

**WHOLE GENOME SEQUENCING, CHARACTERIZATION AND
ANALYSIS OF CORONENE DEGRADING BACTERIAL STRAIN,
HALOMONAS CASEINILYTICA: A BIOINFORMATICS APPROACH**

BY

Thasneema Rafic

A Thesis Presented to the
COLLEGE OF GRADUATE AND INTERDISCIPLINARY STUDIES
KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

In

LIFE SCIENCE

December 2024

KING FAHD UNIVERSITY OF PETROLEUM and MINERALS

DHAHRAN- 31261, SAUDI ARABIA

DEANSHIP OF GRADUATE STUDIES

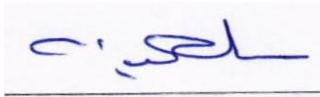
This thesis, written by **Thasneema Rafic** under the direction of her thesis advisor and approved by her thesis committee, has been presented and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE IN LIFE SCIENCE**.

A handwritten signature in black ink, with the date "Jan 23rd, 2025" written to the right of the signature.

Dr. Yaqub Mahnashi
Department Chairman

A handwritten signature in black ink, consisting of a stylized name followed by a horizontal line.

Dr. Alexis Nzila
(Advisor)

A handwritten signature in blue ink, appearing as a stylized name.

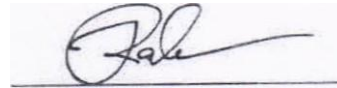
Dr. Suliman Saleh Al-Homidan
Dean of Graduate Studies

A handwritten signature in black ink, featuring a large circular loop at the end.

Dr. Assad Al - Thukair
(Member)

25 December 2024

Date

A handwritten signature in black ink, appearing as a stylized name.

Dr. Tanzilur Rahman
(Member)



© Thasneema Rafic

2024

To

my parents, for their faith that never faltered,

my advisor, for his invaluable guidance and wisdom,

and my friends, for bringing light and laughter to this journey.

ACKNOWLEDGMENTS

In the name of Allah, the Most Gracious, the Most Merciful, I begin by thanking Allah (SWT) for granting me the strength, guidance, and opportunity to undertake and complete this journey. Without his countless blessings and mercy, none of this would have been possible.

I am profoundly thankful to my advisor, Dr. Alexis Nzila, for his unwavering support, guidance, and for having faith in me throughout this journey. His mentorship, patience, and encouragement were pivotal in the completion of this work. I would also like to express my gratitude to Dr. Omer Al Khnbashi, who has been the driving force behind this research, consistently pushing me to explore new dimensions and inspiring me with his passion for the subject. I extend my appreciation to Mr. Mohammed Alarawi from KAUST, whose assistance in the initial sequencing was invaluable.

I am sincerely thankful to my committee members for their invaluable contributions. Special appreciation goes to Dr. Assad Al-Thukair, whose keen observations and insightful guidance have meaningfully elevated the quality of this study. I am grateful to Dr. Tanzilur Rahman, who joined the committee at the last moment and provided the opportunity for me to defend my thesis on time.

I extend my appreciation to the department of Bioengineering and our chairman, Dr. Yaqub Alhussain, for providing the support, resources, and academic environment necessary for this research. My heartfelt thanks also go to KFUPM For offering me the privilege to pursue my master's degree and complete this important milestone in my academic journey.

A special note of thanks to my KFUPM family, who have been my backbone throughout this journey. The friendships, mentorships, and support I have received here have been invaluable, making this experience both enriching and memorable.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	v
TABLE OF CONTENTS.....	vi
LIST OF TABLES.....	viii
LIST OF FIGURES	ix
LIST OF ABBREVIATIONS	x
ABSTRACT.....	xii
ملخص الرسالة.....	xiii
CHAPTER 1: INTRODUCTION.....	1
1.1. Background.....	1
1.2. Objectives	4
CHAPTER 2: LITERATURE REVIEW	5
2.1. Omics studies	9
2.2. Research related to the bacterial strain used in the proposed study	13
CHAPTER 3: MATERIALS AND METHODS.....	15
3.1. Whole genome sequencing and Quality assessment	15
3.2. Taxonomical classification and Strain Identification.....	16
3.3. Gene Prediction and Functional Annotation	16
3.4. Identification of Genomic features.....	18
3.5. Pathway analysis	19
CHAPTER 4: RESULTS	20
4.1. Whole genome sequencing and Quality Assessment	20

4.2. Taxonomical classification.....	22
4.3. Gene Prediction and Functional Annotation	23
4.4. Identification of Genomic features.....	28
4.5. Functional and Pathway analysis	31
CHAPTER 5: DISCUSSION	38
CHAPTER 6: CONCLUSION.....	43
CHAPTER 7: FUTURE PROSPECT	44
REFERENCES.....	46
Appendices	67
Vitae.....	70

LIST OF TABLES

Table 1: tools used for gene prediction and annotation	17
Table 2: Comparison of the Gene prediction and annotation results	27
Table 3: CRISPR sites.....	29
Table 4: RAST Annotation corresponding to aromatic compound metabolism	32

LIST OF FIGURES

Figure 1: Graphical representation of the general protocol used in this research	15
Figure 2: Evaluation of Quality of the genome assembly	21
Figure 3: BUSCO assessment results	22
Figure 4: kraken2 taxonomical classification	23
Figure 5: Proteins associated to Superfamilies	24
Figure 6: Proteins associated with superfamilies	25
Figure 7: Proteins associated with superfamilies	25
Figure 8: Proteins associated with superfamilies	26
Figure 9: An example of the alignment of the gene prediction using three different prediction tools. Here, the gene prediction for Glutaredoxin 1 is shown	27
Figure 10: Gene ontology distribution in the genome	28
Figure 11: Whole genome Annotation of Halomonas elongata	30
Figure 12: Naphthalene degradation pathway as predicted on KEGG database	36
Figure 13: Pathway mapping of metabolism of xenobiotics by cytochrome P450.	37

LIST OF ABBREVIATIONS

PAHs	:	Polycyclic aromatic hydrocarbons
WGS	:	Whole genome sequencing
LMW-PAHs	:	Low molecular weight- polycyclic aromatic hydrocarbons
HMW-PAHs	:	High molecular weight- polycyclic aromatic hydrocarbons
KFUPM	:	King Fahd University of Petroleum and Minerals
ONS	:	Oxford Nanopore Sequencing technology
KAUST	:	King Abdullah University of Science and Technology
BUSCO	:	Benchmarking Universal Single-Copy Orthologs
BLAST	:	Basic Local Alignment Search Tool
CDD	:	Conserved Domain Database
GO	:	Gene ontology
KEGG KAAS	:	KEGG Automatic Annotation Server
KO	:	KEGG Orthology
HGT	:	Horizontal Gene Transfer
RAST	:	Rapid Annotations using Subsystems Technology
CDS	:	Coding sequences
rRNA	:	ribosomal RNA
tmRNA	:	transfer-messenger RNA
tRNA	:	transfer RNA
LINEs	:	Long Interspersed Nuclear Elements

SSRs	:	simple sequence repeats
SINEs	:	Short Interspersed Nuclear Elements
NRPS	:	Non-Ribosomal Peptide Synthetase
RiPP-like	:	Ribosomally synthesized and post-translationally modified

ABSTRACT

Full Name : Thasneema Rafic
Thesis Title : *Whole Genome sequencing, Characterization and Analysis of coronene degrading bacterial strain, Halomonas caseinilytica: A Bioinformatics Approach*
Major Field : Life Science
Date of Degree : December 2024

Polycyclic aromatic hydrocarbons (PAHs) are persistent organic pollutants with significant environmental and health risks. Among them, high molecular weight PAHs such as coronene present a unique challenge due to their stability and resistance to degradation. Bioremediation using microbial strains offers a sustainable and cost-effective approach to PAH remediation, yet only a few bacterial species have been identified with the capability to degrade coronene. This study employs whole genome sequencing (WGS) and bioinformatics analysis to characterize a coronene-degrading bacterial strain, initially identified as *Halomonas caseinilytica* through 16S rRNA sequencing. Using Oxford Nanopore sequencing technology, the complete genome was assembled and analysed, with phylogenetic classification suggesting a reclassification of the strain as *Halomonas elongata*. Gene prediction and annotation identified 4,308 genes, including those associated with hydrocarbon degradation, stress adaptation, and metabolic versatility. Functional and pathway analysis using KEGG and RAST revealed multiple genes involved in xenobiotic metabolism and aromatic compound degradation, while secondary metabolite biosynthetic clusters, operons, and mobile genetic elements were also detected.

The findings of this study adds to the expanding body of knowledge on halophilic bacteria with bioremediation potential, particularly in saline environments affected by oil pollution. These insights pave way for further experimental validation and potential biotechnological applications in environmental cleanup strategies.

ملخص الرسالة

الاسم الكامل: تسنيمه رفيق

عنوان الرسالة: تسلسل الجينوم الكامل، توصيف وتحليل سلالة بكتيرية محللة للكورونين، *Halomonas caseinilytica*: نهج المعلوماتية الحيوية

التخصص: علوم الحياة

تاريخ الدرجة العلمية: ديسمبر 2024

الهيدروكربونات العطرية متعددة الحلقات (PAHs) هي ملوثات عضوية دائمة تشكل مخاطر بيئية وصحية كبيرة. ومن بينها، تُعدّ الهيدروكربونات العطرية متعددة الحلقات ذات الوزن الجزيئي العالي مثل الكورونين تحديًا فريدًا نظرًا لاستقرارها ومقاومتها للتحلل. تُعدّ المعالجة البيولوجية باستخدام سلالات ميكروبية نهجًا مستدامًا وفعالًا من حيث التكلفة لإزالة PAHs، ومع ذلك فقد تم تحديد عدد قليل فقط من أنواع البكتيريا التي تمتلك القدرة على تحلل الكورونين.

تعتمد هذه الدراسة على تسلسل الجينوم الكامل (WGS) والتحليل المعلوماتي الحيوي لتوصيف سلالة بكتيرية قادرة على تحلل الكورونين، والتي تم تحديدها في البداية على أنها *Halomonas caseinilytica* من خلال تسلسل جين *rRNA16*. باستخدام تقنية تسلسل Oxford Nanopore، تم تجميع الجينوم الكامل وتحليله، حيث أشار التصنيف الوراثي إلى ضرورة إعادة تصنيف السلالة إلى *Halomonas elongata*. أدت تنبؤات الجينات والتعليقات التوضيحية إلى تحديد 4,308 جينًا، بما في ذلك تلك المرتبطة بتحلل الهيدروكربونات، والتكيف مع الإجهاد، والتنوع الأيضي.

كشفت التحليلات الوظيفية ومسارات الأيض باستخدام قواعد بيانات KEGG و RAST عن عدة جينات متورطة في أيض المواد الغريبة وتحلل المركبات العطرية، بينما تم الكشف أيضًا عن تجمعات تصنيع المستقبلات الثانوية، والأوبيرونات، والعناصر الجينية المتنقلة.

تساهم نتائج هذه الدراسة في إثراء المعرفة المتزايدة حول البكتيريا المحبة للملوحة ذات الإمكانيات في المعالجة البيولوجية، لا سيما في البيئات المالحة المتأثرة بتلوث النفط. كما تمهد هذه النتائج الطريق لمزيد من التحقق التجريبي والتطبيقات المحتملة في التكنولوجيا الحيوية لاستراتيجيات تنظيف البيئة.

CHAPTER 1:

INTRODUCTION

1.1. Background

Polycyclic aromatic hydrocarbons (PAHs) are a prominent assemblage of organic chemicals comprising 2 or more aromatic rings that are fused together in different orientations, namely, clustered, angular or linear (Kadri et al., 2017). These compounds are broadly categorised based on the amount of constituent aromatic rings. Those that comprise less than 4 aromatic benzene rings are called LMW-PAHs (low molecular weight PAHs) while those with more than 4 rings are called HMW-PAHs (high molecular weight PAHs) (Lawal, 2017). They tend to have high boiling and melting temperatures in contrast to their solubility and vapour pressure, which remain low. An increase in molecular weight correlates with higher melting and boiling points, while simultaneously reducing vapor pressure and solubility. Although PAHs originate from both natural and anthropogenic sources, their formation is predominantly attributed to incomplete decomposition of organic compounds such as petroleum gas, oil, wood and coal hence accounting for their widespread occurrence (Haritash & Kaushik, 2009; Kadri et al., 2017). PAHs demonstrate greater structural diversity than any other category of non-halogenated molecules in the environment and the biological systems. Under natural conditions, they are challenging to degrade and tend to persevere. The retention of these compounds increases with increasing molecular weight, thereby posing a threat due to their omnipresent nature, persistence, resistance to natural degradation, tendency to bioaccumulate and carcinogenic and/or mutagenic effects (Kadri et al., 2017). In the recent years, increased anthropogenic activities has resulted in the accumulation of PAHs in the ecosystem (Ghosal et al., 2016)

The Arabian Gulf region is a leading producer of oil and gas, holding approximately 65% of the world's crude oil reservoirs (Modarress et al., 2016). This leads to a greater sector of associated activities such as transportation, refining or even extraction that are instrumental in the scenarios of accidental spillage from tankers, ships or even pipelines that discharge the PAHs into the soil or coastal environment of the region (Hosseini et al., 2021). Furthermore, with perpetual production and transportation of petroleum and related oil commodities, the concentration of PAHs is only expected to increase in the future (Haritash & Kaushik, 2009).

In the past few decades, strategies of different nature have been propositioned to alleviate the detrimental effects of PAHs. However, biological strategies have an upper hand over physical and chemical approaches by providing lucrative and long term solution (Okeyode et al., 2023). Different scientific investigations have explored the potential of different microorganisms' ability to digest PAHs. Organisms such as algae, fungi as well as bacteria have been studied and assessed for their propensity to metabolise PAHs as a source of energy (Kadri et al., 2017). These investigations predominantly show the competence of microbial species to remediate LMW-PAHs. In case of HMW-PAHs, pyrene and benzopyrene have been thoroughly studied and reported. On the contrary, data is limited on the bioremediation of coronene. This can be accredited to the complexity of the seven ringed compound which makes it difficult to metabolise it (Ghosal et al., 2016; Juhasz et al., 2000).

As contamination due to oil spillage has caused a large quantity of undesirable hydrocarbons, including PAHs, to accrue in the ecosystem, bioremediation has been a consistent topic of research. A major challenge in bioremediation is to understand the physiology, biochemical pathways, metabolism, proteins and genetics that are fundamental to a bacterium's ability to bioremediate. Although most mechanisms are yet to be elucidated, the need to develop successful strategies to eliminate PAHs from the ecological community is imperative, especially when the extend of global oil usage is considered (Hidalgo et al., 2020). Identification and classification

of novel remediating species is pivotal to further engineer industrial applications of biodegrading bacteria. Detailed understanding of the whole bioremediation process will improve our expertise in monitoring and handling cases of accidental oil spills. It can also be used to eliminate the chances of oil degradation in reservoirs that could reduce the desirable property of the crude oil (Hidalgo et al., 2020). Recent developments in biotechnology have enabled precise taxonomical identification of organisms and hence, commendably reducing errors. For instance, whole genome sequencing technologies accurately analyses the whole genome of a cell. This form of sequencing technique overcomes the shortcoming of the 16S RNA sequencing that is conventionally used to identify bacterial strains but has a resolution limited to genus level (Raina et al., 2019). 16S amplicon sequencing method also sequences a specialized yet restricted region of the genome whereas Whole Genome Sequencing covers the whole genome, hence increasing the accuracy. In addition to enhanced detection of species diversity, longer read length improves the accuracy of species identification (Ranjan et al., 2015).

Cognising the underlying genes, proteins and biochemical pathways is necessary to improve bacterial based bioremediation strategies. In addition, the comprehension of distinct metabolic pathways facilitates the formulation of strategies by researchers to enhance and streamline these processes, thereby ensuring effective removal of PAHs. It also allows to foresee the competence of specific bacterial strain to biodegrade various contaminants. For instance, the potential of a known LMW-PAHs degraders to degrade HMW-PAHs such as coronene can be detected. This prognostic ability is indispensable for the discerning selection of bacteria most apt for focused remediation initiatives. In addition, genomic and proteomic knowledge allows researchers to genetically engineer microorganisms with enhanced bioremediation capabilities.

As mentioned earlier, microorganisms degrading HMW-PAHs, especially coronene are rare. However, in 2023, a coronene degrading strain was isolated from fuel station soil of KFUPM (King Fahd University of Petroleum and minerals), Saudi Arabia. This strain was then identified

as *Halomonas caseinilytica* by 16s rRNA sequencing. In this proposed study, we further expand on the isolate by performing whole genome sequencing and bioinformatic analysis to accurately identify, compile and report relevant Genomic and proteomic profile.

1.2. Objectives

- 1) Whole genome sequencing and assessment of the completeness of the sequence
- 2) Taxonomical classification and identification of the strain
- 3) Identification of genes and proteins related to bioremediation of PAHs and of other importance

CHAPTER 2:

LITERATURE REVIEW

In a 2003 estimate, an annual seepage volume of naturally occurring crude oil was recorded to be 600,000 metric tons, with an error gamut of 200,000 metric tons (Fates, 2003). The environmental discharge of hydrocarbons stemming either from accidental events or anthropogenic activities is a primary contributor to water and soil pollution. In 2023 itself, the quantity of oil spill worldwide from oil tankers was recorded to be 2000 metric tons. Soil contamination with hydrocarbons poses significant hazard to local ecosystems, as the amassment of pollutants in biological tissues can induce mutagenicity or even death (Alvarez & Vogel, 1991; Holliger et al., 1997). Commonly employed technologies for soil remediation include mechanical method as well as evaporation, dispersion, washing, and burial. However, these approaches are not cost effective and may not fully eradicate contaminants, leaving remediation incomplete (Das & Chandran, 2011). In a 2002 study by Verrhiest and colleagues, the interaction between microbial communities and PAH mixtures was assessed and the study concluded that (a) PAHs agitate microbial communities, (b) Toxic assessment can identified using indigenous bacteria, and (c) Depending on the physiochemical properties, indigenous bacteria have high proclivity to degrade PAHs (Verrhiest et al., 2002).

Jones et al. reported the biological degradation of alkyl aromatics in marine sediments and bacterial species of *Sphingomonas*, *Mycobacterium*, *Burkholderia*, *Pseudomonas*, *Arthrobacter* and *Rhodococcus* were related to the alkyl aromatic biodegradation (D. M. Jones et al., 1983). Bioremediation of petroleum hydrocarbons was observed in a contaminated humid stream in Lagos, Nigeria where a total of nine bacterial strains were isolated of which *Micrococcus roseus*, *Pseudomonas fluorescens*, *Acinetobacter lwoffii*, *Bacillus subtilis*, *P. aeruginosa* were a few of

them (Adebusoye et al., 2007). The efficiency of bioremediation by bacteria is recorded to range between 0.13% to 50% for bacteria thriving in soil while marine bacteria have a staggering range between 0.003% to 100%.(Hollaway et al., 1980; J. G. Jones et al., 1970; Mulkins Phillips & Stewart, 1974; Pinholt et al., 1979). Bacteria, in fact, are observed to be the most active of all the biological bioremediating agents, where they play the role of primary degraders in oil spilled environment. *Daugulis and McCracken* observed the biodegradation of PAHs by species of *Sphingomonas* (Brooijmans et al., 2009; Daugulis & McCracken, 2003; Rahman et al., 2003). Most bacterial species that have PAH biodegrading abilities have been isolated from contaminated soil exposed to petrochemical waste for a very long time. For instance, *Pseudomonas aeruginosa* was sequestered from a stream severely contaminated by a petroleum refinery (Romero et al., 1998). This strain was growing on high concentration of phenanthrene with a competency to completely eradicate phenanthrene in 30 days. In the same year, *Mycobacterium species* was identified in an area near petrogenic chemical source capable of mineralizing pyrene(Heitkamp et al., 1988; Heitkamp & Cerniglia, 1989). Rehmann and co-workers isolated and classified *strain KR2* of *Mycobacterium* that metabolises pyrene as the sole carbon resource. The metabolic products detected included 1- hydroxy-2-naphthoic acid, 4-5-phenanthrene dicarboxylic acid, phthalic acid and others, all of which were mentioned in the proposed degradation pathway (Rehmann et al., 1998). In 2002, Six bacterial gram-negative strains were isolated from the soil of a petrochemical litter disposal site (Yuan et al., 2002). These strains could degrade phenanthrene, acenaphthene, fluorene, pyrene and anthracene by 70-100% in a duration of 40 days from the time of first exposure. Additionally, bacteria not typically growing in contaminated zones were also observed to have the competence to degrade PAHs. For instance, a consortium of *Rhodococcus sp.*, *Acinetobacter sp.*, and *Pseudomonas sp.* enriched from mangrove sediments achieved 100% degradation of fluorene and phenanthrene within 4 weeks while pyrene took longer to degrade (Yu et al., 2005).

In the past decade, studies on PAH degradation in Saudi Arabia have seen a significant rise. As one of the leading oil and petroleum-producing countries, it is inevitable for the region to harbour naturally occurring bacteria that either possess or acquire the ability to degrade PAHs. Three different PAH degraders were identified from crude oil samples in 2012. These three strains, namely, *Acinetobacter* sp., *Klebsiella pneumonia* and *Klebsiella oxytoca* were able to utilise 4 different PAHs (benzene, pyrene, phenanthrene and fluorene) (Eman et al., 2012). In 2016, a *Methylobacterium radiotolerans* isolated from Jubail industrial area was first documented to be able to exclusively use naphthalene as carbon resource (Nzila et al., 2016). *Pseudomonas citronellolis* and *Ochrobactrum intermedium* were identified as effective PAH degraders. *Ochrobactrum intermedium* showed efficient pyrene degradation which was about 62% in 15 days whereas *Pseudomonas citronellolis* degraded 94% of phenanthrene within 15 days (Oyehan & Al-Thukair, 2017). Co-cultures of *Thermomonas* and *Ralstonia* species and that of *Stenotrophomonas* and *Pseudomonas* species were seen to effectively anthracene and phenanthrene (Nzila, Sankara, et al., 2018). Novel strain of *Caulobacter* and *Burkholderia* were tested for the effect pH and temperature on their ability to degrade pyrene. This study emphasised the possibility of HMW-PAH degradation by these strains (Al-Thukair & Malik, 2016). More recently, *Sphingomonas* species under optimal conditions presented a promising approach for the efficient degradation of anthracene (Al Farraj et al., 2021). A halophilic consortium was found to be able to degrade compounds such as fluorene, pyrene, phenanthrene and hexadecane at different degrees based on their concentration (Amran et al., 2022). *Stenotrophomonas maltophilia* strain AJH1 was studied to be capable of degrading both LMW-PAHs as well as HMW-PAHs up to 95% and 80%, respectively (Arulazhagan et al., 2017). *Rhodococcus qinshengi* was found to entirely break down naphthalene and over 50% of pyrene (Al-Thukair et al., 2020). Species of *Proteus* and *Brevibacillus* were identified to be able to degrade pyrene and naphthalene (Al-Thukair et al., 2020). Another study observed the degradation potent of both LMW-PAHs and HMW-PAHs by a consortium of halothermophilic bacteria in a tank reactor

(Pugazhendi et al., 2017). In samples collected from the shoreline of the Arabian Gulf, Jubail, Nzila et. al identified an *Achromobacter* strain with the competence to degrade pyrene (Nzila, Ramirez, et al., 2018). In 2021, *Staphylococcus haemolyticus* strain 10SBZ1 was identified to degrade LMW-PAHs efficiently (Nzila et al., 2021). In another study, 70% of pyrene from petroleum polluted soil was degraded by *Klebsiella oxytoca* within 12 days (Alfaify et al., 2022). *Halomonas shengliensis* and *Halomonas smyrnensis* were observed to be rapidly growing in the presence of pyrene at concentrations less than 50ppm (Budiyanto et al., 2018). Studies have also suggested the use of PAHs degrading bacteria as an indicator to the presence of crude oil reservoir (Ibrahim et al., 2015; Jamal & Pugazhendi, 2018). And example of such studies was done at the King Abdulaziz university, Jeddah, where a consortium of halophilic *Marinobacter* were pronounced to be good indicators of crude oil (Jamal, 2022). Most studies in the country have confirmed the use of halophilic bacteria in the treatment of saline petroleum wastewater system where these bacteria are mainly working on the degradation of PAHs (Amran et al., n.d., 2022; Jamal, 2022; Jamal & Pugazhendi, 2018; Pugazhendi et al., 2017). Other studies have also seen that thermophilic bacteria have a higher tendency to possess the ability to degrade PAHs (Amran et al., 2022; Pugazhendi et al., 2017).

Benzo(a)pyrene (BaP) is deemed among the lethal and carcinogenic HMW-PAHs. 11 BaP degrading strains were isolated by Aitken et al. from sites of refinery, motor oil, oil and wood treatment of which at least 3 species of *Pseudomonas* and other species of *Burkholderia*, *Agrobacterium*, *Sphingomonas* and *Bacillus* were detected (Aitken et al., 2011). In addition, *Mycobacterium*, *Rhodococcus* as well as mixed cultures of *Flavobacterium* and *Pseudomonas* have demonstrated BaP degrading potency (Schneider et al., 1996; Trzesicka-Mlynarz & Ward, 1995; Walter et al., 1991). Grosser and co-workers, in 1991, observed co-metabolic remediation of benzo[a]pyrene by a pyrene degrading bacterial and yeast extract (Grosser et al., 1991). The strain was identified to be a *Mycobacterium* species. This had an advantage over a previous study

where the degradative species could not be identified from pure culture (Heitkamp & Cerniglia, 1989). Recently, Nzila and colleagues assessed three novel bacterial strains of *Bacillus cereus*, *Micrococcus luteus* and *Brad rhizobium japonicum* for their ability to use BaP as the sole carbon source (Nzila et al., 2023). Another study by the same group identified a *Staphylococcus* strain capable of BaP along with LMW-PAHs (Nzila et al., 2021). Juhasz and co-workers confirmed the ability of 3 strains of *Burkholderia cepacia* to slowly degrade benzo[a]pyrene, dibenz[u,h]anthracene and coronene. This was the first study to attain degradation of the seven-ring compound, coronene. One of these strains was later reclassified based on 16s RNA gene sequencing as *Stenotrophomonas maltophilia* (Juhasz et al., 1996, 2000). Another mix-culture study of bacterial species of *Klebsiella* and *Acinetobacter* was carried out to decontaminate PAHs polluted soil where a substantial degradation of 98% of the overall PAHs was observed in a span of six months. The LMW PAHs were reduced to 0.5% of initial concentration while the HMW PAHs were reduced to 5% of the initial concentration (Haritash & Kaushik, 2009). In 2023, Okeyode et al. reported the capability of a halophilic bacteria *Halomonas caseinilytica* to utilise coronene as the only supply of carbon under highly saline condition (Okeyode et al., 2023). This was the second strain reported to be adept in degrading coronene.

2.1. Omics studies

To accomplish total bioremediation of a contaminated environment, it is necessary to institute and sustain favourable conditions to expediate the rate of PAH degradation. A number of literatures dating as far back as 1940s have shone light on the different factors that impact biodegradation rates (Atlas & Bartha, 1992; FOGHT & WESTLAKE, 1987; Mahmoud & Bagy, 2019; ZOBELL, 1946). To begin with, the presence of microorganisms with suitable metabolic functions is vital provided these organisms are catered with sufficient oxygen, nutrient and appropriate pH. It has been understood that biodegradation of Petroleum Hydrocarbons is a

multifaceted course of action which relies on various parameters such as its nature and quantity, where quantity has a crucial role in limiting the rate of microbial degradation. In addition, petroleum hydrocarbons bind to soil and become challenging to degrade (Barathi & Vasudevan, 2001). This makes microorganism-based biodegradation the ultimate natural means to eliminate hydrocarbon waste products from the site of contamination (Amund & Nwokoye, 1993; Lal & Khanna, 1996; Wackett, 2012).

Bacterial species use diverse biochemical pathways such as anoxygenic, phototrophic, aerobic, chemotrophic and/ or anaerobic to degrade hydrocarbons (Varjani, 2017). As mentioned earlier, these pathways are curtailed by external factors such as accessibility of oxygen and nutrients, degree of temperature, pH, PAHs concentration and their respective chemical and physical characteristics as well as microbial richness (Hidalgo et al., 2020). Internally, enzymes, genes and genetic regulations play a superior role in enabling a bacterium to degrade PAHs and hence this has been a topic of focus of many researchers over the past few years (Ghosal et al., 2016). State of the art meta-analysis technologies, mainly in the fields of genomics and taxonomy can be applied to monitor and dissect the diversity of bacterial genes and proteins (Thomas et al., 2012). High throughput sequencing has been successfully used to screen bioremediation in environments contaminated with hydrocarbons (Coulon et al., 2012; Dos Santos et al., 2011; Simon & Daniel, 2011). Improvement in omics have given better insights into the dynamics of diverse microorganisms and their specific functional genes and proteins competent for the degradation of hydrocarbons polluted biological networks (Pieper & Reineke, 2000; Ron & Rosenberg, 2014). This has commendably increased our knowledge about the bioremediating abilities of inhabitant bacteria flourishing in contaminated environments (Urgun-Demirtas et al., 2006) and made the underlying PAH related metabolic, genetic and enzymatic profiles perspicuous. In context of biochemical degradation pathways of PAHs; pyrene, anthracene,

phenol, naphthalene, BTEX (monoaromatics) and phenanthrene have been decoded by many reviews at the enzyme level (Heider, 2007; Meckenstock et al., 2016; Waigi et al., 2015).

Usually, aerobic conditions are prerequisite for complete degradation of organic pollutants. Enzymatic action of oxygenases and peroxidases catalyse the initial incorporation of oxygen into the organic pollutant. Secondary pathways then convert these compounds into intermediates and derivative compounds which are eventually infused into cellular biomass. Under aerobic conditions, after being catalyzed by an oxygenase or a dioxygenase enzyme, a cis- dihydrodiol compound is produced (Cerniglia, 1992; Cerniglia & Sutherland, 2010; Ghosal et al., 2016; Juhasz et al., 2000) which submits to the enzymatic action of cis-dihydriol dehydrogenase resulting in the rearrangement of the aromatic ring (Shahsavari et al., 2019). Following this, the position of -OH (hydroxyl group) determines the actions of enzymes intradiol cleaving dioxygenase and extradiol cleaving dioxygenase (Cerniglia, 1992; Juhasz et al., 2000). Similar process continues into each ring of the PAH until a metabolic intermediate is produced which is then utilised by the bacterium. In gram positive bacteria, *nid* and *pdo* genes determine the dioxygenases related to degradation of HMW-PAH (Ghosal et al., 2016; Gupta et al., 2015). Conversely, the discovery and quantification of pyrene degrading bacteria was much more efficacious using the *gene pah-rdh* which are genes for PAH-ring hydroxylating dioxygenase than using *nidA gene* (Bacosa & Inoue, 2015). Dioxygenases enzymes relevant to LMW-PAH are reported to be coded by *nah* (naphthalene dioxygenase) *genes* which have been extensively studied in gram negative species of *Pseudomonas* (Gupta et al., 2015; Zhou et al., 2012). Enzymes encoded by *nahAaAbAcAdBCDEF genes* and *nahGTHINLOMKJ genes* are responsible for the degradation of LMW-PAHs. Additionally, *pah* and *phn genes* have also been reported in *Comamonas testosterone* that degrades phenanthrene and naphthalene as well as in *Burkholderia* sp., that seemingly degrades phenanthrene, anthracene and naphthalene (Gupta et al., 2015). On the other hand, degradation of PAHs under anaerobic conditions is relatively complex and has

been chiefly studied on diaromatic structures such as naphthalene, validated by the knowledge that naphthalene usually is an intermediate form in the degradative pathways of PAHs that have over two aromatic rings (Von Netzer et al., 2016). Enzymes such as naphthylmethyl succinate synthase and aryl-CoA reductase play a fundamental role in anaerobic degradation (Eberlein et al., 2013; Meckenstock et al., 2016; Musat et al., 2009). Some studies have suggested the possibility of hydroxylation; however, this has not been proven in anaerobic pathways (Bedessem et al., 1997). Dearomatization of the rings are carried out by 2-naphthoyl-CoA (Von Netzer et al., 2016). Phenanthrene has degradative enzymes akin to naphthalene where some scientific investigations have shown parallel degradation of both PAHs (Coates et al., 1997; Rockne & Strand, 2001).

Genomic analysis in the marine environment has remarkably contributed to the understanding of PAH biodegradation where studies have found growth of PAH degrading bacteria at the site of oil spills, natural sediment or oil reservoirs. For instance, Deepwater Horizon well explosion in 2010 spread as far as 1100 km of the coast. In the same year, Hazen and colleagues surveyed the oil trail samples from the surroundings of the spill site and observed *Oceanospirillales* predominantly populating the area, which generally consists of species such as *Thalassolituus oleivorans*, *Oleiphilus messinensi* and *Oleispira antarctica* that are well-known psychrophilic degraders of hydrocarbons. Additionally, genes such as *phdCI* gene translating carboxylate isomerase that facilitate hydrocarbon degradation were at high incidence in these samples (Atlas & Hazen, 2011; Hazen et al., 2010). Other genomics-based investigations at the same site also observed higher incidence of hydrocarbon remediators such as *Pseudomonadaceae* and genes associated with toluene, n-alkane, and PAH degradation in the sediments closer to the spill (Lamendella et al., 2014). In deeper sediments, functional gene assays identified the genes *assA* and *bssA* connected to hydrocarbon degradation (Kimes et al., 2014). 16s RNA sequencing of

the microbial communities in groundwater tainted with diesel testified the presence of anaerobic degraders *Desulfitobacterium* and *Geobacter* species (Ramos et al., 2014).

Overall, genomic analysis has shown genera such as *Acinetobacter* and *Pseudomonas* to be the most prevalent in multiple environments. Oil reservoirs and seawater samples are mainly populated by *Marinobacter*, *Dietzia* and *Bacillus* sp. Whereas Genus *Syntrophus* was identified in both groundwater and oil reservoirs. Exclusively, genera such as *Halomonas*, *Cycloclasticus* and *Colwellia* were found in seawater (Hidalgo et al., 2020). As scarce as it is, most genomic research on PAHs have successfully developed data on organisms and genes associated with degradation of LMW-PAHs. One such example is the whole genomic sequencing of *Idiomarina piscisalsis* strain that is capable of degrading pyrene (Nzila, Kwon Jung, et al., 2018). Although knowledge on naphthalene degradative pathways can be used to understand HMW-PAHs remediation as it forms an intermediate compound, a complete apprehension is yet to be investigated. In case of coronene, only *Burkholderia cepacia* and as of recent, *Halomonas caseinilytica* has shown to degrade coronene. By combining modern molecular science technologies such as metabolomics, transcriptomics and genomics and further applying bioinformatic tools, novel strains and biodegradative pathways can be identified which in turn will contribute to improve bioremediation strategies by either improving strains or by artificially developing enzymes that can annihilate the PAHs completely from the polluted environment.

2.2. Research related to the bacterial strain used in the proposed study

In the 1996 study of Juhansz and co-workers, the first strain to degrade coronene was identified. However, this degradation occurred only under high bacterial density or by co-metabolism (Juhasz et al., 1996, 1997). In 2023, Okeyode et al. investigated the soil samples from a fuel station of KFUPM (King Fahd University of Petroleum and Minerals), Saudi Arabia. They were

able to isolate and identify a coronene degrading species, *Halomonas caseinilytica* 10SCRN4D (Okeyode et al., 2023). 16s rRNA gene sequencing was employed to recognize the strain. This strain demonstrated competence to use coronene as the lone carbon source under an environment with high salinity of 10% NaCl, w/v. This was the first report of a *Halomonas strain* degrading coronene, specifically under halophilic condition.

CHAPTER 3:

MATERIALS AND METHODS

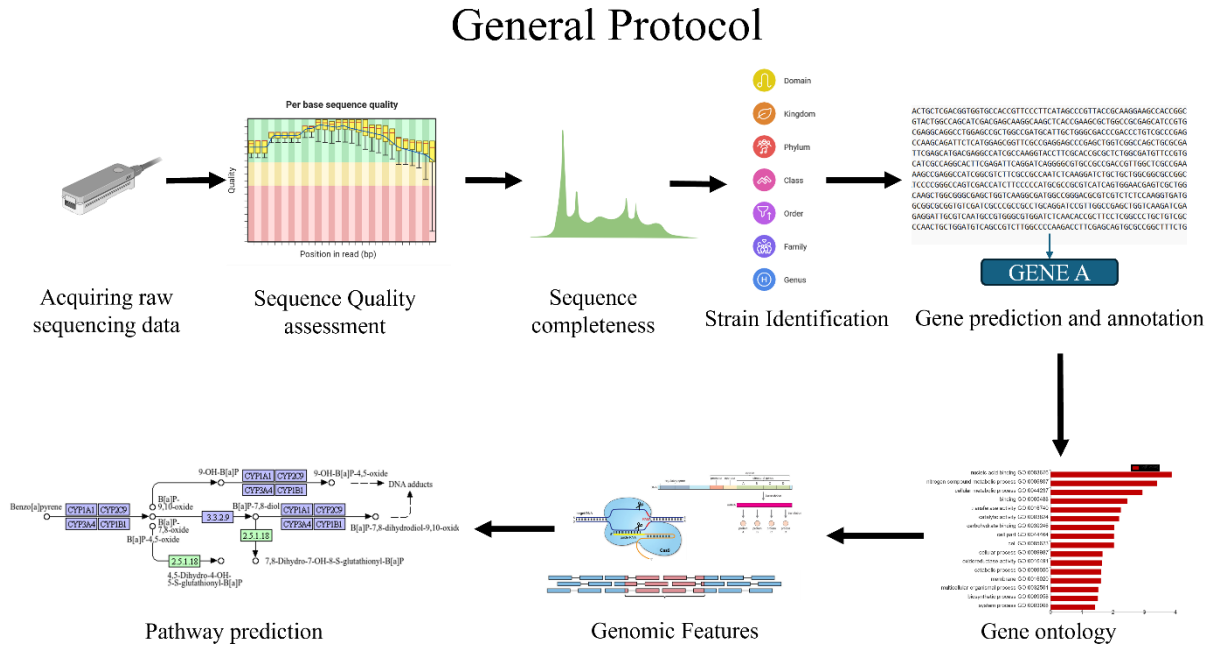


Figure 1: Graphical representation of the general protocol used in this research

3.1. Whole genome sequencing and Quality assessment

Third generation Oxford Nanopore Sequencing (ONS) was used to sequence the whole genome of the bacteria at King Abdullah University of Science and Technology (KAUST). Post sequencing, the quality of the generated reads was assessed using NanoPlot, a tool that provides comprehensive quality statistics and graphical reports. This step ensured that only high-quality reads were used for subsequent analyses, optimizing the reliability of the assembly and annotations (De Coster & Rademakers, 2023). The completeness of the assembled genome was evaluated using Benchmarking Universal Single-Copy Orthologs (BUSCO) v 5.4.6, with an E-

value cutoff of 0.001 for BLAST searches (Simão et al., 2015). This step verified the assembly's completeness and the presence of conserved orthologous genes.

3.2. Taxonomical classification and Strain Identification

Taxonomic classification was performed using Kraken2 (Lu et al., 2022), which assigns taxonomic labels to short DNA sequences with high accuracy. This step facilitated the identification of the microbial strain and its phylogenetic placement based on obtained genomic sequences. A confidence score of 0.5 and minimum hit group of 2 was applied.

3.3. Gene Prediction and Functional Annotation

To maximize the accuracy and comprehensiveness of gene annotation, three distinct gene prediction tools were employed, each coupled with a specific annotation tool tailored to its output. First, gene prediction was performed using PROKKA (Seemann, 2014) with a e-value cut off of $1e-06$ followed by immediate annotation using its own integrated annotation system. Conserved Domain Database (CDD) (Marchler-Bauer et al., 2011; Wang et al., 2022) searches were performed for hypothetical proteins predicted by PROKKA to identify potential functions based on conserved protein domains. An e-value threshold of 0.001 and maximum number of hits of 500 was applied to increase the accuracy of the results. The hypothetical genes identified through CDD were then combined with the annotations from PROKKA. In addition, PRODIGAL (Hyatt et al., 2010) and GeneMarkS2 (Lomsadze et al., 2018) were used to predict the genes which were then annotated using EggNOG-mapper v2 (Cantalapiedra et al., 2021).

MAFFT, a multiple sequence alignment program, was utilized through its command-line interface to align the gene sequences predicted by the aforementioned tools to ensure precise annotation (Kato et al., 2019). To achieve this, the three annotations were first combined and

then MAFFT algorithm was used to align the predictions of each tool with one another. These preliminary alignments were then visualized using MEGA, which provided an intuitive graphical interface to inspect and manually refine alignments (Tamura et al., 2021). Finally, Gene ontology (GO) based functional annotation was performed using InterProScan and Blast2GO (Paysan-Lafosse et al., 2023). *Table 1* compares the different tools used for gene prediction and annotation.

Table 1: tools used for gene prediction and annotation

Tools or algorithm used	Purpose	Benefit	Reference
PROKKA	Gene prediction and annotation	Fast and capable of producing detailed annotations including coding sequences, rRNA, tRNA, and non-coding RNA	(Seemann, 2014)
PRODIGAL	Gene prediction	Highly accurate at predicting protein-coding genes in bacterial and archaeal genomes	(Hyatt et al., 2010)
GeneMarkS2	Gene prediction	Uses both intrinsic (sequence-based) and extrinsic (homology-based) methods to predict genes	(Lomsadze et al., 2018)
EggNOG-mapper	Annotation	Maps sequences to orthologous groups and functional annotations directly from the eggNOG database.	(Cantalapiedra et al., 2021)

RASTk	Functional Annotation	Provides annotations within the context of subsystems, allowing for a layered understanding of biological functions.	(Aziz et al., 2008)
-------	--------------------------	--	---------------------

3.4. Identification of Genomic features

To investigate gene organization and regulation, Operon Mapper was used to predict operons, offering insights into gene clustering and transcriptional control (Taboada et al., 2018). Detection of CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) arrays was performed using CRISPRCasFinder with the following parameters: flanking region size of 100, repeat lengths ranging from 23 to 55, and spacer sizes between 0.6 and 2.5 (Couvin et al., 2018). The Resistance Gene Identifier in CARD database facilitated the identification of genes conferring antibiotic resistance (Alcock et al., 2023).

For a deeper understanding of genome structure and evolution, RepeatMasker v4.1.5 was used to detect repeat elements (Tempel, 2012), and RepeatModeler v2.0.5 facilitated *de nova* annotation (Flynn et al., 2020). Palindrome v5.0.0.1 was applied to detect inverted repeats, with set parameters: repeat lengths of 10–100 bp, a maximum gap of 100 bp, and no mismatches allowed (Rice et al., 2000). Recognizing the importance of mobile genetic elements in bacterial evolution, MobileOG-DB was used (e-value: 1.0e-05, k-value: 1) to pinpoint these elements (Brown et al., 2022).

Additionally, Alien Hunter identified potential horizontal gene transfer (HGT) events, critical for the acquisition of novel traits (Vernikos & Parkhill, 2006). antiSMASH web-based tool assessed the genome for secondary metabolite biosynthesis gene clusters (Blin et al., 2023).

3.5. Pathway analysis

For pathway analysis, two complementary approaches were employed: the RAST (Rapid Annotations using Subsystems Technology) server and KEGG KAAS (KEGG Automatic Annotation Server). The RAST server was used to annotate genes based on curated subsystems and protein families, providing a comprehensive metabolic reconstruction (Aziz et al., 2008; Overbeek et al., 2014). KEGG KAAS was utilized to perform functional annotation of genes by comparing amino acid sequences against the manually curated KEGG GENES database (Moriya et al., 2007). KAAS assigned KEGG Orthology (KO) identifiers using the bi-directional best hit (BBH) method. The resulting KO assignments were then used to automatically generate KEGG pathways and functional classifications.

CHAPTER 4:

RESULTS

4.1. Whole genome sequencing and Quality Assessment

ONS facilitated real-time genome sequencing and yielded long-read assemblies, resulting in a total of six contigs with a combined length of 3,966,854 nucleotides. Among these, the largest contig encompassed 1,702,422 bases, contributing significantly to the total assembly length. A key quality metric, the count of ambiguous ‘N’ bases per 100 kilobase pairs (kbp), was zero, indicating that the assembly contained no unresolved nucleotide positions over this length scale. The absence of ambiguous bases suggests a high-continuity assembly with minimal sequencing errors and no apparent gaps, reinforcing the overall high quality and completeness of the genome (figure 2).

The GC content of the assembled genome was determined to be 63.04%, which, although within the general acceptable range of 40%–80%, is notably elevated. A high GC composition is often correlated with greater genomic stability due to the increased number of hydrogen bonds in guanine-cytosine base pairs, which enhance DNA duplex stability. This property is frequently observed in thermophilic organisms, suggesting that the studied species may be adapted to high-temperature environments where increased thermal stability is advantageous.

To further evaluate the completeness and accuracy of the genome assembly, BUSCO analysis was conducted, which assesses genome quality by detecting the presence of highly conserved orthologous genes across evolutionary lineages. Out of the 619 BUSCO groups searched, 506 genes were identified as complete, of which 505 were present as single-copy genes, while one gene was detected as duplicated. Additionally, 72 BUSCO genes were classified as fragmented, indicating partial assembly, while 41 genes were entirely absent, suggesting potential gaps or

lineage-specific gene loss (Figure 3). These results collectively indicate that the genome assembly exhibits high completeness, with a strong representation of core genes, minimal fragmentation, and a reliable structural organization.

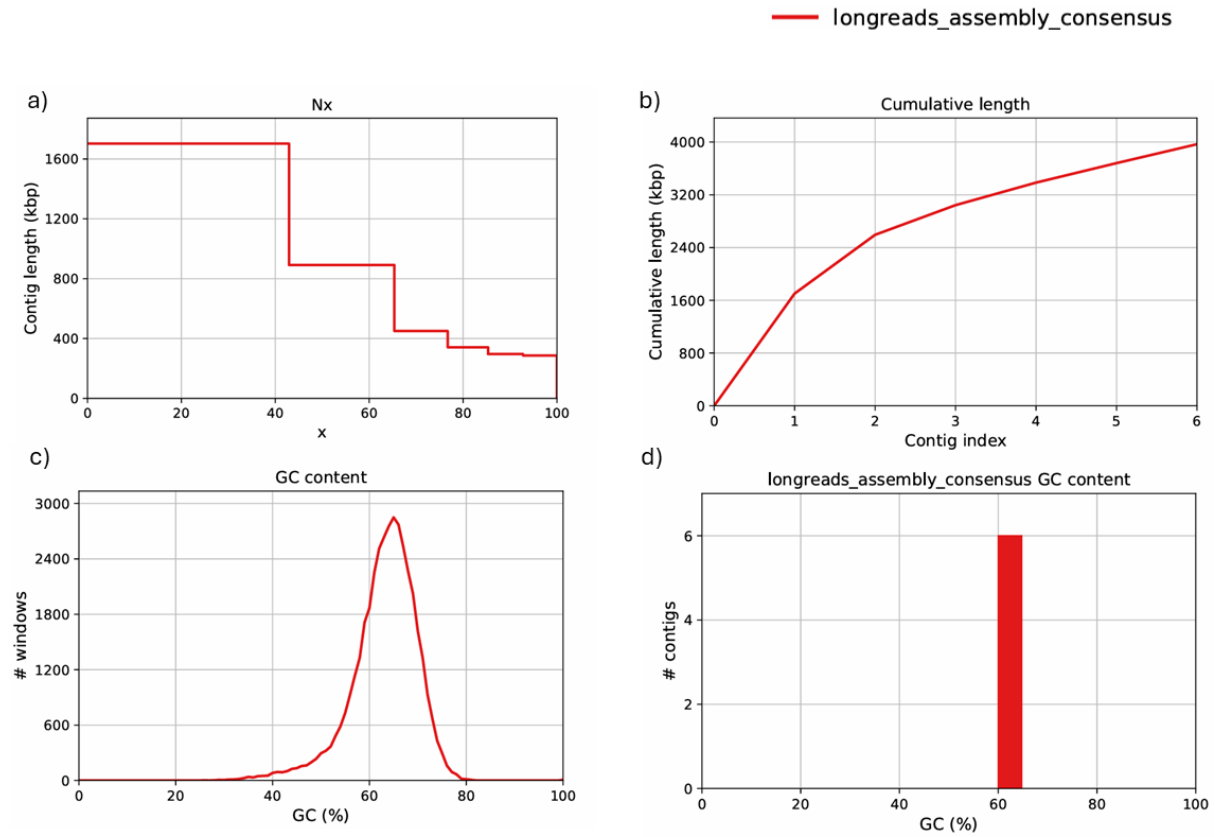


Figure 2: Evaluation of Quality of the genome assembly

a) An Nx plot illustrating assembly continuity, demonstrating that the genome assembly is of high quality, with substantial coverage contributed by large contigs. The high Nx values suggest a well-assembled genome with minimal fragmentation.. b) A cumulative contig length plot, representing the total length of assembled contigs as a function of the contig index. The presence of a steep initial slope, which subsequently plateaus, indicates that a small number of long contigs account for the majority of the genome, a characteristic of a well-assembled, contiguous genome with minimal fragmentation. C) A GC content distribution plot, depicting the variation in GC content across different genomic windows. The data reveal a predominant and consistent GC content around 60%, suggesting a relatively uniform nucleotide composition across the genome with no major deviations. d) A GC content frequency distribution plot, showing the peak GC content across different contigs within the assembled genome. The observed peak, centered slightly above 60%, suggests a uniform and stable GC composition across the genome, reinforcing the structural integrity of the assembly.

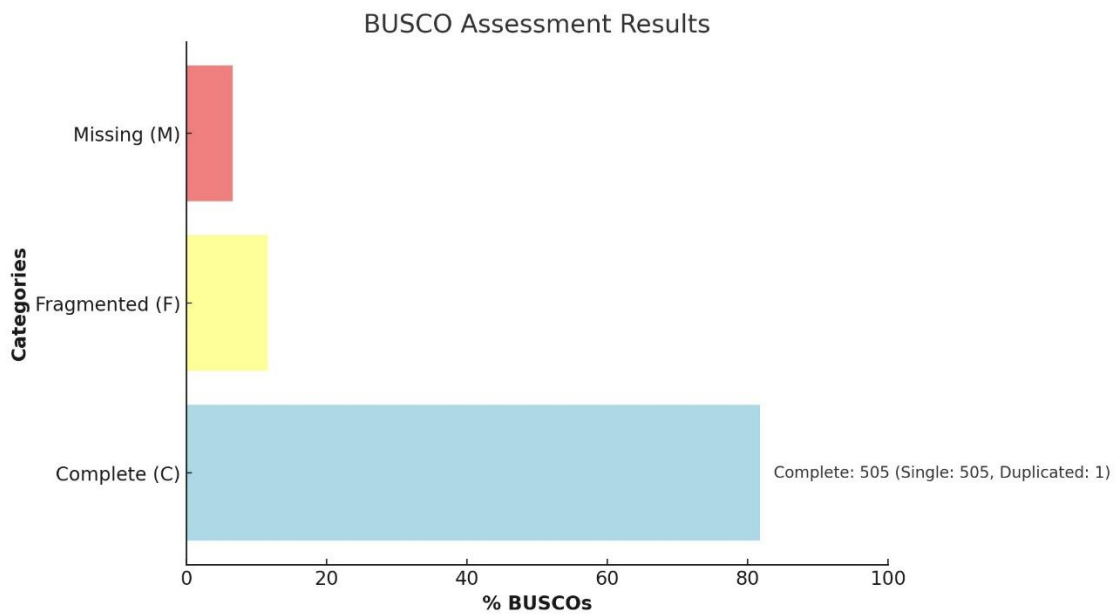


Figure 3: BUSCO assessment results

4.2. Taxonomical classification

The taxonomic classification of the assembled genome was performed reclassified the bacterium as *Halomonas elongata*. This result contradicted the previous taxonomic identification based on 16S rRNA sequencing, which had originally assigned the strain as *H. caseinilytica 10SCRN4D*

Figure 4 provides the detailed output of the taxonomic assignment, where *Halomonas elongata* received the highest confidence score, indicating the strongest association based on sequence similarity and classification parameters.

13.48	1248	1248	U	0	unclassified
86.52	8012	0		1	root
86.49	8009	2	1	131567	cellular organisms
86.45	8005	15		2	Bacteria
86.03	7966	22	P	1224	Proteobacteria
84.94	7865	26	C	1236	Gammaproteobacteria
83.46	7728	1	O	135619	Oceanospirillales
83.38	7721	3	F	28256	Halomonadaceae
83.14	7699	45	G	2745	Halomonas
79.01	7316	0	S	2746	Halomonas elongata
79.01	7316	7316	S1	768066	Halomonas elongata DSM 2581
1.48	137	22	G1	2609666	unclassified Halomonas
0.23	21	21	S	1761789	Halomonas sp. hl-4
0.23	21	21	S	2306583	Halomonas sp. JS92-SW72
0.21	19	19	S	2562282	Halomonas sp. Y2R2
0.17	16	16	S	2587849	Halomonas sp. THAF12
0.14	13	13	S	2576841	Halomonas sp. PA16-9
0.10	9	9	S	2497861	Halomonas sp. BC-M4-5
0.05	5	5	S	2136172	Halomonas sp. SF2003
0.04	4	4	S	1504981	Halomonas sp. K0116
0.03	3	3	S	1346287	Halomonas sp. A3H3
0.02	2	2	S	1666906	Halomonas sp. HL-93
0.01	1	1	S	2661553	Halomonas sp. AM6
0.01	1	1	S	2587844	Halomonas sp. THAF5a
1.47	136	136	S	1178482	Halomonas huangheensis

Figure 4: kraken2 taxonomical classification

Here the first column is the percentage of the fragments that are rooted to the taxon, the second column is the number of fragments covered, the third column indicates the number of fragments that re directly assigned to the said taxon, the fourth column contains rank codes where S-species, G-Genus, F-Family, O-Order, C-Class, P-Phylum, D-Domain, R-Root, U-Unclassified, the fourth column consists NCBI taxonomic IDs and the last column has the Scientific Names

4.3. Gene Prediction and Functional Annotation

Genome annotation using PROKKA identified a total of 4,308 genes, which were classified into distinct functional categories. Among these, 4,227 genes were annotated as coding sequences (CDS), while 12 genes were associated with ribosomal RNA (rRNA), 68 genes with transfer RNA (tRNA), and one gene with transfer-messenger RNA (tmRNA). Additionally, a substantial proportion of the CDS (1,659 genes) were designated as hypothetical proteins, indicating sequences with unknown or uncharacterized functions.

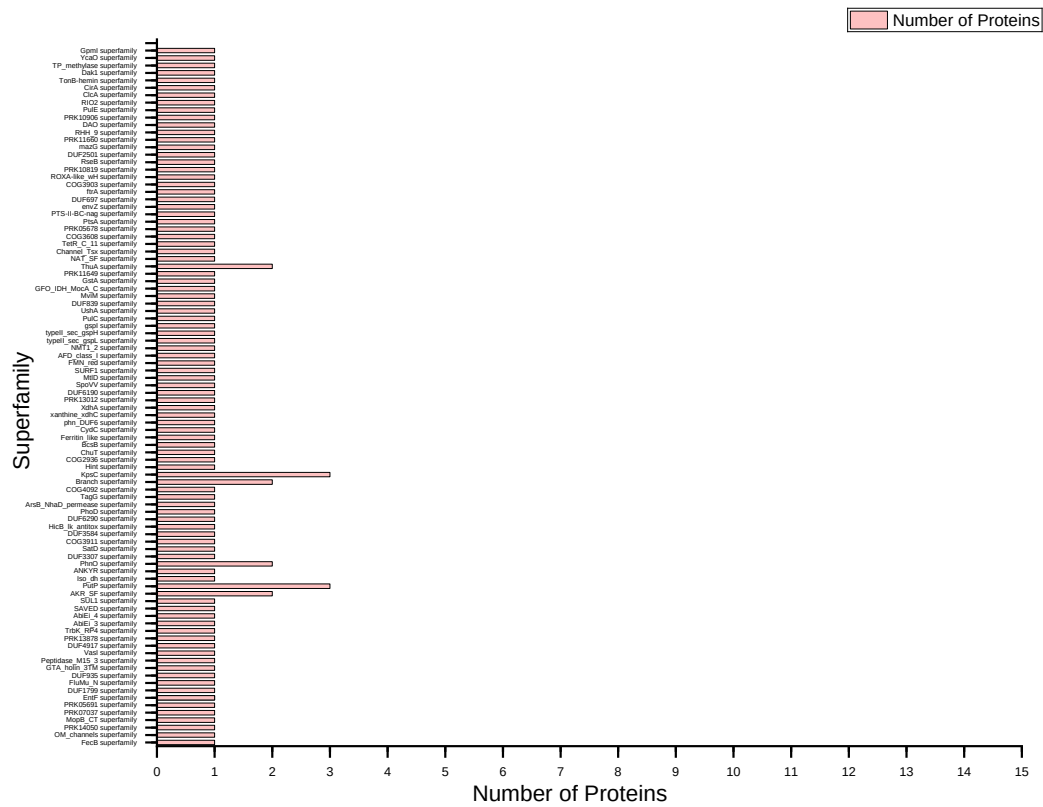


Figure 8: Proteins associated with superfamilies

To cross-validate the gene predictions, PRODIGAL and GeneMarkS2 were also employed. PRODIGAL predicted 4,234 genes, of which 3,785 were successfully annotated using EggNOG-mapper. Similarly, GeneMarkS2 identified 4,280 genes, with 3,861 of them annotated through EggNOG-mapper. Figure 9 shows the alignment result of the three prediction tools for the gene Glutaredoxin 1.

Table 2: Comparison of the Gene prediction and annotation results

Gene Prediction and Annotation Tool	Total length	No. of genes predicted	No. of genes annotated	No. of CDS	No. of rRNAs	No. of tRNAs	No. of Hypothetical Proteins	No. of predicted genes not annotated
PROKKA	1152142	4308	4308	4227	12	68	1659	-
PRODIGAL+	1157291	4234	3785	3657	-	-	128	449
EggNOG-mapper		4280	3861	3732	-	-	129	419
GenemarkS2+		4280	3861	3732	-	-	129	419
EggNOG-mapper	1161458							

Figure 9: An example of the alignment of the gene prediction using three different prediction tools. Here, the gene prediction for Glutaredoxin 1 is shown

GO analysis was conducted using InterProScan and Blast2GO, classifying genes based on their Biological, Cellular, and Molecular functions (Figure 10). A total of 100 proteins were assigned to GO:0006805, which corresponds to xenobiotic degradation, suggesting a potential role in detoxification processes. Notably, no proteins were mapped to GO:0019439, which is associated with aromatic compound catabolism, indicating an absence of direct functional annotations related to this metabolic pathway.

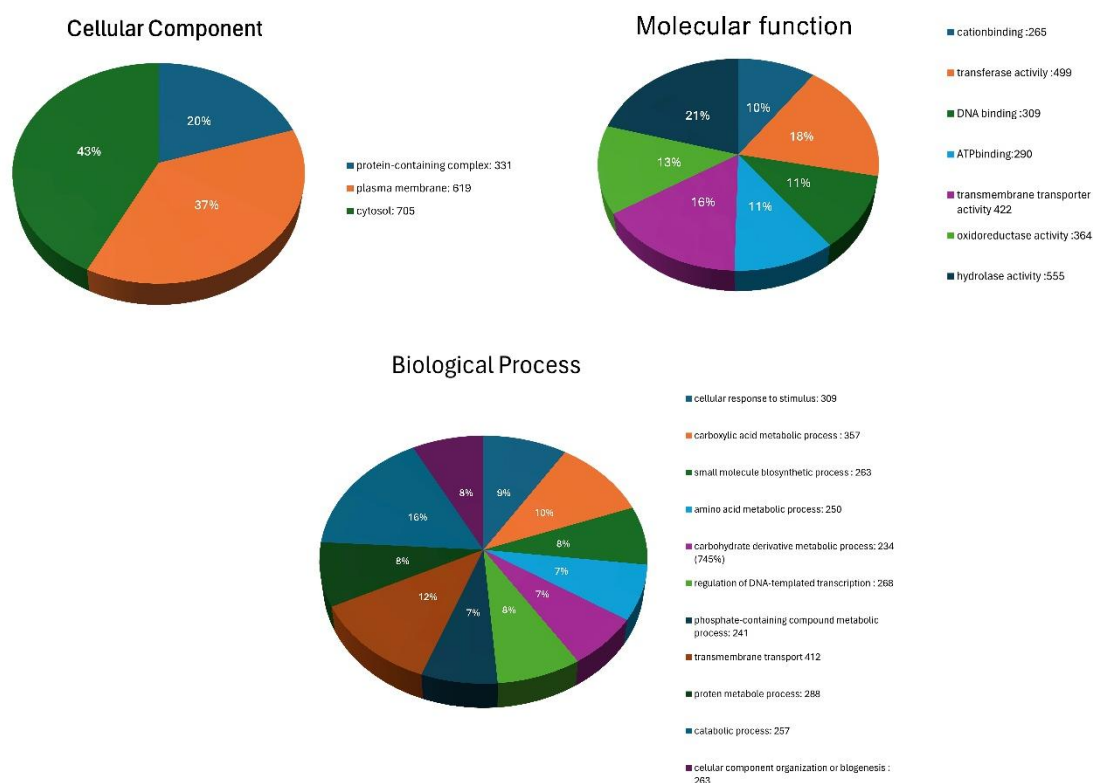


Figure 10: Gene ontology distribution in the genome

4.4. Identification of Genomic features

Operon Mapper identified a total of 2,013 operons within the genome. Among these, at least nine operons were specifically associated with aromatic compound degradation, suggesting potential metabolic pathways involved in biodegradation or detoxification of aromatic compounds.

In addition to operon identification, CRISPR array analysis detected two CRISPR loci, located within the utg0000011 and utg0000051 contigs (Table 3). However, no *cas* genes were identified, indicating that while CRISPR elements are present, the genome lacks the necessary Cas proteins required for an active CRISPR-mediated adaptive immune system.

Table 3: CRISPR sites

Locus	Start	End	Length	orientation	Consensus Repeat sequence
utg000001 l_1	8202 00	8202 88	88	Forward	TGATTACAAGTCAGCTGCTCTACCA
utg000005 l_1	2684 55	2685 74	119	Unknown	AGGACCAGTGCCCGGGTACCCCGG CCGGTG

A total of 436 repeat regions were identified within the genome, predominantly consisting of simple sequence repeats (SSRs). Additionally, other repeat elements, including Long Interspersed Nuclear Elements (LINEs), Short Interspersed Nuclear Elements (SINEs), rRNA repeats and tRNA repeats, were also detected.

Further structural analysis revealed the presence of 1,288 palindromic sequences, 121 mobile genetic elements, and 47 genes associated HGT, indicating potential mechanisms for genomic plasticity and adaptation. antiSMASH analysis identified three secondary metabolite biosynthetic gene clusters:

1. Ectoine biosynthetic gene cluster – associated with osmoprotection and stress adaptation
2. Non-Ribosomal Peptide Synthetase (NRPS) / NRPS-metallophore cluster – potentially involved in metal chelation and microbial interactions
3. Ribosomally synthesized and post-translationally modified peptides (RiPP-like protein cluster) – suggesting bioactive compound synthesis

Additionally, the ectoine hydroxylase enzyme, responsible for the conversion of ectoine to 5-hydroxyectoine, was detected. This enzyme is particularly noteworthy, as it is not universally present in ectoine-producing bacteria (Karna Radjasa et al.,2023). Figure 11 provides a graphical representation of the genome annotation, illustrating the identified features, including repeat elements, mobile genetic elements, and secondary metabolite clusters.

4.5. Functional and Pathway analysis

Annotation using RAST platform revealed that only 32% of the genome could be assigned to known subsystem categories, indicating that a significant portion of the genetic content remains functionally unclassified. The RASTk pipeline identified a total of 4,393 CDS, of which 1,398 genes were mapped to one or more subsystems within the database.

To evaluate the biodegradative potential of the organism, genes involved in aromatic compound metabolism were specifically examined. A total of 30 genes were identified under this category, suggesting a potential role in the degradation of aromatic compounds. Additionally, within the miscellaneous subsystems, two genes were classified under aromatic dioxygenase metabolism, which is critical for the oxidative cleavage of aromatic rings (Table 4).

Other notable subsystems identified include resistance to antibiotics and toxic compounds, which suggests the presence of mechanisms involved in antimicrobial resistance and detoxification. Additionally, genes associated with invasion and intracellular resistance were detected, indicating potential strategies for virulence or host adaptation. The genome also harbors genes related to prophage and phage packaging machinery, suggesting the presence of prophage elements that may contribute to genetic variability and horizontal gene transfer, further enhancing the organism's adaptability.

Table 4: RAST Annotation corresponding to aromatic compound metabolism

Category	Subcategory	Subsystem	Role
Metabolism of Aromatic Compounds	Peripheral pathways for catabolism of aromatic compounds	Biphenyl Degradation	biphenyl-2,3-diol 1,2-dioxygenase III-related protein
		Benzoate degradation	Benzoate transport protein
	Metabolism of central aromatic intermediates	Quinate degradation	3-dehydroquinatase II (EC 4.2.1.10)
		Catechol branch of beta-ketoadipate pathway	Muconate cycloisomerase (EC 5.5.1.1)
			Succinyl-CoA:3-ketoacid-coenzyme A transferase subunit A (EC 2.8.3.5)
			Succinyl-CoA:3-ketoacid-coenzyme A transferase subunit B (EC 2.8.3.5)
		Salicylate and gentisate catabolism	Fumarylacetoacetase (EC 3.7.1.2)
			Fumarylacetoacetate hydrolase family protein
			Maleylacetoacetate isomerase (EC 5.2.1.2)
		4-Hydroxyphenylacetic	3,4-dihydroxyphenylacetate 2,3-dioxygenase (EC 1.13.11.15)

acid catabolic pathway	Homoprotocatechuate degradative operon repressor
	2-hydroxyhepta-2,4-diene-1,7-dioate isomerase (EC 5.3.3.-)
	5-carboxymethyl-2-hydroxymuconate semialdehyde dehydrogenase (EC 1.2.1.60)
	Transcriptional activator of 4-hydroxyphenylacetate 3-monooxygenase operon, XylS/AraC family
	5-carboxymethyl-2-oxo-hex-3-ene-1,7-dioate decarboxylase (EC 4.1.1.68)
	2-oxo-hepta-3-ene-1,7-dioic acid hydratase (EC 4.2.-.-)
	5-carboxymethyl-2-hydroxymuconate delta-isomerase (EC 5.3.3.10)
Homogentisate pathway of aromatic compound degradation	Maleylacetoacetate isomerase (EC 5.2.1.2)
	Homogentisate 1,2-dioxygenase (EC 1.13.11.5)

		Fumarylacetoacetase (EC 3.7.1.2)
		4-hydroxyphenylpyruvate dioxygenase (EC 1.13.11.27)
		Transcriptional regulator, IclR family
Central meta-cleavage pathway of aromatic compound degradation		5-carboxymethyl-2-hydroxymuconate semialdehyde dehydrogenase (EC 1.2.1.60)
		3,4-dihydroxyphenylacetate 2,3-dioxygenase (EC 1.13.11.15)
		2-oxo-hepta-3-ene-1,7-dioic acid hydratase (EC 4.2.-.-)
		5-carboxymethyl-2-hydroxymuconate delta-isomerase (EC 5.3.3.10)
Metabolism of Aromatic Compounds - no subcategory	Aromatic Amin Catabolism	Phenylacetaldehyde dehydrogenase (EC 1.2.1.39)
		3,4-dihydroxyphenylacetate 2,3-dioxygenase (EC 1.13.11.15)
	Gentisate degradation	Fumarylacetoacetate hydrolase family protein

Pathway analysis using InterProScan identified a total of 266 KEGG pathways, with 1,866 sequences associated with one or more metabolic or functional pathways. Additionally, KASS-KEGG was used to generate a KO list, facilitating pathway mapping. In total, 2,101 genes were successfully annotated with KO numbers, providing insight into the diverse biochemical and metabolic pathways present in the bacterial genome.

KEGG pathway mapping enabled the identification of various pathways aligning with the KEGG database, offering a comprehensive overview of the organism's metabolic capabilities. Notably, within the xenobiotic degradation category, 13 partial KEGG pathways were detected. The presence of incomplete pathways suggests the possibility of alternative or non-canonical degradation routes, which may not be fully represented within the standard KEGG modules. This observation highlights the potential for novel or species-specific metabolic adaptations in xenobiotic processing.

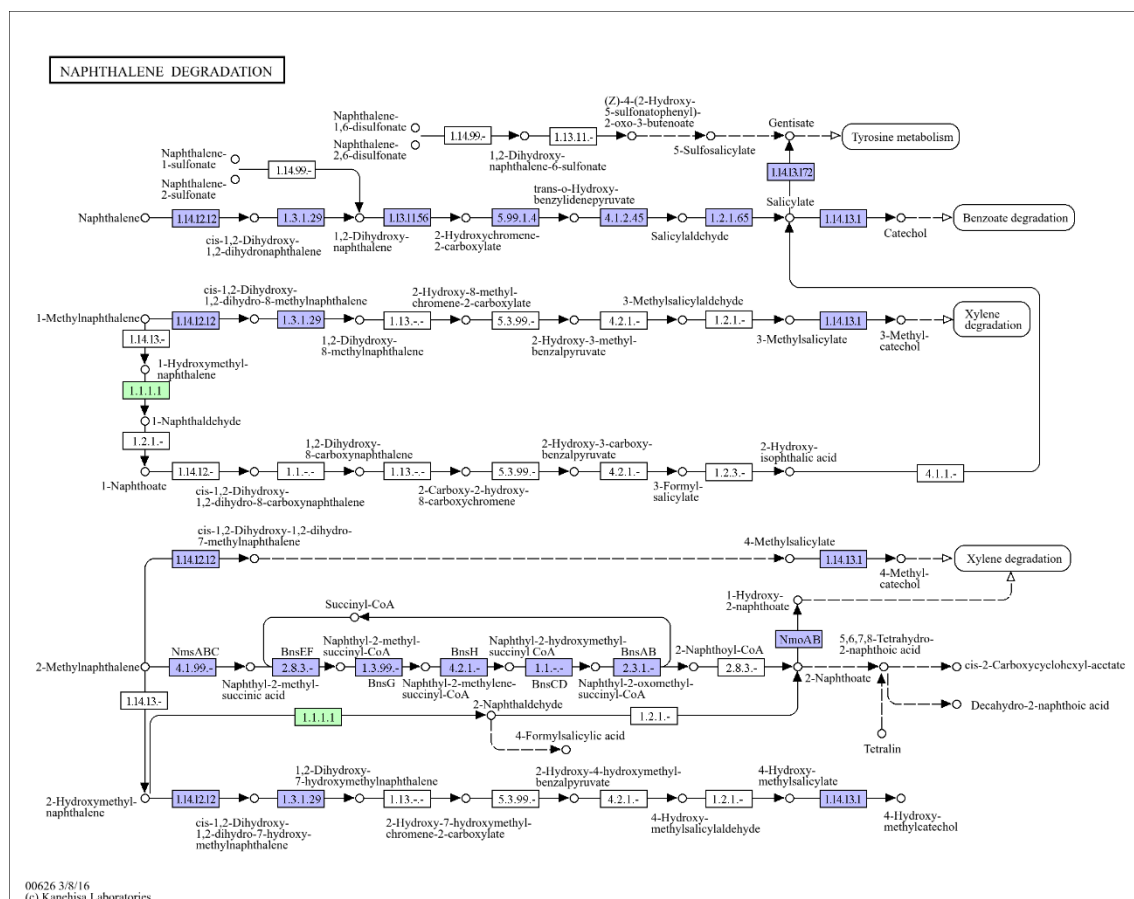


Figure 12: Naphthalene degradation pathway as predicted on KEGG database.

The green boxes indicate the genes that were present in our gene list while the blue boxes indicate those that were absent but should have been present. Here the enzyme with E.C. no 1.1.1.1 found in our genome is Alcohol dehydrogenase

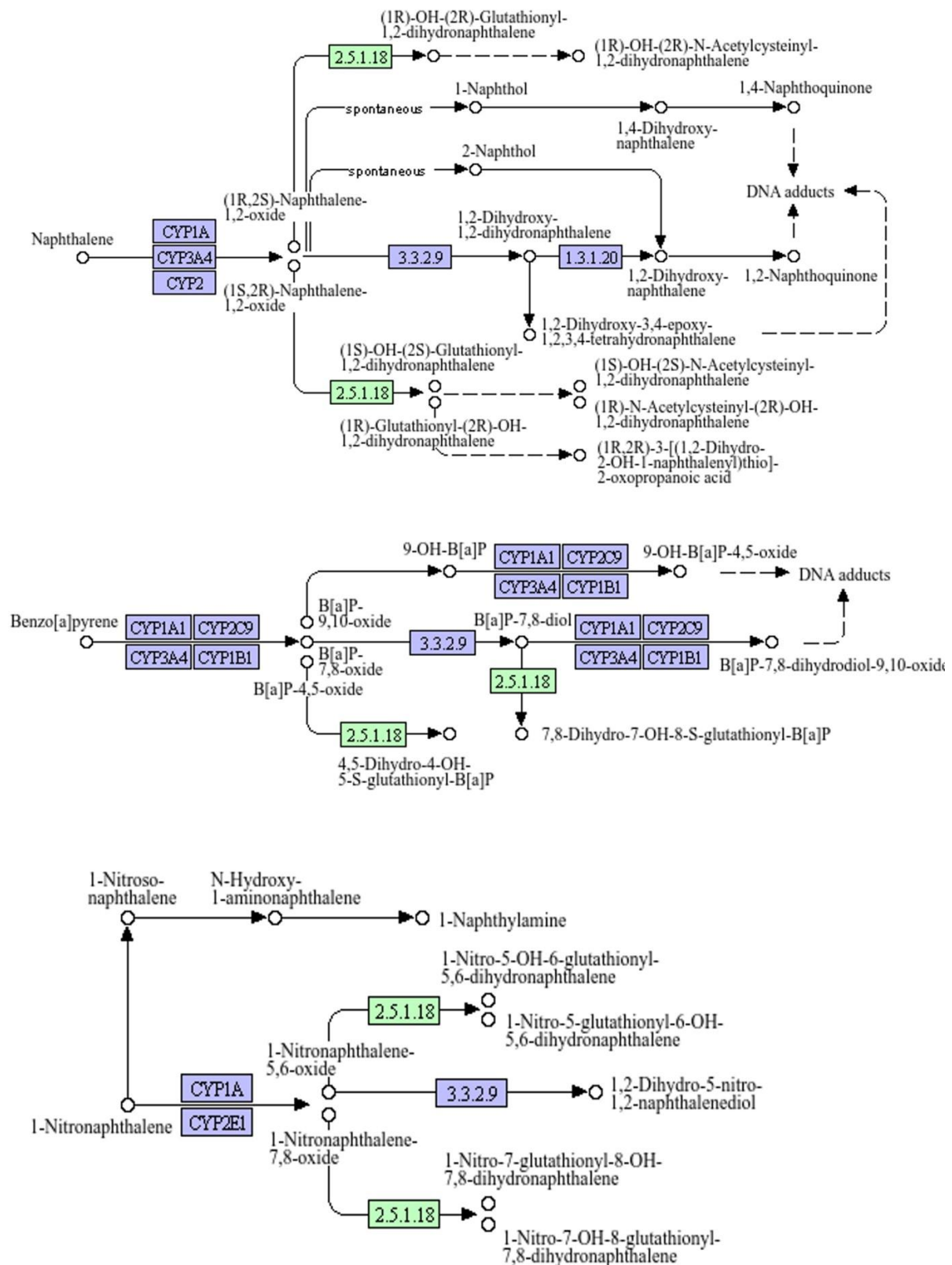


Figure 13: Pathway mapping of metabolism of xenobiotics by cytochrome P450.

The green boxes indicate the genes that were present in our gene list while the blue boxes indicate those that were absent but should have been present. glutathione S-transferase is the gene marked in green with an E.C. no. of 2.5.1.18

CHAPTER 5:

DISCUSSION

The reclassification of our strain from *Halomonas caseinilytica* to *Halomonas elongata* as we transitioned from 16S rRNA sequencing to WGS illustrates the dynamic nature of bacterial taxonomy and the limitations of 16S rRNA based identification methods (Hobmeier, Cantone, et al., 2022a). While 16S rRNA sequencing has long been a staple for bacterial identification, its resolution is often insufficient for distinguishing closely related species, especially within genetically complex genera like *Halomonas* (Jeong et al., 2021; Lafi et al., 2016). This genus exhibits significant genetic similarity among its species, complicating accurate classification when relying solely on the 16S rRNA gene (Hobmeier, Cantone, et al., 2022a). In literature, there have been many instances of reclassification of bacteria, for example, *Elizabethkingia miricola* strain was reclassified as *E. bruuniana* based on WGS results (Liang et al., 2019). The reclassification of our strain from *Halomonas caseinilytica* to *Halomonas elongata* as we transitioned from 16S rRNA sequencing to WGS illustrates the dynamic nature of bacterial taxonomy and the limitations of 16S rRNA based identification methods (Hobmeier, Cantone, et al., 2022a). While 16S rRNA sequencing has long been a staple for bacterial identification, its resolution is often insufficient for distinguishing closely related species, especially within genetically complex genera like *Halomonas* (Jeong et al., 2021; Lafi et al., 2016). This genus exhibits significant genetic similarity among its species, complicating accurate classification when relying solely on the 16S rRNA gene (Hobmeier, Cantone, et al., 2022a). In literature, there have been many instances of reclassification of bacteria, for example, *Elizabethkingia miricola* strain was reclassified as *E. bruuniana* based on WGS results (Liang et al., 2019). Unlike 16S rRNA sequencing, WGS offers a more comprehensive genetic landscape, facilitating

precise species identification through comparative genomic analysis (Li et al., 2020). The discrepancy observed in this study reinforces the advantages of WGS in identifying subtle genomic differences that are essential for accurate taxonomic classification, further demonstrating its superiority over traditional 16S rRNA-based methods.

The integration of multiple gene prediction and annotation tools provided a detailed and comprehensive genomic profile of the *Halomonas elongata* strain. PROKKA, PRODIGAL, and GeneMarkS2 predicted a similar number of genes (4,308, 4,234, and 4,280, respectively), suggesting a consistent gene density across different prediction algorithms. This consistency reinforces the credibility of the gene count and supports the robustness of the genome assembly. PROKKA's annotation revealed a genome structure characterized by a high proportion of protein-coding sequences (4,227 CDS) alongside essential RNA genes, including 12 rRNA, 68 tRNA, and 1 tmRNA. The presence of these RNA elements indicates a fully functional translational machinery, essential for cellular function and protein synthesis (Lafi et al., 2016). However, a notable limitation in the initial annotation was the high number of hypothetical proteins, with 1,659 out of 4,227 CDS lacking functional characterization. This reflects the current gaps in bacterial gene function knowledge, particularly in less-characterized genera like *Halomonas*.

To refine functional predictions, these hypothetical proteins were further analyzed using CDD, reducing the number of truly uncharacterized proteins from 1,659 to 526. This substantial reduction underscores the importance of leveraging multiple annotation tools and databases to improve functional assignments. The remaining 526 hypothetical proteins, which lack identifiable domains or superfamily associations, represent potential candidates for future experimental characterization. These genes may be unique to *Halomonas* or even strain-specific adaptations, possibly linked to the bacterium's ecological niche. In our study, their presence

suggests a potential role in the biodegradation of coronene, an aspect that warrants further investigation (Gasperotti et al., 2018; Jeong et al., 2021).

GO analysis using InterProScan and Blast2GO provided a broad functional overview of the genome, offering insights into its metabolic capabilities. A particularly noteworthy finding was the identification of 100 proteins assigned to GO:0006805, which corresponds to xenobiotic metabolic processes. This suggests a potential role for the strain in polycyclic aromatic hydrocarbon (PAH) degradation, aligning with its hypothesized bioremediation capabilities. However, an unexpected contradiction arose from the absence of proteins categorized under GO:0019439 (aromatic compound catabolism), which is directly associated with aromatic hydrocarbon degradation. This discrepancy can be interpreted in several ways. One possibility is that the strain lacks the specific pathways required for aromatic compound breakdown, indicating a metabolic specialization in alternative xenobiotic degradation processes (Hobmeier, Cantone, et al., 2022a). Another explanation could be limitations in current GO annotations, where essential genes for aromatic compound degradation may not yet be well-represented in existing functional databases. Lastly, the strain may employ alternative or novel pathways for degrading aromatic compounds, which are not yet captured by current GO terms. This underscores the need for further experimental validation and pathway characterization to elucidate the precise mechanisms employed by *Halomonas elongata* in xenobiotic degradation.

Combining the results from Operon Mapper, CRISPR analysis, repeat region identification, and antiSMASH provides a comprehensive perspective on the genomic structure and functional capabilities of the *Halomonas elongata* strain. Notably, the identification of nine operons associated with aromatic compound degradation aligns with the GO findings, which indicated potential xenobiotic degradation activity. This suggests that the strain may harbor specialized metabolic pathways for the breakdown of aromatic compounds, reinforcing its potential role in bioremediation. An intriguing observation was the detection of CRISPR sites in the genome

without associated *cas* genes. This raises two possibilities: (1) the identified CRISPR arrays may be non-functional, lacking the necessary Cas proteins for adaptive immunity, or (2) *cas* genes may be present but were not detected due to annotation limitations in the current dataset. Despite this uncertainty, the presence of CRISPR loci is still noteworthy, as it provides opportunities for genome editing and functional studies in future research (Takeuchi et al., 2012).

The identification of secondary metabolite biosynthetic gene clusters provides key insights into the ecological adaptability and biotechnological potential of the strain. Notably, the presence of an ectoine biosynthesis cluster aligns with the halophilic nature of the strain, as ectoine serves as an osmoprotectant, enabling bacterial survival in high-salinity environments (Hobmeier, Oppermann, et al., 2022; Schwibbert et al., 2010). Additionally, the detection of Non-Ribosomal Peptide Synthetase (NRPS) clusters, including one encoding a metallophore, suggests the strain's ability to produce complex secondary metabolites involved in metal acquisition or ecological interactions. Such mechanisms play a crucial role in bioremediation, particularly in the sequestration and detoxification of heavy metals from contaminated environments. The presence of a Ribosomally Synthesized and Post-Translationally Modified Peptide (RiPP-like) cluster draws attention to the strain's potential for bioactive peptide production, which could be explored for antimicrobial applications (Zytnick et al., 2024). Most notably, the identification of four genes involved in ectoine biosynthesis positions this strain as an important candidate for biotechnological research, particularly in the development of cosmetics and pharmaceuticals, where ectoine is valued for its protective and stabilizing properties in skincare and medical formulations (Hobmeier, Cantone, et al., 2022b; Hobmeier, Oppermann, et al., 2022).

The results from RAST and KEGG pathway analysis provide valuable insights into the functional capabilities and metabolic potential of the strain. The relatively low percentage (32%) of genes assigned to RAST subsystems suggests that a substantial portion of the genome consists of unique or poorly characterized genes, a common occurrence in less-studied bacterial species

(Aziz et al., 2008). This backs the idea to further explore novel genes in the organism. On the other hand, the identification of 13 partial KEGG pathways related to xenobiotic degradation is of interest. The absence of complete pathways could be due to the use of unique pathways or modified pathways by the species which are not defined in KEGG modules or as seen, only 2101 genes were annotated with KO numbers, indicating that there could be relevant genes in the non-annotated genes but due to slight variations they are not categorized under the KEGG modules (Vera et al., 2022). The existence of discrete pathways has been experimentally proven for naphthalene degrading bacteria via a liquid chromatography mass spectrometry study by *Muccee et al.* in 2021 (Muccee et al., 2021). This highlights the need for experimental validation of computational predictions of the pathways identified in our study. Future investigations should focus on biochemical and transcriptomic analyses to confirm the functionality of the predicted xenobiotic degradation mechanisms in *Halomonas elongata*.

CHAPTER 6:

CONCLUSION

This study provides a comprehensive genomic characterization of *Halomonas elongata* (previously misidentified as *H. caseinilytica*), elucidating its capacity to degrade coronene, HMW-PAHs, under saline conditions. The application of WGS and advanced bioinformatics analyses facilitated the identification of specific genes and metabolic pathways involved in PAH degradation, accentuating the strain's potential for bioremediation in marine and hypersaline environments. Key findings from this study include the identification of xenobiotic degradation-associated genes, which support the metabolic potential of *Halomonas elongata* for PAH breakdown. Additionally, the presence of secondary metabolite biosynthetic clusters suggests the strain's capability for bioactive compound production, which may play a role in ecological interactions and stress adaptation. Furthermore, genomic features associated with environmental adaptability were identified, including CRISPR elements and HGT mechanisms, both of which contribute to genome plasticity and resilience in extreme habitats, reinforcing the organism's potential for survival and functionality in saline and contaminated environments.

While the genomic insights presented here provide strong evidence for the metabolic potential of *H. elongata*, further experimental validation is required to confirm the functionality of the predicted pathways, particularly those involved in coronene degradation. Future research integrating transcriptomics, proteomics, and metabolomics will be essential to characterize gene expression dynamics, enzyme activity, and metabolic fluxes under relevant environmental conditions. The findings of this study establish a scientific framework for the development of sustainable biotechnological strategies aimed at mitigating PAH contamination in saline and extreme environments.

CHAPTER 7:

FUTURE PROSPECT

The findings of this study provide a strong foundation for future research aimed at further elucidating the genetic, biochemical, and physiological mechanisms underlying the coronene degradation potential of *Halomonas elongata*. While whole-genome sequencing and bioinformatics analyses have identified key genes and pathways involved in xenobiotic metabolism, secondary metabolite biosynthesis, and environmental adaptation, additional experimental validation is required to confirm their functional roles.

Future research should focus on multi-omics approaches, including transcriptomics, proteomics, and metabolomics to gain deeper insights into gene expression dynamics, enzyme functionality, and metabolic fluxes under environmentally relevant conditions. The application of gene knockout and overexpression studies could help validate the involvement of specific genes in coronene degradation and stress adaptation mechanisms.

Moreover, the biotechnological potential of *H. elongata* in bioremediation applications should be explored through bioreactor studies, optimizing conditions for PAH degradation in saline and contaminated environments. The presence of CRISPR elements in the genome, despite the absence of Cas proteins, also opens avenues for genome engineering, allowing researchers to enhance the strain's metabolic capabilities for targeted bioremediation efforts. Additionally, the ectoine biosynthesis cluster identified in this study suggests a potential application of *H. elongata* in industrial biotechnology, particularly in the production of osmoprotectants for use in cosmetics, pharmaceuticals, and agriculture. Further studies are needed to optimize ectoine production and assess its commercial viability.

Finally, comparative genomic and evolutionary studies across *Halomonas* species could provide insights into strain-specific adaptations, facilitating the identification of novel genes and pathways that contribute to hydrocarbon degradation and environmental resilience. The integration of synthetic biology and metabolic engineering approaches could further enhance the strain's utility in bioremediation and bio-based product development.

REFERENCES

- Adebusoye, S. A., Ilori, M. O., Amund, O. O., Teniola, O. D., & Olatope, S. O. (2007). Microbial degradation of petroleum hydrocarbons in a polluted tropical stream. *World Journal of Microbiology and Biotechnology*, 23(8), 1149–1159. <https://doi.org/10.1007/S11274-007-9345-3/METRICS>
- Aitken, M. D., Stringfellow, W. T., Nagel, R. D., Kazunga, C., & Chen, S. H. (2011). Characteristics of phenanthrene-degrading bacteria isolated from soils contaminated with polycyclic aromatic hydrocarbons. *Https://Doi.Org/10.1139/W98-065*, 44(8), 743–752. <https://doi.org/10.1139/W98-065>
- Al Farraj, D. A., Alkufeidy, R. M., Alkubaisi, N. A., & Alshammari, M. K. (2021). Polynuclear aromatic anthracene biodegradation by psychrophilic *Sphingomonas* sp., cultivated with tween-80. *Chemosphere*, 263, 128115. <https://doi.org/10.1016/J.CHEMOSPHERE.2020.128115>
- Alcock, B. P., Huynh, W., Chalil, R., Smith, K. W., Raphenya, A. R., Wlodarski, M. A., Edalatmand, A., Petkau, A., Syed, S. A., Tsang, K. K., Baker, S. J. C., Dave, M., Mccarthy, M. C., Mukiri, K. M., Nasir, J. A., Golbon, B., Imtiaz, H., Jiang, X., Kaur, K., ... Mcarthur, A. G. (2023). CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Research*, 51(D1), D690–D699. <https://doi.org/10.1093/NAR/GKAC920>
- Alfaify, A. M., Mir, M. A., & Alrumman, S. A. (2022). *Klebsiella oxytoca*: an efficient pyrene-degrading bacterial strain isolated from petroleum-contaminated soil. *Archives of Microbiology*, 204(5). <https://doi.org/10.1007/S00203-022-02850-9>

- Al-Thukair, A. A., & Malik, K. (2016). Pyrene metabolism by the novel bacterial strains *Burkholderia fungorum* (T3A13001) and *Caulobacter* sp (T2A12002) isolated from an oil-polluted site in the Arabian Gulf. *International Biodeterioration & Biodegradation*, 110, 32–37. <https://doi.org/10.1016/J.IBIOD.2016.02.005>
- Al-Thukair, A. A., Malik, K., & Nzila, A. (2020). Biodegradation of selected hydrocarbons by novel bacterial strains isolated from contaminated Arabian Gulf sediment. *Scientific Reports* 2020 10:1, 10(1), 1–12. <https://doi.org/10.1038/s41598-020-78733-0>
- Alvarez, P. J. J., & Vogel, T. M. (1991). Substrate interactions of benzene, toluene, and para-xylene during microbial degradation by pure cultures and mixed culture aquifer slurries. *Applied and Environmental Microbiology*, 57(10), 2981. <https://doi.org/10.1128/AEM.57.10.2981-2985.1991>
- Amran, R. H., Jamal, M. T., Pugazhendhi, A., Al-Harbi, M., & Bowrji, S. (2022). Petroleum Hydrocarbon Degradation and Treatment of Automobile Service Station Wastewater by Halophilic Consortia Under Saline Conditions. *Nature Environment and Pollution Technology*, 21(4), 1629–1637. <https://doi.org/10.46488/NEPT.2022.V21I04.015>
- Amran, R. H., Jamal, M. T., Pugazhendhi, A., Al-Harbi, M., & Bowrji, S. (n.d.). *Nature Environment and Pollution Technology An International Quarterly Scientific Journal Open Access Journal Petroleum Hydrocarbon Degradation and Treatment of Automobile Service Station Wastewater by Halophilic Consortia Under Saline Conditions*. 21, 1629–1637. <https://doi.org/10.46488/NEPT.2022.v21i04.015>
- Amund, O. O., & Nwokoye, N. (1993). *Hydrocarbon Degradation Potentials of Yeast Isolates from a Polluted Lagoon*.
- Arulazhagan, P., Al-Shekri, K., Huda, Q., Godon, J. J., Basahi, J. M., & Jeyakumar, D. (2017). Biodegradation of polycyclic aromatic hydrocarbons by an acidophilic *Stenotrophomonas*

- maltophilia strain AJH1 isolated from a mineral mining site in Saudi Arabia. *Extremophiles*, 21(1), 163–174. <https://doi.org/10.1007/S00792-016-0892-0/METRICS>
- Atlas, R. M., & Bartha, R. (1992). *Hydrocarbon Biodegradation and Oil Spill Bioremediation* (pp. 287–338). Springer, Boston, MA. https://doi.org/10.1007/978-1-4684-7609-5_6
- Atlas, R. M., & Hazen, T. C. (2011). Oil biodegradation and bioremediation: A tale of the two worst spills in U.S. history. *Environmental Science and Technology*, 45(16), 6709–6715. https://doi.org/10.1021/ES2013227/ASSET/IMAGES/LARGE/ES-2011-013227_0002.JPEG
- Aziz, R. K., Bartels, D., Best, A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., ... Zagnitko, O. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics*, 9. <https://doi.org/10.1186/1471-2164-9-75>
- Bacosa, H. P., & Inoue, C. (2015). Polycyclic aromatic hydrocarbons (PAHs) biodegradation potential and diversity of microbial consortia enriched from tsunami sediments in Miyagi, Japan. *Journal of Hazardous Materials*, 283, 689–697. <https://doi.org/10.1016/J.JHAZMAT.2014.09.068>
- Barathi, S., & Vasudevan, N. (2001). Utilization of petroleum hydrocarbons by *Pseudomonas fluorescens* isolated from a petroleum-contaminated soil. *Environment International*, 26(5–6), 413–416. [https://doi.org/10.1016/S0160-4120\(01\)00021-6](https://doi.org/10.1016/S0160-4120(01)00021-6)
- Bedessem, M. E., Swoboda-Colberg, N. G., & Colberg, P. J. S. (1997). Naphthalene mineralization coupled to sulfate reduction in aquifer-derived enrichments. *FEMS Microbiology Letters*, 152(2), 213–218. <https://doi.org/10.1111/J.1574-6968.1997.TB10430.X>

- Blin, K., Shaw, S., Augustijn, H. E., Reitz, Z. L., Biermann, F., Alanjary, M., Fetter, A., Terlouw, B. R., Metcalf, W. W., Helfrich, E. J. N., Van Wezel, G. P., Medema, M. H., & Weber, T. (2023). antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Research*, 51(W1), W46–W50. <https://doi.org/10.1093/NAR/GKAD344>
- Brooijmans, R. J. W., Pastink, M. I., & Siezen, R. J. (2009). Hydrocarbon-degrading bacteria: the oil-spill clean-up crew. *Microbial Biotechnology*, 2(6), 587. <https://doi.org/10.1111/J.1751-7915.2009.00151.X>
- Brown, C. L., Mullet, J., Hindi, F., Stoll, J. E., Gupta, S., Choi, M., Keenum, I., Vikesland, P., Pruden, A., & Zhang, L. (2022). mobileOG-db: a Manually Curated Database of Protein Families Mediating the Life Cycle of Bacterial Mobile Genetic Elements. *Applied and Environmental Microbiology*, 88(18). <https://doi.org/10.1128/AEM.00991-22>
- Budiyanto, F., Thukair, A., Al-Momani, M., Musa, M. M., & Nzila, A. (2018). Characterization of Halophilic Bacteria Capable of Efficiently Biodegrading the High-Molecular-Weight Polycyclic Aromatic Hydrocarbon Pyrene. *Https://Home.Liebertpub.Com/Ees*, 35(6), 616–626. <https://doi.org/10.1089/EES.2017.0244>
- Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., & Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution*, 38(12), 5825–5829. <https://doi.org/10.1093/MOLBEV/MSAB293>
- Cerniglia, C. E. (1992). Biodegradation of polycyclic aromatic hydrocarbons. *Biodegradation*, 3(2–3), 351–368. <https://doi.org/10.1007/bf00129093>

- Cerniglia, C. E., & Sutherland, J. B. (2010). Degradation of Polycyclic Aromatic Hydrocarbons by Fungi. *Handbook of Hydrocarbon and Lipid Microbiology*, 2079–2110. https://doi.org/10.1007/978-3-540-77587-4_151
- Coates, J. D., Woodward, J., Allen, J., Philp, P., & Lovley, D. R. (1997). Anaerobic degradation of polycyclic aromatic hydrocarbons and alkanes in petroleum-contaminated marine harbor sediments. *Applied and Environmental Microbiology*, 63(9), 3589–3593. <https://doi.org/10.1128/AEM.63.9.3589-3593.1997>
- Coulon, F., Chronopoulou, P. M., Fahy, A., Païssé, S., Goñi-Urriza, M., Peperzak, L., Alvarez, L. A., McKew, B. A., Brussaard, C. P. D., Underwood, G. J. C., Timmis, K. N., Duran, R., & McGenity, T. J. (2012). Central role of dynamic tidal biofilms dominated by aerobic hydrocarbonoclastic bacteria and diatoms in the biodegradation of hydrocarbons in coastal mudflats. *Applied and Environmental Microbiology*, 78(10), 3638–3648. https://doi.org/10.1128/AEM.00072-12/SUPPL_FILE/AEM00072-12_SUPPLEMENTAL.PDF
- Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Néron, B., Rocha, E. P. C., Vergnaud, G., Gautheret, D., & Pourcel, C. (2018). CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Research*, 46(W1), W246–W251. <https://doi.org/10.1093/NAR/GKY425>
- Das, N., & Chandran, P. (2011). Microbial Degradation of Petroleum Hydrocarbon Contaminants: An Overview. *Biotechnology Research International*, 2011, 1–13. <https://doi.org/10.4061/2011/941810>
- Daugulis, A. J., & McCracken, C. M. (2003). Microbial degradation of high and low molecular weight polyaromatic hydrocarbons in a two-phase partitioning bioreactor by two strains of

- Sphingomonas sp. *Biotechnology Letters*, 25(17), 1441–1444.
<https://doi.org/10.1023/A:1025007729355/METRICS>
- De Coster, W., & Rademakers, R. (2023). NanoPack2: population-scale evaluation of long-read sequencing data. *Bioinformatics*, 39(5).
<https://doi.org/10.1093/BIOINFORMATICS/BTAD311>
- Dos Santos, H. F., Cury, J. C., do Carmo, F. L., Dos Santos, A. L., Tiedje, J., van Elsas, J. D., Rosado, A. S., & Peixoto, R. S. (2011). Mangrove Bacterial Diversity and the Impact of Oil Contamination Revealed by Pyrosequencing: Bacterial Proxies for Oil Pollution. *PLOS ONE*, 6(3), e16943. <https://doi.org/10.1371/JOURNAL.PONE.0016943>
- Eberlein, C., Estelmann, S., Seifert, J., Von Bergen, M., Müller, M., Meckenstock, R. U., & Boll, M. (2013). Identification and characterization of 2-naphthoyl-coenzyme A reductase, the prototype of a novel class of dearomatizing reductases. *Molecular Microbiology*, 88(5), 1032–1039. <https://doi.org/10.1111/MMI.12238/SUPINFO>
- Eman, Mohamed, A. H., Yousef, N. M. H., Azza, & Farag, G. (2012). Isolation and molecular identification of polyaromatic hydrocarbons- utilizing bacteria from crude petroleum oil samples. *African Journal of Microbiology Research*, 6(49), 7479–7484.
<https://doi.org/10.5897/AJMR12.810>
- Fates, and E. N. R. C. (US) C. on O. in the S. I. (2003). Natural seepage of crude oil into the marine environment. In *Geo-Marine Letters* (Vol. 23, Issues 3–4, pp. 140–146). National Academies Press (US). <https://doi.org/10.1007/s00367-003-0135-0>
- Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences of the United States of America*, 117(17),

9451–9457.

https://doi.org/10.1073/PNAS.1921046117/SUPPL_FILE/PNAS.1921046117.SAPP.PDF

FOGHT, J. M., & WESTLAKE, D. W. S. (1987). BIODEGRADATION OF HYDROCARBONS IN FRESHWATER. In *Oil in Freshwater: Chemistry, Biology, Countermeasure Technology* (pp. 217–230). Pergamon. <https://doi.org/10.1016/b978-0-08-031862-2.50022-8>

Gasperotti, A. F., Revuelta, M. V., Studdert, C. A., & Herrera Seitz, M. K. (2018). Identification of two different chemosensory pathways in representatives of the genus *Halomonas*. *BMC Genomics*, 19(1), 266. <https://doi.org/10.1186/S12864-018-4655-4>

Ghosal, D., Ghosh, S., Dutta, T. K., & Ahn, Y. (2016). Current State of Knowledge in Microbial Degradation of Polycyclic Aromatic Hydrocarbons (PAHs): A Review. *Frontiers in Microbiology*, 7(AUG). <https://doi.org/10.3389/FMICB.2016.01369>

Grosser, R. J., Warshawsky, D., & Vestal, J. R. (1991). Indigenous and enhanced mineralization of pyrene, benzo[a]pyrene, and carbazole in soils. *Applied and Environmental Microbiology*, 57(12), 3462–3469. <https://doi.org/10.1128/AEM.57.12.3462-3469.1991>

Gupta, S., Pathak, B., & Fulekar, M. H. (2015). Molecular approaches for biodegradation of polycyclic aromatic hydrocarbon compounds: a review. *Reviews in Environmental Science and Biotechnology*, 14(2), 241–269. <https://doi.org/10.1007/S11157-014-9353-3/TABLES/9>

Haritash, A. K., & Kaushik, C. P. (2009). Biodegradation aspects of Polycyclic Aromatic Hydrocarbons (PAHs): A review. In *Journal of Hazardous Materials* (Vol. 169, Issues 1–3, pp. 1–15). Elsevier. <https://doi.org/10.1016/j.jhazmat.2009.03.137>

Hazen, T. C., Dubinsky, E. A., DeSantis, T. Z., Andersen, G. L., Piceno, Y. M., Singh, N., Jansson, J. K., Probst, A., Borglin, S. E., Fortney, J. L., Stringfellow, W. T., Bill, M., Conrad, M. E., Tom, L. M., Chavarria, K. L., Alusi, T. R., Lamendella, R., Joyner, D. C., Spier, C., ...

- Mason, O. U. (2010). Deep-sea oil plume enriches indigenous oil-degrading bacteria. *Science*, 330(6001), 204–208.
https://doi.org/10.1126/SCIENCE.1195979/SUPPL_FILE/PAPV2.PDF
- Heider, J. (2007). Adding handles to unhandy substrates: anaerobic hydrocarbon activation mechanisms. *Current Opinion in Chemical Biology*, 11(2), 188–194.
<https://doi.org/10.1016/J.CBPA.2007.02.027>
- Heitkamp, M. A., & Cerniglia, C. E. (1989). Polycyclic aromatic hydrocarbon degradation by a *Mycobacterium* sp. in microcosms containing sediment and water from a pristine ecosystem. *Applied and Environmental Microbiology*, 55(8), 1968–1973.
<https://doi.org/10.1128/AEM.55.8.1968-1973.1989>
- Heitkamp, M. A., Franklin, W., & Cerniglia, C. E. (1988). Microbial metabolism of polycyclic aromatic hydrocarbons: Isolation and characterization of a pyrene-degrading bacterium. *Applied and Environmental Microbiology*, 54(10), 2549–2555.
<https://doi.org/10.1128/AEM.54.10.2549-2555.1988>
- Hidalgo, K. J., Sierra-Garcia, I. N., Dellagnezze, B. M., & de Oliveira, V. M. (2020). Metagenomic Insights Into the Mechanisms for Biodegradation of Polycyclic Aromatic Hydrocarbons in the Oil Supply Chain. In *Frontiers in Microbiology* (Vol. 11, p. 561506). Frontiers Media S.A. <https://doi.org/10.3389/fmicb.2020.561506>
- Hobmeier, K., Cantone, M., Nguyen, Q. A., Pflüger-Grau, K., Kremling, A., Kunte, H. J., Pfeiffer, F., & Marin-Sanguino, A. (2022a). Adaptation to Varying Salinity in *Halomonas elongata*: Much More Than Ectoine Accumulation. *Frontiers in Microbiology*, 13, 846677.
<https://doi.org/10.3389/FMICB.2022.846677/BIBTEX>
- Hobmeier, K., Cantone, M., Nguyen, Q. A., Pflüger-Grau, K., Kremling, A., Kunte, H. J., Pfeiffer, F., & Marin-Sanguino, A. (2022b). Adaptation to Varying Salinity in *Halomonas*

- elongata: Much More Than Ectoine Accumulation. *Frontiers in Microbiology*, 13, 846677. <https://doi.org/10.3389/FMICB.2022.846677/FULL>
- Hobmeier, K., Oppermann, M., Stasinski, N., Kremling, A., Pflüger-Grau, K., Kunte, H. J., & Marin-Sanguino, A. (2022). Metabolic engineering of *Halomonas elongata*: Ectoine secretion is increased by demand and supply driven approaches. *Frontiers in Microbiology*, 13, 968983. <https://doi.org/10.3389/FMICB.2022.968983/BIBTEX>
- Hollaway, S. L., Faw, G. M., & Sizemore, R. K. (1980). The bacterial community composition of an active oil field in the northwestern Gulf of Mexico. *Marine Pollution Bulletin*, 11(6), 153–156. [https://doi.org/10.1016/0025-326X\(80\)90141-1](https://doi.org/10.1016/0025-326X(80)90141-1)
- Holliger, C., Gaspard, S., Glod, G., Heijman, C., Schumacher, W., Schwarzenbach, R. P., & Vazquez, F. (1997). Contaminated environments in the subsurface and bioremediation: organic contaminants. *FEMS Microbiology Reviews*, 20(3–4), 517–523. <https://doi.org/10.1111/J.1574-6976.1997.TB00334.X>
- Hosseini, H., Saadaoui, I., Moheimani, N., Al Saidi, M., Al Jamali, F., Al Jabri, H., & Hamadou, R. Ben. (2021). Marine health of the Arabian Gulf: Drivers of pollution and assessment approaches focusing on desalination activities. *Marine Pollution Bulletin*, 164, 112085. <https://doi.org/10.1016/J.MARPOLBUL.2021.112085>
- Hyatt, D., Chen, G. L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, 11. <https://doi.org/10.1186/1471-2105-11-119>
- Ibrahim, M. M., Al-Turki, A., Al-Sewedi, D., Arif, I. A., & El-Gaaly, G. A. (2015). Molecular application for identification of polycyclic aromatic hydrocarbons degrading bacteria (PAHD) species isolated from oil polluted soil in Dammam, Saud Arabia. *Saudi Journal of Biological Sciences*, 22(5), 651–655. <https://doi.org/10.1016/J.SJBS.2015.04.018>

- Jamal, M. T. (2022). Enrichment of Potential Halophilic *Marinobacter* Consortium for Mineralization of Petroleum Hydrocarbons and Also as Oil Reservoir Indicator in Red Sea, Saudi Arabia. *Polycyclic Aromatic Compounds*, 42(2), 400–411. <https://doi.org/10.1080/10406638.2020.1735456>
- Jamal, M. T., & Pugazhendhi, A. (2018). Degradation of petroleum hydrocarbons and treatment of refinery wastewater under saline condition by a halophilic bacterial consortium enriched from marine environment (Red Sea), Jeddah, Saudi Arabia. *3 Biotech*, 8(6), 1–10. <https://doi.org/10.1007/S13205-018-1296-X/FIGURES/8>
- Jeong, J., Yun, K., Mun, S., Chung, W. H., Choi, S. Y., Nam, Y. do, Lim, M. Y., Hong, C. P., Park, C. H., Ahn, Y., & Han, K. (2021). The effect of taxonomic classification by full-length 16S rRNA sequencing with a synthetic long-read technology. *Scientific Reports* 2021 11:1, 11(1), 1–12. <https://doi.org/10.1038/s41598-020-80826-9>
- Jones, D. M., Douglas, A. G., Parkes, R. J., Taylor, J., Giger, W., & Schaffner, C. (1983). The recognition of biodegraded petroleum-derived aromatic hydrocarbons in recent marine sediments. *Marine Pollution Bulletin*, 14(3), 103–108. [https://doi.org/10.1016/0025-326X\(83\)90310-7](https://doi.org/10.1016/0025-326X(83)90310-7)
- Jones, J. G., Knight, M., & Byrom, J. A. (1970). Effect of gross pollution by kerosine hydrocarbons on the Microflora of a moorland soil. *Nature*, 227(5263), 1166. <https://doi.org/10.1038/2271166A0>
- Juhasz, A. L., Britz, M. L., & Stanley, G. A. (1996). Degradation of high molecular weight polycyclic aromatic hydrocarbons by *Pseudomonas cepacia*. *Biotechnology Letters*, 18(5), 577–582. <https://doi.org/10.1007/BF00140206/METRICS>

- Juhasz, A. L., Britz, M. L., & Stanley, G. A. (1997). Degradation of benzo[a]pyrene, dibenz[a,h]anthracene and coronene by *Burkholderia cepacia*. *Water Science and Technology*, 36(10), 45–51. [https://doi.org/10.1016/S0273-1223\(97\)00641-0](https://doi.org/10.1016/S0273-1223(97)00641-0)
- Juhasz, A. L., Stanley, G. A., & Britz, M. L. (2000). Microbial degradation and detoxification of high molecular weight polycyclic aromatic hydrocarbons by *Stenotrophomonas maltophilia* strain VUN 10,003. *Letters in Applied Microbiology*, 30(5), 396–401. <https://doi.org/10.1046/J.1472-765X.2000.00733.X>
- Kadri, T., Rouissi, T., Kaur Brar, S., Cledon, M., Sarma, S., & Verma, M. (2017). Biodegradation of polycyclic aromatic hydrocarbons (PAHs) by fungal enzymes: A review. *Journal of Environmental Sciences*, 51, 52–74. <https://doi.org/10.1016/J.JES.2016.08.023>
- Karna Radjasa, O., Steven, R., Humaira, Z., Martha Dwivany, F., Nugrahapraja, H., Pebrianto Trinugroho, J., Kristianti, T., Chahyadi, A., Natanael, Y., Priharto, N., Amalia B Sembiring, F. P., Dwijayanti, A., Kusmita, L., Moeis, M. R., & Sri Harjati Suhardi, V. (123 C.E.). *Biosynthetic gene cluster profiling from North Java Sea *Virgibacillus salarius* reveals hidden potential metabolites*. <https://doi.org/10.1038/s41598-023-44603-8>
- Katoh, K., Rozewicki, J., & Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, 20(4), 1160–1166. <https://doi.org/10.1093/BIB/BBX108>
- Kimes, N. E., Callaghan, A. V., Suflita, J. M., & Morris, P. J. (2014). Microbial transformation of the deepwater horizon oil spill-past, present, and future perspectives. *Frontiers in Microbiology*, 5(NOV), 104794. <https://doi.org/10.3389/FMICB.2014.00603/ABSTRACT>
- Lafi, F. F., Ramirez-Prado, J. S., Alam, I., Bajic, V. B., Hirt, H., & Saad, M. M. (2016). Draft Genome Sequence of *Halomonas elongata* Strain K4, an Endophytic Growth-Promoting

- Bacterium Enhancing Salinity Tolerance In Planta. *Genome Announcements*, 4(6), e01214-16. <https://doi.org/10.1128/GENOMEA.01214-16>
- Lal, B., & Khanna, S. (1996). Degradation of crude oil by *Acinetobacter calcoaceticus* and *Alcaligenes odorans*. *The Journal of Applied Bacteriology*, 81(4), 355–362. <https://doi.org/10.1111/J.1365-2672.1996.TB03519.X>
- Lamendella, R., Strutt, S., Borglin, S., Chakraborty, R., Tas, N., Mason, O. U., Hultman, J., Prestat, E., Hazen, T. C., & Jansson, J. K. (2014). Assessment of the deepwater horizon oil spill impact on gulf coast microbial communities. *Frontiers in Microbiology*, 5(APR), 75227. <https://doi.org/10.3389/FMICB.2014.00130/ABSTRACT>
- Lawal, A. T. (2017). Polycyclic aromatic hydrocarbons. A review. In *Cogent Environmental Science* (Vol. 3, Issue 1). Cogent. <https://doi.org/10.1080/23311843.2017.1339841>
- Li, R., Feng, Y., Chen, H., Zhang, C., Huang, Y., Chen, L., Hao, Q., Cao, D., Yuan, S., & Zhou, X. (2020). Whole-Genome Sequencing of *Bradyrhizobium diazoefficiens* 113-2 and Comparative Genomic Analysis Provide Molecular Insights Into Species Specificity and Host Specificity. *Frontiers in Microbiology*, 11, 576800. <https://doi.org/10.3389/FMICB.2020.576800/BIBTEX>
- Liang, C. Y., Yang, C. H., Lai, C. H., Huang, Y. H., & Lin, J. N. (2019). Comparative Genomics of 86 Whole-Genome Sequences in the Six Species of the *Elizabethkingia* Genus Reveals Intraspecific and Interspecific Divergence. *Scientific Reports* 2019 9:1, 9(1), 1–11. <https://doi.org/10.1038/s41598-019-55795-3>
- Lomsadze, A., Gemayel, K., Tang, S., & Borodovsky, M. (2018). Modeling leaderless transcription and atypical genes results in more accurate gene prediction in prokaryotes. *Genome Research*, 28(7), 1079–1089. <https://doi.org/10.1101/GR.230615.117>

- Lu, J., Rincon, N., Wood, D. E., Breitwieser, F. P., Pockrandt, C., Langmead, B., Salzberg, S. L., & Steinegger, M. (2022). Metagenome analysis using the Kraken software suite. *Nature Protocols* 2022 17:12, 17(12), 2815–2839. <https://doi.org/10.1038/s41596-022-00738-y>
- Mahmoud, G. A. E., & Bagy, M. M. K. (2019). Microbial degradation of petroleum hydrocarbons. In *Microbial Action on Hydrocarbons* (Vol. 45, Issue 1, pp. 299–320). https://doi.org/10.1007/978-981-13-1840-5_12
- Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., Fong, J. H., Geer, L. Y., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Jackson, J. D., Ke, Z., Lanczycki, C. J., Lu, F., Marchler, G. H., Mullokandov, M., Omelchenko, M. V., ... Bryant, S. H. (2011). CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Research*, 39(Database issue). <https://doi.org/10.1093/NAR/GKQ1189>
- Meckenstock, R. U., Boll, M., Mouttaki, H., Koelschbach, J. S., Cunha Tarouco, P., Weyrauch, P., Dong, X., & Himmelberg, A. M. (2016). Anaerobic Degradation of Benzene and Polycyclic Aromatic Hydrocarbons. *Journal of Molecular Microbiology and Biotechnology*, 26(1–3), 92–118. <https://doi.org/10.1159/000441358>
- Modarress, B., Ansari, A., & Thies, E. (2016). Outsourcing in the Persian Gulf petroleum supply chain. *Strategic Outsourcing*, 9(1), 2–21. <https://doi.org/10.1108/SO-08-2015-0019/FULL/XML>
- Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., & Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research*, 35(Web Server issue). <https://doi.org/10.1093/NAR/GKM321>

- Muccee, F., Ejaz, S., Riaz, N., & Iqbal, J. (2021). Molecular and functional analysis of naphthalene-degrading bacteria isolated from the effluents of indigenous tanneries. *Journal of Basic Microbiology*, 61(7), 627–641. <https://doi.org/10.1002/JOBM.202100123>
- Mulkins Phillips, G. J., & Stewart, J. E. (1974). Distribution of hydrocarbon utilizing bacteria in Northwestern Atlantic waters and coastal sediments. *Canadian Journal of Microbiology*, 20(7), 955–962. <https://doi.org/10.1139/m74-147>
- Musat, F., Galushko, A., Jacob, J., Widdel, F., Kube, M., Reinhardt, R., Wilkes, H., Schink, B., & Rabus, R. (2009). Anaerobic degradation of naphthalene and 2-methylnaphthalene by strains of marine sulfate-reducing bacteria. *Environmental Microbiology*, 11(1), 209–219. <https://doi.org/10.1111/J.1462-2920.2008.01756.X>
- Nzila, A., Kwon Jung, B., Kim, M.-C., Conrad Ibal, J., Budiyo, F., Musa, M. M., Thukair, A., Kim, S.-J., Shin, J.-H., & 생명과학부 키파드 석유광물대학교. (2018). Complete genome sequence of the polycyclic aromatic hydrocarbons biodegrading bacterium *Idiomarina piscisalsi* strain 10PY1A isolated from oil-contaminated soil 기름으로 오염된 토양에서 분리된 다환방향족탄화수소 분해 세균 *Idiomarina piscisalsi* 10PY1A의 유전체 염기서열 해독. *Korean Journal of Microbiology*, 54(3), 289–292. <https://doi.org/10.7845/kjm.2018.8034>
- Nzila, A., Musa, M. M., Afuecheta, E., Al-Thukair, A., Sankaran, S., Xiang, L., & Li, Q. X. (2023). Benzo[A]Pyrene Biodegradation by Multiple and Individual Mesophilic Bacteria under Axenic Conditions and in Soil Samples. *International Journal of Environmental Research and Public Health*, 20(3), 1855. <https://doi.org/10.3390/IJERPH20031855/S1>

- Nzila, A., Musa, M. M., Sankara, S., Al-Momani, M., Xiang, L., & Li, Q. X. (2021). Degradation of benzo[a]pyrene by halophilic bacterial strain *Staphylococcus haemolyticus* strain 10SBZ1A. *PLOS ONE*, *16*(2), e0247723. <https://doi.org/10.1371/JOURNAL.PONE.0247723>
- Nzila, A., Ramirez, C. O., Musa, M. M., Sankara, S., Basheer, C., & Li, Q. X. (2018). Pyrene biodegradation and proteomic analysis in *Achromobacter xylosoxidans*, PY4 strain. *International Biodeterioration & Biodegradation*, *130*, 40–47. <https://doi.org/10.1016/J.IBIOD.2018.03.014>
- Nzila, A., Sankara, S., Al-Momani, M., & Musa, M. M. (2018). Isolation and characterisation of bacteria degrading polycyclic aromatic hydrocarbons: Phenanthrene and anthracene. *Archives of Environmental Protection*, *44*(2), 43–54. <https://doi.org/10.24425/119693>
- Nzila, A., Thukair, A., Sankara, S., Chanbasha, B., & Musa, M. M. (2016). Isolation and characterization of naphthalene biodegrading *Methylobacterium radiotolerans* bacterium from the eastern coastline of the Kingdom of Saudi Arabia. *Archives of Environmental Protection*, *42*(3), 25–32. <https://doi.org/10.1515/AEP-2016-0028>
- Okeyode, A. H., Al-Thukair, A., Chanbasha, B., Nazal, M. K., Afuecheta, E., Musa, M. M., Algarni, S., & Nzila, A. (2023). Degradation of the highly complex polycyclic aromatic hydrocarbon coronene by the halophilic bacterial strain *Halomonas caseinilytica*, 10SCRN4D. *Archives of Environmental Protection*. <https://doi.org/10.24425/aep.2023.147330>
- Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A. R., Xia, F., & Stevens, R. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology

- (RAST). *Nucleic Acids Research*, 42(Database issue).
<https://doi.org/10.1093/NAR/GKT1226>
- Oyehan, T. A., & Al-Thukair, A. A. (2017). Isolation and characterization of PAH-degrading bacteria from the Eastern Province, Saudi Arabia. *Marine Pollution Bulletin*, 115(1–2), 39–46. <https://doi.org/10.1016/J.MARPOLBUL.2016.11.007>
- Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B. L., Salazar, G. A., Bileschi, M. L., Bork, P., Bridge, A., Colwell, L., Gough, J., Haft, D. H., Letunić, I., Marchler-Bauer, A., Mi, H., Natale, D. A., Orengo, C. A., Pandurangan, A. P., Rivoire, C., ... Bateman, A. (2023). InterPro in 2022. *Nucleic Acids Research*, 51(D1), D418–D427. <https://doi.org/10.1093/NAR/GKAC993>
- Pieper, D. H., & Reineke, W. (2000). Engineering bacteria for bioremediation. *Current Opinion in Biotechnology*, 11(3), 262–270. [https://doi.org/10.1016/S0958-1669\(00\)00094-X](https://doi.org/10.1016/S0958-1669(00)00094-X)
- Pinholt, Y., Struwe, S., & Kjøller, A. (1979). Microbial changes during oil decomposition in soil. *Ecography*, 2(3), 195–200. <https://doi.org/10.1111/J.1600-0587.1979.TB00701.X>
- Pugazhendi, A., Qari, H., Al-Badry Basahi, J. M., Godon, J. J., & Dhavamani, J. (2017). Role of a halothermophilic bacterial consortium for the biodegradation of PAHs and the treatment of petroleum wastewater at extreme conditions. *International Biodeterioration & Biodegradation*, 121, 44–54. <https://doi.org/10.1016/J.IBIOD.2017.03.015>
- Rahman, K. S. M., Rahman, T. J., Kourkoutas, Y., Petsas, I., Marchant, R., & Banat, I. M. (2003). Enhanced bioremediation of n-alkane in petroleum sludge using bacterial consortium amended with rhamnolipid and micronutrients. *Bioresource Technology*, 90(2), 159–168. [https://doi.org/10.1016/S0960-8524\(03\)00114-7](https://doi.org/10.1016/S0960-8524(03)00114-7)

- Raina, V., Nayak, T., Ray, L., Kumari, K., & Suar, M. (2019). A Polyphasic Taxonomic Approach for Designation and Description of Novel Microbial Species. *Microbial Diversity in the Genomic Era*, 137–152. <https://doi.org/10.1016/B978-0-12-814849-5.00009-5>
- Ramos, D. T., da Silva, M. L. B., Nossa, C. W., Alvarez, P. J. J., & Corseuil, H. X. (2014). Assessment of microbial communities associated with fermentative-methanogenic biodegradation of aromatic hydrocarbons in groundwater contaminated with a biodiesel blend (B20). *Biodegradation*, 25(5), 681–691. <https://doi.org/10.1007/S10532-014-9691-4/FIGURES/4>
- Ranjan, R., Rani, A., Metwally, A., Mcgee, H. S., & Perkins, D. L. (2015). *Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing*. <https://doi.org/10.1016/j.bbrc.2015.12.083>
- Rehmann, K., Noll, H. P., Steinberg, C. E. W., & Ketrup, A. A. (1998). Pyrene degradation by *Mycobacterium* sp. strain KR2. *Chemosphere*, 36(14), 2977–2992. [https://doi.org/10.1016/S0045-6535\(97\)10240-5](https://doi.org/10.1016/S0045-6535(97)10240-5)
- Rice, P., Longden, L., & Bleasby, A. (2000). EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics*, 16(6), 276–277. [https://doi.org/10.1016/S0168-9525\(00\)00204-2](https://doi.org/10.1016/S0168-9525(00)00204-2)
- Rockne, K. J., & Strand, S. E. (2001). Anaerobic biodegradation of naphthalene, phenanthrene, and biphenyl by a denitrifying enrichment culture. *Water Research*, 35(1), 291–299. [https://doi.org/10.1016/S0043-1354\(00\)00246-3](https://doi.org/10.1016/S0043-1354(00)00246-3)
- Romero, M. C., Cazau, M. C., Giorgieri, S., & Arambarri, A. M. (1998). Phenanthrene degradation by microorganisms isolated from a contaminated stream. *Environmental Pollution*, 101(3), 355–359. [https://doi.org/10.1016/S0269-7491\(98\)00056-6](https://doi.org/10.1016/S0269-7491(98)00056-6)

- Ron, E. Z., & Rosenberg, E. (2014). Enhanced bioremediation of oil spills in the sea. *Current Opinion in Biotechnology*, 27, 191–194. <https://doi.org/10.1016/J.COPBIO.2014.02.004>
- Schneider, J., Grosser, R., Jayasimhulu, K., Xue, W., & Warshawsky, D. (1996). Degradation of pyrene, benz[a]anthracene, and benzo[a]pyrene by *Mycobacterium* sp. strain RJGII-135, isolated from a former coal gasification site. *Applied and Environmental Microbiology*, 62(1), 13. <https://doi.org/10.1128/AEM.62.1.13-19.1996>
- Schwibbert, K., Marin-Sanguino, A., Bagyan, I., Heidrich, G., Lentzen, G., Seitz, H., Rampp, M., Schuster, S. C., Klenk, H.-P., Pfeiffer, F., Oesterhelt, D., & Kunte, H. J. (2010). *A blueprint of ectoine metabolism from the genome of the industrial producer Halomonas elongata DSM 2581 T*. <https://doi.org/10.1111/j.1462-2920.2010.02336.x>
- Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 30(14), 2068–2069. <https://doi.org/10.1093/BIOINFORMATICS/BTU153>
- Shahsavari, E., Schwarz, A., Aburto-Medina, A., & Ball, A. S. (2019). Biological Degradation of Polycyclic Aromatic Compounds (PAHs) in Soil: a Current Perspective. *Current Pollution Reports*, 5(3), 84–92. <https://doi.org/10.1007/S40726-019-00113-8/FIGURES/3>
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19), 3210–3212. <https://doi.org/10.1093/BIOINFORMATICS/BTV351>
- Simon, C., & Daniel, R. (2011). Metagenomic analyses: Past and future trends. *Applied and Environmental Microbiology*, 77(4), 1153–1161. <https://doi.org/10.1128/AEM.02345-10/ASSET/C1E302BD-8CC3-4054-B12D-95AB7C54848C/ASSETS/GRAPHIC/ZAM9991017850001.JPEG>

- Taboada, B., Estrada, K., Ciria, R., & Merino, E. (2018). Operon-mapper: a web server for precise operon identification in bacterial and archaeal genomes. *Bioinformatics*, 34(23), 4118–4120. <https://doi.org/10.1093/BIOINFORMATICS/BTY496>
- Takeuchi, N., Wolf, Y. I., Makarova, K. S., & Koonin, E. V. (2012). Nature and Intensity of Selection Pressure on CRISPR-Associated Genes. *Journal of Bacteriology*, 194(5), 1216. <https://doi.org/10.1128/JB.06521-11>
- Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution*, 38(7), 3022–3027. <https://doi.org/10.1093/MOLBEV/MSAB120>
- Tempel, S. (2012). Using and understanding RepeatMasker. *Methods in Molecular Biology (Clifton, N.J.)*, 859, 29–51. https://doi.org/10.1007/978-1-61779-603-6_2
- Thomas, T., Gilbert, J., & Meyer, F. (2012). Metagenomics - a guide from sampling to data analysis. *Microbial Informatics and Experimentation*, 2(1), 1–12. <https://doi.org/10.1186/2042-5783-2-3/FIGURES/1>
- Trzesicka-Mlynarz, D., & Ward, O. P. (1995). Degradation of polycyclic aromatic hydrocarbons (PAHs) by a mixed culture and its component pure cultures, obtained from PAH-contaminated soil. *Canadian Journal of Microbiology*, 41(6), 470–476. <https://doi.org/10.1139/M95-063>
- Urgun-Demirtas, M., Stark, B., & Pagilla, K. (2006). Use of Genetically Engineered Microorganisms (GEMs) for the Bioremediation of Contaminants. *Critical Reviews in Biotechnology*, 26(3), 145–164. <https://doi.org/10.1080/07388550600842794>
- Varjani, S. J. (2017). Microbial degradation of petroleum hydrocarbons. *Bioresource Technology*, 223, 277–286. <https://doi.org/10.1016/J.BIORTECH.2016.10.037>

- Vera, A., Wilson, F. P., & Cupples, A. M. (2022). Predicted functional genes for the biodegradation of xenobiotics in groundwater and sediment at two contaminated naval sites. *Applied Microbiology and Biotechnology*, 106(2), 835–853. <https://doi.org/10.1007/S00253-021-11756-3>
- Vernikos, G. S., & Parkhill, J. (2006). Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the Salmonella pathogenicity islands. *Bioinformatics (Oxford, England)*, 22(18), 2196–2203. <https://doi.org/10.1093/BIOINFORMATICS/BTL369>
- Verrhiest, G. J., Clément, B., Volat, B., Montuelle, B., & Perrodin, Y. (2002). Interactions between a polycyclic aromatic hydrocarbon mixture and the microbial communities in a natural freshwater sediment. *Chemosphere*, 46(2), 187–196. [https://doi.org/10.1016/S0045-6535\(01\)00064-9](https://doi.org/10.1016/S0045-6535(01)00064-9)
- Von Netzer, F., Kuntze, K., Vogt, C., Richnow, H. H., Boll, M., & Lueders, T. (2016). Functional Gene Markers for Fumarate-Adding and Dearomatizing Key Enzymes in Anaerobic Aromatic Hydrocarbon Degradation in Terrestrial Environments. *Journal of Molecular Microbiology and Biotechnology*, 26(1–3), 180–194. <https://doi.org/10.1159/000441946>
- Wackett, L. P. (2012). Petroleum microbiology: An annotated selection of World Wide Web sites relevant to the topics in Microbial Biotechnology. *Microbial Biotechnology*, 5(4), 579. <https://doi.org/10.1111/J.1751-7915.2012.00350.X>
- Waigi, M. G., Kang, F., Goikavi, C., Ling, W., & Gao, Y. (2015). Phenanthrene biodegradation by sphingomonads and its application in the contaminated soils and sediments: A review. *International Biodeterioration & Biodegradation*, 104, 333–349. <https://doi.org/10.1016/J.IBIOD.2015.06.008>

- Walter, U., Beyer, M., Klein, J., & Rehm, H. J. (1991). Degradation of pyrene by *Rhodococcus* sp. UW1. *Applied Microbiology and Biotechnology*, 34(5), 671–676. <https://doi.org/10.1007/BF00167921/METRICS>
- Wang, J., Chitsaz, F., Derbyshire, M. K., Gonzales, N. R., Gwadz, M., Lu, S., Marchler, G. H., Song, J. S., Thanki, N., Yamashita, R. A., Yang, M., Zhang, D., Zheng, C., Lanczycki, C. J., & Marchler-Bauer, A. (2022). The conserved domain database in 2023. *Nucleic Acids Research*, 51(D1), D384. <https://doi.org/10.1093/NAR/GKAC1096>
- Yu, S. H., Ke, L., Wong, Y. S., & Tam, N. F. Y. (2005). Degradation of polycyclic aromatic hydrocarbons by a bacterial consortium enriched from mangrove sediments. *Environment International*, 31(2), 149–154. <https://doi.org/10.1016/J.ENVINT.2004.09.008>
- Yuan, S. Y., Shiung, L. C., & Chang, B. V. (2002). Biodegradation of polycyclic aromatic hydrocarbons by inoculated microorganisms in soil. *Bulletin of Environmental Contamination and Toxicology*, 69(1), 66–73. <https://doi.org/10.1007/S00128-002-0011-Z>
- Zhou, L., Li, K. P., Mbadinga, S. M., Yang, S. Z., Gu, J. D., & Mu, B. Z. (2012). Analyses of n-alkanes degrading community dynamics of a high-temperature methanogenic consortium enriched from production water of a petroleum reservoir by a combination of molecular techniques. *Ecotoxicology*, 21(6), 1680–1691. <https://doi.org/10.1007/s10646-012-0949-5>
- ZOBELL, C. E. (1946). Action of microorganisms on hydrocarbons. *Bacteriological Reviews*, 10(1), 1–49. <https://doi.org/10.1128/mmbr.10.1-2.1-49.1946>
- Zytnick, A. M., Gutenthaler-Tietze, S. M., Aron, A. T., Reitz, Z. L., Phi, M. T., Good, N. M., Petras, D., Daumann, L. J., & Martinez-Gomez, N. C. (2024). Identification and characterization of a small-molecule metallophore involved in lanthanide metabolism. *Proceedings of the National Academy of Sciences of the United States of America*, 121(32), e2322096121. <https://doi.org/10.1073/PNAS.2322096121>

Appendices

Combining the three annotations

!pip install biopython

```
from Bio import SeqIO
```

```
import os
```

```
# For manually uploaded files
```

```
prokka_file = r'C:\Users\thasn\OneDrive\Desktop\prokka.fasta'
```

```
prodigal_file = r'C:\Users\thasn\OneDrive\Desktop\prodigal.prot'
```

```
genemark_file = r'C:\Users\thasn\OneDrive\Desktop\GeneMark.faa'
```

from Bio import SeqIO

```
# Define the filenames of your uploaded files directly
```

```
prokka_file = 'prokka.fasta'    # Use the exact name after uploading
```

```
prodigal_file = 'prodigal.prot' # Use the exact name after uploading
```

```
genemark_file = 'GeneMark.faa' # Use the exact name after uploading
```

Output file for the combined sequences

```
combined_file = 'combined_predictions.fasta'
```

Lists to hold sequences from each tool

```
prokka_sequences = []
```

```
prodigal_sequences = []
```

```
genemark_sequences = []
```

Function to read sequences from a file and store them in a list

```
def read_sequences(file_path):
```

```
    sequences = []
```

```
    with open(file_path, 'r') as handle:
```

```
        for record in SeqIO.parse(handle, 'fasta'):
```

```
            sequences.append(record)
```

```
    return sequences
```

Read sequences from each tool

```
prokka_sequences = read_sequences(prokka_file)
```

```
prodigal_sequences = read_sequences(prodigal_file)
```

```
genemark_sequences = read_sequences(genemark_file)
```

Output list to hold the interleaved sequences

```
interleaved_sequences = []
```

Determine the maximum number of sequences to interleave

```
max_length = max(len(prokka_sequences), len(prodigal_sequences), len(genemark_sequences))
```

Interleave sequences

```
for i in range(max_length):
```

```
    if i < len(prokka_sequences):
```

```
        prokka_sequences[i].id = f'{prokka_sequences[i].id}_Prokka' # Modify ID for Prokka
```

```
        prokka_sequences[i].description = ""
```

```
        interleaved_sequences.append(prokka_sequences[i])
```

```
    if i < len(prodigal_sequences):
```

```
        prodigal_sequences[i].id = f'{prodigal_sequences[i].id}_Prodigal' # Modify ID for Prodigal
```

```
        prodigal_sequences[i].description = ""
```

```
        interleaved_sequences.append(prodigal_sequences[i])
```

```
    if i < len(genemark_sequences):
```

```
        genemark_sequences[i].id = f'{genemark_sequences[i].id}_GeneMark' # Modify ID for GeneMark
```

```
        genemark_sequences[i].description = ""
```

```
        interleaved_sequences.append(genemark_sequences[i])
```

Write the interleaved sequences to the combined FASTA file

```
with open(combined_file, 'w') as output_handle:
```

```
    SeqIO.write(interleaved_sequences, output_handle, 'fasta')
```

```
print(f'Combined file created as '{combined_file}' with interleaved sequences.")
```

Aligning using MAFFT

Install required packages

```
!apt-get install mafft
```

```
!pip install biopython
```

Import necessary modules

```
import subprocess
```

```
import io # This import was missing
```

```
from Bio import SeqIO
```

Define a function to perform alignments using MAFFT and handle input/output sequences

```
def align_sequences(sequences):
```

```
    # Write sequences to a temporary input file
```

```
    temp_input = '/content/temp_input.fasta'
```

```
    SeqIO.write(sequences, temp_input, 'fasta')
```

Build the command for MAFFT

```
    command = ['mafft', '--auto', temp_input]
```

```

# Execute the command and capture the output
result = subprocess.run(command, capture_output=True, text=True)

# Parse aligned sequences from the output using io.StringIO
aligned_sequences = SeqIO.parse(io.StringIO(result.stdout), 'fasta')
return list(aligned_sequences)

# Load all sequences from the input FASTA file
input_fasta = '/content/aligned.fasta'
sequences = list(SeqIO.parse(input_fasta, 'fasta'))

# List to store all aligned sequences
all_aligned_sequences = []

# Align each group of three sequences and store results
for i in range(0, len(sequences), 3):
    group = sequences[i:i+3] # Get a group of three sequences
    if len(group) == 3: # Ensure there are exactly three sequences in the group
        aligned_group = align_sequences(group)
        all_aligned_sequences.extend(aligned_group)

# Save all aligned sequences to a single file
output_path = '/content/combined_aligned_sequences.fasta'
SeqIO.write(all_aligned_sequences, output_path, 'fasta')

# Output success message
print(f"Aligned sequences saved to {output_path}")

```

Vitae

Name : Thasneema Rafic

Nationality : Indian

Date of Birth : 06/19/1999

Email : thasneemarafic@gmail.com

Address : House no. 14 98, opposite masjid, Shirva
post, Shirva, Udupi, Karnataka, India,
574116

Academic Background : M.S. Life sciences, Department of
Bioengineering, King Fahd University of
Science and Technology, Dhahran, Saudi
Arabia

B.Sc. Biotechnology, Manipal School of
Life sciences, Manipal Academy of Higher
Education, Manipal, India

Mobile : 00966555690564

Projects : Isolation of microRNAs from blood sample
infected with Plasmodium and in silico
target prediction (2021)

Whole genome sequencing,
characterization and analysis of coronene
degrading bacterial strain, halomonas
caseinilytica: a bioinformatics approach
(2024)

Publications

Targeting Receptor-Ligand Chemistry for
Drug delivery Across Blood-Brain Barrier
in Brain Disease (2021)