GENOME CHARACTERIZATION OF *HAHELLA CHEJUENSIS* STRAIN HN01, A PATHOGEN OF RED EGG DISEASE IN TILAPIA (*OREOCHROMIS SP.*)

Miss Putu Cri Devischa Gallantiswara

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Veterinary Science and technology
Common Course
Faculty of Veterinary Science
Chulalongkorn University
Academic Year 2018

การจำแนกคุณลักษณะทางจีโนมของ *ฮาเฮลล่า เจจูเอ็นซิส* สายพันธุ์ เอชเอ็น 01
เชื้อแบคทีเรียก่อโรคเรดเอ้ก ในปลานิล

น.ส.พูตู ศริ เดวิสชา กัลลันติสวารา

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาศาสตร์ทางการสัตวแพทย์และเทคโนโลยี ไม่สังกัดภาควิชา/เทียบเท่า
คณะสัตวแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2561

Thesis Title	GENOME CHARACTERIZATION OF *HAHELLA CHEJUENSIS* STRAIN HN01, A PATHOGEN OF RED EGG DISEASE IN TILAPIA (*OREOCHROMIS SP.*)
By	Miss Putu Cri Devischa Gallantiswara
Field of Study	Veterinary Science and technology
Thesis Advisor	Assistant Professor CHANNARONG RODKHUM, D.V.M., Ph.D., D.T.B.V.P
Thesis Co Advisor	Pattanapon Kayansamruaj, D.V.M., Ph.D.

Accepted by the Faculty of Veterinary Science, Chulalongkorn University in Partial Fulfillment of the Requirement for the Master of Science

..... Dean of the Faculty of Veterinary Science
(Professor ROONGROJE THANAWONGNUWECH, D.V.M., Ph.D., D.T.B.V.P.)

THESIS COMMITTEE

..... Chairman
(Associate Professor NAPADON PIRARAT, D.V.M., Ph.D., D.T.B.V.P)

..... Thesis Advisor
(Assistant Professor CHANNARONG RODKHUM, D.V.M., Ph.D., D.T.B.V.P)

..... Thesis Co-Advisor
(Pattanapon Kayansamruaj, D.V.M., Ph.D.)

..... Examiner
(Teerapong Yata, M.Sc., Ph.D.)

..... External Examiner
(Ha Thanh Dong, M.Sc., Ph.D.)

พูตู คริ เดวิสชา กัลลันติสวารา : การจำแนกคุณลักษณะทางจีโนมของ *ฮาเฮลล่า เจจูเอ็นซิส* สายพันธุ์ เอชเอ็น 01 เชื้อแบคทีเรียก่อโรคเรดเอ้ก ในปลานิล. (GENOME CHARACTERIZATION OF *HAHELLA CHEJUENSIS* STRAIN HN01, A PATHOGEN OF RED EGG DISEASE IN TILAPIA (*OREOCHROMIS SP.*)) อ.ที่ปรึกษาหลัก : ชาญณรงค์ รอดคำ, อ.ที่ปรึกษาร่วม : พัฒนพล ขยันสำรวจ

โรคเรดเอ้ก เป็นโรคอุบัติใหม่ที่เกิดขึ้นกับฟาร์มเพาะเลี้ยงปลานิล ซึ่งมีสาเหตุเกิดจากเชื้อแบคทีเรียแกรมลบที่ถูกเรียกว่า ฮาเฮลล่า เจจูเอ็นซิส (*Hahella chejuensis*) โรคนี้สามารถเกิดขึ้นได้กับไข่ปลาทุกระยะและทำให้ลดอัตราการเพาะฟักลง อย่างไรก็ตามการศึกษาเกี่ยวกับความรุนแรงในการก่อโรคของเชื้อนี้ยังมีอยู่ค่อนข้างจำกัด ในการศึกษานี้เราได้วิเคราะห์ความสัมพันธ์ทางอนุกรมวิธานและศึกษาจีนปัจจัยความรุนแรงในการก่อโรคของเชื้อแบคทีเรียฮาเฮลล่า เจจูเอ็นซิส สายพันธุ์ เฮชเอ็น 01 โดยใช้วิธีการวิเคราะห์รายละเอียดในจีโนมของเชื้อ จีโนมของเชื้อถูกหาลำดับนิวคลีโอไทด์โดยใช้ Illumina Miseg plateform จากการวิเคราะห์ Phylogenetic tree โดยอาศัยจีน housekeeping หลายจีนสามารถยืนยันได้ว่าแบคทีเรียนี้อยู่ในจีนัส ฮาเฮลล่า และ สปีชีส์ เจจูเอ็นซิส ค่าเฉลี่ยการคำนวณความเหมือนของลำดับนิวคลีโอไทด์กับเชื้อสายพันธุ์อ้างอิงได้แก่ ฮาเฮลล่า จีจูเอ็นซิส เคซีทีซี 2396 คือ 89.40 % ซึ่งค่านี้แสดงให้เห็นถึงความแตกต่างระหว่างจีโนมของแบคทีเรียทั้ง 2 สายพันธุ์ การทำ annotation กับ Virulence Factor Data Base พบว่า ภายในจีโนมของฮาเฮลล่า เจจูเอ็นซิส สายพันธุ์ เฮชเอ็น 01 มีจีนปัจจัยก่อความรุนแรงที่จำเพาะอยู่ถึง 19 จีน ได้แก่ capsule, pigments, adhesion proteins, enzymes, toxins, และ secretion system productions เป็นที่น่าสนใจว่าภายในจีโนมของฮาเฮลล่า เจจูเอ็นซิส สายพันธุ์ เฮชเอ็น 01 พบจีน 2 จีนที่มีความคล้ายคลึงสูงกับจีน toxin ในวิบริโอ คลอเรอเร่ ได้แก่ cholera enterotoxin (*ctx*A) และ zona occludens toxin (*zot*) การศึกษานี้ได้ค้นพบสิ่งใหม่ที่นำมาซึ่งจุดเริ่มต้นในการศึกษาคุณสมบัติทางพันธุกรรมในการก่อความรุนแรงของเชื้อฮาเฮลล่า เจจูเอ็นซิส สายพันธุ์ เฮชเอ็น 01

สาขาวิชา วิทยาศาสตร์ทางการสัตวแพทย์และเทคโนโลยี ลายมือชื่อนิสิต .....

ปีการศึกษา 2561 ลายมือชื่อ อ.ที่ปรึกษาหลัก .....

ลายมือชื่อ อ.ที่ปรึกษาร่วม .....

6075402731 : MAJOR VETERINARY SCIENCE AND TECHNOLOGY

KEYWORD: Genome, Hahella chejuensis, Red Eggs Disease, Tilapia

Putu Cri Devischa Gallantiswara : GENOME CHARACTERIZATION OF *HAHELLA CHEJUENSIS* STRAIN HN01, A PATHOGEN OF RED EGG DISEASE IN TILAPIA (*OREOCHROMIS SP.*) . Advisor: Asst. Prof. CHANNARONG RODKHUM, D.V.M., Ph.D., D.T.B.V.P Co-advisor: Pattanapon Kayansamruaj, D.V.M., Ph.D.

Red eggs disease is an emerging disease occurred in Tilapia hatcheries which caused by a Gram-negative bacterium called *Hahella chejuensis* HN01. The disease infects to all the stage of fish eggs and reduces the hatchability. However, there are limited study about the bacterial pathogenicity. In this study, we analyzed the phylogenetic relationship and characterized the virulence traits of *Hahella chejuensis* by using genomic approach. The bacterial genome was sequenced using Illumina Miseq platform. The phylogenetic tree based on several housekeeping genes confirmed that the bacterium was belong to Genus *Hahella* and species *chejuensis*. the Average Nucleotide Identity calculation with the reference strain, *Hahella chejuensis* KCTC2396 was 89.40% which still indicates the variability between two genomes. The annotation against the Virulence Factor Data Base revealed the genome of *Hahella chejuensis* HN01 specifically harbors 19 virulence genes contributed to capsules, pigments, adhesion proteins, enzymes, toxins, and secretion system productions. Interestingly, two genes highly similar to toxin genes in *Vibrio cholerae*; cholera enterotoxin (*ctxA*) and zona occludens toxin (*zot*), were found exclusively in the strain HN01. This study brings a new insight in preliminary exploring the genotypic virulence properties of *Hahella chejuensis* HN01.

Field of Study:	Veterinary Science and technology	Student's Signature .....
Academic Year:	2018	Advisor's Signature .....
		Co-advisor's Signature .....

ACKNOWLEDGEMENTS

This thesis will not come true without the support of every party. My deepest gratitude for my advisor, Assistant Professor Channarong Rodkhum, D.V.M., Ph.D., D.T.B.V.P., and Instructor Pattanapon Kayansamruaj, Ph.D., for their excellent guidance, advice and encouragement throughout the period of my study. Associate Professor Napadon Pirarat, D.V.M., Ph.D., D.T.B.V.P., the chairman of thesis committee for his splendid guidance and advice. Mr. Teerapong Yata, Ph.D, the thesis committee, for their valuable suggestions and advices. Instructor Ha Thanh Dong, Ph.D., external thesis examiner, for his kind support and patientness.

Also ASEAN Scholarship of Chulalongkorn University that has provided the funding for Master degree. The 90th Anniversary of Chulalongkorn University Scholarship that has provided the research funding. And all my colleagues in Department of Veterinary Microbiology, Fish Infectious Diseases Research Unit (FIDs RU), international students of Veterinary Science and Technology Program and all staff whose kindly helpfull.

Last but not the least, my beloved family, my biggest motivation.

Putu Cri Devischa Gallantiswara

TABLE OF CONTENTS

Page

ABSTRACT (THAI) iii

ABSTRACT (ENGLISH) iv

ACKNOWLEDGEMENTS v

TABLE OF CONTENTS vi

LIST OF TABLES ix

LIST OF FIGURES x

LIST OF ABBREVIATIONS 1

CHAPTER I. INTRODUCTION 2

1. Importance and rationale 2
2. Hypothesis 5
3. Objectives of Study 5

CHAPTER II. LITERATURE REVIEW 6

1. Tilapia (*Oreochromis sp.*) is an important fish in worldwide aquaculture 6
2. The management of tilapia hatcheries 9
3. Red Egg Disease is a novel disease reported in Thailand Tilapias hatcheriess .. 10
4. *Hahella chejuensis,* the causative agent of Red Eggs Disease in Tilapia eggs ... 11
5. General concept of bacterial virulence properties 13
6. General concept of bacterial genome characterization 16

CHAPTER III. MATERIAL AND METHODS 20

1. Genome sequencing preparation 21

    1.1 Bacterial culture condition and DNA extraction 21

1.2 Library preparation, sequencing and assembly 21
2 Phylogenomic study 21
2.1 Digital DNA-DNA Hybridization (DDH), Average Nucleotide Identity (ANI) estimation, and Phylogenetic analysis 21
3. Genome characterization and comparison 26
3.1. Genomic features *of Hahella chejuensis* HN01 and *Hahella chejuensis* KCTC2396 26
3.2. Putative virulence properties identification 26
3.3. Other genome elements prediction 27
4. Data Analysis 27
CHAPTER IV. RESULT 28
1. The phylogenomic analysis of *Hahella chejuensis* HN01 28
2 Genome characterization and comparison 33
2.1 Genomic features *of H. chejuensis* HN01 and *H. chejuensis* KCTC2396 33
2.2 Putative virulence properties identification 37
CHAPTER V. DISCUSSION AND CONCLUSION 47
1. Discussion 47
2. Conclusions and Suggestions 53
3. Advantages of the study 53
REFERENCES 55
VITA 67
APPENDIX 68
Appendix 1. The media for *Hahella chejuensis* culture 68
Appendix 2. The best model applied in every house keeping gene based phylogenetic tree 68

Appendix 3. COG Automatic Classification of Hahella sp. HN01 and Hahella chejuensis KCTC2396 69

Appendix 4. Kyoto Encyclopedia Genes and Genomes (KEGG) Functional Annotations 71

Appendix 5. The putative virulence genes listed by VFDB of *Hahella chejuensis* HN01 and KCTC3296 78

LIST OF TABLES

Page

Table 1. Worldwide Tilapia Production (thousand tonnes) (FAO, 2018a) 8

Table 2 Production of Inland Freshwater Fish in Thailand from 2012-2016 (thousand tonnes) (FAO, 2018a)..... 9

Table 3 Phenotypic characteristics of Hahella chejuensis KCTC2396 (Lee et al., 2001) 12

Table 4. Member of Proteobacteria used in phylogenomic analysis..... 24

Table 5 *Oceanospirillales* member used in the phylogenomic analysis 25

Table 6. Genome profile of H. chejuensis. HN01 and H. chejuensis KCTC2396..... 35

Table 7 Putative virulence gene identified as unique gens in Hahella chejuensis HN01 40

Table 8 Putative virulence gene identified as unique gens in *Hahella chejuensis* KCTC2396 42

LIST OF FIGURES

Page

Figure 1 Conceptual Framework 20

Figure 2 A maximum likelihood tree based on almost complete 16S rRNA gene sequences of phylum Proteobacteria member. 29

Figure 3. The phylogenetic relationship of strain HN01 among other *Oceanospirillales* member based on complete 16S rRNA sequence: 30

Figure 4. The phylogenetic relationship of strain HN01 among other *Oceanospirillales* member based on complete *rpoD* sequence: 30

Figure 5. The phylogenetic relationship of strain HN01 among other *Oceanospirillales* member based on complete *atpD* sequence: 31

Figure 6. The phylogenetic relationship of strain HN01 among other Oceanospirillales member based on complete infB sequence: 31

Figure 7 The phylogenetic relationship of strain HN01 among other Oceanospirillales member based on complete recA sequence: 32

Figure 8. The maximum likelihood tree generated from concatenated housekeeping sequences. 32

Figure 9. MicroScope functional categories of H. chejuensis HN01 and H. chejuensis KCTC2396 36

Figure 10. Circular genome retrieved from GenoScope visualized by CGView 37

Figure 11. Pan-genome analysis of Hahella chejuensis KCTC2396 and Hahella chejuensis. HN01 37

Figure 12 Two regions of prophage found in Hahella chejuensis HN01 genome by PHASTER webtool 45

Figure 13 Four regions of prophage found in *Hahella chejuensis* KCTC2396 genome by PHASTER webtool 46

LIST OF ABBREVIATIONS

ANI	Average Nucleotide Identity
ASFIS	Aquatic Sciences and Fisheries Information System
BLAST	Basic Local Alignment Search Tool
CARD	Comprehensive Antibiotic Resistence Database
CDSs	Coding Sequences
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
dDDH	Digital DNA DNA Hybridization
GI	Genomic Island
KEGG	Kyoto Encyclopedia Genes and Genomes
LCB	Local Colinear Block
MICFAM	MicroScope gene Families
MLSA	Multilocus Sequence Analysis
NGS	Next Generation Sequencing
PAIDB	Pathogenicity Island Database
RAST	Rapid Annotation Subsystem Technologies
TSA	Tryptic Soy Agar
TTSSs	Type III Secretion Systems
VFDB	Virulence Data Base

CHAPTER I. INTRODUCTION

1. Importance and rationale

As the increasing of human consumption towards aquaculture products, it is also contributing to the increasing of aquaculture production. Of 80.0 million tonnes (73%) from 110.2 million tonnes global aquaculture is belong to food fish which mostly are supported by freshwater aquaculture. Tilapia is one of the favorite fish species in worldwide freshwater aquaculture (FAO, 2018b). The fish has a good characteristic for farming, including fast-growing, wide range of food types, and able to survive in poor water condition (Amal and Zamri-Saad, 2011). In Thailand, the production of Tilapia is the second highest after white shrimp. Based on Thailand Department of Fisheries report in 2009, the production of Tilapia achieved 221,042 ton/year from 521,880 ton/year of total inland aquaculture production (Ferreira et al., 2015).

However, the high fish production can lead to many disease outbreaks in fish farms. The high stocking densities will trigger rapid infection transmission and drive clinical disease. The disease can be spread by water column from ponds or cages within a farm and between farms. There are many pathogen have been reported causing disease outbreak in Tilapia farm, including *Streptococcus agalactiae*, *Flavobacterium columnare, Aeromonas veronii, Francisella noatunensis* subsp. *orientalis,* and *Edwardsiella ictaluri (Dong et al., 2015a; Dong et al., 2015b)*. Tilapia Lake Virus and Megalocytivirus also has been reported causing high mortality in growing Tilapia (Subramaniam et al., 2012; Dong et al., 2017).

While there are many diseases are reported in growing stage fish, in 2016, a novel disease was reported in Thailand tilapia hatcheries systems. It is also claimed as the first disease found in freshwater fish eggs. The disease was typical with the change of infected eggs color from normal yellow to be red and reduce the eggs

hatchability rate, thus called as Red Eggs Disease. The fry production loss is 10% and will increase during cold season up to 50%. Red Eggs Disease was caused by a Gram-negative bacterium, *Hahella chejuensis* (Senapin et al., 2016).

Hahella chejuensis is belong to marine bacteria group, though it has been reported to cause a disease in freshwater system. *Hahella chejuensis* is a halophilic group that the first species was firstly discovered in the coastal marine sediment of Cheju Island, South Korea, as *Hahella chejuensis* (Lee et al., 2001). *Hahella chejuensis* produces a red pigment or also known as prodigiosin. The pigment has lytic activity towards certain microalgae in the ocean. Study about genomic blueprint of *H. chejuensis* revealed that the bacterium produces a large number of extracellular polysaccharides which has been known responsible for the development of biofilms and often act as a virulence factor of pathogenic bacteria (Jeong et al., 2005). The occurrence of the bacterium among Thailand tilapia hatcheries may be related to salt usage in the system. The condition may enable bacterial colonization and contamination (Senapin et al., 2016).

Although *Hahella chejuensis* has been reported as a causative agent of red egg disease and marine algicidal agent, the molecular mechanism of red egg disease and bacterial virulence determinants are yet to be understood. The number of studies about *H. chejuensis* causing red egg disease in fish is very limited. The information of virulence properties of *H. chejuensis* to fish egg is also still unclear (Jeong et al., 2005; Senapin et al., 2016).

Since the introduction of high-throughput sequencing or Next Generation Sequencing (NGS) in 2005, it increases the number of sequence data thousands of times in one sequence run rather than Sanger sequencing. NGS delivers bacterial genome sequencing faster (in hours or days) and cheaper (Loman et al., 2012). Different with Sanger sequencing which need specific primer and certain condition for

each pathogen, NGS can be applied for all pathogens with a single protocol. Since then, many complete and draft genomes submitted in online database, such as National Center for Biotechnology Information (NCBI, www.ncbi.nlm.nih.gov), Virulence Data Base (VFDB), Comprehensive Antibiotic resistance Database (CARD), and many more (Chen et al., 2005; Dark, 2013; Jia et al., 2017). This development of sequencing technology has greatly improved the understanding of bacterial genome arrangement and its genomic contents (Loman and Pallen, 2015). NGS has been used widely in clinical microbiology and infection prevention, such as outbreak management, molecular case finding, characterization and surveillance of pathogens, rapid bacterial identification in clinical specimen, pathogen taxonomy, and determining the transmission of zoonotic microorganism (Meier-Kolthoff et al., 2013; Yoon et al., 2017)

NGS data can be analyzed with bioinformatic tools to access many information related to the bacterial lifestyle, such as metabolism, antimicrobial resistance determinant and virulence associated gene (Burrack and Higgins, 2007; Rouli et al., 2015; McDermott et al., 2016). There are many available tools that are free, accessible, and user friendly (Edwards and Holt, 2013). Genome annotation is 'gene-finding' process which can be performed by web-based tools RAST or MicroScope (Aziz et al., 2008; Vallenet et al., 2009). Genome comparison can be done with a series of computational tool, such as MAUVE, MEGA, Artemis and many more (Rutherford et al., 2000; Darling et al., 2004; Kumar et al., 2016). Therefore, in this study we combine NGS data of *Hahella chejuensis* HN01 with bioinformatics which can bring a new insight in preliminary exploring the genotypic virulence properties of *Hahella chejuensis* HN01.

2. Hypothesis

There are several changes inside of *Hahella chejuensis.* strain HN01 genome which may affect to the bacterial virulence associated genes

3. Objectives of Study

To characterize the bacterial virulence genes of *Hahella chejuensis* strain HN01 using a genomic approach.

CHAPTER II. LITERATURE REVIEW

1. Tilapia (*Oreochromis sp.*) is an important fish in worldwide aquaculture

Tilapias refer to a group of fish within the family Cichlidae. The term of Tilapia was designated as genus level by Smith (1840). There are more than 70 species of Tilapias around the world and they have similar morphology and undetermined characterization. Even though several alternatives classification had been introduced, however many taxonomists and scientist are still using *Tilapia* genus to name all tilapias species. Nile tilapia (*Oreochromis niloticus*) is the most popular Tilapias among worldwide aquaculture, besides that other species identified in Tilapias genus including Redbelly tilapia (*Tilapia zillii*), Blue tilapia (*Oreochromis aureus*), Galilee tilapia (*Sarotherodon galilaeus*), Mozambique tilapia (*Oreochromis mossambicus*), and also including Tilapia hybrid species (*Oreochromis mossambicus* x *Oreochromis niloticus*) (El-Sayed, 2006; Ferreira et al., 2015).

In 2016, Tilapias (Nile tilapia and other Tilapias) production achieved 10% of world finfish aquaculture. This number was the second highest after Carps (Grass carp, Silver carp, and other common carp) which achieved up to 37% of the 54,091 tonnes world aquaculture production. Furthermore, the development of Tilapias farming was quite interesting in recent four years. Its production was around 3,165 tonnes in 2010 and gradually increased to be more than 4000 tonnes in 2016. Tilapias have good characteristics and easy to grow, therefore they do not need a complicated system for culturing and cost effective. Moreover, the trend of fish or seafood eating is greatly increasing nowadays, and this phenomenon indirectly has contributed to Tilapias production (El-Sayed, 2006; Amal and Zamri-Saad, 2011; FAO, 2018b).

At the beginning, Tilapias habitat was originally from Egypt and come countries in Africa. Several years after its first introduction to over Africa, then to South and Central America, and some parts of Asia, the fish has been spread around the world,

especially in East and South East Asia (El-Sayed, 2006; Wang and Lu, 2016). Nowadays, China has accounted for about 28% (1.4 million tonnes) of world Tilapia production for over five years, then followed by Indonesia, Egypt, Bangladesh, Brazil, Thailand, Viet Nam, Philippines, and Mexico (Table 1) (FAO, 2018a).

As one of main tilapia supplier country in South East Asia, Thailand's Tilapias is the most inland freshwater fish production. Even though the number was not consistently increase, however Tilapia production is still dominating the countries commodity (Table 2). Of the 384 thousand tonnes freshwater fish production, tilapias production achieves 208 thousand tonnes in 2016. After Nile tilapia, there are Africa-bighead carp (112 thousand tonnes), Silver barb (31 thousand tonnes), Striped catfish (19 thousand tonnes), Snake Skin gourami (14 thousand tonnes), and other freshwater species (15 thousand tonnes) (FAO, 2018a).

Commonly, Tilapias are grown alone (monoculture system) in semi or intensive culture which is believed can produce high yield of fish. However, polyculture system has been spreading widely in many countries. The trend of polyculture arose since it can increase the profit, improve the health status, and make efficient feeding (Wang and Lu, 2016). Many farms among countries culture Tilapias together with prawn like in Amazon river, Brazil, goose in China, tiger shrimp in Philippine, or white shrimp in Thailand (Cruz et al., 2008; Ferreira et al., 2015; Zhou et al., 2018; Rodrigues et al., 2019). For the last two polyculture can be applied in brackish water system. Even though they are known as freshwater fish, but Tilapias also can adapt in high salinity water up to 35 ppt (Suresh and Lin, 1992).

Table 1. Worldwide Tilapia Production (thousand tonnes) (FAO, 2018a)

No	Country	Species[a]	Production source[b]	2012	2013	2014	2015	2016
1	China	Nile tilapia	Aquaculture production (freshwater)	1.165	1.243	1.278	1.334	1.400
2	Indonesia	Nile tilapia	Aquaculture production (freshwater)	661	833	947	1.039	1.102
3	Egypt	Nile tilapia	Aquaculture production (brackishwater)	719	601	715	800	840
4	China	Blue-Nile tilapia, hybrid	Aquaculture production (freshwater)	388	414	420	445	466
5	Bangladesh	Tilapias nei[c]	Aquaculture production (freshwater)	124	210	284	324	343
6	Brazil	Tilapias nei[c]	Aquaculture production (freshwater)	182	169	200	219	239
7	Thailand	Nile tilapia	Aquaculture production (freshwater)	203	198	190	206	208
8	Viet Nam	Tilapias nei[c]	Aquaculture production (freshwater)	197	216	244	283	184
9	Philippines	Nile tilapia	Aquaculture production (freshwater)	161	165	165	164	157
10	Mexico	Tilapias nei[c]	Capture production	56	70	72	83	122
11	Others	Others	Others	1.415	1.464	1.526	1.505	1.626
	Total			3.855	4.119	4.515	4.898	5.061

[a]Aquatic Sciences and Fisheries Information System (ASFIS) species

[b]FAO major fishing area

[c]Tilapias nei (not elsewhere included), it is the FAO term to classify Tilapias which has not been identified in species level yet

Table 2 Production of Inland Freshwater Fish in Thailand from 2012-2016 (thousand tonnes) (FAO, 2018a)

No	Species[a]	2012	2013	2014	2015	2016
1	Nile tilapia	203	198	190	206	208
2	Africa-bighead catfish, hybrid	124	120	114	114	112
3	Silver barb	33	30	29	30	31
4	Striped catfish	26	23	23	19	19
5	Snake skin gourami	27	27	23	15	14
6	Others	17	16	17	15	15
Total		414	397	378	385	384

[a]Aquatic Sciences and Fisheries Information System (ASFIS) species

2. The management of tilapia hatcheries

In the semi intensive culture, growing stage of Tilapias are feed by natural food in a fertilized pond. This method is suitable in a rural area and small-scale farmer. Usually not only Tilapias, but also another omnivore or herbivore fish such as carps are cultured in the same pond. Many farmers have applied more artificial feeding, water reused system, and biosecurity in order to maximize the production. There are a lot of intensive culture developed in earthen ponds, tanks, cages, raceways, recirculating and aquaponic systems. It needs around 73 – 220 days per each culture period in earthen ponds or can be faster in intensive cages (El-Sayed, 2006).

Tilapia hatcheries can be operated in several type of cages, such as concentrate tanks, hapa net cages, or in earth ponds. The natural cages are the seasonal dependent management, where the temperature cannot be adjusted and only rely on the changing climate. Sexually, the Tilapia broodstocks are mature for breeding on three-month-old or the broodfish weight is around 150-250 g and 20-30 cm in length (El-Sayed, 2006). Within the breeding system, a male Tilapia is introduced with two until three female broodfish in the same pond. Normally, the

female can spawn up to 1,000 eggs depending on its size. The male constructs nests on the bottom of pond to attract the females (Surtida, 1998; El-Sayed, 2006).

The fertilized eggs will be incubated in female's mouth or artificial incubation. Commonly, spawning can be carried in earthen ponds, tanks, or hapas. During the artificial incubation, eggs will be incubated in aerated ponds or tanks with a high circulation for 3 days until they are hatched. The hatched fry are moved to another container for hormonal sex reversion (21 days) and nursed with general feed for 7 – 10 more days before they sold (Rakocy, 2005; El-Sayed, 2006; Uppanunchai et al., 2015). According to , it is common to use saline at 5-10 ppt in Thailand's hatchery system in order to prevent parasite infestation (Senapin et al., 2016).

The most important factors to drive the successful of Tilapias culture are water quality (including pH, temperature, ammonia, metabolite residue, etc.), nutrition and feeding management, and stocking densities (El-Sayed, 2006). Mismanagement that occur during the intensive culture can cause many problems, especially disease outbreak. There are many infectious diseases caused by bacteria, virus, fungi, and parasite that have been identified in most of growing stage fish (Noga, 2010).

3. Red Egg Disease is a novel disease reported in Thailand Tilapias hatcheriess

Hahellosis or red egg disease is a novel disease reported in Tilapias hatcheries. It decreases the hatchability rate of Tilapia eggs and occurs at all stages of fish eggs during incubation periods. The disease is typical with the change color of fish eggs from yellow to red. Up to date, it is reported only infect to Red Tilapia and Nile Tilapia eggs, and firstly detected among hatcheries in middle part of Thailand in 2000. Mostly, the disease outbreak occurs during a cold season when the temperature is under 24°C, around December to February. It is caused by a Gram-negative bacteria, *Hahella chejuensis.* (Senapin et al., 2016).

The previous study reported a reduced hatchability of the eggs in the challenged group compare to the control on day two (4.0% and 8.5% respectively) and day three (10.5% and 20.0% respectively). It also successfully identified the presence of the bacteria in broodstocks gonad tissue. However, there is still no symptom or disease observed in living fish. The pathogenesis of Red Eggs Disease is still not clearly understood (Senapin et al., 2016).

Indeed, there were two bacterial species already successfully identified as a pathogen in marine fish eggs, *Tenacibaculum ovolyticus* (*Flexibacter ovolyticus*) (Hansen et al., 1992; Suzuki et al., 2001) and *Pseudoalteromonas piscicida* (Nelson and Ghiorse, 1999). *Tenacibaculum ovolyticus* is a pathogen in Atlantic Halibut (*Hippoglossus hippoglossus* L.) that usually occurs at the early and late stage of hatching eggs. The bacterium cause damage on the egg surface through exoenzymatic activity (Bergh et al., 1992). While *Pseudoalteromonas piscicida* is an opportunistic pathogen in damselfish (*Amblyglyphidodon clarkia*) eggs (Nelson and Ghiorse, 1999). Several Gram-negative bacteria are also found on died Turbots' eggs surface (*Scophthalmus maximus*), including *Aeromonas hydrophila, Moxarella sp., Pseudomonas aeruginosa,* and *Pseudomonas fluorescens.* Nonetheless, there is no clear explanation about their direct role to dead Turbots' eggs (Keskin et al., 1994).

4. *Hahella chejuensis,* the causative agent of Red Eggs Disease in Tilapia eggs

Hahella chejuensis is a member of Proteobacteria phylum, *Gammaproteobacteria* class, *Oceanospirillales* order, and *Hahellaceae* family (Lee et al., 2001; Jeong et al., 2005; Brenner et al., 2008; Senapin et al., 2016). It is a Gram-negative, rod-shaped, facultatively anaerobic, and motile bacterium. The color of its colony is pale orange when it is still young and changed to be pinkish red when it is becoming older. The bacterium can use some group of carbohydrates as carbon sources and produce acid. It can reduce nitrate to nitrite and hydrolyzes esculin and

gelatin. It grows at 10-45°C with the presence of NaCl around 1-8% (optimally with 2%) at pH 6-10 (optimally at pH7), therefore *Hahella chejuensis* is also called as halophilic bacterium (Table 3) (Lee et al., 2001; Jeong et al., 2005; Soliev et al., 2011).

Table 3 Phenotypic characteristics of *Hahella chejuensis* KCTC2396 **(Lee et al., 2001)**

Characteristics	*Hahella chejuensis* KCTC2396
Motility	+, with a single polar flagellum
Catalase	+
Oxidase	+
Indole production	-
H_2S production	-
Utilization of sole carbon:	
Adonitol	+/+
Arabinose	-/-
Cellobiose	+/ND
Citrate	-/ND
Fructose	+/+
Galactose	-/-
Glucose	+/+
Glycerol	+/ND
Inositol	+/+
Lactose	-/-
Malate	-/ND
Malonate	-/ND
Maltose	+/+
Mannitol	+/+
Mannose	+/+
Melibiose	-/ND
Raffinose	-/-
Rhamnose	-/-
Ribose	-/V
Sorbitol	+/+
Sucrose	+/+
Trehalose	+/+
Xylose	-/V

Hahella chejuensis was firstly discovered in coastal marine sediment, Cheju Island, South Korea in 2001 by a group of researchers from Korea Ocean R & D Institute and Korean Collection for Type Cultures (Lee et al., 2001). The researchers found that the bacterium produces a red pigment which known as prodigiosin. Prodigiosin is one of the secondary metabolites produced by certain bacteria which has anticancer, immunosuppressive, algicidal, and anti-malarial activity (Kim et al., 2008; Chawrai et al., 2012). Prodigiosin produced by *Hahella chejuensis* has an algicidal activity. Photomicrograph taken during a laboratory challenged experiment of the prodigiosin from *Hahella chejuensis* showed an acute lytic effect against microalgae *C. polykrikoides* after 30 minutes exposure (Jeong et al., 2005; Kim et al., 2008). Prodigiosin cause loss of cell membrane integrity which allows it to penetrate inside algae *Microcystis aeruginosa* cell after 12 hours exposure. The continuous flowing of prodigiosin will lead to cell burst (Yang et al., 2017).

Study about bacterial genomic blueprint reveals several potential virulence-associated genes on *H. chejuensis* KCTC 2396 genome that suggests the bacteria may be a pathogen in eukaryote organism. There are several clusters of the gene that potentially involved in the producing of bacterial exopolysaccharides. Other genes that homolog with hemolysin and RTX toxin which usually contribute to cytotoxic activity also was found in the bacterial genome. Other virulence associated genes that are found in the bacterial genome are two type of type III secretion systems (TTSSs). The researcher expected that prodigiosin, TTSSs, and other virulence properties contribute to the pathogenic lifestyle of *H. chejuensis* KCTC 2396 (Jeong et al., 2005).

5. General concept of bacterial virulence properties

Pathogenic bacteria use weapons provided in their body to deliver infection, called as virulence properties. Wu et al. (2008) categorized these virulence properties

based on their virulence mechanism and function in to three groups, membrane proteins, polysaccharide capsules that located along the cell, and secretory proteins that mostly act as toxin. Proteins in membrane cell have important role adhesion, colonization, and invasions to the host cell at the beginning stage of infection (Finlay and Falkow, 1997).

The secretion system apparatus facilitates secretory proteins injection into the host cell or extracellular matrix. There are five secretion systems (I-VI Secretion System) assembled at the bacterial membrane surface which are differentiated based on their mechanism and structure. Some systems are only existed in Gram negative bacteria, such as type 1 secretion system (T1SS), T2SS, T3SS, T4SS, T5SS, T6SS, and T9SS, while T7SS is occurred only in Gram positive bacteria (Depelteau et al., 2019). T1SS is a C-terminal secretion signal dependent. The system is composed by three proteins, including ATP-binding cassete (ABC) transporter and a membrane fusion protein (MFP) which are in the inner membrane of Gram-negative bacteria and another protein is in outer membrane *(Thomas et al., 2014)*. T1SS secrete hemolysin toxin in *Escherichia coli, Vibrio cholerae,* or *Bordetella pertussis (Finlay and Falkow, 1997).* T2SS transport unfolded proteins from inner membrane to outer membrane and change it to be folded proteins. A good example of T2SS role in disease pathogenesis was cholera toxin secretion along *Vibiro cholerae* infection. T3SS has similar mechanism with needle and syringe, this it is also called as injectisome. T3SS has been determined as the major virulence trait of *Aermonas salmonicida* (Frey and Origgi, 2016). T4SS plays role during bacterial conjugation. T6SS has a unique structure, tube-shaped structure inside the bacterial cell. (Depelteau et al., 2019).

When the bacteria cause an infection, they will manage a complex virulence mechanism facilitating them to invade their host and establish a disease. There are five basic steps for a bacterium to cause a disease: 1) attachment or bacterial entry

to the body or host cell, 2) evasion from the host immune system, 3) multiplication or colonization at the site of infection and then spreads to other sites, 4) cause damage to the host cells and systems, 5) Spreading from the infected animal to other healthy susceptible animals, for the continuous infection cycle (Gyles and Prescott, 2004). There are many virulence properties work together contributing to the mechanism. The attachment of a bacteria to the target surface appears to be an important prerequisite for successful infection (Ben Hamed et al., 2018). *T. ovolyticus,* a pathogen in Halibut's eggs, attaches and colonize on the mucous surface of eggs during on its early phase of infection. Later the bacterium accomplishes a proteolytic activity and generates ulceration on the chorion layer (the outermost layer of fish eggs). The radiate zone is damaged due to the bacterial exoenzymatic activity which is conceivably resulting in egg puncture, leakage of cell constituents, and larval death (Bergh et al., 1992; Hansen et al., 1992).

Aeromonas hydrophila, a well-known freshwater fish pathogen, is a motile bacterium. It uses both lateral and polar flagella to support its motility in solid surface and watery environment repectively. Qin et al. (2016) reported that these flagellar mediated motilities is essential in bacterial adherence to the host mucus at early infection stage. Then, the bacteria secretes a bunch of toxins including adhesins, cytotoxins, hemolysins, lipases, and proteases which are regulated by type II, III, and VI secretion systems (Rasmussen-Ivey et al., 2016). Biofilm formation is also acknowledged as one of the virulence properties, such as in *Francisella noatunensis* subsp. *orientalis, Vibrio fischeri, Aeromonas* sp.*,* and other pathogenic bacteria (Soto et al., 2015; Ben Hamed et al., 2018; Dias et al., 2018). Virulence Factor Data Base (VFDB) categorizes virulence factors into three group, offensive, defensive, and nonspecific virulence factor. Offensive virulence factor is those related to adherence, invasion, toxin, actin-based motility, and secretion system. Defensive virulence factor

is likely for bacterial survival, including anti-phagocytosis, anti-proteolysis, cellular metabolism, serum resistance, and many more. Other factors which do not belong to the previous group including iron uptake system, magnesium uptake system, and exoenzyme (Chen et al., 2005).

All these virulence properties are diverse among bacteria which give various patterns and characteristics in each bacterial pathogenicity. A bacterium can gain or loss its virulence and affect to their pathogenicity through several mutations or mobile genetic elements, including insertion sequence (IS), bacteriophages, pathogenicity island, and plasmid (Pallen and Wren, 2007). Bacteriophage also can transfer a virulence associated gene and increase the bacterial pathogenicity. *Vibrio harveyi* Y6, a causative agent of piscine scale drop and muscle necrosis syndrome in Vietnam, carried CTXϕ or zonula occludens toxin (*zot*) which is homolog to *V. cholera*. The toxin associated genes were transferred by VHY6ϕ phage (Kayansamruaj et al., 2018). *Streptococcus agalactiae* has hemolysin toxin encoded by 12 genes clustered in *cyl* operon. The presence of 1252 bp IS in the *cyl*F region reduced the expression of hemolysin toxin and expressed incompletely (α-hemolysin). While the absence of *cyl* operon and replaced by 14 kb genomic island (GI) lost the expression of hemolysin at all (γ-hemolysin) (Chou et al., 2019).

6. General concept of bacterial genome characterization

Since the first whole genome sequencing project was established in *Haemophilus influenza* in 1995, there have been many genomic projects are created. More than thousand genes and genome are sequenced and collected into genome database. At the same time, sequencing technology has been developed greatly, faster and cheaper (Deurenberg et al., 2017). There are many high throughput sequencing tools provided by Illumina, Thermofisher, Oxford Nanopore, and Pacific Biosciences (Loman et al., 2012; Deurenberg et al., 2017).

Comparing with Sanger sequencing, Next Generation Sequencing (NGS) use different protocol, consisting DNA library preparation, DNA amplification, and DNA sequencing. DNA library preparation should be performed by extracting the bacterial DNA form a fresh inoculation and measure the extracted DNA quality and quantity by fluorometry. The amount of extracted DNA can be varied according to NGS equipment used in the research. Then, the extracted DNA is fragmented by several enzymatic reagents and amplified randomly in particular duration before sequencing. As mentioned above, there are some platforms have been developed to perform whole genome sequencing. Each platform has different characteristics, such as running time, initial read length, speed (Gb per run), and chemical methods. Illumina Miseq is more cost effective, short running time, lowest error rate, and applicable for microbial project. The short reads mostly produced during sequencing can be filtered by computational software, FastQC or CLC Genomic Workbench (Loman et al., 2012; Edwards and Holt, 2013).

There are many methods to characterize bacterial virulence properties. Moreover, the developing genome sequencing technology has made it possible to study whole bacterial genome. Through bioinformatics or computational analysis, such as annotation, between unknown and closed reference genome from an internet database, we can define virulence-associated genes by their orthologous. Recently, there are many software and websites developed by certain genome study center to facilitate genome annotation such as Basic Local Alignment Search Tools or BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi), Rapid Annotation using Subsystem Technology or RAST (http://rast.nmpdr.org), Microscope (http://www.genoscope.cns.fr/agc/microscope), and many more. It has been applied in *Flavobacterium columnare*, one of the important pathogens in freshwater fish. Study about *F. columnare*'s complete genome sequence reveals virulence-

associated genes which are classified into four classes, virulence factors, gliding motility proteins, adhesins, and putatively secreted proteases. These genes are suggested to work in colonization, invasion, and destruction of fish tissue during infection (Zhang et al., 2017).

Mobile genetic elements such as prophage and pathogenic island can be determined using several computational platforms developed by genetic research center. A successfully persisted phage in a bacterial genome or can also called as prophage can be investigated PHAST web based tool and its upgraded version, PHASTER (Zhou et al., 2011; Arndt et al., 2016). It is a rapid and accurate annotation for finding prophage sequence inside a bacterial genome and plasmid. The prophage detection method used in this website are knowledge-based rules/metrics and gene function based on the database (Arndt et al., 2017). We can also observe its location inside the genome from graphical picture provided by the webtool (Arndt et al., 2016).Pathogenic islands, gained by horizontally transference among bacteria, produce some essential material which are needed in a disease development. Pathogenic islands Identification can be done by annotate the studied bacterial genome against the database (Pathogenicity island database, PAIDB). It can be freely accessed at http://www.gem.re.kr/paidb (Yoon et al., 2007). Another web server called as IslandViewer 4 has an expanded database (Bertelli et al., 2017a). These websites provide an illustration about the pathogenic island location and the homolog islands from other bacteria (Senapin et al., 2016). Another genomic element such as CRISPR (Clustered regularly interspaced short palindromic repeats) can be detected by annotate a bacterial genome at https://crisprcas.i2bc.paris-saclay.fr, an improved webtool (Couvin et al., 2018).

Pangenome is whole repertoire genome from a certain group of bacteria. It is also known as one of a genomic approach to study bacterial nature and describe the

whole genomic composition of studied bacteria. Based on the ability to achieve new genes, pangenome is defined as open or closed pangenome. This open and closed pangenome can determine the bacterial lifestyle. Pangenome is composed of three-part, core, secondary or accessory, and unique genes. Core genome is those genes persists in all strain of studied group, while secondary or accessory genes can be found in several strains of studied group. Unique gene is only present in one strain of the studied group. The unique gene is considered to determine specific characteristic from certain bacteria. Whole genome analyzing by bioinformatics tools can determine the composition of the bacterial genome (Rouli et al., 2015; Kayansamruaj et al., 2018).

CHAPTER III. MATERIAL AND METHODS

In a short explanation, the experiment had been concluded in the concecptual framework below (Figure 1.)

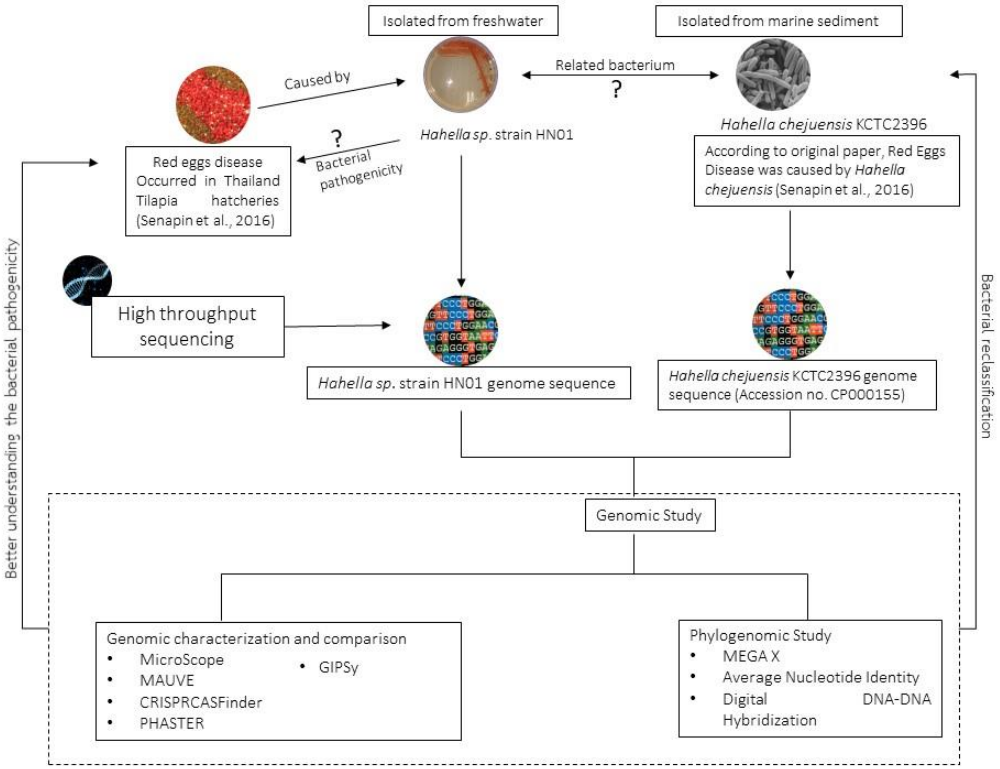


Figure 1 Conceptual Framework

1. Genome sequencing preparation

1.1 Bacterial culture condition and DNA extraction

The bacterium used in this study was *Hahella chejuensis* HN01 which was deducted as strain HN01 during this study, isolated from infected Tilapia hatcheries in Prachinburi province from a previous study (Senapin et al., 2016). The bacterium kept in -80°C bacterial stock was inoculated in Tryptic Soy Agar (TSA; Difco™) supplemented with 1.5% NaCl at 28°C for 48 hours. The ingredients used for bacterial culture was written in Appendix 1. The bacterial DNA was extracted from the bacterial suspension using Wizard™ Genomic DNA Purification Kit (Promega™ , USA). The concentration of extracted DNA was quantified using Qubit™ Fluorometric Quantitation (Thermo Fisher Scientific, Waltham, MA, USA) before stored at -20°C.

1.2 Library preparation, sequencing and assembly

Genome library amplification was done using Nextera XT (Illumina, San Diego, CA, USA). The whole genome sequencing was performed by Illumina Miseq. The good quality of sequencing product, also called as reads was selected (Q score > 30) using software CLC Genomic Workbench ver 6.9 (CLC bio, Aarhus, Denmark) and *de novo* assembly was conducted using SPAdes Genome Assembler. Assembled contigs was scaffolded using SSPACE ver 2.0 software, and the gaps (unknown nucleotide) within the scaffolds was filled automatically by GapFiller ver 2.1 (Boetzer et al., 2011; Nadalin et al., 2012). Finally, the scaffold quality was examined by QUAST program (Gurevich et al., 2013).

2 Phylogenomic study

2.1 Digital DNA-DNA Hybridization (DDH), Average Nucleotide Identity (ANI) estimation, and Phylogenetic analysis

To determine the taxonomic position of strain HN01, we first constructed phylogenetic trees using housekeeping genes. The 16s rRNA sequence of phylum

Proteobacteria member were extracted from the database (https://www.ncbi.nlm.nih.gov/), including strain HN01 from previous study. The sequences were aligned by using MUSCLE method (Edgar, 2004) and trimmed to uniform the sequences length (1,334 nucleotides). The phylogenetic tree was generated with maximum-likelihood method based on the Kimura 2 parameter. A discrete Gamma (+*G*) distribution was used to model the non-uniformity of evolutionary rate among sites and by assuming that a certain fraction of sites are evolutionarily invariable (+*I*) (Nei and Kumar, 2000; Kumar et al., 2018). Bootstrap analysis was applied with 1,000 replicates (Felsenstein, 1985). The phylogenetic tree construction was performed by using MEGA X software (Kumar et al., 2018). The detail of used bacteria is showed in the table 4.

The phylogenetic relation of our strain with other related bacteria was further supported with a tree constructed using multiple housekeeping genes: *16s rRNA*, *rpoD*, *recA, atpD,* and *infB* of the closest species. The housekeeping genes sequences were extracted from the genome of the reference strains through the NCBI database with accession number shown in the table 5. The selected gene sequences were extracted from National Center Biotechnology Information (https://www.ncbi.nlm.nih.gov/) database, aligned, and trimmed same way with the previous steps. The 8,802 of total concatenated gene nucleotides analyzed with MEGA X program in order to select the best model. The tree was generated based on general time reversible (GTR) model. A discrete Gamma (+*G*) distribution was used to model the non-uniformity of evolutionary rate among sites and by assuming that a certain fraction of sites are evolutionarily invariable (+*I*) (Nei and Kumar, 2000; Kumar et al., 2018). Bootstrap value was applied with 1,000 replicates.

Each gene sequence was also used to generate a phylogenetic tree independently in order to observe if the branch was consistent using single gene or

multiple genes. The tree was constructed using maximum likelihood methods with a bootstrap value of 1000 by using MEGA X software. The best model calculated by MEGA X and the length of aligned sequences for every gene was showed in Appendix 2. The phylogram information was supported by nucleotide and amino acid-level comparisons for every pairwise genes calculated with Sequence Identities and Similarities webtool (http://imed.med.ucm.es/Tools/sias.html).

For further clarification of bacterial delineation, the nucleotide level comparisons of two genomes were performed with digital DNA DNA Hybridization (dDDH) and Average Nucleotide Identity (ANI). Genomic distance between strain HN01 and the related bacteria was calculated by *in sillico* DNA-DNA hybridization (dDDH) using web service Genome to Genome Distance Calculator (https://ggdc.dsmz.de/). The calculation was performed based on BLAST+ local alignment tool and recommended setting, formula 3 (Meier-Kolthoff et al., 2013; Meier-Kolthoff et al., 2014a; Meier-Kolthoff et al., 2014b). ANI calculation was performed by using EZBioCloud website (https://www.ezbiocloud.net/tools/ani) (Yoon et al., 2017).

Table 4. Member of Proteobacteria used in phylogenomic analysis

Strain	Species	Order	Type strain	GenBank accession no
HN01	*Hahella chejuensis*	*Oceanospirillales*	-	KT971015.1
KCTC2396	*Hahella chejuensis*	*Oceanospirillales*	√	NR_074812.1
DSM17046	*Hahella ganghwensis*	*Oceanospirillales*	√	AY676463.1
NBRC102683	*Allohahella antartica*	*Oceanospirillales*	√	NR_114177.1
H94	*Allohahella marinomesophila*	*Oceanospirillales*	√	HQ834530.1
AJ275	*Halomonas saccharevitans*	*Oceanospirillales*	√	NR_044117.1
TRM0175	*Halomonas xinjiangensis*	*Oceanospirillales*	√	EU822512.1
JAMM0745	*Neptunomonas japonica*	*Oceanospirillales*	-	NR_041567
4CA	*Nitrincola lacisaponensis*	*Oceanospirillales*	√	NR_042984.1
ATCC11336	*Oceanospirillum linum*	*Oceanospirillales*	√	M22365.1
ME102	*Oleiphilus messinensis*	*Oceanospirillales*	√	NR_025432.1
JC2044	*Zooshikella ganghwensis*	*Oceanospirillales*	-	AY130994.2
DSM 6062	*Alteromonas macleodii*	*Alteromonadales*	-	Y18228.1
YCSA40	*Marinobacter daqiaonensis*	*Alteromonadales*	√	NR_108457.1
T5054	*Marinobacter lutaoensis*	*Alteromonadales*	√	NR_025116.1
UST090418-1611	*Marinobacter xestospongiae*	*Alteromonadales*	√	NR_109066.1
CL-YJ9T	*Marinobacterium rhizophilum*	*Alteromonadales*	-	NR_044157.1

Table 5 *Oceanospirillales* member used in the phylogenomic analysis

No	Species	Strain	Family	Reference sequence*	Type strain
1	*Hahella chejuensis*	HN01	*Hahellaceae*	This study	-
2	*Hahella chejuensis*	KCTC3296		NC_007645.1	√
3	*Hahella ganghwensis*	DSM 17046		NZ_AQXX00000000.1	√
4	*Zooshikella ganghwensis*	DSM 15267		NZ_AUAF00000000.1	√
5	*Endozoicomonas montiporae*	CL-33	*Endozoicomonadaceae*	NZ_CP013251.1	√
6	*Endozoicomonas numazuensis*	DSM 25634		NZ_JOKH00000000.1	√
7	*Endozoicomonas elysicola*	DSM 22380		NZ_JOJP00000000.1	√
8	*Alteromonas macleo;dii*	ATCC 27126	*Alteromonadaceae* (outgroup)	NC_018632.1	√

*National Center for Biotechnology Information

3. Genome characterization and comparison

3.1. Genomic features *of Hahella chejuensis* HN01 and *Hahella chejuensis* KCTC2396

Hahella chejuensis HN01 genome sequences was annotated and analyzed by MicroScope (http://www.genoscope.cns.fr/agc/microcope/usepanel/genebasket.php) (Vallenet et al., 2017) and Rapid Annotation using Subsystem Technology (RAST) (Aziz et al., 2008). The strains HN01 genome was compared with published *Hahella chejuensis* KCTS2396 (Accession no. NC_007645.1) as the reference genome, and then it was deducted as strain KCTC2396. Regarding to the lack of reference genome in the MicroScope database, we also submitted the genome of *Hahella chejuensis* KCTC2396 at the same time. Two genome of the strain HN01 and KCTC2396 were submitted to RAST webtool (http://rast.nmpdr.org/) with default setting. Circular genome was visualized by a Java application called Circular Genome Viewer Cluster (Stothard and Wishart, 2004). Orthologous Groups of strain HN01 and KCTC2396 were automatically classified by COGNiTOR software. The genome repertoire was analyzed based on MicroScope gene families (MICFAM) and computed with the SiLix software provided by the same platform. The MICFAM parameters were 50% of amino acid identity and 80% of amino acid alignment coverage (Miele et al., 2011). These analysis tools were provided in MicroScope website. The amino acid sequences of gene classified in the core and unique genes were extracted for putative virulence properties identification.

3.2. Putative virulence properties identification

The amino acid sequences retrieved from previous study were locally aligned against Virulence Factor Data Base (VFDB) (Chen et al., 2005) by using Blast2GO ver. 5 software. It was performed against VFDB core dataset with BLASTp setting, minimum 30% identity, and 1.0E-3 of e-value (Conesa et al., 2005; Conesa and Götz, 2008; Götz

et al., 2008; Götz et al., 2011). The identified virulence properties were grouped into shared and specific virulence properties.

3.3. Other genome elements prediction

The presence of mobile genetic elements in *Hahella chejuensis* HN01 genome, such as genomic islands (GI), clustered regularly interspaced short palindromic repeats (CRISPR), and prophage were analyzed using different website tools. Genomic islands which are showed the acquired traits of the bacterium were analyzed by using IslandViewer ver 4 (Bertelli et al., 2017b). CRISPRs were analyzed by submitting the genome to CRISPRCasFinder web service (https://crisprcas.i2bc.paris-saclay.fr/CrisprCasFinder/Index) (Couvin et al., 2018). While PHAge Search Tool Enhanced Release (PHASTER) web server (http://phaster.ca/) was used with the same protocol of aforementioned webtool for putative prophage identification (Arndt et al., 2016).

4. Data Analysis

The data analysis in this research used descriptive statistical analysis and bioinformatics analysis.

CHAPTER IV. RESULT

1. The phylogenomic analysis of *Hahella chejuensis* HN01

The position of *Hahella chejuensis* HN01 in the *Proteobcateria* phylum was observed through 16S rRNA gene sequences based maximum likelihood tree. It showed that strain HN01 formed a separated lineage from other lineages composed of related species with *Hahella chejuensis* KCTC2396 (Figure 2) supported with very strong bootstrap value. This phylogenetic tree then was supported with more phylogenetic tree models based on the different housekeeping gene sequences of *Ooceanospirillales* order member. The use of several housekeeping genes sequence individually was assessed in order to reveal the potency of different housekeeping genes to determine strain HN01 taxonomy. Comparison among five phylogenetic trees showed similar result (Figure 3-7 a), which illustrates strain HN01 is always in the same branch with strain KCTC2396 with high value of bootstrap.

The percent identity of 16s rRNA sequences showed a high identity (99.7%) between our isolates and *H. chejuensis* KCTC2396, and the value was decreasing under 97% when it compared to *H. ganghwensis* DSM7046 (94.7%), *Z. ganghwensis* DSM7046 (90.6%), *E. elysicola* DSM22380 (88.5%), *E. montiporae* CL-33 (87.9%), *E. numazuensis* DSM25634 (88.3%), and *A. macleodii* ATSS27126 (85.7%). The comparison among amino acids sequence of our isolate housekeeping gene revealed a high percent identity and similarity to *H. chejuensis* KCTC2396. These results were corelated to the phylogram data that constructed based on the gene sequences (Figure 3-7 b).

To provide a higher resolution of phylogenetic relationship within family, we generated a maximum likelihood model of phylogenetic tree by concatenated

aligned housekeeping genes as recommended by Glaeser et al. (2015). The result was consistent with previous trees which showed the strain HN01 forms a distinct branch with KCTC2396 (Figure 8). The result from digital DNA DNA Hybridization and Average Nucleotide Identity of the strain HN01 and KCTC2396 were 70.70% and 89.04% respectively.

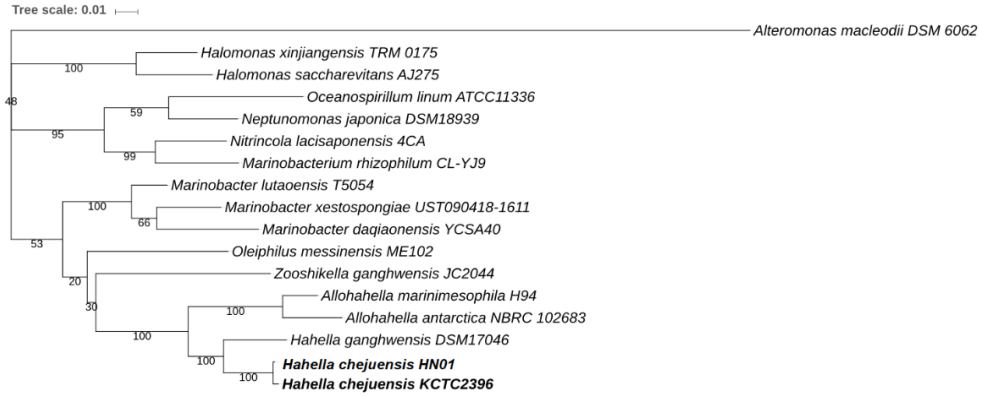
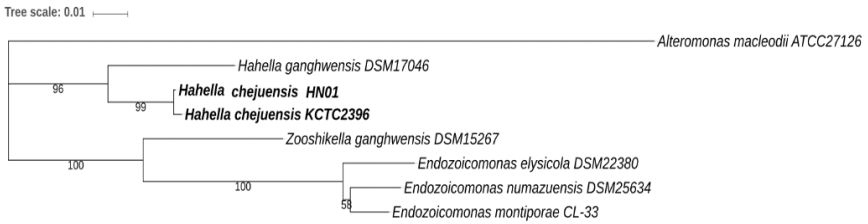


Figure 2 A maximum likelihood tree based on almost complete 16S rRNA gene sequences of phylum Proteobacteria member.

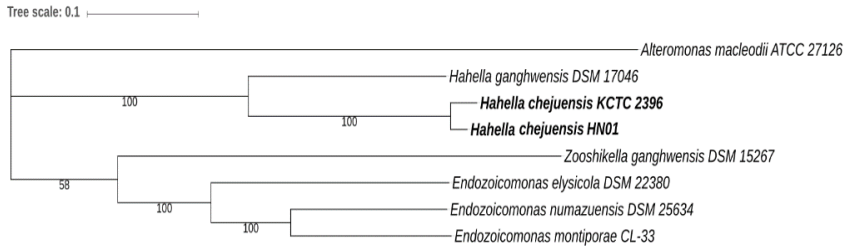
The tree showed the position of HN01 among related bacteria with bootsrap value at the branch point. *Alteromonas macleodii* DSM 6062 was used as an out group. The tree scale, 0.01 showed the subtitutions per nucleotide position.



	H. chejuensis HN01	H. chejuensis KCTC2396	H. ganghwensis DSM17046	Z. ganghwensis DSM15267	E. elysicola DSM22380	E. montiporae CL33	E. numazuensis DSM25634	A. macleodii ATCC27126
H. chejuensis HN01								
H. chejuensis KCTC2396	99.7%							
H. ganghwensis DSM17046	94.7%	94.6%						
Z. ganghwensis DSM15267	90.6%	90.4%	89.9%					
E. elysicola DSM22380	88.5%	88.2%	88.5%	91.1%				
E. montiporae CL33	87.9%	87.8%	88.0%	90.3%	95.6%			
E. numazuensis DSM25634	88.3%	88.2%	88.2%	91.2%	96.7%	96.6%		
A. macleodii ATCC27126	85.7%	85.4%	85.4%	84.7%	83.6%	82.8%	83.9%	

Figure 3. The phylogenetic relationship of strain HN01 among other *Oceanospirillales* member based on complete 16S rRNA sequence:

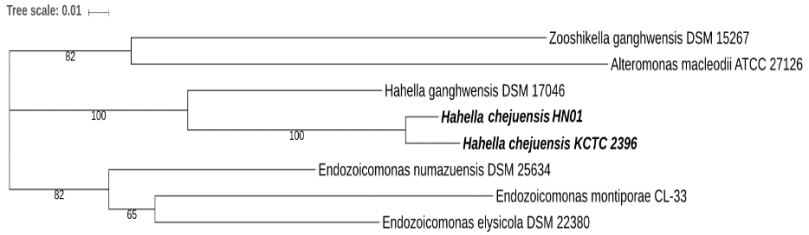
(a) The phylogenetic tree constructed with a maximum likelihood method and (b) a sequence identity matrix reconstructed from gene sequence.



	H. chejuensis HN01	H. chejuensis KCTC2396	H. ganghwensis DSM17046	Z. ganghwensis DSM15267	E. elysicola DSM22380	E. montiporae CL33	E. numazuensis DSM25634	A. macleodii ATCC27126
H. chejuensis HN01		99.84%	86.87%	78.95%	79.18%	77.08%	77.79%	75.12%
H. chejuensis KCTC2396	98.75%		86.71%	78.95%	79.01%	77.08%	77.79%	75.12%
H. ganghwensis DSM17046	80.62%	80.78%		77.65%	78.85%	77.90%	77.30%	75.77%
Z. ganghwensis DSM15267	70.63%	70.47%	70.47%		83.44%	83.79%	83.22%	75.61%
E. elysicola DSM22380	69.83%	70.32%	72.13%	78.85%		89.50%	88.98%	77.21%
E. montiporae CL33	68.24%	68.41%	70.37%	78.39%	85.73%		91.44%	74.63%
E. numazuensis DSM25634	68.42%	68.75%	70.88%	78.94%	85.69%	87.99%		75.82%
A. macleodii ATCC27126	65.13%	65.13%	66.77%	67.43%	69.18%	66.28%	68.09%	

Figure 4. The phylogenetic relationship of strain HN01 among other *Oceanospirillales* member based on complete *rpoD* sequence:

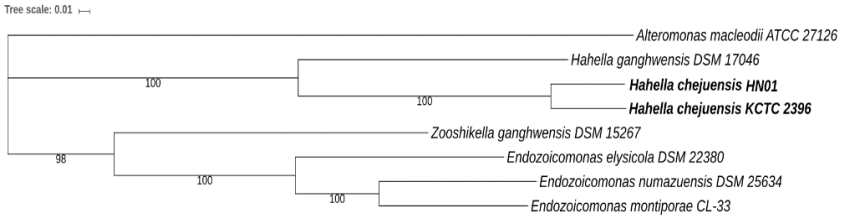
(a) The phylogenetic tree constructed with a maximum likelihood method and (b) a sequence identity (grey region) and similarity (white region) matrix reconstructed from protein sequence.



	H. chejuensis HN01	*H. chejuensis KCTC2396*	*H. ganghwensis DSM17046*	*Z. ganghwensis DSM15267*	*E. elysicola DSM22380*	*E. montiporae CL33*	*E. numazuensis DSM25634*	*A. macleodii ATCC27126*
H. chejuensis HN01		99.56%	96.95%	91.72%	92.37%	91.06%	91.28%	90.86%
H. chejuensis KCTC2396	99.34%		96.73%	91.50%	92.15%	90.84%	91.06%	90.65%
H. ganghwensis DSM17046	94.56%	94.13%		92.37%	92.15%	91.72%	91.28%	90.43%
Z. ganghwensis DSM15267	88.45%	88.45%	88.67%		94.11%	94.11%	93.02%	92.81%
E. elysicola DSM22380	89.54%	89.54%	88.45%	91.72%		96.29%	96.07%	92.37%
E. montiporae CL33	88.23%	87.79%	88.67%	91.06%	94.98%		94.98%	91.28%
E. numazuensis DSM25634	87.58%	87.58%	86.92%	89.54%	93.24%	92.59%		91.50%
A. macleodii ATCC27126	87.60%	87.17%	86.73%	89.32%	88.45%	87.79%	87.36%	

Figure 5. The phylogenetic relationship of strain HN01 among other *Oceanospirillales* member based on complete *atpD* sequence:

(a) The phylogenetic tree constructed with a maximum likelihood method and (b) a sequence identity (grey region) and similarity (white region) matrix reconstructed from protein sequence.



	H. chejuensis HN01	*H. chejuensis KCTC2396*	*H. ganghwensis DSM17046*	*Z. ganghwensis DSM15267*	*E. elysicola DSM22380*	*E. montiporae CL33*	*E. numazuensis DSM25634*	*A. macleodii ATCC27126*
H. chejuensis HN01		96.28%	81.76%	71.20%	66.82%	66.20%	65.38%	65.38%
H. chejuensis KCTC2396	95.23%		81.09%	71.20%	66.93%	66.47%	65.42%	64.84%
H. ganghwensis DSM17046	74.33%	73.89%		72.42%	66.47%	66.24%	65.78%	66.16%
Z. ganghwensis DSM15267	62.50%	62.62%	63.84%		76.10%	76.22%	77.08%	68.38%
E. elysicola DSM22380	57.15%	57.50%	56.46%	68.25%		85.56%	83.35%	66.12%
E. montiporae CL33	58.07%	58.23%	57.63%	68.75%	81.02%		85.84%	65.13%
E. numazuensis DSM25634	57.25%	57.42%	56.60%	69.73%	77.53%	82.78%		65.24%
A. macleodii ATCC27126	55.51%	54.75%	55.46%	59.31%	54.94%	54.89%	54.31%	

Figure 6. The phylogenetic relationship of strain HN01 among other *Oceanospirillales* member based on complete *infB* sequence:

(a) The phylogenetic tree constructed with a maximum likelihood method and (b) a sequence identity (grey region) and similarity (white region) matrix reconstructed from protein sequence.

	H. chejuensis HN01	*H. chejuensis* KCTC2396	*H. ganghwensis* DSM17046	*Z. ganghwensis* DSM15267	*E. elysicola* DSM22380	*E. montiporae* CL33	*E. numazuensis* DSM25634	*A. macleodii* ATCC27126
H. chejuensis HN01		99.42%	95.66%	85.54%	83.52%	83.52%	83.81%	33.06%
H. chejuensis KCTC2396	98.55%		95.66%	85.54%	83.52%	83.52%	83.81%	33.06%
H. ganghwensis DSM17046	92.77%	93.35%		85.59%	83.86%	82.99%	83.28%	33.46%
Z. ganghwensis DSM15267	79.47%	79.19%	79.25%		85.67%	85.91%	86.36%	30.67%
E. elysicola DSM22380	77.16%	76.58%	77.52%	81.46%		91.95%	93.75%	29.88%
E. montiporae CL33	78.03%	78.32%	78.38%	81.32%	88.50%		91.37%	29.88%
E. numazuensis DSM25634	77.45%	77.16%	77.23%	82.10%	90.62%	89.36%		29.48%
A. macleodii ATCC27126	21.11%	20.71%	20.71%	18.32%	17.52%	17.13%	17.52%	

Figure 7 The phylogenetic relationship of strain HN01 among other *Oceanospirillales* member based on complete *recA* sequence*:*

(a) The phylogenetic tree constructed with a maximum likelihood method and (b) a sequence identity (grey region) and similarity (white region) matrix reconstructed from protein sequence.

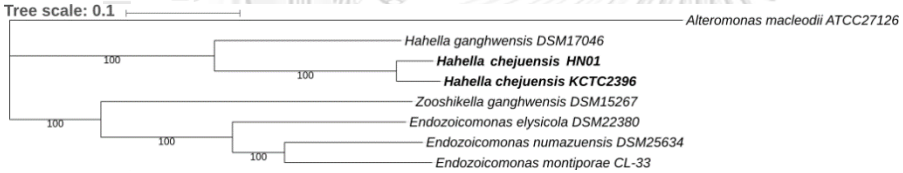


Figure 8. The maximum likelihood tree generated from concatenated housekeeping sequences.

The tree showed the position of strain HN01 among related bacteria with bootsrap value at the branch point. Alteromonas macleodii DSM 6062 was used as an out group. The tree scale, 0.01 showed the subtitutions per nucleotide position.

2 Genome characterization and comparison

2.1 Genomic features *of H. chejuensis* HN01 and *H. chejuensis* KCTC2396

The complete genome assembly of *H. chejuensis* HN01 using SPAdes and SSPACE ver 2.0 software generated 134 scaffolds. All the scaffolds had more than 1,000 bp except Scaffold_133 (991 bp). The whole genome annotation computed by MicroScope webtool of *H. chejuensis* HN01 contained 7.14 Mb in total length with 53.92 of % GC content and contained 6,777 coding sequence of gene (CDS) or 88.25% protein coding density with average length was 935 bp. The size of strain KCTC2396 was 7.2 Mb with 53.87 of % GC content. The genome contained 7,262 CDSs (88.78% of total chromosome size) with average length is 895.35 bp. The genome of *H. chejuensis* HN01 contains 4 rRNA genes and 58 tRNA genes, whereas *H. chejuensis* KCTC2396 genome contains 15 rRNA genes and 67 tRNA genes. The detail about both genome profile retrieved from MicroScope and RAST website are shown in Table 6.

From 6,777 CDSs in *H. chejuensis* HN01 genome, there were 68.26% of the total or 4,626 CDSs were classified in at least one of orthologous group (COG) identified in the genome. From 7,262 CDSs identified in *H. chejuensis* KCTC2396 genome, 4550 CDSs (62.65%) are classified in at least one COG group. COG automatic classification and comparison performed by MicroScope pipeline defined the bacterial genome into 24 clusters (Figure 9, Appendix 3). There are several clusters that have different gene number between the two genomes, including those who responsible in amino acid and carbohydrate metabolism; replication, recombination, and repair; cell wall/membrane/envelope biogenesis, and signal transduction mechanisms. Based on circular genome, we observed multiple regions in *H. chejuensis* HN01 were absent in *H. chejuensis* KCTC2396 genome as well as the opposite condition (Figure 10).

The proteins were mostly found in the replicon according to Kyoto Encyclopedia Genes and Genome (KEGG) functional annotation were carbohydrate and amino acid metabolism. This result was consistent with COG functional annotation. Comparing the result between two genomes showed a significant number of genes in lipid metabolism, 60 and 152 genes for *H. chejuensis* HN01 and *H. chejuensis* KCTC2396 respectively, and xenobiotics biodegradation and metabolism, 46 and 142 genes for *H. chejuensis* HN01 and *H. chejuensis* KCTC2396 respectively. The detail about KEGG functional annotation can be observed in Appendix 4.

The pangenome analysis performed by MicroScope clustered the protein coding genes into MicroScope gene families (MICFAM families) based on their homologous: amino acid alignment coverage and identity. In this study, the homologous genes were filtered based on 50% amino acid identity and 80% amino acid alignment coverage (Miele et al., 2011). The pangenome analysis revealed that the genome of *H. chejuensis* HN01 and *Hahella chejuensis* KCTC2396 share 4,929 gene families containing 5,574 and 5,544 CDSs of *H. chejuensis* HN01 and *H. chejuensis* KCTC2396 respectively. Of the 7,592 pangenome, 1,093 families (1,146 CDSs) were not matched with *H. chejuensis* KCTC2396 and designated as *H. chejuensis* HN01 specific genes. Conversely, 1,570 families (1,665 CDSs) in the *H. chejuensis* KCTC2396, out of the 7,592 pangenome, were not matched with *H. chejuensis* HN01 and designated as *H. chejuensis* KCTC2396 specific genes (Figure 11).

Table 6. Genome profile of *H. chejuensis.* HN01 and *H. chejuensis* KCTC2396

	MicroScope		RAST	
	H. chejuensis **HN01**	***H. chejuensis*** **KCTC2396**	***Hahella sp.*** **HN01**	***H. chejuensis*** **KCTC2396**
Chromosome Size (bp)	7,141,876	7,215,267	7,128,576	7,215,267
G+C content (%)	53,92	53,87	53.9	53,9
Total number of CDS	6,777	7,262	6,888	6,575
Average CDS length	934.86	895.35	-	-
Average intergenic length	144.61	134	-	-
Protein coding density (%)	88.25	88.77%	-	-
Pseudogenes	6	5	-	-
rRNA	1-1-2	5-5-5	-	-
tRNA	58	67	-	-
No of RNAs	-	-	62	82

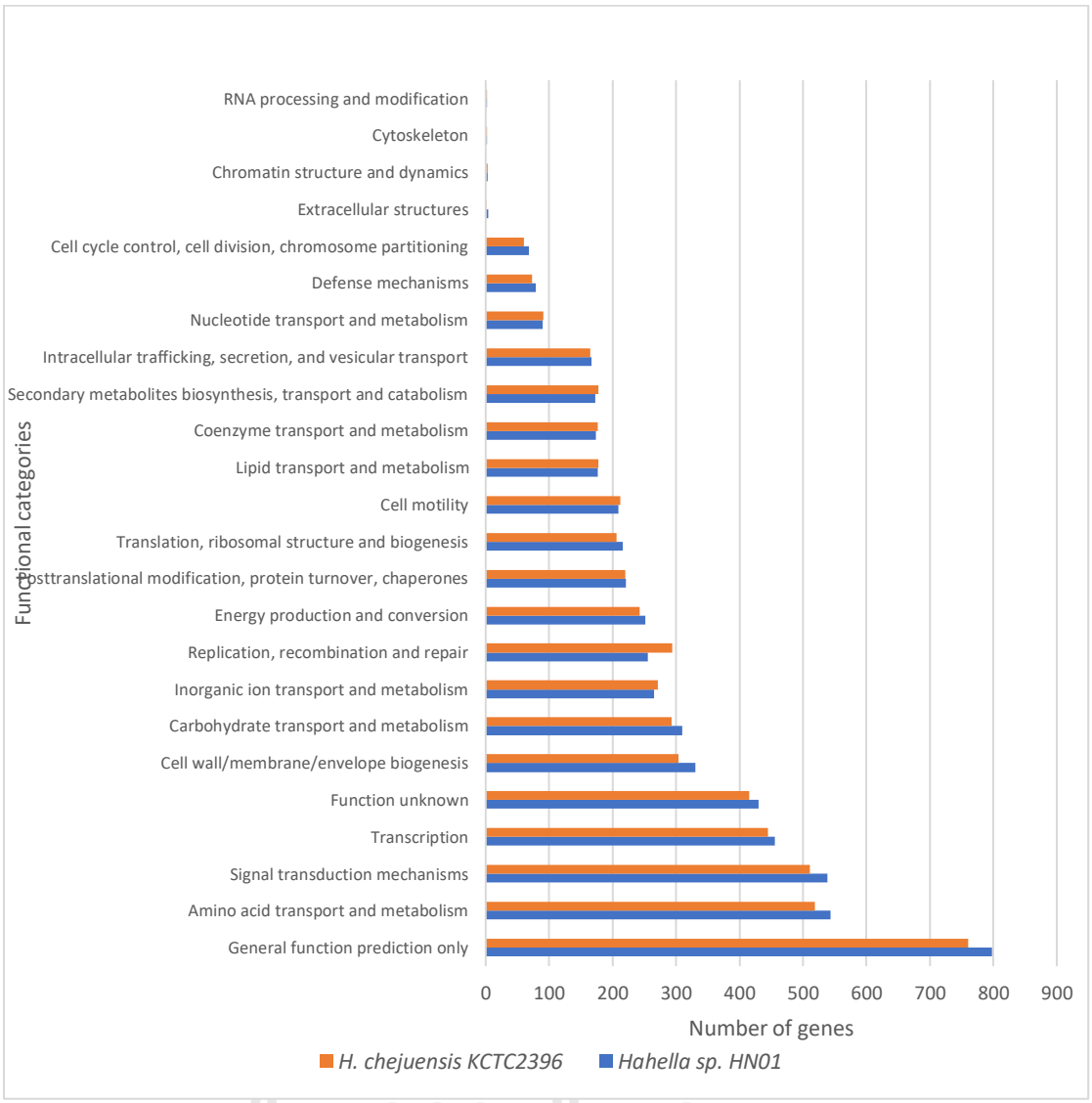
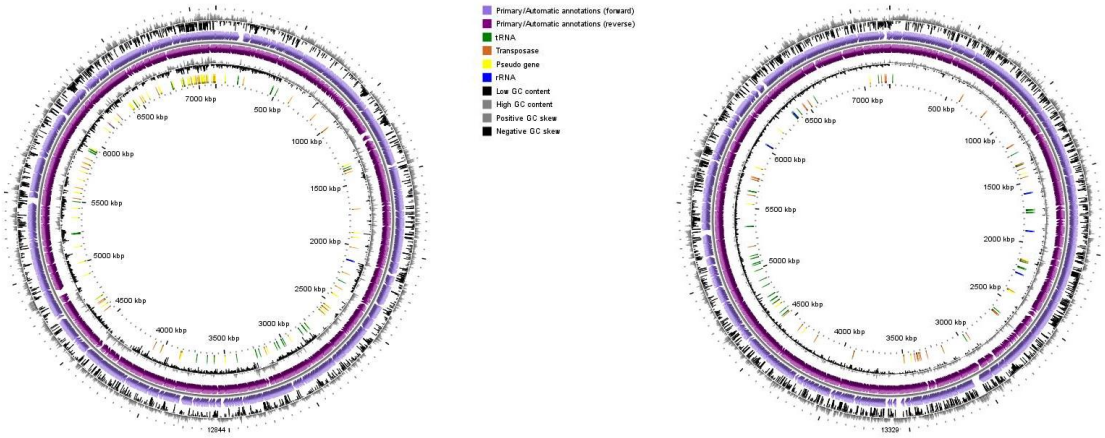


Figure 9. MicroScope functional categories of *H. chejuensis* HN01 and *H. chejuensis* KCTC2396



(a) (b)

Figure 10. Circular genome retrieved from GenoScope visualized by CGView
The left picture belongs to *H. chejuensis* and the right picture belongs *to H. chejuensis* KCTC2396

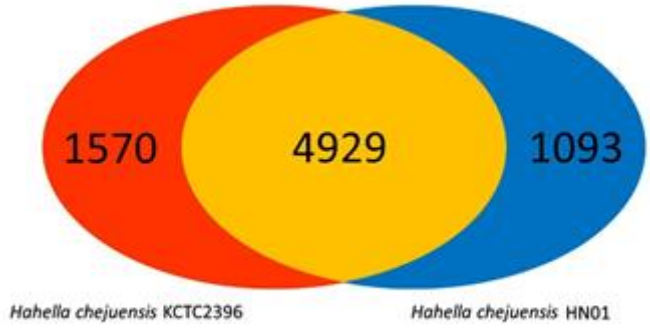


Figure 11. Pan-genome analysis of *Hahella chejuensis* KCTC2396 and *Hahella chejuensis.* HN01

2.2 Putative virulence properties identification

Both core and unique families retrieved from previous step were further analyzed in order to identify the virulence genes. According to pangenome analysis from previous section, there are 4,929 families which are containing 11,118 genes shared between *Hahella chejuensis* HN01 and reference bacterium. From 2663 families or 2808 genes, 1,145 genes are unique to *Hahella chejuensis* HN01 and 1,663 genes are unique to the reference. The local alignment of these genes repertoire against Virulence Factor Database (VFDB) based on the protein

sequence (BLASTp) hit to 2,877 virulence genes, ranged from 30% up to more than 80% identity to the original bacteria. After we filtered the redundant genes, we identified 589 genes associated to virulence properties. The full list of virulence genes observed in the *Hahella chejuensis* HN01 and KCTC2396 genome can be found in the Appendix 4.

As we mentioned above, there are 589 genes identified as repertoire virulence genes in both compared genome, strain HN01 and strain KCTC396. The identified virulence genes were classified in three categories listed in VFDB, offensive, defensive, nonspecific virulence factor, and regulation of virulence-associated genes. Those putatively categorized in the offensive virulence factor were mostly involved in adherence, invasion, motility activity, biofilm formation, production of toxin, endotoxin, and secretion systems (type II, III, IV, VI, and VII). Several genes were putatively identified as defensive virulence factor including, antiphagocytosis, stress protein, serum resistance, immune evasion and intracellular survival activity. The genes identified as nonspecific virulence factor were mostly involved in enzyme production, efflux pump, iron acquisition and uptake system, magnesium uptake system, and manganese uptake system. Several genes were also putatively identified as regulatory genes.

For further analysis, the amino acid sequences of *Hahella chejuensis* HN01 unique family genes from pangenome analysis were extracted in multi FASTA format. The deducted sequences were BLAST against VFDB by using BLAST2Go ver. 5 software. From the analysis, the were 19 putative genes which hit to the virulence data base, including capsule (*hscB, cps4D,* and *wcsT*), cholera toxin (*ctxA*), zona occludens toxin (zot), secretion system (*ankY/legA9, icmE,* and *sspH1*), intracellular adhesion protein (*icaA*), iron uptake (*iroN*), stress protein (*katA*), pyocyanin (*phzS* and phzM), lipopolysaccharide (*icaA*), iron uptake system

(*iroN*), Pseudaminic acid biosynthesis protein (*flgR* and *eptC*), and regulation gene (*rcsB*). The detail of unique genes identified in *Hahella chejuensis* HN01 genome are presented in table 7.

There are 22 genes identified specifically in the strain KCTC 2396 genome. Adherence associated genes such as *shdA, hmw2A, afaG-VII, inlF*, and *etpB* were identified in the unique region. Other virulence features including capsule (*cpsE, cpsJ*, and *cps4H*), lipooligosaccharide (*opsX/rfaC* and *lsgA*), iron uptake (*iutA* and *isdE*), manganese uptake (*psaA*), flagella (*ptmB* and *motB*), secretion system (*vgrG, lpg2370, icmJ/dotN*, and *eccA5*), cytolisin (*cylR2*) and colibactin (*clbE* and *clbH*). The detail of unique genes identified in *H. chejuensis* KCTC2396 genome are presented in table 8

Table 7 Putative virulence gene identified as unique gens in Hahella chejuensis HN01

No	NCBI ID	Homolog Gene*	Vir Name	Product	Original Bacteria
1	AAP42197	*hscB*	Capsule	capsular polysaccharide export protein HscB	*Haemophilus influenzae*
2	NP_344882	*cps4D*	Capsule	capsular polysaccharide biosynthesis protein Cps4D	*Streptococcus pneumoniae*
3	YP_002920359.1	*wcsT*	Capsule	galactoside O-acetyltransferase	*Klebsiella pneumoniae*
4	NP_231100	*ctxA*	CT	cholera enterotoxin, A subunit	*Vibrio cholerae*
5	YP_094446	*ankY/legA9*	Dot/Icm	Dot/Icm type IV secretion system effector LegA9/AnkY, ankyrin repeat-containing protein targeting the LCV for autophagy uptake	*Legionella pneumophila*
6	NP_820609	*icmE*	Dot/Icm T4SS	Type IVB secretion system protein IcmE/DotG, inner membrane protein of core-transmembrane complex	*Coxiella burnetii*
7	NP_647403	*icaA*	Intercellular adhesion proteins	N-acetylglucosaminyltransferase, involved in polysaccharide intercellular adhesin (PIA) synthesis	*Staphylococcus aureus*
8	NP_753164	*iroN*	IroN	salmochelin receptor IroN	*Escherichia coli*
9	NP_273273	*katA*	KatA	catalase	*Neisseria meningitidis*
10	NP_540321	*wbkA*	LPS	mannosyltransferase	*Brucella melitensis*
11	NP_207497	*flgR*	Pse5Ac7Ac	response regulator	*Helicobacter pylori*

12	YP_002343698	*eptC*	Pse5Ac7Ac, Pse5Ac7Am, Pse8OAc, Pse5Am7AcGlnAc	phosphoethanolamine transferase	*Campylobacter jejuni*
13	NP_252907	*phzS*	Pyocyanin	flavin dependent hydroxylase PhzS	*Pseudomonas aeruginosa*
14	NP_252898	*phzM*	Pyocyanin	phenazine-specific methyltransferase PhzM (adenosylmethionine dependent methyltransferase)	*Pseudomonas aeruginosa*
15	YP_002920501.1	*rcsB*	RcsAB	transcriptional regulator RcsB	*Klebsiella pneumoniae*
16	YP_695432	*nanH*	sialidase	sialidase	*Clostridium perfringens*
17	AAD40326	*sspH1*	SPI-1 encoded T3SS	type III secretion system effector SspH1, E3 ubiquitin ligase	*Salmonella enterica*
18	NP_253236	*pilS*	Type IV pili	two-component sensor PilS	*Pseudomonas aeruginosa*
19	NP_231101	*zot*	Zot	zona occludens toxin	*Vibrio cholerae*

*The virulence genes were determined by BLASTp against VFDB by using BLAST2Go ver 5 software

Table 8 Putative virulence gene identified as unique gens in *Hahella chejuensis* KCTC2396

No	NCBI ID	Homolog Gene*	Vir Name	Product	Original Bacteria
1	NP_461448	*shdA*	ShdA	AIDA autotransporter-like protein	*Salmonella enterica*
2	AAA20524	*hmw2A*	HMW1/HMW2	adhesin HMW2A, two-partner secretion (TPS) pathway TpsA exoprotein	*Haemophilus influenzae*
3	AAD44027	*afaG-VII*	Afimbrial adhesin, AFA-VII	AfaG-VII	*Escherichia coli*
4	NP_463939	*inlF*	InlF	internalin F	*Listeria monocytogenes*
5	CBJ04459	*etpB*	EtpA	two-partner secretion transporter EtpB	*Escherichia coli (ETEC)*
6	NP_816137	*cpsE*	Capsule	glycosyl transferase, group 2 family protein	*Enterococcus faecalis*
7	NP_688173	*cpsJ*	Capsule	glycosyl transferase CpsJ(V)	*Streptococcus agalactiae*
8	NP_344886	*cps4H*	Capsule	capsular polysaccharide biosynthesis protein Cps4H	*Streptococcus pneumoniae*
9	NP_438430	*opsX/rfaC*	LOS	heptosyltransferase I	*Haemophilus influenzae*
10	NP_439842	*lsgA*	LOS	lipopolysaccharide biosynthesis transporter	*Haemophilus influenzae*
11	NP_709458	*iutA*	Aerobactin	aerobactin receptor IutA	*Shigella flexneri*
12	YP_001332078	*isdE*	Isd	iron-regulated surface determinant protein E	*Staphylococcus aureus*

13	NP_346089	*psaA*	PsaA	manganese ABC transporter, manganese-binding adhesion liprotein	*Streptococcus pneumoniae*
14	NP_282477	*ptmB*	Flagella	acylneuraminate cytidylyltransferase	*Campylobacter jejuni*
15	YP_095806	*motB*	Flagella	flagellar motor rotation protein MotB	*Legionella pneumophila*
16	YP_856361	*vgrG*	T6SS	VgrG protein	*Aeromonas hydrophila*
17	YP_096378	*lpg2370*	Dot/Icm	Dot/Icm type IV secretion system effector	*Legionella pneumophila*
18	YP_094499	*icmJ/dotN*	Dot/Icm	Dot/Icm type IV secretion system protein IcmJ/DotN	*Legionella pneumophila*
19	NP_216314	*eccA5*	ESX-5	ESX-5 type VII secretion system AAA+ ATPase EccA5	*Mycobacterium tuberculosis*
20	AAM75247	*cylR2*	Cytolysin	cytolysin regulator R2	*Enterococcus faecalis*
21	YP_006635477.1	*clbE*	Colibactin	colibactin biosynthesis aminomalonyl-acyl carrier protein ClbE	*Klebsiella pneumoniae*
22	YP_006635480.1	*clbH*	Colibactin	colibactin non-ribosomal peptide synthetase ClbH	*Klebsiella pneumoniae*

*The virulence genes were determined by BLASTp against VFDB by using BLAST2Go ver 5 software

Prediction of genomic island using islandViewer ver 4 did not identify the presence of virulence associated genes. Two regions of incomplete prophage also were found in the genome. The first prophage was found in 67,058 – 92,011 region position with 24.9 Kb. The prophage harbored 30 proteins. The second prophage was found in 167,416 – 179,324 region position with 11.9 Kb. The prophage harbored 12 proteins. The more detail about the prophage was written on the following figure 10. There were four prophage regions identified in *Hahella chejuensis* KCTC2396 genome (Figure 11). The first region was identified at position 371,860 – 418,565 and contained 38 putative proteins. The length of this region was 46.7 Kb. The second region was identified at position 4,663,633 – 4,680,174 and contained 17 putative proteins. The length of this region was 16.5 Kb. The third region was located at 5,772,524 – 5,796,084 and contained 29 putative proteins. The length of the region was 23.5 Kb. The fourth region was identified at 6,867,447 – 6,901,202 and contained 37 putative proteins. The region length was 33.7 Kb. The prophage harbored in *H. chejuensis* HN01 and *H. chejuensis* KCTC2396 did not contained any virulence associated genes.

CRISPRCasFinder found one sequence with CRISPR in *H. chejuensis* HN01 genome. However, the sequence contained only one spacer and no CAS gene was found. In *Hahella chejuensis* KCTC2396 genome, we identified 4 sequences with CRISPR and 1 sequence with Cas cluster. Each CRISPR sequence from the two bacteria were not identical.

(a)

(b)

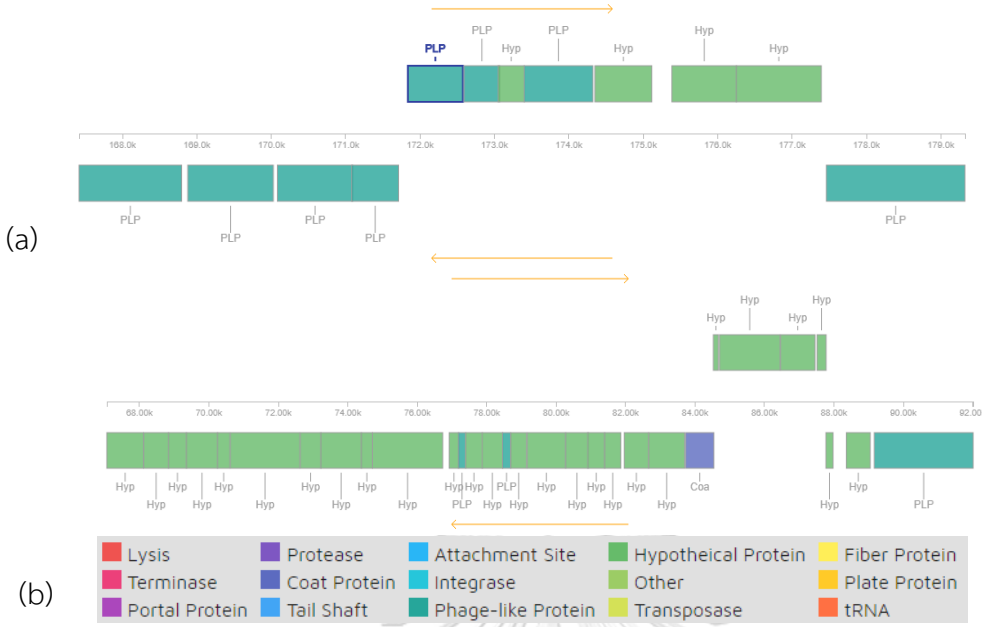
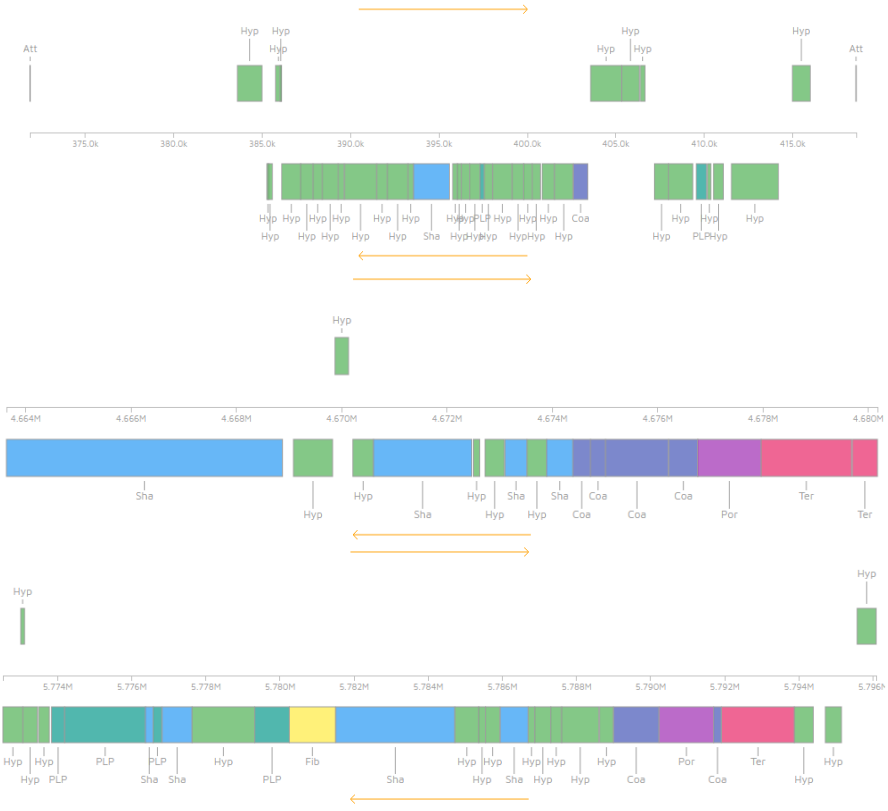


Figure 12 Two regions of prophage found in *Hahella chejuensis* HN01 genome by PHASTER webtool

(a)

(b)

(c)



(d)

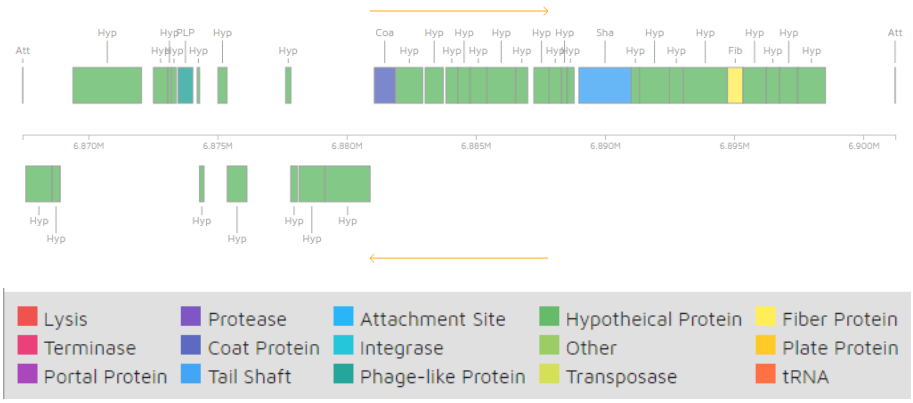


Figure 13 Four regions of prophage found in *Hahella chejuensis* KCTC2396 genome by PHASTER webtool

CHAPTER V. DISCUSSION AND CONCLUSION

1. Discussion

Hahella chejuensis HN01 was firstly isolated from Tilapia hatcheries in Prachinburi, Thailand during 2014 – 2015. The bacterium was reported as a causing agent of Red Eggs Disease. The name of the disease was come from the typical symptom observed at diseased eggs which turn their color to be red. The disease can infect to all egg stages, turned them to be red, and finally make them unhatched. The economic loss due to the reduced egg hatchability can be 10% and the number can be increasing up to 50% in cold season (Senapin et al., 2016). In this study, we characterized the genome of *Hahella chejuensis* HN01 which isolated from previous study and compared the genomic content to closely related bacteria, *Hahella chejuensis* KCTC2396, especially the virulence associated gene. We also study the bacterial phylogenetic relationship with closely related bacteria.

Family *Hahellaceae* contains four genera, including *Hahella, Zooshikella, Halospina, Endozocoimonas,* and *Kistimonas.* While genus *Hahella* contain only two species, *Hahella chejuensis* and *Hahella ganghwensis* which are previously have not reported as pathogens. Based on the original paper, strain HN01 isolated from Red Eggs disease has been designated as *Hahella chejuensis* according to the 16s rRNA sequence. Multiple genes based phylogenetic tree is recommended in order to obtain a higher resolution result. The genes sequence used in the analysis should be those who encode proteins with conserved function or also called as housekeeping genes (Glaeser et al., 2015). In this study, we not only analyzed the genes as a concatenated sequence, but also as individual dataset. We compared the phylogenetic relationship of our strain with close related bacteria based on each housekeeping genes sequence. This analysis was performed in order considering different evolutionary rate

The identical of amino acid sequences is also important in species delineation, especially for the closed related species who only can be differentiated by nucleotide sequence-based analysis. If two bacteria are clearly distinct based on their amino acids sequence, it indicates that they are different species. While, intra species relationship can be defined by the substitution occurred in nucleotide sequence. Therefore, amino acid based phylogenetic calculation should be performed together with nucleotide sequence-based analysis (Glaeser et al., 2015).

The percent identity or similarities of 16S rRNA between *Hahella chejuensis* HN01 and *H. chejuensis* KCTC2396 was 99.7%. It is related to Stackebrandt et al. (1994) that those who share the nucleotide similarities up to 97% or higher are considered to be the same species. Individual gene analyses indicate that the strain HN01 is closely related to *Hahella chejuensis* KCTC2396. The phylogram data were supported with high percent identical and similarity of each amino acid sequence: 98.75% AND 99.84% for *rpoD*, 99.34% and 99.56% for *atpD,* 95.23% and 96.28% for *infB*, and 98.55% and 99.42% for *recA*, respectively. These data are relatable with phylogram information that the strain HN01 and *Hahella chejuensis* KCTC2396 clustered together have high percent identity and similarity.

According to Tindall et al. (2010), it is recommended to do DNA-DNA hybridization when the strains of bacteria share are similar more than 97% of their 16S rRNA gene sequence. In this study we used Genome to Genome Distance Calculator 2.1 webtool. The webtool was proceed based on the DNA DNA hybridization principal. The result of dDDH was 70.70% which indicates that the strain HN01 is under *Hahella chejuensis* species name. To classify a strain of bacteria under a species name, it should have DDH value ≥70% when compared with the reference genome. Different with phylogenetic tree, DNA-DNA hybridization use the whole genome sequence to compare. This method has been established since 1987 and used to determine the relationship among organisms, especially bacteria (Wayne et al., 1987). However, dDDH calculation result was close to the threshold. Besides that,

Average nucleotide identity or ANI is a computer-based analysis that do pairwise alignment to all genome fragment sequence and calculate its identity which means this method has more precise value comparing to dDDH (Yoon et al., 2017). The result from ANI calculation was 89.40%. This result is lower than the standard, 95%, and indicates the variability between two genome sequence. From this study, we confirmed that strain HN01 is *Hahella chejuensis*.

The result from genome annotation of *Hahella chejuensis* HN01 and *H. chejuensis* KCTC2396 from two pipeline, MicroScope and RAST, showed similar size and G+C content. The number of coding sequences are slightly different since they are closely related bacteria. The functional annotation against the COG and KEGG data bases to reveal the function of proteins is also a great concern, as they grouped the varying features into families and super families (Do et al., 2017; Tripathi et al., 2017). The result comparison of COG classification from the two genomes share similar functional distribution. Most protein coding genes grouped in cellular processes and signaling were involved in the basic cellular function such as translation, transcription, and metabolism. 21% of genes have unknown function were grouped in “general function prediction only” and “function unknown”. Orthologs are the genes considered have the same function during evolution. Thus, the determination of orthologs assist to the gene function prediction in a newly identified species (Koonin, 2005). The pangenome analysis clearly showed that the two strains shared high number of common gene families (4,929), which accounted for 64.8% of the total families and considered to be conserved among them. There are 1,570 gene families which are unique in *H. chejuensis* KCTC 2396 genome, whereas there are 1,093 gene families which are unique in *H. chejuensis* HN01.

The local alignment against VFDB of protein sequence (BLASTp) showed many genes hit to the virulence properties. Total identified virulence are 589 genes and these number contribute to more than 80 potential virulence factors. Among all these, protein involved in bacterial structural production such as capsule, flagella,

pili, and secretion system, also biosynthesis of lipopolysaccharide (LPS), lipooligosaccharide (LOS), alginate, toxin, mycobactin, pyochelin, and protease were mostly identified in this study.

According to Gyles and Prescott (2004), bacterial pathogenesis is summarized in five basic steps, consisting (1) attachment and invasion; (2) host immune system evasion and survival; (3) bacterial replication and biofilm formation at the site of infection; (4) damage to the host; and (5) disease transmission from the infected animal to other susceptible animal.

Bacterial flagella and pili play an important role during the early stage of invasion such as adherence and motility. Bacterial capsule has been known to play an important role in bacterial survival and persistence in the environment. It is an indeterminate outer part of bacterial membrane consists of extracellular polysaccharide (ESP) (Roberts, 1996). Bacterial capsular involved in *Campylobacter jejuni* serum resistance and intestinal epithelial cell invasion during *in vitro* experiment (Bacon et al., 2001). The flagella allow a bacterium to move and approach its preferred substrates (Feldman et al., 1998). Flagellar and type IV pili (TFP)-mediated-twitching-motility play an important role in *Pseudomonas aeruginosa* biofilm formation (O'Toole and Kolter, 1998). Flagella supports the movement of the bacterium to the target surface and the type IV pili stabilize the interaction between bacteria and the target surface during biofilm formation (O'Toole and Kolter, 1998). Type IV pili allows the bacteria to move upstream against flow in microfluidic devices (Siryaporn et al., 2015).

After initial contact with the host, bacterial pathogens need to evade from the host immune system. According to Magnadóttir (2006), the innate immune system has been developed since eggs stage. There is phagocytic activity detected in zebrafish since embryo stage. The presence of complement component C3 at 7-9 days post-fertilization has been reported in cod eggs. Cathepsins responsible in

proteolytic digestion also present in embryo and larval stages of cod. The presence of lysozyme in sea bass, tilapia, and salmonid eggs has been studied. In salmonid eggs the enzyme can prevent vertical bacterial disease transfer.

In this study, we identified genes involved in capsule and alginate production. These virulence factors are responsible in phagocytosis evasion. In type III group B streptococci, the binding of complement component C3 to the bacterial surface can be prevented by the presence of capsule (Marques et al., 1992). Alginate has multifunction during *P. aeruginosa* infection in lung. It makes the bacteria attach to the cell surface tightly and being difficult to be removed. Alginate enfolds surrounding the bacteria and make a mucoid appearance. This form makes the pathogen cannot be evaluated by phagocytosis easily. Therefore, it also supports the biofilm production at the infection site (Stapper et al., 2004). Poly-beta-(1-6)-N-acetylglucosamine (PNAG) and adeFGH efflux pump are virulence features that associated with biofilm production in *Acinetobacter baumannii* (Choi et al., 2009; He et al., 2015). The lack of available iron within the host body that limit the bacterial growth is often overcome by the iron uptake and acquisition system (Gyles and Prescott, 2004)

To continue its pathogenic lifestyle for immediate or longer-term, the bacteria should gain more nutrients from the host and leads to the bacterial damage. There are several toxins identified in both *H. chejuensis* HN01 and KCTC2396 genome, including hemolysin and colibactin. Hemolysin form pore in host cell membrane in *S. agalactiae* (Nizet, 2002). In *Klebsiella pneumoniae,* colibactin cause DNA damage to the host cell (Lu et al., 2017). Lipopolysaccharide (LPS) may also presence on the bacterial cell surface and associated to O-antigen biosynthesis (Roberts, 1996)

Secretion system is a complex attribute applied in a pathogen and increase its virulence, especially in Gram negative bacteria. In this study, we find many putative genes associated with type II secretion system, type III secretion system, type IV

secretion system, and type VI secretion system. Type II secretion system inject the effector protein produced by the bacterium into extracellular environment (Nivaskumar and Francetic, 2014). Generally, type II, III, and IV secretion system work as a 'syringe'. The apparatus injects effector protein or toxin produced by a bacterium into a host cell which lead to the cell damage and inflammation, such as in *Aeromonas hydrophila, Aeromonas salmonicida, Bordetella pertussis, Chlamydia trachomatis, Pseudomonas aeruginosa,* and *Legionella pneumophila* (Ninio and Roy, 2007; Bohn et al., 2019).

The presence genes associated to hemolysin and secretion system has been described on previous study about genomic blueprint of *Hahella chejuensis* KCTC2396 (Jeong et al., 2005). Interestingly, the analysis of unique gene in *H. chejuensis* HN01 revealed the presence of zona occludens toxin or zot (*zot*) and cholera toxin (*ctxA*) associated genes which are not presence in *H. chejuensis* KCTC2396. The toxins originally presence in *Vibrio cholerae.* Zot toxin changes the tight junction of intestinal epithelial, leading to the flow of macromolecules to the mucosal barrier (Marinaro et al., 1999). However, there is no study about zot effect to Tilapia eggs. Therefore, we are not sure about its role in red eggs disease pathogenesis.

As the prokaryotic organism, bacteria have a 'simple' organization of their genome that allows it to achieve mutation and genomic element insertion. This genomic plasticity gives the astonishing adaptation ability of bacteria to their new environment and host. However, in this study, we did not identify the presence of virulence associated gene in the genomic island. The presence of CRISPR and CRISPR-associated genes usually are related to the bacterial immune system. The structure of CRISPR-Cas contains associated genes which encode the cutting enzyme, a leader sequence which act as a promoter for the pre-RNA synthesis, and repeats and spacer regions which is specific to certain DNA target (Couvin et al., 2018). The spacer is a short identical sequence taken from a virus that previously ever infect to a bacterium

(Louwen et al., 2014). Prophage is one of mobile genetic element that allows a bacterium to acquire some novel treats, including virulence properties and resistance determinant. Prophage is a genome segment that previously inserted from a bacteriophage (Canchaya et al., 2003). In this study, most of the identified prophages harbor unspecific protein, consisting phage-like and hypothetical protein which did not indicate the presence of mobile virulence determinant genes.

2. Conclusions and Suggestions

The findings of this study support the conclusions that:

a. Based on the genomic analysis, *Hahella chejuensis* HN01 is identic with *Hahella chejuensis* KCTC2396.
b. *Hahella chejuensis* HN01 genome harbored various virulence gene that may associated to the red eggs disease pathogenesis
c. Application of comparative genomics to *Hahella chejuensis* HN01 and *Hahella chejuensis* KCTC2396 revealed a new toxin, zot and cholera toxin (CT) harbored by HN01 genome.

With all the conclusions, a prior knowledge related to the *Hahella chejuensis* strain HN01 pathogenicity can be generated. As the study of a bacterial characterization is very wide, virulence determinant will be never enough. Therefore, to achieve a whole picture of the study, herewith the following recommendations and suggestions:

a. In order to confirm the role of virulence gene in the Red Eggs Disease pathogenesis, it is necessary to conduct a further study consisting biochemical test and molecular research.

3. Advantages of the study

Trough studying bacterial genomic characterization focusing on the virulence properties, as we propose here, can reveal new information related to the pathogenicity of *Hahella chejuensis.*

REFERENCES

Amal MNA and Zamri-Saad M 2011. Streptococcosis in Tilapia (Oreochromis niloticus): A Review. Pertanika Journal of Tropical Agricultural Science. 34: 195-206.

Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y and Wishart DS 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 44(W1): W16-21.

Arndt D, Marcu A, Liang Y and Wishart DS 2017. PHAST, PHASTER and PHASTEST: Tools for finding prophage in bacterial genomes. Brief Bioinform. bbx121.

Aziz RK, Bartels D, Best AA, 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 and Zagnitko O 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics. 9: 75.

Bacon DJ, Szymanski CM, Burr DH, Silver RP, Alm RA and Guerry P 2001. A phase-variable capsule is involved in virulence of Campylobacter jejuni 81-176. 40(3): 769-777.

Ben Hamed S, Tavares Ranzani-Paiva MJ, Tachibana L, de Carla Dias D, Ishikawa CM and Esteban MA 2018. Fish pathogen bacteria: Adhesion, parameters influencing virulence and interaction with host cells. Fish & Shellfish Immunology. 80: 550-562.

Bergh Ø, Hansen GH and Taxt RE 1992. Experimental infection of eggs and yolk sac larvae of halibut, Hippoglossus hippoglossus L. Journal of Fish Diseases. 15(5): 379-391.

Bertelli C, Laird MR, Williams KP, Simon Fraser University Research Computing G, Lau BY, Hoad G, Winsor GL and Brinkman FSL 2017a. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. Nucleic Acids Res. 45(W1): W30-W35.

Bertelli C, Laird MR, Williams KP, Simon Fraser University Research Computing Group, Lau BY, Hoad G, Winsor GL and Brinkman FS 2017b. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. Nucleic Acids Research. 45(W1): W30-W35.

Boetzer M, Henkel CV, Jansen HJ, Butler D and Pirovano W 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics. 27(4): 578-579.

Bohn E, Sonnabend M, Klein K and Autenrieth IB 2019. Bacterial adhesion and host cell factors leading to effector protein injection by type III secretion system. International Journal of Medical Microbiology.

Brenner DJ, Krieg NR and Staley JT 2008. Bergey's Manual of Systemic Bacteriology. Vol 2. 2 ed. In: Springer, Berlin. 1106.

Burrack LS and Higgins DE 2007. Genomic approaches to understanding bacterial virulence. Curr Opin Microbiol. 10(1): 4-9.

Canchaya C, Proux C, Fournous G, Bruttin A and Brüssow HJMMBR 2003. Prophage genomics. 67(2): 238-276.

Chawrai SR, Williamson NR, Mahendiran T, Salmond GPC and Leeper FJ 2012. Characterisation of PigC and HapC, the prodigiosin synthetases from Serratia sp. and Hahella chejuensis with potential for biocatalytic production of anticancer agents. Chemical Science. 3(2): 447-454.

Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y and Jin Q 2005. VFDB: a reference database for bacterial virulence factors. Nucleic Acids Res. 33(Database issue): D325-328.

Choi AHK, Slamti L, Avci FY, Pier GB and Maira-Litrán T 2009. The pgaABCD Locus of *Acinetobacter baumannii* Encodes the Production of Poly-β-1-6-N-Acetylglucosamine, Which Is Critical for Biofilm Formation. 191(19): 5953-5963.

Chou CC, Lin MC, Su FJ and Chen MM 2019. Mutation in cyl operon alters hemolytic phenotypes of Streptococcus agalactiae. Infect Genet Evol. 67: 234-243.

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M and Robles MJB 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. 21(18): 3674-3676.

Conesa A and Götz SJIjopg 2008. Blast2GO: A comprehensive suite for functional analysis in plant genomics. 2008.

Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Neron B, Rocha EPC, Vergnaud G, Gautheret D and Pourcel C 2018. CRISPRCasFinder, an update of

CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic Acids Res. 46(W1): W246-W251.

Cruz PS, Andalecio MN, Bolivar RB and Fitzsimmons K 2008. Tilapia-shrimp polyculture in Negros Island, Philippines: A review. Journal of the World Aquaculture Society. 39(6): 713-725.

Dark MJ 2013. Whole-genome sequencing in bacteriology: state of the art. Infect Drug Resist. 6: 115-123.

Darling AC, Mau B, Blattner FR and Perna NT 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 14(7): 1394-1403.

Depelteau JS, Brenzinger S and Briegel A 2019. Bacterial and Archaeal Cell Structure. In: Reference Module in Life Sciences. Elsevier.

Deurenberg RH, Bathoorn E, Chlebowicz MA, Couto N, Ferdous M, Garcia-Cobos S, Kooistra-Smid AMD, Raangs EC, Rosema S, Veloo ACM, Zhou K, Friedrich AW and Rossen JWA 2017. Reprint of "Application of next generation sequencing in clinical microbiology and infection prevention". J Biotechnol. 250: 2-10.

Dias C, Borges A, Saavedra MJ and Simoes M 2018. Biofilm formation and multidrug-resistant Aeromonas spp. from wild animals. J Glob Antimicrob Resist. 12: 227-234.

Do J, Zafar H and Saier Jr MHJMp 2017. Comparative genomics of transport proteins in probiotic and pathogenic Escherichia coli and Salmonella enterica strains. 107: 106-115.

Dong HT, LaFrentz B, Pirarat N and Rodkhum C 2015a. Phenotypic characterization and genetic diversity of Flavobacterium columnare isolated from red tilapia, Oreochromis sp., in Thailand. J Fish Dis. 38(10): 901-913.

Dong HT, Nguyen VV, Le HD, Sangsuriya P, Jitrakorn S, Saksmerprome V, Senapin S and Rodkhum C 2015b. Naturally concurrent infections of bacterial and viral pathogens in disease outbreaks in cultured Nile tilapia (Oreochromis niloticus) farms. Aquaculture. 448: 427-435.

Dong HT, Siriroob S, Meemetta W, Santimanawong W, Gangnonngiw W, Pirarat N, Khunrae P, Rattanarojpong T, Vanichviriyakit R and Senapin S 2017. Emergence

of tilapia lake virus in Thailand and an alternative semi-nested RT-PCR for detection. Aquaculture. 476: 111-118.

Edgar RC 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 5(1): 113.

Edwards DJ and Holt KE 2013. Beginner's guide to comparative bacterial genome analysis using next-generation sequence data. Microb Inform Exp. 3(1): 2.

El-Sayed A-FM 2006. Tilapia culture. In: CABI.

FAO 2018a. Fishery and Aquaculture Statistics: Global production by production source 1950-2016 (FishstatJ). *in* FAO Fisheries and Aquaculture Department [online]. 21/07/2019 ed, Rome.

FAO 2018b. The State of World Fisheries and Aquaculture 2018 - Meeting the sustainable development goals. Rome.

Feldman M, Bryan R, Rajan S, Scheffler L, Brunnert S, Tang H and Prince A 1998. Role of Flagella in Pathogenesis of *Pseudomonas aeruginosa* Pulmonary Infection. 66(1): 43-51.

Felsenstein JJE 1985. Confidence limits on phylogenies: an approach using the bootstrap. 39(4): 783-791.

Ferreira JG, Falconer L, Kittiwanich J, Ross L, Saurel C, Wellman K, Zhu CB and Suvanachai P 2015. Analysis of production and environmental effects of Nile tilapia and white shrimp culture in Thailand. Aquaculture. 447: 23-36.

Finlay BB and Falkow S 1997. Common themes in microbial pathogenicity revisited. Microbiol Mol Biol Rev. 61(2): 136-169.

Frey J and Origgi FC 2016. Type III Secretion System of Aeromonas salmonicida Undermining the Host's Immune Response. 3(130).

Glaeser SP, Kämpfer PJS and microbiology a 2015. Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. 38(4): 237-245.

Götz S, Arnold R, Sebastián-León P, Martín-Rodríguez S, Tischler P, Jehl M-A, Dopazo J, Rattei T and Conesa AJB 2011. B2G-FAR, a species-centered GO annotation repository. 27(7): 919-924.

Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J and Conesa AJNar 2008. High-throughput functional annotation and data mining with the Blast2GO suite. 36(10): 3420-3435.

Gurevich A, Saveliev V, Vyahhi N and Tesler G 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 29(8): 1072-1075.

Gyles CL and Prescott JF 2004. Themes in Bacterial Pathogenomic Mechanisms. In: Pathogenesis of Bacterial Infections in Animals. 3 ed. Iowa: Blackwell Publishing. 3-12.

Hansen GH, Bergh O, Michaelsen J and Knappskog D 1992. Flexibacter ovolyticus sp. nov., a pathogen of eggs and larvae of Atlantic halibut, Hippoglossus hippoglossus L. Int J Syst Bacteriol. 42(3): 451-458.

He X, Lu F, Yuan F, Jiang D, Zhao P, Zhu J, Cheng H, Cao J and Lu G 2015. Biofilm Formation Caused by Clinical *Acinetobacter baumannii* Isolates Is Associated with Overexpression of the AdeFGH Efflux Pump. 59(8): 4817-4825.

Jeong H, Yim JH, Lee C, Choi SH, Park YK, Yoon SH, Hur CG, Kang HY, Kim D, Lee HH, Park KH, Park SH, Park HS, Lee HK, Oh TK and Kim JF 2005. Genomic blueprint of Hahella chejuensis, a marine microbe producing an algicidal agent. Nucleic Acids Res. 33(22): 7066-7073.

Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, Lago BA, Dave BM, Pereira S, Sharma AN, Doshi S, Courtot M, Lo R, Williams LE, Frye JG, Elsayegh T, Sardar D, Westman EL, Pawlowski AC, Johnson TA, Brinkman FS, Wright GD and McArthur AG 2017. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Res. 45(D1): D566-D573.

Kayansamruaj P, Dong HT, Hirono I, Kondo H, Senapin S and Rodkhum C 2018. Genome characterization of piscine 'Scale drop and Muscle Necrosis syndrome'-associated strain of Vibrio harveyi focusing on bacterial virulence determinants. Journal of Applied Microbiology. 124(3): 652-666.

Keskin M, Keskin M and Rosenthal H 1994. Pathways of bacterial contamination during egg incubation and larval rearing of turbot, Scophthalmus maximus. Journal of Applied Ichthyology. 10(1): 1-9.

Kim D, Kim JF, Yim JH, Kwon SK, Lee CH and Lee HK 2008. Red to red - the marine bacterium Hahella chejuensis and its product prodigiosin for mitigation of harmful algal blooms. J Microbiol Biotechnol. 18(10): 1621-1629.

Koonin EV 2005. Orthologs, Paralogs, and Evolutionary Genomics. 39(1): 309-338.

Kumar S, Stecher G, Li M, Knyaz C and Tamura K 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol. 35(6): 1547-1549.

Kumar S, Stecher G and Tamura K 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol. 33(7): 1870-1874.

Lee HK, Chun J, Moon EY, Ko SH, Lee DS, Lee HS and Bae KS 2001. Hahella chejuensis gen. nov., sp. nov., an extracellular-polysaccharide-producing marine bacterium. Int J Syst Evol Microbiol. 51(Pt 2): 661-666.

Loman NJ, Constantinidou C, Chan JZ, Halachev M, Sergeant M, Penn CW, Robinson ER and Pallen MJ 2012. High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. Nat Rev Microbiol. 10(9): 599-606.

Loman NJ and Pallen MJ 2015. Twenty years of bacterial genome sequencing. Nat Rev Microbiol. 13(12): 787-794.

Louwen R, Staals RH, Endtz HP, van Baarlen P and van der Oost JJMMBR 2014. The role of CRISPR-Cas systems in virulence of pathogenic bacteria. 78(1): 74-88.

Lu M-C, Chen Y-T, Chiang M-K, Wang Y-C, Hsiao P-Y, Huang Y-J, Lin C-T, Cheng C-C, Liang C-L, Lai Y-CJFic and microbiology i 2017. Colibactin contributes to the hypervirulence of pks+ K1 CC23 Klebsiella pneumoniae in mouse meningitis infections. 7: 103.

Magnadóttir B 2006. Innate immunity of fish (overview). Fish & Shellfish Immunology. 20(2): 137-151.

Marinaro M, di Tommaso A, Uzzau S, Fasano A and de Magistris MT 1999. Zonula Occludens Toxin Is a Powerful Mucosal Adjuvant for Intranasally Delivered Antigens. 67(3): 1287-1291.

Marques MB, Kasper DL, Pangburn MK and Wessels MR 1992. Prevention of C3 deposition by capsular polysaccharide is a virulence mechanism of type III group B streptococci. Infection and Immunity. 60(10): 3986.

McDermott PF, Tyson GH, Kabera C, Chen Y, Li C, Folster JP, Ayers SL, Lam C, Tate HP and Zhao S 2016. Whole-Genome Sequencing for Detecting Antimicrobial Resistance in Nontyphoidal Salmonella. Antimicrob Agents Chemother. 60(9): 5515-5520.

Meier-Kolthoff JP, Auch AF, Klenk H-P and Göker MJBb 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. 14(1): 60.

Meier-Kolthoff JP, Hahnke RL, Petersen J, Scheuner C, Michael V, Fiebig A, Rohde C, Rohde M, Fartmann B and Goodwin LAJSigs 2014a. Complete genome sequence of DSM 30083 T, the type strain (U5/41 T) of Escherichia coli, and a proposal for delineating subspecies in microbial taxonomy. 9(1): 2.

Meier-Kolthoff JP, Klenk H-P, Göker MJIjos and microbiology e 2014b. Taxonomic use of DNA G+ C content and DNA–DNA hybridization in the genomic age. 64(2): 352-356.

Miele V, Penel S and Duret LJBb 2011. Ultra-fast sequence clustering from similarity networks with SiLiX. 12(1): 116.

Nadalin F, Vezzi F and Policriti A 2012. GapFiller: A de novo assembly approach to fill the gap within paired reads. BMC Bioinformatics. 13(SUPPL 1).

Nei M and Kumar S 2000. Molecular evolution and phylogenetics. In: Oxford university press.

Nelson EJ and Ghiorse WC 1999. Isolation and identification of Pseudoalteromonas piscicida strain Cura-d associated with diseased damselfish (Pomacentridae) eggs. Journal of Fish Diseases. 22(4): 253-260.

Ninio S and Roy CR 2007. Effector proteins translocated by Legionella pneumophila: strength in numbers. Trends in Microbiology. 15(8): 372-380.

Nivaskumar M and Francetic O 2014. Type II secretion system: A magic beanstalk or a protein escalator. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research. 1843(8): 1568-1577.

Nizet V 2002. Streptococcal β-hemolysins: genetics and role in disease pathogenesis. Trends in Microbiology. 10(12): 575-580.

Noga EJ 2010. Fish Disease: Diagnosis and Treatment. 2 ed. In: Blackwell Publishing, Iowa. 538.

O'Toole GA and Kolter R 1998. Flagellar and twitching motility are necessary for Pseudomonas aeruginosa biofilm development. 30(2): 295-304.

Pallen MJ and Wren BW 2007. Bacterial pathogenomics. Nature. 449(7164): 835-842.

Qin Y, Lin G, Chen W, Xu X and Yan Q 2016. Flagellar motility is necessary for Aeromonas hydrophila adhesion. Microbial Pathogenesis. 98: 160-166.

Rakocy JE 2005. "Subject: Cultured Aquatic Species Information Programme" (online). Available: http://www.fao.org/fishery/culturedspecies/Oreochromis_niloticus/en.

Rasmussen-Ivey CR, Figueras MJ, McGarey D and Liles MR 2016. Virulence Factors of Aeromonas hydrophila: In the Wake of Reclassification. Front Microbiol. 7: 1337.

Roberts IS 1996. The Biochemistry and Genetics of Capsular Polysaccharide Production in Bacteria. 50(1): 285-315.

Rodrigues CG, Garcia BF, Verdegem M, Santos MR, Amorim RV and Valenti WC 2019. Integrated culture of Nile tilapia and Amazon river prawn in stagnant ponds, using nutrient-rich water and substrates. Aquaculture. 503: 111-117.

Rouli L, Merhej V, Fournier PE and Raoult D 2015. The bacterial pangenome as a new tool for analysing pathogenic bacteria. New Microbes and New Infections. 7: 72-85.

Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream M-A and Barrell BJB 2000. Artemis: sequence visualization and annotation. 16(10): 944-945.

Senapin S, Dong HT, Meemetta W, Siriphongphaew A, Charoensapsri W, Santimanawong W, Turner WA, Rodkhum C, Withyachumnarnkul B and Vanichviriyakit R 2016. Hahella chejuensis is the etiological agent of a novel red egg disease in tilapia (Oreochromis spp.) hatcheries in Thailand. Aquaculture. 454: 1-7.

Siryaporn A, Kim Minyoung K, Shen Y, Stone Howard A and Gitai Z 2015. Colonization, Competition, and Dispersal of Pathogens in Fluid Flow Networks. Current Biology. 25(9): 1201-1207.

Smith A 1840. Pisces. In: Illustrations of the Zoology of South Africa. London: Smith, Elder and Co. Cornhill. 5.

Soliev AB, Hosokawa K and Enomoto K 2011. Bioactive pigments from marine bacteria: applications and physiological roles. Evid Based Complement Alternat Med. 2011: 670349.

Soto E, Halliday-Simmonds I, Francis S, Kearney MT and Hansen JD 2015. Biofilm formation of Francisella noatunensis subsp. orientalis. Vet Microbiol. 181(3-4): 313-317.

Stackebrandt E, GOEBEL BMJIjos and microbiology e 1994. Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. 44(4): 846-849.

Stapper AP, Narasimhan G, Ohman DE, Barakat J, Hentzer M, Molin S, Kharazmi A, Høiby N and Mathee K 2004. Alginate production affects Pseudomonas aeruginosa biofilm development and architecture, but is not essential for biofilm formation. 53(7): 679-690.

Stothard P and Wishart DSJB 2004. Circular genome visualization and exploration using CGView. 21(4): 537-539.

Subramaniam K, Shariff M, Omar AR and Hair-Bejo M 2012. Megalocytivirus infection in fish. Reviews in Aquaculture. 4(4): 221-233.

Suresh AV and Lin CK 1992. Tilapia culture in saline waters: a review. Aquaculture. 106(3-4): 201-226.

Surtida MB 1998. How to Operate Tilapia Hatcheries. *in* SEAFDEC Asian Aquaculture. Vol. XX. Aquaculture Department, Southeast Asian Fisheries Development Center.

Suzuki M, Nakagawa Y, Harayama S and 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(Pt 5): 1639-1652.

Thomas S, Holland IB and Schmitt L 2014. The Type 1 secretion pathway - the hemolysin system and beyond. Biochim Biophys Acta. 1843(8): 1629-1641.

Tindall BJ, Rosselló-Móra R, Busse H-J, Ludwig W, Kämpfer PJIjos and microbiology e 2010. Notes on the characterization of prokaryote strains for taxonomic purposes. 60(1): 249-266.

Tripathi C, Mishra H, Khurana H, Dwivedi V, Kamra K, Negi RK and Lal RJFiM 2017. Complete genome analysis of Thermus parvatiensis and comparative genomics of Thermus spp. provide insights into genetic variability and evolution of natural competence as strategic survival attributes. 8: 1410.

Uppanunchai A, Apirumanekul C and Lebel L 2015. Planning for Production of Freshwater Fish Fry in a Variable Climate in Northern Thailand. Environ Manage. 56(4): 859-873.

Vallenet D, Calteau A, Cruveiller S, Gachet M, Lajus A, Josso A, Mercier J, Renaux A, Rollin J, Rouy Z, Roche D, Scarpelli C and Médigue C 2017. MicroScope in 2017: an expanding and evolving integrated resource for community expertise of microbial genomes. Nucleic acids research. 45(D1): D517-D528.

Vallenet D, Engelen S, Mornico D, Cruveiller S, Fleury L, Lajus A, Rouy Z, Roche D, Salvignol G, Scarpelli C and Medigue C 2009. MicroScope: a platform for microbial genome annotation and comparative genomics. Database (Oxford). 2009: bap021.

Wang M and Lu M 2016. Tilapia polyculture: a global review. Aquaculture Research. 47(8): 2363 - 2374.

Wayne L, Brenner D, Colwell R, Grimont P, Kandler O, Krichevsky M, Moore L, Moore W, Murray R, Stackebrandt EJIJoS and Microbiology E 1987. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. 37(4): 463-464.

Wu H-J, Wang AHJ and Jennings MP 2008. Discovery of virulence factors of pathogenic bacteria. Current Opinion in Chemical Biology. 12(1): 93-101.

Yang K, Chen Q, Zhang D, Zhang H, Lei X, Chen Z, Li Y, Hong Y, Ma X and Zheng WJSr 2017. The algicidal mechanism of prodigiosin from Hahella sp. KA22 against Microcystis aeruginosa. 7(1): 7750.

Yoon S-H, Ha S-m, Lim J, Kwon S and Chun J 2017. A large-scale evaluation of algorithms to calculate average nucleotide identity. Antonie van Leeuwenhoek. 110(10): 1281-1286.

Yoon SH, Park YK, Lee S, Choi D, Oh TK, Hur CG and Kim JF 2007. Towards pathogenomics: a web-based resource for pathogenicity islands. Nucleic Acids Res. 35(Database issue): D395-400.

Zhang Y, Zhao L, Chen W, Huang Y, Yang L, Sarathbabu V, Wu Z, Li J, Nie P and Lin L 2017. Complete genome sequence analysis of the fish pathogen Flavobacterium columnare provides insights into antibiotic resistance and pathogenicity related genes. Microb Pathog. 111: 203-211.

Zhou M, Wan Q, Sarath Babu V, Qiu Q, Kou H, Lin C, Zhao L, Yang L, Li J, Huang Y and Lin L 2018. Bacterial features in tilapia (Oreochromis niloticus) and environments in a goose-tilapia polyculture model. Aquaculture. 497: 313-319.

Zhou Y, Liang Y, Lynch KH, Dennis JJ and Wishart DS 2011. PHAST: a fast phage search tool. Nucleic Acids Res. 39(Web Server issue): W347-352.

VITA

NAME	Putu Cri Devischa Gallantiswara
DATE OF BIRTH	10 February 1992
PLACE OF BIRTH	Blitar
INSTITUTIONS ATTENDED	Universitas Gadjah Mada, Indonesia (Bachelor's degree)
HOME ADDRESS	Diana court Petchburi soi *6* Ratchathewi, Bangkok, Thailand 10400
PUBLICATION	a. Poster Presentation: The 18th Chulalongkorn Universiy Veterinary Conference “Genome Characterization on The Virulence Determinants of Hahella chejuensis HN01 Causing Red Eggs Disease in Tilapia Hatcheries” b. Proceeding: The 18th Chulalongkorn Universiy Veterinary Conference “Genome Characterization on The Virulence Determinants of Hahella chejuensis HN01 Causing Red Eggs Disease in Tilapia Hatcheries”

APPENDIX

Appendix 1. The media for *Hahella chejuensis* culture

1. Tryptic Soy Agar + 2% NaCl

Tryptic Soy Agar – TrypticaseTM Soy Agar	40	grams
Distilled water	962.5	ml
NaCl 40%	37.5	ml

2. NaCl 40% (m/v)

NaCl	40	grams
Distilled water	1000	ml

Appendix 2. The best model applied in every house keeping gene based phylogenetic tree

No	Species	Protein Product	Alignment length	Model*
1	*16s rRNA*	-	1959	TN93+G+I
2	*rpoD*	Sigma D of RNA polymerase	1164	K2+G
3	*atpD*	The β subunit of ATP synthase F0F1	1389	GTR+G+I
4	*infB*	Translation initiation factor IF-2	2873	TN93+G+I
5	*recA*	Recombinase A	1417	TN93+G

*****The best model was applied according to MEGA X analyzing. The model abbreviation: GTR: General Time Reversible; K2: Kimura 2-parameter; TN93: Tamurai-Nei. The rate among site: G: gamma distribution; I: evolutionarily invariable

Appendix 3. COG Automatic Classification of Hahella sp. HN01 and Hahella chejuensis KCTC2396

No	Process	Class ID	Description	*Hahella sp.* HN01*		*H. chejuensis* KCTC2396*	
				CDS	%	CDS	%
1	Cellular processes and signaling	D	Cell cycle control, cell division, chromosome partitioning	68	1.0034	60	0.8262
2		M	Cell wall/membrane/envelope biogenesis	330	4.8694	304	4.1862
3		N	Cell motility	209	3.0840	212	2.9193
4		O	Posttranslational modification, protein turnover, chaperones	221	3.2610	220	3.0295
5		T	Signal transduction mechanisms	539	7.9534	511	7.0366
6		U	Intracellular trafficking, secretion, and vesicular transport	167	2.4642	165	2.2721
7		V	Defense mechanisms	79	1.1657	73	1.0052
8		W	Extracellular structures	4	0.0590	1	0.0138
9		Z	Cytoskeleton	2	0.0295	2	0.0275
10	Information storage and processing	A	RNA processing and modification	2	0.0295	2	0.0275
11		B	Chromatin structure and dynamics	3	0.0443	3	0.0413
12		J	Translation, ribosomal structure and biogenesis	216	3.1873	206	2.8367
13		K	Transcription	456	6.7286	445	6.1278
14		L	Replication, recombination and repair	255	3.7627	294	4.0485
15	Metabolism	C	Energy production and conversion	252	3.7185	243	3.3462
16		E	Amino acid transport and metabolism	543	8.0124	519	7.1468

17		F	Nucleotide transport and metabolism	90	1.3280	91	1.2531
18		G	Carbohydrate transport and metabolism	310	4.5743	293	4.0347
19		H	Coenzyme transport and metabolism	174	2.5675	177	2.4373
20		I	Lipid transport and metabolism	177	2.6118	178	2.4511
21		P	Inorganic ion transport and metabolism	265	3.9103	271	3.7318
22		Q	Secondary metabolites biosynthesis, transport and catabolism	173	2.5528	178	2.4511
23	Poorly characterized	R	General function prediction only	798	11.7751	760	10.4654
24		S	Function unknown	430	6.3450	415	5.7147

*MicroScope annotation

Appendix 4. Kyoto Encyclopedia Genes and Genomes (KEGG) Functional Annotations

No	MAP Number	Metabolic pathway	Number Identified Gene	
			H. chejuensis HN01*	*H. chejuensis* KCTC2396*
A		**Amino acid metabolism**		
1	MAP00220	Arginine biosynthesis	29	35
2	MAP00250	Alanine, aspartate and glutamate metabolism	27	31
3	MAP00260	Glycine, serine and threonine metabolism	37	47
4	MAP00270	Cysteine and methionine metabolism	33	42
5	MAP00280	Valine, leucine and isoleucine degradation	15	34
6	MAP00290	Valine, leucine and isoleucine biosynthesis	16	18
7	MAP00300	Lysine biosynthesis	12	20
8	MAP00310	Lysine degradation	10	26
9	MAP00330	Arginine and proline metabolism	13	27
10	MAP00340	Histidine metabolism	13	16
11	MAP00350	Tyrosine metabolism	7	12
12	MAP00360	Phenylalanine metabolism	8	33
13	MAP00380	Tryptophan metabolism	11	27
14	MAP00400	Phenylalanine, tyrosine and tryptophan biosynthesis	23	25
		Total	**254**	**393**
B		**Biosynthesis of other secondary metabolites**		

1	MAP00254	Aflatoxin biosynthesis	2	4
2	MAP00261	Monobactam biosynthesis	7	7
3	MAP00311	Penicillin and cephalosporin biosynthesis	2	3
4	MAP00332	Carbapenem biosynthesis	2	2
5	MAP00333	Prodigiosin biosynthesis	1	3
6	MAP00401	Novobiocin biosynthesis	3	3
7	MAP00402	Benzoxazinoid biosynthesis	4	4
8	MAP00405	Phenazine biosynthesis	2	2
9	MAP00521	Streptomycin biosynthesis	9	13
10	MAP00524	Neomycin, kanamycin and gentamicin biosynthesis	1	1
11	MAP00525	Acarbose and validamycin biosynthesis	2	4
12	MAP00901	Indole alkaloid biosynthesis	0	1
13	MAP00940	Phenylpropanoid biosynthesis	2	4
14	MAP00941	Flavonoid biosynthesis	0	3
15	MAP00945	Stilbenoid, diarylheptanoid and gingerol biosynthesis	0	1
16	MAP00950	Isoquinoline alkaloid biosynthesis	3	5
17	MAP00960	Tropane, piperidine and pyridine alkaloid biosynthesis	4	10
18	MAP00965	Betalain biosynthesis	0	2
19	MAP00966	Glucosinolate biosynthesis	2	2
20	MAP00999	Biosynthesis of secondary metabolites - unclassified	4	5
		Total	**50**	**79**
C		**Carbohydrate metabolism**		
1	MAP00010	Glycolysis / Gluconeogenesis	27	30
2	MAP00020	Citrate cycle (TCA cycle)	25	28

3	MAP00030	Pentose phosphate pathway	21	23
4	MAP00040	Pentose and glucuronate interconversions	5	16
5	MAP00051	Fructose and mannose metabolism	7	14
6	MAP00052	Galactose metabolism	9	13
7	MAP00053	Ascorbate and aldarate metabolism	2	7
8	MAP00500	Starch and sucrose metabolism	14	21
9	MAP00520	Amino sugar and nucleotide sugar metabolism	20	35
10	MAP00562	Inositol phosphate metabolism	8	8
11	MAP00620	Pyruvate metabolism	37	50
12	MAP00630	Glyoxylate and dicarboxylate metabolism	29	40
13	MAP00640	Propanoate metabolism	21	35
14	MAP00650	Butanoate metabolism	13	34
15	MAP00660	C5-Branched dibasic acid metabolism	7	8
		Total	**245**	**362**
D		**Energy metabolism**		
1	MAP00680	Methane metabolism	19	24
2	MAP00710	Carbon fixation in photosynthetic organisms	15	13
3	MAP00720	Carbon fixation pathways in prokaryotes	24	36
4	MAP00910	Nitrogen metabolism	14	7
5	MAP00920	Sulfur metabolism	16	18
		Total	**88**	**98**
E		**Glycan biosynthesis and metabolism**		
1	MAP00510	N-Glycan biosynthesis	0	2
2	MAP00513	Various types of N-glycan biosynthesis	1	2
3	MAP00531	Glycosaminoglycan degradation	1	2

4	MAP00540	Lipopolysaccharide biosynthesis	12	8
5	MAP00550	Peptidoglycan biosynthesis	12	14
6	MAP00603	Glycosphingolipid biosynthesis - globo and isoglobo series	1	4
7	MAP00604	Glycosphingolipid biosynthesis - ganglio series	1	2
		Total	**28**	**34**
F		**Lipid metabolism**		
1	MAP00061	Fatty acid biosynthesis	12	27
2	MAP00062	Fatty acid elongation	4	14
3	MAP00071	Fatty acid degradation	13	27
4	MAP00072	Synthesis and degradation of ketone bodies	4	12
5	MAP00100	Steroid biosynthesis	0	1
6	MAP00120	Primary bile acid biosynthesis	0	2
7	MAP00140	Steroid hormone biosynthesis	0	4
8	MAP00561	Glycerolipid metabolism	6	15
9	MAP00564	Glycerophospholipid metabolism	12	21
10	MAP00565	Ether lipid metabolism	0	3
11	MAP00590	Arachidonic acid metabolism	1	1
12	MAP00591	Linoleic acid metabolism	0	1
13	MAP00592	alpha-Linolenic acid metabolism	5	18
14	MAP00600	Sphingolipid metabolism	0	1
15	MAP01040	Biosynthesis of unsaturated fatty acids	3	5
		Total	**60**	**152**
F		**Metabolism of cofactors and vitamins**		
1	MAP00130	Ubiquinone and other terpenoid-quinone biosynthesis	6	11
2	MAP00670	One carbon pool by folate	12	12

3	MAP00730	Thiamine metabolism	4	12
4	MAP00740	Riboflavin metabolism	8	7
5	MAP00750	Vitamin B6 metabolism	9	6
6	MAP00760	Nicotinate and nicotinamide metabolism	10	17
7	MAP00770	Pantothenate and CoA biosynthesis	18	17
8	MAP00780	Biotin metabolism	10	17
9	MAP00785	Lipoic acid metabolism	0	2
10	MAP00790	Folate biosynthesis	19	18
11	MAP00830	Retinol metabolism	2	4
12	MAP00860	Porphyrin and chlorophyll metabolism	22	25
		Total	**120**	**148**
G		**Metabolism of other amino acids**		
1	MAP00410	beta-Alanine metabolism	7	15
2	MAP00430	Taurine and hypotaurine metabolism	5	4
3	MAP00440	Phosphonate and phosphinate metabolism	3	3
4	MAP00450	Selenocompound metabolism	8	9
5	MAP00460	Cyanoamino acid metabolism	8	10
6	MAP00471	D-Glutamine and D-glutamate metabolism	14	18
7	MAP00472	D-Arginine and D-ornithine metabolism	0	1
8	MAP00473	D-Alanine metabolism	3	4
9	MAP00480	Glutathione metabolism	21	31
		Total	**69**	**95**
H		**Metabolism of terpenoids and polyketides**		
1	MAP00281	Geraniol degradation	4	12
2	MAP00522	Biosynthesis of 12-, 14- and 16-membered macrolides	0	1

3	MAP00523	Polyketide sugar unit biosynthesis	5	8
4	MAP00900	Terpenoid backbone biosynthesis	10	14
5	MAP00903	Limonene and pinene degradation	1	10
6	MAP00908	Zeatin biosynthesis	1	2
7	MAP00981	Insect hormone biosynthesis	0	2
8	MAP01051	Biosynthesis of ansamycins	1	1
9	MAP01053	Biosynthesis of siderophore group nonribosomal peptides	3	4
10	MAP01055	Biosynthesis of vancomycin group antibiotics	3	3
		Total	28	57
I		**Nucleotide metabolism**		
1	MAP00230	Purine metabolism	50	62
2	MAP00240	Pyrimidine metabolism	33	34
		Total	**83**	**96**
J		**Translation**		
1	MAP00970	**Aminoacyl-tRNA biosynthesis**	**26**	**26**
K		**Xenobiotics biodegradation and metabolism**		
1	MAP00361	Chlorocyclohexane and chlorobenzene degradation	0	3
2	MAP00362	Benzoate degradation	4	18
3	MAP00363	Bisphenol degradation	0	1
4	MAP00364	Fluorobenzoate degradation	0	2
5	MAP00623	Toluene degradation	1	2
6	MAP00625	Chloroalkane and chloroalkene degradation	2	6
7	MAP00626	Naphthalene degradation	2	3
8	MAP00627	Aminobenzoate degradation	2	16
9	MAP00642	Ethylbenzene degradation	2	3

10	MAP00643	Styrene degradation	2	4
11	MAP00791	Atrazine degradation	3	3
12	MAP00930	Caprolactam degradation	2	11
13	MAP00980	Metabolism of xenobiotics by cytochrome P450	5	16
14	MAP00982	Drug metabolism - cytochrome P450	6	20
15	MAP00983	Drug metabolism - other enzymes	15	32
16	MAP00984	Steroid degradation	0	2
		Total	**46**	**142**

*Performed by MicroScope pipeline