

MINING AND UNEARTHING POTENTIAL BIOSYNTHETIC GENE CLUSTERS IN THE SALT TOLERANT FUNGI

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

By

SUBHIKSHA N IYER (41739017)



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 **SUBHIKSHA N IYER (41739017)** who carried out the final year project entitled "**MINING AND UNEARTHING POTENTIAL BIOSYNTHETIC GENE CLUSTERS IN THE SALT TOLERANT FUNGI**" under our supervision from January 2023 to May 2023.

Dr. JEMMY CHRISTY

Internal Guide

External Guide

Dr. JEMMY CHRISTY

Head of the Department

Submitted for Viva voce Examination held on

Internal Examiner

External Examiner



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 **"Mining and unearthing potential biosynthetic gene clusters in salt tolerant fungi"** was carried out by **Ms. SUBHIKSHA N IYER (Reg. no. 41739017)**, Department of Bioinformatics and Data Sciences, Sathyabama Institute of Science and Technology, Chennai in partial fulfilment 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

Dr. Amit Kumar

Date: 10th April 2023

DECLARATION

I, **SUBHIKSHA N IYER (41739017)**, hereby declare that the Project Report entitled ***“MINING AND UNEARTHING POTENTIAL BIOSYNTHETIC GENE CLUSTERS IN THE SALT TOLERANT FUNGI”*** 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.



DATE:

SIGNATURE OF THE CANDIDATE

PLACE: CHENNAI

ACKNOWLEDGEMENT

I would like to acknowledge the support of the Management of Sathyabama institute of science and technology for their encouragement in completing this project work successfully.

I express my sincere gratitude to my guide, Dr. Amit Kumar for his invaluable guidance, encouragement, patience and support throughout my project work. His expertise, insights and constructive feedback have been instrumental in shaping my work and helping me achieve my research goals.

I would also like to convey my thanks to Dr. Jemmy Christy, Head of the Bioinformatics Department and my internal guide, for providing me with the necessary support and details at the right time during the progressive reviews and invaluable support.

Sincere thanks to my internal and external guide, whose diverse perspectives and expertise have helped me develop a more nuanced understanding of the research problem.

Finally, I would like to acknowledge the support of my family, the teaching and non teaching staff of the department of bioinformatics whose unwavering support has been a constant source of inspiration and I am grateful for their understanding and patience throughout the process of my master's project.

ABSTRACT

For years, fungi have been known to be prolific producers of natural products, which have numerous applications in therapy and industry. These natural products or secondary metabolites have been a means of adapting to their environment and defending them against competing organisms. The discovery of biosynthetic gene clusters responsible for producing secondary metabolites has been revolutionized by the process of genome mining. This approach involves the systematic analysis of an organism's genome sequence to identify the biosynthetic pathways responsible for producing these metabolites. By analysing the genes encoding the enzymes and regulatory elements involved in secondary metabolism, it is possible to predict the potential products of a given organism's biosynthesis. In addition, genome mining and in-silico techniques has uncovered previously unknown biosynthetic pathways revealing silent gene clusters that were not previously expressed. Halophilic fungi are of significant interest due to their unique adaptations to high salt environments. These organisms produce a wide variety of secondary metabolites with potent bioactivity and potential applications. Insilico analysis can help gain insight into the molecular mechanisms that underlie the unique adaptations and evolution of halophilic fungi. Here, for better understanding of halophilic fungi, we investigated the relationship between halophilic species and distribution of biosynthetic gene clusters using computational analysis of genome sequences of 23 fungal halophiles. We uncovered various types of BGCs in the halophilic fungi taken from Halodom database, like Terpenes, NRPS-like, Siderophores and others, including hybrid BGCs. This paper highlights the diversity of Biosynthetic gene clusters, their mining and their evolutionary relationships. The discovery of cryptic biosynthetic gene clusters may unlock their potential in development of new drugs, antibiotics and also their uses in industry. Overall, the discovery of biosynthetic gene clusters in halophilic fungi represent a significant breakthrough in the field of natural product discovery and opens up new doors for drug discovery and other biotechnological applications.

TABLE OF CONTENTS

CHAPTER	TITLE	PAGE NO
	ABSTRACT	i
	LIST OF ABBREVIATIONS	lii
	LIST OF TABLES	iv
	LIST OF FIGURES	iv
1	INTRODUCTION	
	1.1 Natural products	1
	1.2 Traditional Versus Genome Approaches from natural product	1
	1.3 Microbes and secondary metabolites	3
	1.4 Fungi and Hypersaline Environment.	4
2	REVIEW OF LITERATURE	
	2.1 Importance of Natural Products	6
	2.2 Biosynthetic gene clusters	7
	2.3 Tools and Databases	7
	2.4 Halophiles	18
	2.5 Halodom Database	19
	2.6 Halotolerant Fungi	19
3	AIM AND SCOPE	
	3.1 Aim	22
	3.2 Scope	22
4	MATERIALS AND METHODS	
	4.1 Mining of Halophilic Fungi and their natural products	23
	4.2 Identification of Biosynthetic Gene Clusters in Fungal Halophiles	23
	4.3 Manual Interpretation of Dataset	25
	4.4 Phylogenetic Tree Construction	25
	4.5 Visualization of phylogenetic Tree using ITOL	26
5	RESULTS AND DISCUSSION	
	5.1 Results	27
	5.2 Discussion	32
6	CONCLUSIONS	
	REFERENCES	

LIST OF ABBREVIATIONS

ABBREVIATIONS	FULL FORM
NP	Natural Product
NRPS	Nonribosomal Peptide Synthetases
PKS	Polyketide Synthases
T1PKS	Type 1 Polyketide Synthases
T3PKS	Type 3 Polyketide Synthases
BGC	Biosynthetic Gene Clusters
NCBI	National Centre For Biotechnology Information
DNA	Deoxyribonucleic acid
M	Molar
Ph	Potential of Hydrogen
Na ⁺	Sodium
TCM	Traditional Chinese Medicine
RiPP	Ribosomally synthesized and post-translationally modified peptide
PRISM	Prediction Informatics for Secondary Metabolomes
AntiSMASH	Antibiotics and Secondary Metabolite Analysis Shell
MiBIG	Minimum Information about a Biosynthetic Gene Cluster
BAGEL	Bacteriocin Genome mining tool
RODEO	Rapid ORF Description and Evaluation Online
NaPDoS	Natural Product Domain Seeker
NPD	Natural product domains
CLUSEAN	Cluster Sequence Analyser
BIGFAM	Bacterial and Fungal Biosynthetic Gene Family Database
BIGSCAPE	Biosynthetic Gene cluster - Structural Clustering Analysis Pipeline
MMBGA	Marine Metagenomic BGC Analyser
GC-OSP	Gene Cluster Orthologous Set Profiling
ARTS	Antibiotic Resistance Target Seeker
CARD	Comprehensive Antibiotic Resistance Database
DO-BISCUIT	Domain-Oriented Biosynthesis of Secondary Compounds using Inferred Trajectories
HMM	Hidden Markov Models
pHMM	Proile Hidden Markov Models
MIDDAS-M	Metabolite-Informed Discovery and Design of Antibiotic Synthesizing Microbes
NaCl	Sodium Chloride
w/v	Weight by volume
ITS	Internal Transcribed Spacer
ITOL	Interactive tree of Life
SVG	Scalable vector graphics
PNG	Portable network gaphics

JPG	Joint photographic experts group
PDF	Portable document format
MEGA	Molecular evolutionary genetics analysis

LIST OF TABLES

Table.1. List of tools and databases for BGC mining and prediction

LIST OF FIGURES

FIGURE LEGEND

- Fig.1.** The five steps for extraction and analysis of Biosynthetic gene clusters in halophilic fungi
- Fig.2.** Assembly level percentages of fungal halophiles mined from Halodom
- Fig.3.** Salt tolerance level of fungal halophiles mined from halodom
- Fig.4.** The distribution of BGCs across the strains with the phylogenetic tree and the number of BGCs
- Fig.5.** The diversity of BGCs in halophilic fungi: the number and type of BGC clusters per organism
- Fig.6.** The structure and biosynthesis of Ochratoxins

CHAPTER 1

INTRODUCTION

1.1 Natural Products

Natural products are small molecules, also known as secondary metabolites that are produced by all living organisms. They are formed from primary metabolism [47] and are an important source for a variety of bioactive compounds. They facilitate the survival of organisms in harsh environmental conditions under extreme stress and contend their applications in pharmaceuticals and industry. The natural products are chemically extremely diverse, and have an immense impact towards production of antibiotics, drugs, food, cosmetics and detergents [1]. Natural products have been carefully studied for the past few years and this study has led to understanding their mechanisms in terms of action, pathways or targets.

Experimental evidences state that these can be either used directly as drugs or can be used as targets or lead molecules for discovery of new drugs.[9] Microbial natural products are used in medicine for allograft rejections, cancer and infectious diseases. They also have a part in the agricultural sector in insect, weed and plant pathogen management in insecticides and fungicides [2,3].

Historically, natural products have played a key role in discovery of new drugs. They are extremely complex and difficult to synthesize, which are both a benefit and not. Their complexity gives the idea of various mechanisms of action and also the number of ways they may be utilized. They have a therapeutic application in different disease areas, like cardiovascular diseases and multiple sclerosis.[10]. Since the early 20th century, when penicillin was first discovered, many NPs started being identified and being worked towards their use as drugs [13].

1.2 Traditional Versus Genome Mining Approaches For Natural Product Extraction

The traditional methods for extraction of natural products involved long processes like solvent extraction [6]. NP leads are often identified by phenotypic assay and identifying their molecular mechanism of action can be time consuming. An outline of the traditional bioactivity guided isolation steps involve extraction of NPs from organisms

such as fungi and archaea. There are different choices of extraction based on what the extract could be. Some traditional extraction methods include Solvent extraction, Soxhlet extractor, ultrasonic extraction, microwave assisted extraction, liquid-liquid extraction, solid-phase micro extraction, etc., following extraction, the crude extracts undergo fractionation to give single compounds. After multiple rounds of fractionation, Bioactivity evaluation is done [10].

The major challenges that occur during these traditional extraction methods is that it may not always be possible to culture the organism out of its natural habitat. When these organisms are taken out of their natural habitat, they may not be able to produce their relevant NPs. There may be the presence of NPs that are not known. If one is working to understand the drug like property of the natural products and there are no drug like molecules obtained and when bioactive NPs are present in insufficient amounts [10]. The traditional methods for their extraction has not been perfect, there are some natural products that are under explored.

Bioinformatics and recent genomic advancements have eased the extraction of natural products through biosynthetic gene clusters. A **Biosynthetic Gene Cluster** 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) [5] The biosynthesis of natural products are employed by a number of pathways and each protein has a specific pathway encoded by a number of genes. These genes are clustered together and are contiguously arranged for the production of secondary metabolites [4]. The advancements in Next generation sequencing techniques has revolutionized the discovery of biosynthetic gene clusters by reducing the speed and cost.

An approach by Cacho R.A et.al gave a general workflow of how an NGS approach for connecting secondary metabolites to biosynthetic gene clusters in fungi would look like.[8]. The new bioactive compounds such as Non ribosomal peptides (NRPS), polyketides (PKS) and their hybrid discovery have been aided by genome mining. Previous studies have stated that only 1% of the microbial community has been cultured and used for biosynthetic gene cluster exploitation and most of it has still not been cultured under laboratory conditions [7]. The now emerging methods in genomics

and bioinformatics have enabled the extraction of previously failed BGC discovery with ease.

There are huge amounts of genomic sequence data generated that gives an ocean of information and the development of bioinformatic tools have helped analyse the data. There are a number of software tools that can be relied on to extract the cryptic gene clusters. Alongside, there are a number of databases with genome sequences that are open source for public access. These tools and databases can be used for genome mining of genes and gene clusters that are involved in secondary metabolite biosynthesis. With great amount of information comes great challenges. The load of data that was generated needs to be analysed, for which there are tools and databases that help researchers understand and analyse sequence data. Sequences can be found in NCBI Genbank, an annotated collection of all publicly available DNA sequences [14].

1.3 Microbes and Secondary Metabolites

Microorganisms are the leading producers of Natural products. They have been considered a rich source of unique bioactive compounds since the discovery of antibiotics. The majority of these microorganisms cannot be cultivated in laboratory conditions and only a small number of them can be cultured. These are called uncultivated microorganisms [15].

There are several databases focused on the mining of Biosynthetic gene clusters. Fungi produce a wealth of Biosynthetic gene clusters that give rise to natural products. Databases like Bactibase, antiSMASH database, ClusterMine360, ClustScan database, DoBISCUIT and IMG-ABC/ Integrated Microbial Genomes: Atlas of Biosynthetic Gene Clusters are genome mining tools and provide information about biosynthetic gene clusters and secondary metabolites. Tools used for mining secondary metabolite biosynthesis gene clusters are antiSMASH, ARTS, BAGEL, BiG-SCAPE, ClusterFinder, MIBiG, PKMiner, NaPDoS, EvoMining, RiPPMiner, SMURF etc.

Recent sciences have focussed on extremophile microorganisms with the inquisitiveness to understand the genes present in extremophiles. In both bacteria and fungi, BGCs encode for secondary metabolites that facilitate their survival in such harsh environments that are extreme for any non-halophile. BGCs resemble “selfish

operons" [12] with exchange among closely related strains due to DNA mismatch repair [11].

Halophilic organisms are extremophiles that thrive in environments with high salt concentration, like salt pans, salt lakes and pickled food products. Halophilic organisms are classified into 3 groups based on their degree of salt requirements: Slight halophiles (0.34-0.85 M), Moderate halophiles (0.85-3.4 M) and extreme halophiles (3.4-5.1 M). Halophiles exclude salt from their cytoplasm to avoid salting out. This is called high-salt-in-strategy. They don't have high nutritional requirements and have the capacity to balance the osmotic pressure of the environment. Usually halophiles inhabit ecosystems that have higher salinity than sea water. These ecological niches go from hypersaline soil, oil fields, marshes, sabkhas etc. Halophiles, be it Bacteria, fungi or archaea have the capacity to produce biomolecules of pharmacological importance for therapeutic applications. [16] These organisms can adapt to extreme conditions of salt, and other ions, often, like ultraviolet radiation, and in some cases, extreme pH.

1.4 Fungi and Hypersaline Environment

Fungi that inhabit the natural hypersaline environments do not require salt for their viability, as they can grow and adjust to the whole salinity range, from freshwater to almost saturated salt solutions. The ability to sense changes in Na⁺ concentrations in the environment is of vital importance in cell survival [17]. Halophilic fungi are well known producers of amylases, lipases, cellulases, proteases and others. Those enzymes have industrial applications in various sectors such as biofuel production, bioremediation, food, cosmetics, detergent and pharmaceutical processes [18]. Halophilic eukaryotic organisms have several advantages, like they identify genes that might enhance the properties of industrial yeasts and of plants used in agriculture [17].

Halophilic fungi are a resource for novel compounds with applications bending towards pharmaceutical and agriculture sciences. Fungi is known to produce bioactive compounds. Fungal genomes when analysed can give room for a number of biosynthetic gene clusters that contain the secondary metabolites. The aim of this work is to explore, understand and analyse the signatures of biosynthetic gene clusters

present in halophilic conditions, with the outburst of flood of genomic data in reach in the recent years.

CHAPTER 2

REVIEW OF LITERATURE

2.1 Importance of Natural Products

Natural products and their biological activity are currently of great interest in medicine, the natural food industry, and cosmetics in the field, and several scientific studies reproduce quickly. Most modern studies on the biological activity of natural products focus on different plant species [32]. Natural products remain a consistent source of drug leads with more than 40% of new chemical entities reported since 1981 being derived from microbial natural products. Microbial natural products, their semisynthetic derivatives, and their synthetic analogues are widely used in medicine for the treatment of a range of clinical disorders, including infectious diseases, cancers, high cholesterol, and allograft rejection [1,7].

Microorganisms have provided abundant sources of natural products which have been developed as commercial products from human medicine, animal health, and, plant crop protection. NPs lead 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 [33,34]. The biosynthesis of natural products usually involves multistep enzymatic reactions which are often called biosynthetic pathways. However, the identification of biosynthetic pathways in the genomic context is not simple, requiring the development of computational tools for genome mining. Because computational approaches that can optimize the target trajectory at different levels for the natural overproduction of products are highly desirable.[13]. Natural medicines, such as traditional Chinese medicine (TCM) and Ayurveda, were formed and developed in the daily life of ancient people and in the process of their fight against diseases over thousands of years, and they have produced a positive impact on the progress of human civilization. Natural products offer more drug-like properties molecules from combinatorial chemistry functional groups, chirality and structural complexity. Quantities of active substances in natural medicines are always quite low. The bottleneck was the laboratory and time-consuming extraction and isolation process and the use of natural products in the development of medicines.[6] 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 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 [35,36].

2.2 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. 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. 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. Biosynthetic gene clusters (BGCs) encode for the biosynthesis of a complex range of natural products, which can be classified into several different structural classes based on their chemical compositions includes non-ribosomal peptide synthetases (NRPS), polyketide synthases (PKS), terpenes, and ribosomally synthesized and post-translationally modified peptides (RiPPs) [5].

2.3 Tools and Databases

There are several databases related to biosynthetic gene clusters (BGCs) encoding secondary metabolites Antismash, ClusterFinder, PRISM, MIGiB, BIG SLICE, BAGLE 4, RODEO, BIG SCAPE, NaPDoS, NPD tools, PkMiner, CLUSEAN, RIPPMINER, NP linker, BIG FAM, Biosynthetic gene cluster Network (BGCNet), Np.searcher, Artemis, EvoMining, Marine metagenomic BGC Analyser(MMBGA), BioMass, Gene cluster orthologous Set profiling(GC-OSP), ARTS, SBSPKS, DO BISCUIT, Deep BGC,

MIDDAS M, ClusterMine 360, are some tools for predicting biosynthetic gene clusters (BGCs).

➤ **Anti Smash:**

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. Antismash uses a combination of sequence similarity, hidden Markov models, and machine learning approaches to identify and classify BGCs based on their genetic composition and predicted chemical structures. The tool can detect a wide range of BGCs encoding diverse classes of natural products, such as polyketides, non-ribosomal peptides, terpenes, and hybrid compounds.

➤ **Cluster Finder:**

ClusterFinder is a web-based tool developed by the Sunagawa Lab at the Okinawa Institute of Science and Technology in Japan for the detection and analysis of biosynthetic gene clusters (BGCs) in metagenomic data. The tool allows users to upload and analyze large-scale metagenomic datasets for the presence of BGCs encoding diverse natural products. It uses a novel algorithm that integrates sequence similarity and network analysis to identify BGCs from metagenomic data. The tool first clusters genes based on their sequence similarity, and then analyzes the co-occurrence patterns of these gene clusters across different samples to identify recognised BGCs. The tool generates interactive heatmaps and networks that allow users to identify potential ecological and evolutionary relationships among different natural product biosynthetic pathways.

➤ **PRISM:**

PRISM (PRediction Informatics for Secondary Metabolomes) is a computational tool developed by the Medema Lab at Wageningen University in the Netherlands for the prediction and annotation of secondary metabolites in microbial genomes. The tool is designed to identify and classify biosynthetic gene clusters (BGCs) based on their genomic context and predicted chemical structures. In addition to predicting BGCs and their products, PRISM also provides tools for exploring the evolutionary relationships and functional diversity of different natural product biosynthetic pathways. The tool

generates interactive graphical displays of BGCs, including gene maps, phylogenetic trees, and chemical structure predictions, allowing users to investigate the biosynthetic pathways and functions of different natural products.

➤ **MIBiG:**

MIBiG (Minimum Information on a Biosynthetic Gene Cluster) is a database developed by the University of California at San Diego that contains information on experimentally characterized biosynthetic gene clusters (BGCs) and the natural products they produce. The database contains more than 1,500 bacterial and fungal BGCs covering a wide range of natural products such as antibiotics, antifungals, and anticancer drugs [38].

➤ **Big Slice:**

BIG-Slice is a computational tool developed by the University of California, San Diego, USA to identify and analyze biosynthetic gene clusters (BGCs) within large and complex genomic datasets. The tool is designed to scale to datasets containing hundreds or thousands of genomes, making it suitable for large-scale comparative genomics studies. BIG-Slice uses a combination of gene clustering, phylogenetic analysis, and network inference to identify putative BGCs and predict their functions. The tool incorporates multiple algorithms for gene prediction, gene clustering, and chemical structure prediction to identify and annotate BGCs encoding different types of natural products, such as polyketides, non-ribosomal peptides, and terpenes.

➤ **BAGEL4:**

BAGEL4 (BActeriophage Genomic mining Enhanced by machine Learning) is a computational tool developed by the University of Leuven in Belgium for the identification and analysis of bacteriophage-encoded biosynthetic gene clusters (BGCs). The tool combines sequence analysis, machine learning, and network inference to identify putative BGCs and predict their functions. BAGEL4 starts by identifying bacteriophage genomes in metagenomic datasets using sequence-based methods. The tool then performs gene prediction and clustering to identify putative BGCs, which are subsequently analyzed using machine learning algorithms and helps to predict their functions and chemical structures.

➤ **RODEO:**

RODEO (Rapid ORF Description and Evaluation Online) is a computational tool developed by the University of California, San Diego for the identification and analysis of biosynthetic gene clusters (BGCs) in microbial genomes. The tool uses a combination of sequence-based and comparative genomics approaches to identify putative BGCs and predict their functions. RODEO starts by performing gene prediction and clustering to identify potential BGCs. The tool then performs comparative genomics analyses to identify conserved biosynthetic gene clusters across different microbial genomes, which can provide insights into the functions and evolution of these clusters. RODEO also uses machine learning algorithms to predict the chemical structures and bioactivities of recognized BGCs.

➤ **Big Scape:**

BiG-Scape (Biosynthetic Gene cluster - Structural Clustering Analysis Pipeline) is a computational tool developed by the University of Warwick for the identification and analysis of biosynthetic gene clusters (BGCs) in microbial genomes. The tool uses a combination of sequence-based and structural similarity analyses to cluster BGCs into families and predict their functions. BiG-Scape starts by identifying potential BGCs in microbial genomes using sequence-based methods. The tool then uses AntiSMASH, a widely used BGC prediction tool, to annotate the gene clusters and identify their core biosynthetic genes. BiG-Scape then performs structural similarity analyses on the core biosynthetic genes, using the MIBiG database of known natural products as a reference, to cluster the BGCs into families based on their structural similarities.

➤ **NaPDoS:**

NaPDoS (Natural Products Database Online System) is a web-based tool for identifying and analyzing natural products and their biosynthetic gene clusters, developed by Toyama University, Japan. The database contains over 36,000 natural products from various sources including bacteria, fungi, plants and animals. NaPDoS allows users to search for natural products by chemical structure, molecular weight, and other chemical properties, as well as the name of the producing organism or the biosynthetic gene clusters involved in their synthesis. This tool also provides information on the biosynthetic pathways and genetic background of natural product biosynthetic gene clusters.

➤ **NPD tools:**

NPD (Natural Product Domain) tools are a set of web-based tools developed by the University of California, San Diego for the analysis and annotation of biosynthetic gene clusters (BGCs) involved in the production of natural products. The NPD tools include several modules, such as AntiSMASH, ClusterFinder, and PRISM, which can be used for gene cluster identification, annotation, and comparison, as well as for predicting the chemical structures of natural products based on their biosynthetic gene clusters.

➤ **PkMiner:**

PkMiner is a web-based tool developed by the University of Illinois at Urbana-Champaign for the identification and analysis of biosynthetic gene clusters (BGCs) involved in the production of polyketides, a class of natural products with a wide range of biological activities, including antimicrobial, antifungal, and anticancer properties. PkMiner allows users to search for polyketide BGCs by keyword, structural features, and other criteria, as well as to compare and analyze the BGCs and their associated pathways. The tool also includes features for visualizing the BGCs and their gene products, as well as for predicting the chemical structures of novel polyketides based on their biosynthetic gene clusters.

➤ **CLUSEAN:**

CLUSEAN is a software tool for identifying and analyzing biosynthetic gene clusters (BGCs) involved in the production of secondary metabolites such as antibiotics, antimycotics, and other bioactive compounds. The tool also includes the ability to predict the chemical structures of natural products produced by BGC based on their biosynthetic pathways, as well as the ability to compare and analyze BGCs and their associated pathways. Additionally, CLUSEAN provides a graphical user interface for visualizing predicted BGCs and their gene products [37].

➤ **RIPP Miner:**

RIPP MINER is a web-based tool developed by the University of Warwick for the identification and analysis of biosynthetic gene clusters (BGCs) involved in the production of natural products, such as antibiotics, antifungals, and other bioactive compounds. RIPP Miner uses a computational algorithm that integrates genomic, transcriptomic, and metabolomic data to predict BGCs and their associated natural

products. The tool also includes features for visualizing the BGCs and their gene products, as well as for comparing and analyzing the BGCs and their associated pathway [39].

➤ **NPlinker:**

NPlinker is a software tool developed by the University of California, San Diego, for the identification and analysis of biosynthetic gene clusters (BGCs) involved in the production of natural products, such as antibiotics, antifungals, and other bioactive compounds. NPlinker uses a network-based approach to predict BGCs and their associated natural products by analyzing the relationships between genes, pathways, and chemical structures. The tool also includes features for visualizing the BGCs and their gene products, as well as for comparing and analyzing the BGCs and their associated pathways [40].

➤ **BIG FAM:**

BIG FAM (Bacterial and Fungal Biosynthetic Gene Family Database) is a public database containing information on biosynthetic gene families and their related biosynthetic gene clusters. This database will help researchers study the diversity and evolution of biosynthetic pathways in bacteria and fungi that are involved in the production of many important natural products, including antibiotics, antifungals, and anticancer drugs. It contains information on over 2,000 biosynthetic gene families, groups of genes involved in the production of specific types of natural products. For each gene family, the database provides information on biosynthetic pathways, including the enzymes involved, and biosynthetic gene clusters, including genes involved in the production of natural products.

➤ **Biosynthetic Gene Cluster Network (BGCNet):**

Biosynthetic Gene Cluster Network (BGCNet) is a bioinformatics tool that utilizes a network-based approach to analyze biosynthetic gene clusters. Biosynthetic gene clusters are groups of genes that are responsible for producing a specific natural product, such as a drug or a toxin. BGCNet is designed to help researchers understand the relationships between different biosynthetic gene clusters, which can be useful for predicting the function of unknown gene clusters, identifying potential drug targets, and studying the evolution of biosynthetic pathways. BGCNet works by constructing a

network representation of the biosynthetic gene clusters, where each cluster is represented as a node in the network and the edges represent similarities between the clusters. The similarities are based on shared biosynthetic genes, as well as other factors such as gene order, gene orientation, and cluster boundaries. BGCNet then uses network analysis algorithms to identify clusters that are functionally related or evolutionarily linked, and to visualize the relationships between the clusters.

➤ **Np.Searcher:**

Np.searcher is a bioinformatics tool used for identifying and characterizing natural product biosynthetic gene clusters. Natural products are compounds that are produced by living organisms, and they have many important applications in medicine, agriculture, and other fields. Natural product biosynthetic gene clusters are groups of genes that are responsible for producing these compounds. Np.searcher uses a combination of sequence-based and comparative genomics approaches to identify and analyze biosynthetic gene clusters. The tool can take input data from either genomic or metagenomic datasets, and it uses a machine learning algorithm to classify gene clusters based on their potential natural product class.

➤ **Artemis:**

Artemis is a genome browser and annotation tool that is designed for the visualization and annotation of bacterial and fungal genomes. The software is free and open-source, and it is widely used by researchers and bioinformaticians for genome analysis. Artemis provides a range of tools for the visualization of genomic data, including the ability to display genomic features such as genes, coding regions, and non-coding regions, as well as customizable color-coding of annotations. It also includes tools for comparative genomics, allowing users to compare multiple genomes and identify conserved regions or differences between them.

➤ **EvoMining:**

EvoMining is a bioinformatics tool that uses evolutionary patterns to predict and annotate biosynthetic gene clusters. Biosynthetic gene clusters are groups of genes responsible for producing natural products such as antibiotics, pigments, and toxins. These clusters are often highly conserved across species and can be difficult to identify based on sequence alone. EvoMining takes a comparative genomics

approach, comparing the genomes of closely related species to identify conserved regions that are likely to be biosynthetic gene clusters. The software then uses a machine learning algorithm to classify these regions based on their biosynthetic function

➤ **Marine Metagenomic BGC Analyser (MMBGA):**

Marine Metagenomic BGC Analyser (MMBGA) is a bioinformatics tool designed for the analysis of biosynthetic gene clusters in marine metagenomic datasets. The software is specifically tailored for the analysis of microbial communities in marine environments, which are known to be a rich source of natural products with potential applications in medicine and other fields. MMBGA uses a combination of sequence-based and comparative genomics approaches to identify and annotate biosynthetic gene clusters in metagenomic datasets. The software can take input data in a variety of formats, including raw reads, assembled contigs, or annotated genomes.

➤ **Biomass:**

Biomass is a bioinformatics tool that is designed for the identification and annotation of biosynthetic gene clusters (BGCs) in bacterial and fungal genomes. The software uses a combination of sequence-based and comparative genomics approaches to identify BGCs and predict their functions. BioMass first identifies candidate BGCs based on the presence of genes that are known to be involved in biosynthesis. It then uses comparative genomics to compare the identified BGCs to known BGCs in databases such as MIBiG (Minimum Information about a Biosynthetic Gene cluster) and antiSMASH (antibiotic and Secondary Metabolite Analysis Shell), allowing the software to predict the biosynthetic functions of the BGCs.

➤ **Gene Cluster Orthologous Set Profiling (GC-OSP):**

Gene Cluster Orthologous Set Profiling (GC-OSP) is a bioinformatics tool designed for the comparative analysis of biosynthetic gene clusters (BGCs) based on the presence of orthologous genes. The software uses a combination of sequence-based and comparative genomics approaches to identify and compare BGCs in different genomes. GC-OSP first identifies candidate BGCs based on the presence of genes

that are known to be involved in biosynthesis. It then uses orthologous gene clustering to group the identified BGCs into orthologous sets based on the presence of conserved genes. The software uses a clustering algorithm to identify groups of BGCs that share a high degree of orthology and are likely to be functionally related.

➤ **ARTS: Antibiotic Resistance Target Seeker:**

ARTS (Antibiotic Resistance Target Seeker) is a bioinformatics tool designed for the identification of antibiotic-resistance genes and their associated biosynthetic gene clusters (BGCs). The software uses a combination of sequence-based and comparative genomics approaches to identify and annotate antibiotic resistance genes and BGCs in bacterial genomes. ARTS first identifies putative antibiotic resistance genes based on their similarity to known resistance genes in databases such as CARD (CA) and ResFinder. The software then uses comparative genomics to identify BGCs that are associated with the identified resistance genes.

➤ **SBSPKS:**

SBSPKS is a bioinformatics tool designed for the identification and annotation of Type II polyketide synthase (PKS) biosynthetic gene clusters (BGCs). Type II PKSs are a family of enzymes that are involved in the biosynthesis of a wide range of natural products, including antibiotics, antifungals, and anticancer agents. SBSPKS uses a combination of sequence-based and comparative genomics approaches to identify and annotate Type II PKS BGCs in bacterial and fungal genomes. The software first identifies putative PKS genes based on their similarity to known PKS genes in databases such as antiSMASH and MIBiG. It then uses comparative genomics to identify BGCs that are associated with the identified PKS genes.

➤ **DO BISCUIT:**

DO-BISCUIT (Domain-Oriented Biosynthesis of Secondary Compounds using Inferred Trajectories) is a bioinformatics tool for the prediction of biosynthetic gene clusters (BGCs) based on the domain organization of biosynthetic enzymes. DO-BISCUIT uses a domain-centric approach to identify and classify biosynthetic enzymes into families based on the presence and arrangement of specific domains.

The tool then uses these domain families to predict BGCs by searching for co-occurring domains in genomic data.

➤ **Deep BGCs:**

DeepBGC is a bioinformatics tool that uses deep learning algorithms to identify and annotate biosynthetic gene clusters (BGCs) in bacterial and fungal genomes. BGCs are gene clusters that produce specialized metabolites such as antibiotics, antimycotics, and other bioactive compounds. DeepBGC uses a deep neural network to predict the presence of BGC within a genome sequence. A neural network is trained on a large dataset of annotated BGC and non-BGC so that it can learn BGC patterns and features [41].

➤ **MIDDAS:**

MIDDAS-M (Metabolite-Informed Discovery and Design of Antibiotic Synthesizing Microbes) is a bioinformatics tool for the design of engineered microbial strains capable of producing novel antibiotics. MIDDAS-M uses a metabolite-informed approach to identify potential antibiotic biosynthetic gene clusters (BGCs) and their associated pathways in microbial genomes. First, MIDDAS-M uses comparative genomics and metabolic pathway analysis to identify potential BGCs in known antibiotic-producing microorganisms. Then, MIDDAS-M predicts the metabolic pathways that lead to the production of the target antibiotic and identifies the necessary genetic modifications to enable the production of the antibiotic in the target host organism [37].

➤ **ClusterMine360:**

ClusterMine360 is a bioinformatics tool for the analysis of biosynthetic gene clusters (BGCs) in microbial genomes. It is designed to identify and annotate BGCs involved in the biosynthesis of secondary metabolites, such as antibiotics, antifungals, and antivirals. ClusterMine360 uses a combination of homology-based and de novo methods to predict BGCs in microbial genomes. It also integrates information on gene expression, metabolite production, and pathway analysis to provide a comprehensive analysis of BGCs and their associated pathways [37].

➤ **NRPSPredictor:**

NRSPredictor is a bioinformatics tool used for the prediction of nonribosomal peptide synthetase (NRPS) gene clusters in bacterial and fungal genomes. NRPS gene clusters are responsible for the biosynthesis of a wide range of bioactive peptides, including antibiotics, immunosuppressants, and anticancer agents. NRSPredictor uses a combination of hidden Markov models (HMMs) and machine learning algorithms to identify and classify NRPS gene clusters based on their amino acid sequence patterns and domain structures. The tool also provides information on the predicted chemical structure of the NRPS products and their potential biological activities [37].

Table.1. List of tools and databases for BGC mining and prediction

ANTI SMASH	https://antismash.secondarymetabolites.org/
Cluster Finder	http://magarveylab.ca/clusterfinder.php
PRISM	https://magarveylab.ca/prism/
MIGBIG	https://mibig.secondarymetabolites.org/
BIG SLICE	https://github.com/justinmcleslie/bigslice
BAGLE4	http://bagel4.molgenrug.nl/
RODEO	https://github.com/caufieldjh/Rodeo
BIG SCAPE	https://git.wageningenur.nl/medema-group/bigscape
NapDos	http://napdos.ucsd.edu/

NPD tools	https://www.npdt.org/
PkMiner	http://magarveylab.ca/pkminer.php
CLUSEAN	http://magarveylab.ca/clusean.php
RIPMINNER	https://magarveylab.ca/rippminer/
NP linker	https://magarveylab.ca/np-linker/
BIG FAM	https://bigfam.bioinformatics.nl/
Biosynthetic Gene Cluster Network	https://bgcnet.github.io/
Np.searcher	https://magarveylab.ca/npsearcher/
Artemis	http://sanger.ac.uk/science/tools/artemis

EvoMining	https://evomining.biosynth.eu/
MMBGA	http://mmbga.biosynth.eu/
Biomass	https://bitbucket.org/berkeleylab/biomass/src/master/
GC-OSP	https://github.com/KatelynFerris/GC-OSP
ARTS	http://arts.ziemertlab.com/
SBKPKS	https://github.com/cafinscratch/sbs-pks
DO BISCUIT	https://github.com/cluster411/dobiscuit
Deep BGCs	https://bitbucket.org/berkeleylab/deepbgc/src/master/

MIDDAS	https://middas-m.ceb.cam.ac.uk/
Cluster Mine 360	http://clustermine360.caas.cn/index.html
NRPSPredictor	http://nrps.informatik.uni-tuebingen.de/

2.4 Halophiles

Halophiles are organisms that can tolerate high salinity and are either extremophilic or extremophilic. According to the maximum salt tolerance, he is divided into three categories

Slight, Moderate, Excessive. Halophiles are found in all fields, especially in bacteria, archaea, and eukaryotes, and exhibit a wide range of metabolic processes and physiological responses. Most research on halophiles has focused on understanding the unique adaptations and molecular processes that allow them to maintain osmotic balance in the presence of salt stress. Since most of them belong to ancient evolutionary lineages, there is also much interest in studying 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 because they provide a model of how life can adapt to extreme conditions on other planets. Halophiles are a valuable resource for drug discovery, and the design of some halophiles produces compounds with antibiotic properties that are active against drug-resistant bacteria. 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 variety of enzymes with potential applications in biotechnology, such as biofuel production and food and agricultural processing [19].

Halophiles have a number of potential applications in drug discovery, particularly in the search for novel bioactive compounds. Halophiles are known to produce a wide range of secondary metabolites, including antimicrobial peptides, enzymes, and pigments, which have potential therapeutic applications. These compounds have evolved to help halophiles survive in harsh environments and may have unique chemical structures and biological activities that make them attractive targets for drug development [42].

Another example of a bioactive compound produced by halophiles is halocin C8, which is a cyclic peptide that has been shown to have antifungal activity. Halocin C8 has potential applications in the treatment of fungal infections, which are a significant cause of morbidity and mortality in immunocompromised patients.[43] Halophiles have the potential to bring many benefits to humans, from discovering new drugs to cleaning up pollution. Understanding these unique organisms and their capabilities can provide valuable insight into the nature of life itself and its interaction with the environment. An example of a bioactive compound produced by halophiles is halosine, an antimicrobial peptide that has been shown to be effective against a wide range of bacteria and fungi, including multidrug-resistant pathogens. Halosine has potential applications in the development of new antibiotics and other antibacterial agents [44].

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. There are 282 Archaea, 669 Bacteria, and 314 Eukarya currently registered in the HaloDom database. Three halophilic databases are currently available online.

HaloWeb , HaloBase and HProtDB, HaloWeb focuses on genomic information and provides complete genome sequences for download.

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 [19].

2.6 Halotolerant Fungi

Halophilic Fungi live in hypersaline extreme environments with unique physicochemical properties such as a salinity gradient. Although the investigation of microbiota in salterns has focused on archaea and bacteria, diverse fungi also thrive in the brine and soil of salterns. Fungi isolated from salterns are represented by black yeasts (*Hortaea werneckii*, *Phaeotheca triangularis*, *Aureobasidium pullulans*, and *Trimmatostroma salinum*), *Cladosporium*, *Aspergillus*, and *Penicillium* species. Most studies on saltern-derived fungi gave attention to black yeasts and their physiological characteristics, including growth under various culture conditions.

The requirement for salt to grow is an unusual characteristic for fungi. Two fungal species, *Wallemia ichthyophaga* and *Aspergillus penicillioides*, have been reported as obligate halophiles isolated from salterns. *Wallemia ichthyophaga* is a non-melanized basidiomycetous fungus belonging to the order Wallemiales and the class Wallemiomycetes, and it has rarely been isolated from the environment, with only 24 strains isolated. It is the most halophilic fungus ever described, requiring at least 10% (w/v) NaCl for growth. *Wallemia ichthyophaga* shows optimum growth at 15–20% (w/v) NaCl, and can grow at NaCl saturation (30% [w/v]). It is also one of the most xerophilic (thriving at low water activity) fungal species, which is very rare among the Basidiomycota. *A. penicillioides* is an ascomyceteous fungus present in diverse saline habitats, including the Dead Sea, salterns, estuaries, and mangroves. It was unable to grow without the addition of 10% solar salt, and grew in up to 20% solar salt on CDA. In general, microorganisms growing only in the presence of salt are considered halophiles. Halophilic microorganisms are classified into slight (optimum growth at 3% [w/v] NaCl), moderate (3–15% [w/v] NaCl), and extreme (25% [w/v] NaCl, but failing to grow below 12% NaCl) halophiles. In contrast, halotolerant microorganisms inhabit saline environments but survive without NaCl. Extremely halotolerant microorganisms refer to those that can grow in saline conditions above 15% (w/v) NaCl. Occasionally, the term halophiles has been used for fungi differently from most prokaryotes: fungi thriving in hypersaline environments but viable without salt are designated halophilic fungi. For example, the black yeast *Hortaea werneckii* grows at NaCl concentrations ranging from 0% (w/v) to saturation (32%), with optimum growth at 6–10% (w/v). [45]. These introduced halophilic fungi are a good source of biotechnologically important genes. Studies have shown that when selected genes from *Debaryomyces hansenii* and *Hortaea werneckii* were heterologously expressed in *Saccharomyces cerevisiae*,

some of them significantly improved the salt tolerance of *S. cerevisiae*. Two interesting examples of *Debaryomyces hansenii* are HAK1 and HAL2. *Debaryomyces hansenii* has two different K transporters; while one of them (Trk1p) is present in all yeasts studied so far, the other,...[46] The Ascomycetous yeast order Saccharomycetales contains numerous species that are associated with sugary plant saps and exudates. Osmotolerant yeast taxa belong mainly to the families Saccharomycetaceae and Metschnikowiaceae. However, yeasts that have been isolated from hypersaline waters that have hemi-ascomycetous affinities only belong to the genera *Candida*, *Debaryomyces*, *Metschnikowia* and *Pichia*. Different species of the genus *Pichia* are frequent saline aquatic isolates, with *Pichia membranifaciens* being the most common. The species *Candida glabrata*, *Candida tropicalis*, *Candida parapsilosis* and *Candida krusei* were isolated for the first time from the hypersaline water of the Dead Sea. Among these, only *Candida parapsilosis* was known previously as a food-borne halotolerant yeast, while the others were not known for their halotolerance. The division Basidiomycota contains three orders with halophilic and halotolerant representatives. In the recently separated order Trichosporales the yeast *Trichosporon mucoideshae* has been frequently isolated from hypersaline waters of salterns and from the Dead Sea. In Sporidiales, several *Rhodotorula* species have been isolated from salterns, including the repeatedly isolated *Rhodosporidium sphaerocarpum*, *Rhodosporidium babjevae* and *Rhodotorula larynges*. The order Wallemiales was only recently introduced to accommodate the single genus *Wallemia*, a phylogenetic maverick in the Basidiomycota. Thus, until recently, this genus contained only the species *Wallemia sebi*. However, taxonomic analyses of isolates from sweet, salty and dried food and from hypersaline evaporation ponds in the Mediterranean, Namibia, and the Dead Sea have resolved this genus into three species: *Wallemia sebi*, *Wallemia muriae* and *Wallemia ichthyophaga*. It is remarkable that the entire genus *Wallemia*, and therefore the entire order Wallemiales, are either xerophilic/halophilic or xerotolerant.

CHAPTER 3

AIM AND SCOPE

3.1 Aim

The primary aim of this research endeavour is to delve into the diversity and distribution of biosynthetic gene clusters harboured within the fungal kingdom and appraise their prospective utility for biomedical and biotechnological applications. Leveraging in silico genome mining and robust data analysis techniques, we aim to pinpoint hitherto undiscovered biosynthetic gene clusters that might be engaged in synthesizing innovative bioactive compounds. Our ultimate objective is to unearth novel and promising bioactive agents that could engender a transformative impact on the domains of drug discovery and biotechnology. We also intend to seek expert opinion from specialists in the field of natural product chemistry to corroborate and validate our findings and identify any hidden jewels of bioactivity. This initiative has the potential to radically influence the pharmaceutical and biotechnology industries by ushering in the development of pioneering drugs and biotechnological applications via the discovery of new natural products.

3.2 Scope

The purview of this investigation is to scrutinize the potential biotechnological and medicinal uses of halophilic fungi and their biosynthetic gene clusters. Our primary goal is to scrutinize and appraise these gene clusters in halophilic fungi, probing their diversity and dispersion, and evaluating the likely bioactive compounds that they may produce. In order to accomplish these objectives, we will employ sophisticated bioinformatics tools and techniques, conduct a thoroughgoing review of the scientific literature, and analyse the resultant data. Collaborating with specialists in natural products chemistry will be essential to validate our findings and discern any potential bioactive compounds. The possible identification of novel natural products via this research endeavour could have momentous implications for the pharmaceutical and biotechnology sectors, culminating in the advent of innovative medicines and biotechnological applications.

CHAPTER 4

MATERIALS AND METHODS

4.1 Mining of Halophilic Fungi and Their Genome

The mining of halophilic fungi was done from a database called Halodom. Created in June 2017, halodom is an online database containing more than 1000 halophilic species from all life domains. This resource expands considerably compared with other databases in terms of coverage as it contains information about halophiles from all life domains and the number of entries. The database grows additional metadata and periodical updates are scheduled for this database once in 2 months. It is a useful repository for a wide range of research topics. Halodom currently contains 1267 entries of halophiles across all life domains. There are 282 Archea, 669 Bacteria and 314 Eukarya in the database (version 1.3). The database contains the species, order, class, phylum, domain, Genome availability, Salinity concentration/ range(w/v) and Halotolerant classification of the halophiles [19].

These halophiles were taken and the fungal halophilic species were filtered and curated manually in Microsoft Excel. There were 52 fungal halophilic species filtered out of the 1267 entries and these were checked for the availability of genome sequence. Among the 52 fungal species, we selected 23 fungal species whose genome sequences were available. The species in halodom had links directed to NCBI genome to find the genome sequence of the halophiles. The species with lack of genome sequence on the database were directly mined from NCBI genome in .gbff (genbank) format. NCBI's genome resources include information on large scale genomics projects, genome sequences and assemblies and mapped annotations such as variations, markers and data from epigenomics studies. They include information about organisms' genomes and genomic assembly statistics. Microsoft Excel is a ubiquitous software for data curation and analysis. Excel was used to sort the genomes on the basis of their size and assembly level parameters.

4.2 Identification of Biosynthetic Gene Clusters in the fungal halophiles

A robust web application of antiSMASH 6 was used to extract biosynthetic gene clusters from halophilic fungal genomes. AntiSMASH uses manually curated and

validated 'rules' that define which biosynthetic functions need to exist in a genomic region in order to constitute a BGC. It uses profile hidden Markov models (pHMMs) from PFAM, TIGRFAMs, SMART, BAGEL or custom models. antiSMASH performs module detection, which allows for better prediction of modifications made to the respective monomers during biosynthesis. The ClusterCompare algorithm, similar to ClustalBlast builds on local protein-alignment based sequence comparison as one part of the score. Biosynthetic components are the collection of gene products matching one of antismash's BGC detection profiles. All score parts are scaled to 0 to 1 range and the geometric mean of the parts. Tools like ClusterFinder and DeepBGC offer alternative, machine-learning approaches to predict gene clusters. Antismash fungal version or fungismash can be used to detect BGCs. The job details and the results can be sent to the entered email ID. The data input can either be given as NCBI accession number or the sequence file in Genbank/ FASTA or EMBL format can be uploaded. The detection strictness is maintained at 'relaxed' mode, neither strict nor loose to

- i) Detect well- defined clusters containing all the required parts
- ii) Detect partial clusters missing one or more functional parts.

There are extra features that can be chosen in the database. KnownClusterBlast, ClusterBlast, SubClusterBlast, MIBiG cluster comparison, ActiveSiteFinder, RREFinder, Cluster Pfam analysis, Pfam-based GO term annotation, TIGRFam analysis and TFBS analysis. KnownClusterBlast and MIBiG cluster comparison are methods for identifying and comparing biosynthetic gene clusters involved in the production of secondary metabolites. TFBS analysis identifies transcription factor binding sites within the DNA sequences. Active SiteFinder recognizes active sites within the protein structure. TIGRFam analysis identifies protein families and domains based on their evolutionary relationship. Pfam-based GO term annotation assigns Gene ontology terms to protein domain based on their known functions. SubCluster Blast and RREFinder helps to identify subcluster genes within the gene cluster. The antismash result gives the region, the type of biosynthetic gene cluster, its location, most similarly known cluster and the percentage of similarity. The antismash results can be downloaded- log files, and genbank summary file can be downloaded separately if all results are not needed. Although there are different biosynthetic gene cluster extraction tools like MIBiG, PRISM, BAGEL, NP.searcher, BiG-SLiCE and NaPDoS, antiSMASH is considered as the gold standard tool for the extraction of Biosynthetic

gene clusters because of its ability to accurately predict and annotate a wide range of natural product biosynthetic pathways [20].

4.3 Manual interpretation of the dataset

The manual interpretation of the fungal dataset was done using Microsoft excel. The organisms were separated on the basis of their assembly level- Contig, Scaffold and Chromosomal assembly level and the Biosynthetic gene cluster classes were carefully curated on the basis of antismash results. The organisms were sorted from highest to lowest number of biosynthetic gene clusters for easy interpretation. Excel was used to design pie charts and different graphs for easy understanding and interpretation of the results. Graphical plots and charts visualizing the number of clusters per organism, the number of clusters overall, the correlation between the genome size and the number of biosynthetic gene clusters and other statistics were visualized using Microsoft excel.

4.4 Phylogenetic tree construction

The evolutionary relationships of the fungal halophiles were analysed with the construction of a phylogenetic tree. Phylogenetic trees are either rooted or unrooted depending on the presence of an ancestor. Rooted trees have clearer interpretation of the evolutionary relationships between different taxa. For the construction of phylogenetic trees, the ITS region of the halophilic fungal sequences were extracted and downloaded. The ITS regions of the fungal species were considered for the construction of phylogenetic trees because ITS regions are a genetic marker that is widely sequenced and is also considered as the formal barcode regions of fungal kingdom [21].

The ITS regions (Internal Transcribed Spacer) regions are non coding regions. They evolve quickly and show a high degree of variation between species, which is why they are very useful in studying evolutionary relationships. The ITS regions of the fungal halophiles were downloaded from NCBI genome in fasta format.

To align the fasta sequences for further process, Bioedit, a biological sequence alignment editor for sequence manipulation and analysis was used. [22] Bioedit was used to align the sequences and remove the low quality bases in the sequences. Bioedit was also used to do clustal omega multiple sequence alignment.

To make a reliable phylogeny in the Biosynthetic gene cluster, we constructed a phylogenetic tree with the ITS regions of the halophilic fungi using MEGA 11. MEGA 11 or Molecular Evolutionary Genetics Analysis Version 11 is a comprehensive tool for building timetrees of species, pathogens and gene families using rapid relaxed-clock methods [23]. Using MEGA 11, a phylogenetic tree of the aligned fasta files was constructed in the neighbour joining method with bootstrap value set to 1000. Bootstrapping is a method to quantify uncertainty in Phylogenetic trees. It randomly resamples the data. Bootstrap evaluates the reliability on a particular phylogenetic tree. After the construction of a phylogenetic tree, the tree was exported in Newick format for visualization. Newick is a standard file format used to represent phylogenetic trees. The newick format represents the tree as a series of nested parentheses, with each set of parentheses containing a branch length and the labels of the descendant nodes. Newick format was preferred as it was useful for visualization of the phylogenetic tree in a tool called ITOL [24].

4.5 Visualization of phylogenetic trees using ITOL

After the construction of phylogenetic tree, it was visualized using ITOL. ITOL is an online tool for phylogenetic tree display and annotation [25]. The tree display is implemented in JavaScript and uses HTML5 canvas element for visualization. The control panel of ITOL provides basic and advanced visualization setting, with mode being rectangular, circular or unrooted. The branches can be colored 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 has to be uploaded for the construction of 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.

CHAPTER 5

RESULTS AND DISCUSSION

5.1 RESULTS

We used genome sequences for 23 different halophilic fungi to investigate their biosynthetic potential. Firstly, the genome sequences of the halophilic fungi 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 a prevalent use for the same. Out of 23 genomes, 12 genomes were on scaffold level assembly (52.17%), 6 were contigs (26.08%) and 5 were on chromosomal assembly level (21.73%). 52% of the organisms were on moderate salinity level and 48% were on extreme salinity level. Fig. 2 and Fig. 3 shows the statistical charts of the data described above. In order to understand the evolutionary relationships between the fungal halophiles, a phylogenetic tree was constructed using MEGA 11. The interactive tree of life (ITOL) was used to illustrate the phylogenetic tree. The interactive tree of life of halophilic fungi in Fig. 4a and 4b suggested that there were 4 ancestral groups occupying clear evolutionary positions at the root of the tree.

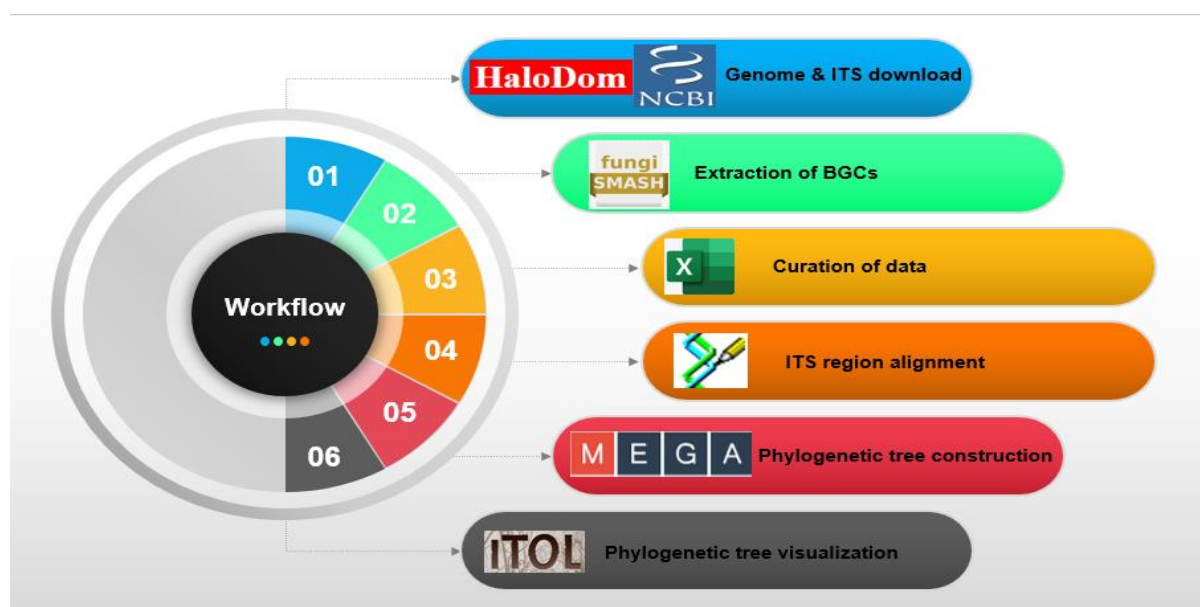


Fig. 1. The five steps for extraction and analysis of biosynthetic gene clusters in halophilic fungi.

374 BGCs were extracted using antiSMASH with a variety of cluster classes including NRPS-like, Terpenes and siderophores. There were a total of 9 major BGC classes.

(shown in Fig. 5). Results show that these halophilic fungi carry about 2 to 82 number of BGCs per genome with mean value being 16.9090 and standard deviation being 24.9378. Among the 23 fungal halophiles, the highest number of BGCs were occupied in 3 positions, the first being *Aspergillus ochraceus* (n=82), then *Aspergillus wentii* (=78) and then *Penicillium chrysogenum* (n=62). Two genomes contain 2 BGCs, *Candida parapsilosis* and *Meyerozyma guilliermondii*. 11 genomes contain 3 BGCs- *Candida glabrata*, *Debaryomyces hansenii*, *Debaryomyces subglobosus*, *Kregervanrija fluxuum*, *Metschnikowia bicuspidate*, *Pichia membranaefaciens*, *Rhodotorula glutinis*, *Rhodotorula mucilaginosa*, *Saccharomyces cerevisiae*, *Zygosaccharomyces mellis* and *Zygosaccharomyces rouxii*. The most common BGC classes recorded were Terpenes and NRPS-like.

The number of NRPS-like ranged from as high as 20 (found in *Aspergillus ochraceus* and *Wickerhamiella versatilis*) and as low as 1 (found in 11 genomes). There were no NRPS-like biosynthetic gene cluster in 2 genomes- *Rhodotorula glutinis* and *Rhodotorula mucilaginosa*. The second highest number of BGCs were T1PKS (79) which ranged from as high as 27 (*Aspergillus wentii*) and as low as 2 (*Hortaea werneckii*). The antismash result of *Aspergillus ochraceus* showed 100% similarity towards T1PKS and NRPS and 71% similarity towards NRPS-like. Most similarly known clusters were also present, (n=27) and the clusters with highest similarity include Cichorine, Aspergillic acid, nidulanin A, Act-Toxin II, viriditoxin, 6-methylsalicylic acid, Ochratoxin A, Terezine D and Xenolozoyenone. The second highest number of BGCs recorded in *Aspergillus wentii* showed 100% similarity towards T1PKS and Terpenes. Most similarly known clusters of *A.wentii* (n=18) include naphthopyrone, clavatic acid, citrinin, patulin, nidulanin A and cytochalasin E/K. The next highest number of BGCs were recorded in *Penicillium chrysogenum*.

P.Chrysogenum has been an organism that gained a significant value in the past several years with its applications in vaccine. *P.chrysogenum* showed 100% similarity to having T1PKS and terpene clusters. NRPS showed 75% similarity. Most similarly known clusters here (n=17) include andrastin A, chrysoxanthone A/B/C, sorbicillin, penicillin and naphthopyrone. There was no specific pattern observed in connection to the number of BGCs and the assembly level of the genome, the highest number of BGCs were recorded in genomes with Chromosomal and Scaffold assembly level. Considering the previous statement, BGC count as low as 2 and 3 were also recorded in the scaffold and chromosomal level. The number of BGCs in contig level assembly

ranged from 2 to 14. There were multiple copies of a single BGC, the highest number was found in NRPS-like. The entire set of genomes harboured 118 NRPS-like. The second highest number was seen in T1PKS, with 79 copies. The third were terpenes, with 76 copies, then NRPS (n=73) and so on.

Aspergillus ochraceus contained the highest number of BGCs (n=82). *A. ochraceus* is generally known as the food spoilage fungus. *Aspergillus* are well known for their roles in drugs, toxins and other enzymes in industry. They are environmentally tolerant and grow fast. Ochratoxin, a secondary metabolite extracted from *Aspergillus ochraceus* have a range of properties. They can be used to control the quality of food and feed products. They are used to test the contamination of products and check for safety standards. They have antifungal properties and can be used to control plant diseases. Biological toxic studies performed on rats, trout and mice showed carcinogenic potency of ochratoxin[31]

There were some uncommon BGCs that were found only in a few genomes, like Fungal RiPP-like and T3PKS were found only in 1 genome. T3PKS (n=1) was found in *P. chrysogenum* and Fungal RiPP-like (n=1) was found in *Aspergillus wentii*. It was also noted that a strain may have multiple classes of BGC classes. For instance, all the genomes have at least 1 cluster class of Terpenes and NRPS-like.

We found that some BGCs, known as hybrid BGCs contain genes that code for more than one enzyme joined together in a variety of combinations. In *A. ochraceus*, hybrids like NRPS+ indole and T1PKS+ NRPS was found with 100% similarity and NRPS+indole was found with 55% similarity. In *Aspergillus wentii*, T1PKS+ NRPS-like were found with 50% and NRPS-like+Terpene was found with 33% similarity. *Penicillium chrysogenum* had NRPS+T1PKS hybrid with 100% similarity.

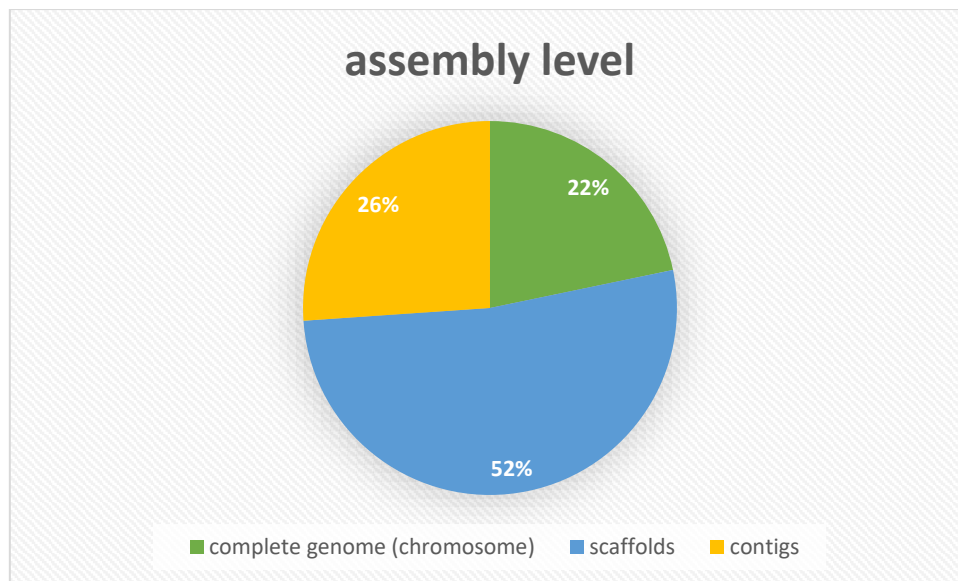


Fig. 2. Assembly level percentages of fungal halophiles mined from Halodom

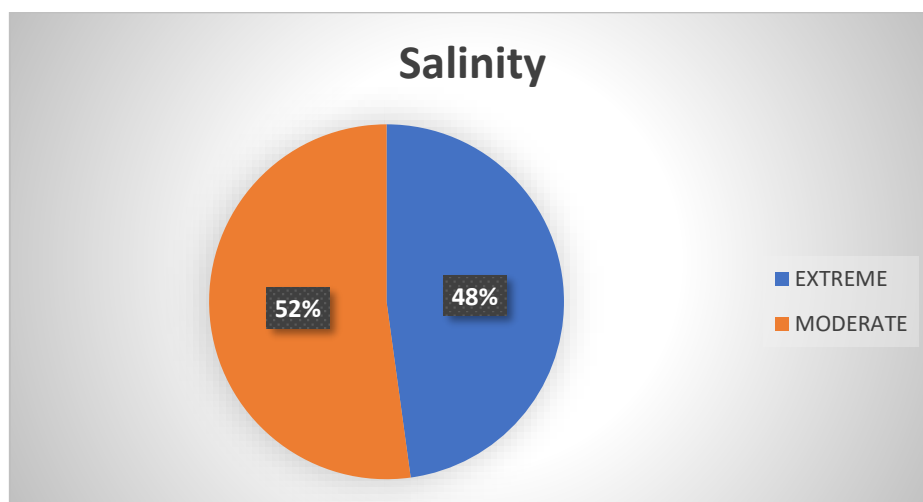


Fig. 3. Salt tolerance level of fungal halophiles mined from halodom

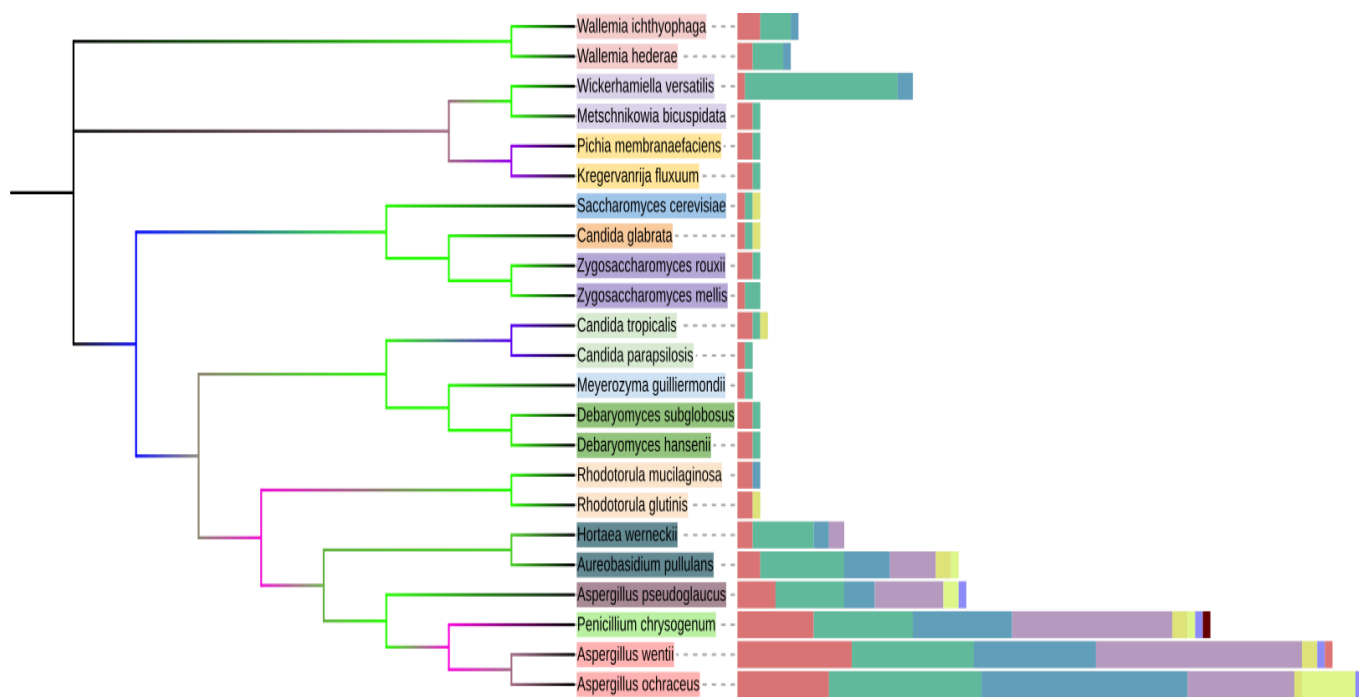


Fig.4. The distribution of BGCs across the strains with the phylogenetic tree and the number of BGCs.

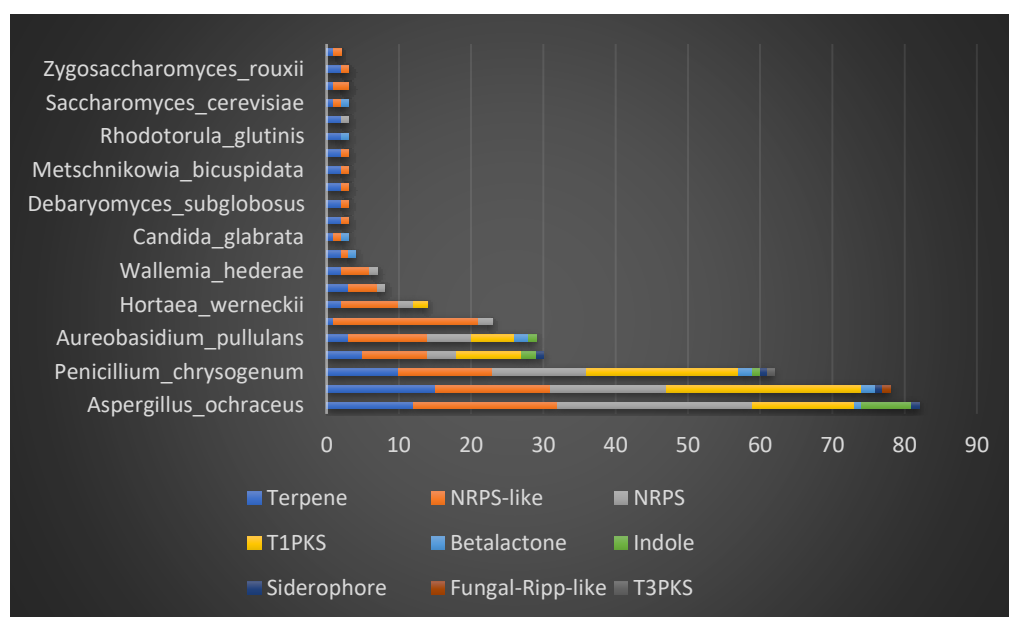


Fig. 5. The diversity of BGCs in Halophilic fungi: the number and type of BGC clusters per organism

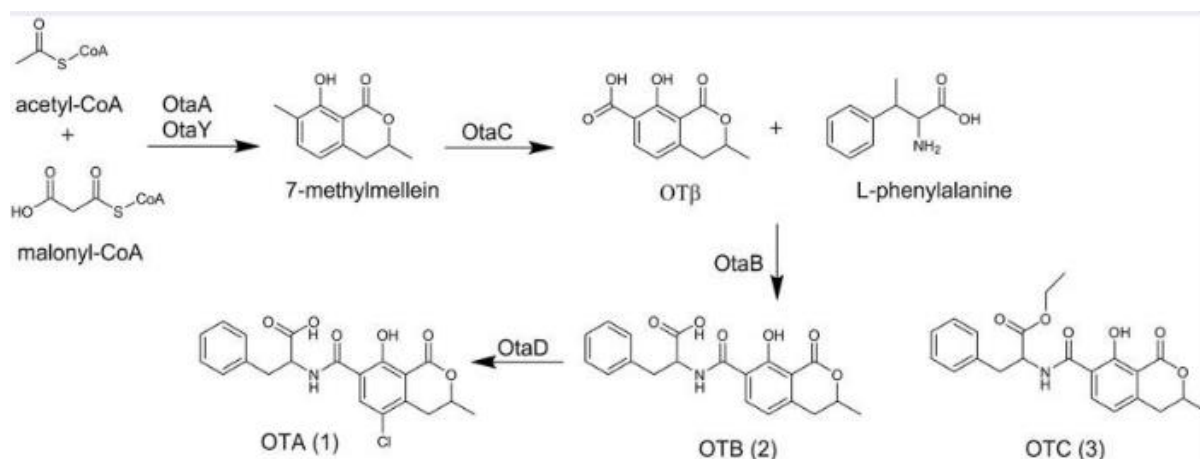


Fig. 6. The structure and biosynthesis of ochratoxins [31]

5.2 DISCUSSION

The secondary metabolite discovery has been limited by many factors. There are a number of applications of natural products that are being used in various sectors of life. The secondary metabolites synthesized by halophilic fungi have diverse potential applications in pharmaceuticals, bioremediation, food industry, agriculture and cosmetics industry. Here, in 23 fungal organisms we found a plethora of cryptic BGCs that may unlock newer potential that has not been discovered yet. The data obtained by in-silico analysis will promote the development of new therapeutics and other new sources of drugs when combining with the usage of these genome mining methods. In this study, the extraction and analysis of BGCs is highly dependent on the data available in the public databases. Halophiles require <3% NaCl to grow [28]. The data we obtained opened up a range of most similarly known clusters which contained NPs like viriditoxin, citrinin, penicillin, sorbicillin and aspergillic acid.

Citrinin is a mycotoxin known to have anticancer [26] and neuroprotective properties [27]. **viriditoxin**, a "classical" secondary metabolite is previously reported to have antibacterial and other properties [29]. It is also active against cancer cell lines and is proposed for therapeutic applications. Apart from these, there was a production of glycerol, erythritol, arabitol and mannitol and the involvement of alkali metal transporters (K^+/Na^+) as osmoadaptation by *Hortaea werneckii*. [30]. The presence of Hybrids increases the number of BGCs present. The discovery of biosynthetic gene clusters have been limited due to traditional practises and many other reasons like the functional or structural integrity of the organism when transferred from its natural

habitat. Taking the examples of halophilic fungi, the salt-tolerant and highly adaptable *H. werneckii* and the strictly halophilic *W. ichthyophaga* inhabit environments with low a_w and high concentrations of toxic inorganic ions. To cope with these harsh environmental conditions, they employ many general and specific molecular mechanisms. When cells are exposed to an environment with high osmotic pressure (salinity), they must react quickly to the resulting loss of water. In *H. werneckii* and *W. ichthyophaga* this is achieved by the synthesis of glycerol and other compatible solutes. In many natural hypersaline environments, the concentration of toxic Na ions is much higher than that of K ions, so mechanisms to maintain a stable and high intracellular K/Na ratio are essential for survival in such environments. Both *H. werneckii* and *W. ichthyophaga* can maintain high K/Na ratios over a wide range of environmental Na concentrations. In *H. werneckii*, this homeostasis is maintained by the regulated transport of K and Na across the plasma membrane, as cation transporters are diverse and highly abundant in this fungus.

The *insilico* methods in use today help uncover a number of Natural products within the organism. The presence of a specific genetic makeup is what might help halophiles survive in extreme salt conditions and produce secondary metabolites for their own survival. These secondary metabolites can also be exploited by humans in order to discover new therapeutics and other industrial products. The number of halophiles present in halodrom contains limited number of genome sequences which is a drawback considering the number of natural products that may be uncovered for a large number of halophiles. Medicinal uses like their roles in anticancer and other neuro diseases may pave a way for fungal halophiles and their culture to prepare drugs in pharmaceutical area.

Industrial uses of these fungal halophiles shan't be ignored, for instance, the use of terpenes in essential oils and their possession of a wide range of biological activities including anticancer, anti-inflammatory, antimicrobial and antioxidants [6]. The presence of a wide range of biological activities of each biosynthetic gene cluster opens up new spectra of how much of it hasn't been exploited yet. A plethora of cryptic BGCs can be discovered and their occurrence and uses can be explored with increasing information and genome mining can be of immense impact to undoubtedly improve the process and increase the spectrum of information available on these extremophiles.

CHAPTER 6

CONCLUSIONS

Large scale projects around fungi like 1000 genomes project was done to find the taxonomic family within fungi. We have mined the biosynthetic gene clusters of 23 fungal organisms that are halophilic in nature. The extraction of biosynthetic gene clusters have given a range of families of gene clusters which give secondary metabolites. These extremophiles live in habitats that probably cannot be created exactly in a scientific laboratory, which means their genomic design has a makeup that helps them survive in such an extreme habitat. For example, siderophore helps in maintaining osmotic pressure in the organism by maintaining salt and water levels in its system. Halophilic fungi have also been widely studied for these purposes and their use to mankind.

A phylogenetic tree was constructed around these clusters. There have been several studies around the phylogeny of biosynthetic gene clusters. The conclusions and analysis for known sequences can be derived easily, but for unknown sequences, the clades that are present in close proximity to a known cluster may indicate that they have a close relationship, meaning they may have similar properties. The close relationship between different biosynthetic gene clusters in an organism can be understood by constructing the phylogenetic tree. These analyses give an understanding that the genomes of fungal halophiles are an abundance of natural products. A lot of them are still unexplored, and the advances in genome mining technologies are paving a way to work around these clusters (mining and analysis) with ease.

REFERENCES

- [1] Śliżewska W, Struszczyk-Świta K, Marchut-Mikołajczyk O. Metabolic Potential of Halophilic Filamentous Fungi-Current Perspective. *Int J Mol Sci.* 2022;23(8):4189. Published 2022 Apr 10. doi:10.3390/ijms23084189
- [2] Rutledge, Peter J.; Challis, Gregory L. (2015). Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nature Reviews Microbiology*, 13(8), 509–523. doi:10.1038/nrmicro3496
- [3] Franck E. Dayan, Charles L. Cantrell, Stephen O. Duke, Natural products in crop protection, *Bioorganic & Medicinal Chemistry*, Volume 17, Issue 12, 2009, Pages 4022-4034, ISSN 0968-0896, <https://doi.org/10.1016/j.bmc.2009.01.046>.
- [4] Dai G, Shen Q, Zhang Y, Bian X. Biosynthesis of Fungal Natural Products Involving Two Separate Pathway Crosstalk. *Journal of Fungi.* 2022; 8(3):320. <https://doi.org/10.3390/jof8030320>
- [5] Medema MH, Kottmann R, Yilmaz P, et al. Minimum Information about a Biosynthetic Gene cluster. *Nat Chem Biol.* 2015;11(9):625-631. doi:10.1038/nchembio.1890
- [6] Zhang, QW., Lin, LG. & Ye, WC. Techniques for extraction and isolation of natural products: a comprehensive review. *Chin Med* 13, 20 (2018). <https://doi.org/10.1186/s13020-018-0177-x>
- [7] Steven G Van Lanen; Ben Shen (2006). Microbial genomics for the improvement of natural product discovery. , 9(3), 0–260. doi:10.1016/j.mib.2006.04.002
- [8] Cacho RA, Tang Y, Chooi YH. Next-generation sequencing approach for connecting secondary metabolites to biosynthetic gene clusters in fungi. *Front Microbiol.* 2015;5:774. Published 2015 Jan 14. doi:10.3389/fmicb.2014.00774

[9] Bauer A, Brönstrup M. Industrial natural product chemistry for drug discovery and development. *Nat Prod Rep*. 2014;31(1):35-60. doi:10.1039/c3np70058e

[10] Atanasov, A.G., Zotchev, S.B., Dirsch, V.M. et al. Natural products in drug discovery: advances and opportunities. *Nat Rev Drug Discov* 20, 200–216 (2021). <https://doi.org/10.1038/s41573-020-00114-z>

[11] Jensen PR. Natural Products and the Gene Cluster Revolution. *Trends Microbiol*. 2016;24(12):968-977. doi:10.1016/j.tim.2016.07.006

[12] Lawrence JG, Roth JR. Selfish operons: horizontal transfer may drive the evolution of gene clusters. *Genetics*. 1996;143:1843–1860.

[13] Ren H, Shi C, Zhao H. Computational Tools for Discovering and Engineering Natural Product Biosynthetic Pathways. *iScience*. 2020;23(1):100795. doi:10.1016/j.isci.2019.100795

[14] Benson DA, Cavanaugh M, Clark K, et al. GenBank. *Nucleic Acids Res*. 2018;46(D1):D41-D47. doi:10.1093/nar/gkx1094

[15] Abdel-Razek AS, El-Naggar ME, Allam A, Morsy OM, Othman SI. Microbial Natural Products in Drug Discovery. *Processes*. 2020; 8(4):470. <https://doi.org/10.3390/pr8040470>

[16] Corral P, Amoozegar MA, Ventosa A. Halophiles and Their Biomolecules: Recent Advances and Future Applications in Biomedicine. *Mar Drugs*. 2019;18(1):33. Published 2019 Dec 30. doi:10.3390/md18010033

[17] Gunde-Cimerman N, Ramos J, Plemenitas A. Halotolerant and halophilic fungi. *Mycol Res*. 2009;113(Pt 11):1231-1241. doi:10.1016/j.mycres.2009.09.002

[18] Śliżewska W, Struszczyk-Świta K, Marchut-Mikołajczyk O. Metabolic Potential of Halophilic Filamentous Fungi-Current Perspective. *Int J Mol Sci*. 2022;23(8):4189. Published 2022 Apr 10. doi:10.3390/ijms23084189

[19] Loukas A, Kappas I, Abatzopoulos TJ. HaloDom: a new database of halophiles across all life domains. *J Biol Res (Thessalon)*. 2018;25:2. Published 2018 Jan 15. doi:10.1186/s40709-017-0072-0

[20] Blin K, Shaw S, Kloosterman AM, et al. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res*. 2021;49(W1):W29-W35. doi:10.1093/nar/gkab335

[21] Nilsson RH, Tedersoo L, Ryberg M, Kristiansson E, Hartmann M, Unterseher M, Porter TM, Bengtsson-Palme J, Walker DM, de Sousa F, Gamper HA, Larsson E, Larsson KH, Kõljalg U, Edgar RC, Abarenkov K. A Comprehensive, Automatically Updated Fungal ITS Sequence Dataset for Reference-Based Chimera Control in Environmental Sequencing Efforts. *Microbes Environ*. 2015;30(2):145-50. doi: 10.1264/jsme2.ME14121. Epub 2015 Mar 19. PMID: 25786896; PMCID: PMC4462924.

[22] Hall, T.A. (1999) BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, 41, 95-98.

[23] Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol Biol Evol*. 2021;38(7):3022-3027. doi:10.1093/molbev/msab120

[24] Cardona G, Rosselló F, Valiente G. Extended Newick: it is time for a standard representation of phylogenetic networks. *BMC Bioinformatics*. 2008;9:532. Published 2008 Dec 15. doi:10.1186/1471-2105-9-532

[25] Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res*. 2021;49(W1):W293-W296. doi:10.1093/nar/gkab301

[26]Chang C-H, Yu F-Y, Wang L-T, Lin Y-S, Liu B-H. Activation of ERK and JNK signaling pathways by mycotoxin citrinin in human cells. *Toxicol. App. Pharmacol.* 2009; 237: 281–287.)and neuroprotective (Nakajima Y, Iguchi H, Kamisuki S, Sugawara F, Furuichi T, Shinoda Y. Low doses of the mycotoxin citrinin protect cortical neurons against glutamate-induced. *The Journal of Toxicological Sciences* 2016; 41(2): 311-319.

[27] de Oliveira Filho JWG, Islam MT, Ali ES, Uddin SJ, Santos JVO, de Alencar MVOB, Júnior ALG, Paz MFCJ, de Brito MDRM, E Sousa JMC, Shaw S, de Medeiros MDGF, Dantas SMMM, Rolim HML, Ferreira PMP, Kamal MA, Pieczynska MD, Das N, Gupta VK, Mocan A, Dos Santos Andrade TJA, Singh BN, Mishra SK, Atanasov AG, Melo-Cavalcante AAC. A comprehensive review on biological properties of citrinin. *Food Chem Toxicol.* 2017 Dec;110:130-141. doi: 10.1016/j.fct.2017.10.002. Epub 2017 Oct 7. PMID: 28993214.

[28] Wilson Z.E., Brimble M.A. Molecules derived from the extremes of life. *Nat. Prod. Rep.* 2009;26:44–71. doi: 10.1039/B800164M.

[29] Urquhart, A.S., Hu, J., Chooi, YH. et al. The fungal gene cluster for biosynthesis of the antibacterial agent viriditoxin. *Fungal Biol Biotechnol* 6, 9 (2019). <https://doi.org/10.1186/s40694-019-0072-y>

[30]Gonçalves, M.F.M.; Hilário, S.; Tacão, M.; Van de Peer, Y.; Alves, A.; Esteves, A.C. Genome and Metabolome MS-Based Mining of a Marine Strain of *Aspergillus affinis*. *J. Fungi* 2021, 7, 1091. <https://doi.org/10.3390/jof7121091>

[31] Chen L, Li E, Wu W, et al. The Secondary Metabolites and Biosynthetic Diversity From *Aspergillus ochraceus*. *Front Chem.* 2022;10:938626. Published 2022 Aug 25. doi:10.3389/fchem.2022.938626

[32] Ekiert HM, Szopa A. Biological Activities of Natural Products. *Molecules.* 2020 Dec 7;25(23):5769. doiss: 10.3390/molecules25235769. PMID: 33297511; PMCID: PMC7730830.

[33] 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

[34] Corson, T. W. & Crews, C. M. Molecular understanding and modern application of traditional medicines: triumphs and trials. *Cell* 130, 769–774 (2007)

[35] .. 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. Chrystal, and N. Nicholson (ed.), *Biodiversity: new leads for pharmaceutical and agrochemical industries*. The Royal Society of Chemistry, Cambridge, United Kingdom

[36] Grabley, S., and R. Thiericke (ed.). 1999. *Drug discovery from nature*, p. 3–33. Springer-Verlag. Berlin, Germany.

[37] 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.

[38] Terlouw BR, Blin K, Navarro-Muñoz JC, Avalon NE, Chevrette MG, Egbert S, Lee S, Meijer D, Recchia MJ, Reitz ZL, van Santen JA, Selem-Mojica N, Tørring T, Zaroubi L, Alanjary M, Aleti G, Aguilar C, Al-Salihi SAA, Augustijn HE, Avelar-Rivas JA, Avitia-Domínguez LA, Barona-Gómez F, Bernaldo-Agüero J, Bielinski VA, Biermann F, Booth TJ, Carrion Bravo VJ, Castelo-Branco R, Chagas FO, Cruz-Morales P, Du C, Duncan KR, Gavriilidou A, Gayrard D, Gutiérrez-García K, Haslinger K, Helfrich EJN, van der Hooft JJJ, Jati AP, Kalkreuter E, Kalyvas N, Kang KB, Kautsar S, Kim W, Kunjapur AM, Li YX, Lin GM, Loureiro C, Louwen JJR, Louwen NLL, Lund G, Parra J, Philmus B, Pourmohsenin B, Pronk LJ, Rego A, Rex DAB, Robinson S, Rosas-Becerra LR, Roxborough ET, Schorn MA, Scobie DJ, Singh KS, Sokolova N, Tang X, Udway D, Vigneshwari A, Vind K, Vromans SPJM, Waschulin V, Williams SE, Winter JM, Witte TE, Xie H, Yang D, Yu J, Zdouc M, Zhong Z, Collemare J, Linington RG, Weber T, Medema MH. MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. *Nucleic Acids Res*. 2023

Jan 6;51(D1):D603-D610. doi: 10.1093/nar/gkac1049. PMID: 36399496; PMCID: PMC9825592.

[39] Agrawal P, Khater S, Gupta M, Sain N, Mohanty D. RiPPMiner: a bioinformatics resource for deciphering chemical structures of RiPPs based on prediction of cleavage and cross-links. *Nucleic Acids Res.* 2017 Jul 3;45(W1):W80-W88. doi: 10.1093/nar/gkx408. PMID: 28499008; PMCID: PMC5570163.

[40] Hjörleifsson Eldjárn G, Ramsay A, van der Hooft JJJ, Duncan KR, Soldatou S, Rousu J, Daly R, Wandy J, Rogers S. Ranking microbial metabolomic and genomic links in the NPLinker framework using complementary scoring functions. *PLoS Comput Biol.* 2021 May 4;17(5):e1008920. doi: 10.1371/journal.pcbi.1008920. PMID: 33945539; PMCID: PMC8130963.

[41] Liu M, Li Y, Li H. Deep Learning to Predict the Biosynthetic Gene Clusters in Bacterial Genomes. *J Mol Biol.* 2022 Aug 15;434(15):167597. doi: 10.1016/j.jmb.2022.167597. Epub 2022 May 6. PMID: 35526560.

[42] 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.

[43] Atri, N. S., Sharma, D., Ahmed, M., Akhter, Y., & Hasnain, S. E. (2018). Halocin C8: a potent antimicrobial peptide from the halophilic archaeon *Haloferax* sp. strain C8. *Microbiology*, 164(9), 1168-1180. DOI: 10.1099/mic.0.000692. PMID: 29968438.

[44] 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.

[45] Chung D, Kim H, Choi HS. Fungi in salterns. *J Microbiol.* 2019 Sep;57(9):717-724. doi: 10.1007/s12275-019-9195-3. Epub 2019 Aug 27. PMID: 31452042.

[46] Nina Gunde-Cimerman, Jose Ramos, Ana Plemenitaš, Halotolerant and halophilic fungi, Mycological Research, Volume 113, Issue 11, 2009, Pages 1231-1241, ISSN 0953-7562, <https://doi.org/10.1016/j.mycres.2009.09.002>.

[47] <https://revive.gardp.org/resource/natural-products/?cf=encyclopaedia>