CAPITAL UNIVERSITY OF SCIENCE AND TECHNOLOGY, ISLAMABAD

Genomic Characterization of Halophilic Bacteria from the Khewra Salt Mine

by

Irum Hassan

A thesis submitted in partial fulfillment for the degree of Master of Science

in the

Faculty of Health and Life Sciences
Department of Bioinformatics and Biosciences

2025

To my father, who always believed in me.

To my husband, who stood firmly by my side.

And to my sister, Arzoo Rehman, for her constant moral and emotional support.

CERTIFICATE OF APPROVAL

Genomic Characterization of Halophilic Bacteria from the Khewra Salt Mine

by

Irum Hassan

(MBS233016)

THESIS EXAMINING COMMITTEE

S. No.	Examiner	Name	Organization
(a)	External Examiner	Dr. Attiya Kanwal	IIU, Islamabad
(b)	Internal Examiner	Dr. Sami Ullah Jan	CUST, Islamabad
(c)	Supervisor	Dr. Sahar Fazal	CUST, Islamabad

Dr. Sahar Fazal
Thesis Supervisor
August, 2025

Dr. Syeda Marriam Bakhtiar
Head
Dept. of Bioinfo. & Biosciences
August, 2025

Dr. Sahar Fazal
Dean
Faculty of Health & Life Sciences
August, 2025

Author's Declaration

I, **Irum Hassan** hereby state that my MS thesis titled "**Genomic Characterization of Halophilic Bacteria from the Khewra Salt Mine**" is my own work and has not been submitted previously by me for taking any degree from Capital University of Science and Technology, Islamabad or anywhere else in the country/abroad.

At any time if my statement is found to be incorrect even after my graduation, the University has the right to withdraw my MS Degree.

(Irum Hassan)

Registration No: MBS233016

Plagiarism Undertaking

I solemnly declare that research work presented in this thesis titled "**Genomic Characterization of Halophilic Bacteria from the Khewra Salt Mine**" is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and Capital University of Science and Technology towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS Degree, the University reserves the right to withdraw/revoke my MS degree and that HEC and the University have the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized work.

(Irum Hassan)
Registration No: MBS233016

Acknowledgement

All praise and gratitude are for Almighty Allah, whose countless blessings, guidance, and mercy enabled me to complete this research work successfully.

I wish to express my profound appreciation to my research supervisor, Dr Sahar Fazal , for their invaluable guidance, insightful suggestions, and continuous encouragement throughout the course of this study. Their support and mentorship played a critical role in shaping the direction and depth of this research.

I am sincerely thankful to the University of Central Science and Technology (CUST) for providing a supportive academic environment and the necessary resources that facilitated the successful completion of this research project. I would also like to extend my deepest gratitude to my parents, Laiq ur Rehman and Andheeda Begum, whose endless prayers, love, and unwavering support have been my strength throughout this journey. Their sacrifices and belief in my potential have always been a source of motivation.

A special note of thanks to my husband,Tahir Anjum, for his constant patience, understanding, and encouragement during this demanding period. His support has been instrumental in helping me stay focused and determined. I am equally grateful to my sister, Arzoo Rehman, and all my siblings, family members, and friends for their consistent support, kind words, and sincere prayers. I would like to acknowledge the sincere help and cooperation of my senior, Muneeba, whose timely advice and guidance were of immense value during the course of this research. Lastly, I extend my thanks to all those who contributed in any way to the successful completion of this thesis.

(Irum Hassan)

Abstract

Halophilic bacteria flourishing in extreme saline environments, such the Khewra salt mines, exhibit distinctive modifications that render them advantageous for biotechnological and industrial uses. This research sought to extract, describe, and conduct genetic analysis of halophilic bacteria from desolate saline ponds in the Khewra mines. Samples were obtained from saline deposits and liquid brines, subsequently cultured on nutrient agar enriched with NaCl (0.5%–2%) to isolate halophilic bacteria. Whole-genome sequencing (Illumina platform) was conducted on a chosen strain, and the raw data were analyzed using Geneious Prime (v2024.0.2) for quality control, de novo assembly, and annotation.The assembled genome (1.14 Mb) exhibited a high coding density, comprising 4289 protein-coding genes, 82 tRNAs, and 3 rRNAs. Functional annotation revealed genes associated with osmotic stress adaptation, encompassing ion transporters and pathways for compatible solute production. CRISPR-Cas study identified a potential Cas8a1a system, indicating adaptive immunity to phage predation. Prediction of antibiotic resistance genes (via CARD-RGI) identified six resistance determinants, including vanY (vanG cluster) and FosB, suggesting potential multidrug resistance. Prophage screening (PHASTEST) revealed five integrated phage areas including 175 connected genes, some of which are related to stress-induced lysis in industrial strains. This research offers genetic insights into the halophilic bacterium *Bacillus licheniformis*, emphasizing its survival strategies under extreme salinity and prospective biotechnological uses. The results highlight the dangers of phage contamination in industrial fermentation, as prophage-associated genes can initiate lytic cycles under stress conditions. Additional investigation of these extremophiles may produce innovative enzymes and stress-resistance genes applicable in industrial contexts.

Keywords: Halophilic bacteria, *Bacillus licheniformis*, whole-genome sequencing, CRISPR-Cas, antibiotic resistance, prophages, Khewra salt mines

Contents

List of Figures

List of Tables

Abbreviations

BLAST	Basic Local Alignment Search Tool
BUSCO	Benchmarking Universal Single-Copy Orthologs
CARD	Comprehensive Antibiotic Resistance Database
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CTAB	Cetyltrimethylammonium Bromide
EDTA	Ethylenediaminetetraacetic Acid
EMBOSS	European Molecular Biology Open Software Suite
Glimmer	Gene Locator and Interpolated Markov ModelER
KEGG	Kyoto Encyclopedia of Genes and Genomes
MAFFT	Multiple Alignment using Fast Fourier Transform
MIRA	Mimicking Intelligent Read Assembly
Pfam	Protein Families Database
Prodigal	Prokaryotic Dynamic Programming Gene-finding Algorithm
ProtBERT	Protein Bidirectional Encoder Representations from Transformers
RPM	Revolutions Per Minute
SDS	Sodium Dodecyl Sulfate
TBE Buffer	Tris-Borate-EDTA
TE Buffer	Tris-EDTA Buffer
tRNA	Transfer RNA
VFDB	Virulence Factor Database

Chapter 1

Introduction

Salt is a great ionic compound that is used in food, medicines, cattle feedstock, detergents, chemicals etc. for its extensive health and chemical properties. Salts come in a variety of forms from throughout the globe, each with unique characteristics, hues, and flavors. The world's salt reserves are found underground, in lakes, or in caves, and are extracted using both conventional and cutting-edge modern techniques.

Pakistan is renowned for its prominent Salt Ranges, which extend from Kalabagh to Jhelum, encompassing districts such as Chakwal and Jhelum in Punjab. The geological study has identified 10.54 billion tons of rock salt reserves in Bahadur Khel in District Kohat, indicating the abundance of rock salt deposits in Pakistan. Pakistan is among the few countries that possess a comprehensive understanding of all types of salt available globally, including rock salt, sea salt, and Salt Lake salt. The salt mines in the Indian salt range formation are among the oldest in the subcontinent, located at the eastern terminus of the salt range. The extraction of salt has been successful in several locations such as Kewra, Orchha, Panabagh, Jatta, and Bahadurfil, from which several thousand tons of salt are mined each year. Significant salt reserves are found in various locations in KPK, specifically at the Jatta, Bahadurfil, and Karak salt mines [1].

Based on the quality of the rock salt, four distinct varieties of crystals are identified: pink, white, soft nungs, and soft crystalline lime. The general situation of the

salt range on the island side is underutilized, and output remains insufficient, considering the deposits of this crucial mineral. The aggregate output estimate for all mines is around 600,000 tons per year. The Salt Lake has exceptional quality and abundance. The lakes are situated approximately 250 km northeast of Karachi, with an extraction of 500 tons occurring day from these 10-12 lakes, resulting in an annual production average of around 150,000 tons. Underneath the Kuldana Formation, the Kohat salt is thought to be a regularly bedded deposit. It is found in a variety of cyclically folded anticline structures that are bordered on all sides by extremely steep, almost vertical Sizalik and Eocene strata. The Kohat rock salt near Bahadur Khel outcrops span an estimated 12 km in length and 0.5 km in width, according to the Geological Survey of Pakistan. An Early Eocene age has been attributed to the Bahadur Khel salt [2].

The Salt Range salt is predominantly pink to red, whereas Kohat salt is usually white to gray. The salt from the Salt Range is typically coarse, crystalline to massive, and comparatively dense. Kohat salt is often fine- to medium-crystalline, friable, loose, and exhibits a sweet look. The kohat salt is characterized by a porphyritic texture, which is rare among the deposits of the Salt Range. The Kohat salt is situated at the base of the Salt Range. The Khewra Salt Mines are regarded as the earliest salt mines in the Indian subcontinent's history. In this region, salt manifests in an uneven, boom-like formation, comprising seven substantial salt seams with a total thickness of roughly 150 meters. The salt has a color range from translucent and white to pink, reddish, and even brownish red. In specific strata, it has a crystalline configuration. Within the mine, there exist remarkable alternating bands of red and white salt, forming a visually captivating pattern.The mine comprises 18 operational levels, with a total drivage length over 40 kilometers [3].

Sodium chloride is essential for maintaining the acid-base equilibrium in the body. It is also crucial for the synthesis of gastric acid, a constituent of hydrochloric acid in the stomach. Moreover, it possesses significant physiological activities. Sodium chloride is an essential component of blood and aids in the regulation of the body's potassium levels. The human body comprises roughly 2.5 grams of salt for every

kilogram of body weight. The kidneys consistently discharge salt as a 2% solution, necessitating continual replenishment. Humans cannot consume seawater due to its salt concentration of approximately 3.5%, which surpasses the body's tolerance and results in dehydration [4].

Salt is utilized in the diets of humans and animals, as well as for the prevention of heat-related ailments, cramping, and other medical applications. A prevalent therapeutic application is for isotonic saline solutions. It is frequently utilized in the preservation and seasoning of meals, including the curing of meat and fish, the preparation of pickles, the preservation of various vegetables, and numerous more commonplace domestic applications. Salt is utilized in the industrial production of various chemical products, such as calcium chloride, chlorine dioxide, chlorine gas, sodium chlorate, sodium fluorosilicate, sodium hypochlorite, sodium perchlorate, soda ash, sodium metal, sodium sulfate, hydrochloric acid, and caustic soda. It functions as a crucial element in multiple processes, including serving as a freezing point depressant in ice cream production and in refrigerating brines [4].

Table salt, or sodium chloride, is widely used in the culinary industry as a flavoring ingredient, textural garnishment, preservative, and food protector. Commercial table salt comes from saline settings such ponds, lakes, brine springs, rock salts, and seawater. Prior to distribution, most marketable salts are purified and finely ground. Furthermore, saline environments can promote the growth of halophiles, or salt-loving microorganisms. Halophiles are recognised as a distinct class of extremophiles with severe salinity requirements, including species from the "Three Domains Classification". Furthermore, these microorganisms are important for use in the biomedical sector because to their unique metabolism, low feeding needs, and ability to adapt to harsh settings [5].

Pakistan is a formidable competitor in the salt mining sector, possessing the world's largest salt resources, with over 10 billion tonnes of verified reserves that are 98 percent pure. The largest salt man deposits are predominantly observable in the Kohat-Potwar plateau of the Upper Indus (Krishnan, 1966). The Punjab Saline Series, also known as The Salt Range Formation, represents the earliest salt deposits dating back to the Edicambrian period, only exposed in the Salt

Range region at the southern boundary of the Potwar sub-basin (PB). The Bahadurkhail salts, referred to as the Kohat saline series, are younger Eocene salt deposits located in the Kohat sub-basin (KB) [6].

The Karak salt mines are located in $32^o39\ \ 0$ N, $73^o1\ \ 0$ E, at an elevation of 678 m (2227 feet), with rock salts varying in color from light grey to dark grey. Precipitation in the region is infrequent and irregular. Moreover, the atmosphere is always evolving. The region is very cold in winter and exceedingly hot in summer. The geological stratum indicates that these rocks are of the Eocene age. The mine contains raw salt with 7.98 mg/kg of water-insoluble impurities and 0.12 percent moisture content. The composition of Karak salt is characterized by a medium to fine crystalline structure, exhibiting a loose and friable texture with a subtly sweet consistency. The porphyritic nature of its halite differentiates it from other salt range crystals [7].

Salt mines serve as the primary supply of rock salt globally, including the Khewra Salt Mine in Punjab, Pakistan, which is the second largest in the world and yields the finest pink salt. Numerous studies have identified viable bacteria within fluid inclusions in halite crystals. During the crystallization process, these bacteria may become ensnared in the fluid inclusions and persist until extraction and packaging. The proteome of *Halobacterium salinarum* cells was examined following two months of entrapment in halite brine inclusions. The results closely resembled those of stationary phase liquid cultures, with notable down-regulation of ribosomal proteins suggesting acclimatization to the halite environment. Likewise, research has shown the persistence of highly halophilic archaeal communities in crystallizer ponds and halite crystals. The carotenoid-rich microbial community enhances salt production and yield by increasing light absorption and local temperature [8].

The salt mines are influenced by connected humidity and temperature conditions. The fluctuations in the mine's interior temperature are affected by the external natural environment, and similarly, humidity levels within mines are determined by external conditions. In a sealed container containing hygroscopic salt, relative

humidity fluctuates at a lower percentage than absolute humidity during temperature variations. Brine solutions, stalactites, salt pans, stalagmites, and holes in salt walls are all present in these locales [9].

Halophiles have developed several adaptations to thrive in saline environments. The "salt in" technique suggests that the presence of K+ and, to a lesser extent, Na+, is essential for equilibrating the microbe's internal osmotic pressure with that of its surroundings. Moreover, certain halophiles collect suitable organic solutes in the cytoplasm to modulate intracellular osmotic pressure. Moreover, Archaea are the most extreme halophiles, requiring optimal salt concentrations surpassing 25% w/v dissolved salts. The study sought to examine hypersaline ecosystems and their habitats, focusing on the physical characteristics, the raw quality of pink salt, and the 16S RNA diversity of microorganisms, particularly halophilic archaea and bacteria from Pakistan's Khewra Salt Mine [10].

Diverse saline conditions have produced a multitude of bacterial species. Gram-negative halophilic microbes encompass *Marinomonas*, *Alteromonas*, *Acinetobacter*, *Vibrio*, and *Pseudomonas*. Similarly, microbes from the genera *Bacillus*, *Salinococcus*, *Marinococcus*, and *Sporosarcina* have been identified in slatterns and saline soils [11]. Bacteria from the *Halomonas*, *Magadiensis*, and *Virgibacillus*, *Halodenitrificans* species, were identified in water samples collected from the Khewra salt mine. An examination was conducted on bacterial diversity in the Khewra salt mine. Khewra is the largest salt mine globally in terms of surface area and the second largest producer of rock salt in the world. The finding of tertiary microfossils in the salt range encompasses about 600 million years of Earth's history. This represents the inaugural research of its kind to investigate the biochemical and molecular diversity of culture able bacteria in the Khewra salt mine.

Halophiles may thrive at NaCl concentrations between 0.2 to 5.1 M and are present in the domains of Eukarya, Archaea, and Bacteria (or 2 to 30%). Halophiles are classified as severe, extreme, or mild based on NaCl content.The *Halobacteriaceae* family is the largest assemblage of halophilic Archaea, with 36 genera and 129 species and contains the largest number of salt-dependent species [12].

The cell encounters restricted water supply and elevated inorganic ion concentrations in a hypersaline environment. Halophilic bacteria have evolved mechanisms to withstand elevated salt concentrations over time. The primary processes involve the expulsion of detrimental inorganic ions and the buildup of suitable solutes. In the cell membrane of some halophiles, a chromo protein termed as bacteriorhodopsin functions as a trans-membrane proton pump to elevated salt concentrations, these bacteria synthesize a substantial quantity of amino acids with a high ratio of acidic to basic types, gas vesicles, salt-resistant proteins, internal and extracellular enzymes, and solutes [7].

The predominant focus of research has been on the Karak salt mine in Pakistan. Research indicates that the archaeal and bacterial communities in the Karak Salt Mine exhibit variability, with Class *Acetothermia* comprising the top five phyla, alongside members of the phyla *Firmicutes*, *Proteobacteria*, *Euryarchaeota*, and *Bacteroidetes*. The predominant bacteria found in the samples were *Salinibacter* and *Marinobacter*, suggesting their significant involvement in this ecosystem. The perchlorate-resistant bacteria (*Halomonas* and *Marinobacter*) and archaea (*Haloferax* and *Haloarcula*) in the salt mine, along with their tolerance to hypersalinity, provide this environment an optimal model for astrobiological investigations. A prospective future endeavor is a comprehensive research and analysis of the microbial cultural variety within Pakistan's salt range, recognized as one of the continent's oldest salt deposits. Despite halophiles' role in the development of terrestrial ecosystems, their products and identities remain largely unknown [13].

The Khewra Salt Mine is the most ancient mountain range in Pakistan, renowned for its mineral-rich salt deposits.The Khewra Salt Mine is presently managed by the Pakistan Mineral Development Corporation (PMDC).Khewra Salt exhibits a spectrum of colors, from translucent to pink and reddish hues, and is extracted using the chamber and pillar process, resulting in an annual production of 400,000 tons of salt. This salt mine possesses unique flora and fauna, including several distinct microbial communities that need further exploration. Nevertheless, the investigation of its microbial richness started only a few years prior. At now, six halophilic microorganisms have been identified: *Bacillus tequilensis*, *Bacillus*

xiamenensis, *megaterium*, *Staphylococcus pasteuri*, *Pseudomonas aeruginosa*, and *Pseudomonas putida*.

The Khewra salt mine remains unexamined for its unique microorganisms, which have developed throughout time and possess distinct salt tolerance capacities. Rich in ancient salt deposits and unique local flora and fauna, these salt mines possess significant social, environmental, scientific, and commercial potential for advancement and development [14].

The study of halophilic bacteria from salt mines has significantly enhanced our understanding of microbial adaptation to extreme salinity conditions. Comparative genomics has shown the evolutionary strategies employed by halophiles, including horizontal gene transfer and genome simplification, which improve survival in hypersaline conditions [15].

These investigations reveal a variety of genes implicated in osmotic balance, stress response, and unique metabolic processes, encompassing the production of compatible solutes, salt-tolerant enzymes, and secondary metabolites. Comparative genomics has shown the evolutionary strategies employed by halophiles, including as horizontal gene transfer and genome streamlining, which augment survival under hypersaline conditions.

These findings augment our understanding of extremophile biology and offer prospects for biotechnological applications in bioremediation, enzyme production, and industrial biotechnology in high-salinity contexts [16].

1.1 Problem Statement

Pakistan's salt mines, particularly the Khewra Salt Mine—one of the largest and oldest in the world—represent unique hypersaline ecosystems with rich but largely unexplored microbial diversity. However, there is a significant gap in our understanding of the genomic architecture and functional potential of these indigenous microbial communities. Comprehensive genomic studies are needed to uncover the molecular mechanisms of adaptation.

1.2 Aim and Objectives

The aim of this study is to perform whole genome sequencing and genomic analysis of halophilic bacteria isolated from the Khewra Salt Mine in Pakistan.

The objectives of this research are as following:

1. To isolate and identify halophilic bacterial strains from the Khewra Salt Mine.
2. To perform whole genome sequencing of selected halophilic isolates using high-throughput sequencing platforms.
3. To perform structural and functional annotation to characterize a bacterial genome.

Chapter 2

Literature Review

Salt occurs as a crystalline mineral, commonly referred to as halite. The composition of salt significantly differs based on the extraction source. The salts from diverse sources comprise numerous minerals, residues of heavy metals, and other substances. The most prevalent compound is sodium chloride, NaCl. NaCl is table salt and appears as salt crystals due to the strong ionic bonds that held atoms together [17].

2.1 Types of Salt

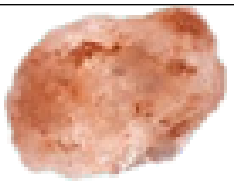
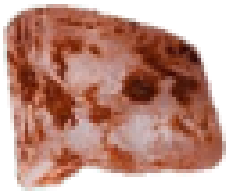
There are various types of salt based on their location of formation and composition. The Table emphasizes the origin, characteristics, and use of five salts used in food.

TABLE 2.1: Types of Salts That Are Found In Food [18]

Name	**Origin**	**Color & Texture**	**Benefits**	**Illustration**
Iodized Salt	Modified Table Salt, contains Iodine	White, Cubic crystals	Cooking, thyroid function	

continued on next page

Table 2.1 continued from previous page

Name	Origin	Color & Texture	Benefits	Illustration
Himalayan Pink Salt	Khewra Salt Mines	White to deep red, less crystalline	Cooking, Beauty products, Lamps and Ornaments	
Celtic Salt (sea salt)	Coastal regions in France	Light grey to grey, high moisture content, crystalline	Cooking, brining	
Hawaiian Red Sea Salt	Hawaii	Deep red, Crystalline	Seasoning	
Hawaiian Black Lava Salt	Hawaii	Black, Crystalline	Seasoning, Aids digestion	

2.2 Health Aspects of Salt

Salt is beneficial for animals and human consumption. The following are a few reasons salt is considered an important mineral along with its health consequences if taken more than the required amount.

2.2.1 Health Benefits of Salt

Salt is essential for regulating fluid levels in the body, transmitting nerve impulses, facilitating muscular contraction and movement, and enabling nutrition availability and delivery to cells. Edible salt is sodium chloride, which is abundant in sodium. Approximately 90% of sodium consumption is derived from sodium chloride. The health issues and dysfunctions linked to excessive salt consumption

in the diet are attributed to processed foods such as white bread, cereals, buns, biscuits, pastries, cakes, poultry, meat, and meat products. The processed food contains excessive non-discretionary added salts that function as preservatives, yet long-term health concerns may ensue [18].

2.2.2 Health Concerns of Salt

Excessive salt consumption is well-known to result in hypertension, renal failure, and cardiovascular disorders. Hypertension results from elevated blood pressure due to the accumulation of salt. Reports indicate that salt adversely affects the kidneys, contributing to the formation of kidney stones, resulting in discomfort, damage, and perhaps culminating in renal failure. Reports have revealed that excessive salt consumption accelerates bone loss, a condition known as osteoporosis, since elevated blood salt levels lead to increased excretion of sodium and chloride, ultimately resulting in calcium depletion from bones. Consequently, health institutions are continually seeking methods to reduce sodium levels in food products. The primary measures include labeling sodium content in packaged goods, substituting salt with spices, herbs, and essential oils, while ensuring product quality and safety [19].

2.3 Halo-therapy

Since 1843, therapists have posited that laborers in salt mines enjoy robust health and seldom suffer from respiratory ailments such as colds and influenza. This was determined to be a consequence of inhaling air saturated with salt from the salt mines. Subsequently, a novel therapy for the treatment of respiratory and dermatological problems was identified termed 'halo-therapy.' It pertains to salt treatment, characterized by its dry nature, conducted in a designated setting equipped with halo-generators that release an aerosol of dry salts into the atmosphere. The premise of halotherapy involves the inhaling of salt particles measuring 2-5 microns and their absorption via the skin during scheduled sessions [20].

Halo-therapy is applied in salt rooms that can be either active or passive. In an active salt room, pure NaCl is dispensed via halo-generators or nebulizers in the room. Whereas, in passive salt room, rooms are charged with specific salts like, Himalayan pink salt, rock salt, Caribbean salt, etc. It appears as artificial salt caves with low humidity and temperature ranging from 22-24^{o}C. The concentration of sodium chloride in passive room is comparatively low in controlled environment [21]..

It has potential uses in the treatment of the following disorders: asthma, sinusitis, tonsillitis, bronchitis, allergies, and Chronic Obstructive Pulmonary Disease (COPD). Halo-therapy operates by promoting free radical oxidation, therefore diminishing oxidative damage, enhancing immune function, alleviating clinical symptoms of illness, detoxifying the body, clearing respiratory passages, and exhibiting anti-inflammatory, antifungal, and antibacterial effects, as well as accelerating skin repair and regeneration [22]. Halo-therapy has only minor side effects such as cough, conjunctivitis or irritation. Besides, the contraindications might affect the process of halo-therapy by blood disorders, tuberculosis, hypertension, cardiovascular disease, fever, etc. To conclude, halo-therapy is a dry salt therapy provided in controlled environment in active or passive salt rooms by salt particles inhalation to cure pulmonary disorders, skin disorders and various other benefits with minimum side effects [23].

2.4 Halophytes

2.4.1 Classification of Halophytes

Halophiles, the extremists that can tolerate highly saline environment, have been studied in many plants. Today, up to 2600 and maybe more halophyte species are found worldwide, with many possible applications in crops. Halophytes are salt-loving plants that are classified into 2 prime categories

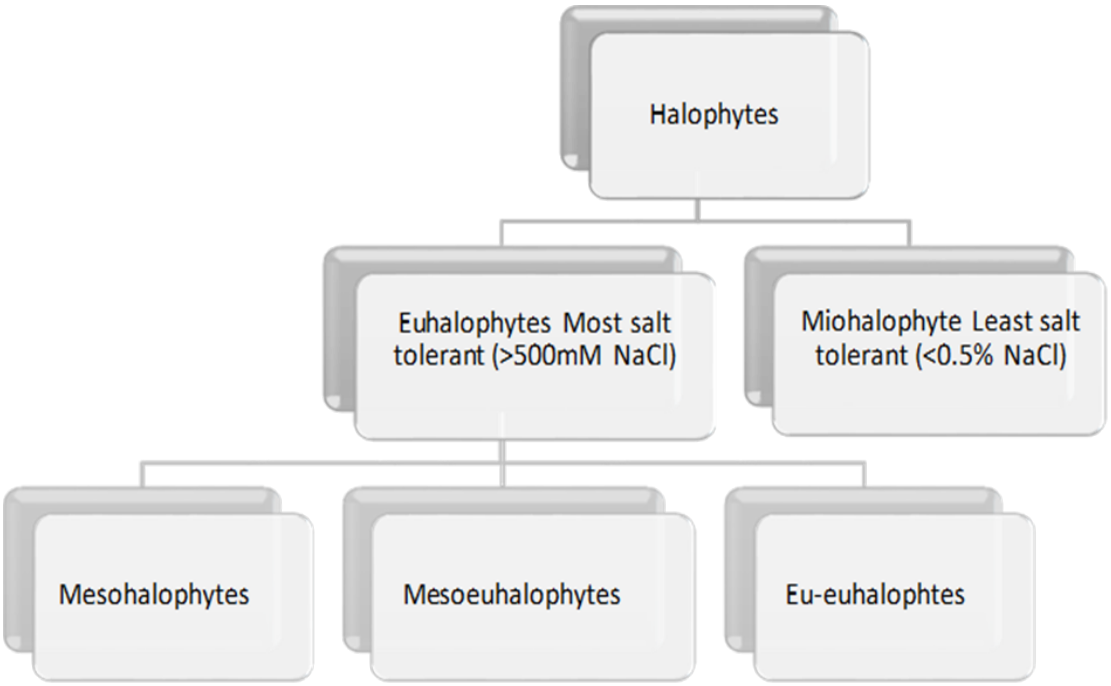


FIGURE 2.1: Classification Of Halophytes [24].

Miohalophytes are least tolerant to salt concentrations, less than 0.5% NaCl or Na_2CO_3. Whereas the euhalophytes can tolerant moderate to extreme saline environment and are subdivided into three categories; Mesohalophytes (5-10% salinity), Mesoeuhalophytes (more than 10% salinity), and Eu-euhalophytes (less than 1% salinity) [24].

2.4.2 Habitats of Halophytes

Typically, halophytes are found in tidal flat saline areas that could be coastal areas (near saline sea water) or mines and their soils. Most recently halophytes have been attributed to be found in wastewater plants and oil fields, which could be due to excess amount of salt left behind that leads to a salty environment. Globally such habitats are found, and they are filled with halophytes, thriving in such salty environments. However, it must be duly noted that coastal areas are now under threat of climate change and that increases the loss of water and accumulation of salt, both on nearby shores and in rocks. Moreover, domestic utilization of detergents comprises several acids and compounds that interact to form salts,

which ultimately migrate to terrestrial and aquatic environments, resulting in elevated salinity levels [25].

2.4.3 Halophyte Rhizosphere and Root Endosphere Microbiome

Rhizospheres are the soil that surrounds plants that are involved in root growth. These rhizospheres have many halophiles residing in halophytes, which adds a great insight into how plants that survive saline conditions actually make good use of salt tolerating microbes. The reason for microbial presence can be attributed to the much added purpose of the rhizosphere, which provides good source of nutrients (carbohydrates, amino acids, vitamins etc.). Such plants also have been known to form symbiotic relationships with bacteria, to help them get nitrogen fixed. Such bacteria are *Halobacillus, Bacillus, Bradyrhizobium, Salinibacter, Serratia, Mesorhizobium,* and *Pseudomonas* (Bharti *et al.*, 2024). These bacteria are known as the PGPR, Plant Growth Promoting Rhizo-bacteria. These bacteria stimulate plant growth by a number of mechanisms.

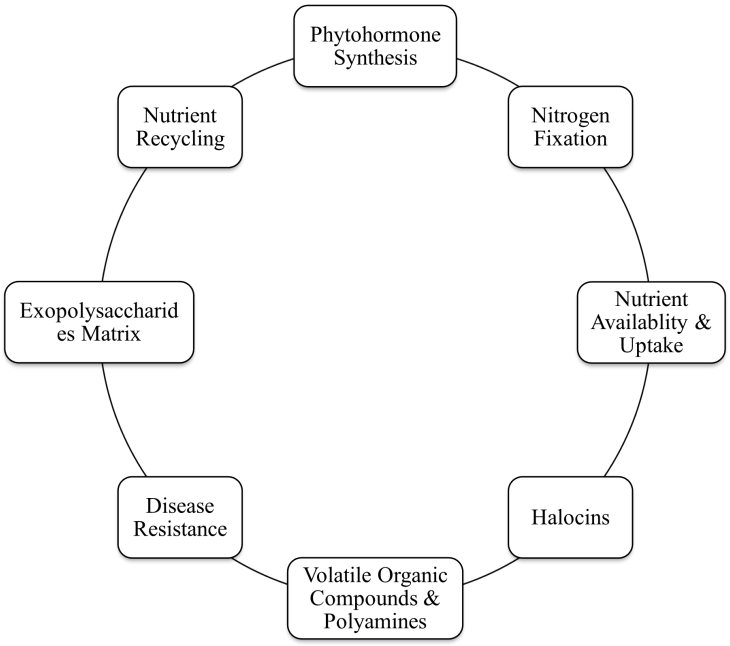


FIGURE 2.2: Plant Growth Stimulated By Bacteria [26].

2.5 Impact of Halophytes

There are number of halophytes found in salt mines. Halophytes have established a variety of tactics to grow well in high-salt surroundings and tolerate salinity. Salt promotes the vegetative growth of halophytes but reduces the growth of non-halophytes [27]. Saline environment is a habitat for halophytes including species, for example, mangroves which can also flourish in seawater [28]. Halophytes have enhanced growth and size when exposed to optimal salt concentrations compared to environments devoid of salt. Under saline circumstances, halophytes actively regulate ion absorption, secretion, storage, and exclusion [29]. To adapt to saltwater conditions, halophytes have evolved several specialized processes or structures. Conversely, non-halophytes lack morphological adaptations to endure salt stress; when compelled to inhabit saline soil, their biomass diminishes, and they cannot complete their life cycles. Halophytes endure severe salinity owing to the operation of specialized organs and leaf succulence [30].

TABLE 2.2: Impact of Halophilic Plants on Health and Environment [26]

Plants	**Uses**
Acacia modesta wall (tree)	Emollient and demulcent.
Acacia nilotica (Linn.) Delile (tree)	Diarrhea and tonic.
Phoenix sylvestris Roxb (tree)	Tonic and fodder.
Zizyphus maurtiana lam (tree)	Fodder, hedge plant and scabies, honey
Capparis spinosa L (shrub)	Pickles, diuretic and tonic.
Periploca aphylla dcne (shrub)	Swollen joints, cough and flu.
Teucrium stocksianum boiss (herb)	Expectorant, fever and sour throat

Halophytes serve as a crucial ecological and economic function. They stabilize soil in saline regions, mitigate erosion, and enhance soil structure and fertility via organic matter deposition. Their capacity to take salt from soil or water renders them viable candidates for the phytoremediation of salt-affected regions. Furthermore, some halophytes are being investigated for application in biosaline agriculture, where they can be grown on marginal saline soils for fodder, food, biofuel,

and pharmaceutical products. *Salicornia* and *Atriplex* are under investigation for oilseed production and animal feed, respectively.

Recent research has emphasized the synergistic interactions between halophytes and halophilic or salt-tolerant microorganisms, including plant growth-promoting rhizobacteria (PGPR), which augment nutrient absorption and stress resilience. These connections are crucial for sustainable agriculture in saline and arid locations, providing environmentally benign alternatives to chemical fertilizers and soil additives.

2.6 Halophiles

Salt mines exemplify a microbiome abundant in halophiles. Halophiles are classified as extremophiles, creatures that thrive in harsh environments inhospitable to other life forms, specifically as bacteria that tolerate high salinity. Halophiles have established their niche, with various capacities that enable them to endure elevated salt concentrations through many evolutionary strategies, including the extrusion of poisonous inorganic ions and the accumulation of essential solutes.

Bacteriorhodopsin serves as a crucial proton pump located within the cell membrane, aiding bacterial cells in resisting salt through proteins and amino acids. Halophiles from the three domains of Archaea, Bacteria, and Eukarya are renowned for inhabiting saline environments and substrates, characterized by varied concentrations of NaCl (2-30%) [31].

These halophilic microorganisms have lately gained attention for their ancient lineage and extensive, unexplored evolutionary development, which possesses several medicinal and commercial applications in biotechnology. By examining the survival of these microbiomes in ancient settings and the evolutionary processes that transpired, with the biochemical and genetic examination of their colonies, we might discover novel bioactive substances that may have contemporary applications [32].

The Asian subcontinent has a historical legacy of salt mining, which extends to Pakistan, where several salt rock formations are found throughout the nation. Khyber Pakhtunkhwa (KPK) has the Khewra Salt Mines located in Khewra, north of Pin Dadan Khan, and the Karak Salt Mines situated in the northwest. The Kohat Range salt mines have verified halophilic organisms within their salt deposits. These bacteria may be up to one million years old, preserved in these salt formations. To far, metagenomic investigations have been conducted on the Karak Salt Mine utilizing Illumina 16S rRNA sequencing and on the Khewra Salt Mine employing barcoded amplicon sequencing [33].

2.6.1 Impact of Halophiles

Halophiles are present in the realms of Eukarya, Archaea, and Bacteria. They are categorized according to their NaCl growth requirements. They are located in salt marshes, saline lakes, saline soils, and salt pans, which are widely spread worldwide and are characterized by hypersalinity. Halophiles are classified as minor, moderate, or intense based on their NaCl concentration tolerance [34]. Halophilic bacteria exhibit tolerance to elevated salt concentrations, little dietary needs, and the capacity to regulate osmotic pressure within their environment.

Their halo-adaptation strategies involve the accumulation of suitable solutes and intracellular storage of KCl to preserve sodium equilibrium in the cytoplasm and counteract the external osmotic pressure induced by elevated salinity. Halophilic bacteria are recognized for synthesizing retinal proteins, carotenoid pigments, hydrolytic enzymes, and suitable solutes that serve as bio-fertilizers, macromolecule stabilizers, and biopolymer. Amylases, lipases, gelatinases, xylanases, and proteases are examples of enzymes that function under salt stress conditions in halophile [35].

Halophilic enzymes function effectively throughout a broad spectrum of salt concentrations, PH levels, and temperatures, conditions under which other proteins denature.Exopolysaccharides and phytohormones are crucial macromolecules produced by halophiles, serving as valuable agents for the bioremediation of pollutants

in saltwater environments. Microbial salt crystals impact the preservation of life in the hyper-saline fluid contained within the crystals.

2.6.2 Interactions Between Flora and Microorganisms in Saline-Affecteds Soils

Exposure of microorganisms to a hyperosmotic environment induces fast efflux of cellular water, leading to decreased turgor and cytoplasmic dehydration. Various forms of adaptation have been developed to counteract cellular water efflux. The osmotic balance between the cell's cytoplasm and the surrounding medium is preserved by subjecting the cytoplasm to elevated ionic strength. The initial cellular response to osmotic upshift involves the outflow of water, the absorption of K+, and the accumulation of compatible solutes within the cell.

Salt stress (50-200 mM NaCl) in legume crops impedes productivity due to its detrimental effects.The proliferation of root nodule bacteria and the host plant affects the symbiotic progression and the capacity for nitrogen fixation. Salinity stress in cultural medium resulted in a reduction of nitrogen fixation by impairing nodule growth and the symbiotic relationship in Vicia faba. Nevertheless, after the full development of the root nodule in optimal conditions, nitrogen fixation continues even after the introduction of salt-stress conditions. The early productive type of Phaseolus vulgaris can withstand low salinity levels (48 mM NaCl) but not elevated levels (72 and 96 mM NaCl). The GRA19 strain of *Rhizobium leguminosarum biovar.* Viciae exhibited tolerance to low salt concentrations (50 mM) by comparing growth under stress conditions to growth in the absence of stress. The symbiotic nitrogen fixation (acetylene reduction activity) in the faba bean cultivar Alameda, infected with GRA19, was diminished under saline circumstances. Different species of Rhizobium exhibit variability in salt tolerance, with NaCl tolerance levels ranging from 100 to 650 mM.

Rhizobia have significant variability in their tolerance to salt. Several strains are inhibited by 100 mM of NaCl, although growth at salt concentrations over 300 mM has been shown for the strains of *Sinorhizobium meliloti* and *Rhizobium*

tropici. Certain strains of alfalfa, acacia, prosopis, and leucaena can withstand 500 $mmol^{-1}$fl NaCl.

2.7 Application Strategy for Halophilic Microorganisms

2.7.1 *Halophilic Bacteria*

These halophilic bacteria can regulate osmotic pressure in their environment. Furthermore, species that thrive in excessively saline environments and necessitate salt for optimal growth and development are termed halophiles. They exhibit considerable diversity, encompassing three realms of life: Bacteria, Eukarya, and Archaea. They reside in soda lakes, salt ponds, and rock salt crystals as dormant cells. Organisms can be classified into two categories: those that can endure salt and those that necessitate salt for growth and development. Halotolerance is a process that enables halophilic bacteria to sustain growth and development in saline environments.

Halophiles are categorized into minor halophiles, moderate halophiles, and extreme halophiles. Slight halophiles thrive optimally within a NaCl content of 0.2-0.5 M (1-3%). Moderate halophiles can thrive in a NaCl concentration of 0.5-2.5 M (3-15%), whereas extreme halophiles can flourish in a concentration of 2.5-5.2 M (15-30%) NaCl. Halophiles encompass aerobic, anaerobic, heterotrophic, phototrophic, and chemoautotrophic varieties located in diverse settings.

In agriculture, plants encounter numerous abiotic environmental challenges, including drought, cold, salt, nutrient deficiencies, infections, and heavy metals. This stressor results in irregularities in plant growth and development. In dry and semi-arid regions, insufficient rainfall, elevated temperatures, and substandard water quality contribute to soil salinity, recognized as a significant environmental stressor. Halophilic bacteria adapt to salinity through distinct mechanisms, aiding the plant in enduring salt stress conditions. Plants employ many biochemical and

physiological methods to survive in salt-stressed soil, including osmolyte synthesis, antioxidant enzyme activity, hormonal regulation, and ion exclusion mechanisms. In addition to these plant defense mechanisms, the bacterial community in the soil, including halophilic bacteria, significantly contributes to enhancing salt tolerance in the soil.

2.8 Classification of Halophilic Bacteria

Halophilic microorganisms are salt-adapted entities classified within the order *Halobacteriales* and the family *Halobacteriaceae*. The inaugural halophilic microorganism, *Halanaerobium praevalens*, was identified in Utah's Great Salt Lake and classified as a genus within the *Bacteroidaceae* family. Subsequently, novel halophilic bacterial species and genera were characterized by 16S rRNA sequencing and membrane lipid profiling. Various halophilic organisms are enumerated in Table 2.3.

TABLE 2.3: Halophilic Bacteria Species With The Salt-Tolerant Range [1].

Halophilic Bacterial Species	**Salinity Range for Growth and Development (%)**
Kangiella spongicola	2-15
Halanaerocella petrolearia	6-26
Salisediminibacterium cookie	3-30
Amphibacillus cookie	6-26
Desulfohalophilus alkaliarsenatis	12.5-33
Halanaerobacter jeridensis	6-30
Natribacillus halophilus	7-23
Fodinibius salinus	10-15
Alkalibacterium gilvum	0-17.5
Halomicroarcula pellucida	20-30
Salinibacter iranicus	12-30
Halanaerobium sehlinen	5-30
Saliterribacillus perciscus	0.5-22.5
Limimonas halopajila	15-30
Aquibacillus halophilus	0.5-20
Halobellus salinus	15-30

Table 2.3 continued from previous page

Halophilic Bacterial Species	**Salinity Range for Growth and Development (%)**
Bacillus daingensis	0-16
Oceanicola flagellatus	0-21
Spiribacter salinus	10-25
Halomonas huangheensis	1-20
Salifodinibacter halophilus	25
Halomonas sambharensis	5-8
Lentibacillus saliphilus (YIM 93176T)	0-22
Halomonas urmiana	0.5-20
Marinobacter halodurans	1-18
Aliifodinibius saliphilus	3-25
Arhodomonas recens	2-25

2.9 Adaptation Strategies of Halophilic Bacteria in Saline Environments

Water is the essential element responsible for life. Living microbes have the capacity to adapt and endure in hostile settings. Microorganisms that are not acclimated to saline environments will experience water loss, leading to cellular shrinkage and ultimately death due to compromised cellular structure and function. To mitigate excessive water loss in such environments and maintain cellular integrity and functionality, halophilic bacteria have developed two types of strategies to cope with elevated salt concentrations. The initial strategy is the salt-in strategy, whereas the subsequent one is the compatible solute strategy. Bacterial cells maintain osmotic equilibrium between internal and external environments by accumulating a high concentration of KCl. This approach is executed by the cell through alterations in numerous physiological metabolisms, including enzyme activity, generation of cellular components, and the morphology and function of certain organelles. The high-salt-in technique safeguards halophiles against saline environments by intracellularly accumulating inorganic ions to maintain equilibrium of salt concentrations in their surroundings. Bacterial cells maintain osmotic

equilibrium between their internal and external environments by accumulating a high concentration of KCl. Halophiles utilize Cl- pumps to transport Cl- from the environment into the cytoplasm during this process. Arginines and lysines are positioned at both termini of the channel to augment the absorption and discharge of Cl-.

Halophilic bacteria safeguard their cells against elevated salt concentrations by accumulating suitable solutes, including organic compounds (proline, betaine, ectoine, trehalose) and inorganic ions (K^+, Mg^{2+}, Na^+). Osmolytes, or suitable solutes, are either synthesized by the bacterial cell in the cytoplasm or absorbed from the surrounding media. Most bacteria lack the internal mechanisms for the active transfer of water to counteract external osmotic pressure. Consequently, the internal environment is regulated through the transport and synthesis of a set of suitable solutes, ensuring no disruption to the cell's metabolic functions. Compatible solutes are categorized based on their chemical composition into anionic solutes, zwitterionic solutes, and non-charged solutes. Organic anions are employed to regulate the internal environment of halophilic bacteria in highly saline circumstances. Halophilic bacteria, including *Halomonas* and *Halobacterium*, produce ectoine and L-glutamate, respectively, to endure salinity-stressed environments. Certain halotolerant bacteria, including *Bacillus*, *Pseudomonas*, *Aeromonas*, and *Zymomonas* utilize polyol molecules such as sorbitol, arabitol, glycerol, and mannitol for osmoadaptation in salt-stressed environments. Halophilic bacteria utilize zwitterionic solutes generated from neutral amino acids as osmolytes under salt-stressed circumstances. Betaine is a naturally occurring anionic molecule utilized as an osmolyte to safeguard cells from osmotic stress by regulating intracellular water balance. Various halophilic bacteria, including *Halomonas*, *Virgibacillus*, *Oceanobacillus*, and *Polaribacter*, can synthesis betaine from glycine by methylating the primary amine to produce a quaternary amine. Certain methanogens, including *Methanohalophilus* and *Methanohalobium*, can accumulate and manufacture betaine by the methylation of glycine or the oxidation of choline. Ectoine (cyclic tetrahydropyrimidine), acquired from the external environment or produced from the medium, serves as an osmolyte for halophilic bacteria to safeguard against salt-induced stress.This was identified from the *Halorhodospora halochloris*

bacteria, which was extracted from the hypersaline Mono Lake. Ectoine osmolytes have been identified in halotolerant and halophilic bacteria, including *Halomonas*, *Oceanobacillus*, *Nesterenenkonia*, *Methylophaga*, and *Methyllarcula*. Certain polar and non-charged organic compounds have been utilized as osmolytes to safeguard the cell during instances of elevated salt stress. Glycerol osmolyte has been identified in some bacteria and halotolerant yeast under situations of salt stress. Certain sugar compounds, including trehalose, have been identified in halotolerant and halophilic organisms, serving as suitable solutes to withstand desiccation, thermal extremes, cold, and hypersaline conditions. Certain proteobacteria and marine cyanobacteria are recognized for accumulating sucrose as an osmolyte under salt-stressed environments. Certain proteins, including proline, acetylated glutamine dipeptide, and carboxamine, function as osmolytes, safeguarding the cell from elevated salt concentrations. They predominantly occur in halophilic purple sulfur bacteria and marine phototrophic bacteria.

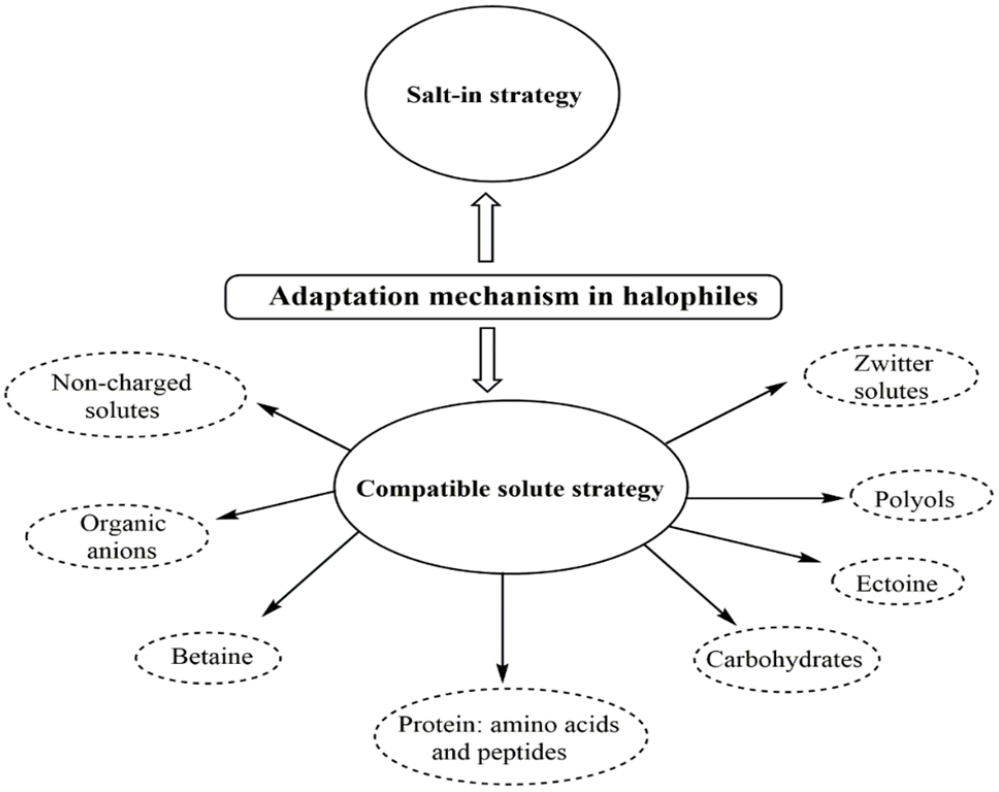


FIGURE 2.3: Various Modes Of Adaption To Saline Conditions By Halophilic Bacteria [1].

2.9.1 Halophilic Bacteria: The Contribution of Halophilic Bacteria to Plant Growth Enhancement Under Saline Stress

Throughout growth and development, every living organism or plant endures the challenging circumstances of the soil. To alleviate the stressful situation, people will either confront it or formulate an alternative strategy. Due to their delicate and sessile nature, plants cannot evade adverse environments; hence, they respond defensively. Halophilic bacteria enhance their tolerance, growth, and productivity through several pathways, enabling them to mitigate the adverse effects of abiotic stress conditions with distinct functional characteristics.

2.9.1.1 The Function of Bacterial Phytohormones

Bacterial phytohormones are chemical substances present in low concentrations that influence physiological and biological processes in plants. These minute concentrations of bacterial phytohormones regulate various processes associated with plant differentiation and development. Bacterial hormones, released near plant roots, can elicit a physiological response in the host plant.

Plant growth-promoting bacteria (PGPB) produce phytohormones, including indole-3-acetic acid (IAA), cytokinins, abscisic acid, gibberellins, and other growth regulators that facilitate plant growth and development. All these phytohormones extend root stimulation by significantly augmenting root length and surface area, resulting in enhanced nutrient absorption and consequently improving plant health in salt-stressed conditions.

2.9.1.2 Aminocyclopropane-1-Carboxylate Deaminase

In minimal concentrations, ethylene is a crucial and volatile bacterial phytohormone that influences the regulation of plant growth. Ethylene phytohormones affect the growth of vegetative plant structures, the rooting of cuttings, nodulation,

and the signaling processes related to salt stress in the root zone. Excessive production of ethylene hormones due to abiotic stress might restrict plant growth and development. Chemical inhibitors, like cobalt ions and aminoethoxyvinylglycine, are frequently employed to address these challenges. Nonetheless, these substances are excessively expensive and detrimental to the environment. Salt-tolerant bacteria produce aminocyclopropane-1-carboxylate (ACC) deaminase, which catalyzes the conversion of ACC to -ketobutyrate and ammonia, hence reducing ethylene levels in salt-stressed plants.

2.9.1.3 Phosphate Solubilization

Phosphorus (P) is a crucial macronutrient essential for the synthesis of several biochemicals, including nucleotides, phospholipids, nucleic acids, and phosphoproteins, as well as for plant growth and development. Under conditions of salt stress, phosphorus availability diminishes, leading to symptoms of phosphorus deficiency. Soil has two forms of phosphorus: organic and inorganic.

The mobility and availability of phosphorus to plants are significantly lower than that of other nutrients, including zinc, iron, copper, and potassium. The predominant type of phosphorus in the soil is insoluble, hindering its mobility and availability to plants.

Halophilic strains facilitate the transformation of insoluble phosphorus into soluble phosphorus and contribute to the preservation of soil phosphorus levels. Extensive study has been conducted on halotolerant strains capable of solubilizing and rendering phosphorus accessible. The black paper displayed phosphorus mobilization and absorption, resulting in enhanced root proliferation and plant growth.

Rhizobacterial strains can flourish in elevated salinity settings (60 g/L NaCl) and are proficient phosphorus solubilizers in soil. Under saline stress conditions, the Pseudomonas strains significantly affected the growth and development of Zea mays L(maize). In saline conditions, the inoculation of PSB *Herbaspirillum seropedicae* and *Burkholderia sp* enhanced crop weight by 1.5-21 percent.

2.9.1.4 Antioxidant Activity

Salt stress triggers the formation of reactive oxygen species, which degrade different biomolecules, including proteins and lipids, ultimately leading to plant mortality. Plants possess antioxidant mechanisms that enable their survival in the presence of reactive oxygen species (ROS). A multitude of antioxidative enzymes (such as superoxide dismutase, peroxidase, and catalase) and non-enzymatic antioxidants (including ascorbic acid and glutathione) facilitate the processes of reactive oxygen species scavenging. Numerous halotolerant plant growth-promoting rhizobacteria (PGPR), including *S. proteamaculans* and *Rhizobium leguminosarum*, are recognized for their production of enzymes (SOD, POX, CAT) that facilitate plant survival under saline stress conditions. Recent findings indicate that salt-tolerant bacteria (*P. simiae AU*) augment peroxidase and CAT gene expression in soybean plants after exposure to 100 mM NaCl stress inoculation. PGPR inoculation alleviates the detrimental effects of oxidative stress through both enzymatic and non-enzymatic pathways in saline-stressed environments. In instances of non-enzymatic mechanisms, they diminished their exposure to reactive oxygen species by relocating to areas with decreased sun radiation. The synthesis of pigments and the encapsulation of DNA with proteins and chromatin create alternative targets for the assault of reactive oxygen species. Certain non-enzymatic antioxidant chemicals inhibit reactive oxygen species. The enzymatic approach yields many enzymes, including superoxide dismutase, catalases, glutathione peroxidases, and peroxiredoxins, without producing additional reactive species. Antioxidant enzymes convert deleterious products into less hazardous molecules or identify and destroy them. These approaches also preserve the optimal physiological concentrations of reactive species, such as reactive oxygen species (ROS).

2.10 Salt Production

Salt manufacturing has swiftly transitioned from old evaporation methods to contemporary technology-driven processes. Traditionally, the method of solar salt

evaporation involved filling enormous pools with saltwater either from the sea or underground. Initially, they were filtered through cloth or strainers to eliminate contaminants. The saltwater was let crystallize under sunlight while evaporation removed the water. This required up to one week, depending upon the season and weather conditions. The residual salt was pulverized into a finer powder and marketed. The brine was handled similarly for the salt in mines. The salt rocks were subjected to significant pressure, crushed, filtered, and subsequently sold. These substances were referred to as evaporated salts. Today, the immense mechanical power of machines has halved the difficult and time-consuming job of salt production. The salt rocks are crushed and ground into a paste, which is then treated with preservatives and other chemicals. Desalination of seawater generates brine, posing an environmental threat. The evaporation-crystallization technique can be employed to treat brine, facilitating the separation of salts and enhancing production output. The residual brine must be neutralized prior to its transformation into valuable goods. The brine is either transformed into inert chemicals or immobilized [36].

2.11 Salt Mines of Pakistan

2.11.1 Nature and Origin of Salt Deposits of Pakistan

Pakistan possesses many types of salt sources, including rock salt, sea salt, and salt lakes located in Khewra, Kalabagh, Karak, Kohat, Bahadur Khel, and Jatta.The salt crystals located in salt mines are classified according to their morphology: white, pink, soft lumps, and soft crystalline aggregates.The salt located in the Kohat salt range is characterized by a white-grey hue, fine to medium crystalline structure, and a loose consistency.The Khewra salt variety consists of pink-red, coarse crystalline, compact salt.In Khewra, salt predominantly exists in three color zones: white, pink, and red to beef-colored, located among formations resembling domes.It possesses a crystalline structure [1].

2.11.2 Elements Found in Salt Mines

The salt samples collected from different salt mines in Pakistan.Trace elements and heavy metals were identified in the samples.Among the identified elements was Iron (Fe), which is linked to metabolism, as well as oxygen storage and transfer.Zinc (Zn) and Iron (Fe) serve as cofactors for certain enzymes, whereas Copper (Cu) modulates enzymatic activity inside the body [37]. Manganese (Mn) is essential for skeletal muscle and bone development, reproduction, growth and cognitive function.Zinc (Zn) contributes to metabolic processes, tissue functionality, growth, and neurological activity.Although all current elements are essential for maintaining normal bodily functions, heavy metals such as Lead (Pb) and Chromium (Cr) are deemed detrimental to health.They induce significant health consequences upon exposure or ingestion. Their Recommended Daily Allowance (RDA) is similarly inferior to the minerals that govern regular bodily activities.All salt samples were determined to be pure and devoid of trace components, with the exception of Kohat and Khallar salt.They were determined to be abundant in Sulphate, Calcium, and Magnesium.These contaminated salts require purification before safe utilization [38].

TABLE 2.4: Trace Elements Of Salt Samples (mg/kg) [21]

Serial No.	**Fe**	**Zn**	**Cu**	**Mn**	**Cr**	**Pb**	**Cd**
I	0.62	0.18	0.01	0.00	0.34	0.03	0.00
II	0.24	0.12	0.00	0.00	0.36	0.10	0.00
III	0.46	0.13	0.01	0.00	0.37	0.02	0.00
IV	0.97	0.19	0.01	0.06	0.19	0.02	0.00
V	0.48	0.20	0.03	0.00	0.37	0.03	0.00
VI	0.29	0.07	0.01	0.00	0.33	0.03	0.00
VII	1.16	0.11	0.02	0.00	0.40	0.03	0.00

The table 2.4 is based on samples taken from various salt mines in Pakistan. **I** - Rock Salt (Kohat), **II** - Pink Salt (Khewra), **III** - Red Salt (Khewra), **IV** - White Hard Salt (Khewra), **V** - White Crystal Salt (Khewra), **VI** - Khallar Salt (Kallar Kahar), and **VII** - Lake Salt (Tharparker).

2.11.3 Chemical Composition

Nearly seven salt samples were procured from Kohat, Khallar, Khewra, and Salt Lake to examine their chemical composition and contaminants by both instrumental and traditional methodologies.The gathered samples were analyzed for contaminants including trace metals, water-insoluble compounds and moisture utilizing an atomic absorption spectrophotometer [39].

TABLE 2.5: Analytical Parameters Of Different Salt Samples (%) [21]

Sr. No.	NaCl	Moisture	Insoluble Matter	Ca	Mg	SO_4	K
I	92.02	0.12	3.50	1.75	0.16	2.15	0.002
II	98.30	0.11	0.40	0.30	0.12	0.28	0.046
III	98.10	0.09	0.50	0.25	0.12	0.58	0.045
IV	97.55	0.12	0.80	0.20	0.12	0.25	0.800
V	98.15	0.13	0.90	0.20	0.09	0.33	0.204
VI	84.15	0.22	5.00	0.90	1.26	8.35	0.024
VII	98.25	0.31	0.30	0.30	0.12	0.50	0.006

The table 2.5 is based on samples taken from various salt mines in Pakistan. **I** - Rock Salt (Kohat), **II** - Pink Salt (Khewra), **III** - Red Salt (Khewra), **IV** - White Hard Salt (Khewra), **V** - White Crystal Salt (Khewra), **VI** - Khallar Salt (Kallar Kahar), and **VII** - Lake Salt (Tharparker).

2.12 Khewra Salt Mine in Pakistan

The Khewra salt mine is situated in the Punjab province, near Jhelum, roughly 160 kilometers from Islamabad.It is the world's second-largest salt mine, sometimes referred to as the 'Museum of Geology,' originating from the Precambrian era.One of the prevalent ways for salt extraction employed there is the chamber and pillar technique. Approximately half of the salt in the mine is mined, while the remainder is retained to sustain the structural pillars.It is situated 288 meters above sea level and comprises nineteen floors, of which eleven are subterranean.The salt obtained from Khewra exhibits a spectrum of colors, including translucent, white, milky

white, dull white, pink, cattle red, and brown.The mine has naturally occurring bands of two colored salts: white and red.The Khewra salt mine comprises around 64 chambers, each containing a saturated brine solution.The water in the roofs and walls of my mine infiltrates to fill the rooms. They are less influenced by variations in humidity and temperature [40].

2.13 Whole Genome Sequencing

Whole metagenome sequencing entails the extraction of DNA from a sample, subsequent fragmentation, and high-throughput sequencing utilizing Next Generation Sequencing (NGS) methodologies.The data acquired from sequenced entire genomes can subsequently be employed for Genome Resolved Metagenomics (GRM) research and to derive Metagenome Assembled Genomes (MAGs) of microbes that cannot be cultivated in laboratory settings. Metagenomic methodologies have facilitated the identification of new microbial species that may contribute to nutrient recycling and ecosystem management. Halophilic bacteria are increasingly being identified more effectively, owing to gene prediction using Hidden Markov Model (HMM) profiles that indicate gene function [41] .

Whole genome sequencing not only helps to provide taxonomic data on halophiles, but also their abundance and genetic diversity.Shotgun metagenomics investigation of saline environments allows for the discovery of dominant genes that code for diverse secondary metabolites and heavy metal breakdown, such as arsenic by *Natronomonas* and *Halorubrum* [42]. Whole metagenome sequencing is critical for investigating the microbial diversity of halophiles in very salty settings like salt mines.However, the high expense of metagenomics sequencing limits its use as the only tool for studying halophilic diversification [43].

Chapter 3

Methodology

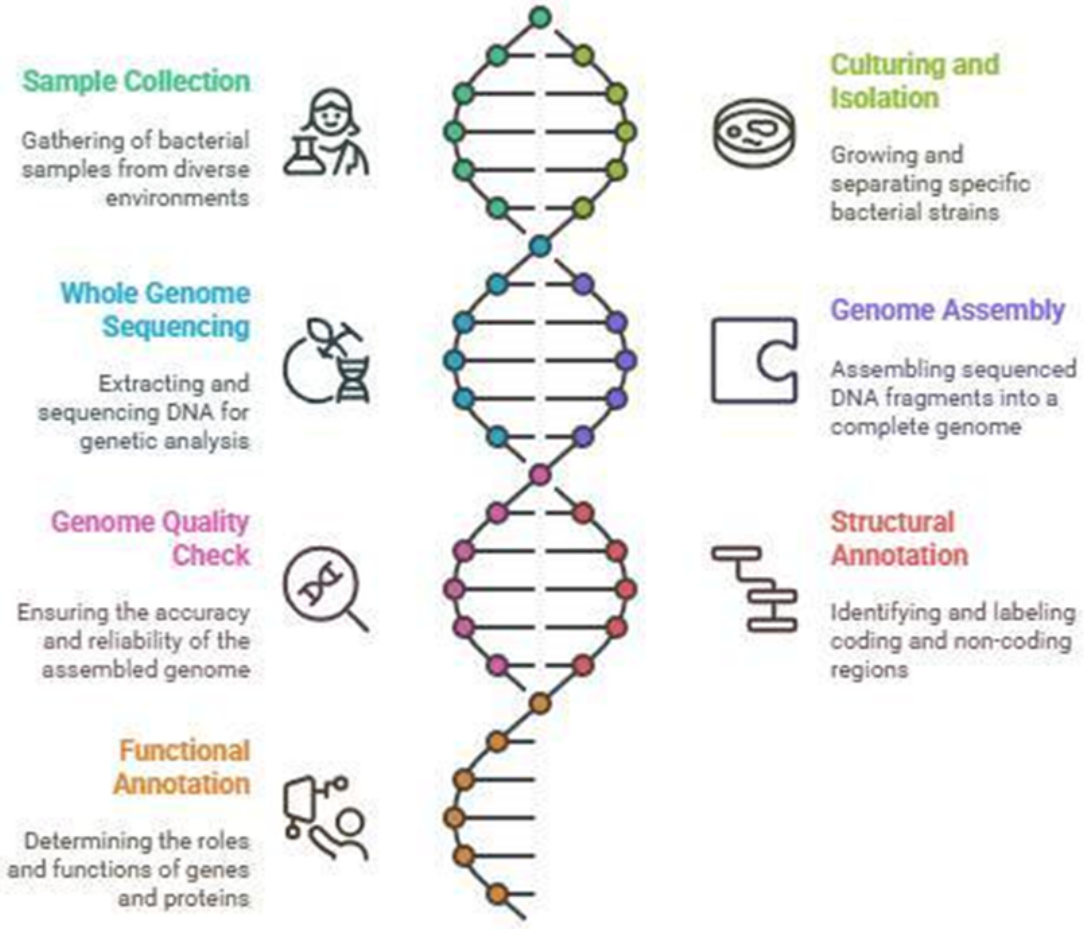


FIGURE 3.1: Flowchart of Methology

3.1 Sample Collection

3.1.1 Sample Location

The location chosen for sample collection in salt mine was an abandoned area. This site was deserted due to minimal appeal to the mine workers. A specific site was located to the right of Chandni Chowk, where various saline ponds were discovered. This saline pond had its water extracted, revealing visible salt crystals at the bottom. This pond was shallower and lacked adequate illumination, resulting in people typically avoiding this dim area.

3.1.2 Salt Sample Collection Procedure

A sterile blade was used to break the salt from the walls of the mine and the resulting salt crystals were collected in sterile polythene bags. Three samples of salt were collected from three different locations. The samples were stored in storage box at 4^oC. The sample color was salmon salt with red hues.

3.1.3 Culturing of Bacteria from Salt Samples

To isolate halophilic bacteria, a medium was created by dissolving 4.2g of nutrient agar in 150ml of distilled water. This medium included a 0.5% NaCl content, corresponding to 5g of NaCl per liter of nutritional agar. Salt samples were first crushed with sterile pestle and mortar and then diluted in 5ml water in test tubes. The four liquid samples were diluted in four test tubes of 5ml water. Four liquid solutions were poured into 33 plates; The plates were incubated at 37^oC for four days.

3.1.4 Isolation of Halophilic Bacteria

To isolate halophilic bacteria, prominent colonies cultivated on 0.5% nutritional agar were subsequently streaked onto fresh agar plates containing 2% NaCl for screening purposes. The revised agar formulation consisted of 5.6g nutrient agar and 3g NaCl in 200ml of distilled water was prepared. Each colony was streaked onto ten plates to get pure isolates. The purpose of employing 2% NaCl was to exclusively isolate halophilic bacteria from the colonies. The plates were then placed in an incubator at 37^{o}C for 24 hours and assessed the next day.

3.2 DNA extraction for Whole Genome Sequencing

High-quality bacterial DNA was isolated for whole-genome sequencing utilizing a standardized CTAB technique. 1.5 ml of nutrient broth was aliquoted into labeled Eppendorf tubes and centrifuged at 7000 rpm for 5 minutes to isolate the bacterial colonies as a pellet. Following centrifugation, the supernatant was meticulously removed, and the pellet was resuspended in 500μl of lysis buffer. The lysis buffer was freshly formulated and pre-heated to 60-65^{o}C using the following proportions: 500μl of 100mM Tris-HCl, 200μl of 20mM EDTA, 1160μl of 1.4M NaCl, and 0.1g of 2% CTAB were combined with 3000μl of sterile distilled water to produce 5ml of lysis buffer. After the addition of lysis buffer, the pellet was homogenized either by vortexing or by pipetting. Subsequently, 20μl of Proteinase K, 40μl of 10% SDS, and 6-8μl of β-Mercaptoethanol were included. Samples were incubated at 95^{o}C for a duration of nearly 2 hours or longer to enhance results. Subsequent to incubation, 500μl of Chloroform: Isoamyl alcohol in a 24:1 ratio was introduced, and the samples were centrifuged at 13,000 rpm for 10 minutes. The centrifugation with chloroform and isoamyl alcohol resulted in the creation of three distinct layers in the Eppendorf tube: an organic layer at the bottom, debris in the middle, and an aqueous layer at the top. In one of the tubes, the layers were non-uniform; thus, it was inverted and subsequently centrifuged again at 13,000 rpm for 10 minutes.

The aqueous layer from each sample was meticulously collected and transferred to newly labeled Eppendorf tubes. Subsequently, 500μl of cold isopropanol was introduced, and the samples were incubated at ambient temperature for 20 minutes, followed by centrifugation at 13,000 rpm for 15 minutes. Following centrifugation, the supernatant was removed, and the pellet was washed twice, initially with 100% ethanol and subsequently with 70% ethanol. 500 μl of ethanol was added, followed by centrifugation at 8000 rpm for 5 to 10 minutes, after which the supernatant was discarded. The washing procedures were identical for both concentrations of ethanol.

The pellet was subsequently permitted to air dry on a paper towel for approximately 40 minutes or until fully dry. The air-dried pellet was subsequently suspended in 40 μl of low TE buffer and incubated at 60^{o}C for a further 30 minutes, after which it was prepared for gel loading. Samples of extracted DNA were cryopreserved with an equal volume of low TE buffer until needed for gel electrophoresis.

3.2.1 DNA Visualization and Amplification

The amplifications of DNA was performed using PCR in a thermocycler. The primer used was 9F (5 -GAGTTTGATCCTGGCTCAG-3) 19mer. The DNA was centrifuged for five seconds. Subsequently, 5μl of DNA was combined with 90μl of Master Mix, 15μl of Forward Primer, 15μl of Reverse Primer, and 30μl of PCR. The cycle duration was 2 hours and 18 minutes. Subsequently, the PCR DNA was subjected to electrophoresis for visualization.

3.2.2 Gel Electrophoresis

Gel electrophoresis was conducted to visualize the amplified product of the gene. For the 1% agarose gel in 1X TBE buffer containing ethidium bromide (5μl per 100ml).

3.2.3 Sequencing

The commercial service of Macrogen Korea was utilized for whole genome sequencing.

3.3 Whole Genome Annotation

The structural and functional annotation of the Bacterial genome was done by using different computational tools

3.3.1 Tools and Databases

3.3.1.1 Genious Prime (version 2024.0.2)

Geneious Prime [44] is a bioinformatics platform that offers researchers extensive tools for the analysis of Sequenced data in molecular biology research. The technology converts raw sequencing data into comprehensible visuals, allowing researchers to assess results effectively via its easy interface.

Genious Prime enhances the insights and visibility of the sequencing process, mitigates risks and errors, and boosts efficiency in sequencing.

Geneious Prime, incorporates an extensive array of tools inside a user-friendly interface, facilitating the effortless conversion of raw sequencing data into significant biological insights. The tools include miRA, Glimmer, EMBOSS, LASTZ, and Maft.

3.3.1.2 MIRA

The utility facilitates De Novo assembly by converting raw data files in text format into a usable and assembled format. MIRA is utilized for various genomic applications, including somatic mutation detection across cell types, cross-species

transcript similarity analysis, and high-fidelity sequence assembly for clinical microarray oligonucleotide design. The software is tailored for datasets ranging from 100 to 150 megabases or 20 to 40 million reads.

3.3.1.3 Cutadapt

It examines all reads for adaptor sequences and eliminates them upon detection. In the absence of a stated filtering option, all reads from the input file are incorporated into the output; some may be truncated, while others stay intact. Even reads reduced to zero length are preserved in the output.

3.3.1.4 BBDuk

A tool using a kmer-based decontamination utility from BBTools, which offers efficient processing of sequencing reads via concurrent quality control processes. The tool's kmer-matching algorithm provides rapid and accurate processing of high-throughput sequencing data, executing these critical preprocessing steps in a singular streamlined operation. The implementation as a Geneious Prime plugin helps researchers easier access to these quality control features without necessitating command-line proficiency.

3.3.1.5 Glimmer

It is utilized for the prediction of genes inside the entire genome. Glimmer (Gene Locator and Interpolated Markov ModelER) is a tool developed for gene prediction in microbial genomes, encompassing bacteria, archaea, and viruses. The approach utilizes interpolated Markov models (IMMs) to statistically differentiate protein-coding sequences from noncoding areas with high precision.

3.3.1.6 EMBOSS

It is utilized for retrieving information regarding proteins in the genome and for analyzing the proteins. Identify and annotate transcription factor binding sites and protein-coding regions. This plugin incorporates two EMBOSS nucleotide analysis tools: tfscan for predicting transcription factor binding sites and tcode for showing protein-coding potential via graphical output.

3.3.1.7 MAFFT

MAFFT (Multiple Alignment using Fast Fourier Transform) is a high-performance technique for multiple sequence alignment that has been effectively included into Geneious Prime. This robust interface allows researchers to conduct swift and precise multiple sequence alignments directly within the Geneious Prime platform, obviating the necessity for separate tools. The MAFFT implementation in Geneious Prime preserves all advanced characteristics of the algorithm.

3.3.1.8 BLAST

The Basic Local Alignment Search Tool (BLAST) is a bioinformatics application that detects regions of local similarity among nucleotide or protein sequences. BLAST identifies statistically significant matches by comparing query sequences with specialized databases, utilizing E-values and bit scores to infer biological links. Researchers utilize BLAST for several applications, such as describing unknown sequences, predicting gene or protein functions via homology, assessing evolutionary conservation, and identifying members of gene families.

Its enhanced algorithm equilibrates speed and sensitivity, establishing it as an essential instrument for genomes, proteomics, and metagenomics research. BLAST accommodates several search types (e.g., BLASTn for DNA, BLASTp for proteins) with adjustable parameters to enhance alignment accuracy for particular research requirements.

3.3.1.9 PROKSEE

Proksee assembles, annotates, analyzes, and visualizes bacterial genomes online. Researchers can create dynamic genomic maps and do advanced studies without command-line experience using its user-friendly interface and many bioinformatics tools [45].

3.4 Whole Genome Analysis

3.4.1 Data Acquisition and Preprocessing

FASTQ format raw sequencing data acquired from Illumina was loaded using Geneious Prime 2024.0.1. Quality control was executed with the integrated Trim & Filter tool with the parameters: a quality score threshold of Q20, a minimum read length of 50 bp, and adapter removal via the BBDuk algorithm.

3.4.2 De Novo Genome Assembly

The entire genome was uploaded into the Genious Prime tool for the purpose of constructing the de novo assembly. Raw sequencing reads in FASTQ format from several platforms underwent quality control prior to assembly, utilizing Cutadapt for trimming and FastQC for quality filtering. Genius Prime utilizes AI-driven *k*-mer optimization specifically designed for bacterial genomes (typically 21–127 bp) and conducts mistake correction prior to contig assembly. Trimming and filtering for adapter removal were conducted using BBDuk with default parameters. Reads with a Q-score below 20 were eliminated during quality reduction.

3.4.3 Structural Annotation

The structural annotation of bacterial genomes using Genius Prime is a multi-faceted computational strategy that integrates evidence-based and ab initio prediction techniques to precisely detect and delineate genomic features. The workflow commences with a high-quality genome assembly as input, which is subjected to meticulous preparation, including contig screening to eliminate sequences less than 500 bp that may signify assembly artifacts.

Genius Prime employs a hybrid methodology for gene prediction, initially utilizing Prodigal's dynamic programming technique to detect open reading frames (ORFs), followed by the enhancement of these predictions with deep learning models trained on bacterial genomic characteristics. These AI algorithms examine sequence attributes such as ribosome binding site motifs, codon use bias, and variations in GC content to enhance prediction accuracy, especially for short genes and those with unusual compositions.

The system concurrently detects non-coding attributes via specific instruments: tRNAscan-SE, utilizing bacterial-specific covariance models, identifies transfer RNAs, whereas Barrnap precisely locates ribosomal RNA operons. For functional characterisation, predicted proteins are aligned with curated databases (UniProt, NCBI NR) with accelerated BLAST algorithms, while HMMER scans detect protein domains from Pfam.

Genius Prime's distinctive machine learning module improves annotation by identifying potential horizontal gene transfer events through unusual sequence composition analysis and recognizing phase-variable locations via tandem repeat detection.The final outputs consist of GFF3-formatted annotation files with standardized feature descriptions, FASTA files with predicted protein sequences, and detailed reports on annotation quality measures.

3.4.4 Functional Annotation

The procedure commenced with a quality-assured genome assembly and structural annotation file, which is subjected to stringent homology-based analysis via BLAST searches against curated databases (UniProt, NCBI NR) and domain identification utilizing Pfam and InterProScan.

The system's machine learning capabilities markedly improve annotation quality by examining genomic context, such as operon structures and syntenic interactions, while deep learning models (ProtBERT) facilitate the prediction of functions for hypothetical proteins using sequence pattern recognition. Specialized modules identify virulence factors (VFDB), antibiotic resistance genes (CARD), and metabolic pathways (KEGG), along with the detection of non-coding components like as tRNAs, rRNAs, and CRISPR arrays using specific tools (CRISPER Cas9, tRNAscan-SE, Barrnap).

3.4.4.1 Bakta

Bakta is a command-line and web-based program for quick, standardized, taxon-independent annotation of bacterial genomes, plasmids, and metagenome-assembled genomes. It prioritizes accuracy, speed, and functional annotations, including tiny proteins (sORFs) ignored by other methods [46].

3.4.4.2 PHASTEST

Web server PHASTEST (PHAge Search Tool with Enhanced Sequence Translation) quickly identifies, annotates, and visualizes prophage sequences in bacterial genomes and plasmids. The revised version of PHAST and PHASTER offers faster processing, higher accuracy, and better visualization [47, 48].

3.4.4.3 Eggnogmapper

eggNOG-mapper uses precomputed orthologous groups (OGs) from the eggNOG database to quickly functionally annotate novel sequences (genomes, transcriptomes, metagenomes). BLAST and InterProScan are outperformed by orthology assignments, which prevent paralogous transfers and enhance annotation precision [49].

3.4.4.4 fIDBAC Fast Bacterial Genome Identification PlatformA

A bioinformatics tool created for swift and precise identification and typing of bacterial genomes, developed by researchers from Zhejiang University and associated institutes. It mitigates the constraints of conventional techniques (e.g., 16S rRNA gene analysis) by utilizing whole-genome data and meticulously curated databases [50].

Chapter 4

Results

4.1 Culturing and Isolation of Bacteria

Visible colony growth was observed on nutritional media after four days. The petri plates exhibiting significant growth were chosen for further isolation, resulting in the acquisition of isolated colonies. (Figure 4.1)

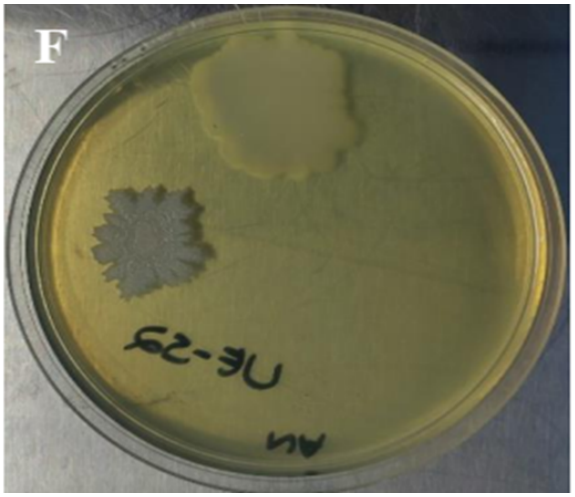


FIGURE 4.1: Colony of Isolated Bacteria from the Salt Mines

4.1.1 Visualization and Amplification of DNA

After the extraction of DNA, the DNA was run on 1% agarose gel electrophoresis for visualization. There were clear bands of DNA of all samples, highlighting that

DNA extraction was successful and amplification of DNA gel electrophoresis also revealed successful results. (Appendix I)

4.2 Whole Genome Sequencing

4.2.1 Raw Data Acquisition & Initial Processing

The Whole Genome Sequencing of halophilic bacteria from the Khewra salt mines was conducted utilizing the Sequencing by Synthesis method (Illumina) of Next Generation Sequencing. Initially, a FastQC report was produced by Genewiz Company, which confirmed the sample's accuracy. The FastQC report furnished critical information regarding the sequence, encompassing basic statistics, quality scores per sequence, quality per base, quality per tile, base composition, N content per base, adapter content, GC content per sequence, duplication levels, length distribution, and overrepresented sequences.

4.2.1.1 Raw Reads Bacterial Genome

The Whole Genome Sequencing (WGS) output initially consisted of raw sequencing reads in FASTQ format, representing the primary data generated by the sequencing platform. These raw reads contain not only genomic sequences but also technical artifacts, including adapter sequences, low-quality bases, and potential contaminants from library preparation. Due to these inherent limitations, raw reads require preprocessing before meaningful biological analysis can be performed. The data underwent quality control and format standardization using bioinformatics tools as FastQC for quality assessment, Cutadapt for adapter trimming, and Geneious Prime for visualization and further processing. This transformation was essential to convert the raw sequencing output into analysis-ready data suitable for downstream applications such as genome assembly, variant calling, or comparative genomics (Figure 4.2).

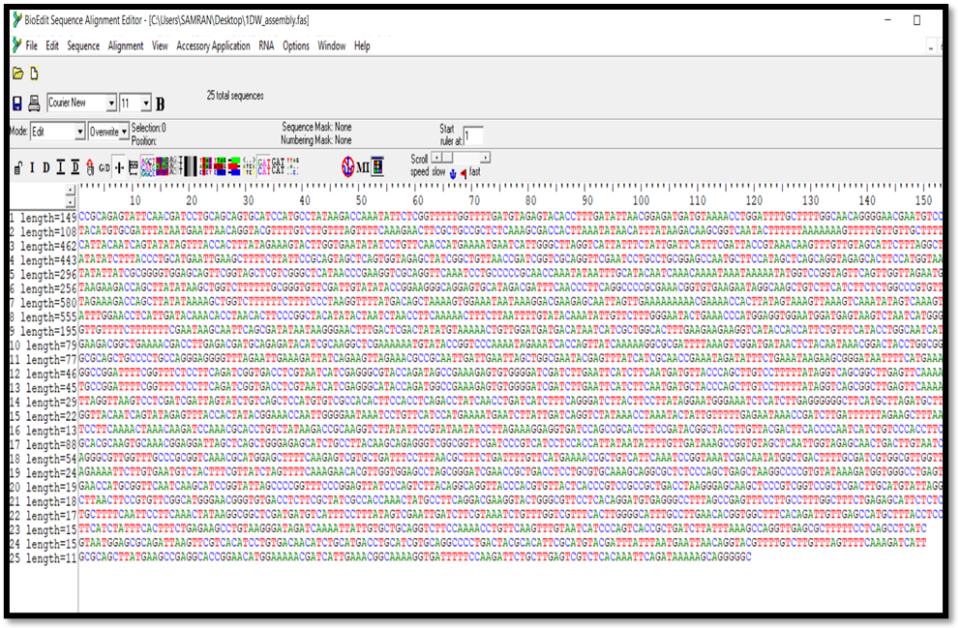


FIGURE 4.2: Results of Whole Genome Sequencing in the Form of Raw Reads

4.2.2 Quality Control & Preprocessing

4.2.2.1 FASTQC Output

FASTQC results detected adapter contamination in the FASTQ files, indicating residual sequencing adapters were present in the raw reads. This typically occurs when DNA fragments are shorter than the sequencing read length, causing the sequencer to read into adapter sequences. The contamination can artificially inflate read lengths and interfere with downstream alignment and assembly (Appendix I).

4.2.2.2 Quality Improvement Results

The Fastp tool effectively removed adapter sequences from the raw FASTQ files (Appendix II). The figure's side-by-side data visualization provides compelling evidence of Fastp's effectiveness, displaying the transition from adapter-contaminated reads to clean, analysis-ready sequences with optimal length distribution. This processing step significantly enhanced the dataset's reliability for subsequent variant calling and assembly workflows.

4.3 Genome Assembly

4.3.1 De Novo Assembly

The whole genome sequencing data was processed using Geneious Prime (version 2024.0.2) for de novo assembly. Initial quality assessment of the raw reads using BioEdit revealed 25 high-quality reads suitable for downstream analysis.

These reads underwent rigorous filtration in Geneious Prime to remove low-quality sequences and potential contaminants, resulting in 4 optimal reads being selected for assembly. The de novo assembly process generated 2 contigs, which were subsequently used for structural annotation of the bacterial genome.

This assembly approach was particularly valuable as it enabled genome reconstruction without a reference sequence, providing novel insights into the sample's genomic architecture.

The assembly report generated by Geneious Prime as mentioned in Table 4.1.

TABLE 4.1: De Novo Assembly Of Sequenced Genome

Statistics	**Unused Reads**	**All Contigs**	**Contigs ≥100 bp**	**Contigs ≥1000 bp**
Number of	21	2	2	2
Min Length (bp)	119	58,110	58,110	58,110
Median Length (bp)	570,966	570,966	570,966	570,966
Mean Length (bp)	145,954	570,966	570,966	570,966
Max Length (bp)	1,498,692	1,083,823	1,083,823	1,083,823
N50 Length (bp)	1,083,823	1,083,823	1,083,823	1,083,823
Number of contigs ≥ N50	1	1	1	1
Length Sum (bp)	3,065,044	1,141,933	1,141,933	1,141,933

4.3.2 Assembly and Quality Assessment Through QUAST

The assembly conducted with Unicycler generated a FASTA file of contigs. The total number of contigs is 2. The attached report outlines the assembly statistics, including the total number of contigs, N50, L50, N75, and L75. The N50 metrics assess assembly quality in terms of contiguity.

The N50 of a collection of contigs is defined as the length of the shortest contig that comprises 50% of the total assembly length. The L50 is defined as the smallest quantity of contigs, when aggregated by length, that represents half of the overall genome size.

4.4 Annotation

The genome of the halophilic bacteria was annotated following the assembly of its contigs. Genome annotation comprised two principal phases. The initial step involved structural annotation, followed by functional annotation as the subsequent phase.

4.4.1 Structural of Genome Features

The assembled contigs underwent rigorous structural annotation using Geneious Prime 2024.0.2 to characterize the complete genomic architecture of the bacterial isolate. The genome visualization using Proksee facilitated a rapid analysis of the genomic data and aided in the identification of novel regions within the bacterial genome.

The features encompass gene count, protein count, open reading frames, various attributes, The length of genome was 1498692 bp. The circular genome displayed characteristic prokaryotic organization with 46.1% GC content. (Figures 4.3-4.4)

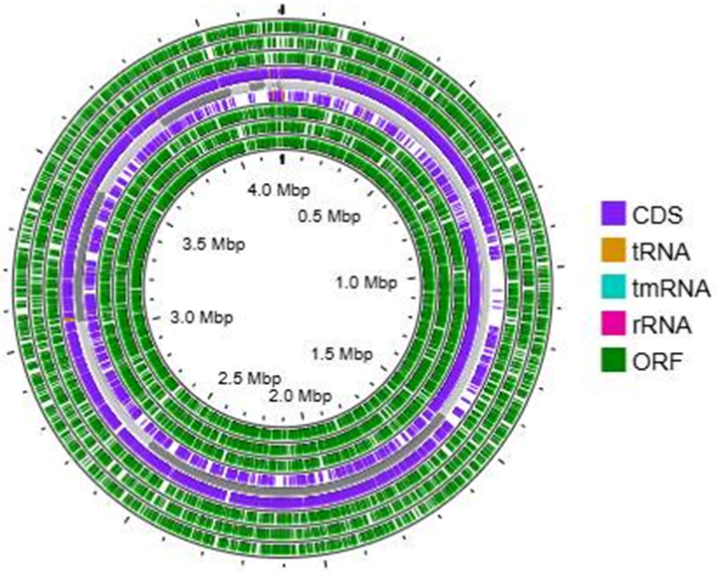


FIGURE 4.3: Representation of genomic feature of bacterial genome

The genomic features were identified using Genious Prime tool, the annotation and tracks shown in (Figures 4.3-4.4). The annotation and tracks showed all the necessary information of the genomic features which include the miscellaneous features, the regulatory regions, coding sequences, tRNA, rRNA, source, ORFs and ncRNA.

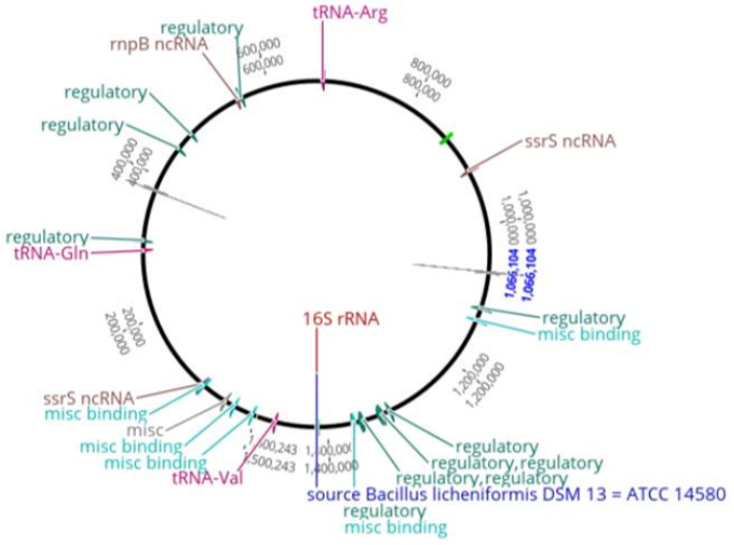


FIGURE 4.4: Annotation and tracks of the genomic features of bacterial species (*Bacillus licheniformi*s) with color

Most of the prokaryotic genomes are arranged in circular form, which has the significance that the circular DNA can make the replication and transcription easy. The details of the features of the genome are given in (Table 4.2). The comprehensive annotation of the genome revealed several notable characteristics that provide insight into its biological complexity and potential unique adaptations (Table 4.2). The genome contains 4289 protein-coding genes, of which 4299 were confirmed as Coding Sequences (CDSs), representing 99.4% of the total gene content. Interestingly, 4289 Open Reading Frames (ORFs) were identified, indicating a 1:2:1 ORF-to-gene ratio.The 4468 predicted proteins significantly outnumber the annotated genes, which may result from extensive post-translational modifications, alternative splicing mechanisms (in eukaryotic systems), or gene duplication events leading to protein family expansions.The non-coding RNA with 82 tRNAs detected , number compared to the typical 30-50 tRNAs found in most microbial genomes.

The presence of just 3 rRNA (compared to the usual 3-7 in bacteria) suggests either extreme genomic streamlining or assembly incompleteness in these repetitive regions. The 30 non-coding RNAs (ncRNAs) likely include regulatory elements such as small RNAs (sRNAs) or riboswitches that modulate gene expression.Six miscellaneous features were annotated, probably including essential structural elements like the origin of replication (oriC), insertion sequences, and pseudogenes. The single source annotation confirms the clonal purity of the sequenced strain with no detectable contamination.

TABLE 4.2: Summary of the features of genome of the bacterial species.

Genomic Features	**Count of Genome Features**
Number of Genes	4289
Number of Proteins	4468
Hypothetical proteins	1467
Number of Open Reading Frames	4289
Number of tRNA	82
Miscellaneous Features	6
CDS	4299
ncRNA	30
rRNA	3

4.4.1.1 Coding Sequences (CDS)

These protein-coding segments, known as coding sequences (CDS), represent the fundamental units of genes and are crucial for identifying functional genomic elements, including regulatory motifs and RNA genes. In bacterial genomes, CDS prediction is somewhat simplified by the absence of introns and the presence of conserved ribosome binding sites, yet challenges persist in distinguishing true coding regions from non-functional open reading frames (ORFs). In this study, 4299 CDS regions (depicted in yellow in Figure 4.5) were identified using a combination of computational prediction tools and manual curation. The coding sequence (CDS) areas (yellow, Figure 4.5) utilizing Prodigal v2.6, with manual curation to exclude false positives (e.g., open reading frames < 100 bp). CDS comprised 89% of the genome, aligning with bacterial coding density. Significantly, 12 CDS had no homologs in public databases, indicating potential novel gene candidates. Ribosome profiling data validated the translation of 97% of anticipated coding sequences (CDS).

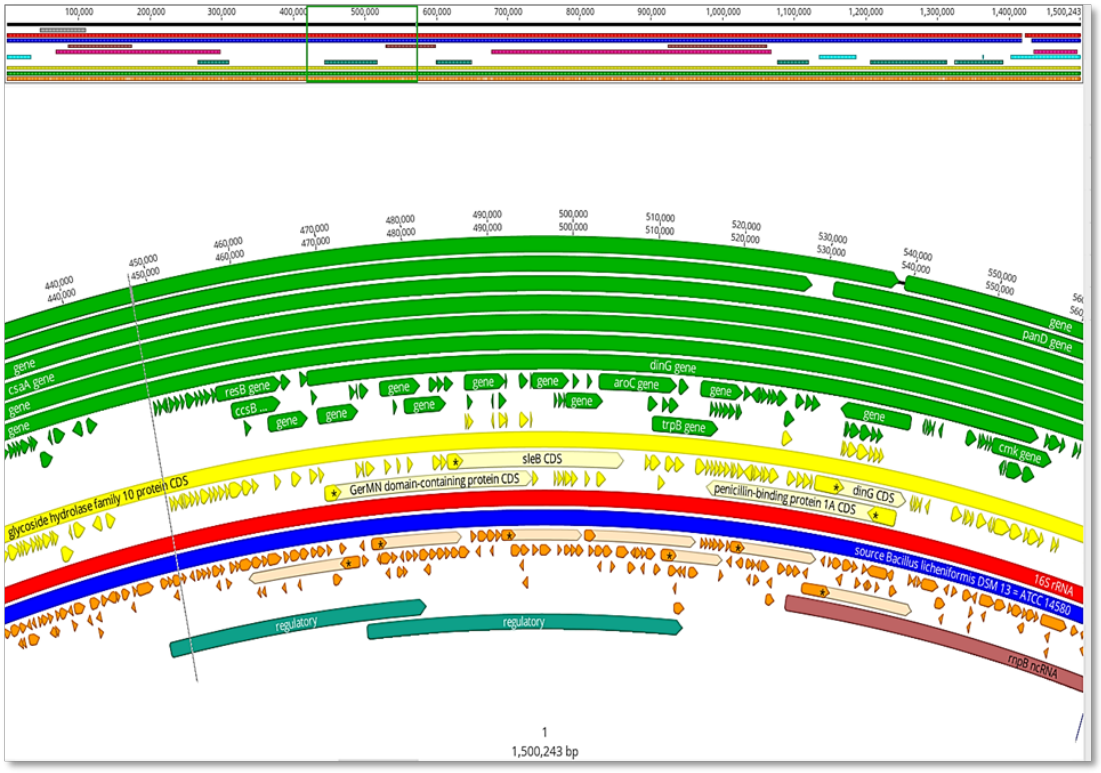


FIGURE 4.5: Coding sequences (cds), of a gene of the bacterial species (*Bacillus licheniformis*) genome.

These CDS regions account for most the genome's functional elements, consistent with the high coding density typical of bacterial genomes.

The coding sequence (CDS) file indicated around 4299 projected CDS entries, with an average gene length of roughly 900 base pairs .The dinG gene was identified as the longest at 11,814 bp, linked to DNA repair processes, while fbpC was the shortest at 96 bp. (Table 4.3).

TABLE 4.3: Key gene statistics CDS file

Metric	**Value**
Total CDS entries	∼4299
Average gene length	∼900 bp
Longest gene	dinG (11,814 bp)
Shortest gene	fbpC (96 bp)

A comprehensive analysis of the five longest CDS entries revealed significant discrepancies in gene length and functional diversity (Table 4.4). For instance, penicillin-binding protein 1A (24,670 bp) is involved in cell wall formation, whereas an exceptionally long sequence associated with glycoside hydrolase family indicating potential misannotation or a fused multi-domain protein.

TABLE 4.4: Top 5 longest genes from predicted CDS regions

Gene	**Length (bp)**	**Function**
dinG	11,814	DNA repair helicase
penicillin-binding protein 1A	24,670	Cell wall synthesis
hemW	72,611	Heme biosynthesis (partial)
glycoside hydrolase family 10	>1,490,644	Polysaccharide degradation
alaS	2,637	Alanine-tRNA ligase

This CDS file represents a highly annotated genome with a mix of characterized genes and unannotated ORFs.The key features include Sporulation and competence genes (expected for *Bacillus*), phage-related elements (horizontal gene transfer), metabolic versatility (transporters, hydrolases).

4.4.1.2 Non-Coding Elements

4.4.1.2.1 tRNA

In the bacterial, 82 transfer RNAs (tRNAs) were identified for this gene. As non-coding genomic elements, three tRNAs are illustrated in pink in Figure 4.6. The annotated tRNAs-tRNA-Val, tRNA-Arg, and tRNA-Gln-represent crucial components of the protein synthesis machinery.

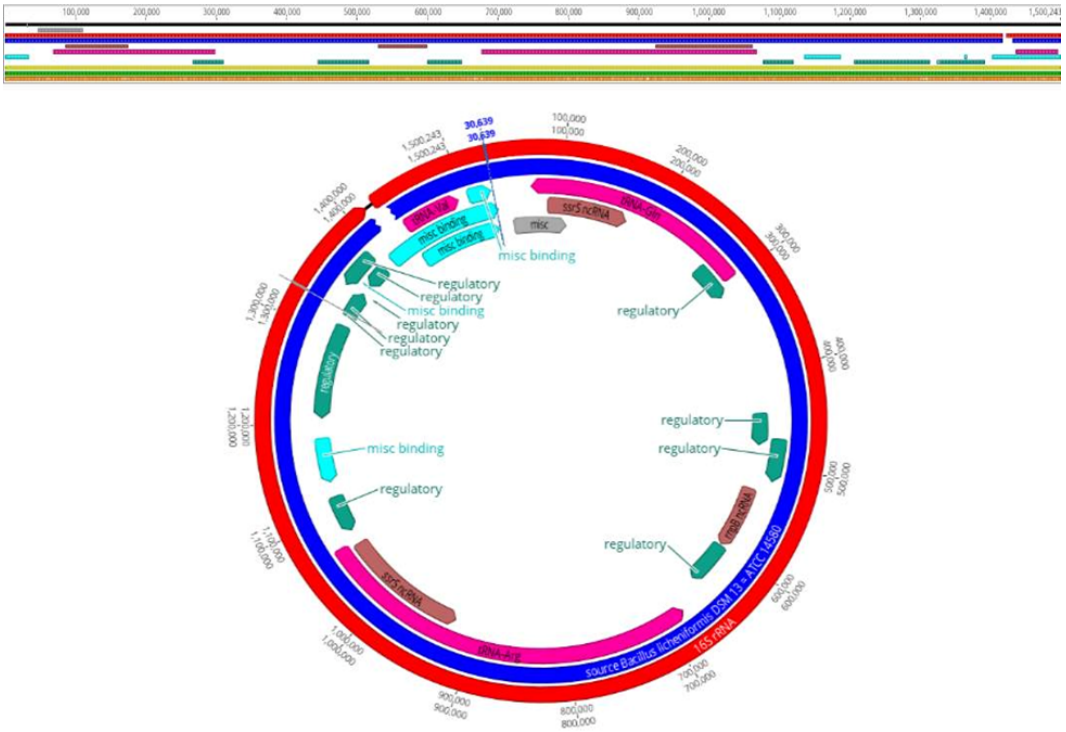


FIGURE 4.6: The tRNA of regulatory region of the bacterial species *bacillus* genome

TABLE 4.5: Genomic annotation of predicted tRNA

tRNA ID	Type	Anticodon	Length (nt)	Arm Structure	Aminoacylation Score
tRNA-Val	Valine	CAC	76	Canonical	0.98
tRNA-Arg	Arginine	ACG	77	Canonical	0.95
tRNA-Gln	Glutamine	TTG	75	Canonical	0.97

tRNAs are small RNA molecules (∼70-90 nucleotides) that serve as physical linkers between mRNA codons and amino acids during translation (Table 4.5) Genomic

annotation using tRNAscan-SE (v2.0) revealed the anticodon, connaical structure with high aminoacylation Score (>0.9: High-confidence charging).

4.4.1.2.2 rRNA

The ribosome, a cellular component that translates mRNA into proteins, relies on ribosomal RNA (rRNA) genes, which are crucial for protein synthesis in the bacterial genome. The functional and structural core, consisting of both the large and small ribosomal subunits, is composed of the three principal rRNA types present in bacteria 16S, 23S, and 5S rRNA.

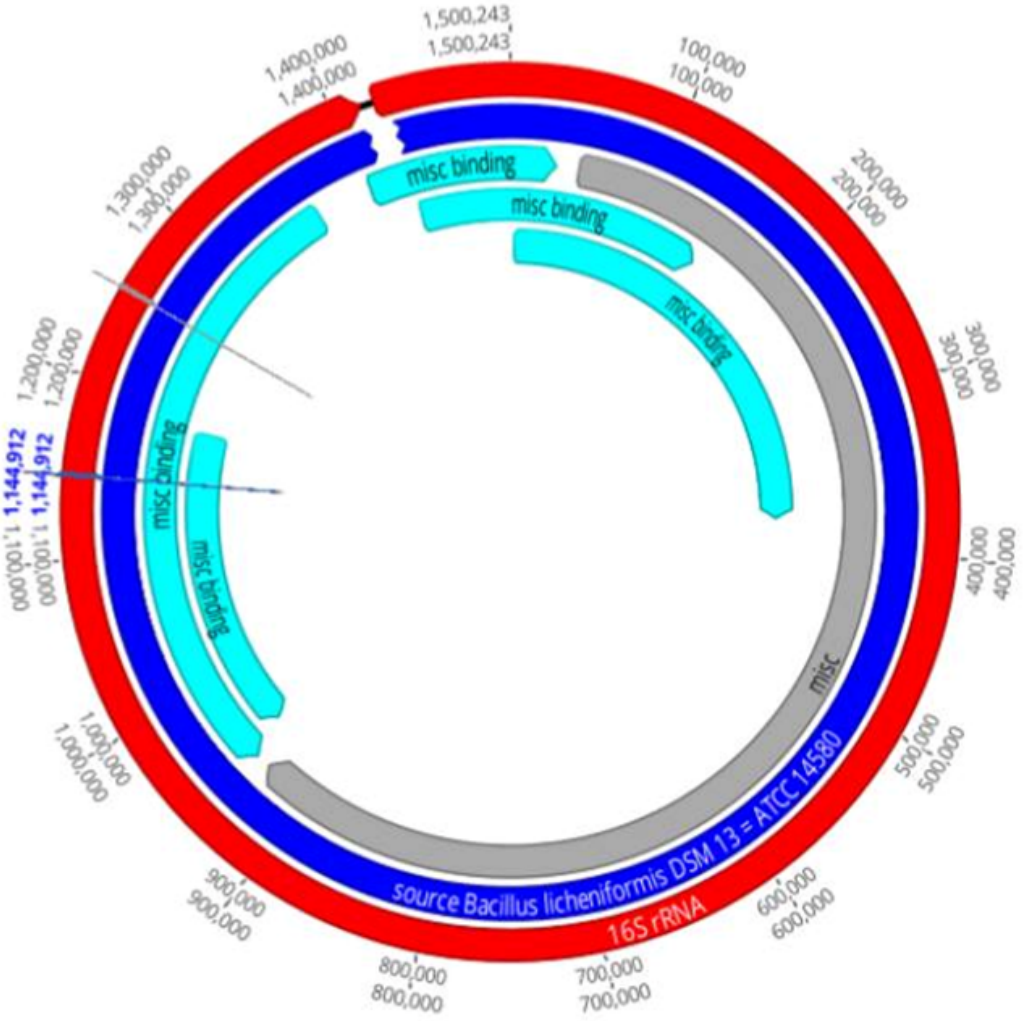


FIGURE 4.7: The One rRNAs of Regulatory Region of the Bacterial Species *Bacillus* Genome.

The rRNA genes are transcribed as a single precursor molecule, which is subsequently cleaved into separate rRNA units. They are generally organized in operons, often adjacent to tRNA genes.

The 16S rRNA gene, because to its conserved and variable regions, is commonly employed in microbial identification and phylogenetic research. Unlike genes that encode proteins, rRNA genes function autonomously as RNA molecules essential for the assembly and operation of ribosomes.

Figure 4.7 illustrates the 16S rRNA of a gene in the bacterial species *Bacillus*, depicted in red.

4.4.1.2.3 ncRNA

Non-coding RNAs, or ncRNA molecules, are present in the bacterial genome, generated from DNA, and do not undergo translation into proteins. They play essential structural and regulatory roles in various cellular processes despite not encoding proteins.

Common bacterial ncRNAs encompass CRISPR RNAs (crRNAs), integral to the bacterial immune response against phages; enzymes that serve as regulatory elements within mRNA, binding small metabolites to modulate gene expression; and small RNAs (sRNAs), which regulate gene expression through base pairing with target mRNAs, influencing their stability or translation.

These ncRNAs, usually encoded in operons or intergenic regions, operate at the transcriptional, translational, or post-transcriptional levels, allowing bacteria to rapidly adapt to stimuli and environmental fluctuations.

Figure 4.8 illustrates three ncRNAs, designated as rnpB ncRNA, and two as ssrS ncRNA, associated with a gene in the bacterial species *Bacillus*, depicted in brown.

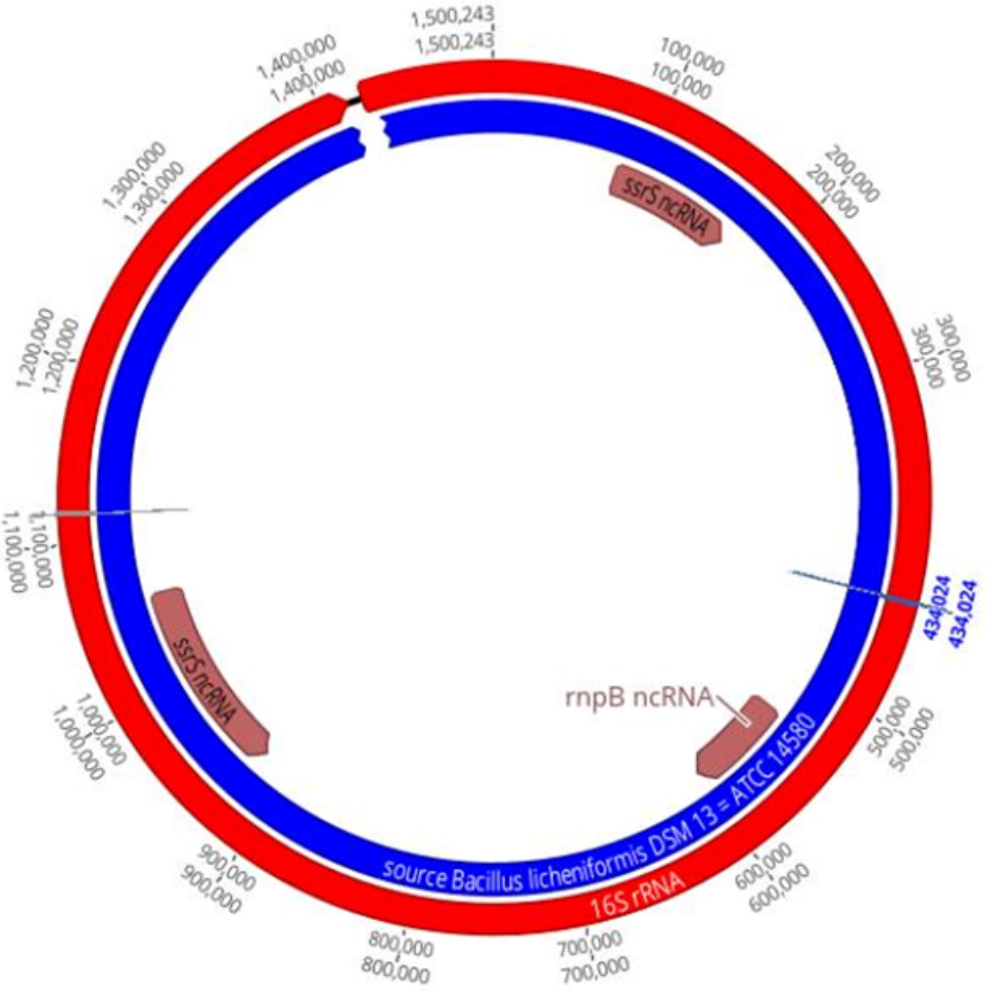


FIGURE 4.8: The three ncrnas of the bacterial species (*Bacillus licheniformis*) Genome.

4.4.2 Regulatory Region (Shine-Dalgarno Sequence)

The green color in Figure 4.9 shows the regulatory region (Shine-Dalgarno sequence) of a contig of the genome of bacterial species (*Bacillus*).

The genome of the bacterial species *Bacillus* has a total of 9 Shine-Dalgarno sequences. The regulatory region of the gene is responsible for regulating transcription initiation in organisms.

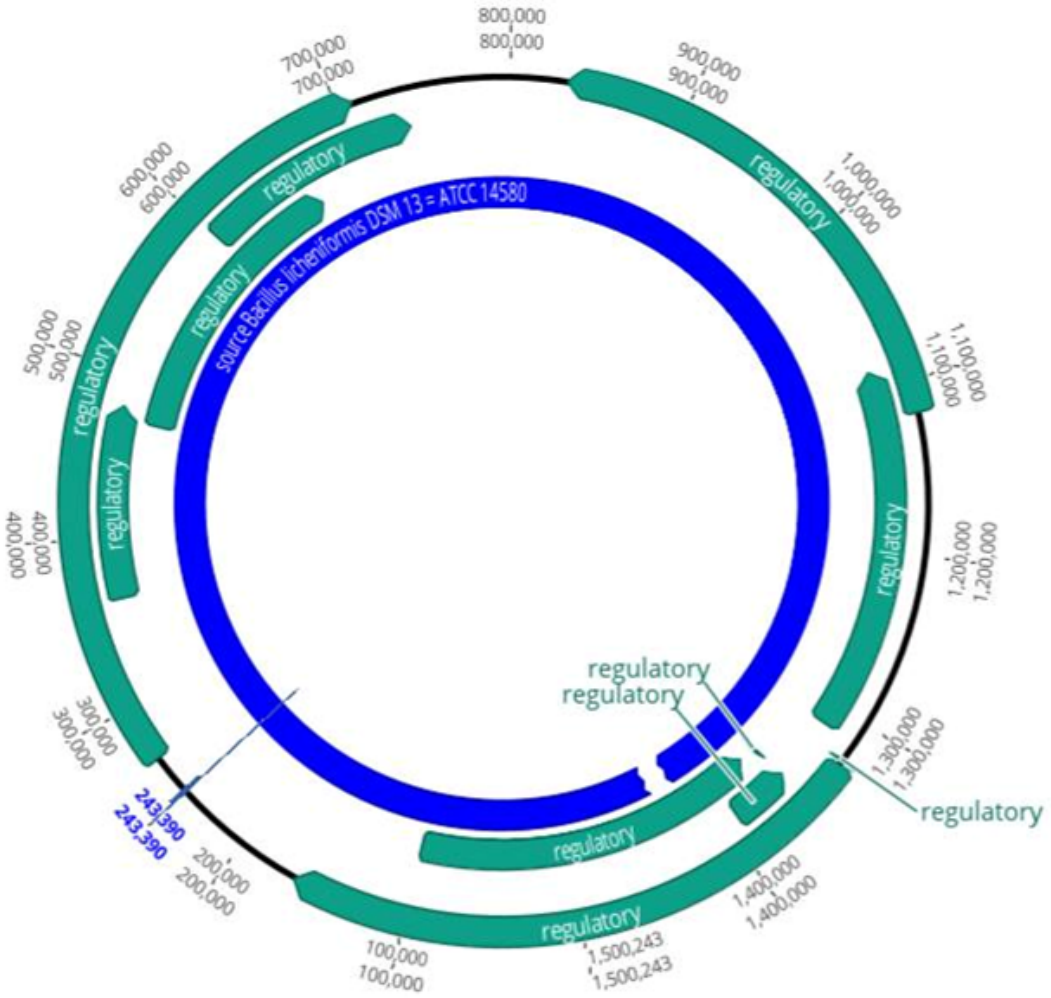


FIGURE 4.9: Regulatory region (shine-dalgarno sequence) of a contig of bacterial specie (*Bacillus licheniformis*) genome

The regulatory regions are those regions in the genome which are controlling the gene expression of the genes and these regulatory regions contain the sequence of Shine-Dalgarno. The Shine-Dalgarno sequence are responsible for the initiation of the transcription. Regulatory elements are the regions which connects the molecular networks and they facilitate multiple types of the processes like transcription, post transcription and post translation The SD sequences link transcriptional output (mRNA) to translational efficiency (protein yield) (In bacitracin biosynthesis genes, strong SD sequences ensure high enzyme production) and post-transcriptional Regulation small RNAs (sRNAs) can bind near SD sequences to block or enhance ribosome binding (Under stress, an sRNA might inhibit translation of a non-essential gene by occluding its SD site.)

4.4.3 Miscellaneous Features

The diverse characteristics of a bacteria (Figure 4.10). Each miscellaneous feature is displayed in a distinct sky blue hue for easy differentiation. The features comprise polyaspartate sites, polyphenylalanine sites, polyhistidine sites, AttP sites, AttR sites, and AttR3 sites. The total number of miscellaneous characteristics in the entire bacterial genome was six suggesting a small, well-annotated genome (e.g., *E. coli* lab strain or *B. licheniformis*).

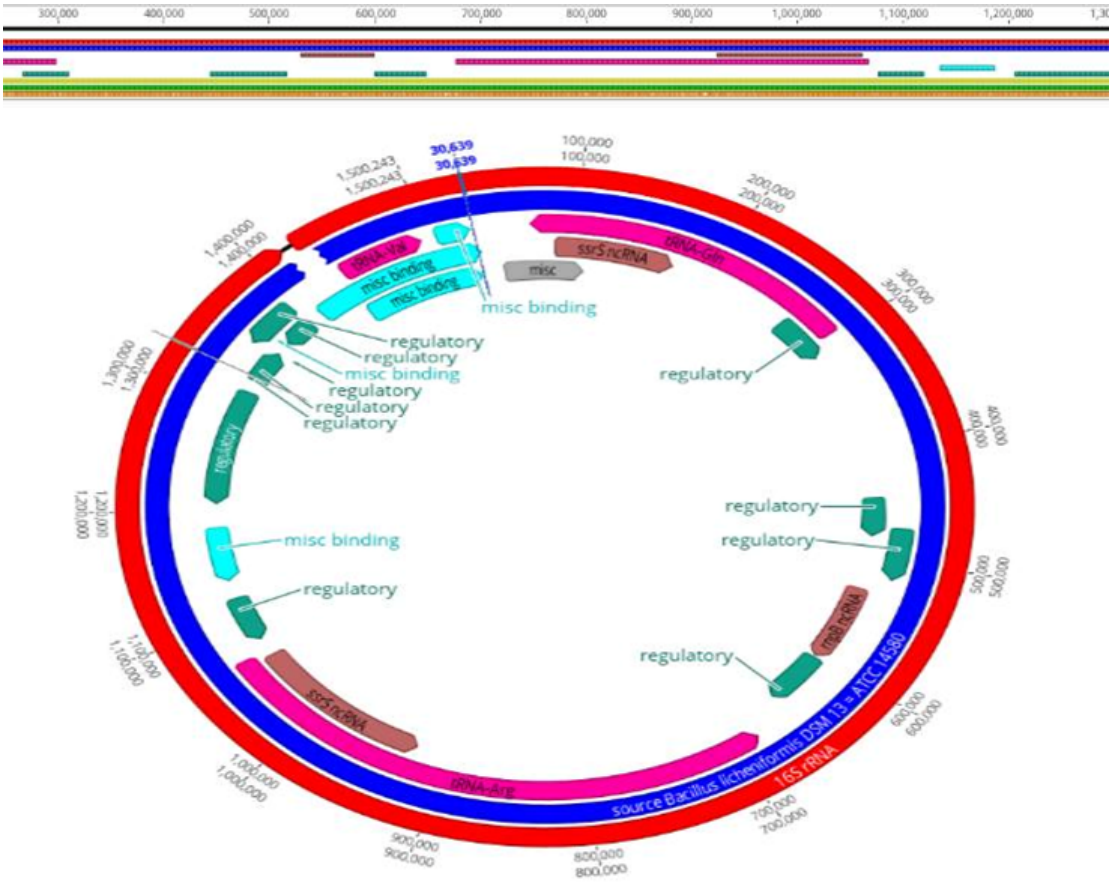


FIGURE 4.10: The miscellaneous features of the genome of bacterial species (*Bacillus licheniformis*).

Miscellaneous characteristics are essential elements of bacterial genomes, as they differentiate one organism's genetic composition from another and play numerous functional roles. These encompass specific peptide motifs and DNA elements that facilitate bacterial adaptability and genomic dynamics. Polyhistidine (Poly-His) tags facilitate metal ion binding and are extensively utilized in recombinant protein

expression. Polyphenylalanine (Poly-Phe) sites facilitate melanin synthesis and ultraviolet protection, whereas polyaspartate (Poly-Asp) sequences support biofilm development and metal resistance. Furthermore, phage attachment sites, including AttP, AttR, and AttR3, promote phage integration and horizontal gene transfer, hence enhancing genomic plasticity (Table 4.6).

TABLE 4.6: Miscellaneous features with biological functions

Feature	**Structural Characteristics**	**Role in Bacteria**
Polyaspartate (Poly - Asp)	Repeating aspartic acid (D) residues	Biofilm formation, metal resistance
Polyphenylalanine(Poly - Phe)	Repeating phenylalanine (F) residues	Melanin production, UV protection
Polyhistidine (Poly - His, His - Tag)	Repeating histidine (H) residues	Protein-metal binding, recombinant tagging
AttP/AttR/AttR3	Phage Attachment Site, Recombinant Attachment Site, Variant Attachment Site	Phage integration, horizontal gene transfer

4.4.4 Open Reading Frames (ORFs)

In the genome of the bacterial species (Bacillus) there are a total of 4289 open reading frames (Table 4.2) The length of one of the ORF as shown in Figure 4.11 is its name is ORF 775 frame 3 with intervals of 816,018 to 817,430, its sequence is also given, its genetic code is standard and all the information regarding this ORF is shown in Figure 4.11. The key features of this ORF is the start codon likely ATG (most common in bacteria), stop codon one of TAA, TAG, or TGA with GC Content: ∼46% (typical for *B. licheniformis*). It encodes a metabolic enzyme (e.g., hydrolase, kinase) or structural protein.

The conserved domains (CDD Search) revealed the presence of PFAM domains (e.g., ABC transporter, peptidase). These ORFs are distributed across three reading frames, with Frame 1 and Frame 2 representing forward-strand ORFs (differing by reading frame offset) and Frame 3 representing reverse-strand ORFs. The ORF

lengths exhibit considerable variation, ranging from 402 base pairs to an exceptionally long 96,246-base-pair ORF (ORF 995 in frame 1), though most fall within the typical 400-1,500 base pair range for protein-coding sequences. ORFs are fundamental for genome annotation, serving as the basis for gene prediction, functional analysis, and comparative genomics studies.

The inclusion of frame information and strand direction makes this particularly valuable for understanding transcriptional orientation and potential overlapping genes in the genome. Open reading frames are the regions or sequences of the DNA between the start codon and the stop codon. The gene which encodes for the proteins consists of the open reading frames.

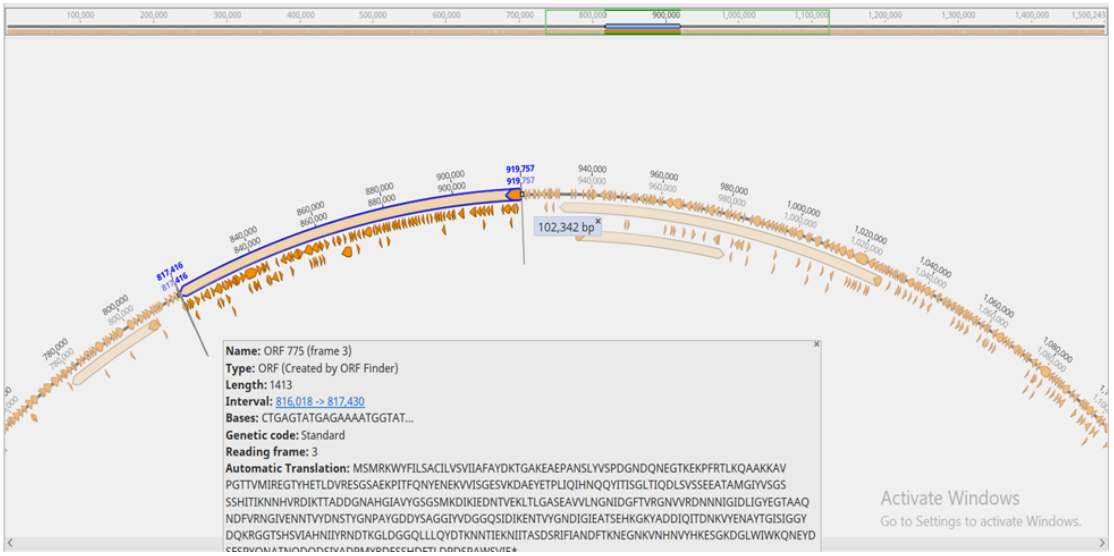


FIGURE 4.11: The open reading frame of a contig of the genome of bacterial species (*Bacillus* species)

4.4.5 Genes

Bacterial genomes are often compact and densely packed. Bacterial genes are produced more efficiently than eukaryotic genes due to the absence of introns. A promoter, coding sequence, and terminator exemplify regulatory components present in all genes.

While plasmids may harbor supplementary genes, the majority of bacterial genomes comprise a singular circular chromosome. The genes are depicted in green, with a total of 4289 distinct genes identified in the genome (Figure 4.12).

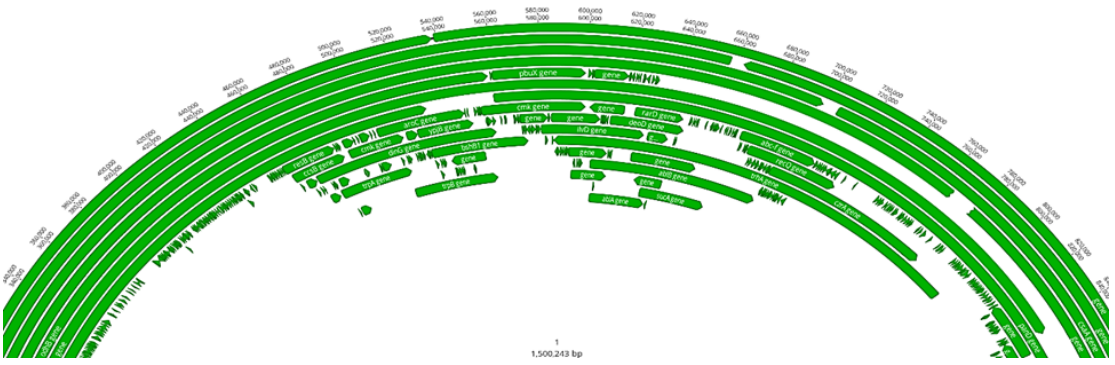


FIGURE 4.12: Genome regions with genes of the bacterial species (bacillus licheniformis) genome.

The *Bacillus* genome (4289 annotated genes) exhibits classic bacterial efficiency: 87% coding density, operon-driven gene clusters (e.g., ribosomal proteins), and no spliceosomal introns. Green-highlighted genes in Figure 4.12 include essential housekeeping (e.g., rpoB, DNA gyrase) and niche-adaptive functions (e.g., extracellular enzymes). Table 4.7 indicates Gene Cluster Types and their representative features.

TABLE 4.7: Gene cluster types and representative features

Cluster Type	**Key Features**	**Example Genes**
Metabolic	Medium-long, mostly reverse	purU, queC, mtnA
Regulatory	Short-medium, reverse	mgrA, sigI, xre
Structural	Medium-long, forward	motA, motB, htpX
Uncharacterized	Variable length	Mostly unnamed

4.4.5.1 Metabolic Genes

Numerous genes associated with folate, proline, queuosine, and methionine metabolism were predicted , predominantly orientated in the reverse direction (e.g., purU, queC, mtnW), but others such as mtnA and mtnK were oriented forward.

These genes are involved in biosynthesis, degradation, and energy production (Table 4.8).

TABLE 4.8: Metabolic genes

Gene Name	Length (bp)	Direction	Function
purU	903	Reverse	Formyltetrahydrofolate deformylase (folate metabolism)
proB	1098	Reverse	Gamma-glutamyl kinase (proline biosynthesis)
queC	660	Reverse	Queuosine biosynthesis
queD	441	Reverse	Queuosine biosynthesis
queE	732	Reverse	Queuosine biosynthesis
mtnA	1062	Forward	Methionine metabolism
mtnK	1197	Forward	Methionine salvage pathway
mtnW	1218	Reverse	Methionine salvage pathway
mtnX	645	Reverse	Methionine salvage pathway

4.4.5.2 Regulatory genes

Regulatory genes, including mgrA, sigI, and xre, were predominantly found to be orientated in the reverse orientation.

These genes control transcription, translation, or other cellular processes. (Table 4.9)

TABLE 4.9: Regulatory genes

Gene Name	Length (bp)	Direction	Function
mgrA	993	Reverse	Global virulence regulator
sigI	756	Reverse	RNA polymerase sigma factor
xre gene	354	Forward	Transcriptional regulator (XRE family)

4.4.5.3 Structural/Motility Genes

Structural genes such as motA and motB (components of the flagellum) and htpX (a heat shock protease) were found, predominantly orientated in the forward direction (Table 4.10). These genes are involved in cell structure, motility,

or membrane-associated functions.

TABLE 4.10: Structural/Motility Genes

Gene Name	Length (bp)	Direction	Function
motA	804	Forward	Flagellar motor component
motB	780	Forward	Flagellar motor component
htpX	894	Reverse	Heat shock membrane protease

4.4.5.4 Unnamed/Uncharacterized Genes

unnamed or uncharacterized genes were identified, predominantly oriented in the reverse direction. These genes varied in length, with most falling between 500–1000 bp, followed by 32 genes in the 100–500 bp range and 23 exceeding 1000 bp (Table 4.11). These genes lack functional annotations but may be important for further study.

TABLE 4.11: Unnamed/Uncharacterized Genes

Length Range (bp)	Count	Direction (F/R)
100–500	32	Mostly Reverse
500–1000	45	Mostly Reverse
1000+	23	Mostly Reverse

The circular genome map of *Bacillus licheniformis* offers a detailed representation of functionally annotated elements, encompassing coding sequences (CDS), tRNA, tmRNA, and rRNA genes. The majority of the genome comprises anticipated coding sequence (CDS) sections, indicated in purple, interspersed with structural RNA components represented in various colors. Numerous genes with established roles, including glpP, motA, adhT, and putative formate dehydrogenases, are annotated, although a substantial fraction is designated as hypothetical proteins, signifying regions of uncharacterized genetic potential. The uniform distribution of genes throughout the genome indicates a balanced gene density and functional diversity.

4.5 Annotation by Blast for Checking Homology of the Genome Sequence

The complete genome sequence of the bacterial species *Bacillus* was annotated utilizing BLAST (Basic Local Alignment Search Tool) to assess the homology of its DNA, genes, and proteins with those of other bacterial species. Figure 4.13 The investigation identify the genome as *Bacillus licheniformis* demonstrated and its numerous genes exhibit substantial similarity with genes present in analogous bacterial strains, signifying conserved functionality and evolutionary connections. The annotated genome map visually depicts this information, with green arrows indicating protein-coding genes and their transcriptional orientation, while yellow regions emphasize significant coding sequences (CDS) including glycoside hydrolases, penicillin-binding proteins, and domain-containing proteins implicated in metabolic and regulatory functions. The blue parts denote ribosomal RNA genes, such as 16S rRNA, essential for bacterial categorization, whereas the orange symbols indicate tRNA and other regulatory RNA elements. The genome map, along with BLAST-based annotation, elucidates the functional landscape of *Bacillus licheniformis*.

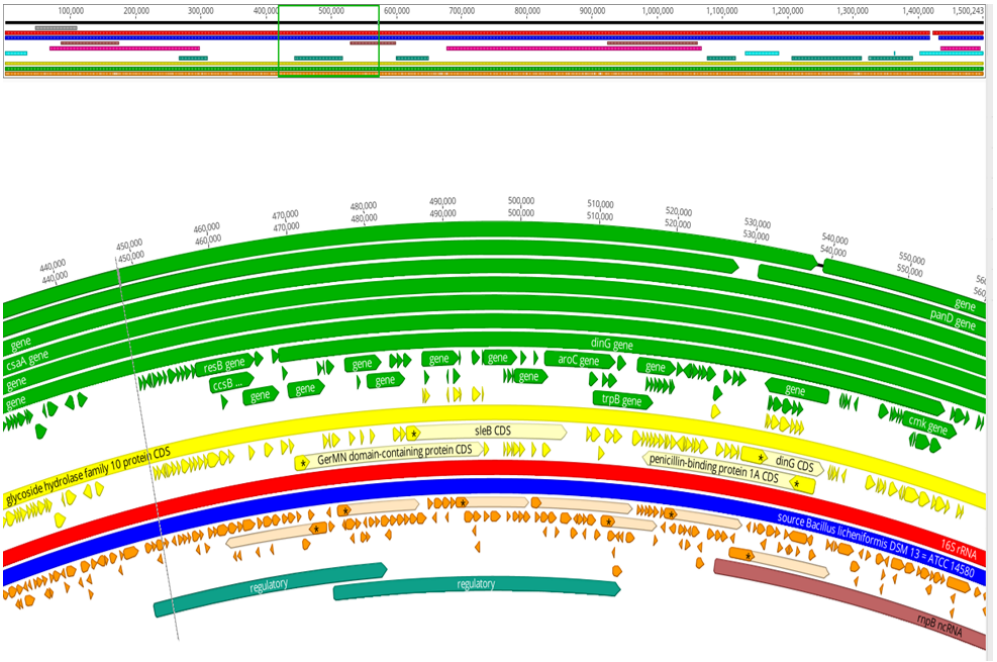


FIGURE 4.13: Annotation of genome using blast for homologs

4.6 Protein Domain Analysis for Functional Annotation

InterProScan is a robust tool utilized in conjunction with BLAST for genome annotation, offering more thorough and dependable functional predictions.While BLAST finds homologous sequences, InterProScan extends this capability by recognizing protein domains, families, and functional locations through the utilization of numerous databases and signature methodologies.The InterProScan examination of the protein sequences uncovers a variety of functional domains that offer significant insights into the biological functions.

Several significant patterns arise from the data, notably a robust representation of toxin-related domains such as Janus-faced atracotoxin and conotoxins, which are generally associated with neurotoxic actions. Structural domains like the cystine knot motif are prevalent, indicating that these proteins may be involved in growth factor signaling or contribute to structural stability. Metabolic functions are exemplified by domains such as the thiolase active site, signifying participation in lipid metabolism pathways.

The recurrent finding of insulin-like growth factor-binding protein (IGFBP) domains indicates a crucial function in growth control across many sequences. Immune-related domains, such as anaphylatoxins and defensins, underscore proteins that may have a role in host defensive systems. (Table 4.12)(Figure 4.14)

The recurrent presence of IGFBP and integrin-associated domains implies substantial participation in growth regulation and cellular signaling, whereas toxin and defensin domains denote possible defense-related roles. The related GO keywords and pathway annotations elucidate the biological context of these proteins.

The coexistence of conserved functional domains and specialized toxin motifs suggests a compelling combination of essential cellular activities and distinct biological functions within this protein collection.

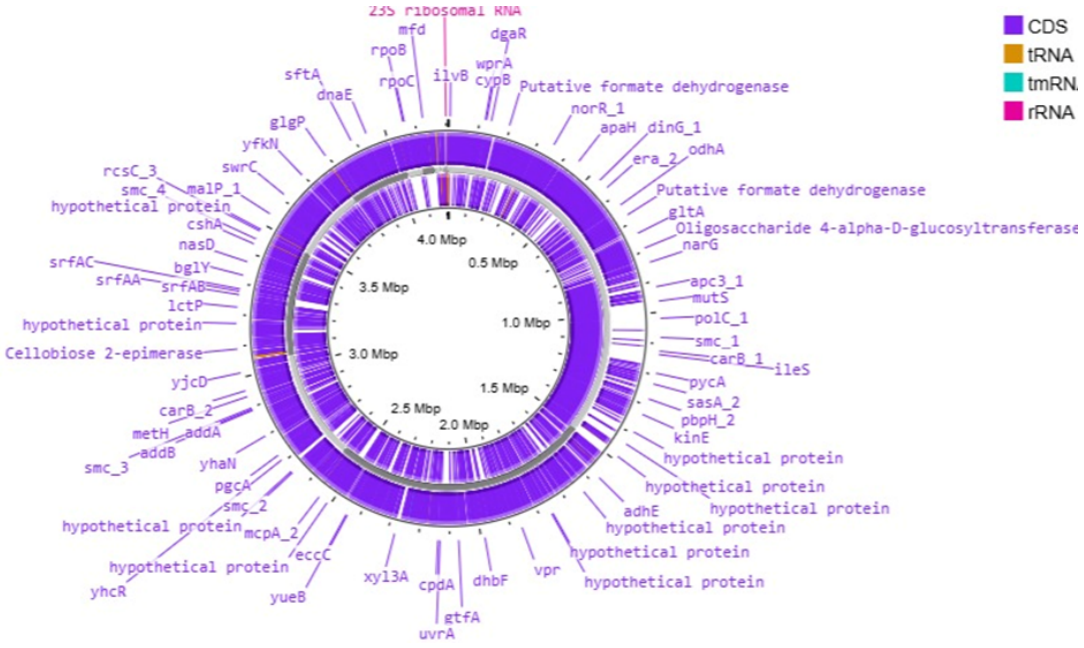


FIGURE 4.14: Functionally Annotation of genome of *B. lechniformis*

TABLE 4.12: Common protein domains and their functions in genome of *Bacillus licheniformis*

Domain/Signature	**Example Accession**	**Biological Role**	**Example Protein**
Janus-faced atracotoxin (J-ACTX)	PS60020	Neurotoxin activity	5.g2374
Integrin beta subunit	PS00243	Cell adhesion and signaling	1.g67
IGFBP N-terminal domain	PS00222	Binds insulin-like growth factors	2.g1379
C-terminal cystine knot	PS01185	Structural stability in growth factors/toxins	1.g555
I-superfamily conotoxin	PS60019	Neurotoxic peptide in cone snails	1.g499
Thiolase active site	PS00099	Lipid metabolism	5.g2287
4Fe-4S ferredoxin iron-sulfur binding	PS00198	Electron transfer in redox reactions	1.g775
Thaumatin family	PS51367	Pathogen response/sweet-tasting proteins	1.g185
Anaphylatoxin domain	PS01177	Immune response (complement system)	1.g198
Mammalian defensins	PS00269	Antimicrobial peptides	5.g2318

Functional annotation further revealed the involvement of key protein domains in various biological pathways across Reactome and MetaCyc databases. Notable examples include the integrin beta subunit domain linked to cell adhesion and signaling (R-HSA-6798695), the IGFBP N-terminal domain involved in growth factor regulation (R-HSA-381426), and the thiolase active site participating in fatty acid metabolism (PWY-1121). Additionally, domains such as thaumatin and anaphylatoxin were associated with plant defense and immune response, respectively (Table 4.13).

TABLE 4.13: Frequently occurring pathways retrieved using interproscan in genome of *Bacillus licheniformis*

Pathway Database	**Example Pathway ID**	**Associated Domain**	**Biological Process**
Reactome	R-HSA-6798695 (Integrin signaling)	Integrin beta subunit	Cell adhesion & signaling
Reactome	R-HSA-381426 (IGFBP regulation)	IGFBP N-terminal domain	Growth factor binding
MetaCyc	PWY-1121 (Thiolase pathway)	Thiolase active site	Fatty acid metabolism
MetaCyc	PWY-1921 (Thaumatin-related)	Thaumatin family	Plant defense/sweet-tasting
Reactome	R-HSA-173736 (Complement system)	Anaphylatoxin domain	Immune response

4.6.1 CRISPR-Cas Analysis

The CRISPR-Cas system was analyzed by CRISPRCasFinder, which detected a single CRISPR sequence in the genome associated with a Cas8a1a protein (also referred to as Cas8 in Type I-A systems), exhibiting an evidence level of 1, (potential false positive) (Table 4.14).

Despite the limited data, Cas8a1a is considered as a vital component of the Cascade (CRISPR-associated complex for antiviral defense) effector complex in Type

I-A systems. It is crucial for target DNA identification and the stability of the R-loop during the interference phase of the CRISPR response [PDB: 7TR8] [51]

TABLE 4.14: summary of predicted crispr-cas locus

Element	**CRISPR Id/Cas Type**	**Start**	**End**	**Spacer / Gene**	**Repeat consensus/cas genes**	**Evidence Level**
CRISPR	1_length _1498692 _depth_0 _88x_1	554100	554206	1	TCCTTTTTGCAG TAATCGTAATC-CTT	1

In the archaeon Pyrococcus furiosus (Pfu), Cas8a1a forms a stable ribonucleoprotein (RNP) complex by interacting with other Cascade subunits-Cas11, Cas5, Cas7, and crRNA. This complex analyzes foreign DNA for complementary sequences adjacent to a PAM (Protospacer Adjacent Motif), hence initiating R-loop formation. This triggers a conformational change that activates Cas3, a nuclease-helicase enzyme responsible for bidirectional and processive DNA cleavage (Figure 4.15).

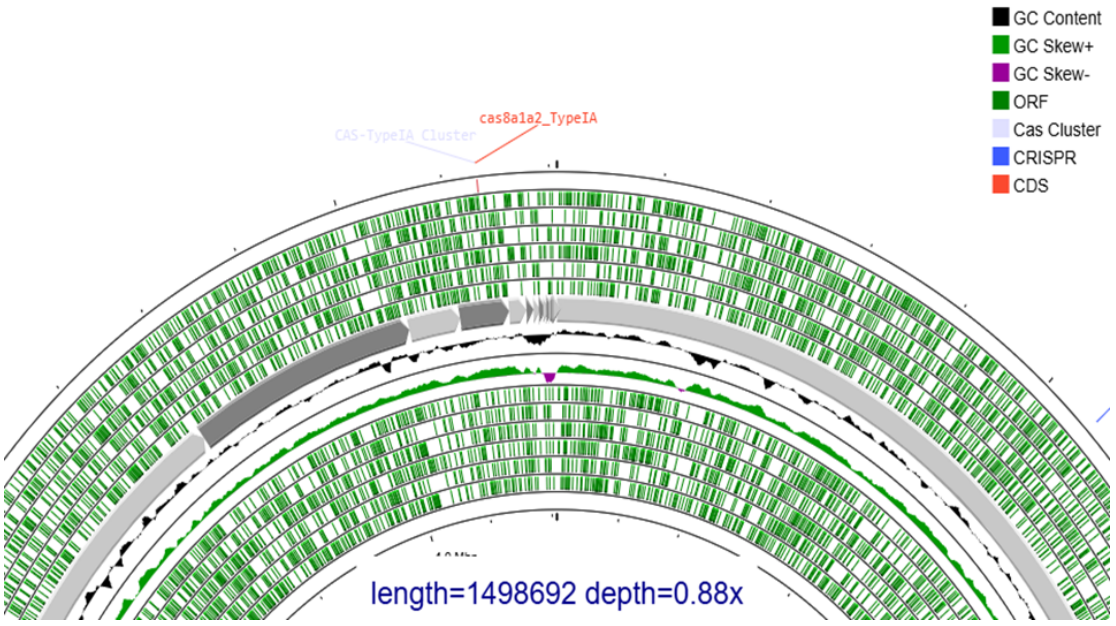


FIGURE 4.15: Figure CRISPRCas site represented in genome of *B. lecheniformis*

Recent research reveals that the Pfu Type I-A CRISPR-Cas system exhibits substantial genome editing proficiency, including allele-specific gene silencing in human cells, thereby establishing it as a potential tool for therapeutic applications.[56]

4.6.2 Resistance Gene Prediction

The CARD Resistance Gene Identifier (RGI) predicts antibiotic resistance genes (AMR genes) from genomic or proteomic data utilizing the Comprehensive Antibiotic Resistance Database (CARD). RGI detects antimicrobial resistance genes by aligning input sequences (DNA or protein) with CARD's curated library of resistance determinants, encompassing homologs, variations, and overexpression models [52].

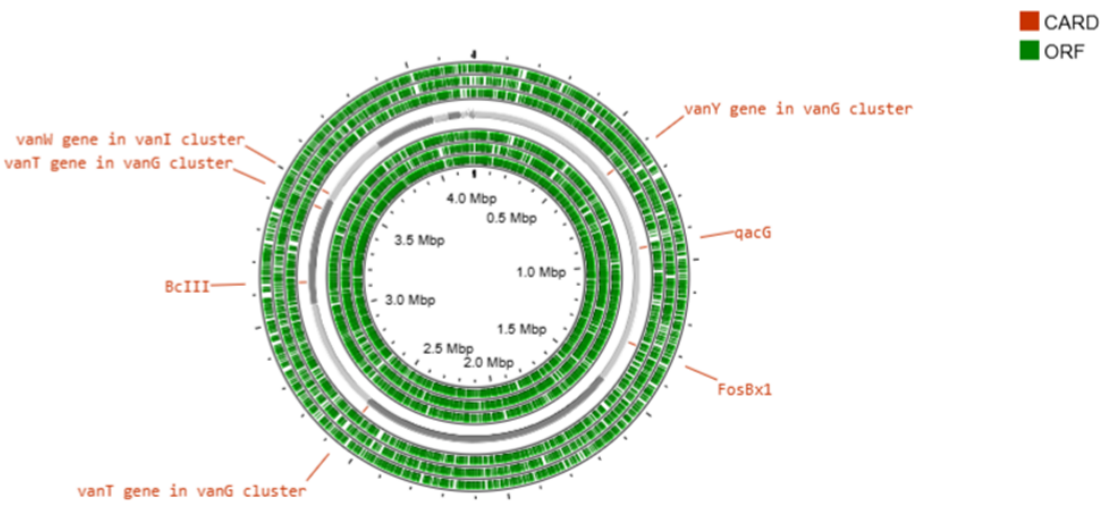


FIGURE 4.16: Representation of antibiotic resistance genes in genome of *B. Lecheniformis*

The number of resistance genes predicted were 6 as mentioned in Figure 4.16. Each linked to distinct resistance mechanisms and bacterial hosts. VanY and vanT, components of the vanG cluster, contribute to glycopeptide resistance via D,D-carboxypeptidase and serine racemase activities, respectively, and are frequently present in Enterococcus faecalis and gut anaerobes. The qacG gene encodes an SMR family efflux pump, frequently found on plasmids in Staphylococcus

species. Furthermore, FosB, which inactivates fosfomycin through thiol transfer, and BCIII, a beta-lactamase that targets cephalosporins, have been documented de *S. aureus*, *E. faecium*, and *Bacillus cereus*. vanW, commonly located inside vanA/vanB clusters, serves a regulatory or structural function in vancomycin resistance in *E. faecium* (Table 4.15).

TABLE 4.15: List of resistance gene predicted from card along with their mechanism of resistance

Gene	**Resistance Mechanism**	**Associated Cluster**	**Clinical Pathogens**
vanY (vanG)	d, d-Carboxypeptidase	vanG	*E. faecalis, Ruminococcus*
qacG	Efflux pump (SMR family)	Plasmid-borne	*Staphylococcus spp.*
FosB	Fosfomycin inactivation (thiol transfer)	Chromosomal/plasmid	*S. aureus, E. faecium*
BCIII	beta-lactamase cephalosporin antibiotics	BcI, BcII, and BcIII	*Bacillus cereus*
vanT (vanG)	Serine racemase	vanG	*E. faecalis, gut anaerobes*
vanW (vanI)	Regulatory/structural role	vanA/vanB	*E. faecium* (VRE)

4.6.3 Prediction and Annotation of Prophage Sequences

PHASTEST (PHAge Search Tool with Enhanced Sequence Translation) is a bioinformatics tool developed for the swift discovery, annotation, and visualization of prophage sequences in bacterial genomes and plasmids. It is essential for examining phage-host interactions, microbial evolution, and bacterial disease.

Phage research verified the existence of five unique prophage areas in the *B. licheniformis* genome, comprising roughly 175 phage-related genes, including structural and regulatory elements (Figure 4.17). These regions encompass distinctive phage modules, including those that code for tail fibers, head proteins, portal complexes, and terminases, hence substantiating their classification as integrated prophages.

The circular genomic map depicts the whole DNA sequence of *Bacillus licheniformis*, emphasizing both essential bacterial genes and incorporated viral (phage) genes. The outer orange rings signify the bacterial coding regions, whereas the inner colored tracks indicate phage-related genes, including capsid proteins, tail components, integrases, and other elements vital for phage construction and function. These viral gene insertions signify historical occurrences of horizontal gene transfer, wherein phage DNA was incorporated into the bacterial chromosome.

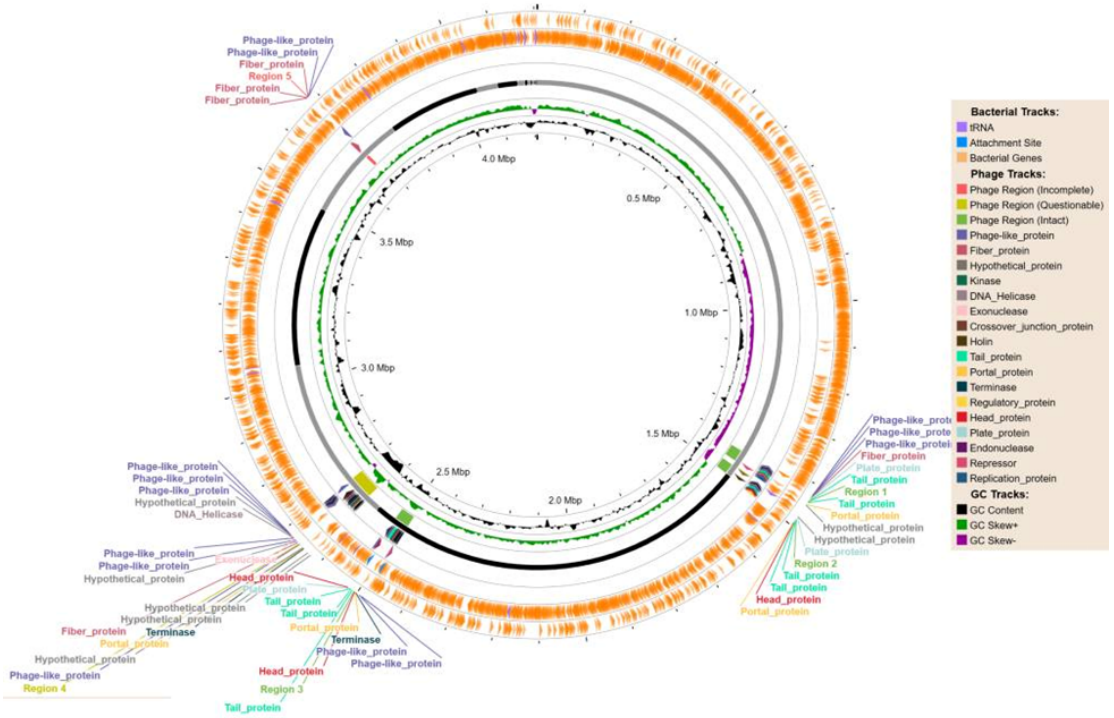


FIGURE 4.17: Circular visualization of prophage sequences in circular genome of *B. Lecheniformis*

The prophage regions found in the *Bacillus licheniformis* genome (Figure 4.18) encompass numerous loci, each expressing phage-related proteins including tail, head, portal, fiber, plate, and terminase components. Numerous genes in these regions are classified as phage-like or hypothetical proteins, signifying either recognized or uncharacterized phage components. The existence of holins and DNA helicases indicates the functionality or remains of these prophages. These components collectively indicate possible functions in genomic plasticity, horizontal

gene transfer, and increased virulence or adaptability of the organism. The linear genome scale at the figure's base offers an exact genomic context for these characteristics.

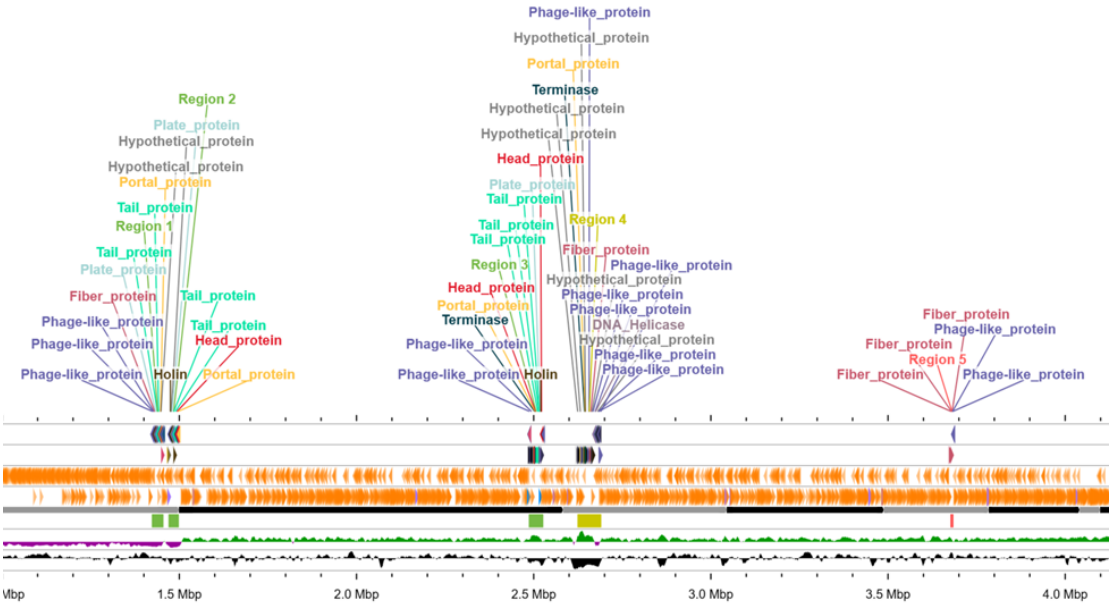


FIGURE 4.18: Linear visualization of prophage sequences in the genome of *B. lecheniformis*

Prophages in *Bacillus licheniformis* are integrated viral entities that influence host biology, stress adaptation, and industrial process results. A specific element, the PBSX-like prophage (BLi_Pp2), bears resemblance to the faulty PBSX prophage identified in *B. subtilis*. While it does not generate infectious viral particles, it activates under stress conditions, such as mitomycin C induction. During activation, it produces phage-like particles that enclose arbitrary segments of host DNA instead of viral genomes. This technique may promote horizontal gene transfer among microbial populations or trigger bacterial cell lysis under stress, thereby affecting the stability and efficiency of industrial fermentation operations involving *B. licheniformis* [53].

Genes from lytic and temperate phages, including those linked to the ϑ phage and the LP52 family, are crucial to the biology of *Bacillus licheniformis*, especially in industrial applications. The ϑ phage is recognized for inducing lysis in industrial strains, particularly those utilized for bacitracin manufacturing. This

phage can exist in both lytic and temperate states. In its temperate condition, certain variations like the LP52 prophage can merge into the host genome. Upon induction-typically initiated by stresses such as mitomycin C-the LP52 prophage activates and generates infectious phage particles that can form plaques on susceptible *B. licheniformis* strains. Phage-induced lysis of production strains poses a significant challenge in commercial fermentation, frequently resulting in process failure and diminished product yields [54].

4.6.4 Prediction of Virulence Factors

For the prediction of virulence factors, the fIDBAC database was utilized. The analysis identified five virulence-associated genes along with their corresponding functions and related *Bacillus licheniformis* strains. (Table 4.16)

TABLE 4.16: table of predicted virulence factors identified

Gene	**Virulence Factor**	**Class**	**Description**	**Strain**
hlyIII	Hemolysin III	Toxin	Putative membrane protein, Hemolysin III homolog [Hemolysin III (CVF560)]	*Bacillus licheniformis* ATCC 14580
capD	Polyglutamic acid capsule	Immune evasion	Ggt [Polyglutamic acid capsule (CVF566)]	*Bacillus licheniformis* DSM 13 (ATCC 14580)
capB	Polyglutamic acid capsule	Immune evasion	YwsC [Polyglutamic acid capsule (CVF566)]	*Bacillus licheniformis* DSM 13 (ATCC 14580)
capC	Polyglutamic acid capsule	Immune evasion	Capsule biosynthesis protein [Polyglutamic acid capsule (CVF566)]	*Bacillus licheniformis* ATCC 14580
capA	Polyglutamic acid capsule	Immune evasion	Poly-gamma-glutamate synthesis protein [Polyglutamic acid capsule (CVF566)]	*Bacillus licheniformis* ATCC 14580

4.6.5 Prediction of Ortholog

The study of ortholog genes in different species that evolved from a common ancestor through speciation is crucial for understanding bacterial genomics, evolution, and function.The eggNOG-mapper analysis of 16 ORFs with 137 orthologs revealed a complex pattern of orthologous relationships, primarily among Firmicutes species including *Bacillus*, *Paenibacillus*, and *Lactobacillus*. The results highlight several important biological insights.

First, conserved core genes were identified, such as ORFs 12_0 and 13_0 which show one-to-one orthology with *Bacillus licheniformis* (BL00759) and *B. pumilus* (BPUM_2770), strongly suggesting these encode essential housekeeping functions like peptidases or metabolic enzymes.

Second, evidence of gene family expansions was found through many-to-many relationships (e.g., 15_0's association with multiple *B. licheniformis* genes BL00225-BL03686) and one-to-many connections (e.g., 12_0's matches to three paralogs in *B. xiamenensis*), indicating evolutionary duplications that may relate to stress response or niche adaptation.

A particularly interesting finding was ORF 14_0, which showed unexpected homology with plant genes from *Medicago truncatula* (AES84212), Gossypium, and Solanum, potentially representing a case of horizontal gene transfer that warrants further investigation.

The analysis also revealed broad taxonomic distribution for some ORFs, such as 16_1's connections to anaerobic species like *Clostridium* and *Heliobacterium*, suggesting metabolic versatility. These results are summarized in a table highlighting key ORFs, their closest orthologs, relationship types, and potential functions, providing a concise overview of the most significant findings from this orthology analysis.

The diverse patterns observed - from highly conserved genes to potential HGT events - demonstrate the value of this approach for both functional annotation and evolutionary studies of bacterial genomes.(table 4.17)

TABLE 4.17: List of ortholog genes identified using eggnog mapper

Query	**Seed Ortholog (Species)**	**Key Orthology Types**	**Potential Function**	**Notable Observations**
12_0	BL00759 (*B. licheniformis*)	one2one, one2many, many2many	Peptidase (UniProt)	Broad conservation in *Firmicutes*; paralogs in *B. xiamenensi*.
13_0	BL00759 (*B. licheniformis*)	one2one, many2one	Peptidase (UniProt)	Similar to 12_0; many2one in *Lactobacillus*/*Staphylococcus*.
14_0	AES84212 (*Medicago truncatula*)	many2one (plants)	Unknown	Possible HGT or contamination.
15_0	BL00083 (*B. licheniformis*)	many2many (*BL00225-BL03686*)	Transporter	Gene family expansion in *Bacillus*.
16_1	HM1_3149 (*Heliobacterium*)	many2many (*Clostridium*)	Redox enzyme	Anaerobic metabolism link.
22_0	BL05294 (*B. licheniformis*)	many2many (*Bacillus*/*Paenibacillus*)	Hydrolase	Complex orthology with soil bacteria.

Chapter 5

Discussion

Halophilic microorganisms flourish in habitats characterized by elevated salinity, including sun saltpans, saline lakes, and coastal marshes. This encompasses both bacteria and archaea that exhibit distinctive physiological adaptations enabling them to endure excessive osmotic pressure, elevated salinity, and various environmental challenges, including temperature and pH fluctuations. Halotolerant and halophilic species generally thrive in environments with over 5% salinity and may endure various stressors concurrently. Recent years have seen the discovery of several novel halophilic taxa, whose metabolic variety renders them particularly valuable for scientific and industrial investigation. Marine bacteria, like *Bacillus pumilus* and other halophiles, exhibit resilience and adaptability, positioning them as viable candidates for diverse uses.

Halophilic bacteria are gaining recognition for their biotechnological potential. They can generate bioactive molecules with antibacterial characteristics and provide sustainable alternatives to synthetic pesticides and pharmaceuticals. These organisms produce suitable solutes and secondary metabolites, including lipopeptides, polypeptides, polyketides, isocoumarins, and macrolactins, which demonstrate a diverse array of biological activity. A significant focus of halophile research is enzyme synthesis. Enzymes such as amylases and proteases sourced from halophiles exhibit remarkable stability and functionality under harsh circumstances, rendering them appropriate for industrial applications in food processing,

pharmaceuticals, detergents, and textiles. The genus *Bacillus* is well recognized for its potent amylase production capability. Notwithstanding the potential, the isolation and preservation of halophilic microorganisms continue to pose significant challenges. A mere proportion of their metabolic and enzymatic potential has been investigated, signifying a substantial, unexploited resource for future inquiry.

Halophilic enzymes, commonly known as extremozymes, have exceptional functional stability in settings characterized by elevated salinity, high temperatures, and low moisture levels. These enzymes are highly valued in industry for their strong performance under extreme processing conditions. Outer membrane vesicles (OMVs) seen in halophilic marine bacteria are predominantly composed of membrane-associated proteins and may function in stress response and protein transport. Various strains of *Bacillus*, encompassing thermophilic and haloalkaliphilic species, are recognized for their production of alkaline proteases and cellulases. These strains can function effectively in situations with elevated salt and high pH levels. Additional significant halophilic and halotolerant species, including *Halomonas heilongjiangensis, Halosiccatus urmianus, Lentibacillus kimchi, Sporohalobacter salinus, Salinicoccus sediminis, Halomonas hydrothermalis, Planococcus maritimus, Virgibacillus dokdonensis,* and *Bacillus aquimaris*, enhance the variety of enzyme-producing organisms appropriate for commercial utilization.

Bacillus licheniformis is a Gram-positive, spore-forming, saprophytic bacteria that naturally resides in soil and on plant surfaces. It is renowned for its industrial significance, especially for its capacity to generate a diverse range of extracellular enzymes, antibiotics, and other biochemicals. The enzymes encompass α-amylase, penicillinase, pentosanase, cycloglucosyltransferase, β-mannanase, and pectinolytic enzymes, utilized across many sectors including detergent production, starch hydrolysis, and the textile and paper industries. Moreover, some strains of B. licheniformis are utilized for the production of peptide antibiotics, specialized compounds, and poly-γ-glutamic acid.

The *B. licheniformis* genome has a circular chromosome of approximately 4.22 Mb, encoding more than 4,200 protein-coding genes, in addition to several tRNAs and rRNA operons. The mean gene length aligns with other *Bacillus* species,

indicating genomic stability and functional conservation. Comparative research reveals significant collinearity and synteny with related species, including *Bacillus subtilis* and *Bacillus halodurans*, with a substantial proportion of coding sequences exhibiting orthologs. Significant distinctions encompass the lack of specific major secondary metabolite operons found in *B. subtilis*, such the polyketide synthase and plipastatin synthase clusters. These discrepancies indicate functional heterogeneity in metabolic capabilities and ecological functions.

The finding of conserved genes via one-to-one orthologs, exemplified by BL00759 in *Bacillus licheniformis*, strongly indicates that these genes fulfill critical biological activities preserved throughout evolutionary history. These highly conserved regions likely encode essential cellular functions such as DNA replication, protein synthesis, or fundamental metabolic pathways crucial for bacterial survival [55]. Comparative genomic studies have shown that important genes exhibiting one-to-one orthology links generally evolve under robust purifying selection, preserving functional stability across species [55, 56].

Conversely, larger gene families exhibiting many-to-many orthologous links, as seen in *s*, suggest recurrent gene duplication events that may facilitate adaptive evolution. Such expansions can facilitate niche specialization, diversify stress responses, or enhance metabolic flexibility [57]. Investigations into *Bacillus* genomes demonstrate that lineage-specific gene duplications frequently result in functional divergence across paralogs, with certain copies adopting novel roles while others preserve ancestral functionality [55, 56].

The identification of anomalies, such as plant-like homologous sequences in ORF 14_0, constitutes particularly fascinating instances that may signify horizontal gene transfer (HGT) events among disparate taxonomic groups. Research has recorded infrequent yet notable instances of horizontal gene transfer (HGT) from plants to bacteria, especially among plant-associated microorganisms, where such transfers may confer adaptive benefits in host relationships. The unforeseen evolutionary patterns necessitate focused examination using complementary methods such as phylogenetic reconstruction, synteny analysis, and experimental validation to ascertain their biological relevance and potential functional consequences in

bacterial genomes [58]. The unforeseen evolutionary patterns necessitate focused examination using complementary methods such as phylogenetic reconstruction, synteny analysis, and experimental validation to ascertain their biological relevance and potential functional consequences in bacterial genomes [58].

TABLE 5.1: Comparison of genome of different strains of *Bacillus lechiformis* with khewra salt mine strains

Feature	**Strain**	**ATCC 14580 (Ref)**	**WX-02 (Industrial)**	**DSM 13 (Type)**	**YB06 (Plant - Associated)**
Total Genes	4,289	4,208	4,320	4,190	4,325
Protein - Coding (CDS)	4,299	4,208	4,270	3,314	4,290
Proteins	4,467	-	-	-	-
Hypothetical Proteins	1,468 (34%)	1,318 (31%)	1,400 (33%)	-	∼1,450 (34%)
Orthologs*	137	3,400 (vs *B. subtilis*)*	-	-	-
tRNA Genes	82	72	72	86	81
ncRNA	30	24 (annotated)	-	1,316^	-
rRNA	3	7	7	8	6

*Orthologs typically shared with *B. subtilis* 168

^ DSM 13 includes 855 antisense RNAs and 461 small RNAs

The strain under investigation exhibits 2-4% more genes than the reference strains ATCC 14580 and DSM 13, and aligns with the industrial strain WX-02 (4,320 genes), surpassing the plant-associated strain YB06 (4,325 genes). The CDS count of 4,299 aligns with the potential for protein coding. (Table 5.1)

The elevated protein count (4,467 compared to 4,299 CDS) indicates the presence of alternative translation initiation sites, which are prevalent in *Bacillus*, as well as annotations related to post-translational modifications. The hypothetical protein proportion of 34% is characteristic of *Bacillus*. The tRNA count (82) exceeds that

of reference strains (72), potentially indicating adaptation to specific codon usage and occurrences of horizontal gene transfer.

The presence of low rRNA operons (3 compared to the typical 6-8) suggests potential assembly gaps in ribosomal regions and a strain-specific reduction, which is improbable. The reported 137 orthologs is insufficient, while approximately 3,400 were anticipated for *B. subtilis*. Potential explanations may include limited analytical parameters or highly diverse genetic material. The strain study's characteristics closely align with industrial strain WX-02, as elevated gene/CDS counts may indicate i) enhanced metabolic capabilities. ii) Prophage or plasmid composition iii) Augmented regulatory networks.

This comparison indicates that this strain falls within the typical range for *B. licheniformis*, with the ortholog count being the sole notable outlier for further examination. The increased tRNA levels and protein annotations may suggest specific metabolic adaptations.

Chapter 6

Conclusion

This research offers an extensive genomic and functional analysis of a halophilic strain of *Bacillus licheniformis* obtained from the hypersaline ponds of the Khewra salt mines in Pakistan. The severe environmental circumstances of this region—marked by high salinity, restricted light, and minimal human disturbance—have prompted the development of distinct microbial adaptations, rendering this bacterium a compelling focus for biotechnological and ecological investigation.

Key findings reveal genetic adaptations to hypersaline environments. The de novo constructed genome (1.14 Mb, GC content 46.1%) exhibited numerous critical genomic characteristics that promote survival in high-salinity environments: Osmoregulatory genes that encode suitable solute transporters, such as those involved in the production routes of glycine betaine, proline, and ectoine, are responsible for maintaining cellular osmotic equilibrium. Ion homeostasis mechanisms, including as Na+/H+ antiporters and K+ uptake systems, are essential for sustaining cytoplasmic ion balance. Stress-response genes, including as heat shock proteins and DNA repair enzymes, that improve survival in variable environmental conditions. These findings correspond with prior research on halophiles while offering new insights into the particular adaptations of *B. licheniformis* in a relatively unexplored saline ecosystem.

CRISPR-Cas System and Its Defense Mechanisms The genome contains a Type I-A CRISPR-Cas system, featuring a Cas8a1a component, which is essential for

DNA target recognition and phage defense. The existence of CRISPR arrays indicates active protection against viral predation, an essential survival characteristic in microbial communities. The limited quantity of spacers (just one detected) may suggest either minimal historical exposure to phages or a highly effective defense mechanism that inhibits phage integration. This discovery has ramifications for industrial applications, as CRISPR-based genome editing may be utilized to improve strain resilience in biomanufacturing. Antimicrobial Resistance and Biotechnological Issues The CARD-RGI analysis identified six genes associated with antibiotic resistance, including: vanY (vanG cluster) – Linked to glycopeptide resistance (e.g., vancomycin). FosB — Imparts resistance to fosfomycin through enzymatic inactivation. qacG – Encodes a small multidrug resistance (SMR) efflux pump, which may enhance tolerance to disinfectants. Although these genes may confer a competitive advantage in natural ecosystems, their existence raises apprehensions over horizontal gene transfer to pathogenic strains, especially in contexts where industrial or agricultural runoff could propagate resistance determinants.

Detection of Prophages and Associated Industrial Risks The PHASTEST study identified five prophage areas containing 175 related genes, which include: Temperate phage sequences associated with the PBSX-like family, recognized for stress-induced pseudo-lysogeny. Genes associated with the lytic cycle (e.g., holins, endolysins) that may induce cell lysis in response to fermentation stress. This presents a considerable risk for industrial applications, as prophage induction (e.g., due to nutrient constraint or chemical stressors) may result in fermentation failure on a massive scale. Cultures of *B. licheniformis* utilized for the synthesis of enzymes or antibiotics (e.g., bacitracin).

6.1 Future Recommendations

This study enhances our comprehension of halophilic *B. licheniformis* while emphasizing its dual function as a biotechnological resource and a potential repository of resistance genes. This strain could be enhanced for sustainable industrial uses by utilizing its extremophilic characteristics while minimizing dangers such

as prophage lysis and the spread of antimicrobial resistance (AMR). Subsequent investigations should concentrate on strain enhancement and ecological effect evaluations to guarantee safe and efficient application.

1. Functional Validation of Stress Knockout studies should be conducted to validate the function of osmoregulatory and CRISPR-associated genes in halotolerance. Additionally, proteomic research must be conducted under salt stress to identify essential survival proteins.

2. Phage Resistance Engineering can be achieved using CRISPR-mediated silencing of prophage sequences to improve strain stability in industrial fermenters. Integrating this technique allows for the optimization of *B. licheniformis* for resilient, phage-resistant biomanufacturing, hence ensuring consistent production of enzymes, antibiotics, and other valuable metabolites in industrial environments.

3. Evaluation of Novel Halophilic Enzymes The very saline conditions of the Khewra salt mines have likely facilitated the evolution of distinctive, salt-tolerant enzymes in *B. licheniformis*, which possess considerable industrial potential. High-throughput functional metagenomics and activity-based screening can be utilized to discover novel halophilic proteases, amylases, lipases, and cellulases that exhibit stability in high-salinity environments and extreme processing conditions (e.g., detergents, organic solvents). Promising candidates can be defined by appropriate pH, temperature stability, and salt activation, with possible uses in food processing (fermentation, meat tenderization), bio-detergents, and leather treatment. Directed evolution can further augment their catalytic efficiency and stability for commercial applications.

4. Metabolic Engineering to Produce Compatible Solutes *B. licheniformis* inherently synthesizes osmoprotectants such as ectoine, hydroxyectoine, and betaine, which possess significant uses in cosmetics (anti-aging, moisturizers) and medicines (protein stabilizers, UV protection). Metabolic engineering techniques can enhance biosynthesis by:

- overexpressing rate-limiting enzymes (e.g., ectABC operon for ectoine),
- inhibiting competitive pathways (e.g., glycine betaine catabolism)
- augmenting precursor availability (aspartate semialdehyde). Optimizing fermentation (e.g., inducing osmotic stress, employing fed-batch culture) can enhance yields. Innovations in downstream processing, including cell-free extraction and secretion engineering, will enable economical manufacturing for commercial applications.

5. Monitoring Environmental and Antimicrobial Resistance (AMR)

   - Monitoring the transmission of resistance genes in saline ecosystems is feasible, as *B. licheniformis* contains clinically significant resistance determinants (e.g., vanY, FosB, qacG) that may transfer to pathogens via mobile genetic elements (plasmids, transposons) in salty environments. Monitoring can utilize metagenomic sequencing of soil and water samples to trace the distribution patterns of antibiotic resistance genes and identify high-risk reservoirs.
   - The evaluation of horizontal gene transfer risks to clinically significant microorganisms can also be conducted. This technique can elucidate whether severe salinity serves as an impediment or enabler of horizontal gene transfer, hence guiding measures to minimize the dissemination of resistance from environmental to clinical contexts. This surveillance is essential for One Health initiatives, as the industrial application of halophiles may expedite the dissemination of antimicrobial resistance genes via wastewater discharge or agricultural runoff.

Bibliography

[1] S. A. Hussain, M. Asif, M. Bilal, M. Imran, M. Arshad, and M. Nadeem, "An overview of pakistan rock salt resources and their chemical characterization," *Pakistan Journal of Scientific & Industrial Research Series A: Physical Sciences*, vol. 64, no. 2, pp. 137–148, Jul 2021.

[2] K. Ahmad, M. Rafique, M. Sohail, A. Jabbar, M. Tariq, and S. U. Rahman, "Dosimetric properties of thermoluminescent nacl pellets from khewra salt mines, pakistan," *Luminescence*, vol. 37, no. 10, pp. 1701–1709, Oct 2022.

[3] H. Ullah, M. Hanif, A. Majid, M. Bilal, S. A. Turab, and M. Nadeem, "Characteristics and prediction of paleo-environment of eocene jatta gypsum, kohat basin, pakistan," *Arabian Journal of Geosciences*, vol. 16, no. 6, pp. 1–19, May 2023.

[4] L. J. Thompson, "Sodium chloride (salt)," in *Veterinary Toxicology*, Jan 2025, pp. 483–486.

[5] Y. Malik, M. Sajid, A. Rehman, F. Ahmad, and M. Imran, "Bacterial diversity at himalayan pink salt extraction site," *Biology (Basel)*, vol. 14, no. 3, Mar 2025.

[6] B. Rehman, S. Mehmood, A. Qadir, M. Nadeem, and F. Ahmad, "Salt tolerance potential of native plant species and halophilic bacteria from bahadur khel and khewra, pakistan," *Kuwait Journal of Science*, vol. 52, no. 3, Jul 2025.

[7] L. M. Cycil, S. DasSarma, W. Pecher, R. McDonald, M. AbdulSalam, and F. Hasan, "Metagenomic insights into the diversity of halophilic microorganisms indigenous to the karak salt mine, pakistan," *Frontiers in Microbiology*, vol. 11, p. 1567, Jul 2020.

[8] F. Nadal-Molero, A. Campos-Lopez, J. Tur-Moya, and A. B. Martin-Cuadrado, "Microbial community on industrial salty bovine hides: From the slaughterhouse to the salting," *Systematic and Applied Microbiology*, vol. 46, no. 4, p. 126421, Jul 2023.

[9] B. Khalid, "Effect of temperature and humidity on salt mine environment," *Pakistan Journal of Meteorology*, vol. 7, no. 13, 2010.

[10] B. Naik, R. Kumar, S. Kumar, P. Kumar, and A. Kumar, "Production, characterization, and application of novel fungal pullulanase for fruit juice processing," *International Journal of Biological Macromolecules*, vol. 248, Sep 2023.

[11] S. W. Noor, S. Arshad, S. M. Abdullah, B. K. Kayani, and S. Fazal, "Comprehensive analysis of bacterial diversity in the salt walls of khewra salt mine using 16s rrna gene sequencing," Jun 2024.

[12] A. Saini, A. Kumar, G. Singh, and S. K. Giri, "Survival strategies and stress adaptations in halophilic archaebacteria," in *ACS Symposium Series*, Jan 2023, vol. 1434, pp. 1–21.

[13] S. Saleem, A. Iqbal, F. Ahmed, and M. Ahmad, "Phytobeneficial and salt stress mitigating efficacy of iaa producing salt tolerant strains in gossypium hirsutum," *Saudi Journal of Biological Sciences*, vol. 28, no. 9, pp. 5317–5324, May 2021.

[14] U. Haroon, M. Farooq, A. Rehman, M. Nadeem, S. Mehmood, and F. Ahmad, "Isolation of halotolerant bacteria from rhizosphere of khewra salt mine halophytes and their application to induce salt tolerance in wheat," *Geomicrobiology Journal*, vol. 38, no. 9, pp. 768–775, 2021.

[15] M. A. Amoozegar, A. Safarpour, K. A. Noghabi, T. Bakhtiary, and A. Ventosa, "Halophiles and their vast potential in biofuel production," *Frontiers in Microbiology*, vol. 10, no. AUG, p. 1895, Aug 2019.

[16] M. M. Naik, M. Imran, D. C. Vaigankar, S. Y. Mujawar, A. D. Malik, and S. K. Gaonkar, "Genome guided bioprospecting of extremely halophilic haloferax sp. as1 for cazymes, bioremediation and study metabolic versatility," *Proceedings of the Indian National Science Academy*, vol. 91, no. 2, pp. 297–305, Jun 2021.

[17] E. D. Salvo, A. Alfonzo, N. Francesca, R. Gaglio, L. Settanni, and G. Moschetti, "Gourmet table salts: The mineral composition showdown," 2023.

[18] A. C. Navarro, K. Gallagher, S. Griffin, C. L. Leydon, I. J. Perry, and J. M. Harrington, "Systematic review on the impact of salt-reduction initiatives by socioeconomic position to address health inequalities in adult populations," *Nutrition Reviews*, vol. 83, no. 3, pp. e1090–e1100, Mar 2025.

[19] T. S. Kenao, J. C. Sossa, M. N. Paraiso, M. Belo, and G. P. Sopoh, "Dietary sodium and potassium intakes and salt reduction strategies: Systematic review in africa," *International Archives of Public Health and Community Medicine*, vol. 6, p. 82, 2022.

[20] R. Crisan-Dabija, I. G. Sandu, I. V. Popa, D. V. Scripcariu, A. Covic, and A. Burlacu, "Halotherapy—an ancient natural ally in the management of asthma: A comprehensive review," *Healthcare*, vol. 9, no. 11, p. 1604, Nov 2021.

[21] S. W. Noor, S. Arshad, S. M. Abdullah, B. K. Kayani, and S. Fazal, "16s rrna gene sequencing reveals bacterial diversity in khewra salt mine walls," *Access Microbiology*, vol. 6, no. 12, p. 000869.v4, Dec 2024.

[22] C. Zhang, L. Wang, Y. Zhang, H. Li, and J. Yang, "Halotherapy relieves chronic obstructive pulmonary disease by alleviating nlrp3 inflammasome-mediated pyroptosis," *Annals of Translational Medicine*, vol. 10, no. 23, p. 1279, Dec 2022.

[23] D. Barber, Y. Malyshev, F. Oluyadi, A. Andreev, and S. Sahni, "Halotherapy for chronic respiratory disorders: From the cave to the clinical," *Alternative Therapies*, 2025, [E-Pub Ahead of Print].

[24] M. M. Rahman, M. A. Hoque, M. A. Kashem, M. S. Islam, and M. A. Kader, "Adaptive mechanisms of halophytes and their potential in improving salinity tolerance in plants," *International Journal of Molecular Sciences*, vol. 22, no. 19, p. 10733, Oct 2021.

[25] S. Gairola, A. Hameed, A. Rasheed, A. Alketbi, M. Aljasmi, and A. El-Keblawy, "Seed germination and salinity tolerance of habitat-indifferent halophytes as associated with geographical distribution," *Seed Science and Technology*, vol. 50, no. 2, pp. 125–140, Sep 2022.

[26] M.-N. Grigore, "Definition and classification of halophytes as an ecological group of plants," in *Handbook of Halophytes: From Molecules to Ecosystems towards Biosaline Agriculture*, May 2021, pp. 3–50.

[27] Q. Zhou, D. Lyu, W. Li, Y. Wen, and Z. Wang, "Effects of irrigation amount and salinity levels on maize (zea mays l.) growth, water productivity and carbon emissions in arid region of northwest china," *Agronomy*, vol. 14, no. 11, p. 2656, Nov 2024.

[28] T. Guo, X. Huang, K. Feng, and X. Mao, "Impact of deficit drip irrigation with brackish water on soil water–salt dynamics and maize yield in film-mulched fields," *Agronomy*, vol. 15, no. 2, p. 379, Jan 2025.

[29] H. Yuan, X. Ma, Y. Li, J. He, and S. Li, "Saline water irrigation changed the stability of soil aggregates and crop yields in a winter wheat–summer maize rotation system," *Agronomy*, vol. 14, no. 11, p. 2564, Oct 2024.

[30] S. Wang, S. Ge, W. Mai, and C. Tian, "Nitrogen promotes the salt-gathering capacity of suaeda salsa and alleviates nutrient competition in the intercropping of suaeda salsa/zea mays l." *International Journal of Molecular Sciences*, vol. 23, no. 24, p. 15495, Dec 2022.

[31] M. Gupta, "Halophilic cyanobacteria and microalgae: Role in environmental clean-up and value-added products," in *Physiology of Halophytes*, Mar 2025, pp. 213–231.

[32] A. Katamadze, O. Vergara-Díaz, E. Uberegui, A. Yoldi-Achalandabaso, J. L. Araus, and R. Vicente, "Evolution of wheat architecture, physiology, and metabolism during domestication and further cultivation: Lessons for crop improvement," *The Crop Journal*, vol. 11, no. 4, pp. 1080–1096, Aug 2023.

[33] R. J. M. Ventayen, R. Basuel, R. Resultay, E. Galas, and D. Buted, "Salt industry in the philippines: Review on trends and intervention," *SSRN Electronic Journal*, Mar 2023.

[34] Y. Arif, P. Singh, H. Siddiqui, A. Bajguz, and S. Hayat, "Salinity induced physiological and biochemical changes in plants: An omic approach towards salt stress tolerance," *Plant Physiology and Biochemistry*, vol. 156, pp. 64–77, Nov 2020.

[35] S. DasSarma and P. DasSarma, "Halophiles and their enzymes: negativity put to good use," *Current Opinion in Microbiology*, vol. 25, pp. 120–126, Jun 2015.

[36] X. Zhang, W. Zhao, Y. Zhang, and V. Jegatheesan, "A review of resource recovery from seawater desalination brine," *Reviews in Environmental Science and Bio/Technology*, vol. 20, no. 2, pp. 333–361, Mar 2021.

[37] N. Masood, K. A. Hudson-Edwards, and A. Farooqi, "Groundwater nitrate and fluoride profiles, sources and health risk assessment in the coal mining areas of salt range, punjab pakistan," *Environmental Geochemistry and Health*, vol. 44, no. 3, pp. 715–728, Mar 2022.

[38] N. Masood, T. Zafar, K. A. Hudson-Edwards, and A. Farooqi, "Spatial distribution and health risk assessment of toxic metal(oid)s in soils of coal mining areas of the salt range, punjab, pakistan," *Physics and Chemistry of the Earth, Parts A/B/C*, vol. 134, p. 103566, Jun 2024.

[39] M. T. Hussain, Q. M. Sharif, and M. Hussain, "Chemical evaluation of major salt deposits of pakistan," *Journal of The Chemical Society of Pakistan*, vol. 29, no. 6, p. 569, 2011. [Online]. Available: https://jcsp.org.pk/home.aspx

[40] M. Waheed, H. Majeed, and S. S. Zahra, "Khewra pink rock salt as a registered geographical indication of pakistan," *Trends in Intellectual Property Research*, vol. 1, no. 2, pp. 1–5, Oct 2023.

[41] E. Brown, U. Dessai, S. Mcgarry, and P. Gerner-Smidt, "Use of whole-genome sequencing for food safety and public health in the united states," *Foodborne Pathogens and Disease*, vol. 16, no. 7, pp. 441–450, Jul 2019.

[42] M. Banar, D. Rokaya, R. Azizian, Z. Khurshid, and M. Banakar, "Oral bacteriophages: metagenomic clues to interpret microbiomes," *PeerJ*, vol. 12, p. e16947, 2024.

[43] C. A. Austin-Tse, H. L. Rehm, B. H. Funke, M. S. Lebo, A. O'Donnell-Luria, and M. W. Snyder, "Best practices for the interpretation and reporting of clinical whole genome sequencing," *NPJ Genomic Medicine*, vol. 7, no. 1, pp. 1–13, Dec 2022.

[44] "Geneious | bioinformatics software for sequence data analysis," 2025, accessed: Jul. 08, 2025. [Online]. Available: https://www.geneious.com/

[45] J. R. Grant, R. G. Beiko, and A. G. McArthur, "Proksee: In-depth characterization and visualization of bacterial genomes," *Nucleic Acids Research*, vol. 51, no. W1, pp. W484–W492, Jul 2023.

[46] O. Schwengers, L. Jelonek, M. A. Dieckmann, S. Beyvers, J. Blom, and A. Goesmann, "Bakta: Rapid and standardized annotation of bacterial genomes via alignment-free sequence identification," *Microbial Genomics*, vol. 7, no. 11, 2021.

[47] D. Arndt, J. R. Grant, A. Marcu, T. Sajed, A. Pon, Y. Liang, and D. S. Wishart, "Phaster: a better, faster version of the phast phage search tool," *Nucleic Acids Research*, vol. 44, no. W1, pp. W16–W21, Jul 2016.

[48] Y. Zhou, Y. Liang, K. H. Lynch, J. J. Dennis, and D. S. Wishart, "Phast: A fast phage search tool," *Nucleic Acids Research*, vol. 39, no. SUPPL. 2, Jul 2011.

[49] J. Huerta-Cepas, K. Forslund, L. P. Coelho, D. Szklarczyk, L. J. Jensen, C. von Mering, and P. Bork, "Fast genome-wide functional annotation through orthology assignment by eggnog-mapper," *Molecular Biology and Evolution*, vol. 34, no. 8, pp. 2115–2122, Aug 2017.

[50] Q. Liang, X. Li, F. Chen, Y. Liao, X. Wu, F. Luo, and X. Wang, "fidbac: A platform for fast bacterial genome identification and typing," *Frontiers in Microbiology*, vol. 12, p. 723577, Oct 2021.

[51] C. Hu, Y. Wang, X. Zhang, J. Yang, and L. Wang, "Cascade complex from type i-a crispr-cas system," Aug 2022.

[52] "The comprehensive antibiotic resistance database," 2025, accessed: Jul. 08, 2025. [Online]. Available: https://card.mcmaster.ca/analyze/rgi

[53] R. Hertel, D. Rodriguez, A. Goesmann, A. Goesmann, J. Blom, and C. Hertel, "Genome-based identification of active prophage regions by next generation sequencing in bacillus licheniformis dsm13," *PLoS One*, vol. 10, no. 3, p. e0120759, 2015.

[54] J. Doskočil, J. Forstová, H. Štorchová, and J. Meyer, "Genomic structure and evolution of bacillus licheniformis and lp52 phage family," in *Gene Manipulation and Expression*, 1985, pp. 3–21.

[55] H. Luo, F. Gao, and Y. Lin, "Evolutionary conservation analysis between the essential and nonessential genes in bacterial genomes," *Scientific Reports*, vol. 5, no. 1, pp. 1–8, Aug 2015.

[56] T. Gabaldón and E. V. Koonin, "Functional and evolutionary implications of gene orthology," *Nature Reviews Genetics*, vol. 14, no. 5, pp. 360–366, May 2013.

[57] M. W. Rey, S. Ramaiya, B. A. Nelson, S. D. Brody-Karpin, E. J. Zaretsky, M. Tang, A. L. de Leon, X. Xiang, V. Clifton, B. J. Drees, T. A. Graddis,

R. Gammie, B. D. Rasmussen, V. S. Bisgaard, S. F. Barrett, T. S. Brettin, S. N. Peterson, E. M. Ronning, W. C. Nierman, M. J. Feldblyum, and C. M. Fraser, "Complete genome sequence of the industrial bacterium bacillus licheniformis and comparisons with closely related bacillus species," *Genome Biology*, vol. 5, no. 10, 2004.

[58] G. Zhang, Y. Chen, Q. Li, J. Zhou, J. Li, and G. Du, "Growth-coupled evolution and high-throughput screening assisted rapid enhancement for amylase-producing bacillus licheniformis," *Bioresource Technology*, vol. 337, Oct 2021.

Appendix-I

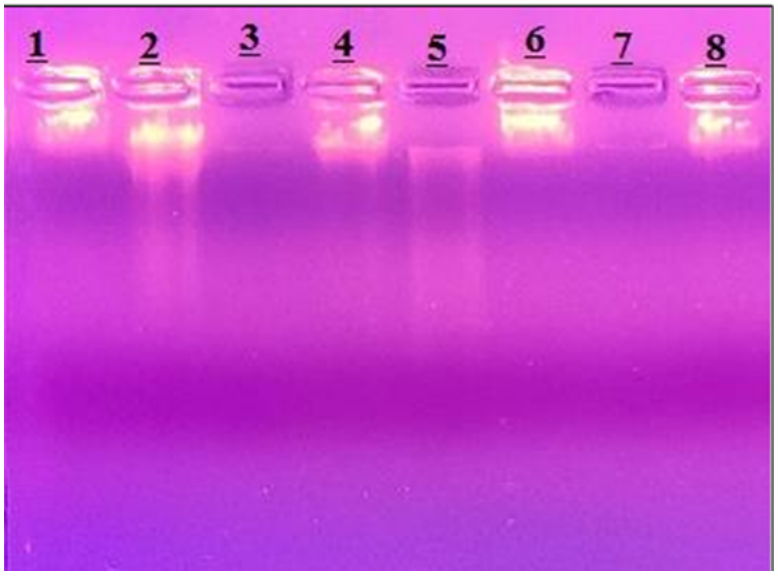


FIGURE 1: The DNA extraction results of isolates 5 (UE-S2-2), The slots 3 and 7 were left empty on purpose.

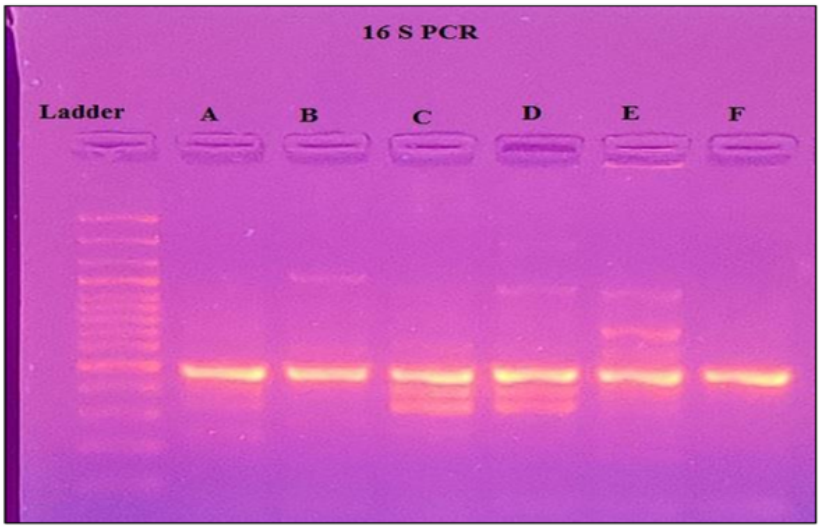


FIGURE 2: Figure II On the right side is the PCR fragment ladder and on the left is D (UE-S2-2) All the bands were of 465bp.

Appendix-II

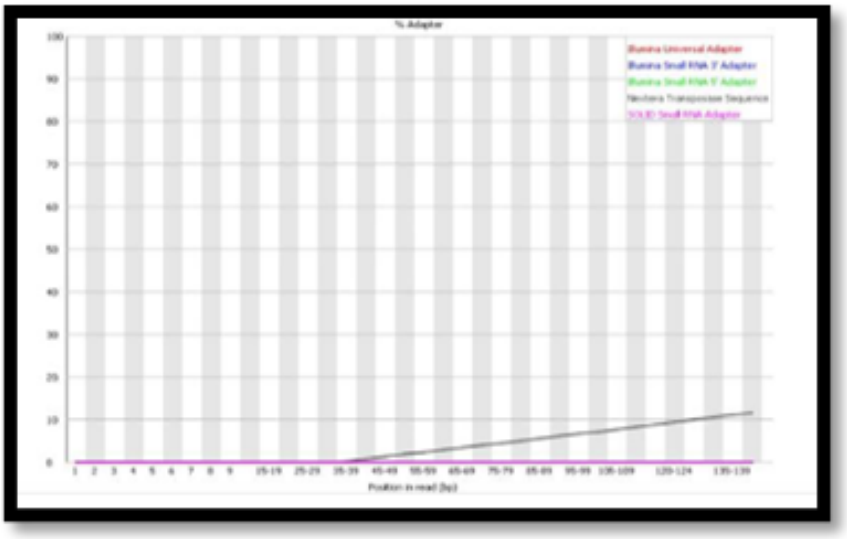


FIGURE 3: Adapter content was present.

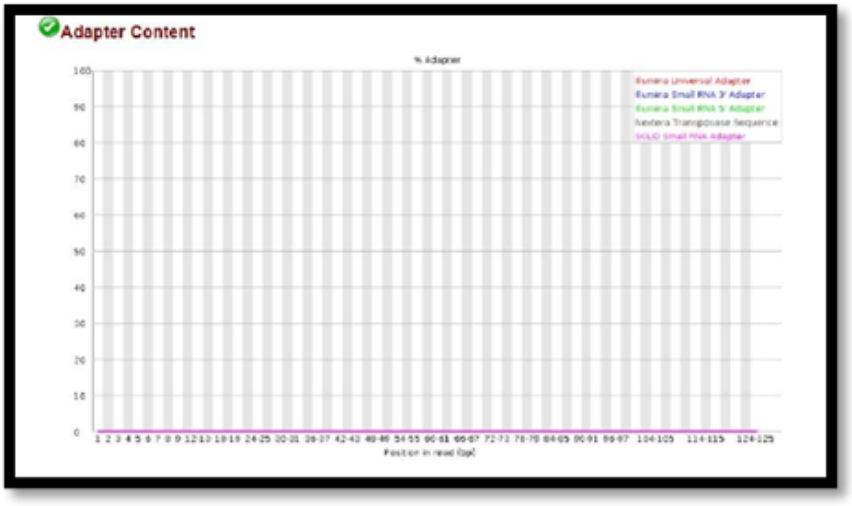


FIGURE 4: Remove the adapter content from the sequenced with the help of Fastp tool