Genome analysis of the alginate degrading marine bacterium *Vibrio halioticoli*

（*Vibrio halioticoli* のゲノム解析）

北海道大学大学院水産科学院

海洋応用生命科学専攻

Graduate School of Fisheries Sciences
Division of Marine Life Science

高 峰

Gao Feng

2018（平成 30 年）

CONTENTS

Chapter 1. General Introduction

In maintaining a sustainable ecosystem in this period of global warming, the development of key technologies for renewable energy sources has become an important challenge. Bioethanol, bio-higher alcohol and biohydrogen, produced from various carbohydrates present in terrestrial plants, are potentially important renewable energy sources using well controlled microbial energy conversion metabolism (Ingram et al., 1987; Ohta et al., 1991; Dien et al., 2003; Stephanopoulos, 2007; Atsumi et al., 2008; Atsumi et al., 2010; Lee et al., 2010; Dürre and Richard, 2011; Geddes et al., 2011). Biofuel production, however, involves a paradox in the energy-food conflict, so we need to find alternative fermentation substrates and unique microbes for the development of food-uncompetitive biofuel production for use in the future.

In the meantime, successful constructions of metabolically engineered *Escherichia coli* and *Sphingomonas* sp. implemented with homo-ethanologenic pathways have been achieved (Takeda et al., 2011; Wargacki et al., 2012). In both strains, alginate degradation and utilization are the key metabolic pathways in bioethanol production using fast growing kelp as feed stocks. Kelp is cultured and/or harvested not only as a food crop but also for industrial purposes such as polysaccharide production, especially in Asian countries such as Japan and China, and also in Chile (McHugh, 2003). In 2014, the global annual production of kelp reached to 7.6 million tonnes wet weight (ww) (Fig. 1.1). Kelp is not a staple crop, and used as animal feeds, agricultural fertilizers, and sources of polymers. The industrial utilization rate of the kelp is only 30% (dry weight (dw)), the major products are algin, mannitol and iodine, and about 60% of the kelp components (such as

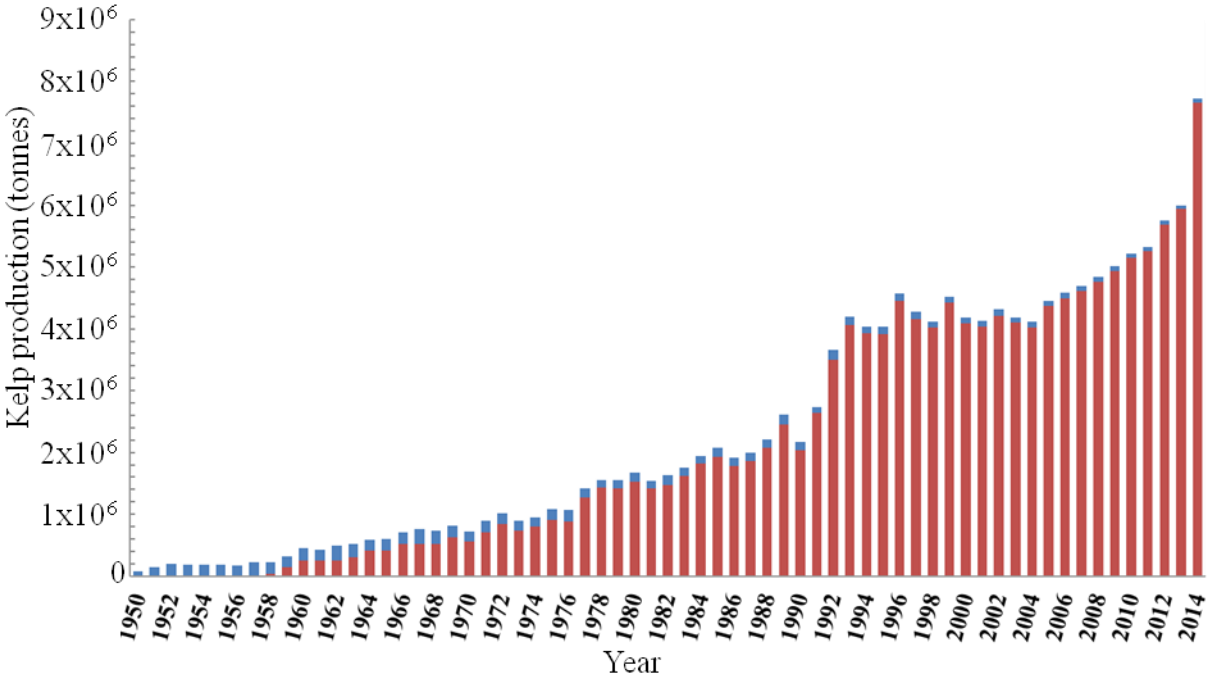

Fig. 1.1. Kelp production in the world from 1950 to 2014. : Culture, : Wild.

sodium alginate, dietary fiber) still cannot be utilized and discharged as waste (Tang, 2003). According to FAO data, China is responsible for 80 % of globally farmed kelp in 2002. In particular, kelp production was 62.2 tonnes ww in 1952 but by 1958 it had dramatically increased over 10000 tonnes ww (Jin et al., 2009). From the 1960s to the 1980s, the development of the algin production sector further pushed Chinese kelp farming to a higher level, and by 2000 Chinese production had reached over 4 million tonnes/yr ww. As kelp grows very quickly, traditional kelp aquaculture could easily be refined to adapt to energy-aimed cultivation systems, recently called "The Energy Line" (Stokstad, 2012). However, *E. coli* and *Sphingomonas* do not naturally occur in marine environments and are incapable of showing their best fermentation performance in saline environments using kelp as feedstock. We need to find more suitable microbes to improve performance in biofuel production from seaweed feedstock under marine conditions (Hidaka and Sakai, 1968).

Alginate (alginic acid) is a major carbohydrate contained in kelp ranging from 20% to 40% (dw) (Takeda et al., 2011). The carbohydrate is liner polysaccharide, composed of 1-4 linked β-D-mannuronic acid (M) and the C-5 epimer α-L-guluronic acid (G) as the unit structures, and generally divided into three types, polyM block, polyG block and MG random block (Gacesa, 1988; Gacesa, 1992; Wong et al., 2000). To date, the degradation of alginate has been characterized in some microorganisms. In the microorganisms, alginate lyase catalyzes the depolymerization of alginate into metabolizable units via a β-elimination reaction (Gacesa, 1992; Wong et al., 2000; Ochiai et al., 2006; Zhu and Yin, 2015). From the mode of degradation of alginate, alginate lyase can be classified as endotype alginate lyase or exotype alginate lyase. Alginate lyase also can be

further classified into polyG and polyM specific lyases that can degrade each block of alginate, respectively. A unique alginate lyase with broad substrate specificity from *Pseudoalteromonas elyakovii* has also been described (Sawabe et al., 1997). In most bacteria, oligomers are further degraded into unsaturated monomers by exotype alginate lyase often described as oligoalginate lyase (Gacesa, 1992; Wong et al., 2000; Yamasaki et al., 2004; Kim et al., 2012a; Kim et al., 2012b). Surprisingly, a bacterium possessing a super channel on the cell surface, *Sphingomonas* sp. A1, which could assimilate insoluble alginate macromolecule using alginate binding proteins and ATP-binding cassette transporters has also been reported (Takeda et al., 2011). In both cases, alginate and/or alginate oligomers are converted into unsaturated uronic acids, and further metabolized to 4-deoxy-L-*erythro*-5-hexoseulose uronic acid (DEH) by alginate lyase. After that, DEH was converted into 2-keto-3-deoxy-D-gluconate (KDG) using dehydrogenases, which is further phosphorylated by a carbohydrate kinase into 2-keto-3-deoxy-phosohogluconate (KDGP). Afterword, through the Entner-Doudoroff pathway the KDGP can be directly assimilated and further converted into ethanol (Takeda et al., 2011). Through the pathway in alginate metabolism, excess one molecule of ATP is generated but no excess reducing equivalents are produced, which means alginate if used as a single substrate is not a suitable, but necessary, substrate for ethanol production (Takeda et al., 2011; Wargacki et al., 2012; Trinh et al., 2008). Therefore, to enable alginate catabolism, a counterbalance to the excess-reducing equivalents must be preconditioned.

Vibrios are a large group containing marine facultative anaerobic bacteria (Farmer et al., 2005). The bacteria are widespread in marine environments, but one of the major habitats is in the gut of marine fish/shellfish. Among marine fish/shellfish species, there are a huge variety of kelp (or other

seaweed) eating animals. We expected the presence of unique marine bacteria which can ferment seaweed carbohydrates in the digestive system (gut) of such herbivorous marine animals (Hehemann et al., 2010). In fact, *Vibrio halioticoli* and related species have been isolated from *Haliotis* abalone (Gastropoda) as a newly described alginate degrading vibrio (Sawabe et al., 1998; Sawabe et al., 2003; Sawabe, 2006). The bacterium shows degradation of alginate, anaerobically produced acid from glucose without gas, Na^+ requirement, oxidase-positive, and non-motile. From the description of Sawabe et al. (2013), this bacterium has been placed in the Halioticoli clade with nine other *Vibrio* species (Sawabe et al., 2004, 2007a and 2013). The bacterial species of the Halioticoli clade are regarded as possible symbionts that have relationships with the host herbivorous animals in the digestion and conversion of alginate to short chain fatty acids (Sawabe et al., 2004, Sawabe, 2006). More recently, the *V. halioticoli* cell has been used to create a new biocatalyst implemented with Production of Ethanol cassette (PET operon) and the engineered vibrio actually produced ethanol from alginate (Inohara, 2014). The pathway implemented by the PET operon can save reducing equivalents to forward in production of ethanol. To improve the biocatalysts in producing bioethanol more efficiently, the detailed metabolic pathway of alginate and the gene expression controls of such genes in the bacterium should be elucidated.

The advent of whole genome sequence provided a more efficient way to elucidate the detailed metabolic pathway and the gene function. Follow with the development of sequencing technical, genome sequence allowed the establishment of taxonomic schemes (such as average amino acid identity, supertrees, genome-to-genome distance, and the Karlin signature) (Amaral et al., 2014). Along with additional tools, they can comprehensively analyze and classify thousands of genomes.

These new tools have led to new understandings of genetic relationships better than the 16S rRNA gene only approximates (Land et al., 2015). Not only this, these also help us to gain a better understanding of the ecology of these microbes rather than the classic phenotypic characterization using time-consuming laborious wet laboratory tests. It is important to obtain phenotypic information based on whole genome sequences (Sim et al., 2012).

In this study, the alginate loving bacterium, *V. halioticoli* IAM 14596^{T}, was used to investigate the alginate metabolic pathway. To achieve this aim, whole genome sequencing was performed and genome wide mining of genes responsible for alginate degradation was conducted.

Chapter 2. Whole genome sequencing and analysis of *Vibrio halioticoli* IAM 14596^T

ABSTRACT

In this chapter, the complete genome sequence of alginate loving *Vibrio halioticoli* IAM 14596^T is determined. The larger (Chromosome 1 (Chr. 1)) and smaller (Chromosome 2 (Chr. 2)) circular chromosomes comprised of 2,785,698 bp and 1,098,310 bp, respectively. In addition to these two chromosomes, one plasmid comprised 244,363 bp. The 21 kb identical region was found between Chr. 2 and the plasmid. Genes responsible for conjugation were found on this plasmid, and suggesting the function as a conjugative plasmid. The average G+C content was 42.9%. Annotation results showed a total of 3,602 coding sequences (CDSs), and 75% of these CDSs are annotated. The vast majority of recognizable genes are for metabolism (such as carbohydrate, energy, nucleotide, amino acid, cofactors and vitamins metabolism), genetic information processing (such as translation, replication, and repair) and environmental information processing (such as membrane transport, and signal transduction). *V. halioticoli* possessed genes responsible for three central carbohydrate metabolic pathways (EM, PP, and ED), TCA cycle, glyoxylate cycle, and mixed acid fermentation pathways. In addition, the variety of transporters, secretion systems, and protein export systems were predicted in this chapter. Pathways responsible to utilization of mannitol and alginate were also reconstructed. The small chromosome contains more CDSs with unknown functions (45% in annotation rate), but Chr. 2 contains many more genes responsible for alginate degradation and assimilation.

INTRODUCTION

Since the first two successful complete genome analyses of bacteria, biology and biotechnology of prokaryotes have dramatically changed due to technical improvements in genome sequence technologies and developments of genome wide gene mining tools in silico (Fraser et al., 1995; Land et al., 2015). This also contributes to establishing the field of "Bioinformatics", the accumulation of genetic data, and the establishment of public databases. In particular, the establishment of nucleotide and protein databases is noteworthy. In 1979, Walter Goad and colleagues at the Theoretical Biology and Biophysics Group at Los Alamos National Laboratory established the Los Alamos Sequence Database, which led to the creation of the public GenBank in 1982 (Strasser, 2011). In the early 1990's this responsibility was awarded to NCBI (The NCBI handbook, available on line at http://www.ncbi.nlm.nih.gov/books/NBK21101/). Currently, GenBank is the "right-hand man" of gene analysis, and the amount of recorded information is huge (Ilene, 2007). GenBank contains publicly available nucleotide sequences for more than 400,000 organisms and helps in both developing and implementing individual and genome sequence submissions and search tools (Benson et al., 2007, 2011 and 2017). The genome sequence data of 50 bacterial phyla and 11 archaeal phyla are also currently available, nevertheless ca. 90% of bacterial genomes have not yet been completed (Land et al., 2015). The advances in genome sequence technologies have directed us to spotlight on uncultured organisms, to establish a portable classification system of prokaryotic taxa, and finding novel prokaryotic immunity against viruses such as the clustered regularly interspaced shot palindromic repeats (CRISPR)-Cas system, to elucidate evolution and population genetics of prokaryotes in the aspect of core- and pan-genomes, and to open state-of-art strategies in genome-scale metabolic modelling of prokaryotes (Land et al.,

2015).

Vibrionaceae is a large group of Gram-negative bacteria, possessing a curved-rod shape and sodium ion requirement for growth (Gomez-Gil et al., 2014; Holt et al., 1994). It is ubiquitous in aquatic environment and can be isolated from both marine and brackish habitats. Several species are associated with host-microbe interactions. For example, *Vibrio coralliilyticus* and *Vibrio harveyi* are marine animal pathogens, *Vibrio cholerae*, *Vibrio parahaemolyticus* and *Vibrio vulnificus* are human pathogens, also some species form mutualistic relationships with marine organisms. As a versatile metabolism, *Vibrionaceae* have attracted a lot of attention towards their physiology and taxonomy. *Vibrionaceae* is also a reading bacterial taxon in which genomic taxonomy is applied (Thompson et al., 2009). On the basis of the methodology, the species delineation with average amino acid identity (ANI), supertrees, genome-to-genome distance, and the Karlin signature becomes a clear, standard and portable scheme in bacterial taxonomy (Thompson et al., 2009). Genome sequence data is also effective in genus description in *Vibrionaceae* (Amin et al., 2017). Currently, 2,401 vibrio genomes are available in the NCBI genome repository, but only 65% of each type strain of the *Vibrionaceae* species has been sequenced, and only 30 species genomes have been completed (https://www.ncbi.nlm.nih.gov/genome/). To push forward the biology of vibrios including their biotechnological application, completed genome sequence and information must be accumulated, and this could contribute to phenotype prediction on the basis of whole genome sequences (Amaral et al., 2014).

V. halioticoli was isolated from the gut of the abalone *Haliotis discus hannai* (Sawabe et al., 1995;

Sawabe et al., 1998; Sawabe et al., 2006). The bacterium has a high native ability to ferment mannitol and alginate. The wild type cannot convert alginate to ethanol directly, but the metabolically engineered cells with installed PET operon were able to produce ethanol (Inohara, 2016). The *V. halioticoli*-like strains are widespread in the gut of abalone living in the word's coastal habitats. Currently 10 species in the Halioticoli clade have been described, including the newly described *Vibrio ishigakensis* (see Chapter 4 in detail), and symbiotic association to the host abalone has been considered (Sawabe et al., 2006). Therefore the *V. halioticoli* is a biotechnologically and ecologically important marine bacterium but the complete genome sequence has not yet been achieved. In this chapter, to understand the alginate metabolic pathway and related metabolic network in this bacterium, I undertook complete genome sequencing of *V. halioticoli* IAM 14596^{T} using Next Generation Sequencer (NGS) including the 3rd generation PacBio (Miyamoto et al., 2014). This work enables us to provide a blue print of not only unique bacteriological and biochemical features but also seaweed carbohydrate metabolic pathways of *V. halioticoli* IAM 14596^{T}.

MATERIALS AND METHODS

Bacterial strains and culture condition

V. halioticoli IAM 14596^{T} was cultured in ZoBell 2216E agar medium containing 0.5% sodium alginate at 20 ℃ (Sawabe et al., 1995). For extraction of genomic DNA, the bacterium was cultured in ZoBell 2216E broth without alginate at 25 ℃ with gentle shaking at 130 rpm for 24 h.

DNA extraction and genome sequence

DNA of this strain was prepared as described by Marmur (1961), with minor modification. The bacterial cells were collected by centrifugation (10,000 rpm), and used for DNA extraction. The purified DNA showed >1.8 of 260/280 ratio was checked by spectrophotometer (Ultrospec 2000, Pharmacia Biotech, USA).

The draft genome sequence of the strain was determined using a 454 FLX Titanium genome sequencer with 37x shotgun and 153x pared end sequence coverage, and assembled using Newbler software. Finally, the sequence contigs were assembled into eight scaffolds (average scaffold size: 509,773 bp). Illumina reads were also obtained using HiSeq2000 platform (Illumina, CA, USA) based on the standard protocol provided by the manufacture. Some gaps were filled by manual Sanger sequencing. The 3rd generation NGS, Pacific Biosciences RSII sequencer (Pacific Biosciences, CA, USA) was also used to obtain the long-read sequences of the genome as detailed in the procedure previously reported (Kamada et al, 2014; Shimizu et al., 2017). Genome sequencings and assembles were performed by Dr. Ogura (Kyusyu Univ., Japan) and Professor Toyoda (National Institute of Genetics, Shizuoka, Japan).

Genome annotation

The CDS prediction was performed using MiGAP (http://www.migap.org/; Noguchi et al., 2008) and Rapid Annotation System Technology (RAST) server (Aziz et al., 2008). After correcting overlapping CDS (selecting lager CDS and removing no blast hit CDS on the opposite strand of longer CDS) obtained by two autoannotation methodologies, all available CDSs were searched

using BlastP (Altschul et al., 1990) against NR nucleotide database to obtain the top 30 sequences. According to the criteria of Blattener et al. (1997), functions were assigned to each CDS; 1) a specific physiological role was assigned if most of the hits were for a specific function, but if the substrates varied among the hits, the common denominator was assigned to the ORF, 2) if less specificity was found among the hits, a general function was assigned to an ORF, and 3) when functions of the hit sequences varied and there was no solid agreement even for type of function, or when only one sequence was hit, no function was assigned to the query ORF and it was counted among the unknowns (or hypothetical).

Genome analysis of *Vibrio halioticoli* IAM 14596^T with the metabolism overview

Genome analysis was first performed using RAST server. This server is based on manually curated subsystems and subsystem-based protein families that automatically guarantee a high degree of assignment consistency (Aziz et al., 2008). The CDSs were also annotated using the pathway mapping tool BlastKOALA in the Kyoto Encyclopaedia of Genes and Genomes (KEGG) database (http://www.genome.jp/kegg/). BlastKOALA is a recently developed automatic annotation server for genome sequences, which perform KO (KEGG Orthology) assignments to characterize individual gene functions and reconstruct KEGG pathways, BRITE hierarchies and KEGG modules to infer high-level functions of the organism (Kanehisa et al., 2016). The genus prokaryotes were used for the search as KEGG GENES database file. The program ExPASy translates (ExPASy Bioinformatics Resource Portal) and the implemented protein similarity search tools were used to confirm the annotation. Finally, all annotation results from the MiGAP, RAST, BLASTP, and KEGG results were unified and checked by eye.

Diagnostic phenotypic features

There are 19 phenotypic characterization features (production of amylase, chitinase, arginine dihydrolase, ornithine decarboxylase, and indole, utilization of sucrose, D-gluconate, cellobiose, acetate, L-glutamate, D-mannitol, D-glucosamine, fumarate, succinate, D-xylose, L-arabinose, citrate and DL-malate, and alginate) that are key phenotypes of *V. halioticoli* differentiating from related species as diagnostic features (Sawabe et al., 1998). Genes responsible for the diagnostic features were searched on the basis of KEGG. It clearly shows each biochemical reaction with substances, products and enzymes.

RESULTS AND DISCUSSION

Genome features

Using genome sequences produced by 454 FLX Titanium genome sequencer, eight scaffolds were produced. However, even my efforts of gap filling using manual Sanger sequencing, 24 kb plasmid-like sequence was only manually closed. On the basis of this knowledge, complete genome sequences of *V. halioticoli* IAM 14596^T, consisting of two chromosomes (Okada et al., 2005) and the plasmid-like structure, were successfully obtained using the PacBio sequencer. After error correction using the Illumina short reads, 2,785,698 bp larger (Chr. 1), 1,098,310 bp smaller (Chr. 2) chromosomes and a 244,363 bp plasmid-like sequences were obtained (Fig. 2.1 and Table 2.1). The sequences were also matched to two larger scaffolds assembled using 454 reads which are likely to correspond to Chr. 1 and Chr. 2. All Illumina reads were mapped evenly on the reconstructed genomes, which also supports the completeness of the genomes (Fig. 2.2). The

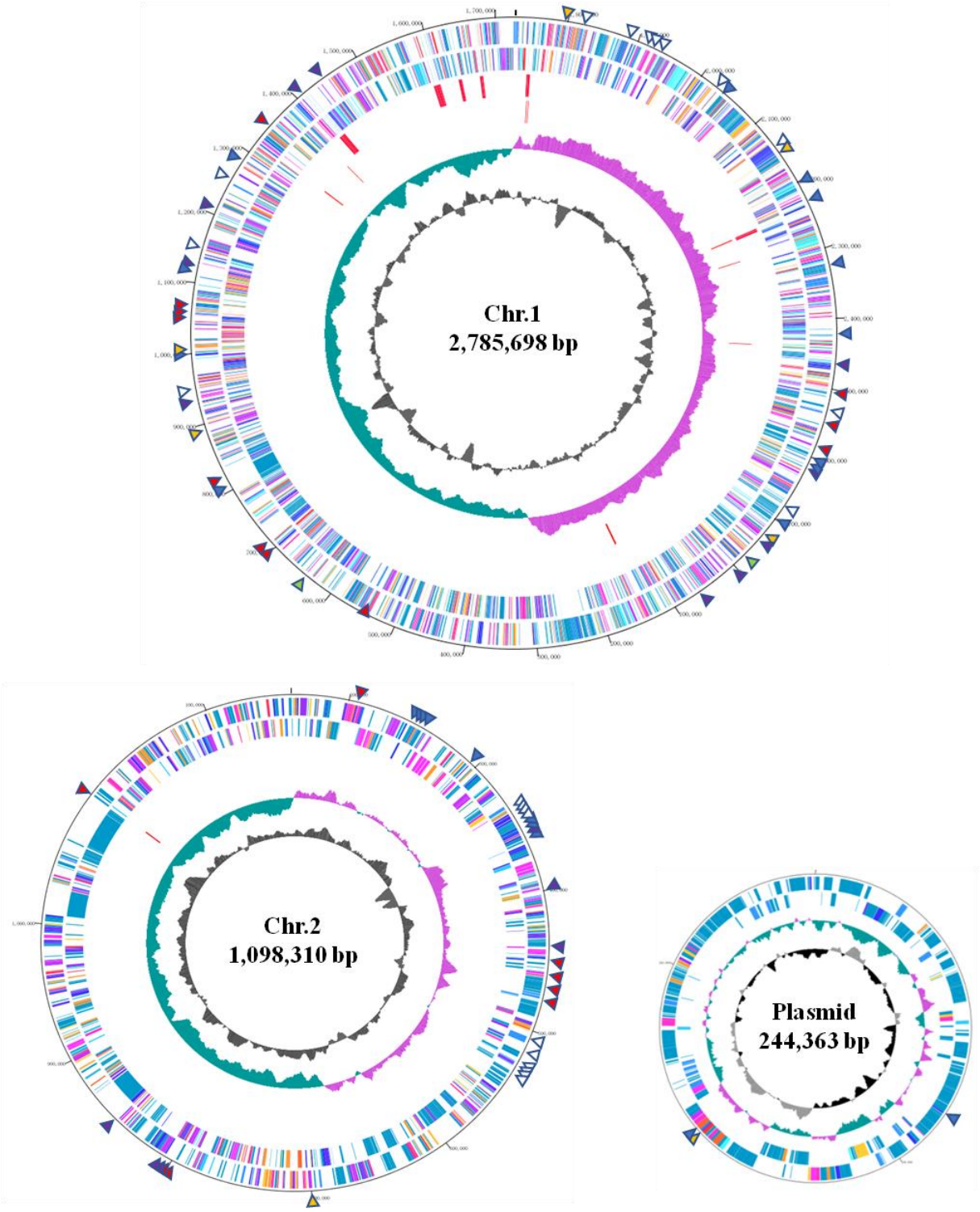

Fig. 2.1. Circular maps of *Vibrio halioticoli* IAM 14596^T genome. From the outside to inward: location of genes classified to CAZyme, forward strand CDS, reverse strand CDS, tRNA, rRNA, GC skew, and GC content. CDS feature was colored by functional category defined in the Seed database. The location of carbohydrate-active enzyme automated using web resource dbCAN.

▲: Glycoside Hydrolases, △: Glycosyl Transferases, ▲: Polysaccharide Lyases,
▲: Carbohydrate Esterases , ▲: Auxiliary Activities , ▲: Carbohydrate-Binding Modules

Table 2.1. General genome features of *Vibrio halioticoli* IAM 14596^T

Feature	Chr. 1	Chr. 2	Plasmid	Whole
Genome size (bp)	2785698	1098310	244363	4128371
GC content (%)	43.23	42.39	41.18	42.88
Number of genes (CDS)	2470	888	244	3602
Protein coding (bases)	2400918	939008	202740	3542666
Protein coding (%)	86.19	85.50	82.97	85.81
tRNAs	29	1	0	30
rRNAs	33	0	0	33

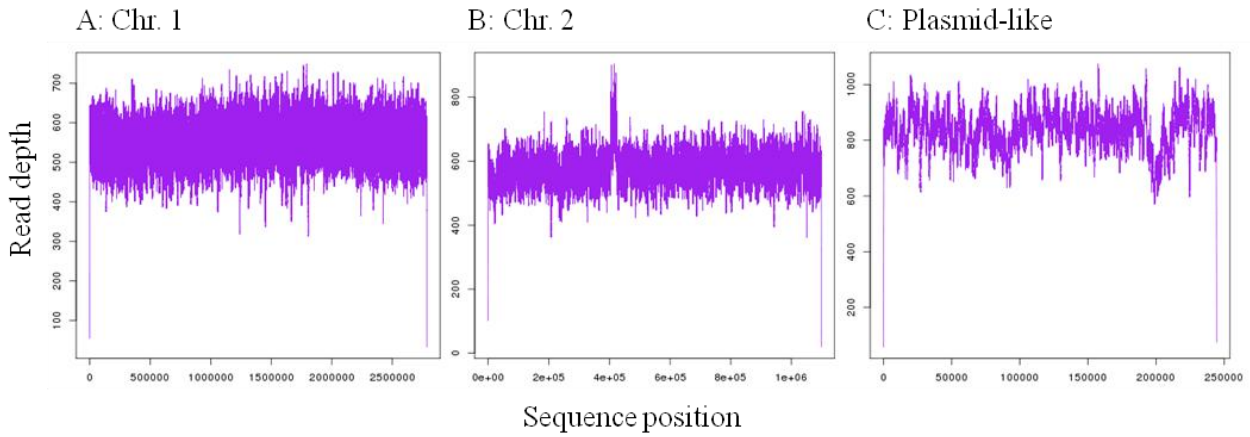

Fig. 2.2. Illumina reads mapping on the complete *Vibrio halioticoli* genomes.

average guanine-plus-cytosine (G+C) contents of each genome were 43.23%, 42.39%, and 41.18%, respectively (Table 2.1), corresponding to the experimentally determined GC moles% (43.1 mol%) using high-performance liquid chromatography (HPLC) (Sawabe et al., 1998). I succeeded for the first time in obtaining the complete genome sequence of *V. halioticoli*, and this is the second complete genome in Halioticoli clade species followed by *V. breoganii* FF50 (NZ_CP016177/NZ_CP016178/NZ_CP016179).

In total 3,602 CDSs were predicted on the genome, which covers 85.8% of the genome (Table 2.1). Unfortunately, approximately 25% of the CDSs were annotated as hypothetical proteins, even these were conserved in other bacterial species (Table 2.2). Details of the functional categories are described below. In total 33 rRNAs were found and all rRNAs were on the Chr. 1 (Table 2.1). A total of 8 rRNA operons were likely to be on the genome. In total 30 tRNAs were found, most of them were on the Chr. 1, and only one tRNA (TCC, Gly) was on the Chr. 2.

Estimation of replication origins

To identify the region of replication origin of Ch. 1, the genes and unique sequences required for replication of each genome were searched. The replication origin contains several common features consisting of chromosome partitioning protein genes, *parB* (sequence specific DNA binding protein) and *parA* (ATPase), rRNA small subunit 7-methylguanosine methyltransferase gene, *gidB*, tRNA uridine 5-carboxymethylaminomethyl modification enzyme gene, *gidA*, chromosomal replication initiator protein gene, *dnaA*, and DNA recombination and repair protein gene, *recF* (Fig. 2.3A) (Duigou et al., 2006; Val et al., 2014). Length of the *V. halioticoli dnaA* gene was 1,374 bp and

Table 2.2. Genome annotation of *Vibrio halioticoli* IAM 14596^T

	Chr. 1	Chr. 2	Plasmid	Whole
Number of genes (CDS)	2470	888	244	3602
Number of annotated CDS	1634	403	63	2100
Function category by KO system				
Metabolism	882	196	17	1095
Carbohydrate metabolism	144	38	6	188
Energy metabolism	94	33	0	127
Nucleotide metabolism	80	19	0	99
Lipid metabolism	55	9	0	64
Amino acid metabolism	160	22	5	187
Metabolism of other amino acids	31	8	2	41
Glycan biosynthesis and metabolism	58	10	0	68
Metabolism of cofactores and vitamins	132	26	2	160
Metabolism of terpenoids and polyketides	22	5	0	27
Biosynthesis of other secondary metabolites	24	8	1	33
Xenobiotics biodegradation and metabolism	22	9	0	31
Enzyme families	60	9	1	70
Genetic Information Processing	508	53	5	566
Environmental Information Processing	378	127	27	532
Cellular Processes	187	30	12	229
Organismal Systems	23	5	2	30
Human Diseases	73	18	1	92
Unclassified	278	96	15	389

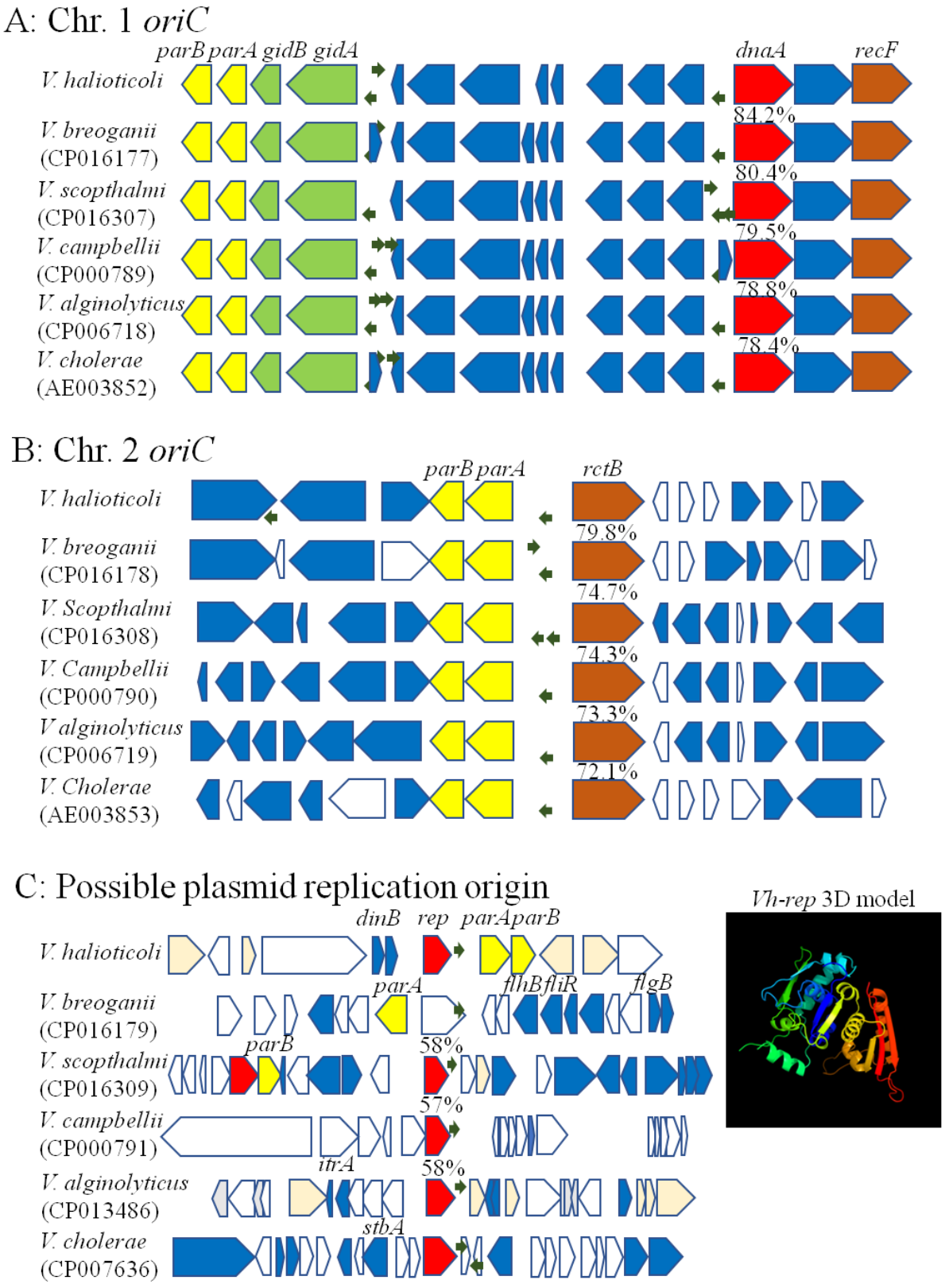

Fig. 2.3. Genome structure of estimated origin of replication region.

Replication initiation protein, Partitioning protein, Mobile element protein

General function Hypothetical Putative DNA box

84.2%, 80.4%, 79.5%, 78.8% and 78.48% nucleotide identity with *dnaA* genes of *V. breoganii* FF50 (NZ_CP016177/NZ_CP016178/NZ_CP016179), *V. scophthalmi* VS-12 (NZ_CP016307/ NZ_CP016308/NZ_CP016309), *V. campbellii* ATCC BAA-1116 (NZ009783_CP000789/ NZ009874_CP000790/NZ009777_CP000791), *V. alginolyticus* ATCC 17749 (NC_022349_CP006718/NC_022359_CP006719/NZ_CP013486), and *V. cholerae* O1 biovar El Tor str. N16961 (NC_002505_AE003852/NC_002506_AE003853/NZ_ CP007636), respectively. Typical DnaA box (TTATCCACA) sequences were found downstream and upstream of *gidA*, and downstream of *dnaA* gene on the *V. halioticoli* genome (Fig. 2.3A). Positions of the DnaA box were likely to be similar in vibrio but the numbers were likely to vary in different species.

To identify the region of replication origin of Ch. 2, the genes and unique sequences required for replication of each genome were searched. The replication origin contains several common features consisting of chromosome partitioning protein genes, *parB* and *parA*, and transcriptional regulator protein, *rctB* gene (Fig. 2.3B) (Duigou et al, 2006; Val et al., 2014). Length of the *V. halioticoli rctB* gene was 1,977 bp and 79.8%, 74.7%, 74.3%, 73.3% and 72.1% nucleotide identity with *rctB* genes of *V. breoganii, V. scophthalmi, V. campbellii, V. alginolyticus* and *V. cholerae*, respectively. Typical DnaA box (TTATCCACA) sequences were found downstream after *parA* on the *V. halioticoli* Ch. 2 (Fig. 2.3B). Positions of DnaA box, *parB, parA* and *rctB* genes were likely to be similar in vibrio but numbers of DnaA box were likely to vary in different species.

Presence of Plasmid-like structure

A plasmid-like sequence (244,363 bp) was found in *V. halioticoli* genome. It shows lower G+C

content (41.18 %) than other two chromosomes, and encodes 238 CDSs. A gene located on a CDS positioned from 101 to 1171 (356 AA, 41028.67 Da) possessed rep-3 Initiator Rep protein domain by Pfam search, and was grouped into Winged helix-like DNA-binding domain superfamily, which could function as a replication initiation (Fig. 2.3C). DNA box was also located downstream of the *rep* gene (Fig. 2.3C). Compared to plasmid structures which possess similar rep genes, no identical structure was found, which means the plasmid was likely to be unique in the strain and/or in the species. Tra gene set was found in the plasmid-like structure, which suggests the plasmid might be a conjugative plasmid. However, low similarity of the rep gene could provide better identification of the incompatibility group of the plasmid-like structure. In previous studies to construct GFP-expressed *V. halioticoli* cells, the R6K plasmid used was successfully transferred to the *V. halioticoli* IAM 14596^{T} cells via conjugation (Dunn et al., 2006; Sawabe et al., 2006). In the aspect of incompatibility of plasmids, the *V. halioticoli* plasmid might not be the IncX type. In fact, the *V. halioticoli* plasmid neither possesses typical core genes set of IncX plasmids consisted of *pir-bis-par-hns-topB-pilX-actX-taxCA* nor failed in silico amplification using *taxC* targeting IncX typing PCR (Johnson et al., 2012; Timothy et al., 2012). A possible Type I-E CRISPR/Cas system set consisted of *cas3-cse12345-cas1-cas2* with CRISPR loci consisted of two repeats (Koonin et al., 2017) was also located on the plasmid. No CRISPR/Cas system was found on the Chr. 1 and 2.

Reconstruction of the metabolic pathway

Among 2,470 CDSs on the *V. halioticoli* Ch. 1, 1634 CDSs were annotated in BlastKOALA approximate to 66.3% (Fig. 2.4 and Table 2.2). Most of the CDSs encode products involved in Metabolism (882, 54.0%), and followed by Genetic Information Processing involved with

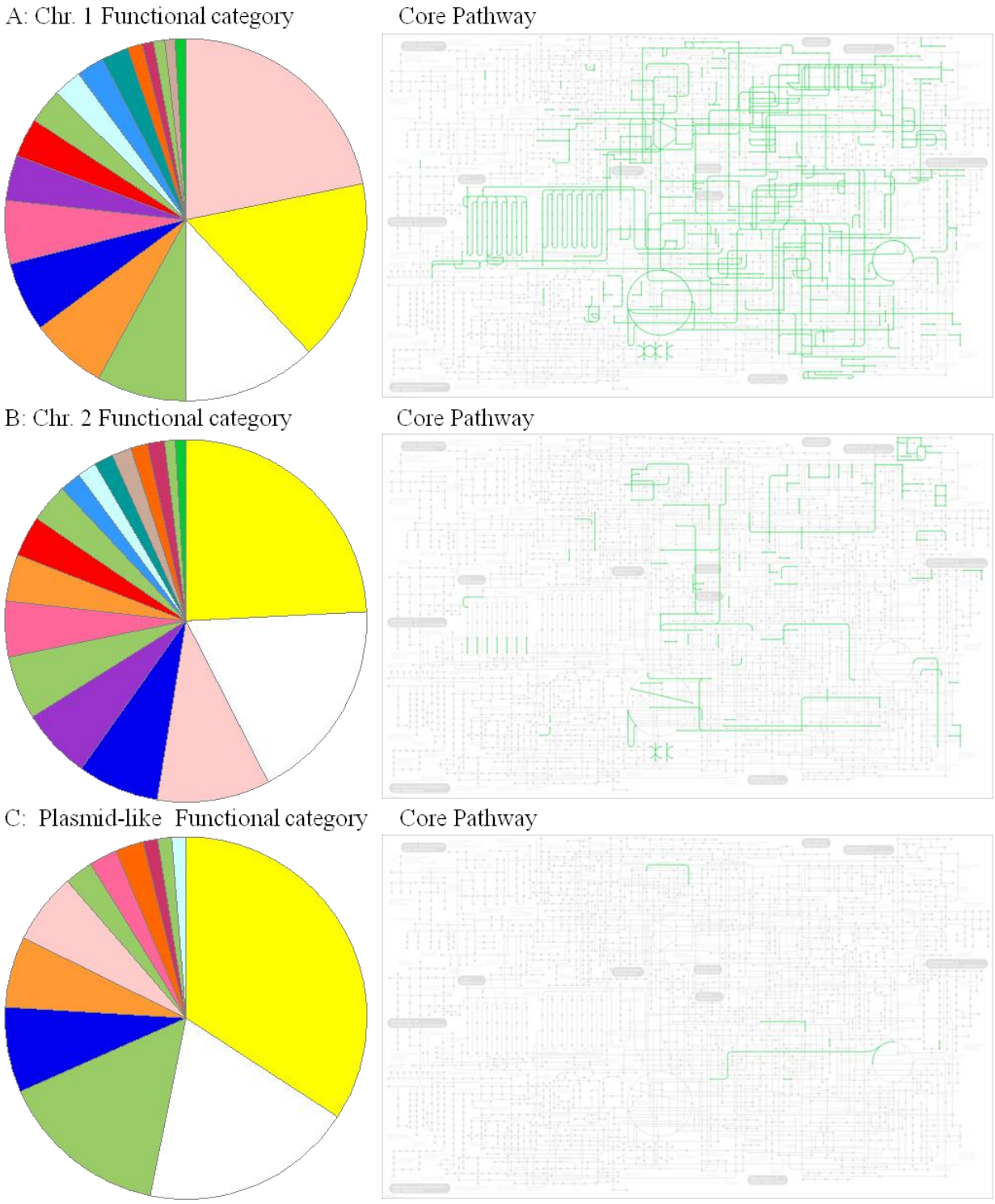

Fig. 2.4. Pathway partitioning of the *Vibrio halioticoli* genomes and plasmid.
Left: KEGG pathway mapping; Right: functional category according to the KO system.

- : Amino acid metabolism
- : Biosynthesis if other secondary metabolism
- : Carbohydrate metabolism
- : Cellular Processes
- : Energy metabolism
- : Environmental Information Processing
- : Enzyme families
- : Genetic Information Processing
- : Glycan biosynthesis and metabolism
- : Human Diseases
- : Lipid metabolism
- : Metabolism of cofactors and vitamins
- : Metabolism of terpenoids and polyketides
- : Metabolism of other amino acid
- : Nucleotide metabolism
- : Organismal System
- : Unclassified
- : Xenobiotics biodegradation and metabolism

Transcription, Translation, Replication and repair, and Folding, sorting, and degradation (508 CDSs, 31.1%), followed by Environmental Information Processing involved with Membrane transport and Signal transduction (378 CDSs, 23.1%), followed by Cellular Processes involved with Transport and catabolism, Cell growth and death and Cellular community (187 CDSs, 11.4%). Next is Human Diseases which are involved with Cancers, Immune diseases, Neurodegenerative diseases, cardiovascular diseases, Endocrine and metabolic diseases, Infectious diseases and Drug resistance (73 CDSs, 4.5%). The last one is Organismal Systems, which are involved with Immune system, Endocrine system, Circulatory system, Digestive system, Excretory system, Nervous system, Aging and Environmental adaptation (25 CDSs, 1.5%).

For *V. halioticoli* Ch. 2, among 888 CDSs on this chromosome, 403 CDSs were annotated in BlastKOALA corresponding to 45.4% Fig. 2.4 and Table 2.2). Most of the CDSs encode products involved in Metabolism (196, 48.6%), and followed by Environmental Information Processing (127 CDSs, 31.5%), followed by Genetic Information Processing (53 CDSs, 11.2%), followed by Cellular Processes (30 CDSs, 7.4%). Next is Human Diseases, (18 CDSs, 4.5%). The last one is Organismal Systems (5 CDSs, 1.2%). In contrast with Ch. 1, Ch. 2 showed more percentage of CDS encoding products involved in Environmental Information Processing and lower percentages of CDS encode products involved in Genetic Information Processing.

For *V. halioticoli* plasmid, among 244 CDSs on this chromosome, 63 CDSs were annotated in BlastKOALA corresponding to 25.8% (Fig. 2.4 and Table 2.2). Most of the CDSs encode products involved in Environmental Information Processing (27 CDSs, 42.9%), and followed by Metabolism

(17, 27.0%), followed by Genetic Information Processing (5 CDSs, 7.9%). Next is Organismal Systems (2 CDSs, 3.2%). The last one is Human Diseases (1 CDSs, 1.6%). In contrast with Ch. 1 and Ch. 2, this plasmid showed highest percentage of CDS encoding products involved in Environmental Information Processing.

Overall metabolic pathways of *V. halioticoli* were reconstructed using KEGG and BLAST search tools (Figs. 2.4 and 2.5). *V. halioticoli* possessed genes responsible for three central metabolic pathways in carbohydrate metabolism; 1) Embden-Meyerhof (EM) pathway, 2) oxidative pentose phosphate (PP) pathway, and 3) Entner-Doudoroff (ED) pathway. Tricarboxylic acid (TCA) cycle, glyoxylate cycle, and mixed acid fermentation pathways were also reconstructed. CAZyme annotation using dbCAN revealed *V. halioticoli* possessed CDS classified into 32 glycoside hydrolase (GH), 26 glycosyltransferase (GT), 20 polysaccharide lyase (PL), 7 carbohydrate esterase (CE) and 2 auxiliary activities (AA) in total (Fig. 2.1, and Table 2.3). Among them, CDSs classified into 9 GH and 6 PL also possessed carbohydrate-binding module families (CBMs), and total CBMs number of 27 were found. The variety of transporters, secretion systems, and protein export systems were predicted. 117 types of ABC transporter genes being capable of transport 30 substrates were found. There are 7 mineral and organic ion transporters (4-Amino-5-hydroxymethyl-2-methylpyrimidine (HMP), glycine betaine, iron (III), molybdate, proline, N-formyl-4-amino-5-aminomethyl-2-methylpyrimidine (FAMP), tungstate), 3 oligosaccharide, polyol, and lipid transporters (maltose, maltodextrin, phospholipid), 4 monosaccharide transporters (autoinducer, D-xylose, ribose and *sn*-glycerol-3-phosphate), 8 phosphate and amino acid transporters (amino acid, arginine, D-methionine, general L-amino acid,

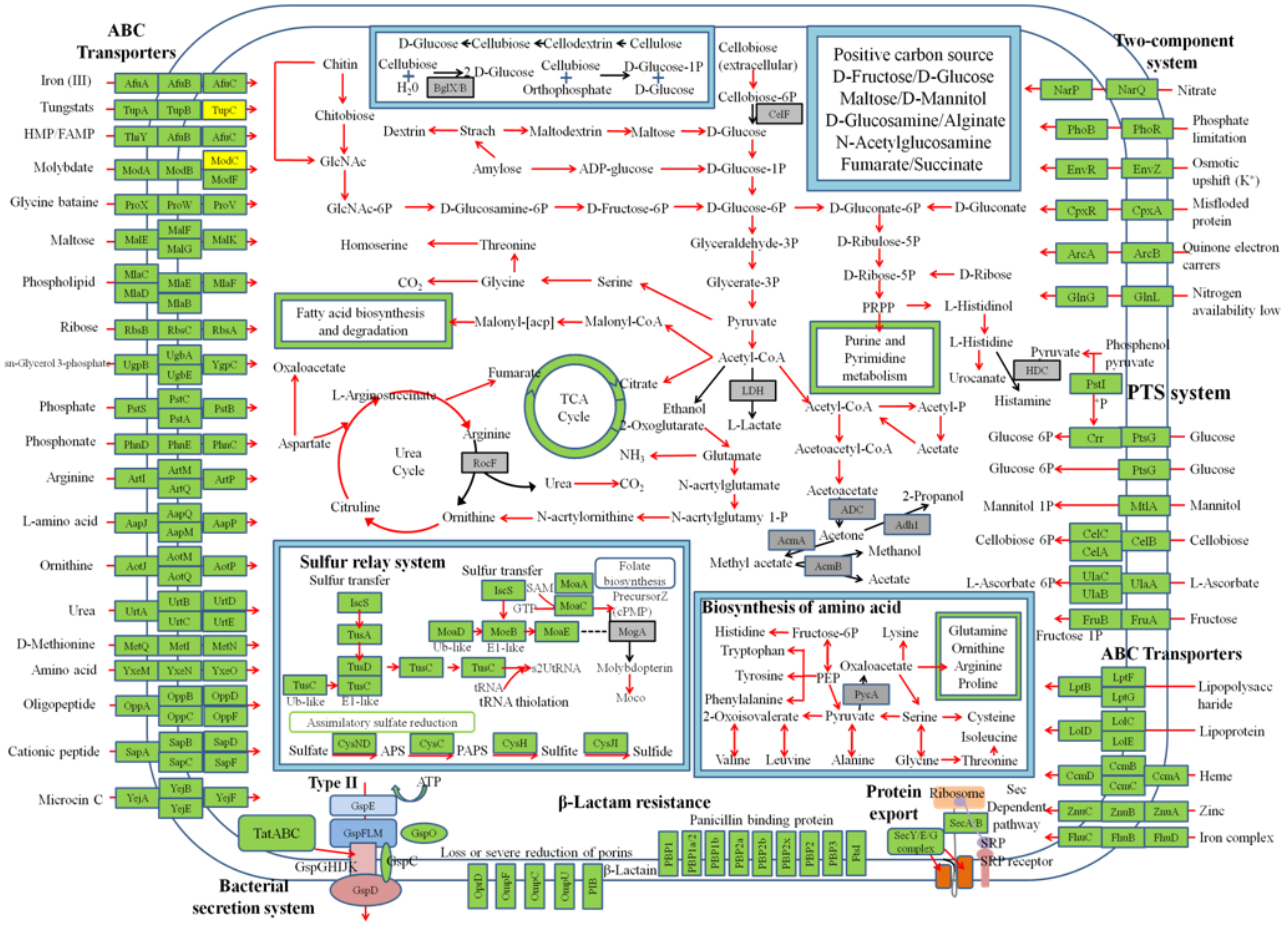

Fig. 2.5. Overview of metabolic pathway of *Vibrio halioticoli* IAM 14596^{T} genome.

Table. 2.3. Carbohydrate-Active enZYmes data of *Vibrio halioticoli* IAM 14596^T

CAZyme	CDS number located on			
	Ch. 1	Ch. 2	Plasmid	Whole
Glycoside Hydrolase	20 (7)	10 (2)	2	32 (9)
GH3	1	0	0	1
GH9	1	0	0	1
GH13	3 (3)	4 (2)	0	7 (5)
GH18	2 (1)	0	0	2 (1)
GH19	1 (1)	0	0	1 (1)
GH20	2 (1)	0	0	2 (1)
GH23	5 (1)	0	1	6 (1)
GH73	1	0	0	1
GH77	0	1	1	2
GH88	0	1	0	1
GH92	0	4	0	4
GH102	1	0	0	1
GH103	1	0	0	1
GH109	2	0	0	2
Glycosyl Transferase	12	14	0	26
GT1	1	0	0	1
GT2	2	4	0	6
GT4	1	6	0	7
GT5	1	0	0	1
GT9	1	0	0	1
GT14	0	1	0	1
GT19	1	0	0	1
GT26	1	0	0	1
GT28	1	0	0	1
GT35	0	1	0	1
GT51	2	1	0	3
GT70	0	1	0	1
GT81	1	0	0	1
Polysaccharide Lyase	11 (4)	9 (2)	0	20 (6)
PL6	3 (2)	2 (1)	0	5 (3)
PL7	5 (2)	4 (1)	0	9 (3)
PL15	1	1	0	2
PL17	1	2	0	3
PL22	1	0	0	1
Carbohydrate Esterase	5	1	1	7
CE1	2	0	0	2
CE3	0	1	0	1
CE9	1	0	0	1
CE10	1	0	1	2
CE11	1	0	0	1
Auxiliary Activity	2	0	0	2
AA6	2	0	0	2
Carbohydrate-Binding Module	20	7	0	27
CBM2	1	0	0	1
CBM5	2	0	0	2
CBM16	5	1	0	6
CBM20	0	1	0	1
CBM32	2	1	0	3
CBM34	1	0	0	1
CBM41	2	0	0	2
CBM48	2	2	0	4
CBM50	3	2	0	5
CBM67	1	0	0	1
CBM73	1	0	0	1

Parentheses, number of enzyme possess CBM.

ornithine, phosphate, phosphonate and urea), 3 peptide and nickel transporters (cationic peptide, microcin C and oliopeptide), 2 metallic cation, iron-siderophore and vitamin B12 transporters (iron complex and zine), and 3 ABC-2 and other transporters (heme, lipopolysaccharide and lipoprotein). 23 types of PTS genes supported utilization of 5 types of substrates (glucose, mannitol, cellobiose, L-ascorbate and fructose). Six two-component systems (nitrate, phosphate limitation, osmotic upshift K^+, misfolded protein, quinone electron carriers and nitrogen availability low) were found to respond to environmental stimuli (Fig. 2.5)

V. halioticoli also possesses genes responsible for mannitol utilization: mannitol specific phosphoenolpyruvate (PEP)-dependent phosphotransferase system (PTS), which has a role in uptake and phosphorylation of mannitol, and a nicotinamide adenine dinucleotide (NADH)-specific mannitol l-phosphate dehydrogenase, which catalyzes oxidation of mannitol l-phosphate to fructose 6-phosphate. The fructose 6-phosphate is further metabolized through the EM pathway. Mannitol operon (*mtlADR* operon) coding mannitol-specific PTS proteins is made up of three genes, annotated as mannitol-specific IIABC component, PTS system (MtlA), mannitol-1-phosphate 5-dehydrogenase (MtlD) and mannitol repressor protein (MtlR).

Identification of a gene responsible for 4-deoxy-L-erythro-hex- 4-enepyranosyluronate (DEH) reductase in *Sphingomonas* sp. strain A-1 revealed that bacteria use the ED pathway in alginate metabolism (Preiss and Ashwell, 1962a; 1962b; Takase et al., 2010). Alginate polymer is degraded in extracellular by end-type alginate lyase, the resultant alginate oligomers are transported into intercellular, and then converted into unsaturated monosaccharide by saccharification process using

exo-type oligoalginate lyase, generating DEH. The conversion of DEH and 2-keto-3-deoxy-D-gluconic acid (KDG) is catalyzed by a nicotinamide adenine dinucleotide phosphate (NADPH)-dependent short-chain dehydrogenase (*sdr*). KDG is phosphorylated by ATP-dependent 2-dehydro-3-deoxygluconokinase (*kdgK*), so the generated 2-dehydro-3-deoxy-D-gluconate 6-phosphate (KDGP) finally can enter the ED pathway in *V. halioticoli*. All genes responsible to the pathway were found (Fig. 2.6)

A total of 15 candidate genes encoding alginate lyases were found on *V. halioticoli* (Fig. 2.7). Among them, three genes (C1gene1623_S2G4, C1gene875_S2G5 and C2gene58_S4G10) were unlikely to be not only located in tandem but also located closely on the related gene responsible for alginate utilization. S2G4 was localized on Ch. 1 locus from 1805696-1806688. Before this area, I also found some genes respond tight adherence proteins (*tad*, associated with Flp pilus assembly) and Flp pilus assembly proteins (Flp). The gene set of *tadG-tadF-tadE-tadD-tadC-tadB-cpaF-tadZ* and *cpaA-pilA-pilA* were found from this locus. S2G5 was localized on Ch. 1 locus from 968803 to 970368. Around this area, genes of *tesB*, NTPase, HPA2, *lysR* and beta-lactamase were present. The genes of C1gene1920_S2G1, C1gene1919_S2G2 and C1gene1914_S2G3 were localized in a gene cluster on Ch. 1 locus from 2150679 to 2160048. Around this area, two genes of PTS-Mtl-EII component proteins and *araD, ulaD, sguB, kduD, uxaC, tctB* and *arsB* gene were found. In Ch. 2, three alginate lyase related genes (C2gene578_S4G1, C2gene580_S4G2 and C2gene581_S4G3) constituted a cluster locus from 672876 to 678503, I also found the HPA2, MAGT, *pyrC*, *glpK* and *glgC* genes present around this area. Interestingly, four genes (C2gene303_S4G4, C2gene304_S4G5, C2gene312_S4G8 and C2gene319_S4G9) were localized within a large cluster

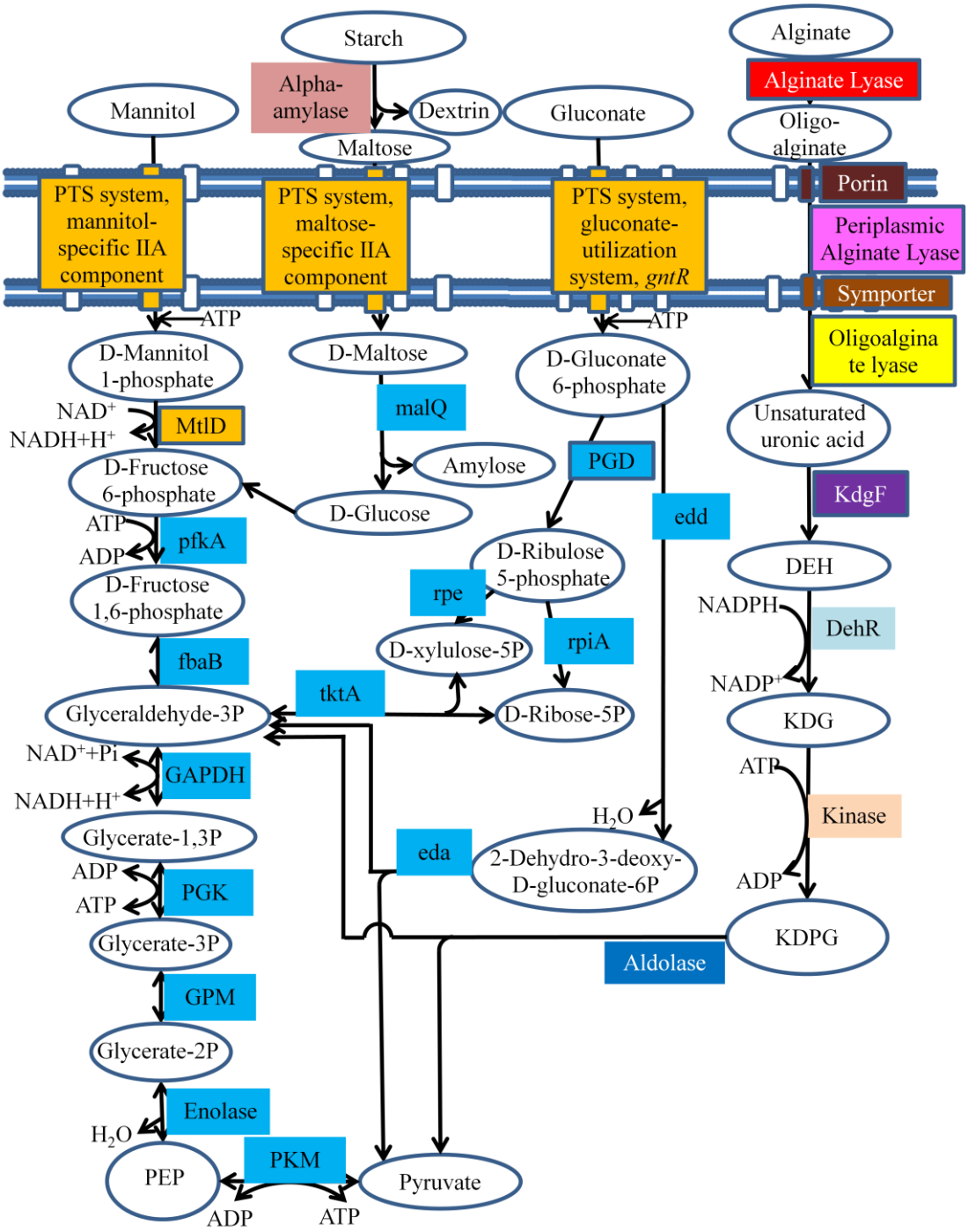

Fig. 2.6. General description of the major seaweed component metabolic pathways in genome of *Vibrio halioticoli*.

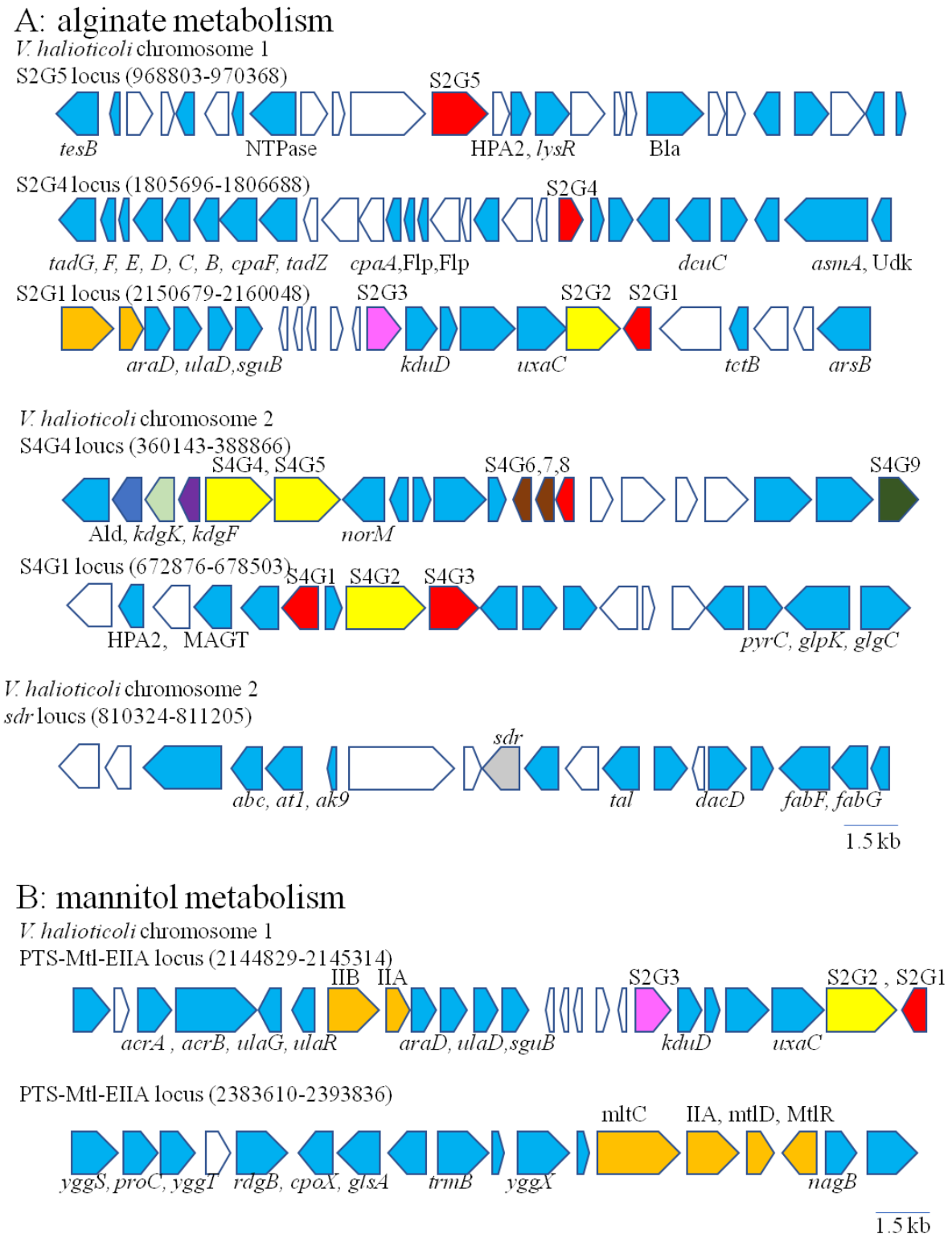

Fig. 2.7. Gene location on the genome of *V. halioticoli* IAM 14596T strain. A, gene related in alginate metabolism; B, gene related in mannitol metabolism.

: Alginate lyase, : Oligoalginate lyase, : Periplasmic alginate lyase, : Porin, : Symporter, : Sugar kinase, : Aldolase, : PTS transporter, : KdgF, : General function, : Hypothetical

comprising additional alginate-assimilating genes including two porin (S4G6 and S4G7), aldolase, *kdgF* and *kdgK*. S4G4 and S4G5 may function as oligoalginate lyase. S4G8 may be an alginate lyase, and S4G9 may work as symporter (Fig. 2.7). In addition, a *sdr*-like gene was found on Ch. 1 locus from 810324 to 811205. Around this area, there were also presented the genes of ABC transporter, aminotransferase, archaeal kinase, *tal*, *dacD*, *fabF* and *fabG*. However, I totally found 9 genes which were identified as functional alginate lyases out of 15 gene candidates. Details of gene location of alginate lyases and the related metabolism are described below and in the next chapter.

Genotype to Phenotype

KEGG pathway prediction was also used to diagnose phenotypic features. *V. halioticoli* shows negative reaction of amylase and chitinase, arginine dihydrolase, ornithine decarboxylase, indole production, utilization of sucrose, D-gluconate, cellobiose, D-glucronate, γ-aminobutyrate, acetate, pyruvate, propinonate, L-glutamate, D-mannitol, D-glucosamine, fumarate, succinate, meso-erythritol, D-xylose, L-arabinose, citrate, DL-malate and δ-aminovarate (Sawabe et al., 1998). Most of the negative reactions were supported by the absence of genes responsible for the diagnostic features except amylase, chitinase, and indole production, utilization of sucrose, D-gluconate, acetate, D-mannitol, fumarate, succinate, D-xylose, L-arabinose, citrate and DL-malate (Table 2.4). Nevertheless, I found two α-amylase genes and the assimilation gene sets (C2gene119, C2gene204) on Chr. 1, but the bacterium showed negative starch degradation (Sawabe et al., 1998). Recently, implementation of amylase gene from *V. harveryi* S20 could enhance the degradation and utilization of starch by the bacterium (data not shown). These suggest the inactivation or less expression of these amylase genes of *V. harveryi* S20. I also found the presence

Table 2.4. Comparison between genotype and phenotype of *Vibrio halioticoli* IAM 14596^T

Characteristic	Pheno-type	Geno-type	Pathway	Associated regulatory genes	Locus tag	Enzyme/Protein responsible
production of:						
Amylase	-	+	Starch and sucrose metabolism	Hydrolases, alpha-amylase	C2Gene119	Alpha-amylase
					C2Gene204	Alpha-amylase
Chitinase	-	+	Amino sugar and nucleotide sugar metabolism	Hydrolases, chitinase	C1gene529	Chitinase
Arginine decarboxylase	-	-	Arginine and proline metabolism	Arginine regulatory pathway, *argR*	Absent	L-arginine:oxygen 2-oxidoreductase
Ornithine decarboxylase	-	-	Arginine and proline metabolism	Arginine regulatory pathway, *argR*	Absent	L-ornithine carboxy-lyase
Indole	-	+	Aromatic amino acid degradation	Tryptophanase, *tnaA*	C1gene1063	Tryptophan synthase
					C1gene1064	Tryptophan synthase
Utilization of						
Sucrose	-	+	Starch and sucrose metabolism	Sucrose operon repressor, scrR	C1gene705	alpha-glucosidase
D-Gluconate			Pentose phosphate pathway	Transferases gluconokinase, *gntK*	C1gene715	Gluconokinase
Cellobiose	-	-	b-Glucoside metabolism	Cellobiose and glucan utilization regulator, *celR*	Absent	beta-D-glucoside glucohydrolase
Acetate	-	+	Methane metabolism	Transferases acetate kinase, *ackA*	C1gene2281	Acetyl-CoA synthetase
Pyruvate			Pyruvate metabolism	Ligases, pyruvate carboxylase		
L-Glutamate	-	-	Alanine, aspartate and glutamate metabolism	Lyases, glutamate decarboxylase	Absent	L-glutamate:NAD^+ oxidoreductase
D-Mannitol	+	+	Fructose and mannose metabolism	Mannitol operon repressor, *mtlR*	C1gene2119	PTS-Mtl-EIIA
Fumarate	-	+	Two-component system	*dcuS-dcuR* two-component regulatory system	C1gene1310	Fumarate hydratase
Succinate	-	+	Carbon fixation pathways	Succinyl-CoA synthetase alpha subunit	C1gene1724	Succinyl-CoA synthetase, alpha subunit
					C1gene1725	Succinyl-CoA synthetase, beta subunit
D-Xylose	-	+	ABC transporters	D-xylose transport system ATP-binding protein, *xylG*	C2gene51	ABC-type sugar transport system, ATPase component
L-Arabinose	-	+	ABC transporters	Arabinose transport system ATP-binding protein		
Citrate	-	+	Citrate cycle (TCA cycle)	ATP citrate (pro-S)-lyase	C1gene1732	Citrate synthase
DL-Malate	-	+	Glyoxylate and dicarboxylate metabolism	Malate dehydrogenase, MDH1	C1gene274	Malate dehydrogenase
Alginate	-	+	Fructose and mannose metabolism	Lyases, poly (M/G) lyase	C2gene59	Oligo alginate lyase (S4G10)
					S2gene320	Alginate lyase (S4G9)

of chitinase genes and the related pathway. The status of chitin utilization might be similar to that of starch utilization in the bacterium. This may be valuable in testing the implementation of active chitinase genes into the bacterium. For indole production, the bacterium possessed complete pathway. As non-flagellation is one of the unique features of this bacterium, no genes responsible for flagellar assembly were found on the genome (West et al., 1977).

Genome comparison

V. halioticoli Ch. 1 was similar to that of *V. breoganii* ZF-55 (comparison score: 546), and followed by *V. cyclitrophicus* 1F273 (comparison score: 496), *V. antiquarius* Ex25 (comparison score: 438), *V. tasmaniensis* 1F-267 (comparison score: 402), and *V. campbellii* DS40M4 (comparison score: 394). Ch. 2 was similar to the genome of *V. harveyi* ATCC BAA-1116 (comparison score: 523), *V. coralliilyticus* ATCC BAA-450 (comparison score: 445), *V. shilonii* AK1 (comparison score: 431), *V. splendidus* 12B01 (comparison score: 397) and *V. breoganii* ZF-55 (comparison score: 395). The plasmid was closely similar to *Photobacterium profundum* SS9 with the comparison score 508.

CONCLUSION

The genome sequence of *V. halioticoli* IAM 14596^{T} was completely finished using the sequencer combination of Sanger, 454 FLX, Illumina, and PacBio. Two chromosomes and one plasmid were reconstructed. This genome sequence provides a starting point for the study of this organism's phenotype and genotype characteristics. It will be interesting to determine the gene expression

patterns that are unique to its function as well as classic phenotypic characterization. Additionally, the genomic sequence of *V. halioticoli* should clearly show the metabolic pathway of this prokaryotic organism, and provide important clues to understanding the metabolic and regulatory networks that link genes on the two chromosomes. Finally, *V. halioticoli* clearly possesses each gene they worked in utilization of mannitol, alginate, and the other components of marine biomass. Cloning and expression of these genes may help us to verify the estimated pathways, and more clearly understand the alginate metabolism in the bacterium.

Chapter 3. Genome wide mining of genes responsible for alginate metabolic pathway of *Vibrio halioticoli* IAM 14596^{T}

ABSTRACT

A total of 15 alginate lyase gene candidates were found from genome analysis, and 11 of them were more likely to be alginate lyase related genes. Among these, three genes, S2G1, S2G4 and S4G8, were similar to previously reported alginate lyase genes of *Vibrio halioticoli* IAM 14569^{T}, designated as *alyVGI* and *alyVGII*. The other genes were newly found alginate lyase gene candidates of this bacterium. The family and domain information search revealed seven genes, S2G1, S2G3, S2G4, S4G1, S4G2, S4G3, and S4G8, which contained the "Alginate lyase2" domain and were classified as PL7. Only the S4G3 gene product contained one of the other domains classified as "F5_F8_type_C" domain. The products of three genes, S2G2, S4G4, and S4G5, contained a domain classified as the "Alginate lyase" and were grouped into PL17. S4G4 and S4G5 were also estimated to be dual domain proteins containing the "Hepar_II_III" domain. Another more probable alginate lyase gene, S4G10, possessed only the domain "Hepar_II_III", and was classified as PL15. Two genes, S4G6 and S4G7, previously reported as an alginate lyase *alyVGIII*, were further identified as an oligosaccharide specific porin gene, *kdgM*. S2G5 and S4G9 were estimated to be pectin lyase not alginate lyase. The 15 alginate lyase gene candidates were successfully cloned into Directional TOPO vector using Lumio technology. Alginate lyase activities were discovered from recombinant proteins from 9 alginate lyase gene candidates on the basis of plate assay and TBA method; 7 clones, S2G1/pET160, S2G3/pET161, S2G5/pET160, S4G1/pET160, S4G2/pET160, S4G3/pET161, and S4G8/pET160/pET161, showed activity in plate

assay, and 2 clones, S2G2/pET161 and S4G10/pET161, showed activity from the cell free extract. The optimum expression was also determined using LB broth with and/or without seawater. Substrate specificity was tentatively identified using plate assay; S2G1/pET160, S4G1/pET160, S4G3/pET161, and S4G8/pET160/pET161 showed polyG specific activities.

INTRODUCTION

Alginate is a linear copolymer widely present in brown algae (Gacesa, 1988). Alginate assimilation process is multi-step and complex, in which endo- and exo- type alginate lyases, ABC and/or energy independent transporters, and many enzymes in glycolysis are involved (Preiss and Ashwell, 1962a, b; Takase et al., 2010; Takeda et al., 2011; Wargacki et al., 2012). Alginate lyases are classified as a polyM-lyase and polyG-lyase (EC 4.2.2.3 and EC 4.2.2.11) according to their cleaving reaction with polyM-rich or polyG-rich block structure of alginates. Additionally, an alginate lyase showing broad substrate specificity has also been described (Sawabe et al., 1997). All alginate lyases reported are no-ATP dependent and irreversible. Alginate lyases function via the endolytic and exolytic β-elimination reaction. Currently known alginate lyases are grouped into PL5, PL6, PL7, PL14 (exolytic oligo-alginate specific), PL15 (oligo-specific), PL17, and PL18 (Garron and Cygler, 2010; Ochial et al., 2010; Park et al., 2014). We also know PL15, PL17, and PL18 are multidomain alginate lyases (Garron and Cygler, 2010; Ochiai et al., 2010; Park et al., 2014; Zhu and Yin, 2015).

The major traits of *V. halioticoli* IAM14596^{T} are non-motile, fermentative, and assimilate alginate.

Bigger growth is observed in a medium containing alginate, showing the cell mass attaching gelling clusters of alginate; this gives an idea that the bacterium may possess efficient alginate assimilating machinery (Sawabe et al., 2006). Formate-acetate fermentation is also observed in the alginate containing medium (Sawabe et al., 2003). To elucidate the alginate assimilating machinery, biochemical and molecular biological studies have been conducted (Sugimura et al., 2000). Due to difficulties in the purification of the enzymes, however, only limited biochemical and genetic information is available; 1) at least 6 alginate lyases were detected in this bacterium using isoelectric focusing, and 2) three polyG specific lyase genes were cloned and sequenced (Sugimura et al., 2000).

Since Preiss and Ashwell (1962a; 1962b) have detected alginate lyase activity in both the accumulation of 4-deoxy-L-erythro-5-hexoseulose uronic acid (DEH) and TPNH-linked dehydrogenase activity in converting DEH to 2-keto-3-deoxy-D-gluconate (KDG) in alginate-induced *Pseudomonas* sp., an unexpectedly large diversity in alginate assimilation machinery and/or biochemical pathways in bacteria have been clarified. The ABC transporter specific to intact alginate molecules, and the further metabolic DEH-KDG pathway, which was proposed by Preiss and Ashwell (1962a; 1962b), are found in a *Sphingomonas* bacterium capable of assimilating intact alginate (Hashimoto et al., 2004; Takase et al., 2010). Functional shotgun screening of *Vibrio splendidus* 12B01 genomic fosmid library in *E. coli* reveals that the presence of oligoalginate specific porin and symporter, which contribute to create metabolically engineered *E. coli* to show alginate assimilation coupled with ethanol production (Wargacki et al., 2012).

In Chapter 2, I succeeded in mapping a complete whole genome of the alginate loving marine bacteria for the first time. Such gap-closed circular genomes provide us with a better blueprint in how alginate assimilation machinery is established in this bacterium. In this chapter, I performed genome wide mining of genes responsible for alginate degradation machinery of the *V. halioticoli* strain using the protein family/domain database (Swiss-Prot/TrEMBL) search and protein localization prediction tool, aiming to increase the probability of the alginate lyase gene candidates and the expressions of these genes to prove if they are active enzymes or not.

MATERIALS AND METHODS

In silico genome wide gene mining of alginate lyase, and structural and protein localization analysis

Alginate lyase gene candidates were retrieved from the complete genome sequence using In Silico Molecular Cloning (ISMC) software (In Silico Biology, Yokohama) with a feature search command. Protein structural information was obtained using the blast searching tool (Altschul *et al*., 1990) in the UniProt Knowledgebase (Swiss-Prot+TrEMBL) implemented in the ExPASy Bioinformatics Resource Portal (http://expasy.org/). Pfam, SuperFamily, and Gene3D information were mainly used for evaluating whether the gene function is categorized into alginate lyase or not. CAZyme classification (Cantarel et al, 2009; http://www.cazy.org/) was conducted with the dbCAN HMMs using HMMER3 (http://csbl.bmb.uga.edu/dbCAN/) (Yin et al., 2012). The prediction of protein localization of these candidate genes was performed using the PSORTB subcellular localization prediction tool with the option of Gram-negative bacteria (http://www.psort.org/psortb/) (Yu et al.,

2010).

Bacterial strains, plasmids, and Media

V. halioticoli IAM 14596^{T} was cultured in ZoBell 2216E agar and/or broth medium containing 0.5% sodium alginate at 20℃. *E. coli* strains TOP10 and BL21 (DE3) were used for transformation and recombinant protein expression, respectively. The recombinant *E. coli* strains were grown on LB agar and/or broth medium containing 100 μg/ml of ampicillin at 37℃. Directional TOPO vectors with Lumio technology (Invitrogen, Carlsbad, USA) were used for cloning the candidate genes and sequencing; one of these vectors, pET160, is designed for N-terminal His- and Lumio-tag, and the other, pET161, is for the C-terminal tags.

For screening positive clones expressing alginate lyase, an alginate lyase activity screening (ALAS) agar plate was used (Sugimura et al., 2000). The composition of the medium was 20g LB-broth (BD, Franklin Lakes, USA), 15 g agar, 5 g alginates, 100 μg/ml ampicillin, 100 mM IPTG, and 1 L of 50% ASW (v/v) in final. Each transformant was spot-inoculated on the ALAS, and incubated at 30℃ for 5 days. After the cultivation, clear zone formation around colonies due to alginate degradation was observed after 5 days culture at 30℃ by flooding the plate with 99.5% ethanol. Shorter time cultivation was also conducted to detect a clone showing strong alginate lyase activity.

Gene manipulations

Genomic DNA of *V. halioticoli* IAM 14596^{T} was isolated using Promega Wizard^{TM} genomic DNA extraction Kit according to the manufacturer's instructions. DNA concentration was determined

using a spectrophotometer (Ultrospec2000, Pharmacia Biotech, Cambridge, UK). The gene primer was designed according to the alginate lyase gene sequencing selected in Table 3.1 using In Silico Molecular Cloning software (ISB) (Table 3.2). These oligonucleotide primers were purchased from Sigma-Aldrich Japan (Ishikari, Japan).

Gene amplification was performed using a thermal cycling PCR system (GeneAmp PCR System 9700, ABI, Foster City, USA) with *pfu* DNA polymerase (Promega, Madison, USA). The *pfu* DNA polymerase reaction mixture consisted of 5 µl *pfu* DNA polymerase 10 × buffers with $MgSO_4$, 1 µl dNTP mix (10mM each), 5-50 pmol forward and reverse primer, 0.5 µg/50 µl DNA template, 1.25 u/50 µl *pfu* DNA polymerase (2-3 u/µl), adjusted to a final volume of 50 µl using nuclease-free water. The reaction cycling condition was as follows; 96℃ for 3 min. for an initial denaturation, 30 cycles of major PCR step by 96℃, 1 min. 50℃, 1 min. 72℃, 1 to 2.5 min. The extension time was set according to 1 kb/min. in each gene of interest. Each amplification product was analyzed by agarose gel electrophoresis using a 1.5% agarose gel and TBE buffer system (Sambrook and Russell, 2001). Ethidium bromide was added to the gel. After the electrophoresis, stained DNA was observed using the Light Capture system (ATTO, Tokyo, Japan) as a fluorescent band. Lamda DNA *Hin*dIII and *Eco*RI digestion maker (Promega) were used as a molecular ruler.

Gene cloning and protein expression were performed using Champion™ pET Directional TOPO Expression Kits with Lumio™ Technology (Invitrogen) according to the manufacturer's instructions. The target genes were cloned into pET vectors and transferred into competent TOP10 *E. coli* cells by heat shock. After the transformation step, recombinant TOP10 *E. coli* cells were

Table 3.1. Family, domain and location information of alginate lyase gene candidates of *Vibrio halioticoli* IAM 14596^T

Gene	Estimated size			The closest gene in Swiss-Prot/ TrEMBLDB				Family and domain information				Localization
	NA	AA	Da	Annotation	Amino acid identity (%)	Score	E-value	Pfam	Super Family	Gene3D	CAZyme	PSORTB predication
S2G1	1059	352	38669	Q9RGQ3_9VIBR Alginate lyase AlyVGI	100%	1868	0.0	Alginate lyase2 (58-351)	Glucanases superfamily (58-352)	Glucanases subgrp (50-351)	PL7	EC, 9.72
S2G2	2001	666	76257	I8PZ55_YERPE Alginate lyase family protein	51%	1734	0.0	Alginate lyase (99-271)	Alginate lyase (8-330)	Alginate lyase (98-341)	PL17	CP, 8.96
S2G3	1044	347	38192	A3UR44_VIBSP Putative alginate lyase	79%	1487	0.0	Alginate lyase2 (34-342)	Glucanases superfamily (34-354)	Glucanases subgrp (28-242)	PL7	PP, 5.91
S2G4	993	330	36464	Q9RGQ3_9VIBR Alginate lyase AlyVGI	57%	912	1.0×10^{-119}	Alginate lyase2 (58-351)	Glucanases superfamily (58-352)	Glucanases subgrp (50-351)	PL7	EC, 9.72
S2G5	1566	521	56404	A3XWI1_9VIBR Putative alginate lyase	39%	819	1.0×10^{-100}	NA	Pectin lyas like (67-537)	Pectin lyas fold (73-521)	PL6	CM=PP=OM=EC, 2.5
S4G1	972	323	35438	G3LI08_9BACT Alginate lyase	99%	1703	0.0	Alginate lyase2 (49-323)	Glucanases superfamily (49-323)	Glucanases subgrp (41-323)	PL7	EC, 9.72
S4G2	1881	626	69916	G4WFC4_9VIBR Alginate lyase	59%	1528	0.0	Alginate lyase2 (5-185, 195-481)	Glucanases superfamily (7-187, 195-483)	Glucanases subgrp (5-184, 190-482)	PL7	CP, 8.96
S4G3	1566	521	57027	C4TJD4_9VIBR Alginate lyase	86%	2396	0.0	Alginate lyase2 (252-516), F5_F8_type_C (45-162)	Glucanases superfamily (252-517), Gal_bind_like(33-168)	Glucanases subgrp (246-516)	CBM32, PL7	EC, 9.72
S4G4	2172	723	81156	G4WFC4_9VIBR Alginate lyase	73%	2903	0.0	Alginate Lyase (47-272), Hepar_II_III (367-506)	Alginate lyase (45-295)	Alginate lyase (55-318)	PL17	CP=CM=PP=OM=EC, 2.0
S4G5	2139	712	80015	G4WFC5_9VIBR Alginate lyase	83%	3271	0.0	Alginate lyase (75-263), Hepar_II_III (368-525)	Alginate lyase (38-305)	Alginate lyase (61-292)	PL17	CP=CM=PP=OM=EC, 2.0
S4G6	705	234	27023	Q9RGQ2_9VIBR Alginate lyase AlyVGIII	75%	986	1.0×10^{-134}	KdgM (16-235)	Oligogalacturonate specific porin	NA	NA	OM, 10.00
S4G7	666	221	25682	Q9RGQ2_9VIBR Alginate lyase AlyVGIII	100%	1199	1.0×10^{-167}	KdgM (16-235)	Oligogalacturonate specific porin	NA	NA	OM, 10.00
S4G8	1017	338	36727	Q9RGQ5_9VIBR Alginate lyase AlyVGII	97%	1696	0.0	Alginate lyase2 (65-322)	Glucanases superfamily (65-331)	Glucanases subgrp (63-330)	PL7	EC, 9.65
S4G9	1569	522	57886	A3XWI1_9VIBR Putative alginate lyase	70%	1992	0.0	NA	Pectin lyas like (67-537)	Pectin lyas fold (73-521)	PL6	CM, 4.90
S4G10	2079	692	79009	A6CUW5_9VIBR Oligo alginate lyase	87%	3366	0.0	Hepar_II_III (415-570)	Alginate lyase (72-418)	NA	PL15	CP, 8.96

Table 3.2. Amplicon size estimation of alginate lyase gene candidates

Lane	Gene	Primer set	Theoretical amplicon size (bp)	Experimental (bp)
1	S2G1	T7	1444	1401
2		Gene	1056	1025
3	S2G2	T7	2386	2421
4		Gene	1998	2007
5	S2G3	T7	1429	1455
6		Gene	1041	1110
7	S2G4	T7	1378	1398
8		Gene	990	1005
9	S2G5	T7	1951	2021
10		Gene	1563	1566
1	S4G1	T7	1357	1371
2		Gene	969	1000
3	S4G2	T7	2266	2298
4		Gene	1878	1902
5	S4G3	T7	1951	1991
6		Gene	1563	1552
7	S4G4	T7	2557	2573
8		Gene	2169	2091
9	S4G5	T7	2524	2512
10		Gene	2136	2133
11	S4G6	T7	1090	1152
12		Gene	702	745
13	S4G7	T7	1051	1088
14		Gene	663	700
15	S4G8	T7	1402	1425
16		Gene	1014	1021
17	S4G9	T7	1954	2001
18		Gene	1566	1603
19	S4G10	T7	2464	2466
20		Gene	2076	2088

cultured in SOC broth, shaken horizontally (200rpm) at 37℃ for 1 hour, then spread on pre-warmed LB agar plate and incubated at 37℃ overnight. The cloning reaction produced over 100 colonies, and 10 to 20 colonies were selected for analysis. Then the positive transformant was purified twice, and stored at -80℃ as glycerol stock.

Each recombinant plasmid was extracted using the Wizard SV Gel and PCR Clean-Up System (Promega, Madison, USA). In brief, each clone was cultured in LB broth containing 100 μg/mL ampicillin with vigorous shaking at 130 rpm (FMC-100, Eyela, Tokyo), and after overnight culture, cells were collected, and the plasmid was extracted using the abovementioned kit. Concentration of the plasmid was measured, and the plasmid was stored at -30℃ before use. Insert fragment was confirmed by PCR with the T7 promoter primer set and gene primer set according to the manufacturer's instructions. The forward primer included the 4 base pair sequences (CACC) necessary for directional cloning on the 5' end. Vector pET160 has the N terminus Lumio tag. Reverse primer was set after gene stop codon. Vector pET161 has the C terminus Lumio tag. The Reverse primer was set not only before the stop codon but also to contain the entire frame. Sequencing was conducted using the BigDye terminator sequencing kit version 3.1 and ABI PRISM 3130XL Genetic Analyzer (ABI, Waltham, Massachusetts, USA).

The purified plasmid was transformed into *E. coli* BL21 (DE3) cells by heat shock procedure according to the manufacturer's instructions. The recombinant BL21 (DE3) *E. coli* cells were first cultured in SOC broth, shaken horizontally (200rpm) at 37℃ for 1 hour, and spread on a pre-warmed LB agar plate. After incubation at 37℃ overnight, 5 colonies were selected for insert

verification, and the positive transformant was also purified twice, and stored at -80℃ as glycerol stock. These transformants were used for the remainder of the tests.

Assay for alginate lyase activity (TBA method)

Alginate lyase activity was assayed using the thiobarbituric acid (TBA) method. All the recombinant BL21 *E. coli* cells were cultured in 50 ml liquid medium with LB broth or LB broth containing 50% (v/v) ASW, 100 mM IPTG and 100 μg/ml ampicillin at 30℃. Incubation time was set as 12, 24, 36, 48, and 60h, respectively. After incubation, cell concentration was determined by a spectrophotometer at OD600 (Ultrospec2000, Pharmacia Biotech, Cambridge, UK). A total of 30ml culture was centrifuged (9,605×g, 15min, 4℃). Supernatant was collected and stored at -30℃. This fraction was named as supernatant fraction (SF). Cells were washed by centrifugation (9,605×g, 5min, 4℃) with 30 ml of 0.1M Tris-HCl buffer (pH 7.5). Then, 100 μl chloroform was added to each cell pellet, and then vortexed for 30 seconds. A total of 750 μl of 0.1 M Tris-HCl buffer was added to the sample, and then centrifuged (19,283×g, 4℃, 5 min) to prepare cell free extracts (CFE).

The reaction mixture for measuring the enzyme activity was composed of 0.1 M Tris-HCl buffer (pH 7.5) and 1% sodium alginate in 2.7 ml volume. A total of 10% (v/v) crude enzyme solution (SF or CFE) corresponding to 0.3 ml was added to the reaction mixture. A subsample of 0.25 ml reaction mixture was taken at each time point (0, 15, 30, 45 min), and then the production of 4-keto-5-deoxy uronic acid was measured by TBA method in Abs549 (Preiss and Ashwell, 1962a). Alginate lyase activity unit (U) was defined as the activity producing 1 μmole β-formylpyruvate

equivalent of 4-deoxy-5-keto-uronic acid per 1 minute in the defined reaction conditions using the following formula; A549/0.290*0.01(μmole)/0.3(enzyme vol. ml)/time.

RESULTS

Family and domain information of alginate lyase gene candidates

In addition to the BLAST search information, the protein family, domain and localization information were further retrieved using the InterProcan search tool in the Swiss-Prot and TrEMBLE protein database. A further in silico search revealed that 11 out of 15 candidates were highly likely to be alginate lyases (Table 3.1). On the chromosome 1 gene candidates, genes encoding S2G1 (1,059 bp, 532 AA, 38.7 kDa), S2G2 (2,001 bp, 666 AA, 76.3 kDa), S2G3 (1,044bp, 347 AA, 38.2 kDa), and S2G4 (993 bp, 330 AA, 36.5 kDa) showed above 50% amino acid identity against known alginate lyase, and contained an alginate lyase domain. The gene S2G5 (1,566 bp, 521 AA, 56.4 kDa) shared 39% amino acid identity with a vibrio putative alginate lyase but a pectin lyase like domain and was grouped into PL6. On the chromosome 2 genes, genes encoding S4G1 (972 bp, 323 AA, 35.4 kDa), S4G2 (1,881 bp, 626 AA, 69.9 kDa), S4G3 (1,566 bp, 521 AA, 57.0 kDa), S4G4 (2,172 bp, 723 AA, 81.2 kDa), S4G5 (2,139 bp, 712 AA, 80.0 kDa), S4G8 (1,017 bp, 338 AA, 36.7 kDa), S4G9 (1,569 bp, 522 AA, 57.9 kDa) and S4G10 (2,079 bp, 692 AA, 79.0 kDa) showed similarity of above 59% amino acid identity, and contained alginate lyase domains. The genes, S4G6 (705 bp, 234 AA, 27.0 kDa) and S4G7 (666 bp, 221 AA, 25.7 kDa), were previously identified as polyG specific alginate lyases (*alyVGIII*), but in this analysis these genes shared 75% and 100% AA identity against oligogalacturonate specific porin (Table 3.1). The gene S4G9 shared

70% AA identity with a vibrio putative alginate lyase but pectin lyase like domain and grouped into PL6.

In previous gene cloning experiments, three alginate lyase genes were successfully cloned from *V. halioticoli* IAM 14596^{T} and sequenced (Sugimura et al., 2000). These sequences have been deposited in GenBank under accession numbers AF114039 (*alyVGI*), AF114037 (*alyVGII*), and AF114040 (*alyVGIII*). These nucleotide sequences were almost identical to the flanking and gene coding region of S2G1, S4G8, and S4G7, respectively. In this search result, similar genes were also found. Gene S2G1 and S2G4 were similar to *alyVGI* sharing amino acid identities of 100% and 57%, respectively. Gene S4G8 shared 97% amino acid identity against *alyVGII*. As previously mentioned, gene S4G6 and S4G7 showed high amino acid identity to *alyVGIII*, but the family/domain database search revealed these genes might be oligosaccharide specific porin, not alginate lyase.

In addition, gene S2G2 shared 51% amino acid identity with I8PZ55_YERPE alginate lyase family protein from *Yersinia pestis* PY-96. Gene S2G3 shared 79% identity with A3UR44 _VIBSP putative alginate lyase from *V. splendidus* 12B01. Gene S2G5 shared 39% identity with A3XWI1 _9VIBR putative alginate lyase from *Vibrio* sp. MED222. Gene S4G1 shared 99% identity with G3LI08 _9BACT alginate lyase (uncultured bacterium). Genes S4G2 and S4G4 shared 59% and 73% identity with G4WFC4 _9VIBR alginate lyase from *Vibrio 'midae'*, respectively. S4G5 shared 83% identity against G4WFC5_9VIBR Alginate lyase of *Vibrio 'midae'*. Gene S4G3 shared 86% identity with C4TJD4 _9VIBR alginate lyase from *Vibrio* sp. A9m. Gene S4G9 shared 70% identity

with A3XWI1_9VIBR putative alginate lyase from *Vibrio* sp. MED222, and gene S4G10 shared 87% identity with A6CUW5_9VIBR oligo alginate lyase from *V. splendidus* 12B01. Most of the gene candidates showed sequence similarity with vibrio genes.

Protein family and domain information was finally obtained from four kinds of database; Pfam, Superfamily, Gene3D, and CAZyme. Pfam is a widely used database of protein families and domains. A total of 16,712 families are included in the latest Pfam 31.0 version. The seed alignments are used to build profile hidden Markov models (HMMs) that can be used to search any sequence database for homologues in a sensitive and accurate fashion (Finn et al., 2016). Superfamily is a database of structural and functional annotation for all proteins and genomes. The annotation is based on a collection of hidden Markov models, which represent structural protein domains at the SCOP superfamily level. A superfamily groups together domains which have an evolutionary relationship. The annotation is produced by scanning protein sequences from over 2,478 completely sequenced genomes against the hidden Markov models (http://supfam.org/SUPERFAMILY/index.html). Gene3D is a data resource based on the CATH classification of protein structures on the basis of four major levels in this hierarchy: Class (secondary structure composition), Architecture (overall shape), Topology (fold family), and Homologous superfamily (http://www.cathdb.info/). Carbohydrate-Active enZymes (CAZymes) build and breakdown complex carbohydrates and glycoconjugates for a large body of biological roles. The Carbohydrate-Active Enzyme (CAZy) database is a knowledge-based resource specializing in these enzymes (Cantrarel et al., 2009); currently, the database contains up to date knowledge on 148 GH, 105 GT, 27 PL, 16 CE, 13 AA and 83 CBM

(http://www.cazy.org/Home.html at 20th Dec. 2017). Known PLs require an acidic group at subsite +1 sugar. Based on the nature of the +1 sugar, the substrates of PLs can be divided into three groups: (1) galacturonic acid, (2) glucuronic/iduronic acid, and (3) mannuronate/guluronate. Polysaccharides in the first group and includes pectate and pectin, the second group includes hyaluronan, chondroitin, chondroitinsulfate (CS), dermatan sulfate (DS), heparin, HS, heparosan, glucuronan, and xanthan, and the third group contains alginates. Currently known alginate lyases are further grouped into PL5 ($(\alpha/\alpha)_4$incomplete toroid, EC 4.2.2.3), PL6 (β-helix, EC4.2.2.3), PL7 (β-jelly roll, EC4.2.2.3 & 4.2.2.11), PL14 (EC4.2.2.3 & EC4.2.2.- (exo-oligo), β-jelly roll), PL15 (EC4.2.2.- (oligo-specific), $(\alpha/\alpha)_{5,6}$ incomplete troid+antiparallel β-sandwich), PL17 (EC4.2.2.3, α/α barrel+β-sandwich), and PL18 (EC4.2.2.3, β-jelly roll) (Garron and Cygler, 2010; Park et al., 2014).

S2G1, S2G3, S2G4, S4G1, S4G2, S4G3, and S4G8 gene products contain the domain classified as "Alginate lyase 2" according to the Pfam database (Table 3.1). The domain is frequently found in poly(beta-D-mannuronate) lyases. All gene products were classified in the polysaccharide lyase family 7 (PL7). PL7 contains both poly(beta-mannuronate) lyase and poly(alpha-glucronate) lyase showing -jelly roll structure, and 1042 bacterial and 8 eukaryotes lyases are found in the database (http://www.cazy.org/PL7.html). Interestingly, S4G2 includes dual Alginate lyase 2 domains; there are four cases possessing such domain architecture. The S4G3 gene product also includes one of the other domains classified as "F5_F8_type_C" domain. The F5/F8 type C domain is known to have a role in binding anionic phospholipid forming an amphipathic alpha-helix (http://pfam.sanger.ac.uk/family/PF00754). There are 8 sequences with F5_F8_type_C & Alginate lyase 2 domains (http://pfam.sanger.ac.uk/family/PF08787#tabview=tab1); *Stigmatella aurantiaca*

(Q094P5 STIAD), *Saccahrophagus degradans* (Q21HU1 SACD2), *Vibrio splendidus* 12B01 (A3UR41 VIBSP), *Vibrio* sp. A9m (C4TJD4_9VIBR), *Pseudoalteromonas* sp. CY24 (D0QMJ7_9GAMM), *Vibrio alginolyticus* 40B (D0X275_VIBAL), *Kribbella flavida* (D2PLH3_KRIFD), and *Geobacillus* sp. strain Y412MC10 (D3EHA3_GEOS4). Various vibrios possess this type of alginate lyase.

S2G2, S4G4, and S4G5 gene products include the domain classified as "Alginate lyase". These gene products were grouped into PL17. S4G4 and S4G5 were also dual domain proteins containing "Hepar_II_III" domain. This family features sequences that are similar to a region of the *Flavobacterium heparinum* proteins heparinase II and heparinase III; the former is known to degrade heparin and heparan sulphate, whereas the latter predominantly degrades heparan sulphate (Su et al., 1996). There are 114 sequences containing Alginate lyase & Hepar_II_III domain architecture in the current Pfam database. I could find the type of genes in bacteria from both marine and terrestrial origins. S4G10 included only a domain "Hepar_II_III". From the Gene3D information, gene S2G1, S2G3, S2G4, S4G1, S4G2, S4G3 and S4G8 were classified as Glucanases subgroup. Genes, S2G2, S4G4 and S4G5, were classified as Alginate lyase. S2G5 and S4G9 were classified as Pectin lyase fold. At here, alginate lyase and alginate lyase2 family pointed to the poly (beta-D-mannuronate) lyase, "F5_F8_type_C" domain known as discoidin domain which is identified in blood coagulation factors, and "Hepar_II_III" domain was the degradation enzyme for heparin and heparan sulphate.

Localization prediction of the alginate lyases in *Vibrio halioticoli* IAM 14569^{T}

As mentioned in the Introduction, the *V. halioticoli* might possess efficient alginate assimilating machinery (Sawabe et al., 2006). So, it is important to know the cellular localization and secretion of alginate lyases. I conducted a PsortB protein localization prediction. Five gene products were exported extracellularly, and three were expressed in cytoplasm. Three genes could not be identified by the localization (Table 3.1, Fig. 2.7).

11 genes were more likely to function as alginate lyase (Table 3.1). At least six types of alginate lyases were present in the genome of *V. halioticoli*; 1) extracellular secreting (or periplasmic) 35-40 kDa PL7 types (S2G1, S2G3, S2G4, S4G1 and S4G8), 2) a cytoplasmic 80 kDa PL17 type (S2G2), 3) a cytoplasmic 70 kDa dual domain PL7 (S4G2), 4) an extracellular 57 kDa dual domain alginate lyase (S4G3), 5) localization unknown 81 kDa dual domain PL17 type (S4G4 and S4G5), and 6) cytoplasmic 79 kDa PL15 type alginate lyases (S4G10) (Fig. 2.6 and 2.7). More interestingly, some genes were likely to be arranged in an operonic structure (Fig. 2.7). In particular, genes responsible for alginate lyase showing different protein family and localization were likely to be arranged in tandem; ex. S2G2 cytoplasmic PL17 and S2G3 extracellular PL7, S4G2 cytoplasmic PL7 and S4G3 extracellualdual domain enzyme. In a different case, S4G4 and S4G5 sharing 67% nucleotide sequence identity were likely to be duplicated. Furthermore, S4G8 extracellular PL7 gene was arranged on the same orientation as S4G6 and S4G7 oligosaccharide porin genes.

The decomposition and metabolism of alginate is a very complicated process. First, the macromolecular alginate cannot be transported into the cell. This leads to the alginate degradation pathway being divided into two parts, the intracellular reaction and extracellular reaction.

Undoubtedly, there require two types of alginate lyase. One type was transferred into the outside of the cell, the macromolecular alginate was broken down into oligoalginate. Next, the oligoalginate was transported into the cell and reacted with the intracellular alginate lyase. Therefore, the localization of these two types of alginate lyase was very important in the gene expression analysis, especially the extracellular alginate lyase. To date, many kinds of extracellular alginate lyases have been reported. This protein always consists of two parts, leader peptide responsible for transport and mature protein responsible for catalytic reaction. All this information is recorded in the database curated in the PSORT bioinformatics tool. Gene analysis according by PSORT was necessary, but it still needs to be validated by activity test.

In chapter two, I showed the genome of *V. halioticoli* IAM 14596^{T} consists of two chromosomes. Genome size are 2.786 Mb for the large chromosome and 1.098 Mb for the small chromosome. In contrast with gene function, the number of the alginate lyase genes in large chromosome was doubled in the small chromosome. In total 4 and 7 alginate lyase genes were found in chromosome 1 and 2, respectively. Therefore, it is interesting to investigate further gene expression, protein translational control and interaction of these enzymes.

Bioinformatical in silico analysis to estimate gene function is always necessary with experimental proof. The real alginate lyase gene must be obtained according to experimental analysis, e.g. protein expression test is substantive work. This suggests the idea that the bacterium possesses efficient alginate assimilating machinery (Sawabe et al., 2006).

Activity screening

A total of 15 candidate genes previously describe in Table 3.1 were successfully cloned into the expression vector pET160 and/or pET161 (Table 3.3). Inserted genes were checked by colony PCR in the first step (Fig. 3.1), and subsequently by sequencing (Table 3.3). In colony PCR screening, fragment length of amplicon produced by T7 promoter primer set was 346 bp and 371 bp larger than those of the original sequence in pET160 and pET161, respectively (Table 3.3). All amplicon size was similar to that estimated by sequences (Fig. 3.1). 11 more probable alginate lyase gene candidates were cloned; 8 genes, S2G1, S2G4, S2G5, S4G1, S4G2, S4G4, S4G5, and S4G8, were cloned into pET160, which means these gene products were N-terminal tag by 6xHis and Lumio tag, and 6 genes, S2G2, S2G3, S4G3, S4G8, S4G9, and S4G10, were cloned into pET161, which also means these gene products were C-terminal 6xHis and Lumio tags (Table 3.3). It is confirmed that both 5' and 3' ends of the insert were identical to the original *V. halioticoli* genome sequence, and especially in S2G1, S2G3, S2G4, S4G1, S4G2, S4G3, and S4G8 genes, full length sequence was confirmed to be identical (Table 3.3). S2G5 and S4G9 genes, putative pectin lyase gene, were also cloned into pET160 and pET161, respectively, and both ends of the inserted sequences were identical. The S4G6 and S4G7 genes, putative KdgM gene encoding oligosaccharide porin, were cloned into pET161 and pET160, respectively, and the sequences were confirmed to be identical to original *V. halioticoli* sequences (Table 3.3).

The alginate lyase activity of each clone was tested by plate assay (Fig. 3.2). Apparent activity was observed in S2G1/pET160, S2G3/pET161, S2G5/pET160, S4G1/pET160, S4G2/pET160, S4G3/pET161, S4G8/pET160, and S4G8/pET161 (Fig. 3.3), of which gene product might be

Table 3.3. Summary of cloning and expression experiments

Gene	Cloning/Activity[a]		Sequence identity (%)		
	pET160	pET161	5'termini	3' termini	full
S2G1	+/EC	NT	100	100	100
S2G2	NT	+/CP	100	100	NT
S2G3	NT	+/EC	100	100	100
S2G4	+/ND	NT	100	100	100
S2G5	+/EC	NT	100	100	NT
S4G1	+/EC	NT	100	100	100
S4G2	+/EC	NT	100	100	100
S4G3	NT	+/EC	100	100	100
S4G4	+/ND	NT	100	100	NT
S4G5	+/ND	NT	100	100	NT
S4G6	NT	+/ND	100	100	100
S4G7	+/ND	NT	100	100	100
S4G8	+/EC	+/EC	100[b]	100[b]	100[b]
S4G9	NT	+/ND	100	100	NT
S4G10	NT	+/CP	100	100	NT

[a] EC: extracelluar, CP: cytoplasmic, ND: not detected, NT: not tested.

[b] Both cloned genes showed identical sequences.

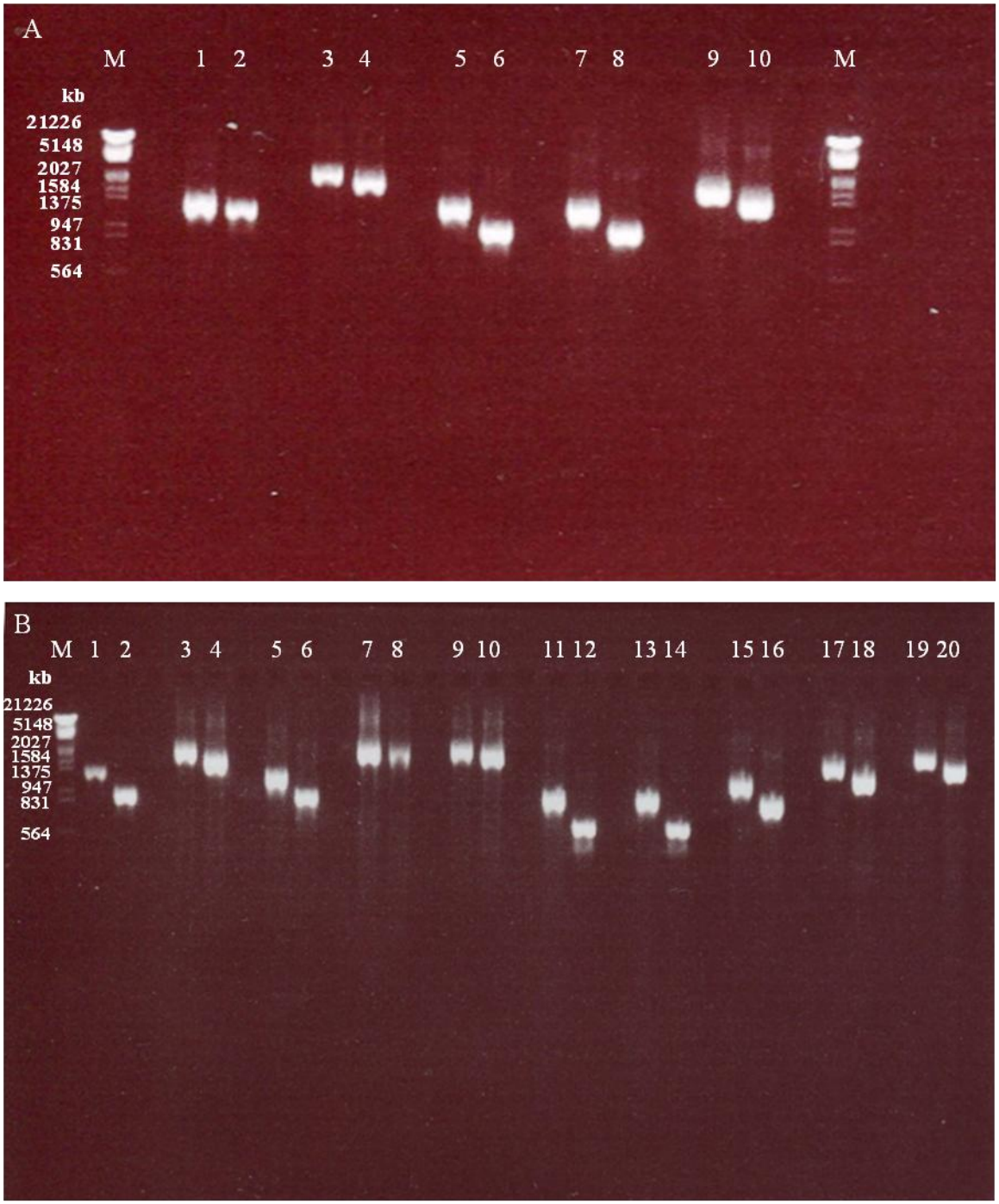

Fig. 3.1. Agarose electrophoresis of PCR products. A: scaffold 2 genes; Lane; M: marker (lDNA *Hin*dIII+*Eco*RI digest), 1-2: S2G1 amplicon with T7 primer set and gene primer set; 3-4: S2G2 amplicon with T7 primer set and gene primer set; 5-6: S2G3 amplicon with T7 primer set and gene primer set; 7-8: S2G4 amplicon with T7 primer set and gene primer set; 9-10: S2G5 amplicon with T7 primer set and gene primer set; B: scaffold 4 genes; Lane; M: marker (lDNA *Hin*dIII+*Eco*RI digest), 1-2: S4G1 amplicon with T7 primer set and gene primer set; 3-4: S4G2 amplicon with T7 primer set and gene primer set; 5-6: S4G3 amplicon with T7 primer set and gene primer set; 7-8: S4G4 amplicon with T7 primer set and gene primer set; 9-10: S4G5 amplicon with T7 primer set and gene primer set, 11-12: S4G6 amplicon with T7 primer set and gene primer set; 13-14: S4G7 amplicon with T7 primer set and gene primer set; 15-16: S4G8 amplicon with T7 primer set and gene primer set; 17-18: S4G9 amplicon with T7 primer set and gene primer set; 19-20: S4G10 amplicon with T7 primer set and gene primer set.

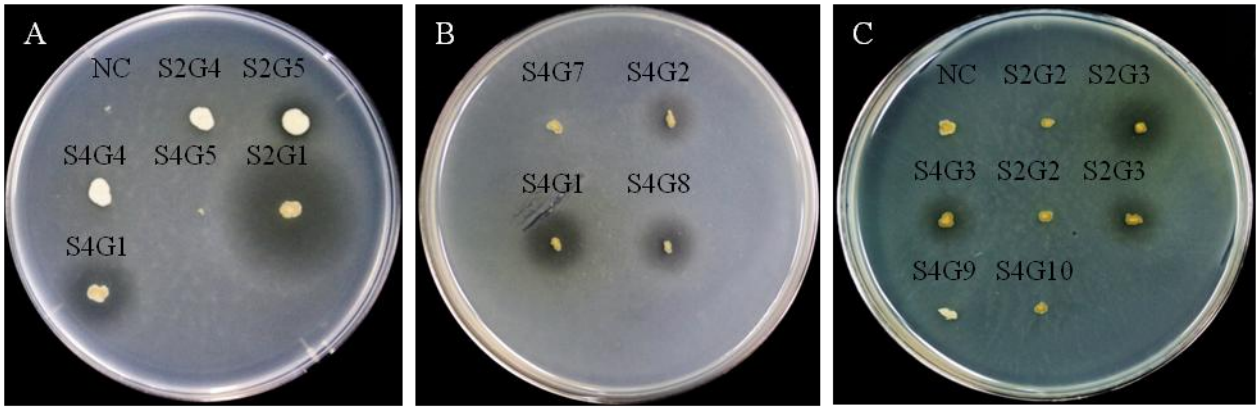

Fig. 3.2. Plate assay results of alginate lyase gene candidate transformants. A and B: pET160, C: pET161. NC: Negative control (pET160/CAT and pET161/CAT).

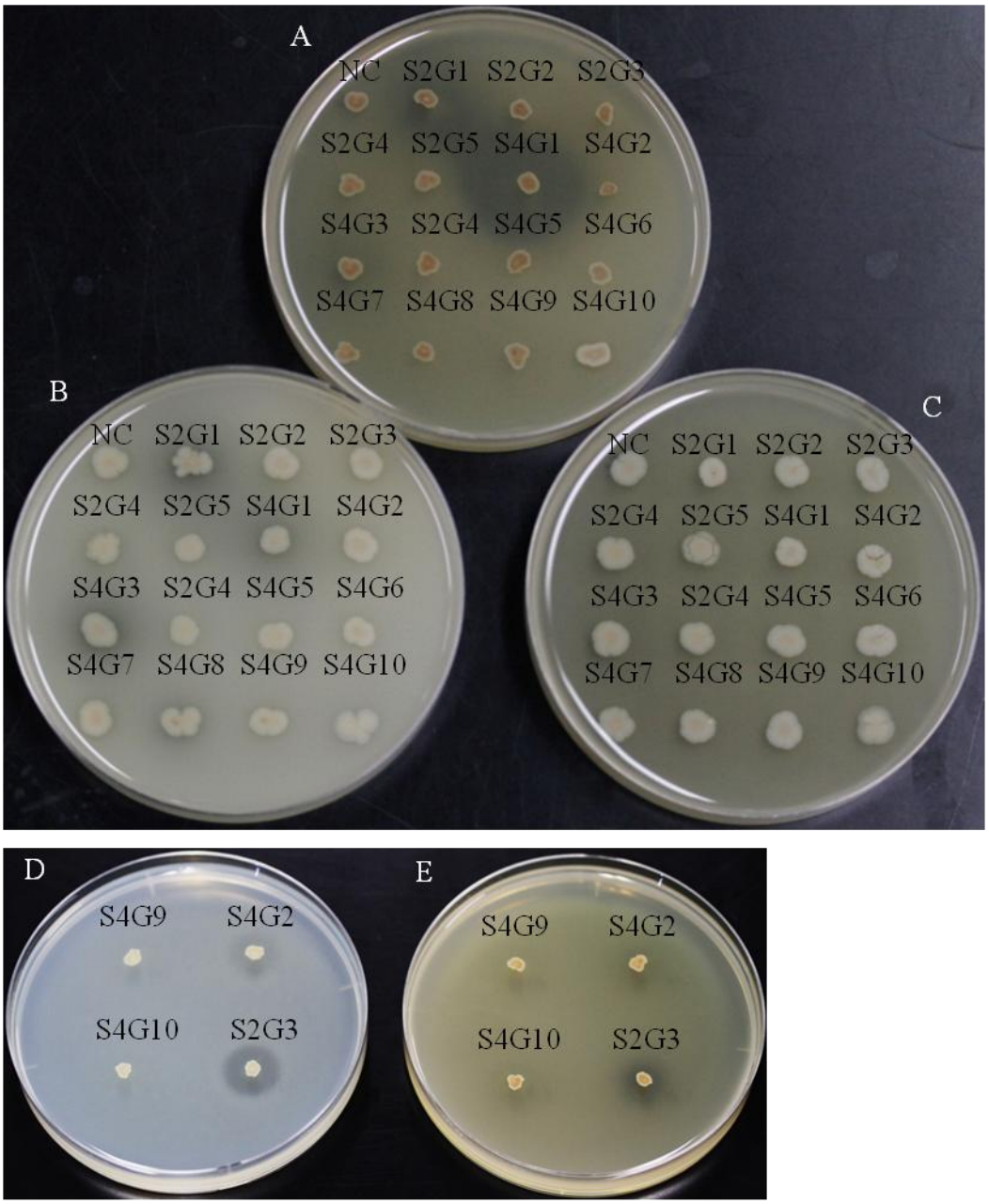

Fig. 3.3. Substrate specific of alginate lyase gene candidate transformants using plate assays. A, LB agar with 1% MG-random medium, B: LB agar with 1% G-rich medium, C: LB agar with 1% M-rich medium, D: culture extension using LB agar with 1% G-rich medium, E: culture extension using LB agar with 1% MG -rich medium.

extracellular secreting alginate lyases. No activity on the plate was observed in S2G2/pET161, S2G4/pET160, S4G4/pET160, S4G5/pET160, S4G6/pET161, S4G7/pET160, S4G9/pET161 and S4G10/pET161.

Moreover, a substrate specificity of each gene product was determined on the plate containing not only sodium alginate but also polyG (F_G=0.769, F_M=0.237, F_{GG}=0.858, F_{MM}=0.32) and polyM (F_G=0.128, F_M=0.872, F_{GG}=0.170, F_{MM}=0.914) substrates (Fig. 3.3). S2G1/pET160, S4G1/pET160, S4G3/pET161, and S4G8/pET160 showed polyG specific activities (Fig. 3.3B). S2G3/pET161 and S4G2/pET160 might be polyG specific, but there were no polyM activities. S2G5/pET160 specificity has not been determined yet.

Measurement of alginate lyase activity in cell free extract and culture supernatant of the clones

A total of 7 alginate lyase gene products showed alginate lyase activity on the plate, meaning these gene products might be secreted proteins. As cytoplasmic and periplasmic alginate lyases were also predicted by bioinformatic tools described in Table 3.1, I tried to measure intracellular alginate lyase activity from the cell free extract of clones of interest.

Before the detecting the alginate lyase activity, the time to get maximum alginate lyase activity was determined using S2G1/pET160 and S4G10/pET161 as representative clones of extracellular and intracellular enzymes, respectively (Table 3.1, Fig. 3.4). These 2 clones were cultured in LB+50% ASW, and then prepared using SF and CFE at 12-hour intervals. In S2G1/pET160, alginate lyase

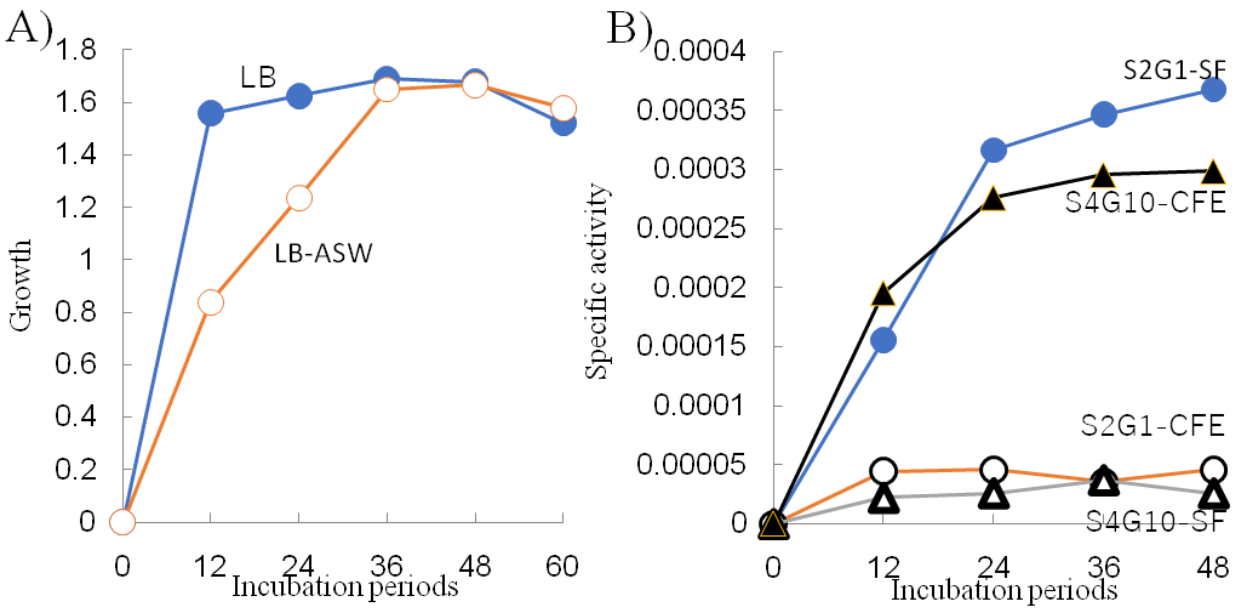

Fig. 3.4. Growth (A) and specific activity change (B) of the transformatns which were cloned by alginate lyase gene. LB: LB-broth without ASW, LB-ASW: LB-broth with 50% ASW, SF: supernatant fraction, CFE: cell free extrect.

Table 3.4. Summary of cloning and expression experiments for alginate lyase gene candidates

Gene	The closest gene in Swiss-Prot/ TrEMBLDB	Family and domain information				Localizzation	Activity	
		Pfam	Super Family	Gene3D	CAZyme	PSORT B predication	Plate assay (diameter)	Localization in *E. coli* determined by TBA
S2G1	Q9RGQ3 _9VIBR Alginate lyase AlyVGI	Alginate lyase2 (58-351)	Glucanases superfamily (58-352)	Glucanases subgrp (50-351)	PL7	EC (9.72)	Positive (2.27 cm)	SF
S2G2	I8PZ55 _YERPE Alginate lyase family protein	Alginate lyase (99-271)	Alginate lyase (8-330)	Alginate lyase (98-341)	PL17	CP (8.96)	Negative	CFE
S2G3	A3UR44 _VIBSP Putative alginate lyase	Alginate lyase2 (34-342)	Glucanases superfamily (34-354)	Glucanases subgrp (28-242)	PL7	PP (5.91)	Positive (1.26 cm)	NT
S2G4	Q9RGQ3 _9VIBR Alginate lyase AlyVGI	Alginate lyase2 (58-351)	Glucanases superfamily (58-352)	Glucanases subgrp (50-351)	PL7	EC (9.72)	Negative	ND
S2G5	A3XWI1 _9VIBR Putative alginate lyase	NA	Pectin lyas like (67-537)	Pectin lyas fold (73-521)	PL6	CM=PP=OM=EC (2.5)	Positive (0.92 cm)	NT
S4G1	G3LI08 _9BACT Alginate lyase	Alginate lyase2 (49-323)	Glucanases superfamily (49-323)	Glucanases subgrp (41-323)	PL7	EC (9.72)	Positive (1.76 cm)	SF
S4G2	G4WFC4 _9VIBR Alginate lyase	Alginate lyase2 (5-185, 195-481)	Glucanases superfamily (7-187, 195-483)	Glucanases subgrp (5-184, 190-482)	PL7	CP (8.96)	Positive (1.33 cm)	NT
S4G3	C4TJD4 _9VIBR Alginate lyase	Alginate lyase2 (252-516), F5_F8_type_C (45-162)	Glucanases superfamily (252-517), Gal_bind_like (33-168)	Glucanases subgrp (246-516)	CBM32, PL7	EC (9.72)	Positive (1.29 cm)	NT
S4G4	G4WFC4 _9VIBR Alginate lyase	Alginate lyase (47-272), Hepar_II_III (367-506)	Alginate lyase (45-295)	Alginate lyase (55-318)	PL17	CP=CM=PP=OM=EC (2.0)	Negative	ND
S4G5	G4WFC5_9VIBR Alginate lyase	Alginate lyase (75-263), Hepar_II_III (368-525)	Alginate lyase (38-305)	Alginate lyase (61-292)	PL17	CP=CM=PP=OM=EC (2.0)	Negative	ND
S4G6	Q9RGQ2 _9VIBR Alginate lyase AlyVGIII	KdgM (16-235)	Oligogalacturonate specific porin	NA	NA	OM (10.00)	Negative	ND
S4G7	Q9RGQ2 _9VIBR Alginate lyase AlyVGIII	KdgM(16-235)	Oligogalacturonate specific porin	NA	NA	OM (10.00)	Negative	ND
S4G8	Q9RGQ5 _9VIBR Alginate lyase AlyVGII	Alginate lyase2 (65-322)	Glucanases superfamily (65-331)	Glucanases subgrp (63-330)	PL7	EC (9.65)	Positive (1.68 cm)	SF
S4G9	A3XWI1_9VIBR Putative alginate lyase	NA	Pectin lyas like (67-537)	Pectin lyas fold (73-521)	PL6	CM (4.90)	Negative	ND
S4G10	A6CUW5_9VIBR Oligo alginate lyase	Hepar_II_III (415-570)	Alginate lyase (72-418)	NA	PL15	CP (8.96)	Negative	CFE

activity was detected in SF from 12 hour-culture cells, and the activity increased up to 0.00035 U/mg protein at 48 hour-cultured cells. But faint activity was detected from CFE. On the other hand, while the activity of S4G10/pET161 was detected from CFE culture after 12 hours, and the activity peaked at 36 hours; the activity of SF was faint.

Results of activity measurements are summarized in Table 3.4. Alginate lyase activity in CFE was also detected from S2G2/pET161 (0.000101559 U/mg). Unfortunately, no activity was detected from S4G4/pET160 and S4G5/pET160, which are encoding multidomain enzymes. In a total of 9 genes alginate lyase were detected; 7 were extracellular, and 2 were intracellular. These activity localizations matched the PsortB prediction results (Table 3.4). Interestingly, S2G5/pET160 showed alginate degrading activity (Fig. 3.2); although this gene was grouped into pectin lyase like family and PL6, and possessed a pectin lyase fold.

Among the extracellular enzymes, the zone diameter of clones S2G1, S2G3, S2G5, S4G1, S4G2, S4G3, and S4G8 were different. They ranged from 1.26 cm (S2G1) to 2.27 cm (S2G3) (Table 3.4).

DISCUSSION

Even in the periods when computation science and bioinformatics are rapidly expanding, nearly 50% of the prokaryotic gene functions cannot be identified by in silico analysis alone. Alginate lyases are core enzymes in the alginate metabolism of *V. halioticoli*, and have important links to the metabolic design of "Alginate to Biofuel" in the bacterial cells. In this study, I determined the more

probable alginate lyase genes using tools which predict protein families and domains of the gene products, and protein localization. A total of 11 genes are found to be more likely to be alginate lyase genes (Table 3.1). Two other genes (S4G6 and S4G7) are oligosaccharide specific porin genes, and the other 2 genes are protein containing PL6 type pectin lyase folds. The actual function of the gene products, however, must be determined to help us understand the whole picture of alginate metabolism in the bacterium.

Not only the 11 more probable alginate lyase genes but also the other 4 genes were successfully cloned in *E. coli* using Lumio expression vectors. Among the 11 genes, 9 gene products showed alginate degrading properties (Table 3.5); 7 identified as extracellular proteins, and 2 were as intercellular. It is interesting to note that the S2G4 gene product, which showed 57% amino acid identity with *alyVGI* (and the gene S2G1), did not show any alginate degradation activity.

Tentative molecular phylogenetic analysis among bacterial PL7 genes may show 1) possible common ancestry among S2G1, S2G3 and S2G4, and 2) convergence between S2G1 related genes and S4G8 (Fig. 3.5). To analyze whether the S2G4 gene might be a pseudogene or to have evolved into a new gene, three dimensional models of S2G1 and S2G4 gene products were also predicted by the Phyre 2 program (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) (Kelley and Sternberg, 2009). Fortunately, 72% of S2G1 residues were modelled showing >90% confidence using three PL7 alginate lyases (d1uaia, c2z42A, and d1vava) as templates and 68% of S2G4 residues were also modelled with 100% confidence using the same structural data as templates. Both models indicated both gene products contained β-sandwich with a deep cleft, which is

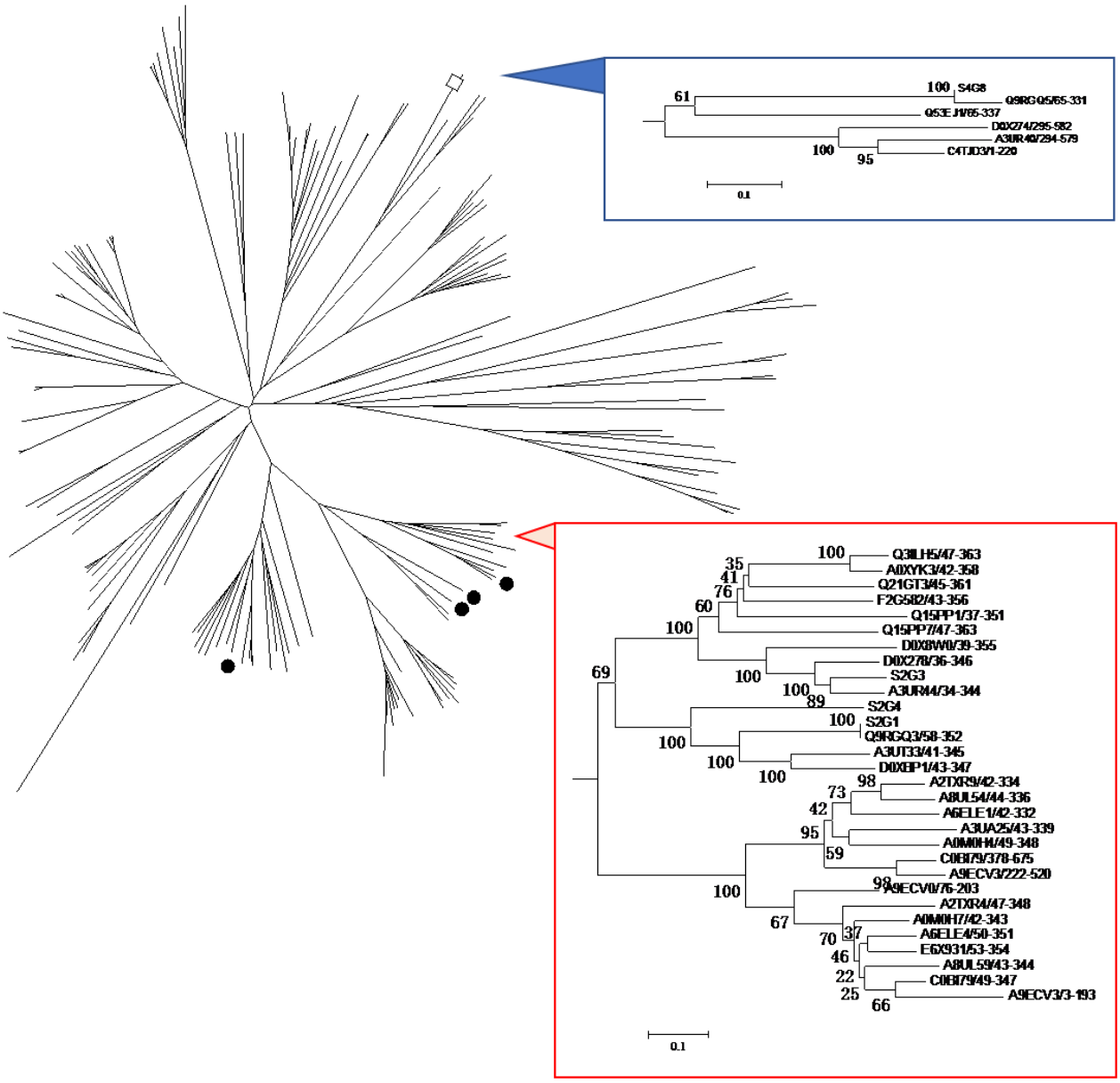

Fig. 3.5. Molecular phylogeny of PL7 alginate lyase on the basis of deduced amino acid sequences. Sequence data were retrieved from Pfam database, and aligned by ClustalX software. The tree was visualized by MEGA5 with boot-strap (100 replications) neighbor joining algorithm with Poisson correction and pairwise deletion.

Accession numbers were follos;

1) *Vibrionaceae*; S2G1, S2G3, S2G4, S4G8, D0X8W0, D0X278, A3UR44, Q9RGQ3, A3UT33, D0XBP1, Q9RGQ5, Q53EJ1, D0X274, A3UR40, C4TJD3.

2) *Flavobacteriaceae*; A2TXR9, A8UL54, A3UA25, A0M0H4, COBI79, A9ECV3, A9ECV0, A2TXR4, A0M0H7, E6X931, A8UL59, C0BI79, A9ECV3.

3) *Pseudoalteromonadaceae*; Q3ILH5.

4) *Alteromonadaceae*; Q21GT3, A0XYK3, F2G582, Q15PP1, Q15PP7.

5) Environmental samples; A6ELE1, A6ELE4.

common structure observed in PL7 alginate lyases (Ogura et al., 2008) (Fig. 3.6). Surprisingly, superimposed modelling among S2G1, S2G4 and *Sphingomonas* A1-II' enzymes also showed similar three dimensional structures (Fig. 3.6). According to the secondary and three dimensional structure comparisons by Yamasaki et al. (2004) and Ogura et al. (2008), two mutations in S2G4, R150H and Y278H (A1-II' numbering), were found in near regions of the possible catalytic residues predicted in the *Sphingomonas* A1-II'. Further characterization of the gene products and structure-function analysis will give new insights into how much alginate lyase genes have diverged. A very high activity was detected from the S2G1 clone, but not in the S2G4, which showed the *V. halioticoli* cell is the naturally evolving gene pool of alginate lyases.

From the clear zone diameter and protein specific activity, the activity localization of each alginate lyase appeared to be similar compared with the PSORT gene analysis (Table 3.4). The accuracy of forecasting is higher than 90%. Four possible extracellular proteins and two possible intracellular proteins displayed the same results as the prediction. This improves the feasibility of further experiments according to gene analysis. Furthermore, many reference materials could be obtained and could contribute to a deeper understanding of the protein-protein interaction and the mode of action. In fact, interesting alginate secretion machinery interacting with alginate synthetic Alg proteins and alginate lyase, AlgL, was recently proposed in a mucoid strain of *P. aeruginosa* (Jain and Ohman, 2005)

Finally, I concluded that *V. halioticoli* possesses at least five types of functional alginate lyases; 1) extracellular secreting (or periplasmic) 35-40 kDa PL7 types (S2G1, S2G3, S4G1 and S4G8), 2) a

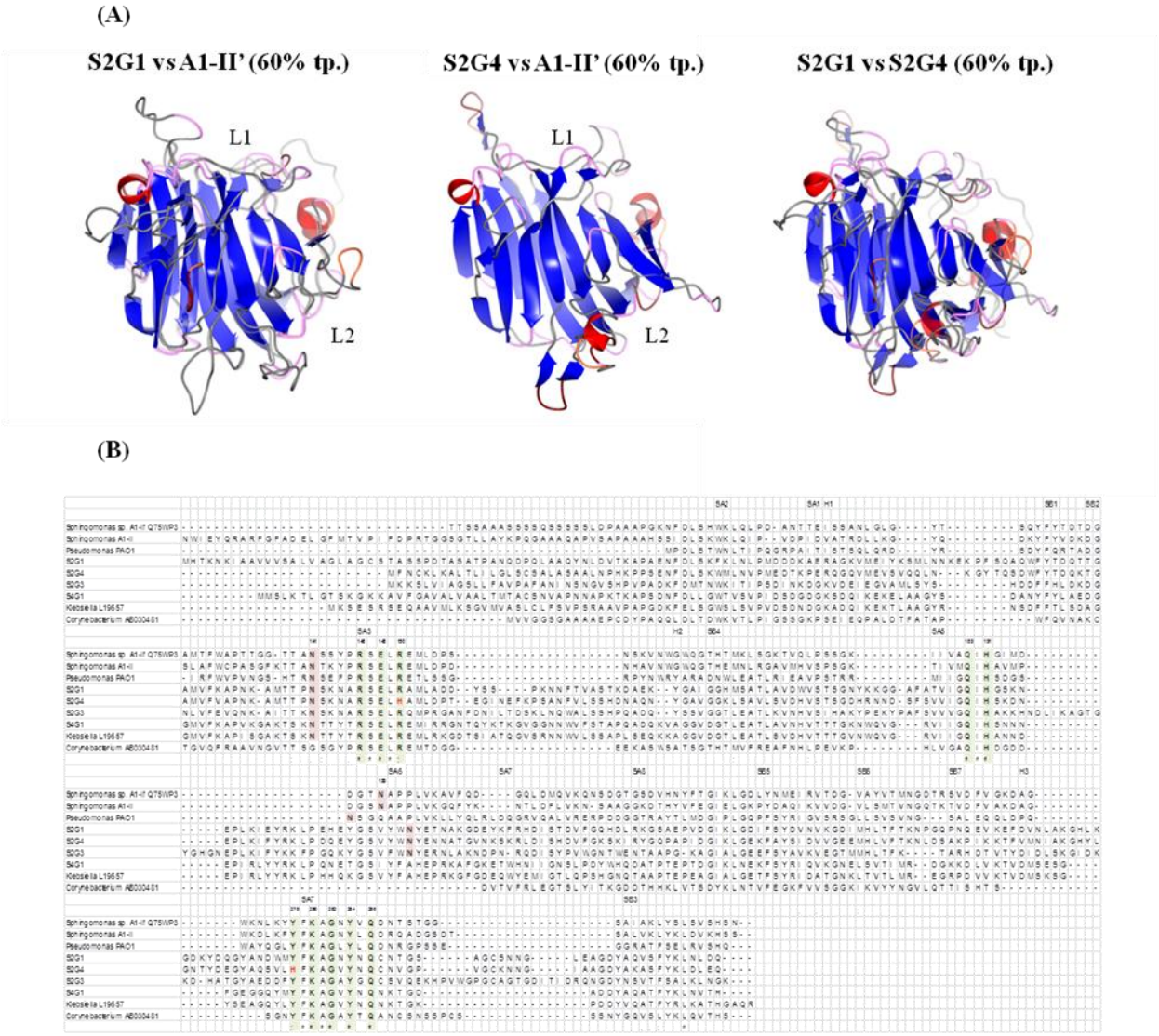

Fig. 3.6. Superposing of 3D models among S2G1, S2G4 and the *Sphingomonas* A1-II' (A) and the observed mutation in S2G4 products (B) (numbering is from Ogura et al., 2008)

cytoplasmic 80 kDa PL17 type (S2G2), 3) a cytoplasmic 70 kDa dual domain PL7 (S4G2), 4) an extracellular 57 kDa dual domain (PL7 and CBM32) alginate lyase (S4G3), and 5) cytoplasmic 79 kDa PL15 type alginate lyases (S4G10) (Table. 3.1). The extracellular functional alginate lyases were polyG specific. Unfortunately, the activities of localization unknown 81 kDa dual domain PL17 type (S4G4 and S4G5) failed. Further biochemical studies for each enzyme are necessary.

Gene S2G5 is a special case; low gene similarity and unidentified hypothetical localization cannot resolve the contribution in the alginate degradation in silico, but now I understood the gene product is alginate lyase extracellular secreted by the expression studies. Genes S2G3 and S4G2 may be new alginate lyases as yet unidentified in the *V. halioticoli* cells. Actual functions of S4G6 and S4G7 will also be studied further due to the successful cloning and the expression in this study. In the future, they all will become the subject of more research.

CONCLUSION

Here, I further describe the alginate metabolic pathway in *V. halioticoli* IAM 14596^T strain using whole genome analysis (chapter 2) and gene cloning, expression (in this chapter) results (Fig. 2.7). As a unique alginolytic marine bacterium, at least 11 activity genes were worded in the first step of alginate degradation (until unsaturated uronic acid) in this bacterium. When the strain growth on medium with alginate, the cell will produce a type of extracellular enzyme (S2G1, S2G5, S4G1, S4G2, S4G3 and S4G8). This enzyme is first transported to outside of cell, and functions in cutting the large molecule alginate to oligoalginate. Small molecule oligoalginate could be carried and

transported into periplasmic through the porin protein (S4G6 and S4G7). In periplasmic, the oligoalginate was degraded once more using periplasmic alginate lyase (S2G3). Associated consumption of one unit ATP, the smaller molecule oligoalginate could be transported into the cell using symporter (predict as S4G9), and all alginate lyase could be directly responsible for this oligoalginate. Under the activity of oligoalginate lyase (S2G2 and S4G10), oligoalginate was degraded twice to unsaturated uronic acid, then 4-deoxy-L-erythro-5-hexoseulose uronic acid (DEH), After that, DEH was converted into 2-keto-3-deoxy-D-gluconate (KDG) using dehydrogenases, which is further phosphorylated by a carbohydrate kinase into 2-keto-3-deoxy-phosohogluconate (KDGP). Then, KDGP was converted into pyruvate using aldolase. More recently, Inohara (2014) solved the problem of key gene absence (ADH and PDC). The *V. halioticoli* cell is used to create a new biocatalyst implemented with Production of Ethanol (PET) operon and the engineered vibrio actually produces ethanol from alginate. In the next stage, efficiency is the key word. The gene expression controls are necessary. In the most direct, feature comparison was performed for Halioticoli clade species.

Chapter 4. *Vibrio ishigakensis* sp. nov., in *Halioticoli* clade isolated from seawater in Okinawa coral reef area, Japan

ABSTRACT

Five novel strains showing non-motile, alginolytic, halophilic and fermentative features were isolated from seawater samples off Okinawa in coral reef areas. These strains were characterized by an advanced polyphasic taxonomy including genome based taxonomy using multilocus sequence analysis (MLSA) and *in silico* DNA-DNA similarity (*in silico* DDH). Phylogenetic analyses on the basis of 16S rRNA gene sequences revealed that the isolates could be assigned to the genus *Vibrio*, however they were not allocated into any distinct cluster with known *Vibrionaceae* species. MLSA based on eight protein-coding genes (*gapA*, *gyrB*, *ftsZ*, *mreB*, *pyrH*, *recA*, *rpoA*, and *topA*) showed the vibrios formed an outskirt branch of *Halioticoli* clade. The experimental DNA-DNA hybridization data revealed that the five strains were in the range of being defined as conspecific but separate from nine *Halioticoli* clade species. The G+C contents of the *V. ishigakensis* strains were 47.3-49.1 mol%. Both Amino Acid Identity and Average Nucleotide Identity of the strain C1^T against *V. ezurae* HDS1-1^T, *V. gallicus* HT2-1^T, *V. halioticoli* IAM 14596^T, *V. neonatus* HDD3-1^{T}and *V. superstes* G3-29^{T}showed less than 95% similarity. The genome-based taxonomic approach by means of *in silico* DDH values also supports the *V. ishigakensis* strains being distinct from the other known *Halioticoli* clade species. Sixteen traits (growth temperature range, DNase and lipase production, indole production, and assimilation of 10 carbon compounds) distinguished these strains from *Halioticoli* clade species. The name *V. ishigakensis* sp. nov. is proposed for the species of *Halioticoli* clade, with C1^T as the type strain (JCM 19231^T=LMG 28703^T).

INTRODUCTION

The genus *Vibrio* is a large group of bacteria isolated from marine, brackish and human origins showing Gram-negative and sodium ion requirement for growth (Gomez-Gil, 2003). These bacteria are associated with host-microbe interactions in the marine ecosystem including fish/shellfish pathogenicity (Austin and Zhang, 2006), symbiosis with marine organisms (Ruby, 1996; Ruby et al., 2005), and make a significant contribution to nutrient cycles by mediating organic matter decomposition. Strains of the genus *Vibrio* also show versatile metabolism ability. They are capable of degrading and assimilating complex organic matters such as polysaccharides both under aerobic and anaerobic conditions (Thompson and Polz., 2006). In addition to that, *Vibrionaceae* have attracted a lot of attention due to its physiological and genomic plasticity. The number of known species showing gas production, nitrogen fixation, phototrophy, and non-motility is also increasing (Thompson et al., 2006). So far, more than 100 *Vibrio* species have been described. Their phenotypic and molecular data show that they are highly heterogeneous (Thompson et al., 2001).

The multilocus sequence analysis (MLSA) technique has greatly improved the taxonomy of vibrios. It has high levels of discriminatory capability and assigns many existing species to specific groups (Sawabe et al., 2013). Currently 25 vibrio clades including the *Salinivibiro-Grimontia-Enterovibrio* super-clade are proposed (Al-saari et al., 2015; Sawabe et al., 2013). The clade *Halioticoli* consists at present of seven species namely *V. ezurae, V. gallicus, V. halioticoli, V. neonatus, V. inusitatus, V. breoganii* and *V. superstes*. Also, *V. comitans* and *V. rarus* were named as possible *Halioticoli* clade members (Sawabe et al., 2007b). All species except *V. breoganii* were isolated from gut of abalone, and *V. breoganii* was isolated from the cultured clams *Ruditapes philippinarum* and *R.*

decussatus (Sawabe et al., 2013). They are alginolytic, non-motile due to lack of flagellation and fermentative examples of marine bacteria. Abalone isolates of *Halioticoli* clade are regarded as possible symbionts that have relationship with the host herbivorous animals in the digestion and conversion of alginate to short chain fatty acids (Sawabe et al., 2003).

In this study, five alginolytic vibrio strains phylogenetically related to *Halioticoli* clade were isolated from seawater samples collected from a coral reef are a near Ishigaki Island, Okinawa, Japan. These bacteria were distinguished from other known vibrio species on the basis of advanced polyphasic taxonomy which integrate the analysis of genomic, phenotypic, genotypic, and phylogenetic, and belong to *Halioticoli* clade. Interestingly, this report described a new species in *Halioticoli* clade isolated from seawaters without association of host marine life, and these bacteria might hold important clues to elucidate the evolutionary history of *Halioticoli* clade strains. Genomic, phenotypic, and systematic features are provided here to propose these strains as *V. ishigakensis* sp. nov.

MATERIALS AND METHODS

Bacterial strains

Five strains, C1^T, C5, C64, C212, and C216, were isolated from seawater samples in the Ishigaki Island coral reef area, Okinawa, Japan. Strains C1^T and C5 were collected from Taketomi Island (24°20.5260' N, 124°05.6443' E), C64 from Miyara outer reef (24°20.5489' N, 124°13.0408' E), and C212 and C216 were collected from the Sekisei outer reef (24°21.7557' N, 124°02.7190' E). The

seawater samples were taken at a depth of 4 m while SCUBA diving. No specific permissions were required for the seawater sampling activities in these locations. All isolates were first cultured on thiosulphate-citrate-bile-sucrose (TCBS) medium (Nissui Pharmacy, Tokyo, Japan) and later purified on ZoBell 2216E agar medium at 25℃. The strains were stored at -80℃with 10% glycerol-supplemented broth.

Phylogenetic analysis based on a 16S rRNA gene

The amplification primers used corresponded to positions 25 to 1521 in the *Escherichia coli* sequence to produce a 1.5 kb PCR product. The sequences were aligned with an aligned data set obtained from the Association of Vibrio Biologists (*AViB*) website (http://www2.ioc.fiocruz.br/vibrio/AVib/AVib.html) using the Clustal X program (Larkin et al., 2007). Finally, 1400 bp 16S rRNA gene segment of 141 *Vibrionaceae* species was used to reconstruct the phylogenetic tree using MEGA programs ver. 6.06 (Tamura et al., 2013). For Fig. 4.1, the phylogenetic tree was reconstructed using three methods: neighbor-joining (NJ) (Saitou and Nei, 1987), maximum parsimony (MP) and maximum likelihood (ML) (Swofford, 1992) methods with *E. coli* K-12 as the outgroup. Evolutionary distances of NJ were corrected using the Jukes-Cantor method.

Multilocus sequence analysis (MLSA)

MLSA was conducted in the same manner as in Sawabe et al. (2007b, 2013). Finally, eight protein-coding genes (*gapA, gyrB, ftsZ, mreB, pyrH, recA, rpoA,* and *topA*) of the candidate strains were analyzed. Bacterial DNA of all the strains used for PCR was extracted using the Promega

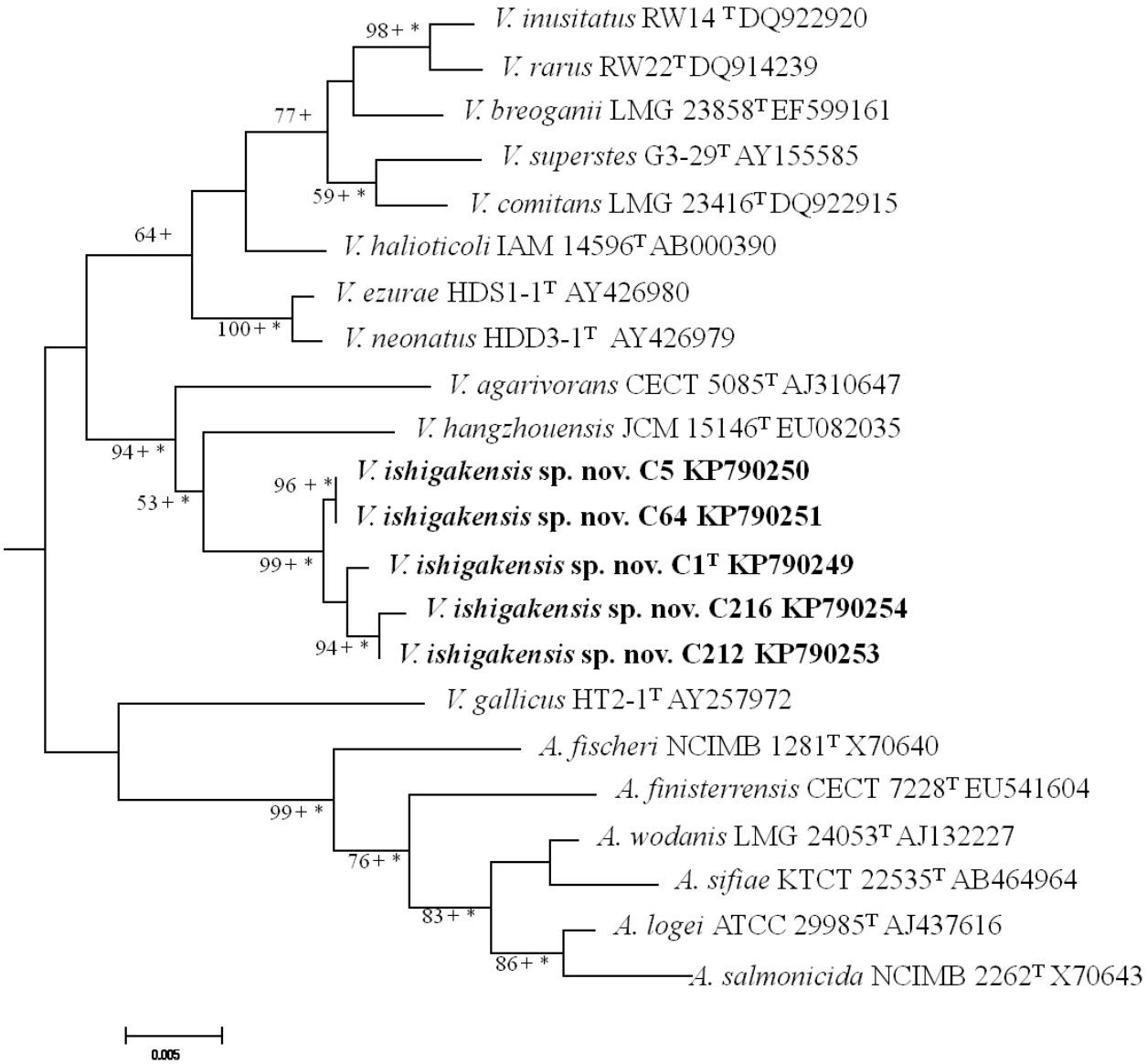

Fig. 4.1. Unrooted phylogenetic tree on the basis of 16S rRNA gene sequences**.** This figure combines the results of three analyses, neighbor-joining (NJ), maximum-parsimony (MP) and maximum-likelihood (ML). The topology shown was obtained by NJ, and the percentage values are the results of a bootstrap analysis using 1000 replications. Branches also obtained in the ML and MP analyses are indicated by + and *, respectively.

Wizard Genomic DNA extraction system according to the protocol provided by the manufacturer (Promega, Madison, WI). Eight protein-coding genes of *V. ishigakensis* strains were amplified using primer sets reported previously using the following PCR steps (Table 4.1); the initial denaturation step consisted of heating the reaction mixture at 96℃ for 120 s, and thermal profiles consisted of 30 cycles of denaturation at 96℃ for 60 s, annealing at 50℃ (or 55℃ for *mreB* amplification using a mreB-985R primer) for 60 s and extension at 72℃ for 120 s. The PCR products were analyzed on a 1.5% agarose gel with a molecular weight standard. The PCR products which produced a single band on agarose gels were purified for sequencing using a Wizard SV Gel and PCR clean-up system (Promega). Then 50 ng of DNA template was directly sequenced using the BigDye terminator sequencing kit (v.3.1) with an Applied Biosystems model 3130x automated sequencer.

Determination of moles percent G+C content and DNA-DNA hybridizations

DNAs of all strains were prepared as described by Marmur (1961), with minor modification. Mole percent G+C contents of DNA were determined using high-performance liquid chromatography (HPLC) (Tamaoka and Komagata, 1984). DNA-DNA hybridization experiments were performed as described by Sawabe et al. (1998) using a fluorometric direct binding method in microdilution wells (Ezaki et al., 1988). Briefly, DNAs of *V. ishigakensis* strain C1^T *V. agarivorans* CETC 5085^T and *V. hangzhouensis* JCM 15146^T were labeled with photobiotin (Vector Laboratories, Inc., Burlingame, CA). Unlabeled single-stranded DNA from target strains were immobilized in microdilution wells (Immuron 200, FIA/LIA plate, black type, Greiner labotechnik, Germany), then 20 ng of labeled DNA was added to each microdilution well and the hybridization was performed under optimal

Table 4.1. Primer sequences for MLSA.

Primer	Sequence	Length	Tm	Anealing temp. (℃)
ftsZ-43F	AAAGTCGTAGGTGTTGG	17	43	50
ftsZ-670R	CCGCTACCCATCATAGC	17	45	
ftsZ-75F	GCTGTTGAACACATGGTACG	20	46	
ftsz-800R	GCACCAGCAAGATCGATATC	20	46	
gapA-150F	AACTCACGGTCGTTTCAAC	19	43	50
gapA-910R	CGTTGTCGTACCAAGATAC	19	43	
gyrB-300F	TCATCATGACGGTACTGCAC	20	44	50
gyrB-1001R	CCTTTACGACGAGTCATTTC	20	44	
mreB-67F	CTGATTTATGTAAAAGG	17	40	50 (55*)
mreB-985R	CATCTCTAGAGCTTTACCGCC	21	56	
mreB-360F	GTGGTTCTACCCAAGTTGAG	20	51	
mreB-999R	CCGTGCATATCGATCATTTC	20	51	
pyrH-80F	GATCGTATGGCTCAAGAAG	19	43	50
pyrH-530R	TAGGCATTTTGTGGTCACG	19	43	
recA-88F	ATGCGCCTTGGTGACAAC	18	45	50
recA-882R	CTTGTCACCATTGTAGCTG	19	43	
recA-130F	GTCTACCAATGGGTCGTATC	20	46	
recA-720R	GCCATTGTAGCTGTACCAAG	20	46	
rpoA-1F	ATGCAGGGTTCTGTAACAG	19	43	50
rpoA-500R	TTGGACGCTCATCTTCTTC	19	43	
topA-400F	GAGATCATCGGTGGTGATG	19	46	50
topA-1200R	GAAGGACGAATCGCTTCGTG	20	51	
topA-407F	GATGAAGAGCGATACAAACG	20	43	
topA-2144R	CCACACTTGTCACACTCAAC	20	49	

* The annealing temperature was used only for mreB-985R primer.

conditions following pre-hybridization. Formamide concentration in the hybridization mixture was determined according to Meinkoth-Wahl (Meinkoth and Wahl, 1984). The hybridization of the biotinylated DNA to immobilized DNAs was performed under optimal conditions following pre-hybridization and detected using fluorometry after binding streptavid in-β-galactosidase to the labeled DNA. 4-Methylumbelliferyl-β-D-galactopyranoside (6 x 10^{-4} M; Wako, Osaka, Japan) was added to each well as fluorogenic substrate for β-galactosidase and incubated at 30℃. Then, fluorescence intensity of each well was measured using a microplate reader (Infinite 200, Tecan, city, country) at wavelength of 360 nm for excitation and 450 nm for emission. DNA-DNA homology was calculated according to Ezaki et al. (1988 and 1989).

Genome sequencing and *in silico* DNA-DNA similarity calculation

Draft genome sequences of strain C1^T, C5 and C216 was obtained using the Ion Torrent PGM platform. The genome sequences were assembled using Newbler ver. 2.8 producing 260, 56 and 33 contigs, respectively. The sequence was deposited in the DDBJ/GenBank/EMBL public database under accession numbers described below.

General genome features were determined using the Rapid Annotations Using Subsystems Technology (The RAST server version 4.0) (Aziz et al., 2008). *In silico* DDH values were estimated to *V. ezurae* HDS1-1^T, *V. gallicus* HT2-1^T, *V. halioticoli* IAM14596^T, *V. neonates* HDD3-1^T and *V. superstes* G3-29^T (deposited under the Bioproject number PRJDB2474) using the Genome-to-Genome Distance Calculator with strain C1^T, C5, and C216 as reference (GGDC 2.0) (Meier-Kolthoff et al., 2013). This online tool infers genome-to-genome distances between pairs of

entirely or partially sequenced genomes. Intergenomic distances are employed for wet-lab DDH prediction. Briefly, genome pairs are aligned with BLAST+ (Camacho et al., 2009) to generate a set of high-scoring segment pairs (HSPs). The information they contained (e.g., the total number of identical base pairs) was transformed into a distance value by the best-fit formula, according to (Meier-Kolthoff et al., 2013). DDH prediction from intergenomic distance, including confidence intervals, was provided by a tested generalized linear model (GLM) (Nelder and Wedderburn, 1972), with log transformation (Meier-Kolthoff et al., 2013). Amino Acid Identity (AAI) and Average Nucleotide Identity (ANI) were calculated according to Konstantinidis and Tiedje (2005) and Thompson et al. (2013) respectively, on the same strains used for *in silico* DDH.

Phenotypic characterization and genome based phenotypes

All strains were cultured on ZoBell 2216E agar medium (Oppenheimer and Zobell, 1952) and the phenotypic characteristics were determined as described previously under the same conditions (Baumann and Schubert, 1984; Gomez-Gil, 2003 and Gomez-Gil et al., 2014; Leifson, 1963; Ostle and Holt, 1982; Thompson et al., 2015). Alginate lyase activity was determined using ZoBell 2216E agar medium with 0.5% alginate (Sugimura et al., 2000). Phenotypic features were also obtained directly from the whole genome sequences by automatically searching the database comparisons of the genes that define the metabolic pathways of each diagnostic feature and their regulatory genes. The method was described by Amaral et al. (2014).

Nucleotide sequence accession number

The genome sequences were deposited at DDBJ/EMBL/GenBank under the accession numbers

BBRZ01000001-260, BBSA01000001-56, and BBSC01000001-33 for *V. ishigakensis* C1^T (JCM 19231), C5 (JCM 19232), and C216 (JCM 19241), respectively.

The five protein-coding gene (*topA, mreB, ftsZ, gapA* and *gyrB*) sequences of *Vibrio breoganii* LMG 23858^T obtained by Sanger sequencing were deposited to GenBank under KU990865, KU990866, KU990867, KU990868, and KU990869, respectively.

The 16S rRNA gene sequences of C1^T, C5, C64, C212 and C216 obtained by Sanger sequencing were deposited to GenBank under KP790249, KP790250, KP790251, KP790253, and KP790254, respectively.

RESULTS AND DISCUSSION

The broad phylogenetic analyses based on 16S rRNA gene sequences of 141 *Vibrionaceae* species clearly showed that these new strains belong to the genus *Vibrio*. Among the strains of *V. ishigakensis* C1^T, C5, C64, C212, and C216, they showed high levels of 16S rRNA gene sequence similarities ranging from 99.6 to 99.9%. These *V. ishigakensis* strains formed a robust cluster. *V. hangzhouensis* JCM 15146^T and *V. agarivorans* CECT 5085^T were related neighbors (Fig. 4.1). However, similarities to *V. hangzhouensis* JCM 15146^T and *V. agarivorans* CECT 5085^T were 97.7% and 97.4%, respectively. These similarities are below the most recent threshold for the species boundary (Kita-Tsukamoto et al., 1993). MLSA was further performed to infer which clades strains of *V. ishigakensis* C1^T, C5, C64, C212, and C216 belong using eight protein-coding genes (Sawabe

et al., 2007b). *V. ishigakensis* strains formed a robust subclade sharing a common ancestry with the previously known *Halioticoli* clade species (Sawabe et al., 2013). The similarity of concatenated sequences of the eight MLSA genes between *V. ishigakensis* and the other *Halioticoli* clade species were 82.0 ±1.6%.

We further determined the level of genomic cohesion among the *Halioticoli* clade species *in vitro* and *in silico*. The DNA G+C content of these novel strains are 47.3, 49.0, 49.1, 48.8 and 47.4 for *V. ishigakensis* C1^T, C5, C64, C212, and C216, respectively. These moles percentages fall within the acceptable range for inclusion in genus *Vibrio* (Gomez-Gil et al., 2014), i.e. 38 to 54% and supported by initial phenotypic and phylogenetic analyses. Experimental DNA-DNA hybridization (DDH) results showed that five strains of *V. ishigakensis* C1^T, C5, C64, C212, and C216 were conspecific with >70% DDH values (Table 4.2). Using C1^T strain as a labeled strain, the DDH values against *V. agarivorans* CECT 5085^T, *V. breoganii* LMG 23858^T, *V. comitans* LMG 23416^T, *V. ezurae* HDS1-1^T, *V. gallicus* HT2-1^T, *V. halioticoli* IAM 14596^T, *V. hangzhouensis* JCM 15146^T, *V. inusitatus* RW14^T, *V. neonatus* HDD3-1^T, *V. rarus* RW22^T, and *V. superstes* G3-29^T, were 18%, 18%, 16%, 14%, 15%, 21%, 13%, 14%, 15%,12%, and 16%, respectively. DDH with *V. agarivorans* CECT 5085^T and *V. hangzhouensis* JCM 15146^T as labeled strains revealed 13% and 10% DDH values against to *V. ishigakensis* C1^T strain.

Genome comparison supported the conspecificity of *V. ishigakensis* strains and the allospecificity against *Halioticoli* clade species. Available draft genome sequences of *V. ishigakensis* C1^T, C5, and C216, *V. ezurae* HDS1-1^T, *V. gallicus* HT2-1^T, *V. halioticoli* IAM 14596^T, *V. neonates* HDD3-1^T

Table 4.2. DNA-DNA hybridization (DDH) results

V. ishigakensis C1^T strain as a labelled strain	DDH values (%)
V. ishigakensis C1^T	100
V. ishigakensis C5	92
V. ishigakensis C64	94
V. ishigakensis C212	92
V. ishigakensis C216	90
V. agarivorans CECT 5085^T	18
V. breoganii LMG 23858^T	18
V. comitans LMG 23416^T	16
V. ezurae HDS1-1^T	14
V. gallicus HT2-1^T	15
V. halioticoli IAM 14596^T	21
V. hangzhouensis JCM 15146^T	13
V. inusitatus RW14^T	14
V. neonatus HDD3-1^T	15
V. rarus RW22^T	12
V. superstes G3-29^T	16
V. agarivorans CECT5085^T strain as a labelled strain	
V. agarivorans CECT 5085^T	100
V. ishigakensis C1^T	13
V. hangzhouensis JCM15146^T strain as a labelled strain	
V. hangzhouensis JCM 15146^T	100
V. ishigakensis C1^T	10

and *V. superstes* G3-29^T were used to provide a more robust argument over their novelty using *in silico* DDH, AAI, and ANI (Table 4.3). *In silico* DDH (%) values (Formula 2, recommended) of C1^T against *V. ezurae* HDS1-1^T, *V. gallicus* HT2-1^T, *V. halioticoli* IAM 14596^T, *V. neonatus* HDD3-1^{T}and *V. superstes* G3-29^T were 28.1 ± 2.4%, 28.6 ± 2.4%, 27.8 ± 2.4%, 28.1 ± 2.4% and 28.8 ± 2.4%, respectively. The *in silico* DDH estimator performed three kinds of calculations (Formula 1 to 3). These interspecific *in silico* DDH values were unlikely to vary. However, interspecific *in silico* DDH values among *V. ishigakensis* strains calculated by the Formula 2 seemed to be lower (60-65%) than those calculated using the other formulae (ca. 85%) (data not shown). We still do not know the reason why the estimates showed such variation, however, experimental DDH provided strong evidence that five *V. ishigakensis* strains were conspecific. Similarly, separate analysis on the AAI and ANI revealed the distant relatedness of C1^T against *V. ezurae* HDS1-1^T, *V. gallicus* HT2-1^T, *V. halioticoli* IAM 14596^T, *V. neonates* HDD3-1^T and *V. superstes* G3-29^T with less than 95% similarity (Table 4.3). These values are below the threshold for species designation and do not fall into the range of conspecific species defined by having >5% mol G+C difference of the genomic DNA, <70% DDH similarity, <95% AAI and ANI against closely related species (Barton, 2005; Thompson et al., 2009 and 2013). The experiments with combination of phylogenetic analyses (Fig. 4.1 and 4.2) showed that strains C1^T, C5, C64, C212, and C216 should be recognized as a new species.

Interestingly, the *V. ishigakensis* strains are likely to share a common ancestry with *Halioticoli* clade species on the basis of the concatenated sequences of eight MLSA genes. In fact, all strains of *V. ishigakensis* possessed alginate lyase activity, and were non-motile (Sawabe et al., 2013). Table

Table 4.3. *In silico* DDH values, average amino acid identity and average nucleotide identity. Species/strains are follows; (1) *V. ishigakensis* C1^T; (2) *V. ishigakensis* C5; (3) *V. ishigakensis* C216; (4) *V. ezurae* HDS1-1^T, (5) *V. gallicus* HT2-1^T, (6) *V. halioticoli* IAM 14596^T, (7) *V. neonatus* HDD3-1^T, (8) *V. superstes* G3-29^T

In silico DDH (%)	1	2	3	4	5	6	7	8
V. ishigakensis C1^T		61±3	65±3	28±2	29±2	28±2	28±2	29±2
V. ishigakensis C5	64±3		61±3	28±2	28±2	28±2	28±2	29±2
V. ishigakensis C216	64±4	61±3		20±2	20±2	20±2	20±2	20±2
Average amino acid identity (AAI) (%)								
V. ishigakensis C1^T		95	95	76	81	76	76	77
V. ishigakensis C5	96		95	76	81	76	76	77
V. ishigakensis C216	95	95		76	80	76	77	77
Average nucleodite identity (ANI) (%)								
V. ishigakensis C1^T		97	96	89	91	89	90	89
V. ishigakensis C5	97		97	89	91	89	89	90
V. ishigakensis C216	99	97		89	90	90	90	90

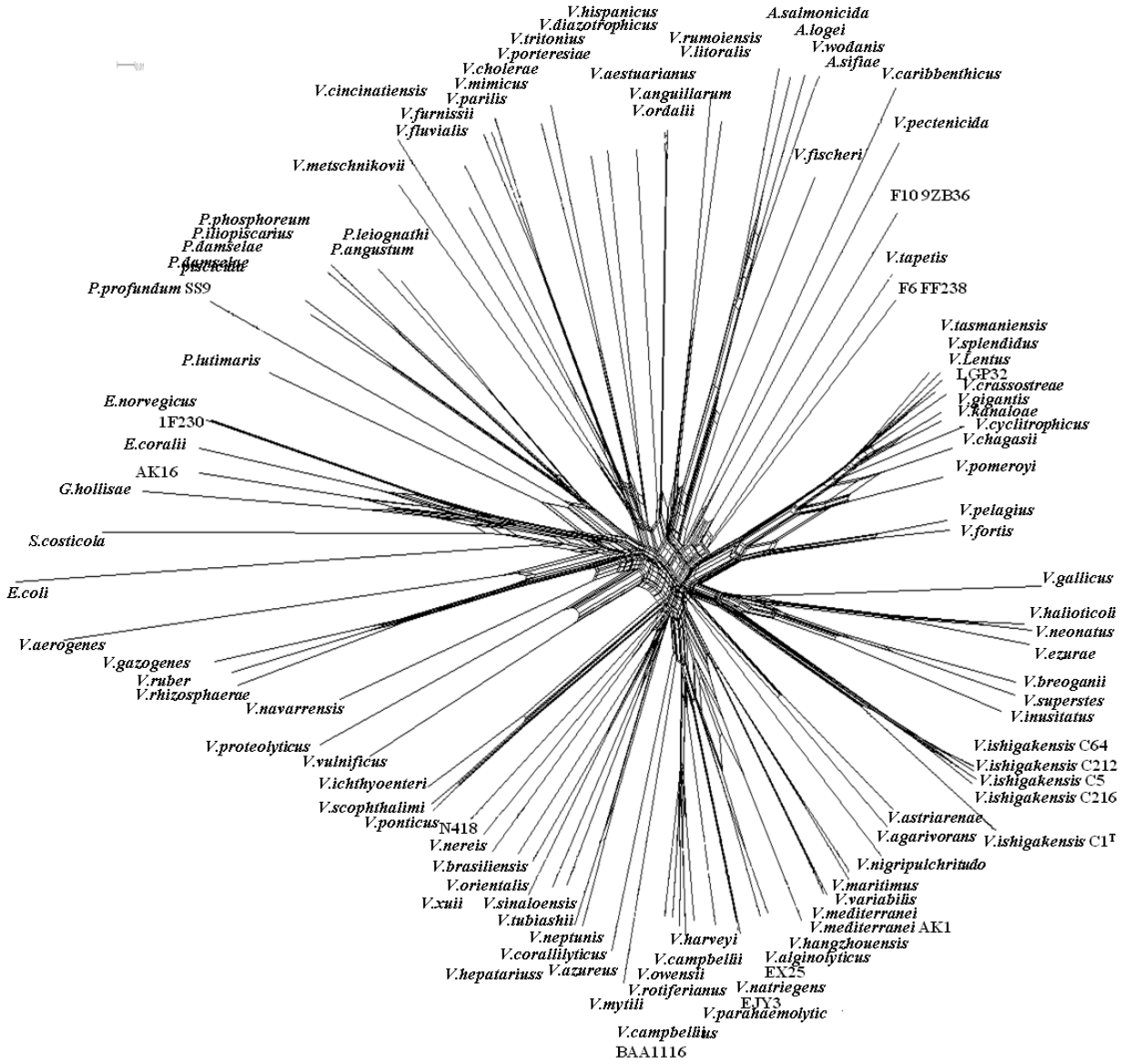

Fig. 4.1. Concatenated split network tree based on eight gene loci. The gapA, gyrB, ftsZ, mreB, pyrH, recA, rpoA, and topA gene sequences of 98 taxa were concatenated including the representative of novel vibrios in the current study (*Vibrio ishigakensis* sp. nov. C1T, C5, C64, C212 and C216). Phylogenetic tree was generated using the SplitsTree4 program. *V. ishigakensis* sp. nov. C1T, C5, C64, C212 and C216 (shown in bold) were protruded from the centre of the tree.

4.4 shows a comparative analysis of phenotypic and biochemical features of novel strains against the type strains. A total of 16 traits distinguished these *V. ishigakensis* strains from other *Halioticoli* clade species (Sawabe et al., 1998). They grow at temperatures between 20 to 40 ℃ and the optimum growth temperature was 30 ℃, which is one of the key features in differentiating them from other species in *Halioticoli* clade. This growth temperature shift (up to 40 ℃) observed in *V. ishigakensis* sp. nov. might reflect their adaptation to subtropical coral reef ecosystems. The *V. ishigakensis* strains tested positive for production of lipase and DNase, and were able to assimilate D-mannose, D-galactose, sucrose, trehalose, and cellobiose, and tested negative for indole production and assimilation of D-gluconate, D-sorbitol, melibiose, lactose and pyruvate, which are also the key differentiating features in carbon assimilations among *Halioticoli* clade species (Table 4.4). It is noted that the phenotypic profiles of each *V. ishigakensis* strain were almost identical except for some variables in amylase production, nitrate reduction, and assimilation of D-glucuronate, γ-aminobutyrate, meso-erythritol, D-xylose and DL-malate. In addition, the gene coding for the major diagnostic phenotypes of the novel species such as fermentation of D-mannitol, maltose and sucrose were detected in the genome sequences of *V. ishigakensis* strain C1^{T} (Amaral et al., 2014). A genome-based taxonomy allows the determination of species-specific phenotypes in silico (Thompson et al., 2015). Positive results in amylase production and negative results in nitrate reduction were observed in *V. ishigakensis* C1^{T} and are also unique clues in differentiating it from the other *Halioticoli* clade species (Table 4.4).

Table 4.4. Phenotypic characteristics for distinguishing *Vibrio ishigakensis* and their closely related species. Taxa are indicated as: (1) *V. ishigakensis* C1^T, (2) *V. ishigakensis* C5, C64, C212, and C216, (3) *V. breoganii* LMG 23858^T, (4) *V. comitans* LMG 23416^T, (5) *V. ezurae* HDS1-1^T, (6) *V. gallicus* HT2-1^T, (7) *V. halioticoli* IAM 14596^T, (8) *V. inusitatus* RW14^T, (9) *V. neonatus* HDD3-1^T, (10) *V. rarus* RW22^T, (11) *V. superstes* G3-29^T.

Characteristics	1	2	3	4	5	6	7	8	9	10	11
Growth at											
15°C	−	−	+	+	+	+	+	+	+	+	+
37°C	+	+	−	−	−	−	−	−	−	−	−
40°C	+	+	−	−	−	−	−	−	−	−	−
Production of											
Amylase	+	d	−	−	−	−	−	−	−	−	−
DNase	+	+	−	+	−	−	−	−	−	−	−
Lipase	+	+	−	−	−	−	−	−	−	−	−
Nitrate reduction	−	d	+	+	+	+	+	+	+	+	+
Indole production	−	−	−	−	+	+	+	−	+	+	−
Utilization of											
D-Mannose	+	+	−	−	+	−	−	+	−	+	+
Sucrose	+	+	−	−	−	−	−	−	−	+	+
D-Gluconate	−	−	+	+	+	−	−	+	−	+	+
D-Sorbitol	−	−	−	−	−	−	−	+	−	+	−
D-Galactose	+	+	+	+	−	−	−	−	−	−	+
Cellobiose	+	+	+	+	−	−	−	+	−	+	+
Melibiose	−	−	+	−	−	−	−	−	−	−	+
Lactose	−	−	+	−	−	−	−	−	−	−	+
D-Glucuronate	−	d	−	+	−	−	−	−	−	−	+
Trehalose	+	+	−	−	−	−	−	−	−	−	+
γ-Aminobutyrate	−	d	+	+	−	−	−	+	−	−	+
Pyruvate	−	−	+	+	−	−	−	+	−	+	−
D-Xylose	−	d	+	+	−	−	−	+	−	+	+

Data for reference species were obtained from Sawabe et al. [30, 31 and 35]. +, all strains tested positive; –, all strains tested negative; NA, not available. d: different reactions. All species are fermentative, non-motile, oxidase- and catalase- positive, require Na^+ for growth, grow at 20–30 °C, grow on TCBS, produce alginase, produce acid from D-glucose, D-mannitol and maltose, utilization of D-fructose, D-glucose, maltose, D-mannitol, D-glucosamine and N-acetylglucosamine. All species are negative for pigmentation, swarming, gas production from D-glucose, arginine dihydrolase, lysine and ornithine decarboxylases, luminescence, hydrolysis of agar, production of gelatinase and acetoin, and utilization of putrescine, L-tyrosine, propionate, meso-erythritol, L-arabinose, DL-malate and citrate.

CONCLUSION

In conclusion, to justify *V. ishigakensis* sp. nov. as a new species of vibrio, 16S rRNA gene and MLSA on the basis of 8 protein-coding gene sequence phylogenies as strong selective measure had grouped the strains into the genus *Vibrio* and distinguished them as different from any known *Vibrio* spp. Despite showing similar alginolytic activity, the strains differed from *V. halioticoli* IAM 14596^T and their other closest relatives. Further genomic analyses including DNA G+C content, DDH, MLSA, genomic sequencing, *in silico* DDH, AAI and ANI of the strain $C1^T$ provided definitive results that *V. ishigakensis* is the newest member of *Vibrionaceae*. Comparative data on the basis of phenotypic characters also placed them in the genus *Vibrio* and simultaneously supports their novelty (Table 4.5). The phenotypic features of *V. ishigakensis* $C1^T$ can be retrieved directly from its genome sequence. *V. ishigakensis* sp. nov. is the first *Halioticoli* clade species not associated with marine invertebrate hosts isolated from coral reef seawater. It is also interesting to compare the genome construction between *V. ishigakensis* and the other *Halioticoli* clade species in the aspects of the unique ecophysiology of *V. ishigakensis*, e.g. adaptation to subtropical marine environments and ecological transition between free-ling and fecal-enteric habitat (Maciá n et al., 2004).

Table 4.5. A list of accession numbers for Figure 1

Bacterial species	Strain	*ftsZ*	*gapA*	*mreB*	*topA*	*gyrB*	*pyrH*	*recA*	*rpoA*
V. ishigakensis	C1=JCM 19231^T=LMG 28703^T	KU682729	KU682734	KU682744	KU682759	KU682739	KU682749	KU682764	KU682754
V. ishigakensis	C212	KU682730	KU682735	KU682745	KU682760	KU682740	KU682750	KU682765	KU682755
V. ishigakensis	C216	KU682731	KU682736	KU682746	KU682761	KU682741	KU682751	KU682766	KU682756
V. ishigakensis	C5	KU682732	KU682737	KU682747	KU682762	KU682742	KU682752	KU682767	KU682757
V. ishigakensis	C64	KU682733	KU682738	KU682748	KU682763	KU682743	KU682753	KU682768	KU682758
V. aerogenes	LMG 19650^T	KF697253	KF697261	KF697278	KF697313	KF697270	KF697286	KF697295	KF697304
V. agarivorans	CECT 5085^T= LMG 21449^T= DSM 13756^T	KT339396	FJ436364	KT339394	KT339395	FJ436362	FJ463225	KT281478	KT281479
V. alginolyticus	LMG 4409^T	EF027344	DQ907274	DQ907405	DQ907472	AB298202	GU266285	AJ842373	AJ842558
V. astriarenae	C7^T = CAIM 1900^T = LMG 28701^T	BBMQ01000001-336 (WGS data)*							
V. azureus	NBRC 104587^T	AB428903	AB428907	AB428905	HE655452	AB428899	AB428901	FR669656	AB465316
V. litoralis	DSM 17657^T	AUFZ01000001 (WGS data)							
V. rumoiensis	FERM P-14531^T= ST16^T	DQ907374	DQ907307	DQ907446	DQ907516	AB298245	NZ_AJYK01000090.1	AJ842503	AJ842690
V. tapetis	LMG 19706^T= CECT 4600^T	DQ907379	DQ907311	DQ907451	DQ907520	AB298249	HE795189	HE795219	AJ842730
V. nigripulchritudo	LMG 3896^T	EF027347	DQ907297	DQ907437	DQ907505	AB298236	GU266290	AJ842480	AJ842667
V. proteolyticus	LMG 3772^T= NBRC 13287^T	EF114210	DQ907305	DQ907444	DQ907514	AB298261	NZ_BATJ01000010	AJ842499	AJ842686
V. gigantis	LMG 22741^T	DQ481629	DQ481617	DQ979358	DQ481655	AB298219	EU871951	EU541593	EU541573
V. pomeroyi	LMG 20537^T	DQ481634	DQ481621	DQ481646	DQ481660	AB298242	EU871960	AJ842497	AJ842684
V. lentus	CAIM 454 = R-3884	DQ914234	DQ481620	DQ914236	DQ914237	AB298226	32*	AJ842453	AJ842640
V. splendidus	LMG 19031^T	DQ481635	DQ481622	DQ481647	DQ481662	AB298248	KF899231	AJ842511	AJ842725
V. kanaloae	LMG 20539^T	DQ481631	DQ481619	DQ481643	DQ481657	AB298225	FN908851	AJ842450	AJ842637
V. cyclitrophicus	LMG21359^T	DQ481625	DQ481613	DQ481639	DQ481651	AB298211	EU871958	AJ842405	AJ842592
V. chagasii	LMG 21353^T	DQ996590	DQ481611	DQ481637	DQ481649	AB298206	EU118252	AJ842385	AJ842572
V. fortis	LMG 21557^T= CECT 8657^T	DQ907346	DQ907282	DQ907417	DQ907484	AB298216	LN713958	AJ842422	AJ842609
V. pelagius	ATCC 25916^T= CECT 4202^T	DQ907369	DQ907302	DQ907442	DQ907511	AB298241	LN713960	AJ580872	AJ842682
V. cholerae	ATCC 39315	NC0_02505 (WGS data)							
V. mimicus	LMG 7896^T= ATCC 33653^T	DQ907357	DQ907292	DQ907430	DQ907498	AB298230	EU118242	EF643485	EF643486
V. 'partilis'	RC586	ADBD01000006 (WGS data)							
V. navarrensis	LMG 15976^T= 1397-6^T	DQ907360	DQ907295	DQ907433	DQ907501	AB298233	KJ807123	AJ842474	KJ807163
V. vulnificus	LMG13545^T= ATCC 27562^T	DQ907382	DQ907313	DQ907454	DQ907522	AB298252	GQ382226	GQ382229	GQ382243
V. maritimus	R40493	GU929927	KF666681	GU929931	GU929939	KF666700	GU929933	GU929935	GU929937
V. variabilis	LMG 25438^T= CAIM 1454 = R-40492^T	GU929926	KF666686	GU929930	GU929938	GU929928	GU929932	GU929934	GU929936
V. mediterranei	AK1 = LMG 19703	DQ907377	DQ907310	DQ907449	DQ907496	NZ_ABCH01000001	GU266289	AJ842507	AJ842695
V. mediterranei	LMG11258^T	DQ907356	DQ907290	DQ907428	DQ907495	AB298228	GU266288	AJ842459	AJ842644
V. owensii	CAIM 1816 = R40496	AB609124	AB609125	GU078686	GU078704	GU078680	GU078692	GU078693	GU078697
V. harveyi	LMG4044^T= NCIMB1280^T= ATCC 14126^T	DQ907350	DQ449616	DQ907422	DQ907488	AB298221	FM202541	DQ648369	KC954196
V. campbellii	ATCC BAA1116	NC_009783 (WGS data)							
V. campbellii	LMG 11216^T	DQ907337	DQ449614	DQ907408	DQ907475	AB298205	EF596641	AJ842377	AJ842564
V. rotiferianus	LMG 21460^T	DQ907372	DQ449619	DQ907445	DQ907515	AB298244	EF596722	AJ842501	AJ842688
V. tasmaniensis	LMG 20012^T	DQ481636	DQ481623	DQ481648	DQ481661	AB298250	EU871961	AJ842515	AJ842731
V. natriegens	LMG 10935^T= CECT 526^T	DQ907359	DQ907294	DQ907432	DQ907500	AB298232	FM202573	FM204826	AJ842658
V. parahaemolyticus	LMG 2850^T	DQ907367	DQ449618	DQ907440	DQ907509	AB298239	GU266286	AJ842490	AJ842677
V. mytili	LMG 19157^T	DQ907358	DQ907293	DQ907431	DQ907499	AB298231	GU266287	HQ455540	AJ842657
V. gazogenes	ATCC 29988^T	KF697255	KF697264	KF697280	KF697314	KF697272	KF697288	KF697297	KF697306
V. ruber	LMG 23124^T	KF697259	KF697268	KF697284	KF697319	KF697276	KF697292	KF697301	KF697310
V. rhizosphaerae	MSSRF 3^T	KF697258	KF697267	KF697283	KF697318	KF697275	KF697291	KF697300	KF697309
V. 'tritonius'	AM2^T	GU951702	GU969215	GU969223	GU969227	GU969219	KF697293	KF697302	KF697311
V. porteresiae	MSSRF 30^T	KF697257	KF697266	KF697282	KF697317	KF697274	KF697290	KF697299	KF697308
V. diazotrophicus	LMG7893^T= CECT 627^T	DQ907342	DQ907280	DQ907413	DQ907480	AB298212	HE805632	AJ842411	AJ842598
V. hispanicus	LMG 13240^T	DQ907353	DQ907286	DQ907425	DQ907492	AB298223	32*	AJ842445	AJ842632
V. fluvialis	LMG 7894^T	DQ907345	DQ907281	DQ907416	DQ907483	AB298215	JN426808	AJ842419	AJ842606
V. furnissii	LMG 7910^T= CAIM 518^T	EF027345	DQ907283	DQ907418	DQ907485	AB298217	JF316672	AJ842427	AJ842614
V. anguillarum	LMG 4437^T	DQ907334	DQ907275	DQ907406	DQ907471	AB298203	32*	AJ580852	AJ842561
V. ordalii	LMG 13544^T= ATCC 33509^T	DQ907364	DQ907298	DQ907438	DQ907506	AB298237	NZ_AEZC01000192	AJ842482	AJ842669
V. aestuarianus	LMG 7909^T	DQ907331	DQ907271	DQ907402	DQ907469	AB298200	32*	AJ842369	AJ842554
V. cincinatiensis	LMG 7891^T	DQ907340	DQ907278	DQ907411	DQ907479	AB298208	32*	AJ580853	AJ842582
V. metschnikovii	LMG 11664^T	EF027346	DQ907291	DQ907429	DQ907497	AB298229	NZ_ACZO01000006		AJ842650
V. ponticus	DSM 16217	DQ907371	DQ907304	DQ907443	DQ907513	AB298243	JF316676	JF316685	JQ308806
V. scophthalmi	LMG 19158^T	DQ907376	DQ907309	DQ907448	DQ907518	AB298247	HM771376	HM771381	HM771386

Continue

V. ichthyoenteri	LMG 19664^T = ATCC 700023^T	DQ907354	DQ907287	DQ907426	DQ907493	AB298224	HM771375	HM771380	HM771385
V. hepatarius	LMG 20362^T	DQ907352	DQ907285	DQ907424	DQ907491	AB298222	JF316674	AJ842444	AJ842631
V. orientalis	LMG 7897^T = ATCC 33934^T	DQ907365	DQ907299	DQ907439	DQ907507	AB298238	EU118243	EU130528	AJ842672
V. tubiashii	LMG 10936^T = ATCC 19109^T	DQ907381	DQ907312	DQ907453	DQ907521	AB298251	JF316670	AJ842518	AJ842734
V. brasiliensis	LMG 20546^T	DQ907335	DQ449619	DQ907407	DQ907473	AB298204	HM771374	AJ842376	HM771384
V. sinaloensis	DSM21326				AEVT01000067 (WGS data)				
V. caribbeanicus	ATCC BAA2122				AEIU01000122 (WGS data)				
V. coralliilyticus	LMG 20984^T	DQ907341	DQ907279	DQ907412	EF114213	AB298210	GU266292	AJ842402	JN039157
V. neptunius	LMG 20536^T	DQ907361	DQ907296	DQ907435	DQ907503	AB298234	GU266291	AJ842478	JN039153
V. pectenicida	LMG 19642^T	DQ907368	DQ907301	DQ907441	DQ907510	AB298240	JN039143	AJ842491	AJ842678
V. nereis	LMG 3895^T	DQ907362	DQ449617	DQ907436	DQ907504	AB298235	JN968379	AJ842479	AJ842666
V. xuii	LMG 21346^T	DQ907384	DQ907315	DQ907456	DQ907524	AB298254	GU266284	AJ842529	AJ842742
V. superstes	CAIM 904^T = LMG 21323^T	KF697260	KF697269	KF697285	KF697320	KF697277	KF697294	KF697303	KF697312
V. breoganii	LMG 23858^T	KU990867	KU990868	KU990866	KU990865	KU990869	EU889130	EU541585	EU541565
V. inusitatus	CAIM1811	KF666653	DQ922912	KF666719	KF666765	KF666698	EU871957	EU541600	EU541579
V. gallicus	HT2-1^T = LMG 21330^T	KF697254	KF697263	KF697279	KF697314	KF697271	KF697287	KF697296	KF697305
V. ezurae	JCM 21522^T = LMG 19970^T = DSM 17533^T	DQ907343	AY546645	DQ907414	DQ907481	AB298213	EU871949	AJ842413	AJ842600
V. halioticoli	IAM 14596^T = LMG 18542	DQ907349	AY546638	DQ907421	DQ907487	AB298220	EU871952	AJ842430	AJ842617
V. neonatus	JCM 21521^T = LMG 19973^T	KF697256	KF697265	KF697281	KF697316	KF697273	KF697289	KF697298	KF697307
V. crassostreae	LMG 22240^T	DQ481624	DQ481612	DQ481638	DQ481650	AB298209	EU871948	EU541594	EU541574
V. sp.	N418				AFWD01000095 (WGS data)				
V. sp.	F6 FF238				AJYW01000195 (WGS data)				
V. sp.	F10 9ZB36				AJYQ01000069 (WGS data)				
V. sp.	LGP32				NC011753 (WGS data)				
V. sp.	EX25				NC013456 (WGS data)				
V. sp.	EJY3				NC_016613 (WGS data)				
Grimontia hollisae	LMG 17719^T = CAIM 625^T	EF027348	DQ907317	DQ907398	EF114215	AB298259	JF739393	AJ842351	AJ842535
EnteroV. coralii	LMG 22228^T	DQ907329	DQ907268	DQ907396	EF114217	AB298198	JF739392	AJ842347	AJ842530
E. norvegicus	LMG 19839^T = CAIM 430^T	EF027349	DQ907269	DQ907397	EF114216	AB298198	JF739391	AJ842348	AJ842531
EnteroV. sp.	1F-230				AJYH01000076 (WGS data)				
EnteroV. sp.	AK16				NZ_ANFM01000001 (WGS data)				
SaliniV. costicola subsp. costicola	LMG 11651^T	EF02735	DQ907316	DQ907399	DQ907468	AB298255	ASAI0100000 1	AJ842367	AJ842552
Photobacterium angustum	ATCC 25915^T = LMG 8455^T	DQ907318	DQ907257	DQ907385	DQ907457	AB298187	EF380235	EF415544	AJ842538
P. damselae subsp. piscicida	DI21				AKYG01000013 (WGS data)				
P. damselae subsp. damselae	ATCC 33539^T = LMG 7892^T	DQ907319	DQ907258	DQ907386	DQ907458	AB298188	EF380236	AJ842357	AJ842541
P. leiognathi	NCIMB 2193^T = ATCC 25521^T	DQ907324	DQ907263	DQ907391	DQ907463	AB298193	EF380238	EF415546	EF415581
P. iliopiscarius	DSM 9896^T = ATCC 51760^T	DQ907322	DQ907261	DQ907389	DQ907461	AB298191	EF380237	EF380245	EF380251
P. lutimaris	CAIM 1851	KF666644	KF666663	KF666709	KF666753	KF666688	KF666723	KF666730	KF666737
P. phosphoreum	IAM 14401^T = ATCC 11040^T	DQ907326	DQ907265	DQ907393	DQ907465	AB298195	EF380239	EF415550	EF415585
P. profundum	SS9 = ATCC BAA-1253				CR378673 (WGS data)				
AltiV. sifiae	KCTC 22535 = H1-1	AB464982	AB464977	AB464992	AB464987	AB464968	AB464972	AB464996	AB465000
A. fischeri	LMG 4414^T = ATCC 7744^T	DQ907344	AY546637	DQ907415	DQ907482	AB298214	EF415528	EU907941	AJ842604
A. wodanis	NCIMB 13582^T = ATCC BAA-104^T	DQ907383	DQ907314	DQ907455	DQ907523	AB298253	EU118246	EU257781	EF380250
A. logei	LMG 14011	DQ907355	DQ907289	DQ907427	DQ907494	AB298227	KF697321	AJ842456	AJ842643
A. salmonicida	LFI 1238				NC_011312 (WGS data)				
Escherichia coli	K-12				NC_010473 (WGS data)				

* these sequences are depositted in TaxVibrio database without not accession number. Sequences are attached as a supporting material.

Chapter 5. General Discussion

The development of key technologies for renewable energy sources has become an important challenge in the era in establishing "Sustainable Society". Renewable energies such as bioethanol, bio-higher alcohol and biohydrogen, are produced from various carbohydrates present in terrestrial plants using well controlled microbial energy conversion metabolism. However, biofuel production, involves a paradox in the energy-food conflict. o we need to find alternative fermentation substrates and unique microbes for the development of food-uncompetitive biofuel production for use in the future (Ingram et al., 1987; Ohta et al., 1991; Dien et al., 2003; Stephanopoulos, 2007; Atsumi et al., 2008; Atsumi et al., 2010; Lee et al., 2010; Dürre and Richard, 2011; Geddes et al., 2011). The challenges also could support to achieve some of goals, "Zero Hunger", "Affordable and Clean Energy", and "Climate Action", setting as the Sustainable Development Goals (SDGs) (http://www.undp.org/content/undp/en/home/sustainable-development-goals.html).

As the next-generation feedstock, marine macroalgae have emerged as a viable option (Stokstad, 2012). It can be used to produce biofuel or precursors, but there are still many difficulties. Seaweed contains a variety of unique carbohydrates. The carbohydrates are used for human consumptions (McHugh, 2003). Among them, alginate is a major component of brown macroalgae, and we now know the oxidized carbohydrate is resistant to biofuel conversion, but it is succeeded in some bacteria form terrestrial sources such as *E. coli* and *Sphingomonas* sp. (Takeda et al., 2011; Wargacki et al. 2012). However, we still do not have many successes using marine microbes.

V. halioticoli IAM14596^{T} was isolated from the gut of the abalone *Haliotis discus hannai* (Sawabe et al., 1995; Sawabe et al., 1998; Sawabe et al., 2006). The bacterium has a high native ability to ferment mannitol and alginate. The wild type cannot convert alginate to ethanol directly, but the metabolically engineered cells with installed PET operon were able to produce ethanol (Inohara, 2016). The *V. halioticoli*-like strains are widespread in the gut of abalone, and currently 10 species in the Halioticoli clade have been described, including the newly described *Vibrio ishigakensis* and symbiotic association to the host abalone has been considered (Sawabe et al., 2006). Therefore the *V. halioticoli* is a biotechnologically and ecologically important marine bacterium but the complete genome sequence has not yet been achieved. To improve the biocatalysts in producing bioethanol more efficiently and to know how symbiotic association has been established between host abalone and the bacteria, the detailed metabolic pathway of alginate and the gene expression controls of such genes in the bacterium should be elucidated.

To understand the detail alginate metabolic pathway in *V. halioticoli*, the technical improvements in genome sequence technologies and developments of genome wide gene mining tools in silico (Fraser et al., 1995; Land et al., 2015) give us many effective means. In this study, I undertook complete genome sequencing of *V. halioticoli* IAM 14596^{T} using Next Generation Sequencer (NGS) including the 3rd generation PacBio (Miyamoto et al., 2014). And then, genome wide mining of genes responsible for alginate degradation was conducted, and the expression of the candidate genes was attempted.

Using a sequencer combination of Sanger, 454 FLX, Illumina, and PacBio, the genome sequence of *V. halioticoli* IAM 14596^{T} was completed. On the basis of this work, two chromosomes and one plasmid were obtained. It consisted of 2,785,698 bp of larger chromosome (Chr. 1), 1,098,310 bp of smaller chromosomes (Chr. 2) and 244,363 bp of plasmid with G+C contents of 43.23%, 42.39%, and 41.18%, respectively. This GC content estimated by the complete genome sequence also showed closer value with the experimentally determination by HPLC. However, this is the first time in obtaining the complete genome sequence of *V. halioticoli*, and the second complete genome in Halioticoli clade species followed by *V. breoganii* FF50.

In total 3,602 CDSs, 33 rRNAs and 30 tRNAs were predicted and covered more than 85% of the genome. The annotation rate is approximately 75%. I also identified, the replication origins of each chromosome by confirming presences of typical genes and unique sequences required for replication origins such as "*parB, parA, gidB, gidA, dnaA, recF* genes and DnaA box" and "*parB, parA, rctB* genes and DnaA box" on the Chr. 1 and Chr. 2, respectively. Fortunately, as a rep-3replication initiation protein gene with Dna box is found in the 3rd reconstructed genome, I decided that it could be functioned to be a plasmid. On the plasmid, as set of genes for conjugation and possible Type I-E CRISPR/Cas system set consisted of *cas3-cse12345-cas1-cas2* with CRISPR loci is found, which means the plasmid also could function to be a conjugative plasmid. A 21,277 bp identical region is also found in both Chr. 2 (position 516756-538033) and the plasmid (position 5028-26305), which may show the plasmid could be integrated in the *V. halioticoli* genome.

Among 3,602 CDSs on the *V. halioticoli* genome, 2100 CDSs were annotated in BlastKOALA corresponding to 58.3% (Fig. 2.4 and Table 2.2). Most of the CDSs encode products involved in Metabolism, Genetic Information, Environmental Information, Cellular Processes, Human Diseases, and Organismal Systems. In metabolic pathways, *V. halioticoli* possessed genes responsible for EM, PP and ED pathway in carbohydrate metabolism. Also, TCA cycle, glyoxylate cycle, and mixed acid fermentation pathways were also reconstructed. From CAZyme annotation, *V. halioticoli* possessed CDS classified into 32 GH, 26 GT, 23 PL, 7 CE and 2 AA in total (Fig. 2.1, and Table 2.3). Some of them possessed CBMs. The variety of transporters, secretion systems, and protein export systems were predicted. 140 types of transporter genes being capable of transport 35 substrates. Six two-component systems were found to respond to environmental stimuli (Fig. 2.6). In addition to this, the genomic sequence of *V. halioticoli* also clearly show the metabolic pathway of this prokaryotic organism, and provide important clues to understanding the metabolic and regulatory networks that link genes on the two chromosomes and plasmid. Finally, *V. halioticoli* clearly represents each gene they worked in utilization of mannitol, alginate, and the other components of marine biomass. Cloning and expression of these genes may help us verify the estimated pathways, and more clearly understand the alginate metabolism in the bacterium.

The gene mining revealed a total of 15 alginate lyase gene candidates in the whole genome data of *V. halioticoli*. Further motif/domain and localization analysis of these gene shows that at least 11 genes were more likely to function as alginate lyase (Table 3.1). At least six types of alginate lyases were contained in the genome of *V. halioticoli*; 1) extracellular secreting (or periplasmic) 35-40 kDa PL7 types (S2G1, S2G3, S2G4, S4G1 and S4G8), 2) a cytoplasmic 80 kDa PL17 type (S2G2),

3) a cytoplasmic 70 kDa dual domain PL7 (S4G2), 4) an extracellular 57 kDa dual domain alginate lyase (S4G3), 5) localization unknown 81 kDa dual domain PL17 type (S4G4 and S4G5), and 6) cytoplasmic 79 kDa PL15 type alginate lyases (S4G10) (Fig. 3.1).

More interestingly, some genes were likely to be arranged in an operatic structure (Fig. 2.2). The operatic structure involving alginate lyases have been reported in *Pseudomonas aeruginosa* (Gacesa, 1998), *Sphingomonas* sp. A1 (Hashimoto et al., 2004), *Zobellia galactanivorans* Dsij^T (Thomas et al., 2012), and *Vibrio splendidus* 12B1 (Wargacki et al., 2012). AlgL of *P. aeruginosa* is in the alginate biosynthetic cluster, and Aly of the *Sphingomonas* is in the super-channel gene cluster for alginate. Only marine flavobacteria (such as *Zobellia*) and proteobacteria (such as *Vibrio*) maintain a variety of alginate utilization system (AUS) gene clusters. The *Zobellia* case study found; 1) unveiled gene duplications in the common ancestor of the AUS-containing flavobacteria (see also Fig. 3.5), 2) several gene duplications are still visible but some are now cryptic, 3) independent LGT even shaping different AUS composition, and 4) different modular structure of alginate utilization system (AUS) operon between marine (*Bacteroidetes* and *Proteobacteria*) and terrestrial (*P. aeruginosa*, *Azotobacter vinelandii*, and *Sphingomonas*), are found (Thomas et al., 2012). Compared to *Zobellia* and *V. splendidus* AUS gene architecture, vibrio does not possess a TonB-dependent receptor (TBDR) and Zg2616-like sugar permease, but maintains outer membrane porins KdgMN and the sodium/solute symporter responsible for the uptake of oligoalginates (Wargacki et al., 2012), which means alginolytic *Vibrio* species may develop different system for the uptake of alginate oligosaccharides from marine bacteroidetes. So it is interesting to undertake

deeper research to elucidate not only AUS gene features but also the evolutionary aspects of these operons, ex. gene duplication, evolutionary history of genes on chromosome 1 and 2 and so on.

In chapter 3, I also attempted to express these genes using the *E. coli* expression system to understand whether each gene product is active or not. Using the *V. halioticoli* IAM 14596^{T} genomic DNA as a template, all 15 target genes were amplified using *pfu* DNA polymerase, and cloned into directional TOPO expression vector with the lumio technology (pET160/161). The recombinant plasmid was first transformed and amplified using TOP10 *E. coli*. The plasmid was purified, and then using the recombinant plasmid, BL21 (DE3) *E. coli* was transformed and used for the protein expression studies. Alginate lyase activity expressed by the transformant was assayed in two-steps; plate assay focuses on extracellular protein and cell fractionation followed by the thiobarbituric acid (TBA) method, which is used for the determination of intercellular expressed enzyme activities. In the result, 7 clones (S2G1, S2G3, S2G5, S4G1, S4G2, S4G3, and S4G8), were positive by the plate assay. Five of these clones (S2G1, S2G2, S4G1, S4G8, and S4G10) were positive for the alginate lyase activity using TBA medium in culture supernatant and/or cell free extract, and 2 (S2G2 and S4G10) of them were positive for intracellular expressed alginate lyases (Table 3.4). The optimum culture condition was obtained in a 36 hours-LB cultured medium in S2G1, S4G1 and S4G8. In the latest summary, gene S2G1, S2G3, S2G5, S4G1, S4G2, S4G3 and S4G8 are extracellular alginate lyase genes; S2G2 and S4G10 are intercellular alginate lyase genes. According to the comparison between the computer prediction results with experimental results, the accuracy of forecasting was very high.

Undoubtedly, the discovery of alginate lyase will promote the development of the clarifying alginate metabolic pathway of *V. halioticoli* IAM14596^{T}. In practical terms, the revealing to the public of the alginate metabolic pathway will be a great breakthrough in bioconversion from alginate to ethanol using this bacterium. Recombinant *V. halioticoli* may produce the high activity alginate lyase to increase production and reduce costs at the same time. Further clone analysis is being conducted to help us understand how multiple alginate lyase system works in *V. halioticoli*.

Substrate specificity was further studied. As far as I know, three types of block structure are present in alginate. And alginate lyases preferentially degrade particular block structures and the enzymes are typically described as either polyguluronate-specific or polymannuronate-specific. Substrate specificity of each gene product was determined on the plate containing not only sodium alginate but also polyG rich and polyM rich substrates (Fig. 3.4). S2G1/pET160, S4G1/pET160, S4G3/pET161, and S4G8/pET160/pET161 showed polyG specific activities (Fig. 3.3B). S2G3/pET161 and S4G2/pET160 might be polyG specific, but there was no polyM plate result currently. This experiment should be carried out repeatedly.

Further clone analysis could start from the analysis of gene specific sites. In chapter 3, genes S2G1 and S2G4 were similar to *alyVGI* gene sharing amino acid identities of 100% and 57%, respectively. However, results of activity test displayed a marked difference. A very high activity was detected from the S2G1 clone, but activity in the S2G4 has not been detected yet. The same and different nucleotide sequence between these two genes will give us one breach for analysis of the specific sites in gene sequences. This analysis could be combined with a 3D model to obtain a highly

possible specific gene site. Also, this gene site could be confirmed according to the activity test of the mutation.

In conclusion, the genome sequence of *V. halioticoli* was completely sequenced. Two chromosomes and one plasmid were obtained. According to this genome, 7 extracellular secreting alginate lyases S2G1, S2G3, S2G5, S4G1, S4G2, S4G3 and S4G8, and 2 oligoalginate lyases S2G2 and S4G10 were discovered from the bacterium *V. halioticoli* IAM 14596^{T}. And S2G1, S4G1, S4G3, and S4G8 showed polyG specific activities in this experiment. Analysis of complete genome detailedly described alginate metabolic pathway in *V. halioticoli*. Also the completely genome sequence provides a new starting point to help us more clearly understand the organism's phenotype and genotype characteristics in the bacterium, and push forward the biofuel production in biology research.

References

Al-saari N, Gao F, Rohul AKM, Sato K, Sato K, Mino S, Suda W, Oshima K, Hattori M, Ohkuma M, Meirelles PM, Thompson FL, Thompson C, Filho GMA, Gomez-Gil B, Sawabe T, Sawabe T (2015) Advanced microbial taxonomy combined with genome-based approaches reveals that *Vibrio astriarenae* sp. nov., an agarolytic marine bacterium, forms a new clade in *Vibrionaceae*. PLoS ONE 10: e0136279.

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403-10.

Amaral GRS, Dias GM, Wellington-Oguri M, Chimetto L, Campe ão ME, Thompson FL, Thompson CC (2014) Genotype to phenotype: identification of diagnostic vibrio phenotypes using whole genome sequences. Int J Syst Evol Microbiol 64:357-365.

Amin AKMR, Tanaka M, Al-Saari N, Feng G, Mino S, Ogura Y, Hayashi T, Meirelles PM, Thompson FL, Gomez-Gil B, Sawabe T, Sawabe T (2017) *Thaumasiovibrio occultus* gen. nov. sp. nov. and *Thaumasiovibrio subtropicus* sp. nov. within the family *Vibrionaceae*, isolated from coral reef seawater off Ishigaki Island, Japan. Syst Appl Microbiol 40:290-296.

Atsumi S, Hanai T and Liao JC (2008) Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. Nature 451: 86-90.

Atsumi S, Wu TY, Eckl EM, Hawkins SD, Buelter T and Liao JC (2010) Engineering the isobutanol biosynthetic pathway in *Escherichia coli* by comparison of three aldehyde reductase/alcohol dehydrogenase genes. Appl Microbiol Biotechnol 85:651-657.

Austin B and Zhang X-H (2006) *Vibrio harveyi*: a significant pathogen of marine vertebrates and invertebrates. Lett Appl Microbiol 43:119-124.

Aziz RK, Bartels D, Best A, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O (2008) The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9:75.

Barton LL (2005) Structural and functional relationships in prokaryotes. Springer Science USA pp:3-43.

Baumann P and Schubert RHW (1984) *Vibrionaceae*. In: Krieg NR, Holt JG (Eds), Bergey's Manual of Systematic Bacteriology, Vol. 1. Williams & Wilkin, Baltimore pp:516-550.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J and Wheeler DL (2008) GenBank. Nucleic

Acids Res 36:D25-D30.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J and Sayers EW (2011) GenBank. Nucleic Acids Res 39:D32-37.

Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Ostell J, Pruitt KD, Sayers EW (2017) GenBank. Nucleic Acids Res 46:D41-47.

Blattner FR, Plunkett G 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y (1997) The complete genome sequence of *Escherichia coli* K-12. Science 277:1453-62.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. BMC Bioinformatics 10:421.

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V and Henrissat B (2009) The carbohydrate-active enZymes database (CAZy): an expert resource for glycogenomics. Nucleic Acids Res 37:D233-238.

Dien BS, Cotta MA and Jeffries TW (2003) Bacteria engineered for fuel ethanol production: current status. Appl Microbiol Biotechnol 63:258-266.

Duigou S, Knudsen KG, Skovgaard O, Egan ES, Løbner-Olesen A, Waldor MK (2006) Independent control of replication initiation of the two *Vibrio cholerae* chromosomes by DnaA and RctB. J Bacteriol 188:6419-24.

Dunn AK, Millikan DS, Adin DM, Bose JL, Stabb EV (2006) New *rfp*- and pES213-derived tools for analyzing symbiotic *Vibrio fischeri* reveal patterns of infection and *lux* expression in situ. Appl Environ Microbiol 72:802-10.

Dürre P and Richard T (2011) Microbial energy conversion revisited. Cur Opin Biotech 22:309-311.

Ezaki T, Hashimoto Y, Takeuchi N, Yamamoto H, Liu S-L, Miura H, Matsui K, Yabuuchi E (1988) Simple genetic method to identify viridans group streptococci by colorimetric dot hybridization and fluorometric hybridization in microdilution wells. J Clin Microbiol 26:1708-1713.

Ezaki T, Hashimoto Y and Yabuuchi E (1989) Fluorometric deoxyribonucleic acid-deoxyribonucleic acid hybridization in microdilution wells as an alternative to membrane filter hybridization in which radioisotopes are used to determine genetic relatedness among bacterial strains. Int J Syst Bacteriol 39:224-229.

Farmer III JJ, Janda JM, Brenner FW, Cameron DN and Birkhead KM (2005) Genus I. *Vibrio Pacini* 1854, 411AL. In Bergey's Manual of Systematic Bacteriology, 2nd ed. Vol. 2

pp:494-546.

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J and Bateman A (2016) The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 44:D279-285.

Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, Bult CJ, Kerlavage AR, Sutton G, Kelley JM, Fritchman RD, Weidman JF, Small KV, Sandusky M, Fuhrmann J, Nguyen D, Utterback TR, Saudek DM, Phillips CA, Merrick JM, Tomb JF, Dougherty BA, Bott KF, Hu PC, Lucier TS, Peterson SN, Smith HO, Hutchison CA 3rd, Venter JC (1995) The minimal gene complement of *Mycoplasma genitalium*. Science 270:397-403.

Gacesa P (1988) Alginates. Carbohydr Polym 8:161-182.

Gacesa P (1992) Enzymic degradation of alginate. Int J Biochem 24:545-552.

Gacesa P (1998) Bacterial alginate biosynthesis-recent progress and future prospects. Microbiology 144:1133-43.

Garron M-L and Cygler M (2010) Structural and mechanistic classification of uronic acid-containing polysaccharide lyases. Glycobiology, 20: 1547-1573.

Geddes CC, Nieves IU and Ingram LO (2011) Advances in ethanol production. Cur Opin Biotech 22:312-319.

Gomez-Gil B (2003) *Vibrio pacinii* sp. nov., from cultured aquatic organisms. Int J Syst Evol Microbiol 53:1569-1573.

Gomez-Gil B, Thompson CC, Matsumura Y, Sawabe T, Iida T, Christen R, Thompson F, Sawabe T (2014) The Family *Vibrionaceae*. In: The Prokaryotes, 4th edition, volume 9. Springer-Verlag, Berlin Heidelberg. pp:659-747.

Hashimoto W, Yamasaki M, Itoh T, Momma K, Mikami B and Murata K (2004) Super-channel in bacteria: structural and functional aspects of a novel biosystem for the import and depolymerization of macromolecules. J Biosci Bioeng 98:399-413.

Hehemann J-H, Correc G, Barbeyron T, Helbert W, Czjzek M and Michel G (2010) Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. Nature 464:908-914.

Hidaka T and Sakai M (1968) Comparative observation of inorganic salt requirement of the marine and terrestrial bacteria. Bull Misaki Mar Biol Inst Kyoto Univ 12:125-149.

Holt JG, Krieg NR, Sneath PHA, Staley JT, Williams ST (1994) Facultatively anaerobic gram-negative rods. Subgroup 2: Family *Vibrionaceae*. In: Bergey's manual of determinative microbiology, 9th edition. Williams and Wilkins, Baltimore. pp:190-253.

Ilene M (2007) GenBank: The Nucleotide Sequence Database. The NCBI Handbook (Internet).

Ingram LO, Conway T, Clark DP, Sewll GW and Preston JF (1987) Genetic engineering of ethanol production in *Escherichia coli*. Appl Environ Microbiol 53:2420-2425.

Jain S and Ohman DE (2005) Role of an alginate lyase for alginate transport in mucoid *Pseudomonas aeruginosa*. Infect Immun, 73:6429-6436.

Jin ZH, Liu Y, Zhang J, Gong QL, Cui JJ, liu T (2009) Current situation and development trend of Chinese seaweed farming. Transactions of Oceanology and Limnology 1:1003-6482

Johnson TJ, Bielak EM, Fortini D, Hansen LH, Hasman H, Debroy C, Nolan LK, Carattoli A (2012) Expansion of the IncX plasmid family for improved identification and typing of novel plasmids in drug-resistant *Enterobacteriaceae*. Plasmid. 68:43-50.

Kanehisa M, Sato Y, Morishima K (2017) BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. J Mol Biol 428:726-731.

Kamada M, Hase S, Sato K, Toyoda A, Fujiyama A, Sakakibara Y (2014) Whole genome complete resequencing of *Bacillus subtilis* natto by combining long reads with high-quality short reads. PLoS One 9:e109999.

Kelley LA and Sternberg MJ (2009) Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc 4:363-371.

Kim HT, Ko H-J, Kim N, Kim D, Lee D, Choi I-G, Woo HC, Kim MD, Kim KH (2012a) Characterization of a recombinant endo-type alginate lyase (Alg7D) from *Saccharophagus degradans*. Biotechnol Lett 34:1087-1092.

Kim HT, Chung JH, Wang D, Lee J, Woo HC, Choi IG, Kim KH (2012b) Depolymerization of alginate into a monomeric sugar acid using *Alg17C*, an exo-oligoalginate lyase cloned from *Saccharophagus degradans* 2-40. Appl Microbiol Biotechnol 93:2233-2239.

Kita-Tsukamoto K, Oyaizu H, Nanba K, Simidu U (1993) Phylogenetic relationship of marine bacteria mainly members of the family *Vibrionaceae*, determined on the basis of 16S rRNA sequences. Int J Syst Bacteriol 43:8-19.

Konstantinidis KT and Tiedje JM (2005) Towards a genome-based taxonomy for prokaryotes. J Bacteriol 18:6258-6264.

Koonin EV, Makarova KS, Zhang F (2017) Diversity, classification and evolution of CRISPR-Cas systems. Curr Opin Microbiol 37:67-78.

Land M, Hauser L, Jun SR, Nookaew I, Leuze MR, Ahn TH, Karpinets T, Lund O, Kora G, Wassenaar T, Poudel S, Ussery DW (2015) Insights from 20 years of bacterial genome sequencing. Funct Integr Genomics 15:141-161.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinform 23:2947-2948.

Lee HS, Vermaas WFJ and Rittmann BE (2010) Biological hydrogen production: prospects and challenges. Trends Biotechnol 28:262-271.

Leifson E (1963) Determination of carbohydrate metabolism of marine bacteria. J Bacteriol 82:33-36.

Macián MC, Garay E, Grimont PAD, Pujalte MJ (2004) *Vibrio ponticus* sp. nov., a neigbour of *V. fluvialis-V. furnissii* clade, isolated from gilthead sea bream, mussels and seawater. Syst Appl Microbiol 27:535-540.

Marmur J (1961) A procedure for the isolation of deoxyribonucleic acid from microorganisms. J. Mol. Biol., 3: 208-218.

McHugh DJ (2003) A guide to the seaweed industy. FAO Fisheries technical pp:441.

Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M (2013) Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 14:60.

Meinkoth J and Wahl G (1984) Hybridization of nucleic acids immobilized on solid supports. Anal Biochem 138:267-284.

Miyamoto M, Motooka D, Gotoh K, Imai T, Yoshitake K, Goto N, Iida T, Yasunaga T, Horii T, Arakawa K, Kasahara M, Nakamura S (2014) Performance comparison of second- and third-generation sequencers using a bacterial genome with two chromosomes. BMC Genomics 15:699.

Nelder JA and Wedderburn RWM (1972) Generalized Linear Models. J R Stat Soc Ser A 135:370-384.

Noguchi H, Taniquchi T and Itoh T (2008) MetaGeneAnnotator: detecting species-specific patterns

of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. DNA Res 15:387-396.

Ochiai A, Yamasaki M, Mikami B, Hashimoto W and Murata K (2006) Crystallization and preliminary X-ray analysis of an exotype alginate lyase Atu3025 from *Agrobacterium tumefaciens* strain C58, a member of polysaccharide lyase family 15. Acta Cryst 62:486-488.

Ochiai M, Yamamoto A, Naito S, Maeyama J, Masumi A, Hamaguchi I, Horiuchi Y, Yamaguchi K (2010) Applicability of bacterial endotoxins test to various blood products by the use of endotoxin-specific lysates. Biologicals 38:629-36.

Ogura K, Yamasaki M, Mikami B, Hashimoto W and Murata K (2008) Substrate recognition by family 7 alginate lyase from *Sphingomonas* sp. A1. J Mol Biol 380:373-385.

Ohta K, Beall DS, Mejia JP, Shanmugam KT and Ingram LO (1991) Genetic improvement of *Escherichia coli* for ethanol production: chromosomal integration of *Zymomonas mobilis* genes encoding pyruvate decarboxylase and alcohol dehydrogenase II. Appl Environ Microbiol 57:893-900.

Okada K Iida T Tsukamoto KK and Honda T (2005) Vibrios commonly possess two chromosomes. J Bacteriol 187:752-757.

Oppenheimer CH and ZoBell CE (1952) The growth and viability of sixty-three species of marine bacteria as influenced by hydrostatic pressure. J Mar Res 11:10-18.

Ostle AG and Holt JG (1982) Nile blue A as fluorescent stain for poly-beta-hydroxybutyrate. Appl Environ Microbiol 44:238-241.

Park D, Jagtap S, Nair SK (2014) Structure of a PL17 family alginate lyase demonstrates functional similarities among exotype depolymerases. J Biol Chem 289:8645-55.

Preiss J and Ashwell G (1962a) Alginic acid metabolism in bacteria. I. Enzymatic formation of unsaturated oligosaccharides and 4-deoxy-L-erythro-5-hexoseulose uronic acid. J Biol Chem 237:309-316.

Preiss J and Ashwell G (1962b) Alginic acid metabolism in bacteria. II. Enzymatic reduction of 4-deoxy-L-erythro-5-hexoseulose uronic acid to 2-keto-3deoxy-D-gluconic acid. J Biol Chem 237:317-321.

Ruby EG (1996) Lessons from a cooperative, bacterial-animal association: the *Vibrio fischeri*-Euprymna scolopes light organ symbiosis. Annu Rev Microbiol 50:591-624.

Ruby EG, Urbanowski M, Campbell J, Dunn A, Faini M, Gunsalus R, Lostroh P, Lupp C, McCann J, Millikan D, Schaefer A, Stabb E, Stevens A, Visick K, Whistler C, Greenberg EP (2005) Complete genome sequence of *Vibrio fischeri*: a symbiotic bacterium with pathogenic congeners. Proc Natl Acad Sci USA 102:3004-3009.

Saitou N and Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406-425.

Sambrook J and Russell DW (2001) Molecular cloning, third edition. Cold Spring Harbor Laboratory Press. New York. USA.

Sawabe T, Oda Y, Shiomi Y and Ezura Y (1995) Alginate degradation by bacteria isolated from the gut of sea urchins and abalones. Microb Ecol 30:192-202.

Sawabe T, Takahashi H, Ezura Y and Gacesa P (1997) Cloning, sequence analysis and expression of *Pseudoalteromonas elyakovii* IAM 14594 gene (*alyPEEC*) encoding the extracellular alginate lyase. Carbohydr Res 335:11-21.

Sawabe T, Sugimura I, Ohtsuka M, Nakano K, Tajima K, Ezura Y, Christen R (1998) *Vibrio halioticoli* sp. nov., a non-motile alginolytic marine bacterium isolated from the gut of abalone *Haliotis discus hannai*. Int J Syst Bacteriol 48:573-580.

Sawabe T, Setoguchi N, Inoue S, Tanaka R, Ootsubo M, Yoshimizu M, Ezura Y (2003) Acetic acid production of *vibrio halioticoli* from alginate: a possible role for establishment of abalone-*V. halioticoli* association. Aquaculture, 219:671-679.

Sawabe T, Hayashi K, Moriwaki J, Fukui Y, Thompson FL, Swings J, Christen R (2004) *Vibrio neonatus* sp. nov. and *Vibrio ezurae* sp. nov. isolated from the gut of Japanese abalones. Syst Appl Microbiol 27:527-34.

Sawabe T, Fukui Y, Stabb EV (2006) Simple conjugation and outgrowth procedures for tagging vibrios with GFP, and factors affecting the stable expression of the *gfp* tag. Lett Appl Microbiol 43:514-22.

Sawabe, T (2006) The mutual partnership between *Vibrio halioticoli* and abalones. In The Biology of Vibrios, F. L. Thompson, B. Austin, and J. Swings (eds). ASM Press Wasington DC pp.219-230.

Sawabe T, Fujimura Y, Niwa K and Aono H (2007a). *Vibrio comitans* sp. nov., *Vibrio rarus* sp. nov. and *Vibrio inusitatus* sp. nov., from the gut of the abalones *Haliotis discus discus*, *H. gigantea*, *H. madaka* and *H. rufescens*. International Journal of Systematic and Evolutionary Microbiology 57:916-922.

Sawabe T, Kita-Tsukamoto K, Thompson FL (2007b) Inferring the evolutionary history of vibrios by means of multilocus sequence analysis. J Bacteriol 189:7932-7936.

Sawabe T, Ogura Y, Matsumura Y, Feng G, Amin AKMR, Mino S, Nakagawa S, Sawabe T, Kumar R, Fukui Y, Satomi M, Matsushima R, Thompson FL, Gomez-Gil B, Christen R, Maruyama F, Kurokawa K, Hayashi T (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.

Shimizu T, Tanizawa Y, Mochizuki T, Nagasaki H, Yoshioka T, Toyoda A, Fujiyama A, Kaminuma E, Nakamura Y (2017) Draft Sequencing of the Heterozygous Diploid Genome of Satsuma (*Citrus unshiu* Marc.) Using a Hybrid Assembly Approach. Front Genet 8:180.

Sim S-J, Baik KS, Park SC, Choe HN, Seong CN, Shin T-S, Woo HC, Cho JY, Kim D (2012) Characterization of alginate lyase gene using a metagenomic library constructed from the gut microflora of abalone. J Ind Microbiol Biotechnol 39:585-593.

Stephanopoulos G (2007) Challenges in engineering microbes for biofuels production. Science 315:801-804.

Stokstad E (2012) Engineered superbugs boost hopes of turning seaweed into fuel. Science 335:273.

Strasser BJ (2011) The experimenter's museum: GenBank, natural history, and the moral economies of biomedicine. Isis 102:60-96.

Su H, Blain F, Musil RA, Zimmermann JJF, Gu K and Bennett DC (1996) Isolation and expression in *Escherichia coli* of *hepB* and *hepC*, genes coding for the glycosaminoglycan-degrading enzymes heparinase II and heparinase III, respectively, from Flavobacterium heparinum. Appl Environ Microbiol 62:2723-2734.

Sugimura I, Sawabe T, Ezura Y (2000) Cloning and sequence analysis of *Vibrio halioticoli* gene encoding three types of polyguluronate lyase. Mar Biotechnol 2:65-73.

Swofford D (1992) PAUP: phylogenetic analysis using parsimony, version 3.0s. Illinois Natural History Survey, Champaign, IL, USA.

Takase R, Ochiai A, Mikami B, Hashimoto W and Murata K (2010) Molecular identification of unsaturated uronate reductase prerequisite for alginate metabolism in *Sphingomonas* sp. A1. Biochim Biophys Acta 1804:1925-1936.

Takeda H, Yoneyama F, Kawai S, Hashimoto W and Murata K (2011) Bioethanol production from

marine biomass alginate by metabolically engineered bacteria. Energy Environ Sci 4:2575-2581.

Tamaoka J and Komagata K (1984) Determination of DNA base composition by reversed-phase high-performance liquid chromatography. FEMS Microbiol Lett 25:125-128.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30:2725-2729.

Tang X (2003) Status and marketing analysis of seaweed chemical industry in China. Fisheries Sci Technol 5:1-4

Thomas F, Barbeyron T, Tonon T, Génicot S, Czjzek M, Michel G (2012) Characterization of the first alginolytic operons in a marine bacterium: from their emergence in marine *Flavobacteriia* to their independent transfers to marine *Proteobacteria* and human gut *Bacteroides*. Environ Microbiol 14:2379-2394.

Thompson CC, Vicente ACP, Souza RC, Vasconcelos ATR, Vesth T, Alves Jr N, Ussery DW, Iida T, Thompson FL (2009) Genomic taxonomy of Vibrios. BMC Evol Biol 9:258.

Thompson CC, Chimetto L, Edwards RA, Swings J, Stackebrandt E, Thompson FL (2013) Microbial genomic taxonomy. BMC Genomics 14:913.

Thompson CC, Amaral GR, Campeão M, Edwards RA, Polz MF, Dutilh BE, Ussery DW, Sawabe T, Swings J, Thompson FL (2015) Microbial taxonomy in the post-genomic era: rebuilding from scratch? Arch Microbiol 197:359-70.

Thompson FL, Hoste B, Vandemeulebroecke K, Swings J (2001) Genomic diversity amongst Vibrio isolates from different sources determined by fluorescent amplified fragment length polymorphism. System Appl Microbiol 24:520-538

Thompson JR and Polz MF (2006) Dynamics of vibrio population and their role in environmental nutrient cycling. In: Thompson FL, Austin B, Swings J, editors. The Biology of Vibrios. ASM Press: Washington, DC. pp:190-203.

Timothy J, Johnson a, Eliza MB, Daniela F, Lars Hestbjerg H, Henrik H, Chitrita D, Lisa KN, Alessandra C (2012) Expansion of the IncX plasmid family for improved identification and typing of novel plasmids in drug-resistant *Enterobacteriaceae*. Plasmid 68:43-50

Trinh CT, Unrean P and Srienc F (2008) Minimal *Escherichia coli* cell for the most efficient production of ethanol from hexoses and pentoses. Appl Environ Microbiol 74:3634-3643.

Val ME, Soler-Bistué A, Bland MJ, Mazel D (2014) Management of multipartite genomes: the

Vibrio cholerae model. Curr Opin Microbiol 22:120-6.

Wargacki AJ, Leonard E, Win MN, Regitsky DD, Santos CNS, Kim PB, Cooper SR, Raisner RM, Herman A, Sivitz AB, Lakshmanaswamy A, Kashiyama Y, Baker D and Yoshikuni Y (2012) An engineered microbial platform for direct biofuel production from brown macroalgae. Science 335:308-313.

West M, Burdash NM, Freimuth F (1977) Simplified silver plating stain for flagella. J Clin Microbiol 6:414-419.

Wong TY, Preston LA and Schiller NL (2000) Alginate lyase: review of major sources and enzyme characteristics, structure-function analysis, biological roles, and applications. Annu Rev Microbiol 54:289-340.

Zhu B and Yin H (2015) Alginate lyase: Review of major sources and classification, properties, structure-function analysis and applications. Bioengineered 6:125-31.

Yamasaki M, Moriwaki S, Miyake O, Hashimoto W, Murata K and Mikami B (2004) Structure and function of a hypothetical *Pseudomonas aeruginosa* protein PA1167 classified into family PL-7. J Biol Chem 279:31863-31872.

Yin Y, Mao X, Yang JC, Chen X, Mao F and Xu Y (2012) dbCAN: a web resource for automated carbohydrate-active enzyme annotation, Nucleic Acids Res 40 (Web Server issue):W445-451

Yu NY, Laird MR, Spencer C and Brinkman FS (2010) PSORTdb--an expanded, auto-updated, user-friendly protein subcellular localization database for Bacteria and Archaea. Nucleic Acids Res 39:241-244.

Acknowledgements

I would like to express my heartfelt respect and deepest sincere gratitude to Prof. **Dr. Sawabe**, and **Dr. Mino** Laboratory of Microbiology, Faculty of Fisheries, Hokkaido University, for his excellent academic advise, valuable suggestions, teaching of scientific technique, support, and understanding throughout the course study for eight years.

My special appreciation is expressed to Assistant Prof. **Dr. Ojima** for their critical review and constructive criticism to improve this report.

And I also would like to express my sincere thank to **Dr. Y. Ogura and Prof. Dr. Hayashi**, at the Kyusyu University, and **Prof. Dr.**, **Toyoda,** at NIG for genome sequence and assemble.