (PF02368) in its C-terminal end.

Gene clusters for regions spanning the eight β -agarase genes of HN897 were compared with 45 bacterial genomes (Supplementary Table S5). The intraspecific pairwise sequence identity for the GH50s from Vas1 ranged between 34% and 48%, and ranged between 37 and 66% for GH16s from Vas2. The four Vas1 GH50-encoding genes were arranged tandemly, with the latter two separated by a non-annotated putative gene followed by an IS110 transposase (upper panel of Figure 3C). A single GH16 gene, Vas1_1339, was located downstream of the above mentioned GH50s cluster, with an interval of 22 protein-coding genes. On Vas2, the three GH16s were also arranged tandemly with the upstream two separated by four non-GH family genes (bottom panel of Figure 3C). As expected, the overall synteny and collinearity between HN897 and C7 were consistent for both Vas1 and Vas2, except for the absence of one GH50 gene, Vas1_1314 (WP_164648139.1), and prevalent frameshifts in several other GH subfamily genes on C7 assembly (Figure 3C and Supplementary Table S5). Moreover, the one-to-one sequence similarities of β -agarases encoding



genes was >99% between these two strains (**Supplementary Table S5**), implying recent ancestry or conserved function. The quadripartite tandem array of GH50s found on Vas1 were only partially evident in other species including *Vibrio sp.* EYJ3

(*Vibrio natriegens* according to GTDB), *Agarivorans gilvus* strain WH0801, *Agarivorans albus* MKT106 and *Vibrio celticus* strain CECT 7224. Notably, the cross-species sequence similarities of encoded GH50 protein sequences between HN897 and EYJ3 were

high (72~92%). Nevertheless, genome-wide synteny was less consistent as evidenced by various degrees of local translocations and inversions (upper panel of **Figure 3C**). Comparatively, the tripartite tandem array of GH16 on Vas2 was only found in two separate contigs of *Aliagarivorans marinus* DSM 23064, but with certain rearrangement (bottom panel of **Figure 3C**).

Functional Verification of the Vas1_1339-Agarolytic Phenotype

We individually constructed knock-out (KO) strains for all the eight putative β -agarase genes found in HN897. Agar-spot assays showed that the deletion mutants for seven of these strains resemble the wild-type strain (**Figure 4A**). In contrast, the agarolytic activity was abrogated in the Vas1_1339 deletion strain (Δ 1339). *In-trans* expression of Vas1_1339 in the knockout strain restored the agarolytic phenotype (**Figure 4A**). Enzymatic activity assays further showed that the agarase activity in culture supernatant of Δ 1339 strain (0.327 U/mL) was significantly reduced comparing to the wild-type strain (1.859 U/mL) and complemented strain (1.408 U/mL) (**Figure 4B**). As control, the blank TSB medium and the culture supernatant of *Vibrio parahaemolyticus* strain RIMD2210633 (clinical isolate O3:K6) (Makino et al., 2003) both exhibited little to no agarase activity (**Figure 4B**).

Using PaperBLAST (Price and Arkin, 2017), we found that Vas1_1339 was most similar to a β -agarase gene (ABL06969.1), *agaV*, from *Vibrio* sp. V134 (similarity: 99%, coverage: 100%) and differing by only two amino acid residuals (Leu²⁵⁹ and Pro³⁴⁷ in Vas1_1339, Ile²⁵⁹ and Leu³⁴⁷ in *AgaV*). The 16S rRNA based ML tree also suggested close relatedness between identified *V. astriarenae* strains (C7, C20, and presumptively, HN897) and strain V134, indicating that the latter was likely a strain of *V. astriarenae* (**Supplementary Figure S5** and **Supplementary Table S6**).

A structure-referenced sequence alignment suggested that the underlying molecular basis for agarolysis was evolutionary conserved, at least within the β -agarase domain (IPR016287) of the sequences that we investigated (**Supplementary Figure S6** and **Supplementary Table S7**). For instance, the active sites for enzymatic reactions were nearly invariant, particularly the catalytic residues that consisted of a tripartite motif "ExDxxE" (marked by black triangles in **Figure 4C**). Likewise, three separate residues on domain termini were highly conserved. These residues were known to function together as a cation-binding site (usually occupied by Ca²⁺, Naretto et al., 2019).

DISCUSSION

Earlier work demonstrated that the marine *Vibrio* isolate HN897 is capable of catabolizing agar as a sole carbon source and it exhibits agarose-degrading activity (Zhang J. Y. et al., 2018). Using sequence data from Illumina and PacBio platforms and other analytic advances and databases such as genome-based taxonomy and quantitative species definition (see Parks et al., 2015, 2018, 2020; Chaumeil et al., 2019), we were able to determine with a high degree of confidence that the strain

HN897 belonged to the *Vibrio astriarenae* (**Figure 1B**). This conclusion was further supported by high matching scores for sequence alignments (**Supplementary Table S3**), and by analysis of 16S rDNA sequences (Zhang J. Y. et al., 2018; **Supplementary Figure S5**).

The maximum likelihood analysis placed both HN897 and *Vibrio astriarenae* C7 within the Marisflavi clade of the *Vibrio* genus (**Figure 1B**). This clade was originally defined as a group of three species, *Vibrio aestivus*, *Vibrio marisflavi* and *Vibrio stylophorae* (Sheu et al., 2011) based on the 16S rDNA comparisons. Subsequent studies have added a new species, *Vibrio mexicanus*, together with *V. aestivus* to form a monophyletic group with *V. marisflavi* that is well separated from *V. stylophorae* (González-Castillo et al., 2015). In contrast, analysis based on multilocus sequence typing suggests that the *V. mexicanus*-*V. aestivus* cluster is well separated from *V. marisflavi*.

While different phylogenetic conclusions may be evident from different methodologies, we submit that genome-level comparisons are likely to be more robust if only because the entirety of genetic data that is considered. Based on our results, the species *V. astriarenae* formed a robust clade (bootstrap 0.99) with both *V. aestivus* and *V. mexicanus* as the closest relatives, and further formed a cluster proximal to *V. marisflavi* (bootstrap 0.95). In addition, an orphan species *Vibrio proteolyticus* defined by Sawabe et al. (2013) was clustered with *Vibrio xiamenensis* as the closest neighbor (bootstrap 0.99), and together with *Vibrio sonorensis* formed another subcluster within the clade Marisflavi. The next closest clades are Scophthalmi-Orientalis that form a deep clade within the family *Vibrionaceae* (**Figure 1B** and **Supplementary Figure S4**).

Other methodological discrepancies have been described including a finding that the strain C7 was closely related to *V. agarivorans* based on analysis of MLST data, but much less so based on DNA-DNA hybridization data (Al-Saari et al., 2015). According to the latest GTDB taxonomy (r89), *V. agarivorans* was reclassified as *Vibrio sagamiensis* and has been moved to the Harveyi clade (**Figure 1B**), indicating the genomic divergence between these two species was much greater than expected despite sharing agarolytic activity. Interestingly, phenotypic data from our group and others has shown that all the members of the Marisflavi clade can catabolize D-galactose as a sole carbon source (Wang et al., 2011; Gao et al., 2012; Lucena et al., 2012; González-Castillo et al., 2015, 2016). It is likely that a combination of phenotypic and genotypic data can be used to provide new insights in resolving the origin and evolution of the Marisflavi clade.

We encountered some challenges with the comparisons between HN897 and C7 that were attributable to an unexpectedly highly prevalence of fragmentation or frameshifts within the coding regions of C7 assembly (see the shattered gene arrows of C7 in **Figure 3C**). The genome of C7 was originally assembled solely by short-read sequencing approach (Ion PGM, 400-bp reads) with low coverage (28x) (Al-Saari et al., 2015), which is expected to contribute to incomplete or fragmented genomes (Batty et al., 2018). By combining short and long-read technologies (Illumina and PacBio, respectively), we were able to

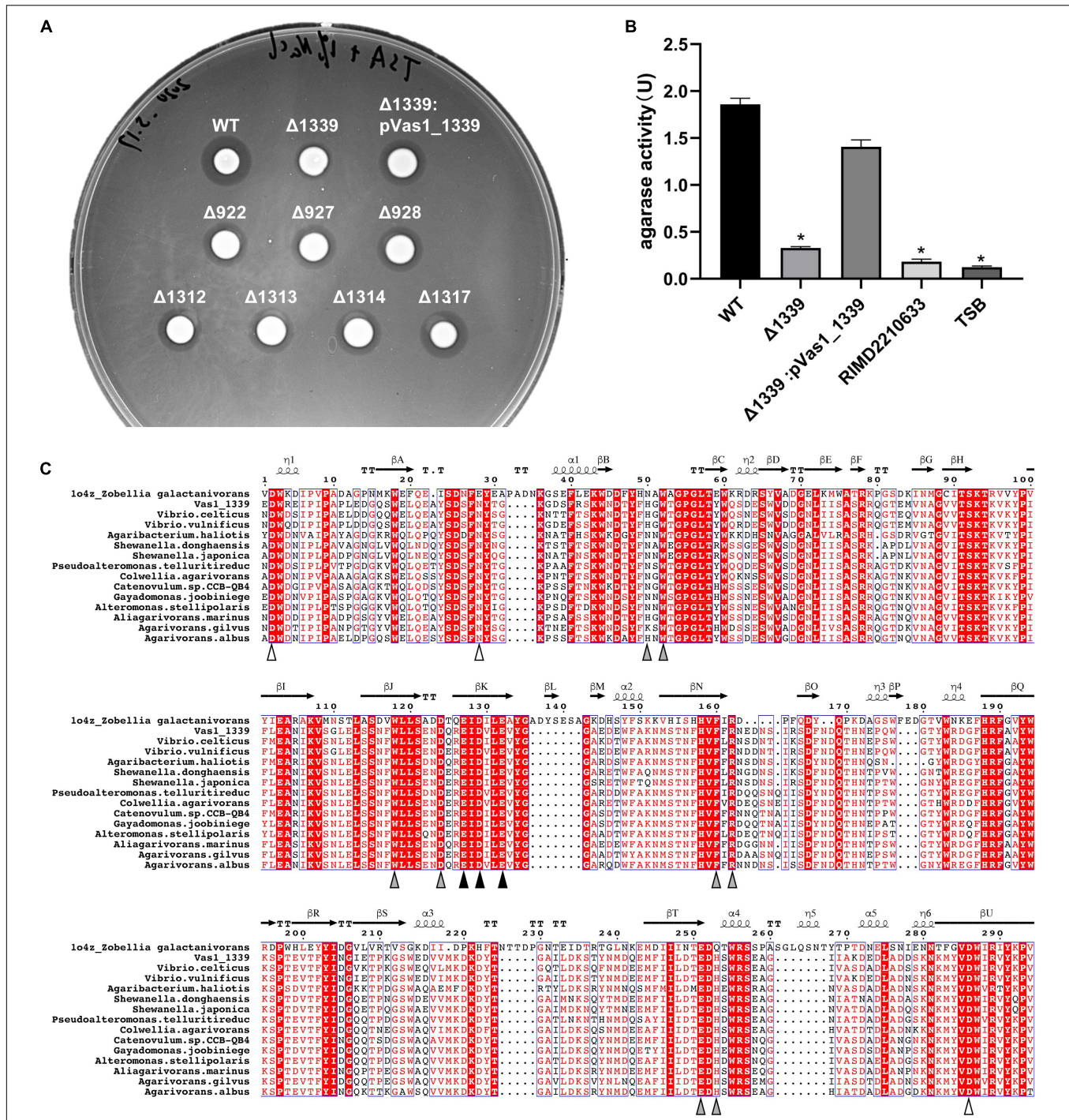


FIGURE 4 | A GH16₁₆-encoding gene *Vas1_1339* is indispensable for the agarolytic phenotype of HN897. **(A)** Results of agar-spot assays comparing the wild type (WT), deletion mutants and complementation strain. Each strain was spotted on a solid medium containing agar as substrate. The colonies that formed transparent lysis zones on agar plates were classified as agarolytic. **(B)** Agarase activity of cell-free culture supernatants was measured as described in section “Materials and Methods” from the wild type (WT), $\Delta 1339$ mutant and $\Delta 1339$:pVas1_1339 strains. The blank TSB medium and the culture supernatant of *V. parahaemolyticus* strain RIMD2210633 were included as controls. The data were expressed as means \pm SEM of agarase activity (U/mL) from three independent experiments. Statistical analysis was performed with one-way ANOVA and $P < 0.05$ were considered statistically significant. Pairwise comparisons were conducted using a Tukey–Kramer multiple comparisons test and the analyses were conducted using the SPSS software. **(C)** Multiple sequence alignment of the GH16 domain comparing *Vas1_1339* to other orthologous β -agarases (see **Supplementary Table S7** for details). The secondary structure information was based on resolved 3D structure of ZgAgaB (Protein Data Bank: 1o4z) from *Zobellia galactanivorans* (Naretto et al., 2019). η -(3)₁₀ and α -Helices were shown as helices; β -Strands were shown as arrows; and β -Turns were labeled with TT. The color-filled pointing out the calcium binding sites (white), active sites (gray) and catalytic residues (black), respectively. The figure was drawn by ESPript 3.0 (Robert and Gouet, 2014). See **Supplementary Figure S6** for the full-length sequence alignment.

assign several unassembled contigs from the original C7 based on the higher fidelity HN897 sequences (marked with asterisks in **Figure 1D**). For example, the first two longest protein-encoding genes in HN897 that contained dozens of tandem repeats were either not included (as for *Vas2_1198*) or not properly assembled (as for *Vas2_433*) in the C7 assembly (marked with shaded boxes in **Figure 1D**). Together these data demonstrated how newer sequencing technology and assembly tools can help retrieve additional insights from earlier sequencing efforts.

Many marine bacteria can be classified into two trophic modes of life namely copiotroph and oligotroph, that have evolved to thrive in distinct ocean habitats with high or low nutrient availability, respectively (Lauro et al., 2009). As one of the well-known clades of marine copiotrophs, the γ -proteobacterial order *Vibrionales* has notable potential to rapidly respond to different environmental clues (e.g., nutrient influx or depletion), which are achieved through a diverse array of metabolic pathways and fine-tuned signaling cascades (Lauro et al., 2009; Zhang X. H. et al., 2018). As mentioned earlier, the HN897 genome encodes two prominent genes *Vas2_1198* and *Vas2_433*, which are enriched with repetitive cadherin domains (**Figure 1D**). Proteins containing such domains may contribute to cell–cell contact of bacteria in the marine environment (Fraiberg et al., 2010), and confer binding to different insoluble carbohydrate substrates as carbohydrate-binding modules (Fraiberg et al., 2011). The HN897 genome encoded 156 CAZymes of which eight genes likely encoded β -agarases that belonged to GH subfamilies GH16 and GH50. Synteny analysis showed that GH16s and GH50s represented two groups of tandemly arrayed genes that were likely paralogous in origin. The only exception to this was *Vas1_1339* that was located on a separate chromosome from the other GH16s.

Interspecific comparisons amongst 46 genome assemblies demonstrated the two β -agarase gene clusters of HN897 shared only partial synteny with a few *Vibrios* (*Vibrio* sp. EJY3; *V. celticus* CECT7224), and species of the family *Alteromonadaceae* (*Agarivorans albus*, *Agarivorans gilvus*, and *Aliagarivorans marinus*), all of which have been functionally demonstrated to be agarolytic or a dominant member of epiphytic bacterial communities associated with marine macroalgae (López-Pérez and Rodríguez-Valera, 2014; Albakosh et al., 2016; Yu et al., 2020). β -agarase genes were frequently associated with insertion sequences, integrases and transposable elements (**Figure 3C**) consistent with an important role for horizontal gene transfer.

We demonstrated that one of the eight β -agarase genes from HN897 (*Vas1_1339*) was required for agarolytic degradation (**Figure 4A**). *Vas1_1339* belonged to the GH16 family that shared a common β -jelly roll fold and a catalytic motif ('ExDxxE' as shown in **Figure 4C** or 'ExDxE') within the active-site β -strand (Naretto et al., 2019; Viborg et al., 2019). Interestingly, two other *Vibrio* β -agarases, namely AgaA from *Vibrio* sp. PO-303 (Dong et al., 2007), and AgaV from *Vibrio* sp. V134 (Zhang and Sun, 2007) were assigned with *Vas1_1339* to the subfamily GH16_16. AgaV is an endo-acting β -agarase that degrades agarose into neoagarotetraose (NeoDP4) and NeoDP6 (Zhang and Sun, 2007).

None of the remaining seven β -agarase genes from HN897 appeared to contribute to agarolysis activity of

HN897. Three of them [*Vas2_922*(WP_164650518.1), *Vas2_927*(WP_164650523.1), *Vas2_928*] were recently classified into subfamily GH16_17 according to the latest CAZy database (update: 2020-05-01). GH16_17 is a small group containing carbohydrate-active κ -carrageenases (EC 3.2.1.83) from both Proteobacteria and Bacteroidetes (Viborg et al., 2019), yet only one *Vibrio*-originated CgkA (QGN18698.1; partial) from *Vibrio* sp. SY01 is recorded in addition to the three from this study. Nevertheless, a recent study identified a κ/ι -carrageenan metabolism pathway in several marine *Pseudoalteromonas* species, with a κ/ι -carrageenan-specific polysaccharide utilization locus (CarPUL) responsible for this phenotype (Hettle et al., 2019). Within the CarPUL of *Pseudoalteromonas fuliginea*, three GH16_17-coding genes namely GH16A (EU509_08815), GH16B (EU509_08870), and GH16C (EU509_08960) were separated with several intervening genes including sulfatase and major facilitator superfamily (MFS) transporter genes. Interestingly, the GH16_17s from HN897 shared 40–57% sequence identity with those of *P. fuliginea*. Given that both HN897 and C7 have orthologous genes associated with carrageenan metabolism (see **Supplementary Table S5**; Hettle et al., 2019), it was tempting to speculate that *V. astriarenae* has the ability to metabolize galactan carrageenan.

The GH50 family of proteins is smaller than the GH16 family, and is considered monospecific to agarose (Kim et al., 2010). To date, eight of eleven *Vibrio*-originated proteins with known β -agarase activity are classified as GH50s (Chi et al., 2012; Yu et al., 2020). For example, The *Vibrio* strain EJY3 has three tandemly arranged GH50s (VejGH50A, VejGH50B, and VejGH50C) that encode extracellular endo-type β -agarases responsible for degrading agarose into neoagarotetraose (NeoDP4) and NeoDP2. Strain EJY3 also has one exo-type β -agarase (VejGH50D) that acts intracellularly to depolymerize NeoDP4 into monomeric sugars (Yu et al., 2020). While most agarose saccharification (agarolysis) appears to involve initial extracellular depolymerization by endo-type β -agarase before transport into the cell (Pluvinage et al., 2013; Yu et al., 2020), such 'pre-depolymerization' may not be necessary for EJY3, considering the absence of both GH16 and GH86 in its genome (Roh et al., 2012). In contrast, AgaV from *V. astriarenae* strain V134 can decompose agarose into NeoDP4 and NeoDP6, besides, additional larger intermediate products including NeoDP8 and NeoDP10 can be also observed (Zhang and Sun, 2007). Consequently, we proposed that β -agarase-mediated agarose degradation in *V. astriarenae* was composed by two steps. This began with degradation of agarose into medium-sized DPs (e.g., NeoDP4 and NeoDP6) based on activity of GH16_16 (*Vas1_1339*) before both secreted endo-type and intracellular exo-type GH50 β -agarases alongside with other CAZymes, as collectively characterized in the strain EJY3, hydrolyze the DPs to yield monomeric sugars as a carbon source. Follow-up studies are needed to functionally characterize these agar-specific polysaccharide utilization loci and clarify the in-depth biochemical pathways of agarose degradation in *V. astriarenae* HN897.

In summary, we presented here the high-quality genomic sequence of an agarolytic *V. astriarenae* strain HN897, which

had an abundance of putative glycoside hydrolases (GH). Eight putative β -agarase encoding genes belonging to the GH16 and GH50 families were further identified based on sequence clustering and comparisons of protein domain architecture. Importantly, GH16_16 β -agarase-coding gene *Vas1_1339* was indispensable for the agarolytic phenotype of HN897, as evidenced by gene deletion and complementation experiments. These results not only provided the basis for understanding the biochemical mechanism of bacterial agarolysis, but also for gaining more detailed insights into the evolution and physiological capability of the genome of *V. astriarenae* HN897.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the NCBI GenBank under accession numbers CP047475 (*Vas1*) and CP047476 (*Vas2*).

AUTHOR CONTRIBUTIONS

YL, CW, XZ, and ML carried out the experiments. XJ and ZZ designed the experiments. XJ and YL analyzed the data. ZZ, XJ, and YL wrote the manuscript. DC contributed to phylogenetic analysis and manuscript preparation. All the authors have read and agreed to the published version of the manuscript.

FUNDING

This study was supported by the National Natural Science Foundation of China (31872597), the Natural Science Foundation of Jiangsu Province (BK20171431), the Jiangsu Agriculture Science and Technology Innovation Fund [CX(19)2033], and the Earmarked Fund for Jiangsu Agricultural Industry Technology System [JATS(2019)477].

ACKNOWLEDGMENTS

We thank Dr. Daofeng Zhang (Hohai University) for his suggestions in the early stage of writing the manuscript. We would like to thank the editor and the two reviewers, who provided comments and suggestions that helped to substantially improve the manuscript. The whole genome sequencing was performed by a commercial vendor (Personalbio, Shanghai, China).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2020.01404/full#supplementary-material>

FIGURE S1 | Two CG maps generated with the GView Server showed the full view of the chromosomes *Vas1* (A) and *Vas2* (B) of HN897, respectively. The contents of the feature rings (starting with the outermost ring) are as follows: Rings

1,2: positive stranded and Rings 3,4: negative stranded CDSs that were color-coded with COGs and RNA species; Ring 5 shows GC content and Ring 6 shows GC skew.

FIGURE S2 | Functional annotations of HN897 genome by COGs. Bar plot showed the numbers of genes that had been classified with each indicated COG group.

FIGURE S3 | Species counts and sequence alignment statistics of top BLAST hits for the genome sequence of *V. astriarenae* HN897. Pie charts showed the composition of hit species of homologous protein sequences aligned by BLASTp against the non-redundant (nr) protein database for *Vas1* (A) and *Vas2* (B). Bar plots showed the average percentage of identity and coverage of sequence alignment (alignment length divided by query length) of *Vas1* (C) and *Vas2* (D) against each indicated species.

FIGURE S4 | Full view of the maximum-likelihood species tree shown in **Figure 1B**. The bootstrap support (0~1) were labeled in each branch node.

FIGURE S5 | ML phylogenetic tree of 16S rRNA gene were estimated by FastTree2 using GTR + gamma model, based on the MAFFT-aligned sequences from Al-Saari et al. (2015) and other *Vibrios* which have high sequence similarity (> 98%) to HN897. Strains that had been referred in the texts were highlighted with red. *V. astriarenae* clade was shaded in gray. Sequence from *E. coli* K-12 substrain DH10B was used as outgroup.

FIGURE S6 | Multiple sequence alignment of full-length *Vas1_1339* with orthologous bacterial β -agarases (see **Supplementary Table S7** for full sequence list). The secondary structure information was based on resolved 3D structure of ZgAgaB (Protein Data Bank: 1o4z) from *Zobellia galactanivorans*. η -(3₁₀) and α -Helices were shown as helices; β -Strands were shown as arrows; and β -Turns were labeled with TT. The figure was drawn by ESPript 3.0 (Robert and Gouet, 2014).

TABLE S1 | Bacterial strains and plasmids used in this study.

TABLE S2 | Sequences for oligo primer pairs used in each indicated experiment.

TABLE S3 | Phylogenomics and BLAST sequence alignments. tab#1: Data related to **Supplementary Figures S3A–D**. Annotation output from BLASTP-NR-based sequence similarity comparisons for both *Vas1* and *Vas2*. tab#2: Data related to **Figures 1A,B**. The full standard metrics for HN897 assembly according to the Genomic Standards Consortium (GSC). Taxonomic classification and genome completeness assessment for *Vibrio astriarenae* HN897 that were performed by GTDB-tk 'classify_wf' and CheckM, respectively. tab#3: Data related to **Figure 1B** and **Supplementary Figure S4**. Metadata of 140 representative *Vibrio* genomes from GTDB release 89. tab#4: Data related to **Figures 1C,D**. tBLASTx hits between HN897 and C7, with threshold of Evalue 1e-10, alignment length 100, percent identity 90.

TABLE S4 | Functional annotations of HN897 genome. tab#1: Related to **Figure 2A**. Output from KAAS-based KEGG Orthology assignment using BLASTp with BBH (bi-directional best hit) method and against representative prokaryotes genes set. tab#2: Related to **Figure 2A**. KEGG pathway maps sorted by count numbers of associated KO terms resulted from KEGG mapper searching. tab#3: Related to **Figure 2B**. Protein domain modules annotated by PfamScan v1.6 searching against Pfam32.0 database. tab#4: Related to **Figure 2B**. Protein domain list sorted by count numbers of domain terms (shortname). tab#5: Related to **Figure 2B**. Gene Ontology annotations for Top 30 counted domain terms. tab#6: Related to **Figure 2C**. Gene counts for each indicated class of carbohydrate-active enzymes, CAZymes, for *Vas1* and *Vas2*. tab#7: Related to **Supplementary Figure S2**. Count numbers of associated genes annotated with indicated categories of Clusters of Orthologous Genes (COGs). tab#8: Related to **Supplementary Figure S3A**. Proteins sequences for a total of 156 identified CAZymes in HN897 genome.

TABLE S5 | Identification of the β -agarase gene repertoire in HN897. tab#1: Related to **Figure 3A**. Annotations for a total of 72 identified GH subfamilies in HN897 (Eight putative beta-agarase were highlighted). tab#2: Related to **Figures 1C,D, 3A,B**. InterProScan outputs for the protein domains of 72 Glycoside Hydrolases, together with the first two longest CDSs that contain the tandem repeat sequences discussed in manuscript. tab#3: Related to **Figure 3A**. Generic Feature Format Version 3 (GFF3) file of the catalytic protein domains for a

total of 72 GHs annotated by CAZy database. tab#4: Related to **Figure 3C**. A total of 45 subject genome assemblies used for sequence similarity searching of β -agarase gene cluster. tab#5: Related to **Figure 3C** Upper panel: The output of MultiGeneBLAST v1.1.14 concerning the comparisons of β -agarases containing regions on Vas1 (Four GH50s and a single GH16 Vas1_1339) among the investigated 46 genomes. tab#6: Related to **Figure 3C** Bottom panel: The output of MultiGeneBLAST v1.1.14 concerning the comparisons of β -agarases containing regions on Vas2 (Three GH16s) among the investigated 46 genomes.

REFERENCES

- Albakosh, M. A., Naidoo, R. K., Kirby, B., and Bauer, R. (2016). Identification of epiphytic bacterial communities associated with the brown alga *Splachnidium rugosum*. *J. Appl. Phy.* 28, 1891–1901. doi: 10.1007/s10811-015-0725-z
- Al-Saari, N., Gao, F., Rohul, A. A., Sato, K., Sato, K., Mino, S., et al. (2015). Advanced microbial taxonomy combined with genome-based approaches reveals that *Vibrio astriarenae* sp. nov., an agarolytic marine bacterium, forms a new clade in vibronaceae. *PLoS One* 10:e0136279. doi: 10.1371/journal.pone.0136279
- Aoki, T., Araki, T., and Kitamikado, M. (1990). Purification and characterization of a novel beta-agarase from *Vibrio* sp. AP-2. *Eur. J. Biochem.* 187, 461–465. doi: 10.1111/j.1432-1033.1990.tb15326.x
- Araki, T., Hayakawa, M., Zhang, L., Karita, S., and Morishita, T. (1998). Purification and characterization of agarases from a marine bacterium, *Vibrio* sp. PO-303. *J. Mar. Biotechnol.* 6, 260–265.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021
- Barbeyron, T., Thomas, F., Barbe, V., Teeling, H., Schenowitz, C., Dossat, C., et al. (2016). Habitat and taxon as driving forces of carbohydrate catabolism in marine heterotrophic bacteria: example of the model algae-associated bacterium *Zobellia galactanivorans* Dsj1(T). *Environ. Microbiol.* 18, 4610–4627. doi: 10.1111/1462-2920.13584
- Batty, E. M., Chaemchuen, S., Blacksell, S., Richards, A. L., Paris, D., Bowden, R., et al. (2018). Long-read whole genome sequencing and comparative analysis of six strains of the human pathogen *Orientia tsutsugamushi*. *PLoS Negl. Trop. Dis.* 12:e0006566. doi: 10.1371/journal.pntd.0006566
- Besemer, J., Lomsadze, A., and Borodovsky, M. (2001). GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* 29, 2607–2618. doi: 10.1093/nar/29.12.2607
- Bowers, R. M., Kyrpides, N. C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T. B. K., et al. (2017). Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and Archaea. *Nat. Biotechnol.* 35, 725–731. doi: 10.1038/nbt.3893
- Chaumeil, P. A., Mussig, A. J., Hugenholtz, P., and Parks, D. H. (2019). GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* 36, 1925–1927.
- Chi, W. J., Chang, Y. K., and Hong, S. K. (2012). Agar degradation by microorganisms and agar-degrading enzymes. *Appl. Microbiol. Biotechnol.* 94, 917–930. doi: 10.1007/s00253-012-4023-2
- Chi, W. J., Seo, Y. B., Chang, Y. K., Lee, S. Y., and Hong, S. K. (2014). Cloning, expression, and biochemical characterization of a novel GH16 β -agarase AgaG1 from *Alteromonas* sp. GNUM-1. *Appl. Microbiol. Biotechnol.* 98, 4545–4555. doi: 10.1007/s00253-014-5510-4
- Chin, C. S., Peluso, P., Sedlazeck, F. J., Nattestad, M., Concepcion, G. T., Clum, A., et al. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* 13, 1050–1054. doi: 10.1038/nmeth.4035
- Delcher, A. L., Kasif, S., Fleischmann, R. D., Peterson, J., White, O., and Salzberg, S. L. (1999). Alignment of whole genomes. *Nucleic Acids Res.* 27, 2369–2376.
- Dong, J., Hashikawa, S., Konishi, T., Tamaru, Y., and Araki, T. (2006). Cloning of the novel gene encoding beta-agarase C from a marine bacterium, *Vibrio* sp. strain PO-303, and characterization of the gene product. *Appl. Environ. Microbiol.* 72, 6399–6401. doi: 10.1128/aem.00935-06
- Dong, J. H., Tamaru, Y., and Araki, T. (2007). A unique β -agarase, AgaA, from a marine bacterium, *Vibrio* sp. strain PO-303. *Appl. Microbiol. Biotechnol.* 74:1248. doi: 10.1007/s00253-006-0781-z
- Ekborg, N. A., Taylor, L. E., Longmire, A. G., Henriessat, B., Weiner, R. M., and Hutcheson, S. W. (2006). Genomic and proteomic analyses of the agarolytic system expressed by *Saccharophagus degradans* 2-40. *Appl. Environ. Microbiol.* 72, 3396–3405. doi: 10.1128/aem.72.5.3396-3405.2006
- Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., et al. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44, D279–D285.
- Fraiberg, M., Borovok, I., Bayer, E. A., Weiner, R. M., and Lamed, R. (2011). Cadherin domains in the polysaccharide-degrading marine bacterium *Saccharophagus degradans* 2-40 are carbohydrate-binding modules. *J. Bacteriol.* 2011, 283–285. doi: 10.1128/jb.00842-10
- Fraiberg, M., Borovok, I., Weiner, R. M., and Lamed, R. (2010). Discovery and characterization of cadherin domains in *Saccharophagus degradans* 2-40. *J. Bacteriol.* 192, 1066–1074. doi: 10.1128/jb.01236-09
- Fu, X., and Kim, S. (2010). Agarases: review of major sources, categories, purification method, enzyme characteristics and application. *Mar. Drugs* 8, 200–218. doi: 10.3390/md8010200
- Fu, X. T., Pan, C. H., Lin, H., and Kim, S. M. (2009). Gene cloning, expression, and characterization of a beta-agarase, agaB34, from *Agarivorans albus* YKW-34. *J. Microbiol. Biotechnol.* 19, 257–264.
- Gao, Z. M., Xiao, J., Wang, X. N., Ruan, L. W., Chen, X. L., and Zhang, Y. Z. (2012). *Vibrio xiamenensis* sp. nov., a cellulase-producing bacterium isolated from mangrove soil. *Int. J. Syst. Evol. Microbiol.* 62(Pt 8), 1958–1962. doi: 10.1099/ijs.0.033597-0
- González-Castillo, A., Enciso-Ibarra, J., Dubert, J., Romalde, J. L., and Gomez-Gil, B. (2016). *Vibrio sonorensis* sp. nov. isolated from a cultured oyster *Crassostrea gigas*. *Antonie Van Leeuwenhoek* 109, 1447–1455. doi: 10.1007/s10482-016-0744-z
- González-Castillo, A., Enciso-Ibarra, J., Bolán-Mejía, M. C., Balboa, S., Lasa, A., Romalde, J. L., et al. (2015). *Vibrio mexicanus* sp. nov., isolated from a cultured oyster *Crassostrea corteziensis*. *Antonie Van Leeuwenhoek* 108, 355–364. doi: 10.1007/s10482-015-0488-1
- Hehemann, J. H., Correc, G., Thomas, F., Bernard, T., Barbeyron, T., Jam, M., et al. (2012). Biochemical and structural characterization of the complex agarolytic enzyme system from the marine bacterium *Zobellia galactanivorans*. *J. Biol. Chem.* 287, 30571–30584. doi: 10.1074/jbc.m112.377184
- Hettle, A. G., Hobbs, J. K., Pluvinage, B., Vickers, C., Abe, K. T., Salama-Alber, O., et al. (2019). Insights into the κ/ι -carrageenan metabolism pathway of some marine *Pseudoalteromonas* species. *Commun. Biol.* 2:474.
- Hsu, P. H., Wei, C. H., Lu, W. J., Shen, F., Pan, C. L., and Lin, H. T. (2015). Extracellular production of a novel endo- β -agarase AgaA from *Pseudomonas vesicularis* MA103 that cleaves agarose into neoagarotetraose and neoagarohexaose. *Int. J. Mol. Sci.* 16, 5590–5603. doi: 10.3390/ijms16035590
- Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., von Mering, C., et al. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. *Mol. Biol. Evol.* 34, 2115–2122. doi: 10.1093/molbev/msx148
- Hunt, D. E., Gevers, D., Vahora, N. M., and Polz, M. F. (2008). Conservation of the chitin utilization pathway in the Vibrionaceae. *Appl. Environ. Microbiol.* 74, 44–51. doi: 10.1128/aem.01412-07
- Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031

- Jonnadula, R., and Ghadi, S. C. (2011). Purification and characterization of β -agarase from seaweed decomposing bacterium *Microbulbifer* sp. strain CMC-5. *Biotechnol. Bioprocess. Eng.* 16, 513–519. doi: 10.1007/s12257-010-0399-y
- Kang, N. Y., Choi, Y. L., Cho, Y. S., Kim, B. K., Jeon, B. S., Cha, J. Y., et al. (2003). Cloning, expression and characterization of a β -agarase gene from a marine bacterium, *Pseudomonas* sp. SK38. *Biotechnol. Lett.* 25, 1165–1170.
- Kim, H. T., Lee, S., Lee, D., Kim, H. S., Bang, W. G., Kim, K. H., et al. (2010). Overexpression and molecular characterization of Aga50D from *Saccharophagus degradans* 2-40: an exo-type beta-agarase producing neoagarobiose. *Appl. Microbiol. Biotechnol.* 86, 227–234. doi: 10.1007/s00253-009-2256-5
- Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27, 722–736. doi: 10.1101/gr.215087.116
- Krogh, A., Brown, M., Mian, I. S., Sjolander, K., and Haussler, D. (1994). Hidden markov models in computational biology: applications to protein modeling. *J. Mol. Biol.* 235, 1501–1531.
- Lauro, F. M., McDougald, D., Thomas, T., Williams, T. J., Egan, S., Rice, S., et al. (2009). The genomic basis of trophic strategy in marine bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 106, 15527–15533. doi: 10.1073/pnas.0903507106
- Lee, C. H., Kim, H. T., Yun, E. J., Lee, A. R., Kim, S. R., Kim, J. H., et al. (2014). A novel agarolytic β -galactosidase acts on agarooligosaccharides for complete hydrolysis of agarose into monomers. *Appl. Environ. Microbiol.* 80, 5965–5973. doi: 10.1128/aem.01577-14
- Lee, D. G., Jeon, M. J., and Lee, S. H. (2012). Cloning, expression, and characterization of a glycoside hydrolase family 118 beta-agarase from *Agarivorans* sp. JA-1. *J. Microbiol. Biotechnol.* 22, 1692–1697. doi: 10.4014/jmb.1209.09033
- Lee, W. K., Lim, Y. Y., Leow, A. T., Namasivayam, P., Ong, A. J., and Ho, C. L. (2017). Biosynthesis of agar in red seaweeds: a review. *Carbohydr. Polym.* 164, 23–30. doi: 10.1016/j.carbpol.2017.01.078
- Letunic, I., and Bork, P. (2018). 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res.* 46, D493–D496.
- Letunic, I., and Bork, P. (2019). Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259.
- Liao, L., Xu, X. W., Jiang, X. W., Cao, Y., Yi, N., Huo, Y. Y., et al. (2011). Cloning, expression, and characterization of a new beta-agarase from *Vibrio* sp. strain CN41. *Appl. Environ. Microbiol.* 77, 7077–7079. doi: 10.1128/aem.05364-11
- Lin, B., Lu, G., Zheng, Y., Xie, W., Li, S., and Hu, Z. (2012). Gene cloning, expression and characterization of a neoagarotetraose-producing β -agarase from the marine bacterium *Agarivorans* sp. HZ105. *World J. Microbiol. Biotechnol.* 28, 1691–1697. doi: 10.1007/s11274-011-0977-y
- Liu, J., Zhao, Z., Deng, Y., Shi, Y., Liu, Y., Wu, C., et al. (2017). Complete genome sequence of *Vibrio campbellii* LMB 29 isolated from red drum with four native megaplasmids. *Front. Microbiol.* 8:2035. doi: 10.3389/fmicb.2017.02035
- López-Pérez, M., and Rodriguez-Valera, F. (2014). “The family Alteromonadaceae,” in *The Prokaryotes*, eds E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson (Berlin: Springer), 575–582.
- Lucena, T., Ruvira, M. A., Arahal, D. R., Macián, M. C., and Pujalte, M. J. (2012). *Vibrio aestivus* sp. nov. and *Vibrio quintilis* sp. nov., related to Marisflavi and Gazogenes clades, respectively. *Syst. Appl. Microbiol.* 35, 427–431. doi: 10.1016/j.syapm.2012.08.002
- Luo, R. B., Liu, B. H., Xie, Y. L., Li, Z. Y., Huang, W. H., Yuan, J. Y., et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1:18.
- Makino, K., Oshima, K., Kurokawa, K., Yokoyama, K., Uda, T., Tagomori, K., et al. (2003). Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism distinct from that of *V. cholera*. *Lancet* 361, 743–749. doi: 10.1016/S0140-6736(03)12659-1
- Medema, M. H., Takano, E., and Breitling, R. (2013). Detecting sequence homology at the gene cluster level with MultiGeneBlast. *Mol. Biol. Evol.* 30, 1218–1223. doi: 10.1093/molbev/mst025
- Mendler, K., Chen, H., Parks, D. H., Lobb, B., Hug, L. A., and Doxey, A. C. (2019). AnnoTree: visualization and exploration of a functionally annotated microbial tree of life. *Nucleic Acids Res.* 47, 4442–4448. doi: 10.1093/nar/gkz246
- Mikkel, S., Stinus, L., and Ludovic, O. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* 9:88. doi: 10.1186/s13104-016-1900-2
- Miller, G. L. (1959). Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal. Biochem.* 31, 426–428. doi: 10.1021/ac60147a030
- Milton, D. L., O’Toole, R., Horsted, T. P., and Wolf-Watz, H. (1996). Flagellin A is essential for the virulence of *Vibrio anguillarum*. *J. Bacteriol.* 178, 1310–1319. doi: 10.1128/jb.178.5.1310-1319.1996
- Morales, V. M., Bäckman, A., and Bagdasarian, M. (1991). A series of wide-host-range low-copy-number vectors that allow direct screening for recombinants. *Gene* 97, 39–47. doi: 10.1016/0378-1119(91)90007-x
- Naretto, A., Fanuel, M., Ropartz, D., Rogniaux, H., Larocque, R., Czjzek, M., et al. (2019). The agar-specific hydrolase ZgAgaC from the marine bacterium *Zobellia galactanivorans* defines a new GH16 protein subfamily. *J. Biol. Chem.* 294, 6923–6939. doi: 10.1074/jbc.ra118.006609
- Ohta, Y., Hatada, Y., Nogi, Y., Miyazaki, M., Li, Z., Akita, M., et al. (2004). Enzymatic properties and nucleotide and amino acid sequences of a thermostable beta-agarase from a novel species of deep-sea *Microbulbifer*. *Appl. Microbiol. Biotechnol.* 64, 505–514. doi: 10.1007/s00253-004-1573-y
- Osterholz, H., Singer, G., Wemheuer, B., Daniel, R., Simon, M., Niggemann, J., et al. (2016). Deciphering associations between dissolved organic molecules and bacterial communities in a pelagic marine system. *ISME J.* 10, 1717–1730. doi: 10.1038/ismej.2015.231
- Papadopoulos, J. S., and Agarwala, R. (2007). COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics* 23, 1073–1079. doi: 10.1093/bioinformatics/btm076
- Parks, D. H., Chuvochina, M., Chaumeil, P. A., Rinke, C., Mussig, A. J., and Hugenholtz, P. (2020). A complete domain-to-species taxonomy for bacteria and Archaea. *Nat. Biotechnol.* doi: 10.1038/s41587-020-0501-8 [Epub ahead of print].
- Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarszewski, A., Chaumeil, P. A., et al. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* 36, 996–1004. doi: 10.1038/nbt.4229
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi: 10.1101/gr.186072.114
- Parte, A. C. (2018). LPSN – list of prokaryotic names with standing in nomenclature (bacterio.net), 20 years on. *Int. J. Syst. Evol. Microbiol.* 68, 1825–1829. doi: 10.1099/ijsem.0.002786
- Pluvinage, B., Hehemann, J. H., and Boraston, A. B. (2013). Substrate recognition and hydrolysis by a family 50 exo-beta-agarase, Aga50D, from the marine bacterium *Saccharophagus degradans*. *J. Biol. Chem.* 288, 28078–28088. doi: 10.1074/jbc.m113.491068
- Price, M. N., and Arkin, A. P. (2017). PaperBLAST: text mining papers for information about homologs. *mSystems* 2:e00039-17.
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. doi: 10.1371/journal.pone.0009490
- Renn, D. (1997). Biotechnology and the red seaweed polysaccharide industry: status, needs and prospects. *Trends Biotechnol.* 15, 9–14. doi: 10.1016/s0167-7799(96)10069-x
- Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* 42, W320–W324.
- Roh, H., Yun, E. J., Lee, S., Ko, H. J., Kim, S., Kim, B. Y., et al. (2012). Genome sequence of *Vibrio* sp. strain EJY3, an agarolytic marine bacterium metabolizing 3,6-anhydro-L-galactose as a sole carbon source. *J. Bacteriol.* 194, 2773–2774. doi: 10.1128/jb.00303-12
- Saurin, W., Hofnung, M., and Dassa, E. (1999). Getting in or out: early segregation between importers and exporters in the evolution of ATP-binding cassette (ABC) transporters. *J. Mol. Evol.* 48, 22–41. doi: 10.1007/pl00006442
- Sawabe, T., Ogura, Y., Matsumura, Y., Feng, G., Amin, A. R., Mino, S., et al. (2013). Updating the *Vibrio* clades defined by multilocus sequence phylogeny: proposal of eight new clades, and the description of *Vibrio tritonius* sp. nov. *Front. Microbiol.* 4:414. doi: 10.3389/fmicb.2013.00414

- Sheu, S. Y., Jiang, S. R., Chen, C. A., Wang, J. T., and Chen, W. M. (2011). *Vibrio stylophorae* sp. nov., isolated from the reef-building coral *Stylophora pistillata*. *Int. J. Syst. Evol. Microbiol.* 61(Pt 9), 2180–2185. doi: 10.1099/ijs.0.026666-0
- Stothard, P., Grant, J. R., and Van Domselaar, G. (2019). Visualizing and comparing circular genomes using the CGView family of tools. *Brief. Bioinform.* 20, 1576–1582. doi: 10.1093/bib/bbx081
- Sugano, Y., Matsumoto, T., Kodama, H., and Noma, M. (1993). Cloning and sequencing of agaA, a unique agarase 0107 gene from a marine bacterium, *Vibrio* sp. strain JT0107. *Appl. Environ. Microbiol.* 59, 3750–3756. doi: 10.1128/aem.59.11.3750-3756.1993
- Thompson, F. L., Iida, T., and Swings, J. (2004). Biodiversity of *Vibrios*. *Microbiol. Mol. Biol. Rev.* 68, 403–431. doi: 10.1128/mmbr.68.3.403-431.2004
- Thompson, J. R., and Polz, M. F. (2006). “Dynamics of vibrio population and their role in environmental nutrient cycling,” in *The Biology of Vibrios*, eds F. L. Thompson, B. Austin, and J. Swings (Washington, DC: ASM), 190–203. doi: 10.1128/9781555815714.ch13
- Tritt, A., Eisen, J. A., Facciotti, M. T., and Darling, A. E. (2012). An integrated pipeline for de novo assembly of microbial genomes. *PLoS One* 7:e42304. doi: 10.1371/journal.pone.0042304
- Viborg, A. H., Terrapon, N., Lombard, V., Michel, G., Czjzek, M., Henrissat, B., et al. (2019). A subfamily roadmap of the evolutionarily diverse glycoside hydrolase family 16 (GH16). *J. Biol. Chem.* 294, 15973–15986. doi: 10.1074/jbc.ra119.010619
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. doi: 10.1371/journal.pone.0112963
- Wang, H., Liu, J., Wang, Y., and Zhang, X. H. (2011). *Vibrio marisflavi* sp. nov., isolated from seawater. *Int. J. Syst. Evol. Microbiol.* 61(Pt 3), 568–573. doi: 10.1099/ijs.0.022285-0
- Xie, W., Lin, B., Zhou, Z., Lu, G., Lun, J., Xia, C., et al. (2013). Characterization of a novel β -agarase from an agar-degrading bacterium *Catenovulum* sp. X3. *Appl. Microbiol. Biotechnol.* 97, 4907–4915. doi: 10.1007/s00253-012-4385-5
- Yu, S., Yun, E. J., Kim, D. H., Park, S. Y., and Kim, K. H. (2020). Molecular and enzymatic verification of the dual agarolytic pathways in a marine bacterium, *Vibrio* sp. strain EJY3. *Appl. Environ. Microbiol.* 86:e02724-19.
- Yun, E. J., Lee, S., Kim, H. T., Pelton, J. G., Kim, S., Ko, H. J., et al. (2015). The novel catabolic pathway of 3,6-anhydro-L-galactose, the main component of red macroalgae, in a marine bacterium. *Environ. Microbiol.* 17, 1677–1688.
- Yun, E. J., Yu, S., and Kim, K. H. (2017). Current knowledge on agarolytic enzymes and the industrial potential of agar-derived sugars. *Appl. Microbiol. Biotechnol.* 101, 5581–5589.
- Zhang, J. Y., Liu, Y. P., Wu, C., Deng, Y. Q., Yin, H. X., and Zhao, Z. (2018). Isolation and enzyme activity assay of an agarase-producing *Vibrio* sp. strain (In Chinese). *J. Trop Biol.* 2, 142–146.
- Zhang, W. W., and Sun, L. (2007). Cloning, characterization, and molecular application of a beta-agarase gene from *Vibrio* sp. Strain V134. *Appl. Environ. Microbiol.* 73, 2825–2831.
- Zhang, X. H., Lin, H. Y., Wang, X. L., and Austin, B. (2018). Significance of *Vibrio* species in the marine organic carbon cycle—a review. *Science China (Earth Sciences)* 61, 1357–1368.
- Zhao, Z., Liu, J. X., Deng, Y. Q., Huang, W., Ren, C. H., Call, D. R., et al. (2018). The *Vibrio alginolyticus* T3SS effectors, Val1686 and Val1680, induce cell rounding, apoptosis and lysis of fish epithelial cells. *Virulence* 9, 318–330.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Liu, Jin, Wu, Zhu, Liu, Call and Zhao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.