



이학석사학위논문

Single Molecule Real Time 기술을 이용한 *Celeribacter marinus* 균주 IMCC12053 의 메틸롬 연구 및 *Flavobacteriales* 목(目) UJ101 균주의 비교유전체 연구



2017 년 2 월

부경대학교 과학기술융합전문대학원

해양바이오융합과학전공

양정안

이학석사학위논문

Single Molecule Real Time 기술을 이용한 *Celeribacter marinus* 균주 IMCC12053 의 메틸롬 연구 및 *Flavobacteriales* 목(目) UJ101 균주의 비교유전체 연구

지도교수 오 현 명

이 논문을 이학석사 학위논문으로 제출함

2017년 2월

부경대학교 과학기술융합전문대학원

해양바이오융합과학전공

양정안

양정안의 이학석사 학위논문을 인준함



목 차

Abstract

Chapter 1. <i>Celeribacter marinus</i> IMCC12053의 Methylome 분석	Chapter	1.	Celeribacter	marinus	IMCC12053 의	Methylome	분석	
--	---------	----	--------------	---------	-------------	-----------	----	--

1.1. 서 론1
1.2. 유전자 서열 분석 데이터5
1.3. SMRT 기술로 <i>C.marinus</i> 의 methylome 인 N ⁶ -methyladenosine(m ⁶ A) 및 N ⁴ -
methylcytosine(m ⁴ C)의 서열분석9
1.4. 다른 <i>Celeribacter</i> 유전자와 비교 분석16
1.5. Nucleotide Accession Number

Chapter 2. Xanthid Crab에 공생하는 *Flavobacteriales bacterium* 균주 UJ101의 비교 유전체 분석

2.1. 서 론
2.2. 유전자 서열 분석 데이터
2.3. 비교 유전체 분석 데이터
2.3.1. Pan/Core-genome 및 게놈을 통한 기능 분석
2.3.2. 대사과정 비교 및 분석45

참고문헌

Methylome study of *Celeribacter marinus* strain IMCC12053 and Comparative Genomics of *Flavobacteriales* strain UJ101 using Single Molecule Real Time Technology

Jhung Ahn Yang

Specialized Graduate School Science & Technology Convergence

Pukyong National University

Abstract

Single Molecule Real Time (SMRT) technology is one of the Next Generation Sequencing (NGS) tool using Pacific Bioscience (PacBio) RS II Sequencing. This sequencing technology has the advantage of possible methylation analysis. In order to analyze 2 Marine Microorganism (*Celeribacter marinus* strain IMCC12053, *Flavobacteriales bacterium* strain UJ101), we used PacBio Sequencing analysis.

First, *Celeribacter marinus* IMCC12053 was isolated from coastal seawater from Yellow Sea of Korea, it was used as the host bacterium for bacteriophage P12053L. We report the complete genome sequence of strain IMCC12053 for further study of the marine bacteriophage P12053L. Genome sequencing data is 3,096,705 base pairs in length by single circular chromosome and the GC content is 56.24%. It contains 3,155 ORFs with 45 tRNAs and 6 rRNAs genes. N6-methyladenosine patterns were also investigated for 32 unmethylated genes and intergenic regions that covered many regulators and phage genes as well as ribosomal RNA genes and tRNA genes. Cryptic N4-methylcytosine pattern was investigated to speculate GpC methylase activity throughout the genome. Comparative genomics with other *Celeribacter* genomes were carried out for polyaromatic hydrocarbon degradation, but there were no aromatic ring oxygenases

in IMCC12053 when compared to Celeribacter indicus P73.

Second, *Flavobacteriales bacterium* strain UJ101was isolated from a xanthid crab shell collected from East Sea of Korea. We report the complete genome sequence of strain UJ101 for the study of metabolic interaction between UJ101 and its host organism. Genome sequencing data is 3,074,209 base pairs in length by single circular chromosome and the GC content was 30.74%. The genome of UJ101 contains 2,698 ORFs with 46 tRNAs and 9 rRNAs genes. According to the annotated list of genes Embden– Meyerhof and pentose phosphate pathway is well conserved, but key enzymes of Entner-Douddoroff pathway were impaired. TCA and glyoxylate cycle were conserved while carbon fixation and one carbon metabolism were mostly lacking except formaldehyde dehydrogenase. UJ101 encodes degradation enzyme including 8 glycosyl transferases, 3 amylases, and 8 peptidases. Biosynthetic enzymes for lysine, tryptophan, phenylalanine, and tyrosine were also impaired. Alcohol and/or organic acid fermentation could not be expected. Strain UJ101 was compared with bacterial genomes isolated from other marine animals (3 strains from invertebrate and 5 strains from fishes). Other related genomes from the same genera were included although they were reportedly isolated from seawater and marine sediments. Chapter 1. Celeribacter Marinus IMCC12053의 Methylome

분석

1.1. 서 론

Celeribacter 속은 해안 해수에서 분리되었으며(Ivanova et al., 2010) Alphaproteobacteria 강, Rhodobacteriaceae 과에 속하는 구성원으로, 바다에서 발견 된 주요 계통 군집 중 하나이다.

첫 번째로 밝혀진 종으로 *Celeribacter neptunius* (Ivanova et al., 2010)로 호주와 뉴질랜드 사이에 있는 태즈먼 바다의 해안에서 분리되었고, 한국 동해 및 황해에서 *Celeribacter baekdonensis* (Lee et al., 2012), *Celeribacter marinus* (Baek et al., 2014) (Table 1) 를 분리하였다.

해안 표면의 해수로부터 단독으로 분리된 상기 균주 와는 다르게 Celeribacter indicus P73T(Lai et al., 2014)와 C. baekdonensis B30은 인도양 심해 퇴적물에서 배양되었다(Cao et al., 2015).

알려진 *Celeribacter* 종 중에서 *C. marinus* IMCC12053 T는 해안의 바다 물에서 분리하였다(Baek et al., 2014) (Table 1). 이 균주는 해양 박테리오파지 P12053L의 호스트 박테리아로 이용되며 (Kang et al., 2012), 호기성이며 화학 종속 영양생물이다. 본 연구에서는 *C. marinus* strain IMCC12053의 전체 유전자 서열을 SMRT기술을 이용하여 완전하게 가까운 유전체(Table 1)를 보고하고, methylation 패턴을 분석하였으며 또한, 동일한 *Celeribacter* 속의 미생물들의 유전체 정보들과 비교하였다. Table 1. *Celeribacter marinus* IMCC12053^T genome assembly and general features.

Genome assembly data					
Item	Description				
Assembly method	SMRT pipe HGAP, SMRTpipe Celera Assembler, SMRTpipe Quiver				
Genome coverage	1063 ×				
Sequencing technology	PacBioRS2				
MIGS data					
Item	Description				
investigation_type	bacteria_archaea				
project_name	Genome sequencing of Celeribacte marinus				
experimental_factor	NA				
lat_lon	37.449722 126.599722				
depth	NA				
alt_elev	NA				
geo_loc_name	South Korea				
collection_date	2009.12				
material	Marine waters				
biome	Ocean				
feature	Water				
num_replicons	NA				

ref_biomaterial	PMID 24425746
pathogenicity	NA
biotic_relationship	free-living
trophic_level	chemoheterotroph
rel_to_oxygen	aerobe
isol_growth_condt	PMID 24425746
seq_math	PacBio RS II sequencing
assembly_name	NA
finishing_strategy	complete; 1063x coverage; 1 contig
annot_source	NA
assembly	SMRT
env_package	missing

1.2. 유전자 서열 분석 데이터

균주 IMCC12053^T의 전체 유전자 서열은 Pacific Biosciences SMRT tool (DNA Link, Inc., Seoul, Korea) (Table 1)를 이용하여 조립되었다: ORFs (Open Reading Frames)는 Glimmer 3 program, 각 서열 기능의 주석은 Blast2Go (SMRT tools pipeline)로 예측되었다. BLASTP는 ORFs를 확인하기 위해 사용되었고, rRNA, tRNA는 RNAmmer(Lagesen et al., 2007)와 tRNAscan-SE(Lowe and Eddy, 1997)를 사용하여 예측되었다.

대사 경로 해석을 위해, 모든 단백질 서열은 National Center for Biotechnology Information (NCBI) 데이터베이스를 기반으로 하여 확인하였으며, RAST 서버를 사용하여 분석 하였다(Aziz et al., 2008). GeneBank 제출 확인 된 데이터는 유전자의 시각화를 위해 CLGenomics 프로그램 형식(ChunLab, Inc., Seoul, Korea)으로 변환하였다. CLGenomics에서 보낸 유전자 테이블은 비교 유전자 형식이 생성되었고, GeneBank에 등록하기 위해 사용하였고, 시각화 된 파일 IMCC12053의 전체 유전자 정보 파일을 이용하여, 염색체 지도 정보를 생성하였다(Figure 1).

균주 IMCC12053T의 유전자(Table 2)는 56.24% 의 GC 함량, 길이는
3,095,705 염기 쌍의 단일 원형 염색체로 구성되었다. 유전자는 두
오페론으로 구성되고, 3,155개의 ORF, 45개의 tRNA, 6개의 rRNA가
포함되어 있다.

유전자의 크기를 비교했을 때, *C. indicus* P73와 *C. baekdonensis* B30가 각 4,969,338 및 4,333,597 염기 쌍인 반면 *C. marinus*는 3,096,705 염기 쌍인 것으로 분석되었다.



Figure 1. Circular map of the chromosome of strain IMCC12053.

From outside to the center: genome size (black line), RNA, forward strand (colored by COG categories), reverse strand (colored by COG categories), GC ratio (brown), GC skew (green and red)

Table 2. Celeribacter marinus IMCC12053 genome summary.

Genome size (bp)	3,096,705
DNA coding region (bp)	2,759,002
DNA G + C content	56.24%
Total genes	3,206
tRNA genes	45
rRNA genes	6
Protein coding genes	3,155
Gene with predicted function only	249

Attribute Value

1.3. SMRT 기술로 C.marinus의 methylome인 N6-methyladenosine (m6A) 및 N4-methylcytosine (m4C)의 서열분석

Alphaproteobacterial ccrM은 5'-GANTC-3'위치의 Adenosine을 메틸화 시키고, 아마도 박테리아 세포 주기의 조절 자로서 중요한 역할을 할 것으로 간주 된다(Kozdon et al., 2013).

IMCC12053의 유전자의 이러한 메틸화 패턴은 SMRT방법에 의해 예측되었고 관찰 되었다. (methylome listed in Supplymentary table 1). GANTC 위치 (N6-methyladenosine, m6A)의 메틸화 상태뿐만 아니라 N4-methylcytosine(m4C)는 코딩 구역 및 유전자간 위치가 정리되어 있다(Table 3). DNA methyltransferase (N6_N4_Mtase; PF01555.14 from pfam.sanger.org)는 PF01555에 대한 E-value 값 순서(Table 4)로 IMCC12053_0024, IMCC12053_0065, IMCC12053_3116로 4개의 ORF가 포함되어 있는 것으로 밝혀졌다.

IMCC12053_1885는 긴 367개의 아미노산이며, Type III restriction modification system의 메틸화 소단위 (EC 2.1.1.72)다. IMCC12053_2040 (411 아미노산), IMCC12053_0065 (411아미노산) 및 IMCC12053_3116 (421아미노산)은 DNA modification methylase로, 자신의 N-terminals은 *ParBc* (ParB-like nuclease domain)로 세포 분열 및 세포주기 조절 (Table 4) 및 염색체 (플라스미드) 파티셔닝 기능을 수행할 수 있다. N6-adenomsine과 N4-cytosine 메틸화 모두 외향고리 변형(exocyclic modification) (Loenen and Raleigh, 2014)을 통해 발생하기 때문에, IMCC12053_1885도 DNA methylase로 예상할 수 있다.

3206개 메틸화 유전자 중 784개 유전자는 메틸화 되지 않은 패턴으로 관찰되었다(Table 3). 그러나, 예상된 GANTC 서열과 32개 위치는 두 가닥 (Table 5)에 메틸화 되지 않은 채로 남았고, 이러한 보호 패턴은 세포주기 또는 전사의 억제와 같은 일부 조절 기능에 관여할 수 있다.

Table 3. IMCC12053 methylome. GANTC (m6A) and m4C methylation patterns were detected by SMRT, which was summarized for 3206 genes and 1334 intergenic regions.

Methylation	m6A	m4C	m6A & m4C	null	Total	
Genes	383	1091	948	784	3206	
Intergenic regions	745	589	-	-	1334	

Table 4. DNA methyltransferase (N6_N4_Mtase) found among IMCC12053 genome. Hmmsearch (v3.1b1) was used for searching IMCC12053 ORF's. Full sequence search results are shown(PFAM accession :bPF01555.14, E-value <1.0x10-6). Query names are according to pfam description.

Locus/CDS	Query name	E-value	Score	bias	Pfam-A matches
IMCC12053_1885	N6_N4_Mtase	1.4x10 ⁻⁵⁸	196.0	0.0	N6_N4_Mtase
IMCC12053_2040	N6_N4_Mtase	4 x10 ⁻³⁶	122.6	0.0	ParBc N6_N4_Mtase
IMCC12053_0065	N6_N4_Mtase	1.5 x10 ⁻³⁴	117.4	0.0	ParBc N6_N4_Mtase
IMCC12053_3116	N6_N4_Mtase	1.6 x10 ⁻³⁴	117.3	0.0	ParBc N6_N4_Mtase

유전자 또는 유전자간 지역의 메틸화되지 않은 m6A 위치는 rRNA 또는 tRNA와 같은 코딩 되지 않은 영역은 파지 잔재(phage remnant)나 이동 할 수 있는 유전적 요소(mobile genetic element) 및 전사 조절 유전자에 추가되어 메틸화되지 않는 것으로 나타났다.(Table 5)

육안으로도 쉽게 관찰될 것 같은 GANTC 서열과는 달리 (서열 로고 프로그램, 예를 들어 weblogo.berkeley.edu), m4C 메틸화 패턴은 애매하게 보이며 m4C 측면 서열 (+-20 염기 쌍)을 쉽게 시각화 할 수 없었다. (Table 6) SMRT로 분석된 m6A 및 m4C 정보를 Gibbs motif sampler 3.1 program (Lawrence et al.,

1993)을 사용하여 m6A 및 m4C motif를 분석했다. 또한 5'-NNNNGmCNNNN-3' (n-

=439) 및 5'-NNNNmCGCNNNN-3' (n=172) 된 m4C modification의 대부분을 식별할

수 있었으며 m6A는 5'-GmANTC-3' (n=2647)로 식별할 수 있었다.

Table 5. Unmethylated GAnTC sites $\begin{pmatrix} 5' - GANTC - 3' \\ 3' - CTNAG - 5' \end{pmatrix}$ without N6-methyladenine (m6A) from IMCC12053 genes and intergenic regions.

gAntc	Strand	ganTc	Strand	Comments
1364135	+	1364137	-	Upstream of IMCC12053_1380, dihydroxy amino acid enzyme
181161 1963620	+ +	181163 1963622	-	Upstream of LacI transcriptional regulator, IMCC12053_0192 Between phage tail fiber protein, IMCC12053_2002 and peptidoglycan binding protein, IMCC12053_2006
1991273	+	1991275	-	Between Holiday Junction resolvase, IMCC12053_2044 and phage helicase, IMCC12053_2045
2184266	+	2184268	-	Upstream of Arsenical resistance operon repressor, IMCC12053_2243
2710246	+	2710248	-	Upstream of IMCC12053_2771 ; TetR transcription regulator
2835314	+	2835316	-	Upstream of IMCC12053_2883 ; integrase
2849521	+	2849523	-	Downstream of IMCC12053_2907 ; phage mobile protein
2862938	+	2862940	-	Upstream of phage integrase family protein IMCC12053_2922
2886980	+	2886982	-	Upstream of IMCC12053_2941~2934; Cell wall synthesis genes cluster
2892583	+	2892585	-	Downstream of IMCC12053_2948/2947 ; phage virulence and mobile element
2895952	+	2895954	-	Upstream of IMCC12053_2952; Lon-protease
2953197	+	2953199	-	Upstream of IMCC12053_3004; Autoinducer producing enzyme
3044382	+	3044384	-	Upstream of IMCC12053_3094; GntR family Transciption factor
3096258	+	3096260	-	Downstream of IMCC12053_3155 ; phage remnant
327210	+	327212	-	Upstream of IMCC12053_0327; HTH-transcriptional regulator
522709	+	522711	-	Upstream of IMCC12053_0518; AraC-transcriptional regulator
60148	+	60150	-	Downstream of IMCC12053_0067/0068 ; phage genes
1286398	+	1286400	-	Malyl-CoA lyase (EC 4.1.3.24); Serine-glyoxylate_cycle for CO ₂ fixation
1433766	+	1433768	-	L-proline glycine betaine ABC transport system permease protein ProV (TC 3.A.1.12.1) ; stress response & upstream of Ribosomal RNA gene cluster (IMCC12053_3180~3175)
1689630	+	1689632	-	Flagellar motor rotation protein MotA; motility and chemotaxis gene cluster

1825088	+	1825090	-	Nucleoside-diphosphate-sugar epimerases
193258	+	193260	-	hypothetical protein; down stream gene of sugar-alcohol utilization operon
202821	+	202823	-	5-aminolevulinate synthase (EC 2.3.1.37)
248052	+	248054	-	Xylose isomerase (EC 5.3.1.5); xylose utilization gene cluster
302158	+	302160	-	Peptide methionine sulfoxide reductase MsrB (EC 1.8.4.12)
3096250	+	3096252	-	Hypothetical protein; phage remnant; down stream of phage tail fiber protein gene
682591	+	682593	-	tRNA-Leu-TAG
1935613	+	1935615	-	Large Subunit Ribosomal RNA; LSU rRNA
2847128	+	2847130	-	tRNA-Phe-GAA
421917	+	421919	-	Transcription-repair coupling factor; RNA metabolism
673352	+	673354	-	Methyl-accepting chemotaxis protein



Table 6. Motif elements flanking m4C (N4-methylcytosine) and m6A (N6-methyladenine) modifications. Gibbs program (v3.1) was used for collecting DNA sequences (10 base pairs long window size) and motif elements occurring greater than 50% of the time was used for visualization of the patterns. Methylated bases are shown in black letters.

Motif	Left	Right	No.	Avg. Score	s.d.
Element	End	End	Motifs	(-10 LogP)	
	-9	0	43	25.7	4.81
	-8	DN	44	24.7	4.30
	-7	2	70	25.1	5.81
	-6	3	10	23.4	2.76
	-5	4	439	25.4	5.34
to the second se	-3	6	172	25.7	5.38
	-2	7	19	25.3	5.69
	-1	8	9	28.1	9.55
CTT+CCAAA	0	9	19	27.8	6.24
	-9	0	65	51.9	24.2
	-8	1	35	51.9	22.7
GATTC	-4	5	1537	59.8	17.7
	-3	6	1110	59.3	16.9
	-2	7	11	26.2	5.81
	-1	8	9	21.8	2.33
	0	9	8	25.5	4.17

1.4. 다른 Celeribiacter 유전자와 비교 분석

Celeribacter 종의 다른 2 유전자 (C. indicus P73과 C. baekdonensis B30)와 함께 다

방향족 탄화수소(Polyaromatic hydrocarbon) 분해 능력을 비교하였다(Cao et al., 2015).

비교 유전체를 분석할 3개의 유전자는 CLGenomics에 의해 시각화 되었고, 이는

BLASTP 및 벤 다이어그램으로 분석 되었다; EzGenome (<u>www.ezbiocloud.net</u>)로부터

얻은 P73과 B30 유전자를 비교를 위하여 사용하였다(Figure 2).

Core 게놈은 *Celeribacter* 종 IMCC12053, P73 그리고 B30은 각 78.1%, 71.1%, and 76.2% 영역의 ORF를 공유하고 있었다. 또한, CLGenomics 보고서의 게놈 요약은 COG-based classification을 생성(Table 7)하여 사용하였다(Tatusov et al., 2000).

상기 감소된 게놈 크기에 관해선 방향족 화합물의 핵심 효소가 변형되어서

IMCC12053의 게놈에선 찾아볼 수 없었고, P73 및 B30이 심해 퇴적물 환경으로

갈수록 게놈 크기가 증가한 것으로 나타났다(Cao et al., 2015; Lai et al., 2014).



Figure 2. Three *Celeribacter* genomes compared. CLGenomics comparative genomics files (*.cg) were generated and were used to count core genes and/or overlapping genes covering the genomes from strains IMCC12053, P73, and B30.

		IMCC	12053	P73		B30	
COG	Description	No. of Genes	%	No. c Genes	of _{0%}	No. of _% Genes	
J	Translation, ribosomal structure and	151	6 46%	159	4 23%	167	5.06%
	biogenesis	101	0.1070	109	1.2370	107	5.0070
K	Transcription	134	5.73%	272	7.23%	263	7.97%
L	Replication, recombination and repair	152	6.50%	247	6.57%	137	4.15%
D	Cell cycle control, cell division, chromosome partitioning	30	1.28%	35	0.93%	31	0.94%
0	Posttranslational modification, protein turnover, chaperones	113	4.84%	158	4.20%	137	4.15%
М	Cell wall/membrane/envelope biogenesis	158	6.76%	216	5.74%	176	5.33%
Ν	Cell motility	76	3.25%	50	1.33%	46	1.39%
Р	Inorganic ion transport and metabolism	86	3.68%	314	8.35%	191	5.79%
Т	Signal transduction mechanisms	82	3.51%	107	2.84%	115	3.48%
С	Energy production and conversion	154	6.59%	259	6.88%	235	7.12%
G	Carbohydrate transport and metabolism	202	8.64%	286	7.60%	250	7.58%
Е	Amino acid transport and metabolism	205	8.77%	387	10.29%	410	12.42%
F	Nucleotide transport and metabolism	69	2.95%	99	2.63%	83	2.52%
Н	Coenzyme transport and metabolism	94	4.02%	140	3.72%	132	4.00%
Ι	Lipid transport and metabolism	90	3.85%	153	4.07%	138	4.18%
Q	Secondary metabolites biosynthesis, transport and catabolism	33	1.41%	83	2.21%	62	1.88%
R	General function prediction only	259	11.08%	421	11.19%	391	11.85%
S	Function unknown	249	10.65%	376	9.99%	336	10.18%
Total		2337	100%	3762	100%	3300	100%

Table 7. Classification of protein-encoding ORF's from *Celeribacter* genomes under functional categories based on COG (<u>http://www.ncbi.nlm.nih.gov/COG/</u>)

게놈 크기에 관계없이 세 Celeribacter 공유하는 유전자는 탄수화물 대사과정인 Embden-Meyerhof 대사경로와 5탄당 인산 대사경로(Pentose phosphate pathway)를 이용하였고, Entner-Doudoroff 대사경로는 IMCC12053에서 d-글루콘산염 탈 수소 효소(d-gluconate dehydratase)가 부재했다. TCA회로는 Celeribacter 종의 Core 게놈에 제시되고 있는데, 아이소사이트레이트 라이아제 (isocitratelyase)가 누락되었지만 ethylmalonyl-CoA 경로(Erb et al., 2007)인 acetate (+3CO2 형성) 동화경로가 존재했다. IMCC125053은 유형 III ribulose-1,5-bisphosphate carboxylase large subunit 만 가지고 있었으며, 반면 C.baekdonensis B30은 RuBisCo의 크고 작은 Subunit을 가지고 있었고, C.indicus P73은 아무것도 가지고 있지 않았다.

Polyaromatic hydrocarbon (PAH) degradation는 *C.indicus* P73의 중요한 기능이지만 *C.baekdonensis* B30는 이전의 연구에서 보고된(Cao et al., 2015) 바와 달리 다방향족 탄화수소 분해(PAH-degradation) 기능이 없었다.

본 연구에서는 IMCC12053 유전자는 P73 (Table 8)의 주요 효소에 대해 검사한 결과 homogentisate, gentisate, 그리고 catechol ortho-cleavage 대사경로는 존재하지 않았다(Table 9). Table 8. Comparison of IMCC12053 and B30 against *Celeribacter* indicus P73. Genomic regions A~D and other extrachromosomal regions (from P73) encoding key enzymes of PAH-degradation paths were used as seed sequence for blastp searching in CLGenomics. (-): not determined, left arrow (\leftarrow): the same annotation as shown in left reference CDS.

Region P73	ⁱⁿ CDS/gene	P73	B30	IMCC12053
A	P73_0177	Fumarylacetoacetate hydrolas family protein	2-Keto-4-pentenoate sehydratase/2-oxohepta-3-ene- 1,7-dioic acid hydratas (catechol pathway)	2,4-Diketo-3- deoxy-L- efuconate hydrolase
А	P73_0175	Gentisate 1,2-dioxygenase (E0 1.13.11.4)	°←	(-)
А	P73_0176	Maleylacetoacetate isomeras (EC 5.2.1.2)	^e Glutathione S-transferase	Maleylacetoaceta te isomerase (EC 5.2.1.2)
В	P73_0352	2-Hydroxychromene-2- carboxylate isomerase	<	2- Hydroxychromen e-2-carboxylate isomerase
В	P73_0349	2,3-dihydroxy-2,3-dihydro- phenylpropionate dehydrogenase (EC 1.3.1)	Dehydrogenases with differer specificities (related to short chain alcohol dehydrogenases	3-Oxoacyl-[acyl- tcarrier protein] reductase (EC) 1.1.1.100)
В	P73_0346	Biphenyl dioxygenase alph subunit (EC 1.14.12.18)	Phenylpropionate dioxygenas aand related ring-hydroxylatin dioxygenases, large termina subunit	eBenzoate 1,2- gdioxygenase alalpha subunit (EC 1.14.12.10)
В	P73_0347	Biphenyl dioxygenase bet subunit (EC 1.14.12.18)	aSmall subunit c phenylpropionate dioxygenase	Benzoate 1,2- ofdioxygenase beta e subunit (EC 1.14.12.10)
В	P73_0348	Biphenyl dioxygenase syster ferredoxin component	Phenylpropionate dioxygenas nand related ring-hydroxylatin dioxygenases, large termina subunit	^e Naphthalene gdioxygenase ¹ ferredoxin
В	P73_0353	Catechol 2,3-dioxygenase (E4 1.13.11.2)	Cbiphenyl-2,3-diol 1,2 dioxygenase	2-(-)
В	P73_0351	Dihydrodipicolinate synthas (EC 4.2.1.52)	"←	÷
В	P73_0354	Ferredoxin reductase	Uncharacterized NAD(FAD) dependent dehydrogenases	-Ferredoxin reductase
В	P73_0330	Fumarylacetoacetate hydrolas family protein	2-Keto-4-pentenoate sehydratase/2-oxohepta-3-ene- 1,7-dioic acid hydratas (catechol pathway)	deoxy-L- efuconate hydrolase
В	P73_0329	Homogentisate 1,2-dioxygenas (EC 1.13.11.5)	se(-)	(-)

В	P73_0331	Maleylacetoacetate isomerase (EC 5.2.1.2) Glutathione S-transferase	Glutathione S- transferase family protein
В	P73_0350	Probable VANILLIN dehydrogenase oxidoreductase dehydrogenases aldehydrogenases	Betaine aldehyde dehydrogenase (EC 1.2.1.8)
С	P73_0845	3-Carboxy-cis,cis-muconate cycloisomerase (EC 5.5.1.2) \leftarrow	(-)
С	P73_0837	3-Oxoadipate CoA-transferaseAcyl CoA:acetate/3-ketoacia subunit A (EC 2.8.3.6) CoA transferase, alpha subunit	Butyrate- acetoacetate CoA-transferase subunit A (EC 2.8.3.9) Suscinul CoA:2
С	P73_0838	3-Oxoadipate CoA-transferaseAcyl CoA:acetate/3-ketoacid subunit B (EC 2.8.3.6) CoA transferase, beta subunit	ketoacid- dcoenzyme A transferase subunit B (EC 2.8.3.5)
С	P73_0839	Acetyl-CoA acetyltransferase (EC 2.3.1.9)	÷
С	P73_0841	Beta-ketoadipate enol-lactone hydrolase (EC 3.1.1.24) Predicted hydrolases o acyltransferases (alpha/beta hydrolase superfamily)	Beta-ketoadipate enol-lactone hydrolase (EC 3.1.1.24)
С	P73_0844	Protocatechuate 3,4- dioxygenase alpha chain (EC← 1.13.11.3)	Benzoate 1,2- dioxygenase alpha subunit (EC 1.14.12.10)
С	P73_0843	Protocatechuate 3,4- dioxygenase beta chain (EC ← 1.13.11.3)	Benzoate 1,2- dioxygenase beta subunit (EC 1.14.12.10)
С	P73_0842	Uncharacterized homolog of gamma-carboxymuconolactone \leftarrow decarboxylase subunit	←
D	P73_2964	4,5-Dihydroxyphthalate decarboxylase (-)	(-)
D	P73_2966	Flavodoxin reductases (ferredoxin-NADPH \leftarrow reductases) family 1	Benzoate dioxygenase, ferredoxin reductase component
D	P73_2965	m-Inositol 2-dehydrogenasePredicted dehydrogenases and related proteins	m-Inositol 2- dehydrogenase (EC 1.1.1.18)
D	P73_2968	Putative phthalate 4,5-Phenylpropionate dioxygenase dioxygenase oxygenase subunit (OhpA2) Phenylpropionate dioxygenases subunit dioxygenases, large terminal subunit	eBenzoate 1,2- gdioxygenase lalpha subunit (EC 1.14.12.10)
etc	P73_1121	Muconate cycloisomerase (EC 5.5.1.1)	(

etc	P73_1122	Benzenetriol dioxygenase	Intradiol ring-cleavag dioxygenase	^{5e} (-)
etc	P73_1454	Gentisate 1,2-dioxygenase (EC 1.13.11.4)	€	(-)
etc	P73_1455	Maleylacetoacetate isomeras (EC 5.2.1.2)	eGlutathione S-transferase	Maleylacetoaceta te isomerase (EC 5.2.1.2)
etc	P73_1456	Fumarylacetoacetate hydrolas family protein	2-Keto-4-pentenoate ehydratase/2-oxohepta-3-ene- 1,7-dioic acid hydratas (catechol pathway)	2,4-Diketo-3- deoxy-L- sefuconate hydrolase
etc	P73_1457	Putative n-hydroxybenzoat hydroxylase	2-Polyprenyl-6- emethoxyphenol hydroxylas and related FAD-depender oxidoreductases	se ^{hydroxylase} (EC ^{nt} 1.14.13.1)
etc	P73_2927	Salicylate hydroxylase (EC 1.14.13.1)	2-Polyprenyl-6- Cmethoxyphenol hydroxylas and related FAD-depender oxidoreductases	se Salicylate hydroxylase (EC 1.14.13.1)
etc	P73_3884	Salicylate hydroxylase (EC 1.14.13.1)	2-Polyprenyl-6- Cmethoxyphenol hydroxylas and related FAD-depender oxidoreductases	Salicylate hydroxylase (EC 1.14.13.1)
etc	P73_4757	Putative n-hydroxybenzoat hydroxylase	2-Polyprenyl-6- emethoxyphenol hydroxylas and related FAD-depender oxidoreductases	Salicylate hydroxylase (EC 1.14.13.1)
etc	P73_4773	Fumarylpyruvate hydrolase	←	÷
etc	P73_4774	Maleylacetoacetate isomeras (EC 5.2.1.2)	eGlutathione S-transferase	Glutathione S- transferase family protein
etc	P73_4775	Gentisate 1,2-dioxygenase (EC 1.13.11.4)	^C ←	(-)

Figure 9. Missing enzymes of IMCC12053 compared to P73 that had PAH degrading functional genes in chromosome and plasmids (etc) (Cao et al., 2015).

Regions in P73	LOCUS from P73	P73	B30	IMCC12053
А	P73_0175	Gentisate 1, 2-dioxygenase (EC 1.13.11.4)	Gentisate 1, 2-dioxygenase	_
В	P73_0353	Catechol 2, 3-dioxygenase (EC 1.13.11.2)	Biphenyl-2, 3-diol 1, 2-dioxygenase	_
В	P73_0329	Homogentisate 1, 2-dioxygenase (EC 1.13.11.5)		_
С	P73_0845	3-Carboxy-cis, cis-muconate cycloisomerase (EC 5.5.1.2)	3-Carboxy-cis, cis-muconate cycloisomerase	_
D	P73_2964	4, 5-Dihydroxyphthalate decarboxylase	- III	_
etc	P73_1122	Benzenetriol dioxygenase	Intradiol ring-cleavage dioxygenase	_
etc	P73_1454	Gentisate 1, 2-dioxygenase (EC 1.13.11.4)	Gentisate 1, 2-dioxygenase	_
etc	P73_4775	Gentisate 1, 2-dioxygenase (EC 1.13.11.4)	Gentisate 1, 2-dioxygenase	_

1.5. Nucleotide Accession Number

Celeribacter marinus IMCC12053 의 complete 게놈 서열의 Genebank database

accession number는 CPO12023이다. 또한, IMCC12053의 균체는 Korean Agricultural

Culture Collection (KACC) 혹은 Biological Resource Center, National Institute of

Technology and Evaluation (NBRC) 두 기관에서 각각 KACC178482, NBRC109702의

등록 번호를 통해서 분양 받을 수 있다.

Chapter 2. Xanthid Crab에 공생하는 *Flavobacteriales bacterium* 균주 UJ101의 비교 유전체 분석

2.1. 서 론

Flavobacteriales는 Flavobacteriia강, Bacteroidetes문에 속하는 구성원으로, 매우 다양한 박테리아 그룹으로 구성되어 있으며 토양, 해양, 식물 그리고 동물의 내장 등 다양한 환경에서 분리되어 왔으며, 그들은 수생 및 육상 환경의 박테리아 군집에서 중요한 역할을 하며, 미생물 군집 구성원의 20%까지 도달 할 수 있을 정도로 매우 풍부하다(Kolton et al., 2013).

몇몇 *Flavobacterium* 종은 심한 물고기의 질병의 원인(Declercq et al., 2013)이 되기도 하며, 식물이나 동물의 보호 및 성장을 촉진하는 역할(Sang and Kim, 2012) 및 토양 혹은 해양 퇴적물의 생물학적 정화에도 영향을 미친다(Fu et al., 2011; Jit et al., 2008).

동해안에 사는 Xanthid Crab 껍질에서 분리된 Flavobacteriales bacterium 균주

UJ101은 아직까지 계통학적 분류 및 생태학적 위치가 모호한 상태이며, 생태계에서

어떠한 기능을 수행하는지 알려지지 않았다.

이에 본 연구에서는 균주 UJ101의 전체 유전자 서열을 SMRT 기술을 이용하여 거의 완전하게 가까운 유전자(Table 1)를 보고한다.

또한, 이 유전체를 이용하여, ANI(Average Nucleotide Identity) 값에 의한 게놈 유사도

및 16S rRNA 유사도에 의한 계통분류 및 분리된 환경 등을 고려하여 비교할

균주의 Flavobacteriales 목의 유전체 선정했다. 또한, Chunlab의 비교 유전체

플랫폼(cg.ezbiocloud.net)을 이용하여, 균주 UJ101 및 Flavobacteriales에 속하는

15균주의 유전체 기능 및 대사과정을 비교 분석하려 시도하였다.

Genome assembly data	
Item	Description
Assembly method	SMRT pipe HGAP, SMRTpipe Celera Assembler, SMRTpipe Quiver
Genome coverage	1063 ×
Sequencing technology	PacBioRS2
MIGS data	
Item	Description
investigation_type	bacteria_archaea
project_name	Genome sequencing of Flavobacteriales bacterium UJ101
experimental_factor	NA
lat_lon	37.10 N 128.38 E
depth	NA
alt_elev	NA
geo_loc_name	South Korea; East sea
collection_date	29-FEB-2012
Country	Korea
Environment	Xanthid Crab gut microbiome
num_replicons	NA
ref_biomaterial	NA
pathogenicity	NA
biotic_relationship	Symbiotic
trophic_level	chemoheterotroph
rel_to_oxygen	aerobe
isol_growth_condt	NA
seq_math	PacBio RS II sequencing
assembly_name	NA
finishing_strategy	complete; 1063x coverage; 1 contig
annot_source	NA
assembly	SMRT
env_package	missing

Table 1. Flavobacteriales bacterium UJ101 genome assembly data and general features.

2.2. 유전자 서열 분석 데이터

균주 UJ101의 전체 유전자 서열은 Pacific Biosciences SMRT tool을 이용하였으며, 분석과정은 주) 디엔에이링크 (DNA Link, Inc., Seoul, Korea) (Table 1)를 이용하여 분석되었다: Open Reading Frames(ORFs)는 Glimmer 3 Program, 각 서열 기능의 주석은 Blast2Go (SMRT tools pipeline)로 예측되었다. BLASTP는 ORFs를 확인하기 위해 사용되었고, ribosomal RNA, transfer RNA는 RNAmmer(Lagesen et al., 2007)와 tRNAscan-SE (Lowe and Eddy, 1997)를 사용하여 예측되었다. 대사 경로 해석을 위해, 모든 단백질 서열은 National Center for Biotechnology Information (NCBI) 데이터베이스를 기반으로 하여 확인하였으며, RAST 서버를 사용하여 분석 하였다 (Aziz et al., 2008). GeneBank 제출 확인 된 데이터는 유전자의 시각화를 위해 CLGenomics 프로그램 형식(ChunLab, Inc., Seoul, Korea)으로 변환하였다. CLgenomics에서 보낸 유전자 테이블은 비교 유전자 형식이 생성되었고,

GeneBank에 등록하기 위해 사용하였고, 시각화 된 파일 전체 유전자 정보 파일을 이용하여, 염색체 지도 정보를 생성하였다 (Figure 1).

균주 UJ101의 유전자는 30.74% 의 GC 함량, 길이는 3,074,209 염기 쌍의 단일 원형 염색체로 구성되었다. 유전자는 46개의 tRNA, 9개의 rRNA가 포함되어 있다 (Table 2).





Figure 1. Circular Map of the Chromosome of strain UJ101. From outside to the center: genome size (black line), RNA, forward strand (colored by COG categories), reverse strand (colored by COG categories), GC ratio (brown), GC skew (green and red)

Genome size (bp)	3,074,209
DNA coding region (bp)	2,803,554
DNA G + C content	30.74%
Total genes	2,740
tRNA genes	46
rRNA genes	9
Protein coding genes	2,684
Gene with predicted function	197

Value

Table 2. Flavobacteriales bacterium UJ101 Genome Summary

Attribute

UJ101의 계통학적 위치 및 생태계에서의 역할 등을 알아보기 위하여 다른 *Flavobacteriales*와 비교 유전체 분석을 실시하였다: 환경 데이터 및 ANI(Average Nucleotide Identity), 16s rRNA 유사성을 고려하여 선정된 15 *Flavobacteriales*의 정보(Table 3)는 15개의 참고문헌을 참고하여 비교하였다.

각 Flavobacteriales의 Ortho Average Nucleotide Identity 값을 이용하여, Phylogenetic tree (Figure 2)를 그렸다. 그 결과 양식어종 등에 기생하는 Flavobacterium 종(genus)과 가까웠다: 제일 가까운 종으론 Flaviramulus ichthyoenteri Th78, 다음으로 Flavobacterium psychrophilum JIP02/86와 Flavobacterium branchiophilum FL-15 순이었다. 그러나 계통 분석에서 균주 UJ101과 다른 균주 들은 과(Family)명만 동등한 수준이므로 유연관계는 그리 가깝다고 볼 수 없다. 비교 유전학은 다른 생물의 유전적 특징을 비교하는 학문이며, 이 균주 들의 비교 유전체 분석은 ㈜ 천랩 (cg.ezbiocloud.net/)의 정보를 이용하여 실시한 결과이다.

도구로서는 pan-/core-genome 분석, 주요 Clusters of Orthologous Group(COG)를

이용한 Pan-genome Orthologous Groups(POGs)의 빈도 분석, Identity differentially

present POGs enrichment pathway 그리고 singleton 분석 등이 있다.



	Organism	Country	Latitude/ Longtitude	Isolation source	Pathogen	No. of contigs	Genome size (bp)	DNA G+C content (%)	No. of CDSs	No. of rRNA genes	No. of tRNA genes	Reference
1	Flavobacteriaceae bacterium UJ101	South Korea; East Sea	37°06'01.1"N 129°22'48.1"E	Marine Invertebrate	х	1	3074209	30.7431928	2683	9	46	This study
2	Flaviramulus ichthyoenteri Th78	China: Qingdao	ND	Marine Fish	Х	5	3953230	32.03400763	3433	6	40	(Zhang et al., 2015)
3	Flavobacterium branchiophilum FL-15	Hungary	ND	Marine Fish	0	2	3563292	32.85829508	3123	9	44	(Touchon et al., 2011)
4	Flavobacterium psychrophilum JIP02/86	France	ND	Marine Fish	0	1	2860382	32.53299734	2510	18	49	(Duchaud et al., 2007)
5	Gramella echinicola DSM 19838	Pacific Ocean	ND	Marine Invertebrate	Х	18	3513826	36.89915209	3184	6	40	(Nedashkovskaya et al., 2005)
6	Gramella forsetii KT0803	Germany	ND	Seawater	Х	1	3798465	36.61076251	3446	9	44	(Kabisch et al., 2014)
7	<i>Gramella portivictoriae</i> DSM 23547	Hong Kong	22°18'37.0"N 114°07'11.9"E	Marine Sediment	х	8	3269398	39.52859211	2985	6	41	(Lau et al., 2005)
8	Nonlabens dokdonensis DSW-6	South Korea: Dokdo	37°14'24.8"N 131°52'04.2"E	Seawater	х	1	3914632	35.31529911	3455	6	41	(Kwon et al., 2013)
9	Nonlabens marinus S1-08	North West Pacific Ocean	30°11'01.0"N 145°05'00.0"E	Seawater	х	1	2915920	39.52515844	2677	6	36	(Park et al., 2012)
10	Nonlabens sediminis JCM19294	Japan	41°48'59.7"N 140°41'59.7"E	Seawater	х	22	2935762	35.33494881	2899	3	32	(Khan et al., 2006)
11	Nonlabens sp. MIC269	Micronesia	7°25'45.3"N 151°51'31.9"E	Marine Sediment	Х	1	2884293	35.45142605	2613	9	36	(Park et al., 2012)
12	Nonlabens ulvanivorans PLR	France; Brittany	48°51'35.2"N 3°04'43.1"W	Marine Invertebrate	x	45	3211390	35.010852	2940	3	33	(Kopel et al., 2014)
13	<i>Tenacibaculum maritimum</i> NBRC 15946	Japan	ND	Marine Fish	0	159	3225035	31.80458507	2915	3	39	(Yoon et al., 2005)
14	Tenacibaculum ovolyticum DSM 18103	Norway	ND	Marine Fish	0	34	4124462	29.53267602	3693	12	45	(Jit et al., 2008)
15	Tenacibaculumsp.47A_GOM-205m	Japan	ND	Seawater	Х	11	2897856	31.81379613	2766	4	45	(Suzuki et al., 2001)
16	Vitellibacter vladivostokensis KMM3516	Japan	ND	Marine Invertebrate	х	25	3269452	40.76019467	3088	4	35	(Nedashkovskaya et al., 2003)

Table 3. Summary of 16 *Flavobacteriales* genome and their contextual data.



Figure 2. Phylogenetic tree by Orthologous Average Nucleotide Identity (OrthoANI) analysis 16 *Flavobacteriales*.

2.3.1. Pan/Core-genome 및 게놈을 통한 기능 분석

Pan-genome은 유전체 세트의 단백질 코딩 유전자(CDS) 들의 합집합으로 16

Flavobacteriia의 Pan-genome 분석을 위해, 유전체 세트에서 CDS를 찾고, 각 CDS가 나타낼 수 있는 기능을 판독했다. 그 이후 BLAST를 통해 대응관계 및

상호관계를 분석했다 (https://chunlabkr.wordpress.com/pan-genome-analysis/).

F.bacterium에서 가능한 게놈 변화를 예측하기 위해, ㈜ 천랩의 Bioinformatics Tool(cg.ezbiocloud.net/)을 이용하여 모든 16 Flavobacteriia 그룹의 Pan-genome을 예측하고, Pan-genome(유전자 레퍼토리), accessory genome (하나의 게놈 상에 있는 특정 유전자) 및 Core-genome (상호 보존된 공통 유전자)를 계산하였다(Castillo et al., 2016).

각 16균주의 유전자 정보는 CLgenomics에 의해 시각화 되었으며, 각 정보를 근거한 유전체인 Pan-genome을 이용하여 각 서식지 및 종별로 그룹을 지어 균주 UJ101과 비교한 벤 다이어그램으로 나타내었다(Figure 3). Pan-genome은 유전체의 수가 증가하면 그 크기도 증가하며, 반대로 Coregenome은 유전체의 수가 증가하면 감소하는 경향을 보인다. 16균주들의 유전체 별 Pan-genome 과 Core-genome 크기 변화 그래프(Figure 4)를 그렸으며 분석 결과 위 경향과 일치하였다.





Figure 3. Venn Diagram of Compared Genomes. A: Comparison between UJ101 and Marine Vertebrate Group, B: Marine Invertebrate Group, C: *Flavoabcterium* Group, D: *Gramella* Group, E: *Nonlabens* Group, F: *Tenacibaculum* Group (*cf.* contextual data in Table 3)



Figure 4. 16 *Flavobacteriales* pan and core-genome according to the number of sequenced genomes. Accumulation curves for the total number of genes (Pan-genome, Blue) or the number of genes in common (Core-genome, Green). This graph is shown for increasing or decreasing values for 16 *Flavobacteriales* genome sequenced based on a power law fit models. The vertical bars correspond to standard deviations after repeating one hundred random input orders of the genomes (Kolton et al., 2013).

상동 유전자 후보 군은 Pan-genome Orthologous Group(POG) 즉 pan-genome과 core-genome의 orthologous group을 생성하는 pan-genome 연산 후에 중복되지 않은 유전자로 묶이게 되고, 유전체의 핵심 부분은 모든 유전체에서 반드시 발견되며, 중요하지 않은 유전자일수록 발견되는 빈도가 낮다. 16균주의 유전자 세트를 가지고 유전자들을 4개의 주요 기능 그룹으로 분류하여 Pan-genome의 각각의 유전자의 상대적인 빈도를 나타내었다 (Figure 5): 각 사각형의 색깔은 기능 유전자들을 나타내고 있으며, 하나의 유전체에서만 독립적으로 발견되는 유전자는 v축 맨 윗부분으로 7,220개의 유전자가 있으며, pan-genome의 약 62%를 차지하고 있었다. 이는 주요 기능들이 밝혀져 있지 않은 비-필수 유전자가 대부분이다. 또한 전체 16개 유전체에서 공통적으로 발견된 유전자는 y축 맨 아랫부분으로 994개가 발견되었으며 이들은 pan-genome의 약 8%만을 차지하고 있었다.



Figure 5. Frequency of genes within the 16 analysed *Flavobacteriales* genomes. (Frequency Plot of POGs with Major COG Category).

16 균주의 Pan-genome 데이터베이스를 이용하여, Pan-/Core-genome을 COG 및

SEED subsystem 범주 내에서 비교하였다(Figure 6) 이 그래프와 범주는 Pan-Genome에 더 많은 Core-genome의 COG 기능 혹은 SEED subsystem 범주의 비율 변화를 보여준다. (대부분이 밝혀지지 않은 기능 혹은 데이터베이스와 일치하지 않는 유전자이다.)

다음으로 한 균주의 고유한 기능을 가지고 있는 유전자의 수를 Singleton이라고 하며, 16 균주의 Singleton 수를 비교한 내용은 Table 4에 나타내었다. 예상대로 이 균주 들의 기능 정보가 많이 부족해 아직 밝혀지지 않은 Singleton이 많았지만, 균주 UJ101은 다른 균주에 비해 Nucleotide transport and metabolism에 관한 Singleton을 많이 가지고 있었다. *Flavobacteria ichthyoenteri* 균주 Th78과 *Flavobacteria psychrophilum* 균주 JIP02/86은 유전자 복제 및 재조합 그리고 복구 기능의 Singleton을 다른 균주 들에 비해 매우 많이 가지고 있는 특징이 관찰되었다.



Figure 6. Comparison 16 *Flavobacteriales* functional Categories by COG (A) and SEED subsystem category (B).

COG	UJ101	Th78	FL-15	JIP02/86	DSM 19838	KT0803	DSM 23547	DSW-6	S1-08	JCM 19294	MIC269	PLR	NBRC 15946	DSM 18103	47A_GOM-205m	KMM 3516
С	5	5	1	1	2	6	1	0	1	6	3	0	1	9	2	5
D	0	0	1	0	0	0	0	1	0	0	0	0	2	1	1	0
E	9	15	7	1	5	3	7	4	1	5	2	1	6	18	2	5
F	10	3	1	0	0	1	0	0	0	0	0	0	3	2	1	2
G	15	48	12	5	11	21	19	1	2	5	1	8	1	8	1	6
Н	3	6	1	1	2	1	0	1	1	2	0	1	3	18	0	4
Ι	3	5	7	2	4	3	0	2	1	5	0	1	1	11	1	0
J	3	0	2	0	1/ (2	0	1	2	2	1	1	0	2	1	3
K	17	25	8	7	4	10	3	2	0	5	0	3	6	20	5	2
L	7	42	36	45	8	29	4	27	2	9	8	4	6	22	5	29
М	16	25	57	12	14	8	9	9	5	8	4	3	21	7	1	16
Ν	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0
0	13	5	6	2	3	5	2	9	2	2	1	2	9	8	5	8
Р	19	32	12	3	3	10	5	4	3	6	0	1	3	11	2	8
Q	6	6	1	1	2	3	2	9	2	2	0	1	0	4	1	2
Т	21	10	25	7	10	9	6	1	4	4	4	2	6	27	5	11
U	0	4	0	0	0	0	0	2	1	1	0	0	0	1	0	4
V	2	5	11	7	4	4	2	6	5	10	1	0	3	13	2	5
Z	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
S	230	311	343	160	93	177	82	165	74	58	44	86	185	258	104	209
Х	279	222	396	194	140	179	131	351	152	114	72	148	358	404	216	224
Total	658	769	929	448	306	472	273	595	258	244	141	262	614	844	356	543

Table 4. Comparison of 16 *Flavobacteriales* strains by Singleton Number. (*cf.* COG functional categories in Figure 6, strain's full name in Table 3).

2.3.2. 대사 경로 비교 및 분석

먼저 균주 UJ101의 주석 달린 유전자 정보를 참고하여 대사과정 분석 결과 Entner-Doudoroff 경로는 결여되어 있었으며, TCA와 glyoxylate cycle은 잘 보존되어 있다. Carbon fixation과 one carbon metabolism에 관련된 유전자는 formaldehyde dehydrogenase를 제외하고 대부분이 결핍되어 있다. 분해 효소는 5개의 glycosyl transferase, 3개의 amylase 그리고 8개의 peptidase가 있었다. 라이신과 트립토판, 페닐알라닌 및 타이로신 생합성 효소는 결여되어 있었다. 알코올과 유기산 발효에 관한 효소 또한 결핍되어 있어 알코올 생성 및 유기산 발효는 하지 않을 것으로 예측된다. 16개의 균주 정보(Table 2)의 분리된 소스(Source) (Marine Vertebrate, Invertebrate and Non-bio) 및 종으로 나누어 통계를 쉽게 육안으로 확인할 수 있는 벤 다이어그램을 그렸다(Figure 7).



Figure 7. Venn diagram of each specific isolation source 16 Flavobacteriales

대사경로 강화축적(Metabolic enrichment)이란 두 유전체 그룹 간에 POG 및 서로 다르게 존재하는 유전자들을 찾은 후 이러한 유전자 그룹이 발견되면 이 정보를 가지고, 풍부하게 존재하는 대사과정을 말하며, 전체 유전자 세트 및 서로 다르게 존재하는 유전자(Singleton) 세트들의 정보를 보여준다.

UJ101을 제외한 다른 15균주의 서식지(Marine Vertebrate, Invertebrate) 및 종(Flavobacterium, Nonlabens, Gramella, Tenacibaculum) 으로 나눠 총 6개 그룹을 각각 만들어 UJ101과 Metabolic Enrichment Singleton을 비교하였다. 중요하다고 고려되는 대사과정(Metabolism pathway)인 Nitrogen(질소), Methane(메탄), Propanoate 및 Butanoate, Sulfur(황)의 Singleton을 비교 분석(Table 5)을 수행하였고, 특히 UJ101이 다른 균주들에 비해 질소 대사과정에 관련된 Singleton이 많이 존재함을 알

수 있었다.

Table 5. Differentially presented KEGG metabolic pathway enrichments between strain UJ101 and respective symbiotic group (+: Different number of genes in the metabolic process).

Α.				Β.				
	Metabolism	UJ101	Marine Vertebrate Group		Metabolism	UJ101	Ir	Marine nvertebrate Group
	Nigtrogen	++	+		Nigtrogen	++++++++	+ ++	
	Methane	+			Methane	++++	+	
	Propanoate	+++	+++		Propanate	++++++	++	+
	Butanoate	+++	+		Butanoate	+++++	++	
	Sulfer	++	+ A HOI		Sulfer	++	++	+
С				D				
0.	Metabolism UJ101		Flavobacterium Group		Metabolism	UJ101		<i>Gramella</i> Group
	Nitrogen -	+++++++++++++++++++++++++++++++++++++++	++		Nitrogen -	+++++++++++++++++++++++++++++++++++++++	++	
	Methane ·	++++			Methane -	+++		+++
	Propionate	+++++	+++		Propionate -	++++++		+++++
	Butanoate -	+++++	+		Butanoate -	+++		++++
	Sulfur ·	++++	+		Sulfur -	++		+++
E.				P				
	Metabolism	UJ101	Nonlabens Group		Metabolism	UJ101	Tena	<i>cibaculum</i> Group
	Nitrogen +++++++++		+++		Nitrogen	+++++	+++	
	Methane	++	++		Methane	++	+	
	Propionate	+++++	+++		Propionate	+++	+++++	ŀ
	Butanoate	+++++	+++		Butanoate	+++		
	Sulfur	+++	++		Sulfur	++	+++++	ŀ

또한, UJ101과 다른 15균주 전체의 대사경로 분포를 비교하여 p value 값이 0.5 이하인 값의 대사과정과 Singleton이 7개 이상 존재하는 대사과정을 고려하여 9 종류의 대사과정을 선택하여 분석하였다: 에너지 대사 2종류(Nitrogen(질소), Carbon fixation Metabolism in Prokaryote (원핵생물 탄소 고정 대사)), 탄수화물 대사 4종류 (Propanoate, Butanoate, Galactose, Starch and Sucrose), 보조인자(Cofactor)에 관한 대사 3종류(Thiamine, Folate Biosynthesis, Biotin)로 분류된다. 이러한 대사과정에서의 균주 UJ101의 Singleton을 비교 분석하였다. 첫째, 에너지 대사(Figure 8)를 분석한 결과는 질소 대사과정 (KEGG number: KO00910)에서는 Nitronate monooxygenase (WP 071671356.1)가 존재함에 따라 Nitroalkane을 Nitrite로 분해해 질소 순환 과정에 이용할 수 있었다. 원핵 생물 탄소고정 대사과정 (KEGG

number: KO00720)에서는 Phosphate acetyltransferase 유전자가 하나 더 많았다(WP 071670495.1).



Figure 8. Compared between strain UJ101 and other 15 *Flavobacteriales* Metabolic enrichment by KEGG Energy metabolism. Strain UJ101's singleton is rectangular box. (1) is Nitronate monooxygenase, (2) is Phosphate acetyltransferase (pta)

둘째, 탄수화물 대사(Figure 9)를 분석한 결과는 Propanoate (KEGG number: KO00640)는 Methylisocitrate lyase (WP 071670334.1)와 2-methylcitrate dehydrate (WP 071670333.1)가 Butanoate (KEGG number: KO00650)는 butyrate kinase (WP 071670494.1)) butanol dehydrogenase (WP 071669855.1, WP 071670146.1, WP 071671039.1), Galactose (KEGG number: KO00052)는 UDP-glucose 4-epimerase (WP 071672050.1), Alpha glucosidase (WP 071671715.1)의 Singleton이 존재했으며, Starch and Sucrose (KEGG number: KO00500)에서는 amylase treX (WP 071669674.1)의 유전자가 하나 더 존재 (KO number: KO2438)하여 다른 15균주에 비해 트레할로스(trehalose)를 많이 이용한다. 그 외에 Pullulanase (WP 071672033.1), αamylase (WP 071671658.1), Alpha-glucosidase 및 4- α -glucanotransferase (WP 071671741.1)가 존재했으며 Amylase가 다른 15개의 균주에 비해 많다는 것이 특징이었다.



Figure 9. Compared between strain UJ101 and other 15 *Flavobacteriales* Metabolic enrichment by KEGG Carbohydrate metabolism. Strain UJ101's singleton is rectangular box. (1) is 2-methylcitrate dehydratase (prpD), (2) is Methylisocitrate lyase (prpB), (3) is Phosphate acetyltransferase (pta), (4) is butanol dehydrogenase, (5) is butyrate kinase, (6) is UDP-glucose 4-epimerase (galE,GALE), (7) is Alpha-glucosidase (malZ), (8) is Isoamylase 1, chloroplastic (treX,glgX), (9) is Pullulanase, (10) is Glycogen phosphorylase, (11) is Alpha-amylase (amyA,malS), (12) is Alpha-glucosidase (malZ), (13) is 4-alpha-glucanotransferase (malQ).

마지막으로 조효소에 관련된 대사과정 (Figure 10)를 분석했다. Thiamine metabolism

(KEGG number: KO00730)에서는 Hydroxymethylpyrimidine kinase (WP_071670608.1),

Folate biosynthesis (KEGG number: KO00790)는 Dihydroneopterin aldolase (WP_071670482.1)의 Singleton이 존재했으며, Biotin metabolism (KEGG number: KO00780)은 UJ101에 대비 *Gramella* 그룹에서 Biotin 합성에 관련된 유전자가 더 많이 발견되었다. Singleton은 *Gramella forsetii* strain KT0803에 Dehydrogenase reductase로 나타났다.



Figure 10. Compared between strain UJ101 and other 15 Flavobacteriales Metabolic enrichment by KEGG Cofactor related metabolism. Strain UJ101's singleton is rectangular box. (1) is Hydroxymethylpyrimidine kinase (thiD), (2) is Dihydroneopterin aldolase (FOL1), (3) is 3-hydroxyacyl-[acyl-carrier-protein] dehydratase (fabZ).

UJ101 균주의 Metabolic enrichment Singleton에서 주목할 만한 부분은 Nitroalkane을 질소 순환 과정에 이용하는 부분과 당 대사과정에서 아밀라아제(amylase)가 더 존재해 녹말 분해 능력, 트레할로즈(trehalose) 이용 능력이 다른 균주에 비해 우위에 있을 것으로 예상된다. 더불어 엽산 생합성(Folate biosynthesis)에 관여된 유전자가 다른 균주에 비해 많이 보이나, UJ101의 전체 게놈의 엽산 생합성 관련 유전자를 살펴본 결과 7,8-Dihydroneopterin 3'triphosphate을 이용해 Dihydroneopterin 생성해 내는 효소가 결여되어 있는 것으로 분석되었다.

참고문헌

1. Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., 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.

2. Baek, K., Choi, A., Kang, I., Cho, J.C., 2014. Celeribacter marinus sp. nov., isolated from coastal seawater. Int J Syst Evol Microbiol 64, 1323– 1327.

3. Cao, J., Lai, Q., Yuan, J., Shao, Z., 2015. Genomic and metabolic analysis of fluoranthene degradation pathway in Celeribacter indicus P73T. Sci Rep 5, 7741.

4. Castillo, D., Christiansen, R.H., Dalsgaard, I., Madsen, L., Espejo, R., Middelboe, M., 2016. Comparative Genome Analysis Provides Insights into the Pathogenicity of Flavobacterium psychrophilum. PLoS One 11, e0152515.

5. Declercq, A.M., Haesebrouck, F., Van den Broeck, W., Bossier, P., Decostere, A., 2013. Columnaris disease in fish: a review with emphasis on bacterium-host interactions. Vet Res 44, 27.

6. Duchaud, E., Boussaha, M., Loux, V., Bernardet, J.F., Michel, C., Kerouault, B., Mondot, S., Nicolas, P., Bossy, R., Caron, C., Bessieres, P., Gibrat, J.F., Claverol, S., Dumetz, F., Le Henaff, M., Benmansour, A., 2007. Complete genome sequence of the fish pathogen Flavobacterium psychrophilum. Nat Biotechnol 25, 763-769.

7. Erb, T.J., Berg, I.A., Brecht, V., Muller, M., Fuchs, G., Alber, B.E., 2007. Synthesis of C5-dicarboxylic acids from C2-units involving

crotonyl-CoA carboxylase/reductase: the ethylmalonyl-CoA pathway. Proc Natl Acad Sci U S A 104, 10631-10636.

8. Fu, Y., Tang, X., Lai, Q., Zhang, C., Zhong, H., Li, W., Liu, Y., Chen, L., Sun, F., Shao, Z., 2011. Flavobacterium beibuense sp. nov., isolated from marine sediment. Int J Syst Evol Microbiol 61, 205–209.

9. Ivanova, E.P., Webb, H., Christen, R., Zhukova, N.V., Kurilenko, V.V., Kalinovskaya, N.I., Crawford, R.J., 2010. Celeribacter neptunius gen. nov., sp. nov., a new member of the class Alphaproteobacteria. Int J Syst Evol Microbiol 60, 1620-1625.

10. Jit, S., Dadhwal, M., Prakash, O., Lal, R., 2008. Flavobacterium lindanitolerans sp. nov., isolated from hexachlorocyclohexanecontaminated soil. Int J Syst Evol Microbiol 58, 1665-1669.

11. Kabisch, A., Otto, A., Konig, S., Becher, D., Albrecht, D., Schuler, M., Teeling, H., Amann, R.I., Schweder, T., 2014. Functional characterization of polysaccharide utilization loci in the marine Bacteroidetes 'Gramella forsetii' KT0803. ISME J 8, 1492–1502.

12. Kang, I., Jang, H., Oh, H.M., Cho, J.C., 2012. Complete genome sequence of Celeribacter bacteriophage P12053L. J Virol 86, 8339-8340. Khan, S.T., Nakagawa, Y., Harayama, S., 2006. Sandarakinotalea sediminis gen. nov., sp. nov., a novel member of the family Flavobacteriaceae. Int J Syst Evol Microbiol 56, 959-963.

13. Kolton, M., Sela, N., Elad, Y., Cytryn, E., 2013. Comparative genomic analysis indicates that niche adaptation of terrestrial Flavobacteria is strongly linked to plant glycan metabolism. PLoS One 8, e76704.

14. Kopel, M., Helbert, W., Henrissat, B., Doniger, T., Banin, E., 2014. Draft Genome Sequence of Nonlabens ulvanivorans, an Ulvan-Degrading Bacterium. Genome Announc 2.

15. Kozdon, J.B., Melfi, M.D., Luong, K., Clark, T.A., Boitano, M., Wang, S., Zhou, B., Gonzalez, D., Collier, J., Turner, S.W., Korlach, J., Shapiro,

L., McAdams, H.H., 2013. Global methylation state at base-pair resolution of the Caulobacter genome throughout the cell cycle. Proc Natl Acad Sci U S A 110, E4658-4667.

16. Kwon, S.K., Kim, B.K., Song, J.Y., Kwak, M.J., Lee, C.H., Yoon, J.H., Oh, T.K., Kim, J.F., 2013. Genomic makeup of the marine flavobacterium Nonlabens (Donghaeana) dokdonensis and identification of a novel class of rhodopsins. Genome Biol Evol 5, 187-199.

17. Lagesen, K., Hallin, P., Rodland, E.A., Staerfeldt, H.H., Rognes, T., Ussery, D.W., 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 35, 3100–3108.

18. Lai, Q., Cao, J., Yuan, J., Li, F., Shao, Z., 2014. Celeribacter indicus sp. nov., a polycyclic aromatic hydrocarbon-degrading bacterium from deep-sea sediment and reclassification of Huaishuia halophila as Celeribacter halophilus comb. nov. Int J Syst Evol Microbiol 64, 4160-4167.

19. Lau, S.C., Tsoi, M.M., Li, X., Plakhotnikova, I., Dobretsov, S., Wong, P.K., Qian, P.Y., 2005. Gramella portivictoriae sp. nov., a novel member of the family Flavobacteriaceae isolated from marine sediment. Int J Syst Evol Microbiol 55, 2497–2500.

20. Lawrence, C.E., Altschul, S.F., Boguski, M.S., Liu, J.S., Neuwald, A.F., Wootton, J.C., 1993. Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. Science 262, 208-214.

21. Lee, S.Y., Park, S., Oh, T.K., Yoon, J.H., 2012. Celeribacter baekdonensis sp. nov., isolated from seawater, and emended description of the genus Celeribacter Ivanova et al. 2010. Int J Syst Evol Microbiol 62, 1359–1364.

22. Loenen, W.A., Raleigh, E.A., 2014. The other face of restriction: modification-dependent enzymes. Nucleic Acids Res 42, 56-69.

23. Lowe, T.M., Eddy, S.R., 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25, 955-964.

24. Nedashkovskaya, O.I., Kim, S.B., Lysenko, A.M., Frolova, G.M., Mikhailov, V.V., Bae, K.S., Lee, D.H., Kim, I.S., 2005. Gramella echinicola gen. nov., sp. nov., a novel halophilic bacterium of the family Flavobacteriaceae isolated from the sea urchin Strongylocentrotus intermedius. Int J Syst Evol Microbiol 55, 391–394. 25. Nedashkovskaya, O.I., Suzuki, M., Vysotskii, M.V., Mikhailov, V.V., 2003. Vitellibacter vladivostokensis gen. nov., sp. nov., a new member of the phylum Cytophaga-Flavobacterium-Bacteroides. Int J Syst Evol Microbiol 53, 1281–1286.

26. Park, S., Yoshizawa, S., Chiura, H.X., Muramatsu, Y., Nakagawa, Y., Kogure, K., Yokota, A., 2012. Nonlabens marinus [corrected] sp. nov., a novel member of the Flavobacteriaceae isolated from the Pacific Ocean. Antonie Van Leeuwenhoek 102, 669-676.

27. Sang, M.K., Kim, K.D., 2012. The volatile-producing Flavobacterium johnsoniae strain GSE09 shows biocontrol activity against Phytophthora capsici in pepper. J Appl Microbiol 113, 383-398.

28. Suzuki, M., Nakagawa, Y., Harayama, S., Yamamoto, S., 2001. Phylogenetic analysis and taxonomic study of marine Cytophaga-like bacteria: proposal for Tenacibaculum gen. nov. with Tenacibaculum maritimum comb. nov. and Tenacibaculum ovolyticum comb. nov., and description of Tenacibaculum mesophilum sp. nov. and Tenacibaculum amylolyticum sp. nov. Int J Syst Evol Microbiol 51, 1639-1652.

29. Tatusov, R.L., Galperin, M.Y., Natale, D.A., Koonin, E.V., 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res 28, 33–36.

30. Touchon, M., Barbier, P., Bernardet, J.F., Loux, V., Vacherie, B., Barbe, V., Rocha, E.P., Duchaud, E., 2011. Complete genome sequence of the fish pathogen Flavobacterium branchiophilum. Appl Environ Microbiol 77, 7656-7662.

31. Yoon, J.H., Kang, S.J., Oh, T.K., 2005. Tenacibaculum lutimaris sp. nov., isolated from a tidal flat in the Yellow Sea, Korea. Int J Syst Evol Microbiol 55, 793-798.

32. Zhang, Y., Liu, J., Tang, K., Yu, M., Coenye, T., Zhang, X.H., 2015. Genome analysis of Flaviramulus ichthyoenteri Th78(T) in the family Flavobacteriaceae: insights into its quorum quenching property and potential roles in fish intestine. BMC Genomics 16, 38.



Acknowledgement

벌써 2 년이라는 시간이 지나고, 이렇게 논문을 제출하기 전이니 저에게 있어서 시간은 엄청나게 빨리 지나가는 손님 같습니다. 이 짧은 시간 가운데서도 많은 사람들을 만났고, 그들을 통해 많은 것을 배우고 희노애락을 같이 할 수 있었음에 감사의 마음을 전합니다.

먼저, 논문이 이렇게 완성되기까지 저를 지켜봐 주시고, 많은 지식을 쌓을 수 있도록 환경을 만들어 주시고, 잘못한 점이 있다면 충고로 끊임없이 도움을 주시고 지도하여주신 오현명 교수님께 진심으로 감사의 말씀을 드립니다. 부족한 저를 잘 인도해 주셨기에 이렇게 석사과정을 다 마쳐갑니다. 앞으로 여기서 배운 지식과 더불어 뭐든 더 열심히 노력하여 교수님의 자랑스러운 제자가 되도록 노력하겠습니다.

둘째로, 열정적인 수업으로 많은 지식을 쌓을 수 있도록, 발표는 어떤식으로 해야하는지 구체적으로 가르쳐 주어 논문발표 때 뿐 아니라 사회에서도 꼭 필요한 자세들을 많이 가르쳐 주신 오현명 교수님, 제제영 교수님께 진심으로 감사의 말씀을 드립니다. 이

- 61 -

논문이 쓰여지기까지 많은 관심과 후원을 해 주신 유영문 교수님, 초안을 끝까지 읽어주시고 관심을 가지고 심사를 해 주신 김경호 교수님과 이대원 교수님께도 감사의 마음을 전합니다.

셋째로, 논문의 첫 챕터를 쓸 때 많은 도움을 주시고, 같이 고민하고 그 고민을 잘 해결할 수 있도록 그리고 해외학회 때도 많은 도움을 주신 문미경 선생님, 류우찬 교수님께도 감사의 마음을 전합니다. 또한, 실험을 도와주고 여러방면에서 챙겨주며 실험에 대해 같이 고민해주며 같이 공부하면서 함께 해 준 정희, 준학이형 그리고 2 년동안 행정으로 많은 도움을 주신 박태인 주무관님께 진심어린 감사를 표합니다.

마지막으로 진심으로 응원해주시고, 용기를 주시고 사랑으로 챙겨주신 부모님, 나를 위해 기도해 주신 3 영도교회 성도님들에게도 감사를 전합니다.

그 외에 많은 분들의 관심과 사랑이 있었기에 논문이 완성될 수 있었습니다. 감사의 말씀을 전하며, 앞으로 더욱 발전하는 모습 보여드리겠습니다. 감사합니다.