

EXPLORING POTENTIAL BIOSYNTHETIC GENE CLUSTERS IN THE HALOPHILIC ARCHAEA

Submitted in partial fulfillment of the requirements for the award of Master of
Science in Bioinformatics and Data Science

By

Murugula Raghavendra (41739012)



**DEPARTMENT OF BIOINFORMATICS
SCHOOL OF BIO AND CHEMICAL ENGINEERING**

SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited with Grade "A" by NAAC | 12B Status by UGC | Approved by AICTE

JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600 119

APRIL – 2023



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)

Accredited with "A" grade by NAAC | 12B Status by UGC | Approved by AICTE

Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai – 600 119

www.sathyabama.ac.in

DEPARTMENT OF BIOINFORMATICS

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **Murugula Raghavendra (41739014)** who carried out the final year project entitled **"EXPLORING POTENTIAL BIOSYNTHETIC GENE CLUSTERS IN THE HALOPHILIC ARCHAEA"** under our supervision from January 2023 to May 2023.

Dr. JEMMY CHRISTY

Internal Guide

Dr. AMIT KUMAR

External Guide

Dr. JEMMY CHRISTY

Head of the Department

Submitted for Viva voce Examination held on

Internal Examiner

External Examiner

DECLARATION

I, **MURUGULA RAGHAVENDRA (41739012)**, hereby declare that the Project Report entitled “**EXPLORING POTENTIAL BIOSYNTHETIC GENE CLUSTERS IN THE HALOPHILIC ARCHAEA**” done by me under the guidance of **Dr. AMIT KUMAR** and **Dr. JEMMY CHRISTY** and is submitted in partial fulfilment of the requirements for the award of Master of Science in Bioinformatics and Data Science.

MURUGULA RAGHAVENDRA

DATE:

SIGNATURE OF THE CANDIDATE

PLACE: CHENNAI

ACKNOWLEDGEMENT

I would like to acknowledge the management of the **Sathyabama Institute of Science and Technology** for their support and encouragement.

I would like to extend my heartfelt gratitude to my guide and mentor, **Dr.Amit Kumar** for his invaluable guidance and support throughout this project. His insights, encouragement, and constructive feedback has been instrumental in shaping my work and helping me achieve my research objectives.

I would also like to express my sincere thanks to **Dr. Jemmy Christy**, the head of the Department of Bioinformatics, and my internal guide for her constant support.

Finally, I express my gratitude to my family and the teaching and non-teaching staff of Sathyabama. Their unwavering belief and confidence in me has shaped my work and I am grateful for their understanding and patience throughout the process of my master's project.



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE

www.sathyabama.ac.in

CERTIFICATE

This is to certify that the dissertation entitled **“Exploring potential biosynthetic gene clusters in the halophilic archaea”** was carried out by **Mr. Murugula Raghavendra (Reg. no. 41739012)**, Department of Bioinformatics and Data Sciences, Sathyabama Institute of Science and Technology, Chennai in partial fulfillment of the requirements of the award of the Degree of Master of Science in Bioinformatics and Data Sciences 2021-2023, is a bonafide record of the project work done under my guidance and supervision in the **Centre for Climate Change Studies, Sathyabama Institute of Science and Technology, Chennai.**

This thesis or part, therefore, has not been submitted elsewhere for any other degree or diploma.

PLACE: Chennai

DATE:

Dr.Amit Kumar

ABSTRACT

Halophilic archaea have been studied for the discovery of biosynthetic gene clusters, and the extraction and analysis of these clusters have provided insights into the evolutionary relationships between these organisms. Research has shown that halophilic archaea produce a wide range of secondary metabolites, and the identification and characterization of these biosynthetic gene clusters can provide valuable information for biotechnological applications. Moreover, the analysis of these gene clusters can help us better understand the evolution of halophilic archaea and the diversity of biosynthetic pathways. Additionally, by comparing the biosynthetic gene clusters of different halophilic archaea using phylogenetic analyses, we can gain insight into the evolutionary relationships and diversification patterns among these organisms. Genome mining has revolutionized the discovery of biosynthetic gene clusters in archaea. With the advent of high-throughput sequencing technologies, it has become easier to sequence and analyze the genomes of microorganisms, including halophilic archaea. This has led to the identification of a large number of biosynthetic gene clusters that were previously unknown. In addition to identifying biosynthetic gene clusters, genome mining has also allowed researchers to study the evolutionary relationships between halophilic archaea. By comparing the genomes of different archaeal species, researchers can identify gene clusters that are conserved across different species. here, we analyze 237 halophilic archaea taken from the halodom database, mined for their biosynthetic gene clusters to understand their diversity and also their evolutionary relationships The study of biosynthetic gene clusters in halophilic archaea can thus contribute to a better understanding of their ecology, evolution, and potential for biotechnological applications.

TABLE OF CONTENTS

CHAPTER NO	TITLE	PAGE NO
	ABSTRACT	i
	TABLES AND FIGURES	iv
	LIST OF ABBREVIATION	v
1	INTRODUCTION	
	1.1 Natural Products	1
	1.2 Primary and Secondary Metabolites	3
	1.3 Biosynthetic Gene Clusters	3
	1.4 Halophiles	4
	1.5 Halophilic Archaea	4
2	REVIEW OF LITERATURE	
	2.1 Importance of Halophilic Archaea	6
	2.2 Biosynthetic Gene Cluster and Classes	7
	2.3 Tools and Databases for Extraction of Biosynthetic Gene Clusters	7
	2.4 Benefits of Halophiles	9
	2.5 Halodom Database	10
	2.6 Halophilic Archaea and Environment	10
3	AIM AND SCOPE	
	3.1 Aim	13
	3.2 Scope	13
4	MATERIALS AND METHODS	
	4.1 Mining of Halophilic Archaea Selection Using Halodom Database	14
	4.2 Genome import from NCBI Genome	14
	4.3 Identification of Biosynthetic Gene Clusters in Halophilic Archaea	15
	4.4 Manual Interpretation of Dataset	16
	4.5 Phylogenetic Tree Construction using MEGA	16

	4.6 Visualization of the phylogenetic tree in ITOL	16
5	RESULTS AND DISCUSSION	
	5.1 Results	17
	5.2 Discussion	21
6	SUMMARY AND CONCLUSION	
	REFERENCES	

LIST OF TABLES AND FIGURES

	DESCRIPTION	PAGE NO
Table 2.1	Databases and tools for biosynthetic gene cluster mining	9
Fig 5.1	Tolerance level	17
Fig 5.2	Phylogenetic tree construction using ITOL	18
Fig 5.3	Major classes of BGCs observed	19
Fig 5.4	Highest number of BGCs	20
Fig 5.5	Prediction of BGCs - process	20
Fig 5.6	Homepage of Halodom database	21

LIST OF ABBREVIATIONS

ABBREVIATION	FULL FORMS
S	
NDA	New drug application
ADME	absorption, distribution, metabolism, and excretion
NP	natural product
BGC	Biosynthetic Gene Cluster
antiSMASH	Antibiotics and Secondary Metabolite Analysis SHell
MIBiG	Minimum Information about a Biosynthetic Gene cluster
Bactibase	Bacterial Antimicrobial Compound Gene cluster database
ClusterMine360	Cluster Mining and Evaluation platform
CSDB	Chemical Structure Database
IMG-ABC	Integrated Microbial Genomes Atlas of Biosynthetic gene Clusters
NRPS	Non-Ribosomal Peptide Synthases
Ripp-like	Repeats-In-Toxin proteins and Proteases-like
RRE	Rare Earth Element
LAP	Linear Azole-containing Peptides
NRPS	Non-ribosomal peptide synthetases
PKS	Polyketide synthases
PKSNRPS	Hybrid PKS-NRPS gene clusters
BGCs	Biosynthetic gene clusters
MIBiG	Minimum Information about a Biosynthetic Gene cluster database
HMMs	Hidden Markov models
NCBI	National Center for Biotechnology Information
M	Molar concentration
HaloDom	Halophilic Domains Database
pH	potential of hydrogen
NaCl	Sodium Chloride
NaCl	sodium chloride
ATP	adenosine triphosphate
GBFF	GenBank Flat File

ABSTRACT

Halophilic archaea have been studied for the discovery of biosynthetic gene clusters, and the extraction and analysis of these clusters have provided insights into the evolutionary relationships between these organisms. Research has shown that halophilic archaea produce a wide range of secondary metabolites, and the identification and characterization of these biosynthetic gene clusters can provide valuable information for biotechnological applications. Moreover, the analysis of these gene clusters can help us better understand the evolution of halophilic archaea and the diversity of biosynthetic pathways. Additionally, by comparing the biosynthetic gene clusters of different halophilic archaea using phylogenetic analyses, we can gain insight into the evolutionary relationships and diversification patterns among these organisms. Genome mining has revolutionized the discovery of biosynthetic gene clusters in archaea. With the advent of high-throughput sequencing technologies, it has become easier to sequence and analyze the genomes of microorganisms, including halophilic archaea. This has led to the identification of a large number of biosynthetic gene clusters that were previously unknown. In addition to identifying biosynthetic gene clusters, genome mining has also allowed researchers to study the evolutionary relationships between halophilic archaea. By comparing the genomes of different archaeal species, researchers can identify gene clusters that are conserved across different species. here, we analyze 237 halophilic archaea taken from the halodom database, mined for their biosynthetic gene clusters to understand their diversity and also their evolutionary relationships The study of biosynthetic gene clusters in halophilic archaea can thus contribute to a better understanding of their ecology, evolution, and potential for biotechnological applications.

TABLE OF CONTENTS

CHAPTER NO	TITLE	PAGE NO
	ABSTRACT	i
	TABLES AND FIGURES	iv
	LIST OF ABBREVIATION	v
1	INTRODUCTION	
	1.1 Natural Products	1
	1.2 Primary and Secondary Metabolites	3
	1.3 Biosynthetic Gene Clusters	3
	1.4 Halophiles	4
	1.5 Halophilic Archaea	4
2	REVIEW OF LITERATURE	
	2.1 Importance of Halophilic Archaea	6
	2.2 Biosynthetic Gene Cluster and Classes	7
	2.3 Tools and Databases for Extraction of Biosynthetic Gene Clusters	7
	2.4 Benefits of Halophiles	9
	2.5 Halodom Database	10
	2.6 Halophilic Archaea and Environment	10
3	AIM AND SCOPE	
	3.1 Aim	13
	3.2 Scope	13
4	MATERIALS AND METHODS	
	4.1 Mining of Halophilic Archaea Selection Using Halodom Database	14
	4.2 Genome import from NCBI Genome	14
	4.3 Identification of Biosynthetic Gene Clusters in Halophilic Archaea	15
	4.4 Manual Interpretation of Dataset	16
	4.5 Phylogenetic Tree Construction using MEGA	16
	4.6 Visualization of the phylogenetic tree in ITOL	16

5	RESULTS AND DISCUSSION	
	5.1 Results	17
	5.2 Discussion	21
6	SUMMARY AND CONCLUSION	
	REFERENCES	

LIST OF TABLES AND FIGURES

	DESCRIPTION	PAGE NO
Table 2.1	Databases and tools for biosynthetic gene cluster mining	9
Fig 5.1	Tolerance level	17
Fig 5.2	Phylogenetic tree construction using ITOL	18
Fig 5.3	Major classes of BGCs observed	19
Fig 5.4	Highest number of BGCs	20
Fig 5.5	Prediction of BGCs - process	20
Fig 5.6	Homepage of Halodom database	21

LIST OF ABBREVIATIONS

ABBREVIATION	FULL FORMS
S	
NDA	New drug application
ADME	absorption, distribution, metabolism, and excretion
NP	natural product
BGC	Biosynthetic Gene Cluster
antiSMASH	Antibiotics and Secondary Metabolite Analysis SHell
MIBiG	Minimum Information about a Biosynthetic Gene cluster
Bactibase	Bacterial Antimicrobial Compound Gene cluster database
ClusterMine360	Cluster Mining and Evaluation platform
CSDB	Chemical Structure Database
IMG-ABC	Integrated Microbial Genomes Atlas of Biosynthetic gene Clusters
NRPS	Non-Ribosomal Peptide Synthases
Ripp-like	Repeats-In-Toxin proteins and Proteases-like
RRE	Rare Earth Element
LAP	Linear Azole-containing Peptides
NRPS	Non-ribosomal peptide synthetases
PKS	Polyketide synthases
PKSNRPS	Hybrid PKS-NRPS gene clusters
BGCs	Biosynthetic gene clusters
MIBiG	Minimum Information about a Biosynthetic Gene cluster database
HMMs	Hidden Markov models
NCBI	National Center for Biotechnology Information
M	Molar concentration
HaloDom	Halophilic Domains Database
pH	potential of hydrogen
NaCl	Sodium Chloride
NaCl	sodium chloride
ATP	adenosine triphosphate
GBFF	GenBank Flat File

CHAPTER 1

INTRODUCTION

1.1 Natural products

Natural Products are biological compounds that are currently a subject of great interest in the pharmaceutical, health food, and cosmetics industries, and the number of biological studies in this field is increasing quickly (1). For more than 50 years, natural products have saved us from infectious bacteria, fungi, and archaea. secondary metabolism has developed in response to the natural environment's needs and challenges.

More than 60% of approved and pre-new drug application (NDA) candidates are either natural products or related to them (2). Natural products have pharmaceutical properties that can help in treating diseases (3,4,5). The majority of the studies on the biological activity of natural products focus on various species of plants (1). Natural products have been the most productive source for the development of drugs (6) Natural products may be extracted from cells, tissues, and secretions of microorganisms (7,8).

Natural products help in chemical research and help in drug discovery and development (9). Natural products are significant for drug development they appear in consistently low amounts in natural medicine (10). The marine environment has a rich source of natural products with respective activities such as anticancer, anti-inflammatory, antiseptic, immunomodulatory, antifungal, antiviral, and antiparasitic (11). In recent decades, the anticancer activities of natural products are reported to have this effect on cancer (12).

Historically, plants have been widely used in traditional medicine and for their biochemical properties. There are an estimated 75,000 species of higher plants on Earth, but only 10% have been utilized in traditional medicine. Of the remaining species, only 1-5% have been scientifically studied and found to have therapeutic value. (13) Many drugs available on the market today were originally discovered from natural sources (14).

Additionally, numerous natural products exhibit favorable ADME and physicochemical properties, making them attractive targets for further research and potential drug development (15)(16).

Archaea have limited associations with medicinal properties and are not commonly used in traditional medicine. However, certain archaeal species have been investigated for their potential in biotechnology and industry, particularly in the production of enzymes and biochemicals. Microbes, such as bacteria and fungi, have been utilized in traditional medicine, particularly in systems such as traditional Chinese medicine. Some species of microbes produce compounds with potential medicinal properties, including antibiotics and other antimicrobial agents. Natural products possess wide structural and chemical diversity that cannot be matched by synthetic libraries of small molecules and continue to inspire novel discoveries in chemistry (17) NPs continue to be an important source of leads for new medicine, despite reduced interest from large pharmaceutical companies. Screening collections of natural products can be assembled economically to provide excellent coverage of drug-like chemical space and in formats that are compatible with high throughput bioassays. (18) In the industrialized nations at the present time, some 50% of all prescribed drugs are derived from natural products (19). Natural products have been a valuable source of drugs throughout history, and there are many different sources of natural product-based drugs. Plants, microbes, marine organisms, animals, minerals, insects, algae, traditional medicines, animal secretions, and bioprospecting are some of the most significant sources of natural product-based drugs. Plants, in particular, have been extensively studied and utilized for their medicinal properties, with many important drugs, such as aspirin and morphine, derived from plant sources. Microbes, including bacteria and fungi, have also been an important source of natural products, with many antibiotics derived from these organisms. Algae, insects, and animal secretions are relatively underexplored sources of natural products but have shown promise in the discovery of new drugs. Bioprospecting involves searching for new drugs in natural environments, such as rainforests and other ecosystems and has led to the discovery of several important natural product-based drugs. Overall, natural products remain an important area of drug discovery research, with many potential sources yet to be fully explored. Microorganisms have provided abundant sources

of natural products which have been developed as commercial products from human medicine, animal health, and, plant crop protection (20).

Natural products have inspired many developments in organic chemistry, leading to advances in synthetic methodologies and to the possibility of making analogues of the lead compound with improved pharmacological or pharmaceutical properties (9). NPs led can be challenging, particularly if synthetic routes are difficult Also NP based drug leads are often identified by phenotypic assays, and deconvolution of their molecular mechanisms of action can be time-consuming (21).

1.2 Primary and Secondary Metabolites

A primary metabolite is a kind of metabolite that is directly involved in growth, development, and reproduction. A primary metabolite is typically present in many organisms or cells. A primary metabolite is also called a central metabolite (22).

Secondary metabolites are compounds that are not necessary for a cell to live, but play a role in the interaction of the cell with its environment. These compounds are often involved in plant protection against biotic or abiotic stresses. Secondary metabolites are from different metabolites families that can be highly inducible in response to stresses. Primary metabolites perform essential metabolic roles by participating in nutrition and reproduction (23).

1.3 Biosynthetic Gene Clusters

A BGC can be defined as a physically clustered group of two or more genes in a particular genome that together encodes a biosynthetic pathway for the production of a specialized metabolite (including its chemical variants (24).

There are several databases related to biosynthetic gene clusters (BGCs) encoding secondary metabolites including antiSMASH, MIBiG, Bactibase, ClusterMine360, CSDB, IMG-ABC, Orphan assembly line polyketide synthases, and DoBiSUIT. BiG-SCAPE, antiSMASH, and BiG-FAM are some tools for predicting biosynthetic gene clusters (BGCs). Generally, bioactive compounds are synthesized by a cluster of genes which is referred to as a biosynthetic gene cluster (BGC). Biosynthetic gene clusters consist of several genes that encode enzymes, regulatory proteins, transporters, and other factors that are involved in the biosynthesis and regulation

of the specialized metabolites. The genes in a BGC are often arranged in a specific order and are regulated by a common set of control elements.

Every class of bioactive compound is synthesized by a specific group of gene clusters. For instance, non-ribosomal peptides are synthesized by non-ribosomal synthases (NRPS). A bacterial genome has multiple classes of BGCs which includes terpenes, RiPP-like, siderophores, thiopeptides, NRPS-independent siderophore, RRE containing, Lanthipeptide, lasso peptides, LAP, Betalactone, thioamitides, hserlactone, linerdin, nucleoside, PKS-like, ectoine (24).

1.4 Halophiles

The organisms growing under extreme conditions are better than any other organism living on Earth. Halophiles are the organisms represented by archaea, eukarya, and bacteria for which the main character is their salinity requirement, halophilic “salt-loving”. Halophilic microorganisms constitute the natural microbial communities of hypersaline ecosystems, which are widely distributed around the world. The general features of halophilic microorganisms are low nutritional requirements and resistance to high concentrations of salt with the capacity to balance the osmotic pressure of the environment (26). These halophiles use a variety of energy sources and can be aerobic or anaerobic halophiles including phototropic, fermentative, sulfate-reducing, and methanogenic species (27,28).

1.5 Halophilic Archaea

Extreme halophilic archaea are a diverse group of eury-archaeota that inhabit hypersaline environments such as salt lakes, ponds, and marine salterns. They are often referred to as “halobacteria” (29).

These diverse halophilic archaeal strains are precious resources for theoretical and applied research but assessing their taxonomic and metabolic novelty and diversity in natural environments has been technically difficult up until recently. As a group, the halophilic archaea (class Halobacteria) are the most salt-requiring and salt-resistant microorganisms within the domain Archaea. Halophilic archaea are flourish in Thalasso haline and athalassohaline environments and require over 100–150 g/L NaCl for growth and structural stability (30). They have been identified in the human

microbiome (31). The metabolic diversity of halophilic archaea has not yet been investigated at the genomic level by metabolic reactions in archaeal genomes and has limited reconstruction of metabolic pathways so far. Archaea also have huge economic potential that to date has only been fully realized in the production of thermostable polymerases. Archaea have furnished us with key paradigms for understanding fundamentally conserved processes across all domains of life (32).

CHAPTER 2

REVIEW OF LITERATURE

2.1 Importance of Halophilic Archaea

Natural products have helped us fight pathogenic bacteria and archaea for more than 50 years (2). The pharmaceutical, health food, and cosmetics businesses are today very interested in natural products and their biological activity, and there are a rising number of scientific research in this area. The majority of studies on the biological effects of natural products concentrate on different plant species. Most naturally occurring substances are beneficial for treating conditions that have anticancer, immune-stimulating, anti-inflammatory,

antioxidant, neuroprotective, and hepatoprotective effects. For thousands of years, people have employed natural resources like plants, animals, and microorganisms for medical purposes.

Bioactive substances found in a variety of natural goods have therapeutic benefits and can aid in the treatment of a number of ailments. For instance, a variety of illnesses, including fever, pain, inflammation, and infections, have been treated with the help of herbs in traditional medicine. Numerous substances produced by plants have been identified, and their medicinal potential has been investigated. (39) Since there are still unmet medical needs, a remarkable diversity of structures and activities, utility as biochemical probes, novel and sensitive assay methods, advancements in isolation, purification, and characterization, and new production techniques, there is a need for more natural product research (40).

Most of the active components in medicines come from natural sources. More than 80% of pharmacological compounds in the 'olden days' before high-throughput screening and the post-genomic era were natural products or inspired by a natural chemical. Natural products are substances that are derived naturally from bacteria, plants, or animals. In many cases, the natural product acts as a lead molecule whose activity can be modified via combinatorial and synthetic chemistry. Natural products have traditionally been the pathfinder molecules because they provide an unfathomable diversity of chemical structures that is unmatched by even the greatest combinatorial databases (42,43).

2.2 Biosynthetic Gene Clusters and Classes

A BGC can be defined as a physically clustered group of two or more genes in a particular genome that together encodes a biosynthetic pathway for the production of a specialized metabolite. Living organisms produce a range of secondary metabolites with exotic chemical structures and diverse metabolic origins. Many of these secondary metabolites find use as natural products in medicine, agriculture, and manufacturing. Research on natural product biosynthesis is undergoing an extensive transformation, driven by technological developments in genomics, bioinformatics, analytical chemistry, and synthetic biology. (24) Generally, bioactive compounds are synthesized by the cluster of genes which is referred to as biosynthetic gene cluster (BGC) Each class of bioactive substance is produced by a unique group of genes (cluster) For instance, non-ribosomal peptides are synthesized by non-ribosomal synthetases (NRPS), while polyketides are synthesized by polyketide synthases (PKS) Polyketide synthases (PKSs), non-ribosomal peptide synthetases (NRPSs), PKS-NRPS hybrids, bacteriocins, siderophores, terpenes, melanin, ectoine, lasso peptide, lantipeptide, thiopeptide, and butyrolactone are just a few of the different groups of BGCs that may be found in bacterial genomes (25).

2.3 Tools and Databases for Extraction of Biosynthetic Gene Clusters

The discovery of novel metabolic pathways is being enabled by the increasing availability of high-quality genome sequencing coupled with the development of powerful computational toolkits to identify metabolic gene clusters. To provide a comprehensive overview of various bioinformatics methods for detecting gene clusters, we compare and contrast key aspects of algorithmic logic behind several computational tools, including,

- **AntiSMASH:** This is a widely used web-based tool that automatically identifies and annotates BGCs in bacterial and fungal genomes. It also provides information on the predicted chemical structures of the compounds that the BGCs produce.
- **MIBiG:** The Minimum Information about a Biosynthetic Gene cluster (MIBiG) database is a curated resource that contains information on experimentally

characterized BGCs from bacterial and fungal genomes. It provides annotations, sequence data, and other metadata for each BGC.

- **PRISM:** This is a tool for predicting and identifying BGCs in metagenomic datasets. It uses a machine-learning approach to identify BGCs based on their genomic context and other features.
- **NaPDoS:** The Natural Product Domain Seeker (NaPDoS) is a database of protein domains that are involved in the biosynthesis of natural products. It can be used to identify BGCs based on the presence of specific domains in the genome.
- **BiG-SCAPE:** This is a tool for clustering and analyzing BGCs based on their similarity. It can be used to identify BGCs that produce similar compounds and to predict the chemical structures of unknown compounds based on the structures of known compounds.
- **NP searcher:** NP.searcher is a bioinformatics tool used for the identification of biosynthetic gene clusters (BGCs) involved in the biosynthesis of natural products. It uses a combination of hidden Markov models (HMMs) and phylogenetic analysis to identify BGCs and predict the chemical structures of the natural products they produce.
- **Clust scan:** ClustScan is a bioinformatics tool used for the identification and analysis of biosynthetic gene clusters (BGCs) in microbial genomes. It uses a combination of sequence analysis and pattern recognition algorithms to identify BGCs that are involved in the biosynthesis of natural products.
- **MIDDAS-M:** MIDDAS-M (Microbial-Derived Drug Discovery Automation System - Metabolomics) is a bioinformatics tool used for the automated identification and characterization of natural products from microbial genomes. It integrates multiple omics data, such as genomics, transcriptomics, proteomics, and metabolomics, to predict the biosynthetic gene clusters (BGCs) and chemical structures of natural products (45).

Table 2.1 Databases and tools for biosynthetic gene cluster mining

AntiSMASH	https://antismash.secondarymetabolites.org/#!/start
MIBiG	https://mibig.secondarymetabolites.org/
PRISM	https://prism-em.github.io/
NaPDoS	https://napdos.bio/
BiG-SCAPE	https://git.wageningenur.nl/medema-group/BiG-SCAPE
NP.searcher	https://www.npsb.univ-lille.fr/NP.searcher/
Clust scan	https://sourceforge.net/projects/clustscan/
MIDDAS-M	https://idda1.hku.hk/MIDDAS-M/

2.4 Benefits of Halophiles

Halophiles are organisms that can endure high salinity and are either extremophiles or extremotolerant. Depending on their maximal salinity tolerance, they are divided into three categories: minor, moderate, and excessive. Halophilic creatures are found in all realms of life, and they exhibit a great range of metabolic processes and physiological reactions, particularly in bacteria, archaea, and eukarya. The majority of halophile studies have concentrated on understanding the unique adaptations and molecular processes that allow them to maintain their osmotic balance in the presence of salt stress. As the vast majority of them belong to ancient evolutionary lineages, there has also been a lot of interest focused on the investigation of their diversity and phylogenetic relationships. Halophiles have been the subject of much scientific research due to their unique adaptations to extreme environments. They are of particular interest in the field of astrobiology, as they provide a model for how life may adapt to extreme conditions on other planets. Halophiles have been a valuable resource for drug discovery and design some halophiles produce compounds with antibiotic properties that are effective against drug-resistant bacteria. Halophilic compounds have many potential applications in drug discovery, especially in the search for new bioactive compounds. Halophiles are known to produce a wide range of secondary metabolites, including antimicrobial peptides, enzymes and pigments with potential therapeutic applications. Designed to help

halophiles survive in harsh environments, these compounds may exhibit unique chemical structures and biological activities that make them attractive targets for drug development (47). These compounds are of great interest to researchers looking for new ways to combat antibiotic resistance. In addition to antibiotic resistance, halophiles have been found to produce a range of enzymes that have potential applications in biotechnology, such as in the production of biofuels or in the processing of food and agricultural products (33).

One example of a bioactive compound produced by halophiles is halocins, which are antimicrobial peptides that have been shown to have activity against a wide range of bacteria and archaea, including multidrug-resistant pathogens. Halocins have potential applications in the development of new antibiotics and other antimicrobial agents (48).

2.5 Halodom Database

Halodom is a database containing information on over 1000 halophilic species from all three domains of life: archaea, bacteria and eukaryotes. The database contains basic information such as the found salinity of a particular organism, its taxonomy, and genomic information via NCBI and other links. HaloDom's data structures are organized into multiple tables to improve database speed and efficiency for user queries and reduce redundancy. The database is (<http://halodom.bio.auth.gr/?view=home>) contains everything useful for researchers, from genome and sequence data to environmental details and metabolic pathways (33).

2.6 Halophilic Archaea and Environments

Extreme halophilic archaea are a diverse group of eury-archaeota that inhabit hypersaline environments such as salt lakes, ponds, and marine salterns. They are often referred to as “halobacteria”. Metabolic reconstruction and comparison analysis have not yet been used to explore the metabolic diversity of halophilic archaea at the genome level. Reconstruction of metabolic pathways has been hampered thus far by the absence of enzyme genes for a number of metabolic activities in archaeal genomes. However, with the recent discovery of novel non-orthologous enzymes in archaea, many of these route gaps have been clarified. (29) Haloarchaea are isolated from hypersaline ecosystems where high salinity

approaches saturation (1.5–5 M salinity), including environments with extreme pH and/or temperature. The unique properties of Haloarchaea and their enzymes that allow them to maintain catalytic activity in these environments make them attractive resources for use in biotransformation processes that need to occur under various industrial conditions. Haloarchaea-derived biocatalysts (enzymes) occupy a unique niche in the organic solvent, salt, and detergent industries. Halophilic microorganisms are producers of antimicrobials, bioplastics, biofuel, extremozymes, extracellular polysaccharides, retinal proteins (bacteriorhodopsin), colored pigments and compatible solutes which has led to their application in agriculture, food and nutrition, biofuel and many more industrial processes such as treatment of saline wastewaters.

These microorganisms live in hypersaline environments such as the Dead Sea, hypersaline lakes, salt pans, saline soils, and salt marshes where the salt concentrations are generally at 3.5% dissolved NaCl. In addition to their stability in high salt, halophilic microorganisms can withstand other harsh conditions such as extremes in pH, temperature, and organic solvent. Based on their optimum growth parameters, halophiles are categorized as extreme at 2.5–5.2 M NaCl, moderate at 0.5–2.5 M, mild at 0.2–0.5 M and are considered extremely halotolerant when they do not require high concentrations (>2.5 M) of salt but grow well under these conditions. Adaptation of halophiles to withstand the adverse consequences when exposed to hypersaline conditions is based on osmoadaptation strategy. Haloarchaea are a target of the scientific community, as these microorganisms produce enzymes, exopolysaccharides, and pigments that may be applied in industrial biotechnology.

Due to the salt-in-cytoplasm strategy, haloarchaeal enzymes/biocatalysts are adapted to remain stable and catalytically active at high ionic strength. As hypersaline and organic solvent conditions share common properties including low water activity, haloarchaeal proteins are reported to tolerate organic solvents. Haloarchaeal enzymes, such as cellulases, xylanases, pectinases, esterases, lipases, catalases, chitinase, amylases, β -galactosidase, β -glucosidases, and laccases, are useful for biomass conversion and herein reviewed. Haloarchaea is also found to glycosylate certain extracellular and surface structural proteins that may facilitate the stability of these proteins in harsh conditions.

Thus, haloenzymes have the potential to enhance the production of renewable fuels and chemicals. This review will focus on the structural components of the various types of biomass, and the role of extremophiles, particularly those of haloarchaea, in generating products and/or enzymes that may be used for biofuels and chemical production (49).

CHAPTER 3

AIM AND SCOPE

3.1 Aim

The main goal of this project is to investigate the biosynthetic gene clusters in archaea using *In silico* genome mining techniques. By analyzing available genomic data, we aim to identify novel biosynthetic gene clusters that could potentially produce bioactive compounds with medicinal and biotechnological applications. Archaea are microorganisms that inhabit extreme environments and are known to produce unique and diverse bioactive compounds that are different from those produced by bacteria and fungi. Therefore, this study is important as it aims to discover and explore novel biosynthetic gene clusters in archaea, which could lead to the development of new drugs and biotechnological products. Furthermore, the findings of this project could also provide valuable insights into the adaptation mechanisms of archaea to extreme environments.

3.2 Scope

The scope of this project involves investigating the biotechnological and medicinal potential of archaea and their biosynthetic gene clusters. Our focus is to identify and analyze these gene clusters in archaea, exploring their diversity and distribution and assessing the bioactive compounds that they could produce. To achieve these objectives, we will use bioinformatics tools and techniques, conduct a thorough review of the literature and analyze the data. The scope of this project involves investigating the biotechnological and medicinal potential of archaea and their biosynthetic gene clusters. Our focus is to identify and analyze these gene clusters in archaea, exploring their diversity and distribution and assessing the bioactive compounds that they could produce. To achieve these objectives, we will use bioinformatics tools and techniques, conduct a thorough review of the literature and analyze the data. Working jointly with natural product chemistry specialists will be essential to validate our results and recognize any promising bioactive compounds. The potential discovery of novel natural products through this project could have significant implications for the pharmaceutical and biotechnology industries, leading to the development of new drugs and biotechnological applications.

CHAPTER 4

MATERIALS AND METHODS

4.1 Mining of Halophilic Archaea Selection Using Halodom Database

Based on the information available in the Halodom database, the data was curated in Excel and separated into bacteria, fungi, and archaea. The sorting and filtering of the data were done in Microsoft Excel. Parameters like genome availability, phylum, class, domain, salinity range, genome availability, and halotolerant classification were analysed in Microsoft Excel.

- Halodom Database: Halodom database is an online database containing more than 1000 halophilic species from all life domains. (33) Archaeal genomes were mined using Halodom, a website that has collected and curated 1267 halophilic species across all life domains so far.
- There are 282 Archaea, 669 Bacteria, and 314 Eukarya currently registered in the Halodom database.

This database is specifically designed to study halophilic microorganisms, which are bacteria, fungi, and archaea that require high salt concentrations for their growth and survival. It provides detailed information about the genome sequences, taxonomy, physiological characteristics, and environmental adaptations of these unique microorganisms. Researchers can use the Halodom database to identify and study halophiles that have potential biotechnological and biomedical applications, such as desalination, bioremediation, and antibiotic production.

4.2 Genome Import from The NCBI Genome

The availability of genome sequences in Halodom was checked. The available genome was linked directly to the NCBI genome website. The corresponding genome for the selected strain isolates was downloaded from the NCBI Genome Database. The NCBI Genome database contains a wide range of information about genomes, including genomic sequences, annotations, and functional information. It includes complete genome sequences for a variety of organisms, as well as draft sequences and genome assemblies. The sequence was downloaded in the GenBank file in GBFF format.

4.3 Identification of Biosynthetic Gene Clusters in Halophilic Archaea

antismash was performed with the input given as GenBank files to extract biosynthetic gene clusters present in the archeal genomes. AntiSMASH (Antibiotics & Secondary Metabolite Analysis Shell) is a bioinformatics tool used for the prediction and analysis of secondary metabolite biosynthesis gene clusters in microbial genomes) For extraction of the Biosynthetic Gene Clusters (A BGC can be defined as a physically clustered group of two or more genes in a particular genome that together encodes a biosynthetic pathway for the production of a specialized metabolite) (24). GenBank files of the archeal halophiles were uploaded onto the AntiSMASH server (<https://antismash.secondarymetabolites.org/#!/start>) Then the relax detection was allowed; not loose or not strict which helps in detecting well-defined clusters containing all required parts we selected all extra features embedded in the website such as KnownClusterBlast, ClusterBlast, SubClusterBlast, MIBiG cluster comparison, ActiveSiteFinder, RREFinder, Cluster Pfam analysis, Pfam-based GO term annotation, TIGRFam analysis, and TFBS analysis. Known cluster blast and MIBiG cluster comparison are methods for identifying and comparing biosynthetic gene clusters involved in the production of natural products such as antibiotics and other secondary metabolites. TFBS analysis involves identifying transcription factor binding sites within DNA sequences, while active site finder identifies active sites within protein structures. Pfam-based GO term annotation assigns Gene Ontology (GO) terms to protein domains based on their known functions, while subcluster blast and RREFinder help to identify subclusters within gene clusters. Finally, TIGRFam analysis is a method for identifying protein families and domains based on their evolutionary relationships. There are other tools instead of the AntiSMASH, like MIBiG (Minimum Information about Biosynthetic Gene Clusters), PRISM (prediction informatics for secondary metabolites), ClustScan (it is a software tool that uses Hidden Markov Models (HMMs) to identify the secondary metabolites), BAGLE4 (BACTERIOCIN Genome mining to Enhance bacterial diversity), SMURF (secondary Metabolite Unknown Regions Finder) These are the alternative tools for the antiSMASH.

4.4 Manual Interpretation of The Dataset

There are a total of 282 archaeal genomes available in the halodom database in we selected 237 archaeal genomes whose sequence was available we extracted the secondary metabolites using AntiSMASH and 16 types of the BGCs were observed.

4.5 Phylogenetic Tree Construction Using MEGA

For the construction of the phylogenetic tree, the 16S rRNA gene was downloaded from the NCBI GenBank These 16srRNA sequences were aligned using a tool called Bioedit. Bioedit helps in the alignment and manipulation of the sequences by allowing the user to trim unnecessary bases. Multiple sequence alignment using the cluster w algorithm was also done. For the phylogenetic tree construction in MEGA, we did the neighbourhood joining method and we set the bootstrap value to 1000 for the robustness of a phylogenetic tree. Then we got the NEWICK format file that was imported into ITOL for the visualization.

4.6 Visualization of The Phylogenetic Tree In ITOL

The interactive Tree of Life (<https://itol.embl.de/>) is an online tool for the display, manipulation, and annotation of the phylogenetic tree. ITOL provides the most common functions available in the phylogenetic tree viewer. The tree display is implemented in JavaScript and uses an HTML5 canvas element for visualization. The control panel of ITOL provides a Basic and advanced visualization setting, with the mode being rectangular, circular, or unrooted. The branches can be coloured and labelled with different styles and fonts. Advanced settings include assignment of taxonomy, display of bootstrap data, and so on. To visualize a phylogenetic tree on ITOL one must first upload the annotation files in the datasets column that may include the colours and the charts on the tree. The NEWICK file must be uploaded to construct the tree, and ITOL takes plain text format only. ITOL also has an export column that exports the constructed tree in SVG, PNG, JPG, and PDF formats (34).

CHAPTER 5

RESULTS AND DISCUSSION

5.1 Results

We used genome sequences for 237 different halophilic archaea to investigate their biosynthetic potential. Firstly the genomes sequences of halophilic archaea available in halodom were mined and used antismash 6.0 to extract predicted secondary metabolites. Although prior investigations have queried diverse approaches to extract BGCs, antismash is regarded as the golden standard tool and has prevalent use for the same. Out of 237 genomes, 72% of organisms were extremely tolerant 27% of organisms were moderately tolerant and 1% of organisms were slightly tolerance Fig 5.1 shows the statistical chart of the tolerance level in halophilic archaea.

In order to understand the relationship between the halophilic archaea, a phylogenetic tree was constructed using MEGA11. The interactive tree of slife (ITOL) was used to illustrate the phylogenetic tree. The interactive tree of life of halophilic archaea in Fig 5.2

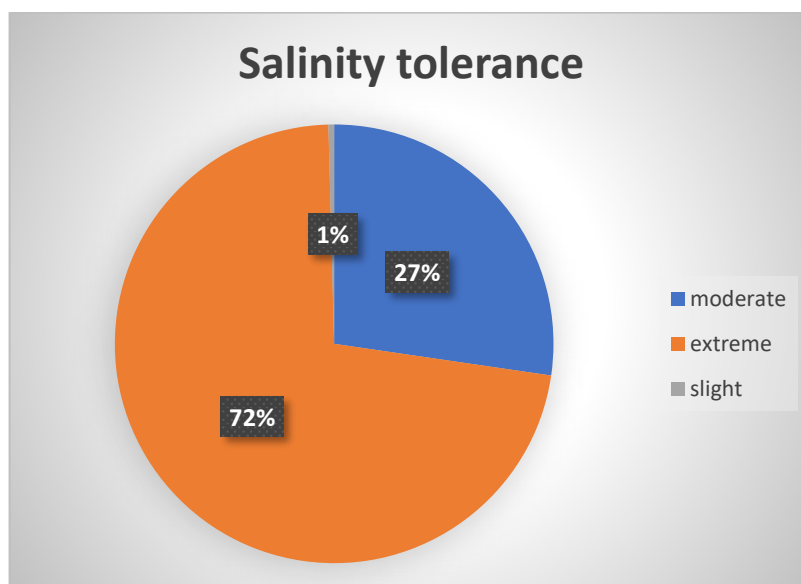


Fig 5.1 Tolerance level

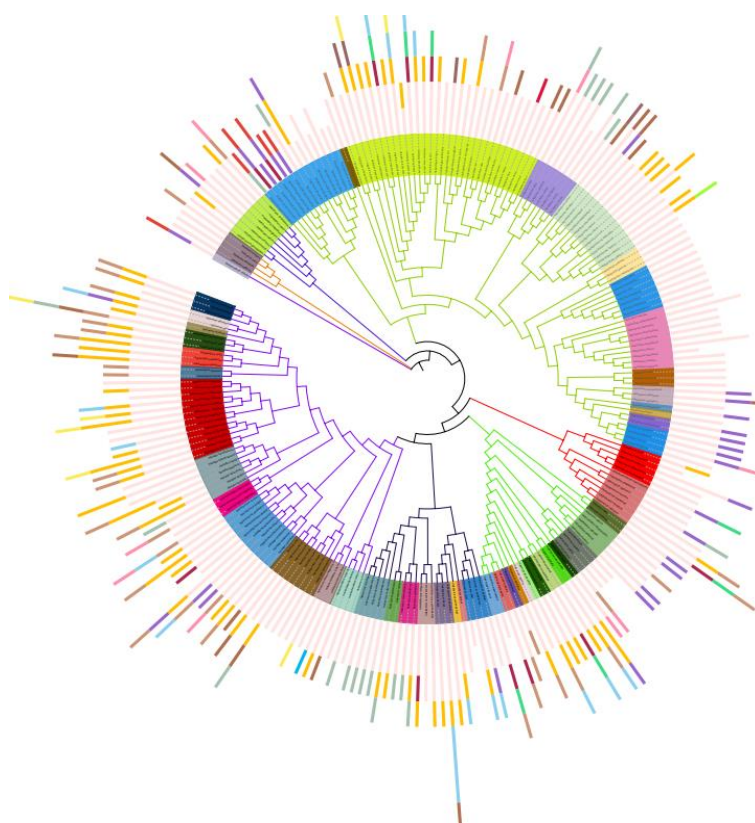


Fig 5. Phylogenetic tree construction using (ITOL).

- TERPENE
- RiPP-like
- Siderophore
- Thiopeptide
- NRPS- independent siderophore
- RRE-containing
- Lanthipeptide-class-ii
- Lasso peptide
- Ranthipeptide
- LAP
- Betalactone
- Thioamitides
- Hserlactone
- Linaridin
- Nucleoside
- PKS-like
- ectoine

In this whole tree there 237 genomes are there and there are 8 major clades and many subclades there the phylogenetic tree of archaea (as mentioned in the above figure 2) suggested that 2 ancestral groups occupy clear evolutionary positions at

the root of the tree. There are majorly 16 BGC classes. (Shown in Fig.3, Fig.4) Results show that halophilic archaea carry about 1 to 8 BGCs per genome with Mean=3.033 and standard deviation =1.157. Among the 237 archeal species the highest number of BGCs were occupied in 3 positions, the first being Haladaptatus cibarius (8), Haloterrigena saccharevitans (7), 6 BGCs are present in Halobacterium salinarum, Halobiforma_haloterrestris .5 BGCs are present in 22 genomes.4 BGCs are present in 48 genomes.3 BGCs are present in 77 genomes.2 BGCs are present in 4 genomes.1 BGCs are present in 12 genomes. The lowest number of BGCs present in Methanohalophilus profundus has only 1 BGCs Thiopeptide. The highest number of BGC classes observed was Terpenes. Some terpenes, such as beta-caryophyllene and humulene, have analgesic properties and are used to relieve pain and in the industrial Terpenes are commonly used as fragrance and flavour compounds in the perfume and food industries due to their strong and pleasant aroma.

BGC class	number of BGCs
Nucleoside	1
PKS-like	1
ectoine	1
Hserlactone	4
Ranthipeptide	5
NRPS- independent siderophore	7
Thioamitides	7
Linaridin	7
Lasso peptide	10
RiPP-like	11
LAP	11
Lanthipeptide-class-ii	18
Betalactone	21
Thiopeptide	31
RRE-containing	36
Siderophore	83
TERPENE	468

Fig 5.3 Major classes of BGCs observed

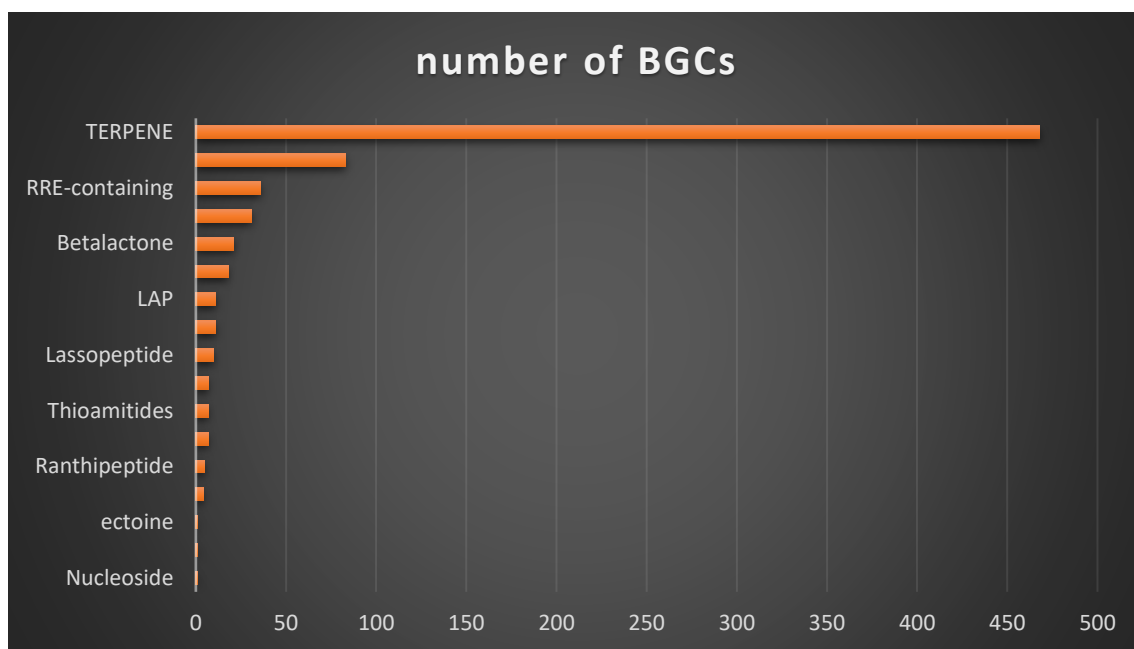


Fig 5.4 Highest number of BGCs



Fig 5.5 Prediction of BGCs – process (44)

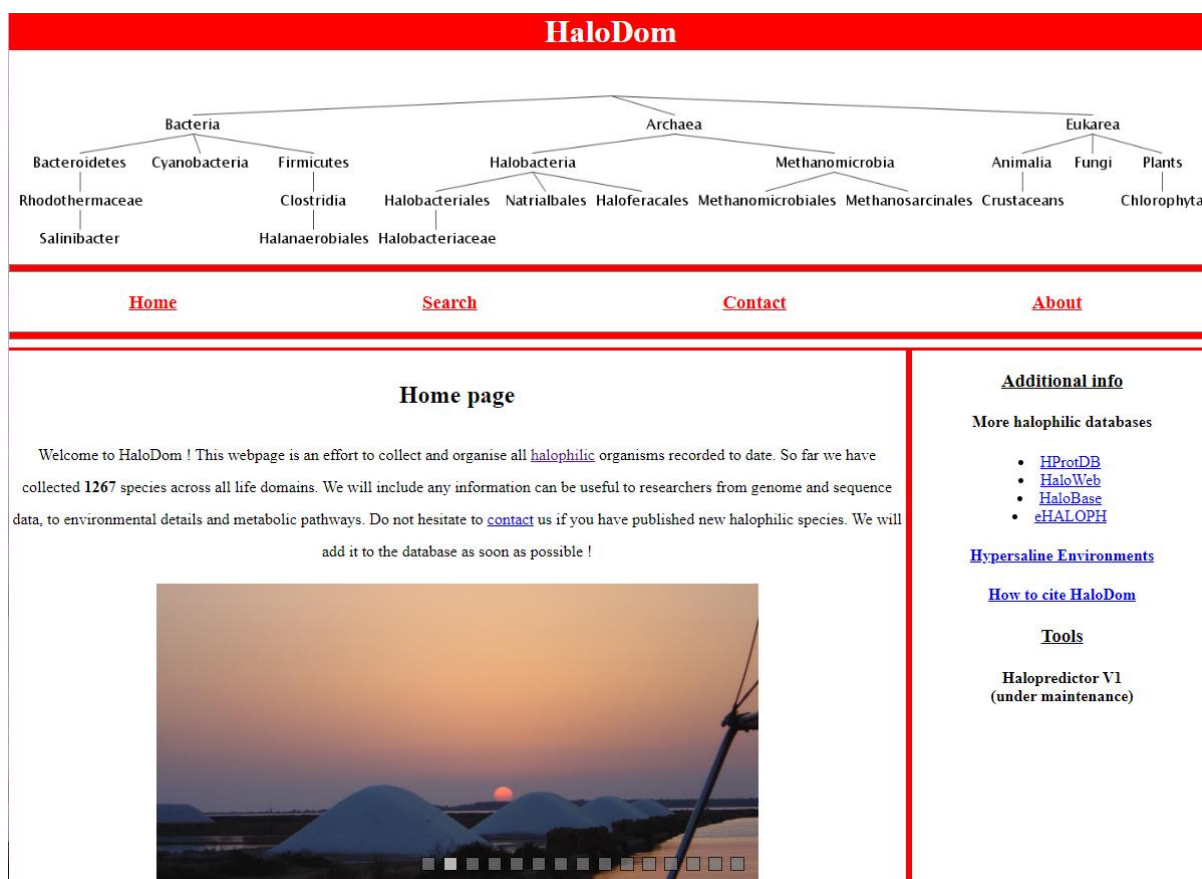


Fig.6. Home page of Halodom database (33)

5.2 Discussion

The secondary metabolite discovery has been limited by many factors. A number of applications of natural products are being used in various sectors of life. The secondary metabolites synthesized by halophilic archaea have diverse potential applications in pharmaceuticals, bioremediation, food, agriculture, and cosmetics. In complete 237 genomes. The highest BGC are terpenes. Terpenes are the largest and most diverse PSMs in nature. And are informative and defensive vehicles used by plants for antagonistic and mutualistic interactions. The production of terpenes by plants to counter biotic (pathogenic microbes, herbivore pests, weeds) and abiotic (water, temperature, light, and salt) stresses is widely studied. While phytoanticipins terpenes are constitutively secreted in absence of pathogen effectors, phytoalexins are elicited in response to inducible pathogenic microbes or feeding herbivores (37). Terpenes constitute a major class of organic molecules produced by diverse ecological contexts.

Although all terpenes originate from the same five-carbon building blocks the structures and functions of terpenes vary widely, and are highly tailored to the requirements of the source organism's environmental pressures and resources. (38). In drug discovery and development, terpenes have been found to have a wide range of biological activities, including anti-inflammatory, analgesic, and antitumor properties. They can also enhance the absorption of other drugs, making them potentially useful as drug delivery agents. One way in which terpenes can help in drug discovery is by serving as a starting point for the development of new drugs. Researchers can isolate and study the chemical structure and biological activity of natural terpenes, and use this information to design and synthesize new compounds with similar properties. Terpenes can also be used as lead compounds in drug discovery, meaning that they can be modified or optimized to create new drugs with improved efficacy and fewer side effects. This approach has been successful in the development of drugs such as paclitaxel, a cancer treatment that is derived from the Pacific yew tree.

We observed a unique BGC in *Methanohalophilus profundus* the BGC is thiopeptide. A novel anaerobic methylotrophic halophilic methanogen strain "SLHTYRO" was isolated from a deep hypersaline anoxic basin called "Tyro" located in the Eastern Mediterranean sea. Cells of "SLYTHYRO" were motile cocci. The strain SLYTHYRO grew between 12 and 37°C (35).

Thiopeptides are members of the ribosomally synthesized and post-translationally modified peptides (RiPP) family of natural products. Most characterized thiopeptides display nanomolar potency towards Gram-positive bacteria by blocking protein translation with several being produced at the industrial scale for veterinary and livestock applications (46). The thiopeptide antibiotics are naturally occurring, sulfur-containing, highly modified, macrocyclic peptides, nearly all of which inhibit protein synthesis in bacteria, and archaea (47).

Thiopeptides (also known as thiazolyl peptides) are structurally complex natural products with rich biological activities. Known for over 70 years for the potent killing of Gram-positive bacteria, thiopeptides are experiencing a resurgence of interest in the last decade, primarily brought about by the genomic revolution of the 21st century. Every area of thiopeptide research from elucidating their biological function

and biosynthesis to expanding their structural diversity through genome mining—has made great strides in recent years. These advances lay the foundation for and inspire novel strategies for thiopeptide engineering. Accordingly, a number of diverse approaches are being actively pursued in the hope of developing the next generation of natural-product-inspired therapeutics. Here, we review the contemporary understanding of thiopeptide biological activities, biosynthetic pathways, and approaches to structural and functional reprogramming, with a special focus on the latter (36).

CHAPTER 6

SUMMARY AND CONCLUSION

Halophilic archaea are a class of extremophilic microorganisms that exhibit a proclivity for survival in environments possessing elevated salt concentrations, such as saline flats and lakes. They belong to the Archaea kingdom and are divided into various genera, such as *Haloferax*, *Halobacterium*, and *Halococcus*. These organisms are recognized for their production of a broad spectrum of metabolites, including antimicrobial agents, enzymes, and pigments, that have promising potential in the domains of biotechnology and pharmaceuticals. The process of extracting biosynthetic gene clusters (BGCs) from halophilic archaea is a means to reveal these valuable compounds. BGC extraction involves the isolation of DNA from the microorganisms, sequencing the genetic material, and scrutinizing the resultant data to identify potential BGCs. One of the techniques employed for BGC extraction is genome mining, which involves seeking BGCs in the microorganism's genome. Genome mining has been fruitful in detecting BGCs in halophilic archaea that are responsible for producing the halovirs A and B antibiotics. Halophilic archaea, which are microorganisms that flourish in high-salt environments, have the ability to produce valuable metabolites with potential uses in the fields of biotechnology and pharmaceuticals. The process of biosynthetic gene cluster (BGC) extraction from these organisms allows for the discovery of such compounds. The approaches of genome mining and metagenomics have proved successful in identifying possible BGCs in halophilic archaea. Ultimately, the exploration of halophilic archaea and the extraction of BGCs may lead to various benefits in the domains of medicine and industry. Metagenomics is another approach employed for BGC extraction that involves scrutinizing the DNA extracted from environmental samples containing halophilic archaea. This technique has also been utilized to identify BGCs in these microorganisms.

REFERENCES

1. Ekiert HM, Szopa A. Biological Activities of Natural Products. *Molecules*. 2020 Dec 7;25(23):5769. doi: 10.3390/molecules25235769. PMID: 33297511; PMCID: PMC7730830.
2. Zhang, Lixin; Demain, Arnold L. (2005). *Natural Products* || . , 10.1007/978-1-59259-976-9(), -. doi:10.1007/978-1-59259-976-9
3. Brahmachari G (2010). *Handbook of Pharmaceutical Natural Products*. Weinheim: Wiley-VCH. ISBN 978-3-52732148-3.
4. Beghyn T, Deprez-Poulain R, Willand N, Folleas B, Deprez B (July 2008). "Natural compounds: leads or ideas? Bioinspired molecules for drug discovery". *Chemical Biology & Drug Design*. 72 (1): 3–15. doi:10.1111/j.1747-0285.2008.00673.x. PMID 18554253. S2CID 20973633
5. Koehn FE, Carter GT (March 2005). "The evolving role of natural products in drug discovery". *Nature Reviews. Drug Discovery*. 4 (3): 206–20. doi:10.1038/nrd1657. PMID 15729362. S2CID 32749678.
6. Alan L. Harvey, *Natural products in drug discovery*, *Drug Discovery Today*, Volume 13, Issues 19–20, 2008, Pages 894–901, ISSN 1359-6446,
<https://doi.org/10.1016/j.drudis.2008.07.004>.
7. Strobel G, Daisy B (December 2003). "Bioprospecting for microbial endophytes and their natural products". *Microbiology and Molecular Biology Reviews*. 67 (4): 491–502. doi:10.1128/MMBR.67.4.491-502.2003. PMC 309047. PMID 14665674.
8. Cushnie TP, Cushnie B, Echeverría J, Fowsantear W, Thammawat S, Dodgson JL, Law S, Clow SM (June 2020). "Bioprospecting for antibacterial drugs: a multidisciplinary perspective on natural product source material, bioassay selection and avoidable pitfalls". *Pharmaceutical Research*. 37 (7): Article 125. doi:10.1007/s11095-020-02849-1. PMID 32529587. S2CID 219590658.
9. Harvey AL (October 2008). "Natural products in drug discovery". *Drug Discovery Today*. 13 (19–20): 894–901. doi:10.1016/j.drudis.2008.07.004. PMID 18691670

10. Abdelmohsen UR, Sayed AM and Elmaidomy AH (2022) Natural Products' Extraction and Isolation Between Conventional and Modern Techniques. *Front. Nat. Products*. 1:873808. doi: 10.3389/fntpr.2022.873808
11. Rateb ME, Abdelmohsen UR. Bioactive Natural Products from the Red Sea. *Mar Drugs*. 2021 May 21;19(6):289. doi: 10.3390/md19060289. PMID: 34064008; PMCID: PMC8224067.
12. Huo JL, Fu WJ, Liu ZH, Lu N, Jia XQ, Liu ZS. Research advance of natural products in tumor immunotherapy. *Front Immunol*. 2022 Sep 2;13:972345. doi: 10.3389/fimmu.2022.972345. PMID: 36159787; PMCID: PMC9494295.
13. Rao AVR, Gurjar MK. Drugs from plant resources: An overview. *Pharmatimes*. 1990;22(5):19-20
14. Farnsworth NR. A computerized data base for medicinal plants. *The Eastern Pharmacist*. 1985;28(326):53-55
15. C. M. Dobson *Nature*, 2004, 432 , 824 —828 CrossRef CAS PubMed.
16. Y. Chen , M. Garcia de Lomana , N.-O. Friedrich and J. Kirchmair , *J. Chem. Inf. Model.*, 2018, 58 , 1518 —1532 CrossRef CAS PubMed .
17. Ben Shen, A New Golden Age of Natural Products Drug Discovery, *Cell*, Volume 163, Issue 6, 2015, Pages 1297-1300, ISSN 0092-8674, <https://doi.org/10.1016/j.cell.2015.11.031>. (<https://www.sciencedirect.com/science/article/pii/S0092867415015500>)
18. Harvey, A., Edrada-Ebel, R. & Quinn, R. The re-emergence of natural products for drug discovery in the genomics era. *Nat Rev Drug Discov* 14, 111–129 (2015). <https://doi.org/10.1038/nrd4510>
19. M.F. Balandrin, A.D. Kinghorn and N.R. Farnsworth. In *Human Medicinal Agents from Plants* (A.D. Kinghorn and M.F. Balandrin, eds.), Symposium Series No. 534, American Chemical Society, Washington, D.C. 2-12 (1993).

- 20.Katz L, Baltz RH. Natural product discovery: past, present, and future. *J Ind Microbiol Biotechnol*. 2016 Mar;43(2-3):155-76. doi: 10.1007/s10295-015-1723-5. Epub 2016 Jan 6. PMID: 26739136.
- 21.Corson, T. W. & Crews, C. M. Molecular understanding and modern application of traditional medicines: triumphs and trials. *Cell* 130, 769–774 (2007)
- 22."Plant Physiology|Digital Textbook Library". www.tankonyvtar.hu. Retrieved 2016-03-16.
- 23.Pagare, Saurabh & Bhatia, Manila & Tripathi, Niraj & Bansal, Y.K.. (2015). Secondary metabolites of plants and their role: Overview. *Current Trends in Biotechnology and Pharmacy*. 9. 293-304.
- 24..Medema, M., Kottmann, R., Yilmaz, P. et al. Minimum Information about a Biosynthetic Gene cluster. *Nat Chem Biol* 11, 625–631 (2015). cc
- 25.Nadarajan Viju;Stanislaus Mary Josephine Punitha;Sathianeson Satheesh; (2021). An Analysis of Biosynthesis Gene Clusters and Bioactivity of Marine Bacterial Symbionts . *Current Microbiology*, (), –. doi:10.1007/s00284-021-02535-4
- 26.Corral P, Amoozegar MA, Ventosa A. Halophiles and Their Biomolecules: Recent Advances and Future Applications in Biomedicine. *Mar Drugs*. 2019 Dec 30;18(1):33. doi: 10.3390/md18010033. PMID: 31906001; PMCID: PMC7024382.
- 27.Ollivier B, Caumette P, Garcia JL, Mah RA (March 1994). "Anaerobic bacteria from hypersaline environments". *Microbiological Reviews*. 58 (1): 27–38. doi:10.1128/MMBR.58.1.27-38.1994. PMC 372951. PMID 8177169
- 28.Oren A (January 2002). "Diversity of halophilic microorganisms: environments, phylogeny, physiology, and applications". *Journal of Industrial Microbiology & Biotechnology*. 28 (1): 56–63. doi:10.1038/sj/jim/7000176. PMID 11938472. S2CID 24223243
- 29.Falb, M., Müller, K., Königsmaier, L. et al. Metabolism of halophilic archaea. *Extremophiles* 12, 177–196 (2008). <https://doi.org/10.1007/s00792-008-0138-x>

- 30.Cui, HL., Dyal-Smith, M.L. Cultivation of halophilic archaea (class Halobacteria) from thalassohaline and athalassohaline environments. *Mar Life Sci Technol* 3, 243–251 (2021). <https://doi.org/10.1007/s42995-020-00087-3>
- 31.Krawczyk KT, Loch C, Kowalewicz-Kulbat M. Halophilic Archaea Halorhabdus Rudnickae and Natrinema Salaciae Activate Human Dendritic Cells and Orient T Helper Cell Responses. *Front Immunol.* 2022 May 26;13:833635. doi: 10.3389/fimmu.2022.833635. PMID: 35720372; PMCID: PMC9204267.
- 32.Jarrell KF, Walters AD, Bochiwal C, Borgia JM, Dickinson T, Chong JPJ. Major players on the microbial stage: why archaea are important. *Microbiology (Reading)*. 2011 Apr;157(Pt 4):919-936. doi: 10.1099/mic.0.047837-0. Epub 2011 Feb 17. PMID: 21330437.
- 33.Loukas A, Kappas I, Abatzopoulos TJ. HaloDom: a new database of halophiles across all life domains. *J Biol Res (Thessalon)*. 2018 Jan 15;25:2. doi: 10.1186/s40709-017-0072-0. PMID: 29796383; PMCID: PMC5957262.
- 34.Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 2021 Jul 2;49(W1):W293-W296. doi: 10.1093/nar/gkab301. PMID: 33885785; PMCID: PMC8265157.
35. L'Haridon S, Haroun H, Corre E, Roussel E, Chalopin M, Pignet P, Balière C, la Cono V, Jebbar M, Yakimov M, Toffin L. *Methanohalophilus profundus* sp. nov., a methylotrophic halophilic piezophilic methanogen isolated from a deep hypersaline anoxic basin. *Syst Appl Microbiol.* 2020 Sep;43(5):126107. doi: 10.1016/j.syapm.2020.126107. Epub 2020 Jun 27. PMID: 32847782.
36. Alexander A. Vinogradov, Hiroaki Suga, Introduction to Thiopeptides: Biological Activity, Biosynthesis, and Strategies for Functional Reprogramming, *Cell Chemical Biology*, Volume 27, Issue 8, 2020, Pages 1032-1051, ISSN 2451-9456,
- 37.Ninkuu V, Zhang L, Yan J, Fu Z, Yang T, Zeng H. Biochemistry of Terpenes and Recent Advances in Plant Protection. *Int J Mol Sci.* 2021 May 27;22(11):5710. doi: 10.3390/ijms22115710. PMID: 34071919; PMCID: PMC8199371.

- 38.Jansen DJ, Shenvi RA. Synthesis of medicinally relevant terpenes: reducing the cost and time of drug discovery. *Future Med Chem.* 2014 Jun;6(10):1127-48. doi: 10.4155/fmc.14.71. PMID: 25078134; PMCID: PMC4162426.
- 39.Subhawa, S.; Chewonarin, T.; Banjerdpongchai, R. The Effects of *Houttuynia cordata* Thunb. and *Piper ribesioides* Wall. Extracts on Breast Carcinoma Cell Proliferation, Migration, Invasion and Apoptosis. *Molecules* 2020,25, 1196.
- 40.Alvarez A, Granizo T. La organización de la información sobre biodiversidad: el Centro de datos para la conservación. In: Mena A, Suárez L (eds), *La investigación para la conservación de la diversidad biológica en el Ecuador*. 1993.
- 41.Sneader, W. (1996) *Drug Prototypes and Their Exploitation*. Wiley, UK
- 42.. Baker, D., U. Mocek, and C. Garr. 2000. Natural products vs. combinatorials: a case study, p. 66–72. In S. K. Wrigley, M. A. Hayes, R. Thomas, E. J. T.Chrysal, and N. Nicholson (ed.), *Biodiversity: new leads for pharmaceutical and agrochemical industries*. The Royal Society of Chemistry, Cambridge, United Kingdom
43. Grabley, S., and R. Thiericke (ed.). 1999. *Drug discovery from nature*, p.3–33. Springer-Verlag. Berlin, Germany.
- 44.for figure(Ren, Hengqian & Shi, Chengyou & Zhao, Huimin. (2019). Computational Tools for Discovering and Engineering Natural Product Biosynthetic Pathways. *iScience*. 23. 100795. 10.1016/j.isci.2019.100795.)
- 45.Chavali AK, Rhee SY. Bioinformatics tools for the identification of gene clusters that biosynthesize specialized metabolites. *Brief Bioinform.* 2018 Sep 28;19(5):1022-1034. doi: 10.1093/bib/bbx020. PMID: 28398567; PMCID: PMC6171489.
46. Schwalen, Christopher J.; Hudson, Graham A.; Kille, Bryce; Mitchell, Douglas A. (2018). *Bioinformatic expansion and discovery of thiopeptide antibiotics*. *Journal of the American Chemical Society*, (), *jacs.8b03896*–. doi:10.1021/jacs.8b03896

47. Singh, P., Kumari, K., & Verma, A. K. (2020). Halophiles and their potential applications in bioremediation and drug discovery. In *Environmental Biotechnology* (pp. 1-18). Springer. DOI: 10.1007/978-981-15-0112-2_1. PMID: 32617944.
48. Wang, G., Wei, Y., Chen, X., Zhang, Y., & Yang, Y. (2021). Halocins: diversity, biosynthesis, and applications. *Applied Microbiology and Biotechnology*, 105(3), 839-851. DOI: 10.1007/s00253-020-11075-y. PMID: 33289046.
49. Kasirajan L, Maupin-Furlow JA. Halophilic archaea and their potential to generate renewable fuels and chemicals. *Biotechnol Bioeng*. 2021 Mar;118(3):1066-1090. doi: 10.1002/bit.27639. Epub 2020 Dec 16. PMID: 33241850; PMCID: PMC7897274.