Coursework Declaration and Feedback Form

The Student should complete and sign this part

Student Number:2511633N	Student Name: FENG NI
Programme of Study (e.g. MSc in Electronics and Electrical Engineering): MSc in Computer system Engineering	
Course Code: ENG5059P	Course Name: MSc Project
Name of **First** Supervisor: Dr Umer Zeeshan Ijaz	Name of **Second** Supervisor:
Title of Project: Whole genome functional analysis of Alcanivorax, an oil degrading bacterium	

Declaration of Originality and Submission Information

I affirm that this submission is all my own work in accordance with the University of Glasgow Regulations and the School of Engineering requirements Signed (Student) : FENG NI	E N G 5 0 5 9 P
Date of Submission : August 20, 2021	

Feedback from Lecturer to Student – to be completed by Lecturer or Demonstrator	
Grade Awarded: Feedback (as appropriate to the coursework which was assessed):	
Lecturer/Demonstrator:	Date returned to the Teaching Office:

University of Glasgow

Whole genome functional analysis of Alcanivorax, an oil degrading bacterium

Student name: FENG NI
Student number: 2511633N
Supervised: Dr Umer Zeeshan Ijaz
Co-supervisor: Ciara Keating

August 20, 2021

A thesis submitted in partial fulfilment of the requirements for the degree of
MASTER OF SCIENCE IN COMPUTER SYSTEM ENGINEERTING

Abstract

Crude oil is one of the importance resources in the Earth People use it to produce fuel, lubricating oil, medicine and plastic. It is highly connected with humans' activities. However, the crude oil is a toxicity substance, which contains benzene, aldehydes and phenols. The toxicity compound is able to damage the health of humans, animals and plants. The serious oil pollution will heavily break the local ecological system. With the increasing requirement of crude oil, the oil trade is more and more frequent, which increases the risk of crude oil spill. The oil degradation is more concerned today. Biodegradation as an environmentally friendly and efficient way to degrade the oil is study in this project. The purpose of this project is to analyze the feasibility of *Alcanivorax* in crude oil degradation. This project analysis 15 genome sequences of *Alcanivorax* by using the software tools PROKKA, ROARY, METABOLIC and R Studio. The ROAYR built a pan genome model of *Alcanivorax*. The pan genome images indicate that the species in *Alcanivorax* have their own particular genes to express the difference from others species in *Alcanivorax*. The METABOLIC analyzed the metabolism of *Alcanivorax*, which demonstrates that *Alcanivorax* has ability to degrade the crude oil component including hydrocarbon, hydrogen sulfide and halogenated compounds degradation. The method Jaccard processing on the R Studio shows that among the *Alcanivorax*, 8 varieties of *Alcanivorax* are dissimilar to each variety in *Alcanivorax*. Others are highly similar to each other.

Acknowledgements

I wish to thank my supervisors, Dr Umer Zeeshan Ijaz and Dr Ciara, for the continuous supports and professional guidance from the very early stage to the completion of this dissertation. Their rich knowledge and expertise in the field of genetics and their instructional feedbacks have encouraged me to explore my research interests in this dissertation and helped me understand how to analyses datasets.

In addition, I would like to thank my parents who gave both financial and emotional support to my pursuit of master degree. At last, I would like to appreciate my dearest friend, Yingying Zhong who helped and encouraged me.

Contents

1. Introduction

1.1 background of crude oil

Crude Oil is one of the important natural resources in the Earth, which is widely utilized in industrial manufacturing. It is a heterogeneous mixture of various hydrocarbons including paraffin, naphthenic hydrocarbon, aromatic hydrocarbon and organohalogen compounds*istry and Technology of Petroleum FOURTH EDITION*, no date). Currently, there are 2 statements of crude oil forming. One is organic petroleum origin and another one is abiogenic petroleum origin. Supporters of organic petroleum origin believe that petroleum is formatted from organic substance, such as bacterium, phytoplankton, zooplankter and higher plants(Kvenvolden, 2006). At the final stage of the life cycle of these creatures, part of their bodies is oxidized and some of their bodies are preserved in sediments under a suitable condition. Then the sediments sink into deeper places through million years. These sediments suffer heat and million tons of pressure from upper layer of the Earth. These forces led the structure of organic compounds be converted. Anaerobic decay is the first step of the conversion. The corpses of creatures are hydrolyzed expect the fats and waxes. Some organic products are created here, such as phenolic and aldehyde. The second step, a waxy substance called kerogen is formed. Eventually, kerogen cleaves in deeper position of earth's mantle with a higher heat and more pressure condition. The form of Kerogen is solid. It can be oxidized when it contacts with oxygen(Vandenbroucke and Largeau, 2007). It can be also further converted to petroleum in the further deeper Earth' crust over million years. For abiogenic petroleum origin, the researchers believe some natural gas deposits were formed in inorganic way deep in the Earth's mantle. However, this theory has not been widely accepted. Due to the long-time forming process, crude oil is a non-renewable resource. Huge number of petroleum exists under the ground. It can be the source to promote the human activities and it can also be the pollutant to damage human.

Back in the 19 century, crude oil had been used in the industrial revolution as a fossil fuel. With the invention of internal combustion engine, the requirement of crude oil is in rapid increase(Safieddin Ardebili *et al.*, 2020). At present, more and more applications of crude oil were found. The function of crude oil beyond to be applied as fuel. The crude oil can be used in multiple aspects. It can be made into plastic, lubricant, bituminous pavement, asphalt, nylon fiber, synthetic rubber and medicine etc(White and Reid, 2018). Today, despite new energy resources contribute to a part of global energy consumption, the human activities still highly rely on crude oil. According to the given data, the world oil tankers (246 million dwt) in seaborne trade has increased exponentially since 1990 and the capacity has risen to 601 million dwt by 2020 (*• Global seaborne trade - oil tanker capacity 2020 | Statista*, no date). In addition, crude oil production shows an upward tendency over decades(*Crude Oil Production*, no date).

With the increasing demand of energy from human activities, more and more crude oil needs to be explored. With the frequent crude oil trade and large crude oil exploitation, the risk of crude oil spill, which posts a detrimental threaten to marine ecosystem, increases.

1.2 Crude oil spill

The natural seep of crude oil is a normal phenomenon. In 2002, an essay estimated that there are about 600,000 tons of crude oil seeping naturally into marine per year. The natural seeping oil approximately accounts for 47% of all crude oil that enters into ocean(Kvenvolden and Cooper, 2003). Because the Earth has the ability to adjust itself, the nature seeping oil will not cause serious ecological damage. The negative impact of crude oil can be counteracted by the carbon cycle system of nature. The carbon cycle system includes spreading, dissolution, photooxidation and biodegradation. At the active natural seep points, microbial communities which feed on oil have even formed.

However, crude oil spill caused by the human activities needs more concerns. Commonly, oil spill accidents occur when oil transport accidents happen and the offshore oil platforms spill. For instance, in 2018, the oil tanker Sanchi collied with Hong Kong-registered freighter CF Crystal. Then Sanchi was on fire and then exploded. This major accident resulted in releasing approximately 136,000 ton of condensate and 2000 ton of oil (Mohammadiun *et al.*, 2021). In 2010, the semisubmersible oilrig Deepwater Horizon exploded. About 700,000 ton of oil spilled into Gulf of Mexico in 12 weeks. More than 1600 miles of coast line are covered by the crude oil. This accident has become one of the most serious oil spill incidents (Zengel *et al.*, 2021).The worse is that in 2004, the Hurricane Ivan heavily damaged America, which causes about 20.5 billion dollars loss. At the same time, a crude oil production platform sank nearby Louisiana coast after the Hurricane Ivan. It is estimated that there are about 300 and 700 barrels of crude oil spilling from the platform to the ocean every day after the Hurricane Ivan since 2004. The continuously spilling can led to a negative effect which is worse than Deepwater Horizon(*Tropical Cyclone Report | Enhanced Reader*, no date).

1.3 The effect of crude oil

The hazardous components in crude oil includes alkanes like various of unbranched alkanes and branched paraffin etc. and aromatic hydrocarbon like polycyclic aromatic hydrocarbon and benzene etc. When crude oil is spilled, the alkane and aromatic with lower carbon number is easy to volatilize. The smell of crude oil is from the volatile

components of itself. The volatile gas is fat-soluble, which is easy to be absorbed by humans and animals. Although the weathering can decrease a part of lighter crude oil compound when the crude oil spill happens, the rest of crude oil compounds still dangerous(*Overview of Petroleum Product Poisoning - Toxicology - Veterinary Manual*, no date). The crude oil component like polycyclic aromatic hydrocarbons can exists in the environment for a long time. The creatures can also be polluted by this component(Borges-Ramírez *et al.*, 2021).

Major crude oil spilling accidents always poured plenty of crude oil into ocean in a short time, which heavily damage the local ecological balance. According to the investigation of Deepwater Horizon incident, investigator found that certain wavelengths of ultraviolet light which reacts with crude oil will increase the toxicity of crude oil by ten times. The toxicity of crude oil will cause injures and even deaths to growing invertebrates and fishes(*Plan for Deepwater Horizon Oil Spill Natural Resource Injury Restoration: An Overview*, no date). In addition, birds' feathers are sensitive to oil. When birds' feathers are stained with oil, their feathers will lose the insulating properties and they cannot maintain their body temperature normally anymore(*Feathers As Insulation*, 2009). As a result, they will die because of the low body temperature. The impact of crude oil on aquatic organisms is also negative. The immune system of aquatic organisms is weakened when the aquatic organisms live in the environment permeated by crude oil for a long period of time. This negative impact will further lead to a malformed growth of the organisms. The injury Quantification of Deep Horizon estimate that about 2 to 5 trillion fish larvae dead in this accident(*Plan for Deepwater Horizon Oil Spill Natural Resource Injury Restoration: An Overview*, no date).

In the survey of the Exxon Valdez incident from NOAA's Auke Bay Laboratory, they found that even the spilled crude oil in sediment has been weathered more than ten years and the concentration of polycyclic aromatic hydrocarbon is in a lower degree, the polycyclic aromatic hydrocarbon still has negative affect to marine organisms with lower toxicity(*The Toxicity of Oil: What's the Big Deal? | response.restoration.noaa.gov*, no date). Some of marine organisms can defense themselves from this negative impact, but the creatures like clams which live in sand are still polluted by the spilled crude oil. Sea otters as the predator of clams can spreading the toxicity of crude oil to other food chains by eating the polluted clams.

The effect of the toxicity of hydrocarbon involves respiratory system, gastrointestinal system, cortical system and central nervous system(Tormoehlen, Tekulve and Nañagas, 2014). The effects of crude oil on people are also serious. For example, benzene, a typical toxic component of crude oil is easy to volatilize. It can be absorbed by skin when it touches the skin. Living in the low concentration of benzene for long-term will lead to chronic intoxication(*CDC | Facts About Benzene*, no date). The symptoms of the chronic intoxication caused by benzene are that the amount of white blood cells and red blood cells will decrease and thus cause hemophilia. The high concentration of benzene can result in acute toxicity which will heavily paralyze the Central Nervous

System.

Other crude oil compounds are also toxicity. For instance, formaldehyde and can increase the risk of having cancer. 1,3-Butadiene can injure the genital system of male and female(*Petroleum Products - Environmental Exposure from Refineries - Proposition 65 Warnings Website*, no date).

Despite the crude oil pollution sometimes happens far away from daily lives, the toxicity of crude oil can still affect human life by biomagnification of food chain. Polycyclic aromatic hydrocarbon can increase the risk of cancer because it enters human bodies through the food chain.

1.4 Measures

There are some methods applied in the degradation of spill crude oil. The first one is Dispersant Chemicals which are also the common measure adopted in the crude oil spill accident Dispersant Chemicals are used to spray on the spilled crude oil. It can separate the crude oil film into small crude oil droplet. It will not directly decrease the crude oil, but it can promote the crude oil spreading into ocean(*Dispersants*, no date). The positive of this measure is that limit the lateral diffusion of spilled crude oil, which protects the coastal plants and animals polluted by the crude oil. The negative aspects are that this measure will put the creatures in vertical areas in dangerous, due to the crude oil vertically sink into the water. In addition, Dispersant Chemicals are still toxicity for creatures, which causes higher toxicity by using with crude oil. Basically, this is the method sacrificing a partial ecological environment to save another ecological environment. The second measure is the degradation of plants. According to the previous study, mangrove forest was proved that it has ability to process the spilled crude oil. The study shows that mangrove forest can largely fixate the toxic crude oil hydrocarbon(Waryszak *et al.*, 2021). On the polluted areas, the mangrove forest converts the toxic crude oil components to high organic matter and accumulate the organic matter as sediment. The third measure is biodegradation. Due to the oil spill has happened for millions of years, the microorganism in the Earth has evolved to the ability of crude oil degradation. The microorganism in crude oil degradation mainly uses as their major source of carbon. Marine bacteria of hydrocarbon degradation have various metabolic mechanism in different positions. Some of them grow in deep ocean are able to degrade in anaerobic environment. On contrary, the species living shallow place can degrade crude oil in aerobic environment. The fourth measure is natural degradation through weathering(Hazen, Prince and Mahmoudi, 2016). However, it is impossible to let the spilled oil be processed by natural degradation because it spends too much time. Therefore, as an economical and environmentally friendly method s the world's n.

In the polluted area, various of microorganisms which feed on oil have been isolated. These kinds of microorganisms include , *Cobetia* and ect(Olivera *et al.*, 2009). *Alcanivorax*, a kind of hydrocarbonoclastic bacteria, is the study object of this dissertation. *Alcanivorax*, a various of hydrocarbonoclastic bacteria, is able to grow under the environment of salt of high concentrations. Its shape is rod-like. It will become red under the Gram Stain. Normally this strain grows in aerobic environment with crude oil as nutriment(Jagtap *et al.*, 2021).

1.5 Alcanivorax

Alcanivorax as a special hydrocarbon degrader rarely exists in the areas where there is no crude oil pollution. While in the crude oil spilled areas such as the Deepwater Horizon and the Sanchi tanker exploration where there was a crude oil incident, this species is found in rich. Under the crude oil spilled areas where nitrogen nutriment and phosphorus nutriment are sufficient, this species can rapidly multiply in quantity. This species also is the dominant strain in crude oil pollution areas(Warr *et al.*, 2018).

1.6 Purpose

This project uses *Alcanivorax* as an experimental subject and explores the feasibility of using it to degrade crude oil biochemically. The purpose of this project is aim to determine the potential of *Alcanivorax* species to degrade crude oil. This will be achieved through the following objectives:

- *Alcanivorax* species genome sequences will be downloaded from NBCI
- The genome sequences will be annotated by PORKKA
- The ROARY will be used to build the model of the pan genome of *Alcanivorax*
- Analyzing the metabolism of *Alcanivorax* by using METABOLIC
- Using R Studio to process the genes data with Jaccard method
- Using R Studio to do the coincidence analysis of the genes of *Alcanivorax*

2.**Methods**

2.1 Data description

The main purpose of this project is to process and analyze the genes of oil degradation bacteria. All the processes and analysis method were with the guidance of Dr Umer Ijaz and Dr Ciara Keating of Glasgow University. The first step was to collect suitable genomes. All the genome sequence data are second-hand sources. The genomes sequences collected from the database of (NCBI) are the reference genomes sequence. There are 15 species considered in this project as shown in table 1, including 15 genomes.

Number	Species	Genomes
1	Alcanivorax borkumensis SK2	RefSeq: GCF_000009365.1
2	Alcanivorax dieselolei B5	RefSeq: GCF_000300005.1
3	Alcanivorax gelatiniphagus	RefSeq: GCF_005938655.1
4	Alcanivorax hongdengensis A-11-3	RefSeq: GCF_000300995.1
5	Alcanivorax indicus	RefSeq: GCF_003259185.1
6	Alcanivorax jadensis T9	RefSeq: GCF_000756655.1
7	Alcanivorax marinus	RefSeq: GCF_016785095.1
8	Alcanivorax mobilis	RefSeq: GCF_002864685.1
9	Alcanivorax nanhaiticus	RefSeq: GCF_000756665.1
10	Alcanivorax pacificus W11-5	RefSeq: GCF_000299335.2
11	Alcanivorax profundi	RefSeq: GCF_003597125.1
12	Alcanivorax profundimaris	RefSeq: GCF_015265435.1
13	Alcanivorax sediminis	RefSeq: GCF_009601165.1
14	Alcanivorax venustensis ISO4	RefSeq: GCF_015356855.1
15	Alcanivorax xenomutans	RefSeq: GCF_900217905.1

Table 1: the genome sequences download from NCBI

2.2 PROKKA and ROARY work flow

PROKKA is the second step to process genome sequences files in this project. The target strain genome sequence files of *Alcanivorax* would be downloaded and will be annotated by PROKKA in Linux system. Genome annotation is proceeded to identify and categorize all the relating features on a genome sequence (Richardson and Watson, 2012). One basic function of it is to present the location of expected coding regions and putative products. Indeed, there are a variety of online annotation severs to be chosen such as Prokaryotic Genomes Automatic Annotation Pipeline provided by the

NCBI, RAST, a web server that can annotate bacterial and genomes and xBASE2 which shares similar functions with RAST. They are though very effective to some extent they are not as useful as PROKKA when the yield of genomes outweighs. In order to fully annotate a genome, PROKKA, a Linux software tool is adopted. It produces standards-compliant output files for further study or visualizing in genome browsers. PROKKA has an integrated software tools which can annotate genomic bacterial sequences in details reliably. If needed, multiple processing cores are feasible. The annotation of a typical bacterial genome sequence can be finished within 10 min on a personal computer equipped with a quad core. Regarding to the input of the alcanivorax, it needs to be preassembled in FASTA (FNA) format. With the external feature prediction tool, PROKKA can recognize the position of the genomic features with contigs of target genome in an accurate and fast way. Afther the genome annotation, there are 10 files exported from PROKKA process as outcome, including three FASTA files, one contig sequence file, one feature table, one sequin editable file, one genbank file, one GFF3 file, one log file and one statistic summary file(Seemann, 2014). The workflow of PROKKA is shown as Figure 1 below.

One of them is GFF(GFF3), which is used to be the input file of ROARY, which is a rapid large-scale prokaryote pan genome analysis system. ROARY can process the exponential data of prokaryotic genomes and identify the major genes accurately. By extracting the coding region from the input, ROARY converts the input into protein sequences and screens DNA fragments and pre-clustered interactivity of CD-HIT (Fu et al., 2012). Sequences are clustered with MCL and the pre-clustered results are combined with the results of MCL. This procedure is significant because it makes it feasible to analyze datasets using normal set-up hardware. Analyzing the similar sequence, homologous groups which contain paralogs are divided into different groups of true orthologs. According to the order of occurrence in the input sequences, a graph is generated to present the relationships between each cluster and provide context of each gene. Given the considerations of accuracy, running time and memory usage of ROARY, it would be the first choice to analyze the pan genome. The most complete standalone pan genome tools are PanOCT using a conserved gene neighborhood to put proteins inro orthologous cluster; LS-BSR using a pre-cluster step to identify genes to families and PGAP taking the annotation, performing an all-against-all BLAST and clustering the outcomes to generate a pan genome. Both PanOCT and PGAP demand an all-against-all comparison through Basic Local Alignment Search Tool and thus the processing time grows on 2x time complexity with the size of input data and are hard to achieve. Further, they also require the 2x memory complexity, which quickly goes beyond RAM available in high performance servers. Although LS-BSR use a pre-clustered method to reduce the running time, it is less sensitive (Sahl et al., 2014) However, ROARY make it possible to construct the pan genome of thousands of prokaryote samples on a standard desktop without sacrificing the accuracy of results. With a single CPU, ROARY can generate a pan genome which includes 1000 isolates within 4.5 hours by using 13GB of RAM\

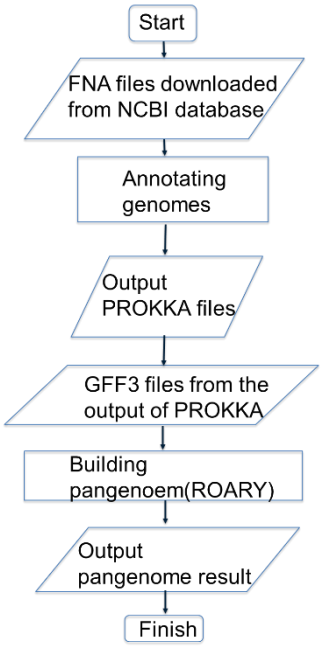

Figure 1: The workflow of PROKKA and ROARY

2.3 METABOLIC

METABOLIC is the abbreviation of Metabolic and BiogeOchemistry analyses in microbes. It is a tool that uses genomes to identify the traits of *Alcanivorax* strain (including 15 genomes) in this project. Aiming at analyze the increasing datasets of metagenomics and single-cell genomes, METABOLIC predicts the microbial metabolism via the annotation of the function of enzyme for *Alcanivorax* strain by a large number of established datasets. Different from other services, the results of METABOLICT are not too detailed for users. Regarding to the working mechanism for METBOLIC, a phenomenon of interdependent and cross-linked metabolic and biogeochemical mechanism and biogeochemical interactions with in a community can present the advantage of metabolic activities which render a high resilience to the community and make it stable. Therefore, it is meaningful to investigate this phenomenon which is also called metabolic handoffs. The genome-scale of METABOLIC process in this project consists of the annotation of genomes of *Alcanivorax*, the verification of the enzymes in *Alcanivorax* community, the analysis of metabolic pathway and the calculation of contributions to individual biogeochemical transformations and cycles. For microbial communities, METABOLIC can analyze the genome abundance in *Alcanivorax*, the potential of promoting the circulation of particular matter. The workflow of METABOLIC shown as Figure 2.

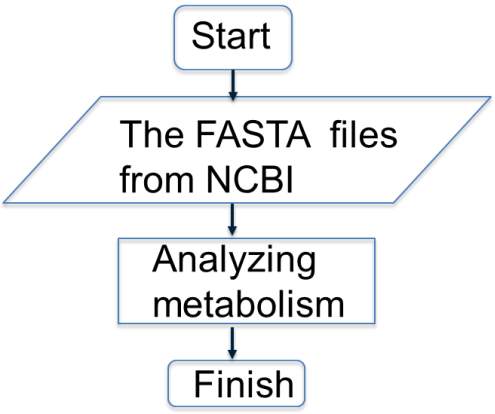

Figure 2: the workflow of METABOLIC

2.4 Statistical analysis

R language as a programming language is outstanding in data statistics, data analysis and machine learning. It is free of charge and open-source. It is compatible with Window, Linux and Mac. The advantage of R language is that it contains a large number of packages involving statistical approach.

R studio as a development platform of R language use in this project. It mainly includes four areas (Figure 3) on its main interface, a console, a source editor, a workspace and history window and a plotting and files management window.

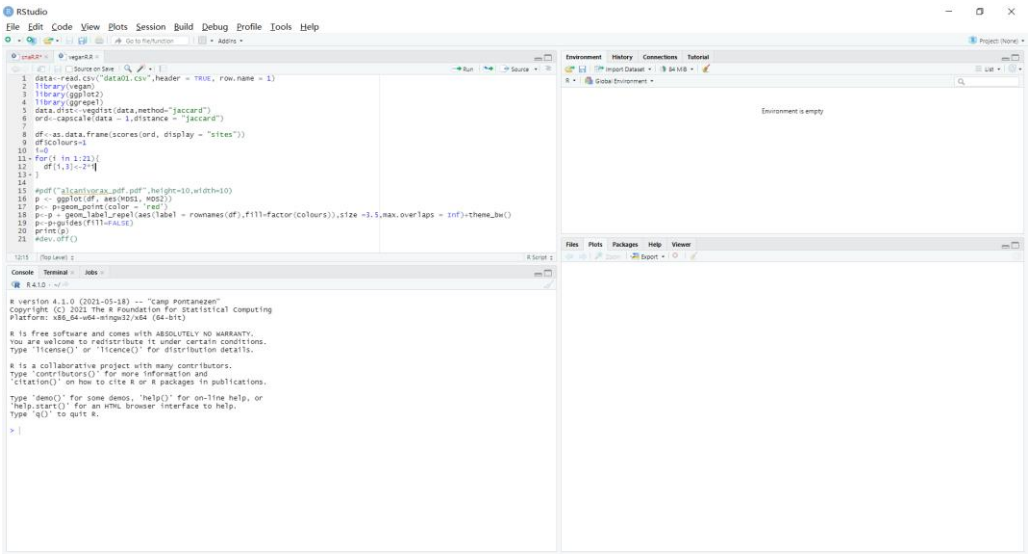

Figure 3: the interface of R Studio

In this project, the distribution diagram of genomes relationship models by R studio by using the R packages vegan, ggplot2 and cna. Vegan is a community ecology package, which offers the tools used to express the community analysis. It contains

diversity algorithms, ordination methods and dissimilarities algorithms. ggplot2 is a plotting tool. It can map the picture in detail according to the requirement of user. In this project, ggplot2 expects the outcome of vegan as input to visualize the genomes data.

The Jaccard as a method of vegan was used in this project. The Jaccard expects the gene absent and present file as input. Jaccard takes each genome as a vector with a group of genes as coordinate. The sample is shown as table 2. Jaccard uses these vectors to calculate the dissimilarities between each genome.

	Gene1	Gene2	…
Genome 1	1	0	…
Genome 2	0	0	…
…	1	1	…

Table 2: the sample input format of vegan

Coincident Analysis (can) is a method which is used to model the casual relationship among each gene. It expects binary file as input, the file format is the with the input file of Jaccard. It infers the casual relationship based on the INUS theory. INUS is the abbreviation of Insufficient but Necessary part of an Unnecessary but Sufficient condition. It is able to be applied in the responsibility determination of criminal law. If incident A is regarded as the INUS condition of incident P, it must exist a conditional combination: $(A \cap B) \cup C \rightarrow P$. B can be an incident or a group of incidents which happens together with the incident A to be able to occur. Incident C represents the others independent incidents which is able to occur the incident P happen. Under this conditional combination, if incident P happens, there is an independent condition for A and B to happen together. One Of A and B independently happens is not able to occur the incident P. This theory is also able to be applied in the causal relationship analysis of genes.

3.Result

3.1 Result of PROKKA

PROKKA successfully annotated 15 genomes of *Alcanivorax*. It exported the FNA file and GFF3 file and others accessory files.

3.2 Result of ROARY

This is what I found after modeling the pangenome with 15 sample genomes by ROARY.

There is a statics summary as shown in table 3. The number of genes including in pan genome analysis is 27299. There are 367 core genes which exist in all individuals. The number of shell gene which present in two or more strains is 3800. The cloud genome which only found in a single strain has 23132 genes. Besides, there is a file containing 27299 genes with their relevant enzymes as outcome.

Core genes	(99% <= strains <= 100%)	367
Soft core genes	(95% <= strains < 99%)	0
Shell genes	(15% <= strains < 95%)	3800
Cloud genes	(0% <= strains < 15%)	23132
Total genes	(0% <= strains <= 100%)	27299

Table 3: the statics summary of pan genome of Alcanivorax

Figure 1 gives a direct description by histogram. It represents the frequency of genes present in genomes. It shows that a large number of genes are belong to cloud genes which only present in one genome. Figure 2 demonstrate the pan genome matrix of *Alcanivorax*. Figure 3: the pan genome pie of *Alcanivorax*

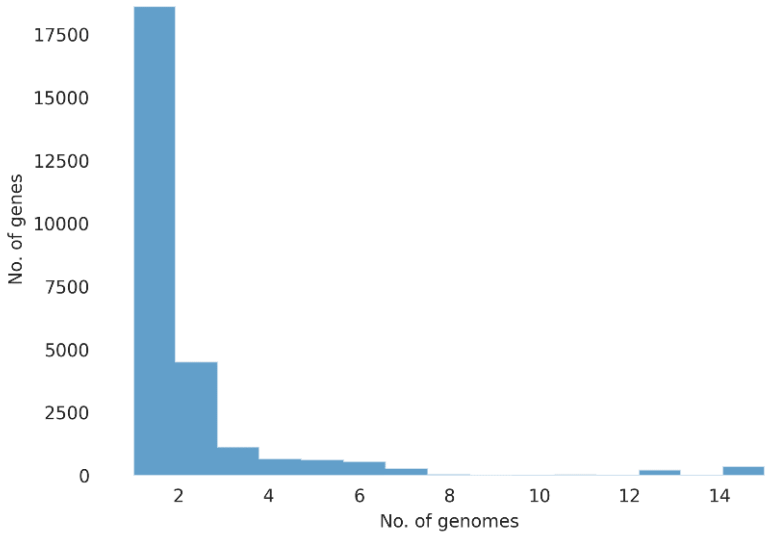

Figure 1: frequency of genes of *Alcanivorax*

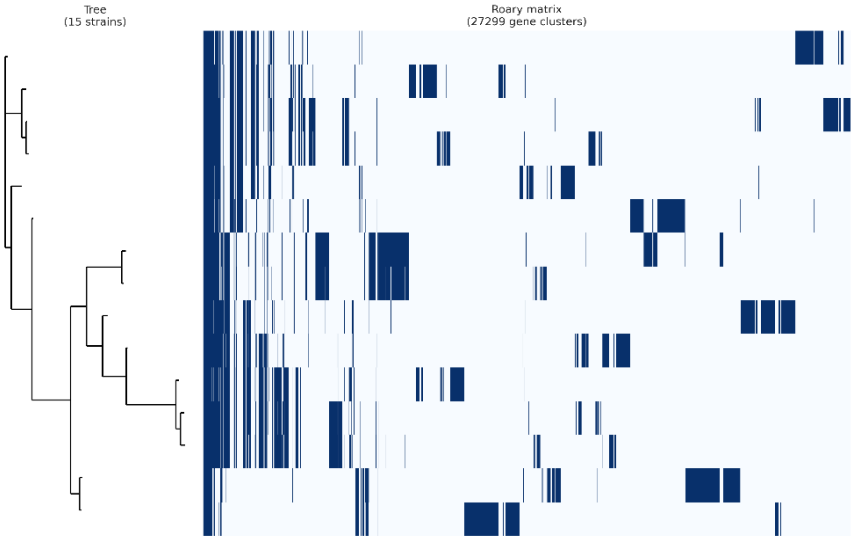

Figure 2: pan genome matrix of Alcanivorax

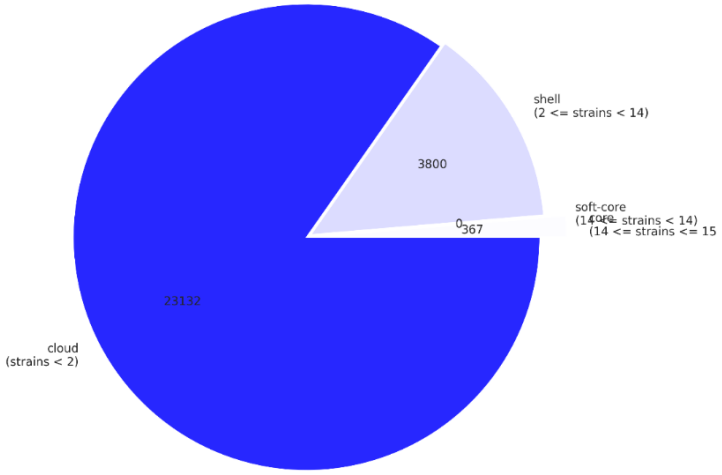

Figure 3: the pan genome pie of Alcanivorax

In this step, the outcome of ROARY export a file showing the presence and absence of genes of *Alcanivorax*. From this file, some important genes which are connected with the crude oil biodegradation are analysis in this report.

①Alkane hydroxylase gene(*alkB*)

In the pangenome export file of *Alcanivorax*, there is a series of genes called *alkB*. *AlkB* includes *alkB1* and *alkB2*. These two genes are important for the crude oil degradation, which exist in many petroleum degrading strains(van Beilen *et al.*, 2004).

AlkB1 series corresponds to the enzyme *alkane 1-monooxygenase 1* and *alkB2* series corresponds to the enzyme *alkane 1-monooxygenase 2*. These two enzymes are used to catalyze the chemical reactions of degrading alkane. Under the aerobic environment, *Alkane 1- monooxygenase* can hydroxylate n-alkanes at the terminal position. This reaction transforms n-alkanes to relevant alcohols. According to the result of previous experiment, *alkane 1-monooxygenase 1* catalyzes the hydroxylation reactions with the alkane of C5 to C12. *Alkane 1-monooxygenase 2* play a part in the hydroxylation reactions using the alkane of C8 to C16(Hara *et al.*, 2004).

②P450 136

In pangenome, 20 genes are related to putative cytochrome *P450 136*. These genes are in shell genes and cloud gene. At present, there is no study shows the function of putative cytochrome *P450 136* in the crude oil degradation of *Alcanivorax*. However, according from the given data, putative cytochrome *P450 136* has similar functions to Putative cytochrome *P450* alkane hydroxylase (*CYP153*) which plays important role in oil degradation. Both includes the function , , and (Gaudet *et al.*, 2011). They are both belong to a superfamily of enzymes *Cytochromes P450* which is also can be found in humans' body.

3.2 Result of METABOLIC

In the analysis of METABOLIC, the outcome shows that there are 282 genes which attend in metabolism. The categories connected with the crude oil degradation are brought into focus in this project, including fatty acid degradation, aromatics degradation, C1 metabolism, methane metabolism and halogenated compound utilization. These genes help *Alcanivorax* to build a system to degrade the crude oil in crude oil pollution areas.

The METABOLIC also exported a outcome of 4 cycles of matter, including carbon cycle (Figure 4), sulfur cycle (Figure 5), nitrogen cycle (Figure 6) and other substance cycles (Figure 7).

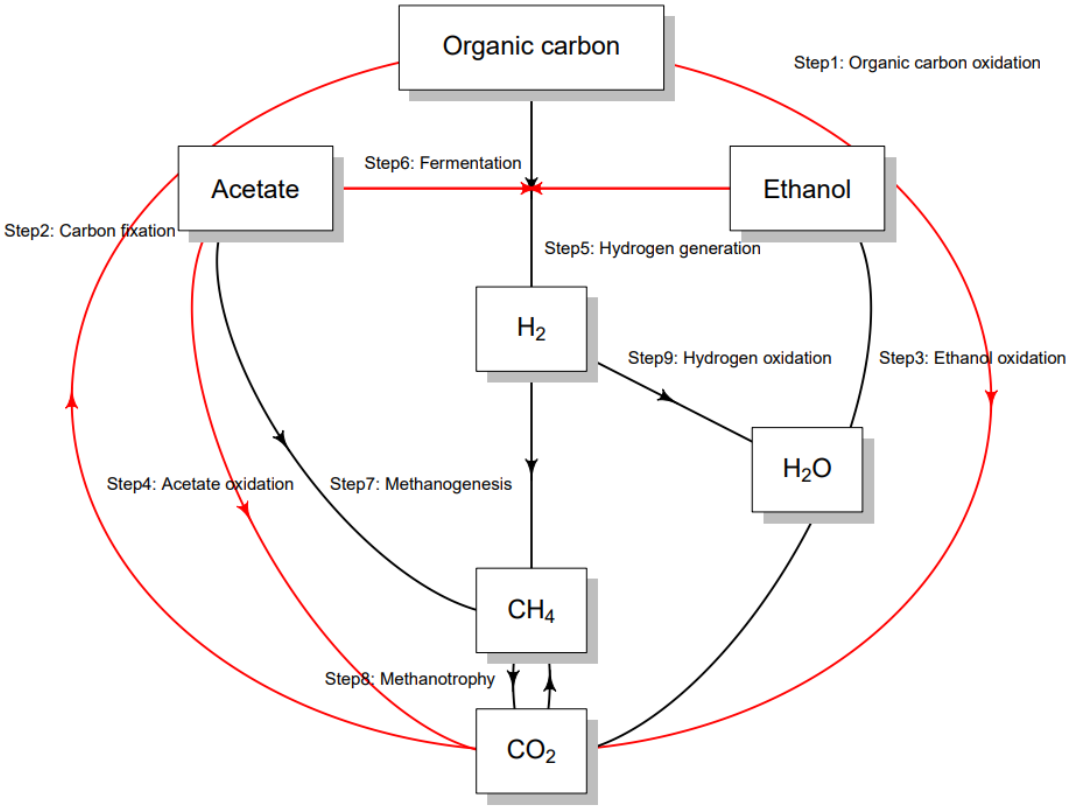

Figure 4: Carbon Cycle of *Alcanivorax*

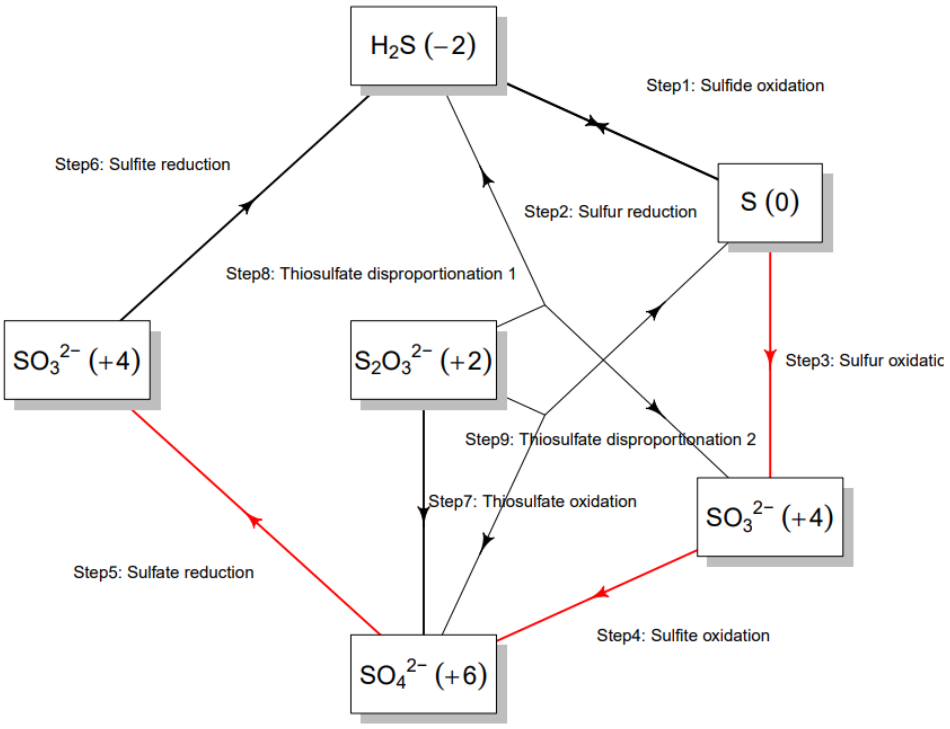

Figure 5: Sulfur cycle of *Alcanivorax*

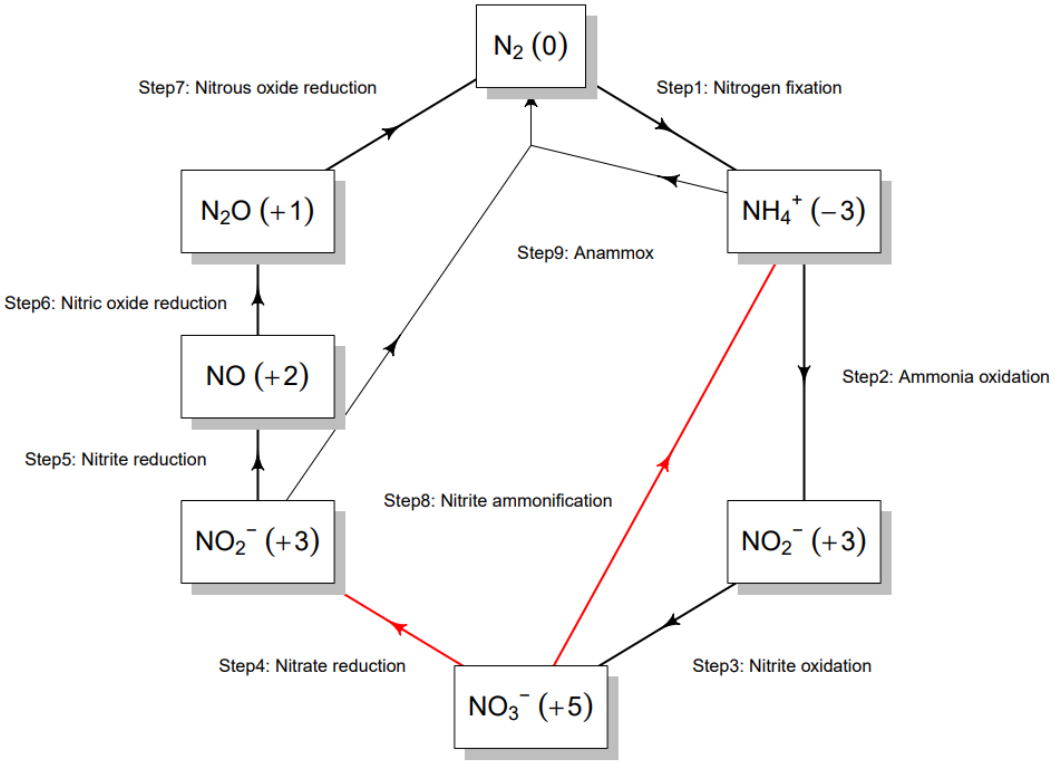

Figure 6: Nitrogen Cycle of Alcanivorax

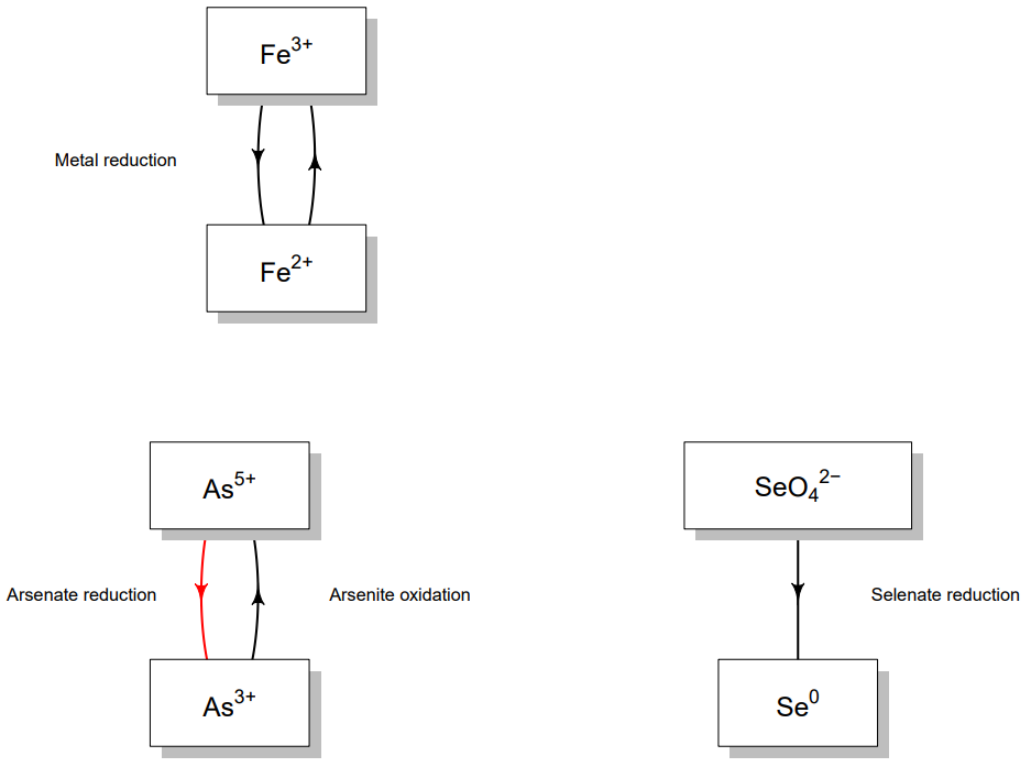

Figure 7: Other Cycle of *Alcanivorax*

This project mainly researched the carbon cycle and sulfur cycle. There are 5 metabolic patterns are connected with the carbon cycle and only 1 metabolic pattern relates with sulfur cycle.

3.2.1 Enzyme in Carbon cycle

3.2.1.1 Fatty acid degradation

Fatty acid is one of the intermediate products in the crude oil degradation. In the process of crude oil degradation, generally, n-alkanes will be oxidized in primary alcohols in first step. In the next step, the primary alcohols will be degraded to aldehyde. After that the aldehyde will be further degraded into fatty acid. Fatty acids toxicity has stronger toxicity when it exists in crude oil. Moreover, the toxicity of short-chain fatty acids is stronger than that of the long-chain(Atlas and Bartha, 1973). Fatty acids are proved that they will inhibit the degradation of crude oil. In the crude oil degradation of *Alcanivorax*, n-alkanes can be oxidized into fatty acids.

① Acyl-CoA dehydrogenase
Acyl-CoA dehydrogenase is a class of enzyme which is particularly used in the catalysis of fatty acid β-oxidation. In the genes of *Alcanivorax*, the genes of *alkB* cluster, *P450 cytochrome monooxygenase* are found that they can convert the alkanes to fatty acids by terminal oxidation(Sabirova *et al.*, 2006). The reaction given by the metabolic result is shows as follow. The *electron-transfer flavoprotein* was found in the gene absent and present file of *Alcanivorax*, which corresponds to 2 genes. One is *etfB*, the genes of electron transfer flavoprotein subunit beta. Another one is *etfA*, the genes of electron transfer flavoprotein subunit alpha. Both *etfA* and *etfB* are in the core pan genome. The reaction uses acyl-CoA and electron-transfer flavoprotein as substrates and a trans-2,3-dehydroacyl-CoA and reduced electron-transfer flavoprotein as products.

3.2.1.2 Aromatics degradation

In the Alcanivorax, there are 7 genes participate in the degradation of Aromatics, including *catechol 1,2-dioxygenase, flavin prenyltransferase, vanillate/4-hydroxybenzoate decarboxylase subunit C, benzoyl-CoA reductase subunit C, benzoyl-CoA reductase subunit B, benzoyl-CoA reductase subunit A* and *benzoyl-CoA reductase subunit D.*

① catechol 1,2-dioxygenase

Catechol 1,2-dioxygenase is a kind of enzyme. Catechol and Protocatechuate are the intermedia products of the aerobic degradation of the aromatic(Rodríguez-Salazar *et al.*, 2020). The *catechol 1,2-dioxygenase*(catA)catalyzes the ortho of catechol broken, which causes the further degradation of catechol into *cis,cis-muconate.*

catechol + oxygen = cis,cis-muconate [RN:R00817]

② flavin prenyltransferase

UbiX flavin prenyltransferase can convert phenol into *Benzoyl-CoA*. *Benzoy-CoA* is the central intermediate compound under the degradation of *Alcanivorax* in low oxygen environment(Boll', Albracht2 and Fuchs', 1997). This metabolic pathway is suitable to actual circumstances. Because in the crude oil polluted areas, some of lighter components will widely cover on the surface of ocean which forms an oil film on between water and air. This film will obstruct the oxygen enter the water. When the spilled crude oil goes through photooxidation, the oxygen is consumed, that lead to an anaerobic environment. The reaction given by the result uses dimethylallyl phosphate and *FMNH2* as substrate and prenylated *FMNH2* and phosphate as products.

③ vanillate/4-hydroxybenzoate decarboxylase subunit C

This substance is particularly used to degrade the *vanillate***.** The degradation substrate only has *4-hydroxybenzoate* which will be converted into phenol.

3.2.1.3 C1 metabolism

According to the result, there is a C1 metabolism system exist in *Alcanivorax*. There are 17 genes participating in the degradation of C1 metabolism, the enzymes are as follow (Table 4).

formate dehydrogenase major subunit
methanol dehydrogenase (cytochrome c) subunit 1
glutathione-independent formaldehyde dehydrogenase
S-formylglutathione hydrolase
S-(hydroxymethyl)glutathione dehydrogenase
S-(hydroxymethyl)mycothiol dehydrogenase
5,6,7,8-tetrahydromethanopterin hydro-lyase
formate dehydrogenase beta subunit
formate dehydrogenase iron-sulfur subunit
formate dehydrogenase (coenzyme F420) alpha subunit
aerobic carbon-monoxide dehydrogenase small subunit
aerobic carbon-monoxide dehydrogenase medium subunit
aerobic carbon-monoxide dehydrogenase large subunit
methanol dehydrogenase

Table 4: the enzyme of C1 metabolism

The oxidation of methane is a typical situation. The methane needs to be oxidized to methyl alcohol in the first step. After that, it can convert into formic acid and then degraded to carbon dioxide and water under the participation of catalyzer.

① methanol dehydrogenase (cytochrome c) subunit 1

The enzyme belongs to the family of oxidoreductases. It catalyzes the primary *alcohols* oxidized including *methanol*. In this project, *methanol dehydrogenase (cytochrome c) subunit* is found that it plays a role in reaction with the substrates of *primary alcohol* and *ferricytochrome cL* and the products of the reaction is *aldehyde*.

② methanol dehydrogenase

This enzyme can catalyze *methanol* convert into *formaldehyde*. The result in this project shows that *methanol dehydrogenase* works under the substrates of *methanol* and *NAD+*. The reaction products are *formaldehyde*, *NADH and H+.*

③ glutathione-independent formaldehyde dehydrogenase

Formaldehyde dehydrogenase is a kind of enzyme using for catalysis. It is a part of family of . *Formaldehyde* and *acetoaldehyde* can be oxidized by the *cata glutathione-independent formaldehyde dehydrogenase*(Ito *et al.*, 1994). According to the outcome of metabolic, the *formaldehyde* can be converted to *formate* in the presence of *water, NAD+* and *glutathione-independent formaldehyde dehydrogenase.*

④ S-formylglutathione hydrolase

S-formylglutathione hydrolase can detoxify the *formaldehyde* by catalyzing the water reacting with the *formaldehyde*. The spontaneous reaction of *formaldehyde* and *glutathione* produces a S-*hydroxymethylglutathione(GS-CH2-OH).* The *GSH* will be converted into *S-formylglutathione* by the oxidation of *formaldehyde dehydrogenase*. Finally, under the catalyst of *S-formylglutathione hydrolase*, *S-formylglutathione* is *hydrolyzed* to *formate* and *GSH*(Gonzalez *et al.*, 2006)

24242424242424. The reaction given by the metabolic result shows that it takes *S-formylglutathione* and water as substrates. The reaction produces the *glutathione* and *formate* as products.

⑤ S-(hydroxymethyl) dehydrogenase

Glutathione and *formaldehyde* spontaneously convert to S-*(hydroxymethyl)glutathione. S-(hydroxymethyl)glutathione* is then hydrolyzed into S-*formylglutathione*. This pathway is one of the ways to detoxify the *formaldehyde*(Pal *et al.*, 2017). The reaction given by the metabolic result points out that it takes *S-(hydroxymethyl)glutathione* and *NAD(P)+* as substrates. The reaction produces *S-*

formylglutathione, NAD(P)H and hydrogen ion as products.

⑥ S-(hydroxymethyl)mycothiol dehydrogenase

Formaldehyde and *mycothiol* is able to spontaneously react into S-*hydroxymethylmycothiol*. *S-(hydroxymethyl)mycothiol* will then be oxidized by *S-(hydroxymethyl)mycothio*l dehydrogenase into *S-formylmycothiol* (Pal *et al.*, 2017). This is one of the pathways to metabolize *formaldehyde*. The reaction given by the metabolic result points out that it takes *S-(hydroxymethyl)mycothiol* and *NAD+* as substrates. The reaction produces *S-formylmycothiol, NADH* and hydrogen ion as products.

⑦ 5,6,7,8-tetrahydromethanopterin hydro-lyase

5,6,7,8-tetrahydromethanopterin hydro-lyase is a catalyst of the reaction of *5,6,7,8-tetrahydromethanopterin* and *formaldehyde*. This reaction can also react spontaneously, but the reactions is much slower without *5,6,7,8-tetrahydromethanopterin hydro-lyase*(Vorholt *et al.*, 2000). The reaction given by the metabolic result points out that it takes *5,6,7,8-tetrahydromethanopterin* and *formaldehyde* as substrates. This reaction produces *5,10-methylenetetrahydromethanopterin* and water as products.

⑧ formate dehydrogenase

In the presence of *formate dehydrogenase major subunit*, *formate* serve as electron donor in aerobic reaction. The reaction given from the outcome shows that the *formate* can be degraded to CO2. The result of metabolic also shows that the beta subunit of *formate* dehydrogenase and iron-sulfur subunit of *formate* dehydrogenase have the same function with the *formate* dehydrogenase major subunit. The reaction given by the metabolic result points out that they take *formate* and NAD+ as substrates. This reaction produces carbon dioxide and NADH as products.

⑨ formate dehydrogenase (coenzyme F420)

Formate dehydrogenase (coenzyme F420) alpha subunit catalyzes the *formate* to be oxidized into carbon dioxide. According to the result of metabolism, *formate dehydrogenase (coenzyme F420)* beta subunit has the same mechanism in the reaction of oxidation of *formate*. The reaction given by the metabolic result shows that they take *formate* and oxidized *coenzyme F420* as substrates. This reaction produces carbon dioxide and reduced *coenzyme F420* as products.

⑩ *formate dehydrogenase (coenzyme F420) beta subunit*

This enzyme catalyzes the oxidation of *formate*. The reaction uses *formate* and oxidized *coenzyme F420* as substrates. The products s carbon dioxide and reduced *coenzyme F420.*

3.2.1.4 Methane metabolism

① Partculate methane/ammonia monooxygenase(pmoA/B/C)
methane/ammonia monooxygenase subunit A, subnit B and *subnit C* have similar function(Myronova *et al.*, 2006). For methane monooxygenase activity, they can oxidize the methane to methanol in the bacteria feed on methane(Stolyar *et al.*, no. However, in *vitro, NADH* can be replaced by the specific *quinols*. The reaction shown as follow.
methane + quinol + O2 = methanol + quinone + H2O [RN:R09518]

② methane monooxygenase regulatory protein B
Soluble methane *monooxygenase regulatory protein B* is the regulator of the coordination complex soluble *methane monooxygenase*. The function of *methane monooxygenase* will be change to hydroxylase when there is a low concentration condition. The reaction in metabolic result shows that is requires *oxygen, methane, NADH and NADPH* as substrates. The reaction produces *methanol, NAD(P)+* and water as products.

③ methane monooxygenase component D
Methane monooxygenase component D is one of the *methane monooxygenases (sMMO)*. According to the known information, *methane monooxygenase* is composed of four parts, part A includes α, β and γ, part B includes *mmoB*, part C includes *mmoC*, part D includes *mmoD*. *mmoD* is a copper-switch for *sMMo* when the copper composition is low, the *mmod* will turn on the switch and the *sMMo* can express. The reaction of *mmoD* is the same as *mmoB*(Semrau *et al.*, 2013). The reaction given by the metabolic result shows that it takes *methane, NAD(P)H,* hydrogen ion and oxygen as substrates. The reaction produces the *methanol, NAD(P)+* and water as products.

3.2.1.5 Halogenated compounds breakdown

Halogenated compounds are also connected with crude oil. Halogenated compounds are also found in drilling platform accidents and spills of petroleum including the mixture of used oil and oil(*Halogens and Waste Oil*, no date).

① 2-haloacid dehalogenase：

In the catalytic reaction of *2-haloacid dehalogenase,* there are two types of *2-haloacid dehalogenase*. The first one is *2-haloacid dehalogenase (configuration-retaining)*. With the participate of this enzyme, the *hydrolytic dehalogenation hydrolytic dehalogenation* of small *(S)-2-haloalkanoic acids* or *(R)-2-haloalkanoic acids* can be

dehalogenated to the relevant *(S)-hydroxyalkanoic* acids or *(R)-hydroxyalkanoic acids*, with reservation of configuration at C-2. The second one is *2-haloacid dehalogenase (configuration-inverting)*. This substance catalyzes the dehalogenation of *(R)-2-haloalkanoic acids* or *(S)-2-haloalkanoic acids* to relevant *(R)-hydroxyalkanoic acids* or *(S)-hydroxyalkanoic acids.* The retention of configuration happens at C-2. These reactions work on the acids of short chain which contains C2-acids to C4-acids. *2-haloacid dehalogenase* is brought into focus for its potential of degrading stubborn halogenated compound. In this project, the functional enzyme is identified as 2-haloacid dehalogenase (configuration-inverting)(Kurihara, Esaki and Soda, 2000). The reaction of the catalysis of 2-haloacid dehalogenase requires *(S)-2-haloacid* and water as substrates, which will produce *(R)-2-hydroxyacid* and halide as products.

⑥ tetrachlorohydroquinone reductive dehalogenase

This enzyme is used to degrade *tetrachlorohydroquinone* in two steps with the reducing agent glutathione. *Tetrachlorohydroquinone* will be degraded to *2,3,6-trichlorohydroquinone* first. In the second step, the *2,3,6-trichlorohydroquinone* will be further converted into *2,6-dichlorohydroquinone*(Mccarthy *et al.*, 1996). The reaction eventually decreases the toxicity of *tetrachlorohydroquinone*. The result given by metabolic shows that there are 2 steps in the degradation of these enzyme. The first step is using *2,6-dichlorohydroquinone, Cl- and glutathione disulfide* as substrates, which will produce *2,3,6-trichlorohydroquinone* and *glutathione* as products. In the next step, *2,3,6-trichlorohydroquinone* will be used as substrates reacting with *Cl-* and glutathione disulfide, the products in this process is *2,3,5,6-tetrachlorohydroquinone and glutathione.*

⑦ 3-chloro-4-hydroxyphenylacetate reductive dehalogenase:

This substance is used in the catalysis of the *C1-OHPA* to *4-hydroxyphenylacetate*(Bisaillon *et al.*, 2010). The reaction analyzed from this project as follow.

1-haloalkane + H2O = a primary alcohol + halide [RN:R02337]

⑧ tetrachloroethene reductive dehalogenase:

The typical reaction of the *tetrachloroethene reductive dehalogenase* is the degradation of *ethylene chloride*. This enzyme makes the microorganism has ability to consuming *ethylene chloride*(Jugder *et al.*, 2015). The real reaction is reverse reaction to the *tetrachloroethene reductive dehalogenase* chemical equation shown as follow. The family of *tetrachloroethene reductive dehalogenase* also contain *trichloroethene dehalogenase* and *vinyl chloride dehalogenase* which can degrade *trichloroethene* and *vinyl chloride*.

trichloroethene + chloride + acceptor = tetrachloroethene + reduced acceptor [RN:R05753]

3.2.1.6 Dissimilatory sulfur metabolism | Sulfide oxidation

Sulfide is a chemical substance, which contains one or more S-2 ions. It also exists in the format of inorganic and organic compounds. Sulfides are common in crude oil. Sulfide is a kind of the dangerous substances in crude oil(Brglez, 2021). For example, hydrogen sulfide is a various of the inorganic compounds, which exists in gaseous state in the nature. It has the properties of high toxicity, causticity and inflammable. The perniciousness of hydrogen sulfide is expressed in injuring eyes, central nervous system and respiratory system(Hemminki and Niemi, 1982). People living in the low concentration of hydrogen sulfide in long term can results in decreasing memory and even the problem of genital system. The acute toxicity of hydrogen sulfide usually happens when people expose themselves under the high-level hydrogen sulfide in short time. People in acute toxicity of hydrogen sulfide will be paralyzed the respiratory system and suffocative. The organic sulfides are flammable(Graphics, no date). The higher molecular weight organic sulfides have, the lower flammability they are. The negative effect of organic sulfides is that they will create sulfur dioxide when they burn. Sulfur dioxide is a gas with irritating odor, which is poisonous for human(*Sulfides, Organic | CAMEO Chemicals | NOAA*, no date). There are two ways sulfur dioxide affecting people. The first one is inhalation which is also the major and the most common way that people contact sulfur dioxide. Sulfur dioxide can convert into sulfurous acid (H2SO3) by reacting with water. Therefore, when people inhale sulfur dioxide, the sulfurous acid will form on mucous membranes to obstruct the breath. The second way is skin and eye contact. The sulfur dioxide can directly irritate the mucous membranes of eyes and skin.

① flavocytochrome c sulphide dehydrogenase, flavin-binding

This enzyme is a catalyst. It participates in the reduction reaction of hydrogen sulfide. The reduction of hydrogen sulfide requires ferricytochrome as reducing agent. The hydrogen will be reduced to sulfur eventually. The reaction shown in the result of metabolic uses *hydrogen sulfide* and *ferricytochrome c* as substrate, which creates *sulfur, ferrocytochrome c* and *hygrogen ion* as prducts.

② sulfide:quinone oxidoreductase

This enzyme catalyzes the reduction reaction of hydrogen sulfide with the substrate of quinone. The reduction reaction can produce a polysulfide which maximum length is 10 sulfur atoms(Nübel *et al.*, 2000). The reaction shown in the result of metabolic uses hydrogen sulfide and quinone as substrate, which creates polysulfide and quinol.

3.3 Result of RStudio

3.3.1 Jaccard

With the help of R Studio, the otherness of *Alcanivorax* is shown as below (Figure 8). There are 15 labels located on the map. Each label corresponds to a genome of one of the *Alcanivorax* species. 7 of them are located together while others spread around.

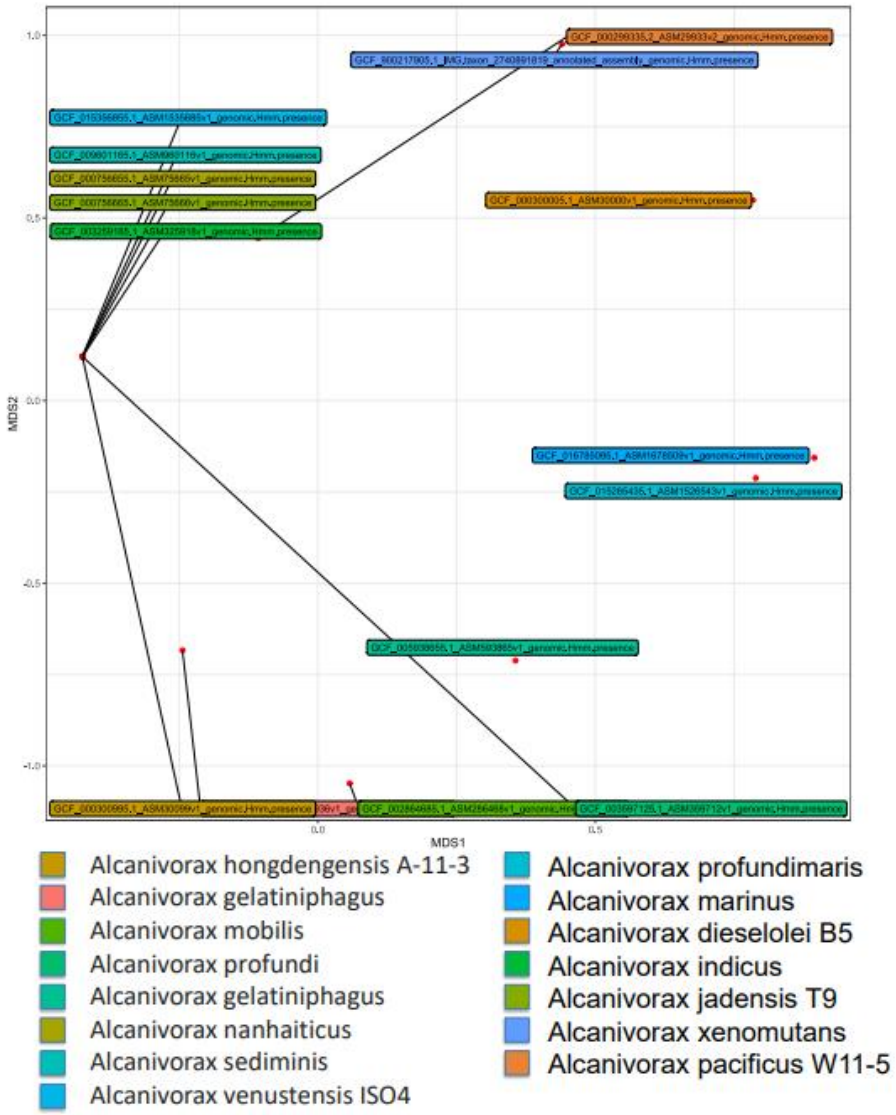

Figure 8: The Jaccard plot

3.3.2 Coincident analysis

In this step, those genes connected with the crude oil degradation were used to analysis the coincident between each other. This part mainly shows that causal relationship of the genes. The complexity threshold is defaulted as 1. The coverage threshold is set as 0.5.

As table 5 shown, there are 7 outcomes of the Minimal Sufficient Condition file. *FDHA.1* is *glutathione-independent formaldehyde dehydrogenase* of Formaldehyde oxidation. *MXAF* is the abbreviation of *methanol dehydrogenase (cytochrome c) subunit 1,* a methanol oxidation enzyme. The outcome of methanol dehydrogenase (cytochrome c) subunit 1 shows that under the condition: *FDHA* is sufficient condition to *MXAF*, the coverage is 0.33 and the complexity is 1. The expression of this condition translated to text: the *FDHA*.*1* leads to the present of *MAXF*, but the confidence of this relationship only has 0.33.

The second instance demonstrate that under the condition: the intersection of *FDOG* and *E3.8.1.2* is sufficient condition to *MXAF*, the coverage is about 0.17 with the complexity is 2. *FDOG* is formate dehydrogenase major subunit, a formate oxidation enzyme. *E3.8.1.2* is the abbreviation of *2-haloacid dehalogenase,* an enzyme of halogenated compounds breakdown. The expression of the condition translates to text: the *FDOG* AND E3.8.1.2 leads to the present of *MAXF*, but the confidence to support this relationship is only 0.17.

The third instance gives the information that under the condition: the intersection of *FDOG* and *E3.8.1.2* is sufficient condition to *FDHA.1,* the coverage is 0.5 and the complexity is 2. The expression of this condition translates to text: the *FDOG* AND E3.8.1.2 leads to the present of *FDHA*. There is 0.5 confidence to support this relationship.

The outcome of the fourth instance shows that under the condition: the *FDHA.1* is sufficient condition to *FDOG*, the coverage is 0.4 with and complexity is 1. The expression of this condition translates to text: the *FDHA.1* leads to the present of *FDOG*, but the confidence to support this relationship is 0.4.

According to the outcome of *FDOG* given by the fifth instance, under the condition: the intersection of *MXAF* and *E3.8.1.2* is sufficient condition to *FDOG*, the coverage is 0.6 with and complexity is 2. The expression of this condition translates to text: the *MXAF* AND *E3.8.1.2* leads to the present of *FDOG* and the confidence to support this relationship is only 0.6.

From the result of *E3.8.1.2* given by the sixth instance, under the condition: the intersection of *FDHA.1* and *FDOG* is sufficient condition to *E3.8.1.2,* the coverage is 0.6

with and complexity is 2. The expression of this condition translates to text: the *FDOG* AND *FDHA.1* leads to the present of *E3.8.1.2* and the confidence to support this relationship is 0.6.

From the outcome of the seventh line, under the condition: the intersection of *MXAF* and *FDOG* is sufficient condition to *E3.8.1.2,* the coverage is 0.2 with and complexity is 2. The expression of this condition translates to text: the *FDOG* AND *MXAF* leads to the present of *E3.8.1.2,* but the confidence to support this relationship is only 0.2.

	Outcome		condition	
1	MXAF		FDHA.1. -> MXAF	
2	MXAF		FDOG * e3.8.1.2 -> MXAF	
3	FDHA.1.		FDOG * e3.8.1.2 -> FDHA.1.	
4	FDOG		FDHA.1. -> FDOG	
5	FDOG		MXAF * E3.8.1.2 -> FDOG	
6	E3.8.1.2		fdha.1. * FDOG -> E3.8.1.2	
7	E3.8.1.2		Mxaf * FDOG -> E3.8.1.2	
	consistency	coverage	complexity	minimal
1	1	0.333333	1	TRUE
2	1	0.166667	2	TRUE
3	1	0.5	2	TRUE
4	1	0.4	1	TRUE
5	1	0.6	2	TRUE
6	1	0.6	2	TRUE
7	1	0.2	2	TRUE

Table 5: the Minimal Sufficient Condition of genes in Alcanivorax

The result of Atomic Solution Formulas shown on table 6. The first instance shows the existence of *INUS* condition. According to the *INUS* theory, the gene *MXAF* and *E3.8.1.2* respectively to be the *INUS* condition with *FDOG*. The coverage of these conditions is 2. The expression of this condition translates to text: the MXAF AND *E3.8.1.2* leads to the present of *MAXF* and the present of *FDOG* must be occurred by *MXAF* OR *E3.8.1.2.* The confidence to support this relationship is 0.6.

The second instance shows the existence of *INUS* condition. According to the *INUS* theory, the *FDHA.1* and *FDOG* respectively to be the *INUS* condition with *E3.8.1.2*. The coverage of these conditions is 2. The expression of this condition translates to text: the *FDHA.1* AND *FDOG* leads to the present of *E3.8.1.2* and the present of *FDOG* must be occurred by *MXAF* OR *E3.8.1.2*. The confidence to support this relationship is 0.6.

	Outcome		Condition	
1	FDOG		MXAF * E3.8.1.2 <-> FDOG	
2	E3.8.1.2		fdha.1. * FDOG <-> E3.8.1.2	
	consistency	coverage	complexity	Inus
1	1	0.6	2	TRUE
2	1	0.6	2	TRUE

Table 6: the Atomic Solution Formulas of genes in Alcanivorax

The result of complex solution formulas shows on table 7. There is only one *INUS* condition found in complex solution formulas, with the coverage is 0.6 and complexity is 4.

Outcome		Condition	
E3.8.1.2, FDOG		(fdha.1. * FDOG <-> E3.8.1.2) * (MXAF * E3.8.1.2 <-> FDOG)	
consistency	coverage	complexity	inus
1	0.6	4	TRUE

Table 7: the Complex Solution Formulas of genes in Alcanivorax

4. Discussion

According to the results, the *Alcanivorax* has the ability to degrade the crude oil. There are 3 major pathways connected with the crude oil degradation, including hydrocarbon, halogenated compound and sulfide degradation. The discussion as follow will based on the absence of genes, the function of enzyme and the casual relationship among the genes of *Alcanivorax*.

AlkB exists in many species. However, the previous report shows that the expression of *alkB* can be different in various species. The same species in various environment can also lead to a different expression(Rojo, 2009). According to the gene absent and present file given by ROARY, the gene *alkB* series is indicated that exists in *Alcanivorax* strain. However, the *alkB* genes were found in the shell pan genome and cloud pan genome, it did not present in the core pan genome. According to the previous searches given by other researches, *alkB* is an important gene which participate in the most situations of crude oil degradation. Therefore, the author in this project giving an assumption. The reason why the alkB did not found in core pan genome is that the lack of the genomes of *Alcanivorax* in this project. *Alcanivorax* totally has 97 genome sequences in NCBI database in current, but only 15 genome sequences are reference. Each species has only one genome sequence. In addition, if the gene sequencing of some species of *Alcanivorax* is not completed, it can also lead to the lack of genes of *Alcanivorax*. Hence the accuracy of the gene absent and present file is limited by above situations. This assumption come from the analysis of *Putative cytochrome P450 136* which is a putative gene in this project. Putative gene means unconfirmed gene. The segment of the sequence is recognized that it is similar to the cytochrome P450 136.

However, at the present, there is no study to prove the presence of *cytochrome P450 136* but *cytochrome P450 153*. Hence it is possible that there is a lack of *Alcanivorax* genomes.

The output images of ROARY demonstrate the distribution of pan genome of *Alcanivorax* in a visualization way. As shown in table 3, a large number of genes of *Alcanivorax* are in the group cloud pan genome which only present in one genome. Figure 3 shows the proportion of the could genes is large. That indicates that there are some alternative genes existing in each species. The alternatively biological nature of each species is reflected by the expression of the alternative genes. The pan genome matrix describes the absent and present of each gene in each genome. It can also explain the otherness between each species in *Alcanivorax*.

According from the outcome of METABOLIC, the *Alcanivorax* has 4 substance cycle including nitrogen cycle, carbon cycle, sulfur cycle and other substance cycle. The carbon cycle and sulfur cycle are mainly focused on this report. The analysis of metabolic results shows that there are 22 metabolic patterns existing in *Alcanivorax*. 6 of the patterns are concerned in this project, including halogenated compound utilization, sulfide oxidation, C1 metabolism, aromatics degradation and fatty acid degradation. The data of the result manifests the strong and perfect ability of degrading crude oil. There are 27 enzymes participating in the degradation of hydrocarbon, including one fatty acid degradation enzyme(*acyl-CoA dehydrogenase*), 7 Aromatics degradation enzymes(*catechol 1,2-dioxygenase, flavin prenyltransferase, vanillate/4-hydroxybenzoate decarboxylase subunit C, benzoyl-CoA reductase subunit C, benzoyl-CoA reductase subunit B, benzoyl-CoA reductase subunit A, benzoyl-CoA reductase subunit D), 12 C1 degradation enzymes(methanol dehydrogenase (cytochrome c) subunit 1, methanol dehydrogenase, glutathione-independent formaldehyde dehydrogenase, S-formylglutathione hydrolase, S-(hydroxymethyl)glutathione dehydrogenase / alcohol dehydrogenase, S-(hydroxymethyl)mycothiol dehydrogenase, 5,6,7,8-tetrahydromethanopterin hydro-lyase, formate dehydrogenase major subunit, formate dehydrogenase beta subunit, formate dehydrogenase iron-sulfur subunit, formate dehydrogenase (coenzyme F420) alpha subunit, formate dehydrogenase (coenzyme F420) beta subunit) and 5 enzymes (methane/ammonia monooxygenase subunit A, methane/ammonia monooxygenase subunit B, methane/ammonia monooxygenase subunit C, methane monooxygenase regulatory protein B, methane monooxygenase component D).*

In the C1 metabolic system, it exists a group of genes which helps *Alcanivorax* to build a robust system to degrade n-alkanes. In the degradation of n-alkanes, the first production usually is primary alcohol. The primary alcohol the alcohol can be oxidized in aldehyde. In the result, there is no evidence show that the aldehyde can be further degraded under the catalysis of particular enzyme. However, the methane can be fully degraded by *Alcanivorax*. In the enzymes working on C1 metabolism, the enzyme methanol dehydrogenase can catalyze the oxidation of methanol to formaldehyde.

The formaldehyde will be further degraded to formate by the glutathione-independent formaldehyde dehydrogenase in the next. In the last step, there are two pathways to metabolize formate in *Alcanivorax*. One way is formate degraded by *formate dehydrogenase major/beta/iron-sulfur subunit* with the substrate of *NAD+,* the product from the reaction is carbon dioxide and NADH. Another way is catalyzed by *formate dehydrogenase (coenzyme F420)* alpha/beta subunit, which also create carbon dioxide. In the above processes of C1 metabolism, *Alcanivorax* does not create harmful gas and intermediate product, which makes it feasible to use the correlative enzymes to degrade crude oil. It is notice that the gene, *fdhA*, participates in the degradation of formaldehyde and *formate*. This mechanism makes the C1 hydrocarbon degraded in a high-efficient pathway.

Alcanivorax has two main pathways to degrade the methane. The first one is the oxidation of particulate methane. This reaction based on three genes *pmoA*, *pmoB* and *pmoC*. These genes correspond to *methane/ammonia monooxygenase subunit A/B/C.* The enzymes catalyze oxidation of methane to be methanol. Another pathway is soluble methane degradation. The enzymes *methane monooxygenase regulatory protein B* and *methane monooxygenase component D* participate in this degradation. The reaction converts methane to be methanol as well. It can be certain that the second pathway do not create any toxic substance. The crude oil pollution areas bring up a large number of bacteria. The ability to degrade different forms of methane increase the range of nutriment of *Alcanivorax*, which makes *Alcanivorax* easier to survives in the crude oil pollution areas.

Not only hydrocarbon can be degraded by *Alcanivorax*, but also Halogenated compound utilization can be degraded. The react condition of the metabolism of *(S)-2-haloacid* is simple. *2-haloacid dehalogenase* catalyzed the reaction with the water as substrate. According to the outcome, *Alcanivorax* can also degrade hydrogen sulfide with the *ferricytochrome* as substrate. Sulfide will be degraded as sulfur in this process. In addition, hydrogen sulfide can also be degraded as polysulfide with quinone as substrate.

This project mainly studies the genes which participate in the crude oil degradation. However, during the research, there is a gene *BLC* which was assumed by previous study can absorb the n-alkanes in the crude oil degradation (Sabirova *et al.*, 2011).

The complex carbon degradation was not study in this project, because the complex carbon metabolism of *Alcanivorax* is not connected with any component of crude oil degradation. It is worth noting that in the complex carbon degradation system of *Alcanivorax*, there is some gene can resolve chitin substance which is an important substance forming the bacterial cell wall. Therefore, the author conjectures that this is the reason why *Alcanivorax* can be the dominant species in polluted areas.

The Jaccard plot export from R studio shows otherness and homoplasy of *Alcanivorax*.

There is a convergence point containing the species *Alcanivorax indicus, Alcanivorax jadensis T9, Alcanivorax nanhaiticus, Alcanivorax sediminis, Alcanivorax venustensis ISO4, Alcanivorax hongdengensis A-11-3 and Alcanivorax profundi.* The others are distributed around this point. Based on the genomes used in this project, the distribution picture shows that there is similarity among the species of *Alcanivorax*. 7 of the species are in high degree of similarity. 8 of them are dissimilar with each species studied in this project. It is unable to come to a result of the degree of dissimilarity from this picture. Because of the lack of genome sequences, this project can not converge area of each genome.

According to the result from coincident analysis, there is evidence that shows the strong causal relationship among the genes of the strain of *Alcanivorax*. The highest coverage is 0.6 from the expression: *MXAF * E3.8.1.2 <-> FDOG and fdha.1. * FDOG <-> E3.8.1.2*. Higher coverage represents higher confidence to support the relationship. According to the expression from complex solution formulas, it is highly complex between these two relationships. The present of gene *FDOG* is determined by gene *MAXF* and *E3.8.1.2*. Only the present of both gene *MAXF* and *E3.8.1.2* can lead to the present of gene FDOG. On contrary, the present of gene *E3.8.1.2* is impacted by the present of gene *FDHA.1* and *FDOG*. Gene *FDOG* and gene *E3.8.1.2* present in the condition of each other, which indicates that they are important to each other. One of these four genes is able to impact the present of others. However, the confidence to support this relationship is 0.6 which is not higher enough to support this relationship.

In summary, there is dissimilarity among the *Alcanivorax*, which are expressed in the large proportion of cloud pan genome, the otherness of each species in pan genome matrix. There also is common among the *Alcanivorax*, which is performed in the Jaccard plot. 7 of 15 species analyzed in this project shows high convergence. The results also show that *Alcanivorax* is able to degrade the components of crude oil in aerobic environment and anaerobic environment. The degradable crude oil components analyzed in this project are hydrocarbon, halogenated compound and sulfide. Although *Alcanivorax* can oxidize many types of hydrocarbon, the efficiency of each enzyme cannot be inferred from the outcome in this project. The compounds like methane can be completely converted to water and carbon dioxide. Hydrogen sulfide can be completely degraded in reduction reaction. *Benzoyl-CoA,* fatty acid and halogenated compound can be degraded by combining with other substance. Besides, the coincident analysis indicates that there is casual relationship among *MXAF*, *FDHA*.1, *E3.8.1.2* and *FDOG*. These four genes can affect the present of each other, but the confidence is not high enough to support this relationship.

5. Conclusion

In conclusion, *Alcanivorax* is feasible to be used in the crude oil biodegradation. It has wide dietary, including hydrocarbon, halogenated compound and sulfide, which makes it in advantage in species competition. Wide dietary also makes it easy to adapt the various crude oil environment. *Alcanvorax* has multiple way to degrade the crude oil component methane and there is no harm gas and substance produced. In addition, *Alcanivorax* is able to produce the substrate, such as electron-transfer flavoprotein, by themselves, which decreases its dependence on other species.

This project was study around the theme crude oil degradation. As a matter of fact, the degradation condition of Alcanivorax is more than crude oil spill, which also including the oil spilled from factors, the hydrocarbon spilled from factor and even plastic pollutants. Due to the plastic has similar structure of hydrocarbon and it is made of crude oil, Alcanivorax can also be used in the degradation of plastic. Previous reports have indicated this ability(Rojo, 2009).

In this research, I found that there is dissimilarity among the *Alcanivorax* species. Various species in *Alcanivorax* is possible to lead to the different expression in crude oil degradation. Therefore, the ability of crude oil degradation among each species of *Alcanivorax* is required further study. Only 7 species show high homoplasy in this research. Due to only few reference genomes in the database of *Alcanvorax*, to verify the homoplasy in *Alcanivorax*, further research can also import the genomes from other species.

6. Reference

• *Global seaborne trade - oil tanker capacity 2020 | Statista* (no date). Available at: https://www.statista.com/statistics/267605/capacity-of-oil-tankers-in-the-world-maritime-trade-since-1980/ (Accessed: August 19, 2021).

Atlas, R. M. and Bartha, A. R. (1973) *Inhibition by fatty acids of the biodegradation of petroleum*, *Antonie van Leeuwenhoek*.

van Beilen, J. B. *et al.* (2004) "Characterization of two alkane hydroxylase genes from the marine hydrocarbonoclastic bacterium Alcanivorax borkumensis," *Environmental Microbiology*, 6(3), pp. 264–273. doi: 10.1111/j.1462-2920.2004.00567.x.

Bisaillon, A. *et al.* (2010) "Identification and characterization of a novel CprA reductive dehalogenase specific to highly chlorinated phenols from desulfitobacterium hafniense strain PCP-1," *Applied and Environmental Microbiology*, 76(22), pp. 7536–7540. doi: 10.1128/AEM.01362-10.

Boll', M., Albracht2, S. S. P. and Fuchs', G. (1997) *Benzoyl-CoA reductase (dearomatizing), a key enzyme of anaerobic aromatic metabolism A study of adenosinetriphosphatase activity, ATP stoichiometry of the reaction and EPR properties of the enzyme*, *Eur. J. Biochem*.

Borges-Ramírez, M. M. *et al.* (2021) "Organochlorine pesticides, polycyclic aromatic hydrocarbons, metals and metalloids in microplastics found in regurgitated pellets of black vulture from Campeche, Mexico," *Science of The Total Environment*, 801, p. 149674. doi: 10.1016/j.scitotenv.2021.149674.

Brglez, Š. (2021) "Risk assessment of toxic hydrogen sulfide concentrations on swine farms," *Journal of Cleaner Production*, 312. doi: 10.1016/j.jclepro.2021.127746.

CDC | Facts About Benzene (no date). Available at: https://emergency.cdc.gov/agent/benzene/basics/facts.asp (Accessed: August 19, 2021).

Crude Oil Production (no date). Available at: https://www.eia.gov/dnav/pet/pet_crd_crpdn_adc_mbbl_a.htm (Accessed: August 13, 2021).

Dispersants (no date). Available at: https://www.biologicaldiversity.org/programs/public_lands/energy/dirty_energy_development/oil_and_gas/gulf_oil_spill/dispersants.html (Accessed: August 19, 2021).

Feathers As Insulation (2009). Available at: http://www.cde.ca.gov/re/pn/fd/documents/elacontentstnds.pdf.

Gaudet, P. *et al.* (2011) "Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium," *Briefings in Bioinformatics*, 12(5), pp. 449–462. doi: 10.1093/bib/bbr042.

Gonzalez, C. F. *et al.* (2006) "Molecular basis of formaldehyde detoxification: Characterization of two S-formylglutathione hydrolases from Escherichia coli, FrmB and YeiG," *Journal of Biological Chemistry*, 281(20), pp. 14514–14522. doi: 10.1074/jbc.M600996200.

Graphics, M. (no date) *Managing Hazardous Materials Incidents*.

Halogens and Waste Oil (no date). Available at: www.dep.state.pa.us.

Hara, A. *et al.* (2004) "Cloning and functional analysis of alkB genes in Alcanivorax borkumensis SK2," *Environmental Microbiology*, 6(3), pp. 191–197. doi: 10.1111/j.1462-2920.2004.00550.x.

Hazen, T. C., Prince, R. C. and Mahmoudi, N. (2016) "Marine Oil Biodegradation," *Environmental Science and Technology*, 50(5), pp. 2121–2129. doi: 10.1021/acs.est.5b03333.
Hemminki, K. and Niemi, M.-L. (1982) *Community Study of Spontaneous Abortions: Relation to Occupation and Air Pollution by Sulfur Dioxide, Hydrogen Sulfide, and Carbon Disulfide*, *International Archives of Int Arch Occup Environ Health*.
Ito, K. *et al.* (1994) *Cloning and High-Level Expression of the Glutathione-Independent Formaldehyde Dehydrogenase Gene from Pseudomonas putida*, *JOURNAL OF BACTERIOLOGY*.
Jagtap, C. B. *et al.* (2021) "Genome sequence of an obligate hydrocarbonoclastic bacterium Alcanivorax marinus NMRL4 isolated from oil polluted seawater of the Arabian Sea," *Marine Genomics*, p. 100875. doi: 10.1016/j.margen.2021.100875.
Jugder, B. E. *et al.* (2015) "Reductive Dehalogenases Come of Age in Biological Destruction of Organohalides," *Trends in Biotechnology*. Elsevier Ltd, pp. 595–610. doi: 10.1016/j.tibtech.2015.07.004.
Kurihara, T., Esaki, N. and Soda, K. (2000) *Bacterial 2-haloacid dehalogenases: structures and reaction mechanisms*, *Journal of Molecular Catalysis B: Enzymatic*. Available at: www.elsevier.comrlocatermolcatb.
Kvenvolden, K. A. (2006) "Organic geochemistry - A retrospective of its first 70 years," in *Organic Geochemistry*, pp. 1–11. doi: 10.1016/j.orggeochem.2005.09.001.
Kvenvolden, K. A. and Cooper, C. K. (2003) "Natural seepage of crude oil into the marine environment," *Geo-Marine Letters*, 23(3–4), pp. 140–146. doi: 10.1007/s00367-003-0135-0.
Mccarthy, D. L. *et al.* (1996) *Exploration of the Relationship between Tetrachlorohydroquinone Dehalogenase and the Glutathione S-Transferase Superfamily †*. Available at: https://pubs.acs.org/sharingguidelines.
Mohammadiun, S. *et al.* (2021) "Intelligent computational techniques in marine oil spill management: A critical review," *Journal of Hazardous Materials*. Elsevier B.V. doi: 10.1016/j.jhazmat.2021.126425.
Nübel, T. *et al.* (2000) "Sulfide:quinone oxidoreductase in membranes of the hyperthermophilic bacterium Aquifex aeolicus (VF5)," *Archives of Microbiology*, 173(4), pp. 233–244. doi: 10.1007/s002030000135.
Olivera, N. L. *et al.* (2009) "Isolation and characterization of biosurfactant-producing Alcanivorax strains: hydrocarbon accession strategies and alkane hydroxylase gene analysis," *Research in Microbiology*, 160(1), pp. 19–26. doi: 10.1016/j.resmic.2008.09.011.
Overview of Petroleum Product Poisoning - Toxicology - Veterinary Manual (no date). Available at: https://www.msdvetmanual.com/toxicology/petroleum-product-poisoning/overview-of-petroleum-product-poisoning (Accessed: August 19, 2021).
Petroleum Products - Environmental Exposure from Refineries - Proposition 65 Warnings Website (no date). Available at: https://www.p65warnings.ca.gov/fact-sheets/petroleum-products-environmental-exposure-refineries (Accessed: August 19, 2021).
Plan for Deepwater Horizon Oil Spill Natural Resource Injury Restoration: An Overview (no date). Available at: www.gulfspillrestoration.noaa.gov.
Rodríguez-Salazar, J. *et al.* (2020) "Characterization of a Novel Functional Trimeric Catechol 1,2-Dioxygenase From a Pseudomonas stutzeri Isolated From the Gulf of Mexico," *Frontiers in Microbiology*, 11. doi: 10.3389/fmicb.2020.01100.

Rojo, F. (2009) "Degradation of alkanes by bacteria: Minireview," *Environmental Microbiology*. Blackwell Publishing Ltd, pp. 2477–2490. doi: 10.1111/j.1462-2920.2009.01948.x.
Sabirova, J. S. *et al.* (2006) "Proteomic insights into metabolic adaptations in Alcanivorax borkumensis induced by alkane utilization," *Journal of Bacteriology*, 188(11), pp. 3763–3773. doi: 10.1128/JB.00072-06.
Sabirova, J. S. *et al.* (2011) "Transcriptional profiling of the marine oil-degrading bacterium Alcanivorax borkumensis during growth on n-alkanes," *FEMS Microbiology Letters*, 319(2), pp. 160–168. doi: 10.1111/j.1574-6968.2011.02279.x.
Safieddin Ardebili, S. M. *et al.* (2020) "A review on higher alcohol of fusel oil as a renewable fuel for internal combustion engines: Applications, challenges, and global potential," *Fuel*. Elsevier Ltd. doi: 10.1016/j.fuel.2020.118516.
Seemann, T. (2014) "Prokka: Rapid prokaryotic genome annotation," *Bioinformatics*, 30(14), pp. 2068–2069. doi: 10.1093/bioinformatics/btu153.
Semrau, J. D. *et al.* (2013) "Methanobactin and MmoD work in concert to act as the 'copper-switch' in methanotrophs," *Environmental Microbiology*, 15(11), pp. 3077–3086. doi: 10.1111/1462-2920.12150.
Sulfides, Organic | CAMEO Chemicals | NOAA (no date). Available at: https://cameochemicals.noaa.gov/react/20 (Accessed: August 20, 2021).
The Toxicity of Oil: What's the Big Deal? | response.restoration.noaa.gov (no date). Available at: https://response.restoration.noaa.gov/about/media/toxicity-oil-whats-big-deal.html (Accessed: August 19, 2021).
Tormoehlen, L. M., Tekulve, K. J. and Nañagas, K. A. (2014) "Hydrocarbon toxicity: A review," *Clinical Toxicology*. Informa Healthcare, pp. 479–489. doi: 10.3109/15563650.2014.923904.
Tropical Cyclone Report | Enhanced Reader (no date).
Vandenbroucke, M. and Largeau, C. (2007) "Kerogen origin, evolution and structure," *Organic Geochemistry*, pp. 719–833. doi: 10.1016/j.orggeochem.2007.01.001.
Warr, L. N. *et al.* (2018) "Nontronite-enhanced biodegradation of Deepwater Horizon crude oil by Alcanivorax borkumensis," *Applied Clay Science*, 158, pp. 11–20. doi: 10.1016/j.clay.2018.03.011.
Waryszak, P. *et al.* (2021) "Planted mangroves cap toxic petroleum-contaminated sediments," *Marine Pollution Bulletin*, 171, p. 112746. doi: 10.1016/j.marpolbul.2021.112746.
White, G. and Reid, G. (2018) *RECYCLED WASTE PLASTIC FOR EXTENDING AND MODIFYING ASPHALT BINDERS*.
Zengel, S. *et al.* (2021) "Planting after shoreline cleanup treatment improves salt marsh vegetation recovery following the Deepwater Horizon oil spill," *Ecological Engineering*, 169. doi: 10.1016/j.ecoleng.2021.106288.

7. Appendices

The codes processed in R Studio

```
#Jaccard
```

```
data<-read.csv("data01.csv",header = TRUE, row.name = 1)
library(vegan)
library(ggplot2)
library(ggrepel)
data.dist<-vegdist(data,method="jaccard")
ord<-capscale(data ~ 1,distance = "jaccard")

df<-as.data.frame(scores(ord, display = "sites"))
df$Colours=1
i=0
for(i in 1:21){
   df[i,3]<-2*i
}

#pdf("alcanivorax_pdf.pdf",height=10,width=10)
p <- ggplot(df, aes(MDS1, MDS2))
p<- p+geom_point(color = 'red')
p<-p + geom_label_repel(aes(label = rownames(df),fill=factor(Colours)),size
=3.5,max.overlaps = Inf)+theme_bw()
p<-p+guides(fill=FALSE)
print(p)
#dev.off()
```

```
#cna
library(cna)
data<-read.csv("data01.csv",header = TRUE, row.name = 1)
threshold_perc <- 0.6
data.cna<-cna(data,cov =threshold_perc)
print(asf(data.cna))

df.msc<-as.data.frame(msc(data.cna))
write.csv(df.msc,"MSC.csv")

df.asf<-as.data.frame(asf(data.cna))
#write.csv(df.asf,"ASF.csv")

df.csf<-as.data.frame(csf(data.cna))
#write.csv(df.csf,"CSF.csv")
```

Declaration of Originality Form

This form **must** be completed and signed and submitted with all assignments.
Please complete the information below (using BLOCK CAPITALS).

Name: FENG NI .....
Student Number: 2511633N .....
Course Name: MSc Project .....
Assignment Number/Name ENG5059P .....

An extract from the University's Statement on Plagiarism is provided overleaf. Please read carefully THEN read and sign the declaration below.

I confirm that this assignment is my own work and that I have:	
Read and understood the guidance on plagiarism in the Student Handbook, including the University of Glasgow Statement on Plagiarism	✓
Clearly referenced, in both the text and the bibliography or references, **all sources** used in the work	✓
Fully referenced (including page numbers) and used inverted commas for **all text quoted** from books, journals, web etc. (Please check with the Department which referencing style is to be used)	✓
Provided the sources for all tables, figures, data etc. that are not my own work	✓
Not made use of the work of any other student(s) past or present without acknowledgement. This includes any of my own work, that has been previously, or concurrently, submitted for assessment, either at this or any other educational institution, including school (see overleaf at 31.2)	✓
Not sought or used the services of any professional agencies to produce this work	✓
In addition, I understand that any false claim in respect of this work will result in disciplinary action in accordance with University regulations	✓

DECLARATION:

I am aware of and understand the University's policy on plagiarism and I certify that this assignment is my own work**,** except where indicated by referencing, and that I have followed the good academic practices noted above

Signed FENG NI .....

The University of Glasgow Plagiarism Statement

The following is an extract from the University of Glasgow Plagiarism Statement. The full statement can be found in the *University Regulations* at https://www.gla.ac.uk/myglasgow/senateoffice/policies/uniregs/regulations2021-22/feesandgeneral/studentsupportandconductmatters/reg32/

This should be read in conjunction with the discipline specific guidance provided by the School at Insert link.

31.1 The University's degrees and other academic awards are given in recognition of a student's **personal achievement**. All work submitted by students for assessment is accepted on the understanding that it is the student's own effort.

31.2 Plagiarism is defined as the submission or presentation of work, in any form, which is not one's own, without **acknowledgement of the sources**. Plagiarism includes inappropriate collaboration with others. Special cases of plagiarism can arise from a student using his or her own previous work (termed auto-plagiarism or self-plagiarism). Auto-plagiarism includes using work that has already been submitted for assessment at this University or for any other academic award.

31.3 The incorporation of material without formal and proper acknowledgement (even with no deliberate intent to cheat) can constitute plagiarism.

Work may be considered to be plagiarised if it consists of:

a direct quotation;

a close paraphrase;

an unacknowledged summary of a source;

direct copying or transcription.

With regard to essays, reports and dissertations, the rule is: if information **or ideas** are obtained from any source, that source must be acknowledged according to the appropriate convention in that discipline; and **any direct quotation must be placed in quotation marks** and the source cited immediately. Any failure to acknowledge adequately or to cite properly other sources in submitted work is plagiarism. Under examination conditions, material learnt by rote or close paraphrase will be expected to follow the usual rules of reference citation otherwise it will be considered as plagiarism. Departments should provide guidance on other appropriate use of references in examination conditions.

31.4 Plagiarism is considered to be an act of fraudulence and an offence against University discipline. Alleged plagiarism, at whatever stage of a student's studies, whether before or after graduation, will be investigated and dealt with appropriately by the University.

31.5 The University reserves the right to use plagiarism detection systems, which may be externally based, in the interests of improving academic standards when assessing student work.

If you are still unsure or unclear about what plagiarism is or need advice on how to avoid it,

SEEK HELP NOW!

You can contact any one of the following for assistance:

Lecturer
Course Leader
Dissertation Supervisor
Adviser of Studies
Student Learning Service