UNIVERSITY OF EMBU

ALEX CHERUTICH KIPNYARGIS

PhD

2024

DIVERSITY AND FUNCTION OF POLYSACCHARIDE DEGRADING BACTERIA FROM LAKE MAGADI, KENYA

ALEX CHERUTICH KIPNYARGIS

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF DOCTOR OF PHILOSOPHY IN APPLIED MICROBIOLOGY OF THE UNIVERSITY OF EMBU

AUGUST, 2024

DECLARATION

This thesis is my original work and has not been presented elsewhere for a degree or any other award.

Signature…… Date…….

Alex Cherutich Kipnyargis
Department of Biological Sciences
B801/226/2018.

Supervisors:

This thesis has been submitted for examination with our approval as the University Supervisors.

Signature…… Date…….

Prof. Romano Mwirichia
Department of Biological Sciences
University of Embu.

Signature…… Date…….

Prof. Eucharia Kenya
Department of Biological Sciences
University of Embu.

Signature……..…… Date……

Dr. Fathiya Khamis
International Centre of Insect Physiology and Ecology
Nairobi, Kenya.

Signature……. Date……

Prof. Rolf Daniel
Department Genomics & Applied Microbiology
Institute of Microbiology and Genetics
Georg-August University, Goettingen, Germany.

DEDICATION

To my lovely family:

Priscillar Cheruiyot, Soila Chemalel, Denzel Kiptoo, and the entire Kipnyargis family.

Their support, love, patience, encouragement, and positive criticism during the enduring times that led to the completion of my Ph.D. work.

ACKNOWLEDGEMENT

I am grateful to a higher power that drove and guided my conscience throughout life and the entire doctoral work. I am deeply indebted to Priscillar Cheruiyot, my loving friend and life partner, who, together with our lovely children: Soila and Kiptoo, made life more bearable in undertaking research during the global COVID-19 pandemic. Immense gratitude goes to Ph.D. supervisors: Prof. Romano Mwirichia, Prof. Rolf Daniel, Prof. Eucharia Kenya, and Dr. Fathiya Khamis. Their relentless efforts in guiding, supervising, critical insights, and rightful guidance in the research phase are wholesomely appreciated. I especially thank Prof. Rolf Daniel for guiding my research work and providing financial and material support during my research stay in Germany. I am grateful to Dr. Robert Hertel, Dr. Birgit Pfeiffer, and Dr. Anja Poehlein for their guidance in optimizing biological methods during the research phase. Their careful attention to detail led to the delivery of the best result. I have become a better researcher because of you all. I am most thankful to the prestigious German Academic Exchange Service (DAAD) and the Management of the University of Embu for offering me a scholarship and a chance to undertake this research in Göttingen, Germany. Many thanks to the National Research Fund (NRF) Kenya and the British Ecological Society for funding for sampling and isolation of the bacterial isolates. I appreciate my fellow Ph.D. cohort: Dr. Ines Friedrich, Dr. Michael Eze, Dr. Tatiana Murillo, Dr. Stefani Diaz, Dr. Dirk Berkelman, Dr. Avril von Hoyningen-Huene, Mechthild Bömeke, Nicole Dörjer, the entire AG-Daniel Group in the Institute of Genomics and Applied Microbiology, and the warm people of the city of Goettingen, Germany. Their support and sessions on work-life balance will forever be cherished. A lot of gratitude to the staff members of the Department of Biological Sciences of the University of Embu for their suggestions and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS AND ACRONYMS

AA	Auxiliary Activity
AAS	Atomic Absorption Spectrometry
ACE	Abundance-based Coverage Estimator
ANI	Average Nucleotide Identity
CAT	Contiq Annotation Tool
CAZyme	Carbohydrate-Active enZymes
CDS	CoDing Sequence
CE	Carbohydrate Esterase
CBM	Carbohydrate-Binding Module
CMC	Carboxymethyl Cellulose
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
BLAST	Basic Local Alignment Search Tool
BRIG	BLAST Ring Image Generator
DAAD	German Academic Exchange Services
GBDP	Genome Blast Distance Phylogeny
DDDH	Digital DNA-DNA Hybridization
DMSO	Dimethyl sulfoxide
DNS	Dinitrosalicylic Acid
DNA	Deoxyribonucleic Acid
DOM	Dissolved Organic Matter
DTT	Dithiothreitol
EC	Enzyme Commission
dNTPs	Deoxynucleotide Triphosphates
EDTA	Ethylene Diamine Tetra-acetic Acid
EAI	Enzyme Activity Index
GGDC	Genome-to-Genome Distance Calculator
GH	Glycosyl Hydrolase
GRAVY	Grand Average of Hydropathicity
GT	Glycosyltransferase
HEC	hydroxyethylcellulose
I-TASSER	Iterative Threading ASSEmbPly Refinement
kDa	kiloDalton
KEGG	Kyoto Encyclopaedia for Genes and Genomes
KO	KEGG Orthology
LB	Luria Bertani
LCB	Lignocellulosic Biomass
MAFFT	Multiple Alignment using Fast Fourier Transform
MAG	Metagenome Assembled Genome
MEGA	Molecular Evolutionary Genetics Analysis
MCS	Multiple Cloning Sites
NCBI	National Center for Biotechnology Information
MIMAG	Minimum Information About Metagenome-Assembled Genome
OD	Optical Density
ORF	Open Reading Frames
OTU	Operational Taxonomic Unit
PCA	Principal Component Analysis
PDB	Protein Databank

PGAP	Prokaryotic Genome Annotation Pipeline
PCR	Polymerase Chain Reaction
PL	Polysaccharide Lyase
PMSF	Phenylmethylsulfonyl Fluoride
RDA	Redundancy Analysis
rRNA	Ribosomal Nucleic Acid
tRNA	Transfer Ribosomal Nucleic Acid
SOB	Sulfur-Oxidizing Bacteria
SRB	Sulfur-Reducing Bacteria
SDS	Sodium Dodecyl Sulfate
SDS-PAGE	Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis
SRA	Sequence Read Archive
TDS	Total Dissolved Solids
TSA	Tryptone Soy Agar
TSB	Tryptone Soy Broth
TYGS	Type (Strain) Genome Server
UV	Ultraviolet

ABSTRACT

Soda lakes are extreme environments characterized by high alkalinity and salt concentrations and moderate to high temperatures. Despite these conditions, soda lakes support distinctive microbial communities that are a potential source of novel biocatalysts for industrial and biotechnological applications. This study investigated the diversity, structure, and function of bacterial isolates recovered from the soda Lake Magadi, particularly isolates that could utilize polysaccharides. Isolation involved enrichment media supplemented with selected polymers and prepared using sterile lake water. The isolates were identified based on sequencing of the 16S rRNA gene region. A plate radial diffusion assay measured enzyme activity against cellulose, carboxymethylcellulose, xanthan, starch, pectin, and xylan. The bacterial isolates were tested for their ability to grow under various pH, salt, and temperature conditions. The results showed that the isolates were closely related to members of the genus *Salipaludibacillus*, *Halomonas*, *Alkalibacterium*, *Alkalihalophilus*, *Evansella*, *Shouchella*, *Halalkalibacterium*, *Halalkalibacter*, *Alkalihalobacterium*, and *Salinicoccus*. The highest enzyme activity was recorded among the isolates belonging to *Salipaludibacillus*, while the least activity was recorded in isolates belonging to *Halomonas*. High cell densities were recorded at pH 7–9, 40°C, and 5% (w/v) NaCl concentrations. Four isolates (LMS6, LMS18, LMS25, and LMS39) were selected for full genome sequencing. Average nucleotide identity (ANI) analysis revealed that LMS6, LMS18, LMS25, and LMS39 were new species of *Shouchella* sp., *Evansella* sp., *Salipaludibacillus* sp., and *Alkalihalobacterium* sp., respectively. Analysis of carbohydrate-active enzymes (CAZymes) revealed a higher number of genes for carbohydrate metabolism in the genomes of *Salipaludibacillus* sp. LMS25 and *Alkalihalobacterium* sp. LMS39. These genes encode for amylases, cellulase, pectinase, xylanase, and chitinase enzymes, further corroborating the findings of plate screening. Furthermore, the study examined the impact of water chemistry on the composition and structure of microbial communities over time. In this case, the 16S rRNA gene amplicons were sequenced using the Illumina MiSeq platform. Results revealed that the most abundant bacterial phyla were Proteobacteria, Cyanobacteria, Bacteroidetes, Actinobacteria, Firmicutes, Verrumicrobia, Deinococcus-Thermus, Spirochaetes, and Chloroflexi. Euryachaeota, Crenarchaeota, and Thaumarchaeota were representative of archaeal diversity. The dominant bacterial species were: *Euhalothece* sp. (10.3%), *Rhodobaca* sp. (9.6%), *Idiomarina* sp. (5.8%), *Rhodothermus* sp. (3.0%), *Roseinatronobacter* sp. (2.4%), *Nocardioides* sp. (2.3%), *Gracilimonas* sp. (2.2%), and *Halomonas* sp. (2%). On the other hand, the dominant archaeal species included *Halorubrum* sp. (18.3%), *Salinarchaeum* sp. (5.3%), and *Haloterrigena* sp. (1.3%). The composition of bacteria was higher than that of archaea, while their richness and diversity varied across the sampling seasons. The alpha diversity indices showed that high diversity was recorded in August, followed by September, June, and July in that order. The findings demonstrated that temperature, pH, P+, K+, NO_3^-, and total dissolved solids (TDS) contributed significantly to the diversity observed in the microbial community. Multivariate analysis revealed that samples were clustered based on salinity and alkalinity rather than the sampling site or season. Furthermore, this study constructed an artificial metagenome by mixing, in equal amounts, the total chromosomal DNA extracted from 40 bacterial isolates. The aim was to evaluate the total genes encoding for hydrolytic enzymes. To achieve this, Genomic libraries were created and sequenced using paired-end 2 x 250 bp runs on Illumina HiSeq 2500 apparatus. The results revealed a total count of 46,641 putative genes, with a proportion of 10% being those

encoding for enzymes involved in the hydrolysis of polysaccharides. Taxonomic assignment showed that more than 99% of the sequences were affiliated with Firmicutes and Proteobacteria. These enzymes included chitinases, cellulases, amylases, and xylanases. In addition, ten metagenome-assembled genomes (MAGs) of medium and high quality were observed. They were affiliated with the genera *Salipaludibacillus*, *Alkalihalophilus*, *Halalkalibacterium*, *Evansella*, *Salinicoccus*, *Alkalibacterium*, and *Halomonas*. Furthermore, a putative gene encoding for an endo-β-1,4-glucanase (LMP_42667) deduced from *Salipaludibacillus* was cloned into a pBAD18 vector and heterologously expressed in Top10 *E. coli* cells. The purified enzyme with 571 amino acids and a theoretical molecular weight of 65.7 kDa demonstrated activity at pH ranges of 4.0 to 8.0 and optimum enzyme activity at 50°C and pH 7.0. The enzyme also demonstrated higher enzyme-specific activity for xylan (9.8 ± 0.1 U/mg). Additionally, enzyme activity was reported against CMC, hydroxyethylcellulose (HEC), microcrystalline cellulose, lichenan, and β-glucan. It depicted stability in the presence of various metallic ions, protein inhibitors, and chemical reagents. Analysis of the structural model showed that the endoglucanase belongs to the glycosyl hydrolase 5 subfamily 4 (GH5_4) group of glycosyl hydrolases. In conclusion, the findings of this study demonstrate that Lake Magadi harbors a rich source of diverse polysaccharide hydrolyzing bacteria with a wide repertoire of genes encoding hydrolytic enzymes with industrial and biotechnological potential. Furthermore, the study reveals the role of salinity and alkalinity in determining the structure of microbial communities in Lake Magadi.

CHAPTER ONE

1.0 INTRODUCTION

1.1 Background of the study

Soda lakes are stable naturally occurring ecosystems defined by high salinity and alkalinity. They are widely distributed across the globe but are mainly found in terrestrial geologies such as arid and semi-arid, steppes and the floor of the East African Rift Valley (Antony et al., 2013; Sorokin et al., 2014). Generally, soda lakes can be of volcanic, tectonic depressions, or meteoric origin. Consequently, the resultant underlying rocks containing groundwater saturated with carbon dioxide (CO_2) increase the concentration of HCO_3^-/CO_3^{2-} relative to Ca^{2+}/Mg^{2+} (Paul et al., 2016; Borzenko & Shvartsev, 2019). Due to high evaporation rates in these areas, saturation and precipitation of the cations are rapidly achieved. Eventually, the concentration of Na^+, Cl^-, and HCO_3^-/CO_3^{2-} increases (Jones et al., 1998; Vavourakis et al., 2016; Rees et al., 2004; Berben et al., 2019). Examples of these lakes include Mono and Summer lakes in North America, Kulunda steppe in Russia, Lake Van in Turkey, Lake Qinghai Hu in China, Lakes Abiata and Shala in Ethiopia, Lake Natron in Tanzania, Lake Wadi Natrun in Egypt, and Lake Magadi (Grant and Sorokin, 2011). The chemical composition and geological origins of soda lakes are crucial in establishing their ecological dynamics and the adaptive mechanisms employed by organisms living in these ecosystems (Krenitz & Shagerl, 2016).

Soda lakes are known habitats for extreme microbial communities that drive nitrogen, carbon, and sulfur cycles, which maintain ecosystem functions such as energy and nutrient sources (Berben et al., 2019). Salt concentrations are among the determinants of the distribution and survival of different microorganisms. For example, slightly halophilic microbes thrive in lower salt conditions (0.2 – 0.5 M), moderately halophilic (0.5 – 2.5 M), and extreme halophiles (2.5 – 5.2 M) or halotolerant (Didari et al., 2020). They have developed unique adaptive features such as the secretion of compatible solutes for maintaining osmotic balance and removing salt from the cytoplasm to prevent aggregation of proteins through a ‘salting out’ strategy (Santos and Da Costa, 2002). Archaea, bacteria, and eukaryotes encompass oxygenic and

anoxygenic phototrophs, aerobic heterotrophs, fermenters, denitrifiers, sulfate reducers, and methanogens (Oren, 2002). The bacterial component is characterized by rich abundance, fast growth, and unique adaptation features. This makes them ideal models for understanding the link between community structures and abiotic stress factors (Lanzén et al., 2013; Vavourakis et al., 2018). Several studies have reported the presence of certain bacteria as the most abundant in soda: Proteobacteria, Actinobacteria, Firmicutes, Bacteroidetes, Cyanobacteria, Chloroflexi, Planctomycetes, Verrucomicrobia, and Tenericutes (Mwirichia et al., 2010; Surakasi et al., 2010; Kambura et al., 2016; Paul et al., 2016; D'haeseleer et al., 2017; Kiplimo et al., 2019; Guan et al., 2020).

Extremely halophilic bacteria are known producers of hydrolytic enzymes that are stable under high pH, salt concentration, and temperature (Martins et al., 2001; Di Donato et al., 2018; Oren, 2010; Sahay et al., 2012; Nyakeri et al., 2018; Arayes et al., 2021). For instance, laccases and esterase have been detected in *Bacillus* sp. strain WT, *Halomonas*, and *Alkalibacterium* sp. (Siroosi et al., 2016; Kiplimo et al., 2019). Additionally, amylases, lipases, proteases, cellulases, xylanases, and pectinases have been described in *Bacillus* sp. from hypersaline lake Chott El Jerid, of Tunisia (Karray et al., 2018; Ruginescu et al., 2020). In the presence of water, these biomolecules hydrolyze polysaccharides such as lignocellulose, pectin, xanthan, xylan, cellulose, and chitin into simple compounds (Babavalian et al., 2013).

Naturally, bacteria secrete extracellular polysaccharides which are crucial in nutrient cycling, biofilm production, and soil aggregation (Morris and Harding 2009; Nwodo et al., 2012). However, naturally occurring lignocellulosic biomass including materials such as agro wastes, dedicated energy crops, and wood is the most abundant and important renewable carbon source (Berlemont and Martiny, 2015; Lovegrove et al., 2017). Complete depolymerization of lignocellulosic biomass is synergistically carried out by bacterial cellulases, namely β-1, 4-endoglucanases (EC 3.2.1.4), β-1, 4-exoglucanases (cellobiohydrolase, EC 3.2.1.91), and β-1, 4-glucosidases (EC 3.2.1.21) (Lin et al., 2016). These enzymes can convert lignocellulosic materials into simple sugars. Different microorganisms ferment these sugars to generate biofuels such as bioethanol. Additional sectors that benefit from applications of glycoside

hydrolases include textile, bio-refineries, fermentation of food and feed, paper industries, agricultural, detergent manufacturing, and pharmaceutical and molecular biotechnology research (Janaki, 2017).

Situated at the southern end of the Kenyan Rift Valley, Lake Magadi is a soda lake with distinctive geochemical properties. The salt concentrations vary from less than 0.1 g/l at the hot springs to more than 300 g/l (w/v) at concentrated brines. Furthermore, the temperature varies from 34ºC in the open waters to 86ºC around the hot springs, while the pH ranges from 9 - 12.5 (Jones et al., 1977). During high evaporation, a crust forms on top of the brine and underneath the trona ($NaHCO_3$ Na_2CO_3-H_2O) (Baumgarte, 2003; Kambura et al., 2016). Previous studies on the lake have isolated bacteria from hot springs and the brine crust but some studies recommended further screening of different sites of the lake by both culture-dependent and culture-independent techniques (Duckworth et al., 2000; Muruga & Anyango, 2013; Kambura et al., 2016; Nyakeri et al., 2018). Therefore, this study aimed at combining both culture-dependent and metagenomic approaches to characterize polysaccharides utilizing bacteria from different sites around the lake. The isolates depicting activity against the polysaccharides were further characterized using genomic techniques to identify genes responsible for the observed enzymatic activity.

1.2 Statement of the problem

Lake Magadi is characterized by high salinity, alkalinity, sunlight, and CO_2 (Grant et al, 1999). Despite this, the lake harbors a high diversity of microbial communities, including polysaccharide-degrading bacteria (Nyakeri et al., 2018). However, the hydrolytic potential of bacteria from the Lake is largely unexplored (Møller et al., 2010; Sorokin et al., 2015). Studying the hydrolytic capability of these microbes is crucial as it may lead to the discovery of novel extremozymes with unique properties that are valuable for industrial applications, such as biofuel production. This is particularly important because the modern world is grappling with increasing energy costs due to fast-depleting oil reserves. Furthermore, the use of fossil fuels releases greenhouse gases that have led to catastrophic repercussions of climate change (Huaidong Zhang et al., 2015). On the other hand, second-generation biofuels make

use of non-food-based lignocellulose derived from agricultural and forest sources. This adds value to waste and thereby protects the environment (Isikgor & Becer, 2015). Production of useful bio-products from these wastes involves lignocellulosic pre-treatment, hydrolysis, and fermentation (Satari et al., 2019). However, at over 35% of the production costs, cellulolytic fermentation is the most cost-intensive step (Ravindran & Jaiswal, 2016). Besides, most industrial cellulases are mesophilic, meaning that they exhibit stability challenges when subjected to industrial operating conditions of temperatures exceeding 45°C, high salt turnover, and pH conditions (Datta et al., 2013; Cabrera & Blamey, 2018). Moreover, the bulk of mesophilic enzymes originate from fungi, which unlike bacterial sources cannot grow faster, are not amenable to genetic engineering, and are generally less adaptable (Maki et al., 2009; Biswas & Paul, 2017). A key ecological question is how microbial diversity changes with the fluctuating physicochemical conditions with seasons. This study hypothesized that microbial communities within the lake shift in response to changes in the water chemistry over time. Lastly, although haloalkaliphilic bacteria have shown to be a promising source of highly stable extremozymes (Zhilina et al., 2005; Rohban et al., 2009; Babavalian et al., 2013; Ruginescu et al., 2020), little is known about the capacity of particular isolates from Lake Magadi to utilize complex polysaccharides. This study isolated bacteria that could utilize various polysaccharides as the source of carbon using an enrichment culture technique. Sequencing of environmental DNA revealed insights into the structure and diversity of bacteria and archaea across sampling seasons. Furthermore, the isolates and genes encoding polysaccharide hydrolyzing enzymes were screened using artificial metagenomic, whole genome sequencing, and expression analyses.

1.3 Justification

Polymer-degrading bacteria in Lake Magadi function in carbon turnover through enzymatic hydrolysis of complex polysaccharides into simple sugars that become a source of carbon and energy for other microorganisms. Understanding the diversity of polysaccharide-degrading bacteria in L. Magadi will help elucidate the mechanism of nutrient cycling and energy flow in extreme environments. Extremoxymes derived from these bacteria have attracted great interest since they provide safe and environment-beneficial solutions under harsh industrial conditions. Indeed, the enzyme market is expected to grow annually by $400 million, with glycosyl

hydrolases accounting for 8% of the demand for enzymes (Shweta, 2014). Thus, continuous research is required to isolate carbohydrate-active enzymes capable of performing under different formulations at physiological conditions commonly encountered in industrial set-ups (Wohlgemuth et al., 2018). Particularly, haloalkaliphilic bacteria obtained from soda lakes can secrete stable extremozymes conversion of biomass in the industry because they have evolved to withstand harsh ecosystem conditions. Additionally, understanding how microbial diversity changes in Lake Magadi with the changes in physicochemical conditions across seasons is important in providing insights into the role of water chemistry in influencing microbial patterns in soda lakes.

1.4 Objectives

1.4.1 General objective

To determine the diversity and function of polysaccharide-degrading bacteria in the hypersaline Lake Magadi in Kenya.

1.4.2 Specific objectives

1. To isolate and characterize the polysaccharide-degrading bacteria from Lake Magadi.
2. To determine the spatiotemporal diversity and composition of the microbial communities in Lake Magadi.
3. To construct an artificial metagenome for polysaccharide-degrading bacteria isolated from Lake Magadi.
4. To clone, express, and characterize a salt-activated endoglucanase gene from the artificial soda lake metagenome.

CHAPTER TWO

2.0 LITERATURE REVIEW

2.1 Genesis and geochemistry of soda lakes

Soda lakes are thalassic highly saline and alkaline ecosystems distributed around the globe (Fig. 2.1). They are mainly established as terrestrial basins in arid and semi-arid regions whose climatic conditions results in evaporation and hence accumulation of salts in local depressions (Foti et al., 2008; Linhoff et al., 2011). Several theories associate soda lakes with tectonic, volcanic, or meteoric origins. These theories propose that water saturated with carbon dioxide (CO_2) increases the molar concentration of HCO_3^-/CO_3^{2-} as compared to Ca^{2+}/Mg^{2+} (Schagerl, 2016; Paul et al., 2016). Nevertheless, some theories associate their formation with the intricate interplay of hydrological, geological, and topological processes (Grant, 2006). When evaporation exceeds inflows, precipitation, and rapid elimination of Ca^{2+}/Mg^{2+} occur, leading to an increase in the concentration of Na^+, Cl^-, and HCO_3/CO_3^{2-}. These elements accumulate in the form of soda ($Na_2CO_3.10H_2O$), sesquicarbonate (trona, $Na_2CO_3.NaHCO_3.2H_2O$), and thermonatrite ($Na_2CO_3.H_2O$) which are commercially exploited in soap, glass, textile, paint, and metal industries (Jones et al., 1998).

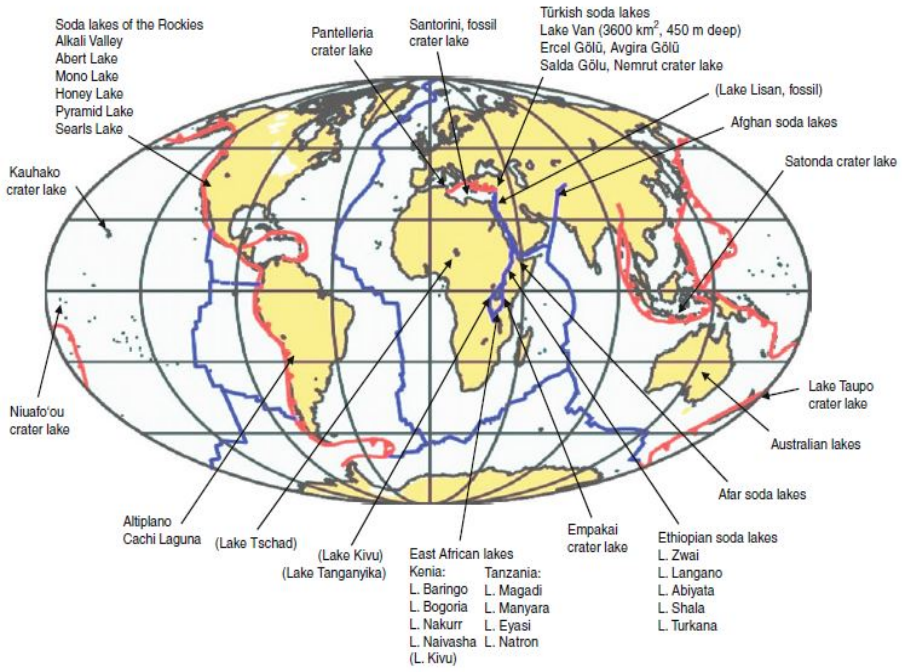


Figure 2.1: Global distribution of soda lakes. Source: Kempe & Kazmierczak (2011).

The term “soda” is deducted from the chemical constituents of the lake’s water, which is considered alkaline due to an abnormally high concentration of inorganic carbon and the availability of sodium as the main cation. When the evaporation process is almost complete, Na_2CO_3 (soda) precipitates on the surfaces (Hugoni et al., 2018). The process of maintaining a strong buffering capacity via the carbonic acid system and CO_2 exchange with the atmosphere results in the accumulation of carbonate and bicarbonate ions. Consequently, the alkalinity is often elevated with pH values ranging from 9.5 to 12 reported in some soda lakes (Foti et al., 2008). However, in more dilute areas alkalinity is lower and alkalinity eventually rises due to rock weathering, solute loss, and evaporative concentrations (Schagerl, 2016). Additionally, continuous evaporation results in an increase in the salt concentration from approximately 5% w/v to saturated sites with 35% w/v where soda lake brines are formed (Rees et al., 2004; Kopejtka et al., 2018). Other characteristic features common to soda lakes include low oxygen levels, and high light and UV intensities (Naghoni et al., 2017). Notably, the physicochemical parameters across the soda lakes vary, thereby affecting the microbial composition and structure (Tazi et al., 2014).

2.2 Microbial diversity in soda lakes

Soda lakes are defined by extreme conditions of high alkalinity and salinity, supply of dissolved inorganic carbon, and often low oxygen availability (Hugoni et al., 2018). Despite such caustic conditions, these ecosystems are among the most productive habitats, hosting diverse and highly adapted microbial and eukaryotic life (Naghoni et al., 2017). The existence of ambient temperatures, high buffering capacity, high incident solar radiation, and constant supply of CO_2 provide the necessary ingredients for the existence of life. The predominant haloalkaliphilic microbial communities thriving in hypersaline soda lakes are bacteria and archaea of major trophic groups (Rees et al., 2004; Vavourakis et al., 2016).

During certain seasons of the year, pigments of pink, purple, red, and green are reported mainly due to blooms of autotrophic micro-organisms (Jones et al.,1998). These blooms are dominated by anoxygenic phototrophic bacteria and/or cyanobacteria belonging to the genus *Spirulina*, *Cyanospira*, and *Arthrospira* in

African ecosystems (Krienitz et al., 2013) and S*ynechococcus* and *Cyanobium* in European alkaline systems (Somogyi et al., 2010). They are phototrophic photosynthetic groups that act as photosynthetic primary producers, supporting the rest of the microbial communities. Here, the daily primary productivity rate is estimated at approximately 10 g Cm^{-2}, much of it acting as a fixed carbon source for aerobic and anaerobic chemoorganotrophic Archaea and Bacteria (Grant et al., 1990).

The soda lakes ecosystem is also a habitat for microbial communities responsible for biogeochemical nitrogen, carbon, and sulfur cycling (Sorokin et al., 2014). These are Chemolithotrophic bacteria, that utilize reduced inorganic compounds as electron donors (Sorokin & Kuenen, 2005). Sulfur-oxidizing bacteria belong to Gammaproteobacteria where the genus *Thioalkalimicrobium*, *Thioalkalivibrio*, and *Thioalkalispira* predominate. However, *Thioalkalivibrio*, a haloalkaliphilic chemolithoautotrophic bacteria is the most diverse group reported so far (Ahn et al., 2021). Nitrogen-fixing bacteria include *Nitrobacter alkalicus* and *Nitrosomonas halophila* belonging to *Alphaproteobacteria* and *Betaproteobacteria*, respectively (Sorokin & Kuenen, 2005). On the other hand, inorganic carbon fixation is performed majorly by cyanobacteria of the genera *Arthrospira* (*Spirulina*), *Anabaenopsis*, and *Cyanospira* (Krienitz et al., 2013).

A combination of salinity and pH in soda lakes confer the major driving forces in microbial composition and structure (Banda et al., 2020; Arayes et al., 2021). Those that require high alkalinity (>pH 9) and salinity up to 33% (w/v) are referred to as the haloalkaliphilic subgroup of extremophiles. Depending on the optimal salt concentration for growth, these bacteria can be slightly halophilic (1 – 3%, 0.2–0.5M NaCl), moderately halophilic (3 – 15%, 0.5 – 2.5M NaCl), and extremely halophilic (15 – 30%, 2.5 – 5.2M NaCl) (Amoozegar et al., 2017; Didari et al., 2020). An increase in salinity particularly above 20% (w/v) correlates with an increase in Archaeal dominance over bacteria (Simachew et al., 2016).

Early isolation surveys of soda lakes of the East African Rift Valley identified many bacterial strains belonging to aerobic heterotrophic haloalkaliphilic organotrophs (Duckworth et al., 1996). With the advancement in technology, recent studies have

revealed genera belonging to Deinococcus-Thermus, Planctomycetes, Actinobacteria, Proteobacteria (α, γ and δ), Bacteroidia, Nitriliruptoria, Oxyphotobacteria, and Verrucomicrobiae have been reported in various soda lakes (Simachew et al., 2016; Kambura et al., 2016; Tandon et al., 2020). On the other hand, major Archaea groups found in these ecosystems are associated with phylum Euryarchaeota (order Halobacteriales, Haloferacales, and Natrialbales), and family *Methanosarcinaceae* and *Methanocaldococcaceae* (Abdallah et al., 2016; Kambura et al., 2016).

Besides bacteria and archaea, fungi and viruses can also be found in these ecosystems. In surveys of the hot springs of the Kenyan Rift Valley, members of the phylum Ascomycota, Basidiomycota, and Glomeromycota have been detected (Salano et al., 2017; Orwa et al., 2020). Similar results were reported in Lonar Lake (Sharma et al., 2016), indicating that specific fungus groups could adapt well to soda lake environments. Little information is published on the diversity of viruses in soda lakes. A few studies have reported the presence of bacteriophages belonging to the family *Myoviridae*, *Podoviridae*, and *Siphoviridae* (Van Zyl et al., 2016; Motlagh et al., 2017).

2.3 Bacterial diversity in Lake Magadi

The bacterial communities of hypersaline Lake Magadi have been studied using both culture-dependent and culture-independent approaches. In surveys using metagenomic amplicon sequencing, the hot springs of the lake were reported to harbor members of the phylum Actinobacteria, Proteobacteria, and Firmicutes (Kambura, 2016). Culture-dependent isolations have identified members of *Dietzia*, *Bacillus*, *Clostridium*, *Norcadia*, *Rhodococcus*, *Halomonas*, *Alkalispirillum*, *Alkalilimnicola*, *Alkalibacterium*, *Staphylococcus*, *Micrococcus* and *Microbacterium* (Duckworth et al., 2000; Rees et al., 2004; Sorokin et al., 2006; Ronoh et al., 2013; Kiplimo et al., 2019) and *Tindallia magadii* (Kevbrin et al., 1998). Further efforts using an assortment of medium formulations yielded isolates related to *Bacillus licheniformis*, *Bacillus subtilis*, *Bacillus cereus*, *Halomonas campisalis*, *Clostridium sporogenes*, and *C. botulinum* (Nyakeri et al., 2018).

Generally, culture-dependent isolations have been the most popular characterization mechanism employed in studying the bacteria of Lake Magadi. This technique is the gold standard for genotyping and physiologically characterizing the pure isolate (Vaz-Moreira et al., 2013). So far, however, no published information is available on whole-genome sequencing (WGS) analyses of these bacteria, hence extrapolation of the genes responsible for adaptation and other important factors is lacking.

2.4 Adaptive mechanisms of bacteria to soda lake ecosystems

Haloalkaliphilic bacteria exhibit modified physiological and structural features to survive under multiple stresses such as elevated salt and high pH or temperature. This shields vital biomolecules such as DNA and RNA from deformation and maintains the activity of cytosolic enzymes (Zahran, 1997; Ruga, 2010). Bacteria use energy towards the removal of salt from the cytoplasm to prevent aggregation of proteins through the 'salting out' strategy (Santos and Da Costa, 2002). Other groups store compatible solutes such as low-molecular-weight compounds like bacteriorhodopsins, biosurfactants, exopolysaccharides, proline, ectoine, and glycine-betaine. They are zwitterionic (neutral) organic compounds whose function is to counter the damaging effects of highly saline conditions on cell physiology and loss of cell water (Santos and Da Costa, 2002; Bursy et al., 2008).

At high temperatures, the fluidity of membranes increases significantly and thus disrupts the cell. Haloalkaliphiles such as *Tindallia* spp., *Natronincola* spp. and *Natroniella* spp. synthesize large amounts of unsaturated fatty acids (UFAs), particularly C16:1ω7 and C18:1ω7 (Banciu & Muntyan, 2015). This mechanism allows bacteria to survive by maintaining the right membrane fluidity (Ulrih et al., 2009). High pH (above 11) and low concentrations of H+, Mg2+, and Ca2+ prevent cells from synthesizing energy. In this case, bacteria actively pump in Mg^{2+} and Ca^{2+} and export out non-essential ions to maintain homeostasis in their interior (Krulwich et al., 2011). Additionally, haloalkaliphilic bacteria have a molecular technique involved in activating antiporter and symporter systems. The antiporter system generates an electrochemical gradient of Na^{+} and H^{+}, while the symporter system facilitates the uptake of Na^{+} and other solutes into the cells (Chinnathambi, 2015).

The cell wall of Gram-positive haloalkaliphiles is a primary barrier cushioning inner cells from salinity and alkalinity stresses. For example, *B. halodurans* has a cell wall with teichuronopeptide (TUP) and teichuronic acid in addition to peptidoglycan (Aono et al., 1990). Similarly, colored pigments produced by primary producers have been postulated to protect the cells against strong UV radiations. Furthermore, they are antioxidants that prevent cell damage due to oxidative stresses (Oren, 2002).

2.5 Lignocellulosic biomass, a potential source of renewable energy

Lignocellulose (Figure 2.2) which is the main constituent of plant biomass consists of cellulose (40 - 50%), hemicellulose (25 - 35%), and lignin (15-20%) intricately linked via covalent and non-covalent bonds (Sun et al., 2016). Altogether, these components comprise approximately 80% of the Earth's total biomass (Bar-On et al., 2018). Cellulose, an insoluble linear homopolymer, is the most abundant component of plant biomass and the most abundant biopolymer on Earth (Behera et al., 2017). The D-glucopyranose monomers are linked together through β-1,4 glycosidic bonds, forming microfibrils as the backbone of the plant cell walls. Hemicellulose is a branched heteropolymer made of short chains of different polysaccharides like xylan, galactomannan, glucuronoxylan, arabinoxylan, glucomannan, and xyloglucan that are linked together by β-(1,4)- and/or β-(1,3)-glycosidic bonds (Zhou et al., 2017).

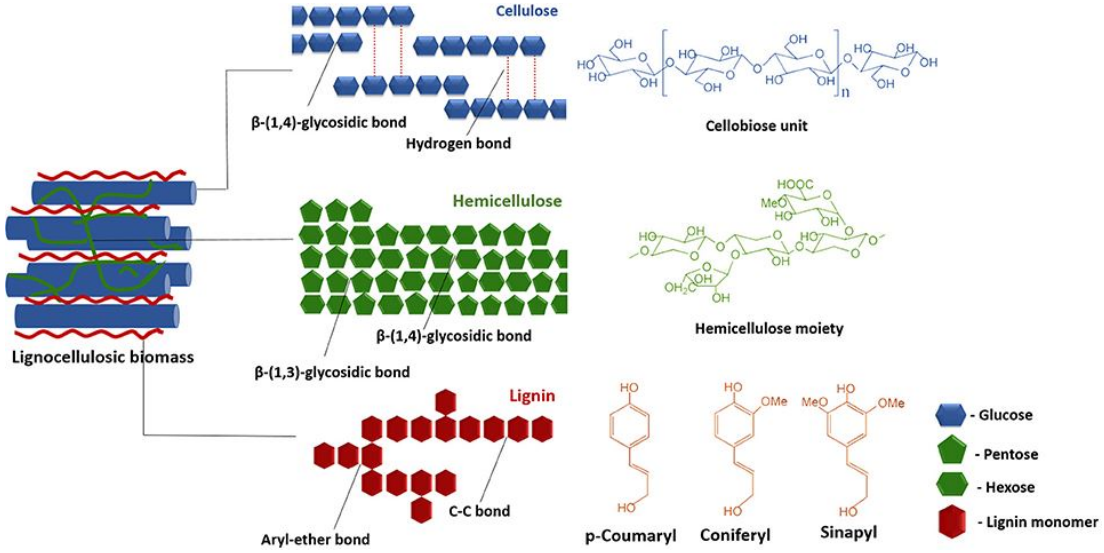


Figure 2.2: Polymer constituents of lignocellulosic materials. (Baruah et al., 2018).

Unlike cellulose, hemicellulose is easily depolymerized into monosaccharides because of its non-crystalline nature and low degree of polymerization (Farhat et al., 2017). On the other hand, lignin, the third most abundant aromatic biopolymer is a highly cross-linked compound made of monolignols, particularly hydroxycinnamyl alcohols, which include *p*-coumaryl, coniferyl, and synapyl alcohols (Pollegioni et al., 2015). It forms the boundary between cellulose and hemicellulose, hence strengthening the overall plant structure. Thus, during lignocellulosic degradation, lignin is the main barrier limiting enzymatic access to cellulose and hemicellulose (Wi et al., 2015).

Due to its abundance, lignocellulosic biomass is a potential alternative to fossil fuels for producing sustainable second-generation biofuels and other bio-based products without competing with food and feed crops (Chandel et al., 2018). However, the recalcitrance of this compound to various pre-treatment methods is a common hindrance to enzymatic hydrolysis and fermentation (Isikgor & Becer, 2015). Before lignocellulosic feedstocks can be used, their crystalline structure, heterogeneous structure, and level of lignification must be overcome.

Several physical, chemical, physicochemical, and biological pre-treatments have been used to achieve structural depolymerization of lignocellulose (Baruah et al., 2018). Once cellulose and hemicellulose are exposed, microorganisms secrete a group of enzymes collectively referred to as Carbohydrate-Active enZymes (CAZymes), breaking down complex polysaccharides into simple monomers (Cantarel et al., 2009). The CAZymes database lists these enzymes as Glycoside Hydrolase (GH), which catalyzes the hydrolysis of glycosidic bonds; glycosyltransferases (GTs), whose role is to synthesize glycosidic bonds from phospho-activated sugar donors; polysaccharide lyases (PLs), which through the β-elimination mechanism, cleave the glycosidic bonds of uronic acid-containing polysaccharides; carbohydrate esterases (CEs), that aid in removing ester-based modification found in mono-, oligo-, and polysaccharides and carbohydrate-binding modules (CBMs) which enhances sustained enzyme-substrate binding, thereby increasing depolymerization (Cantarel et al., 2009).

2.6 Polysaccharide depolymerizing bacteria from soda lakes

Soda lakes, like other extreme ecosystems, have received a lot of attention, primarily because they contain extremophilic bacteria with the potential to produce novel extremozymes. Unlike their mesophilic counterparts, extremozymes demonstrate remarkable stability under extreme conditions such as those portrayed in an industrial setup (Raddadi et al., 2015). An important feature of haloalkaliphilic enzymes is their ability to outcompete salt ions for hydration and maintain their native conformation in the presence of high ionic concentration (Britton et al., 2006). As a result, there has been a notable increase in demand for these enzymes in recent years, with polymer-degrading enzymes such as amylases, cellulases, proteases, cellulases, chitinases, and xylanases leading the pack (Sahay et al., 2012).

Generally, soda lake bacteria with a potential source of polysaccharide degrading enzymes have been associated with members of genera *Marinobacter*, *Bacillus*, *Halobacillus*, *Virgibacillus*, *Clostridium*, *Natroflexus*, *Geomicrobium*, *Nocardiopsis*, *Oceanobacillus*, *Halomonas,* and *Staphylococcus* (Kumar et al., 2012; Babavalian et al., 2013; Nyakeri et al., 2018). Several of these bacteria are aerobic, however, anaerobic bacteria such as low halophilic cellulose hydrolyzing *C. alkalicellulosi* have also been isolated (Zhilina et al., 2005). Several specific haloalkaliphiles are known producers of hydrolytic enzymes. For example, novel pH-stable endoglucanase (JqCel5A) was isolated from a Gram-positive *Jonesia quinghaiensis* (Lin et al., 2016). Chitin degrading *Marinimicrobium*, *Bacillus*, and *Nocardiopsis* spp. have been isolated from hypersaline lakes of Kulunda, Russia (Sorokin et al., 2012). At moderate salinity levels, *Natronoflexus pectinovorans* (*Bacteriodetes*) produced pectinase (Sorokin et al., 2011), while *Natranaerovirga hydrolytica* (Phylum Clostridia) hydrolyzed pectin at high salinity (Sorokin et al., 2012). *Bacillus pumilus* isolated from a salt farm demonstrated xylanolytic activity against xylan from birchwood (Menon et al., 2010).

Halophilic cellulases have been cloned from *Bacillus zhangzhouensis* (Ruginescu et al., 2018), *Marinimicrobium haloxylanilyticum* (Møller et al., 2010) and *Bacillus* sp. B14 (Thirumala & Reddy, 2017). Studies in Lake Magadi established that *Natronococcus* spp. produced a stable α-amylase that could withstand high pH,

salinity, and temperatures (Kobayashi et al., 1992); while *Clostridium alkalicellulosi* secreted cellulases with an affinity for amorphous cellulose (Zhao et al., 2014). Other qualitative assays have indicated that members of the *Bacillus* and *Clostridium* families have the potential to secrete extracellular amylases, lipases, proteases, cellulases, xylanases, and esterases (Nyakeri et al., 2018; Mulango et al., 2020).

2.7 Polysaccharide degrading enzymes

2.7.1 Cellulases

They belong to the glycosyl hydrolase family of enzymes responsible for the degradation of cellulose into simple molecules. Principally, they are categorized into 3 classes (cellulosome) acting synergistically to degrade recalcitrant lignocellulosic materials (Singh et al., 2019). Endoglucanase (EC 3.2.1.4; endo-1,4-β-D-glucan-4-glucanohydrolases) breaks down β-1,4 glucosidic linkages of lignocellulosic biomass to generate oligosaccharides with non-reducing ends. Secondly, exoglucanases (EC 3.2.1.91; Exo-1,4-β-D-glucan-4-cellobiohydrolases) cleave the long-chain oligosaccharides to short-chain cellobiose dimmers, whereas β-glucosidases (EC 3.2.1.21) liberates glucose from cellobiose (Romano et al., 2013; Rafael et al., 2016).

Halophilic cellulases have been utilized in industries like the textile industry for purposes of bioethanol production, 'biopolishing', de-inking in the paper industry, in animal feed for improvement of digestibility, and in the human food industry for processing of wheat products and juice extracts (El-sersyn et al., 2010). Putative cellulases of halophilic nature have previously been isolated from *C. alkalicellulosi* (Zhilina et al., 2005), *C. alkalicellulosi* (Zhao et al., 2014), *Jonesia quinghaiensis* (Lin et al., 2016), *Cellulomonas bogoriensis* (Li et al., 2020) and *B. pumilus* (Ogonda et al., 2021).

2.7.2 Amylases

Responsible for the depolymerization of starch and its derivatives to oligosaccharides and monosaccharides, amylases (E.C.3.2.1.1) constitute about 25% of the industrial enzymes in the global market (Jaralla et al., 2014). They are utilized in fermentation, paper, and textile industries, and the manufacture of syrup (Kobayashi et al., 1992). Haloalkaliphilic amylase-secreting bacteria that have been characterized include *Chromohalobacter* sp. TVSP 10 (Prakash et al., 2009), *Georgenia satyanarayanai* (Srinivas et al., 2012), members of genera *Kocuria*, *Exiguobacterium*, *Citricoccus*,

Oceanobacillus, and *Staphylococcus* (Mukhtar et al., 2018) as well as actinomycetes *Nocardiopsis* spp. and *Streptomyces* spp. (Jenifer et al., 2015).

2.7.3 Chitinases

Chitin is a crystalline polymer comprising 1,4 linked 2-acetamido-2-deoxy-β-D-glucan and acetyl molecules linked to several nitrogen molecules along the chain. It is a widely distributed component in nature providing primary protection to insects, crustaceans, and fungi (Vaijayanthi et al., 2016). Halophilic bacteria utilize chitinases (EC 3.2.11.14) to depolymerize the ecosystem's chitin to generate carbon, nitrogen, and energy, hence helping in carbon and nitrogen cycling (Lindahl & Finlay, 2006). Previous studies elucidate complete chitin degradation paths by halophilic *Chitinivibrio alkaliphilus* (Sorokin et al., 2014). Chitinases have also been associated with *Chitinivibrio* spp. isolated from seafood (Le et al., 2018), *Bacillus mannanilyticus* from a soda lake (Aktuganov et al., 2018), members of the genus *Marinimicrobium*, *Saccharospirillum*, and *Arhodomonas* from hypersaline Kulundu Steppe (Sorokin et al., 2012; Sorokin & Kolganova, 2014). Members of actinobacteria are also known to depolymerize chitin to chitobiose, a compound that has a potential anti-oxidant, and other chito-oligomers that have applications in medical, agricultural, and food industries (Bhattacharya et al., 2008).

2.7.4 Xylanases

Xylanases (EC 3.2.1.8) degrade xylan, the most abundant component of hemicelluloses to monosaccharide xylose. This process is used in textile industries for bio-bleaching and the production of animal feed (Mukhtar et al., 2017; Priya et al., 2012). Hypersaline lakes are a potent source of xylanases with members of Bacteroidetes, Proteobacteria, Actinobacteria, Firmicutes, and Verrucomicrobia being linked to chitinase encoding genes (Wang et al., 2019). However, successful cloning and expression of alkaline-stable xylanases have been reported for *Micrococcus* sp. AR-135 (A. Gessesse & Mamo, 1998), *Bacillus* sp. (Zhang et al., 2010), *Planococcus* sp. SL4 (Xiaoyun Huang et al., 2015), and *Alkalitalea saponilacus* (Liao et al., 2019).

2.7.5 Pectinases

Pectin is a heterogenous polysaccharide that composes 35% of the cell walls of dicotyledonous plants (Kumar and Suneetha, 2015). Pectinases are complex enzymes that include pectin methyl esterases (EC.3.1.1.11), pectin lyases (EC.4.2.2.2), and polygalacturonases (EC.3.3.1.15) which work in synergy to degrade pectin (Jayani et al., 2005). Pectinases are essential enzymes in the clarification of wines, oils, juices, the textile industry, flavoring compounds, and the preparation of linen fabrics (Mukhtar, et al., 2017). In soda lake ecosystems, pectinolytic bacteria that have been isolated belong to the genera *Bacillus*, *Oceanobacillus*, *Gracilibacillus*, *Halobacillus*, *Halomonas*, *Salicola*, *Streptomyces*, and *Natranaerovirga* (Sorokin et al., 2012; Babavalian et al., 2014; Zhou et al., 2017; Pathak & Jadhav, 2019).

2.8 Approaches to the mining of novel haloalkaliphilic bacteria

2.8.1 Culture-dependent technique

Isolation of bacteria via culture methods remains an indispensable tool owing to its contribution towards understanding the strain's attributes such as physiology, pathology, genetics, and their roles in ecology, host health, and the genes encoding for natural bio-products (Vartoukian et al., 2010). Various techniques such as *in vitro* simulation of natural habitats (Kaeberlein and Epstein, 2002) and co-culture (Moore et al., 2007) have proven rather successful culture methods, albeit with a low success rate during sub-culture. To maximize isolation, a combination of methods allowing for continuous manipulation of growth media to contain energy and carbon sources, nutrients, and adequate physicochemical parameters such as air, inoculation size, temperature, and pH is necessary (Pham and Kim, 2012). Additionally, lengthening the incubation period and community co-culture of microorganisms that demonstrate symbiotic relationships enhances cultivation (Kamagata and Tamaki, 2005; Alain and Querellou, 2009).

Selective isolation of strains with unique characteristics is an important research objective in many studies. For instance, selective isolation of fastidious bacteria such as the extremophilic haloalkaliphiles could be achieved through heat treatment of samples to approximately 80^{o}C (Travers et al., 1987). Other techniques entail serial dilutions followed by the introduction of inhibitors of unwanted microorganisms such

as cycloheximide, nystatin (antifungals), and nalidixic acid (anti-Gram negative bacteria), as well as modification of media to simulate the natural nutrient composition (Vartoukian et al., 2010; Pham and Kim, 2012; Bates et al., 2013). When screening isolates for the desired phenotype, interest is in the ability of a given isolate to utilize a particular substrate, hence producing a desired primary and/or secondary metabolite. In this case, media containing the desired substrate and inorganic elements such as $MgSO_4$, K_2HPO_4, inorganic nitrogen source, agar, and trace elements are included. Furthermore, increasing incubation time gives ambient time for the slow growers to adapt to and grow in the given medium (Vartoukian et al., 2010). The cultivation of isolates for enzyme production has been extensively utilized in the isolation of haloalkaliphilic bacteria (Sánchez-Porro et al., 2003; Rohban et al., 2009; Mwirichia et al., 2010; Mulango et al., 2020).

2.8.2 Culture-independent techniques

Even with advancements in culture methods, it is well established that less than 1% of microorganisms can be cultivated *in vitro*. The low success has been attributed to a myriad of challenges such as slower growth amongst some strains relative to fast-growers, biased classification of phylotypes with indistinguishable phenotypes, and the inability of medium ingredients to select for all bacteria in an ecosystem (Vartoukian et al., 2010). Metagenomics is a culture-independent technology that has facilitated the development of metabolic models, modeling microbial interactions with the ecosystem, and elucidation of microbial evolution (Handelsman, 2005; Streit & Daniel, 2010; Cárdenas et al., 2016; Cycil et al., 2020). It entails extraction and direct analysis of total environmental DNA, through the creation of insert DNA libraries and sequencing of these libraries. Metagenomic approaches have aided in the reconstruction of nearly complete metagenome-assembled genomes (MAGs) of uncultured bacteria in various hypersaline environments (Vavourakis et al., 2018; Vera-Gargallo & Ventosa, 2018). In assigning function to a particular enzyme in a species or community of microorganisms the use of environmental genomic expression libraries screening has become the best method (Podar & Reysenbach, 2006). This method has been used to mine halophilic esterase from the hypersaline lake Acıgöl, Turkey (Tutuncu et al., 2019) and cellulases, proteinases, and lipases (Grant & Heaphy, 2010). So far, no study used metagenomic screening to report polysaccharide-degrading enzymes in Lake Magadi.

2.9 Overexpression and purification of halophilic proteins

Since haloenzymes are activated by moderate to high salt concentrations, they require a minimum amount of water to function (Sinha & Khare, 2014). Cloning and expression in heterologous hosts such as *Escherichia coli* is cumbersome because these enzymes may misfold and aggregate in conditions of low ionic strength (Allers, 2010). Nevertheless, halophilic proteins like glutamate dehydrogenase, α-amylase, glucose dehydrogenase, nitrite reductase, and isocitrate lyase from *Haloferax mediterranei* have been cloned and expressed in pET3a vector transformed in *E. coli* BL21(DE3) (Esclapez et al., 2006), cellulases *En5H* and *CelB* from halophilic, alkalithermophilic *Alkalilimnicola* sp., and *Bacillus* sp. BG-CS10, respectively (Mesbah & Wiegel, 2017; Zhang et al., 2010), chitin deacetylase from *Bacillus aryabhattai* B8W22 (Pawaskar et al., 2021), and α-amylase from halophilic *Nesterenkonia* sp. (Solat & Shafiei, 2021).

Earlier methods of purification involved, after cell lysis, the use of organic solvents like acetone in combination with ammonium sulfate precipitation. After an hour of continuous stirring in an ice bath, α-amylase obtained from *Bacillus* sp. was precipitated using 30% and 50% acetone and ammonium sulfate, respectively (Bole et al., 2013). Halophilic pullulanase from haloalkaliphilic *B. halodurans* has been purified using acetone precipitation, centrifugation, and ion-exchange chromatography with DEAE cellulose (Asha et al., 2013). Elsewhere, while isolating a thermostable alkalophilic cellulase from *B. vallismortis*, Gaur and Tiwari (2015) attained a recovery of 28% through a combination of ion-exchange chromatography, gel filtration chromatography, and 80% ammonium sulfate precipitation under sodium phosphate buffer (Gaur and Tiwari, 2015). Further modification of purification has been reported for xylanase characterization where ammonium precipitation dialysis in cellulose tubing (13,000 kDa, Himedia LA393-10 MT) and subsequent purification using ion-exchange chromatography yielded xylanase with an overall specific activity of 299.25 U/mg (Kamble and Jadhav, 2012). Similar techniques have been used to partially purify amylase obtained from soda lake bacteria (Ashabil et al., 2014; Bansod et al., 2014).

Modern-day techniques utilize modification of amplification primers to include histidine tags at the N or C-terminal ends of the protein. This way the target protein can be purified via immobilized metal affinity chromatography. It was employed in the purification of his-tagged recombinant proteins deduced from haloarchaeon *Haloferax volcanii* (Allers, 2010), chitin deacetylase (Pawaskar et al., 2,021), and a cellulase (Zhang et al., 2012). Furthermore, the use of affinity chromatography on the HisTrap column connected to the ÄKTA purifier (GE Healthcare, New York, USA) is a more advanced method of purifying hist-tagged proteins in a series of columns, thus maximizing protein yields (Ishibashi et al., 2011; Aulitto et al., 2021).

CHAPTER THREE

3.0 ISOLATION AND CHARACTERIZATION OF POLYSACCHARIDE-DEGRADING BACTERIA FROM LAKE MAGADI

3.1 Abstract

Bacteria obtained from extreme ecosystems are a potential source of industrially applicable biocatalysts. This study explored the diversity and functional potential of polysaccharide-degrading bacteria isolated from the hypersaline lake Magadi in Kenya. Bacterial isolation and screening for extracellular enzymes were performed using sterile lake water amended with either cellulose, carboxymethylcellulose (CMC), xanthan, pectin, starch, or xylan as the sole carbon source. Phylogenetic analysis of 40 isolates was performed based 16S rRNA gene. The results indicated that the isolates were affiliated to the genera *Salipaludibacillus*, *Halomonas*, *Alkalibacterium*, *Alkalihalophilus*, *Evansella*, *Shouchella*, *Halalkalibacterium*, *Halalkalibacter*, *Alkalibacterium*, and *Salinicoccus.* Most of the isolates exhibited high hydrolytic activities on the test substrates (enzyme activity index of ≥ 2.0). The highest enzyme activity was recorded for the isolates belonging to *Salipaludibacillus* while the least activity was recorded for hamolonads. Further, 4 isolates (LMS6, LMS18, LMS25, and LMS39) were selected for whole genome sequencing based on their ability or lack thereof to degrade the test polysaccharides. LMS25 affiliated to *Salupaludibacillus* sp. and LMS39 affiliated to *Alkalihalobacterium* sp. utilized all test substrates. Sample LMS6 belonging to *Shouchella* utilized only one test substrate while LMS18 affiliated to *Evansella* did not exhibit any activity with the test substrates. Average nucleotide identity (ANI) analysis revealed that LMS6, LMS18, LMS25, and LMS39 were new species associated with *Shouchella* sp., *Evansella* sp., *Salipaludibacillus* sp., and *Alkalihalobacterium* sp., respectively. Screening for genes that encode Carbohydrate-Active enZymes (CAZymes) within the genomes revealed the presence of genes coding for amylase, cellulase, pectinase, xylanase, and chitinase enzymes. This supported the findings of phenotypic screening, demonstrating that bacterial isolates from Lake Magadi are a rich source of polysaccharide-hydrolyzing enzymes with potential industrial significance.

3.2 Introduction

Alkalihalophilic bacteria are double extremophiles requiring an alkaline pH (9 - 13) and high salinity (3 - 30%) to grow (Sorokin et al., 2014; Mukhtar et al., 2017). The bacteria have been isolated from hypersaline environments including soda lakes, carbonate springs, marine environments, coastal lagoons, acid mines, solar salterns, salt brines, and the Dead Sea (El Hidri et al., 2013). To maintain functional and cellular integrity under such conditions, alkalihalophilic bacteria have developed multi-level adaptive strategies such as the 'salt-in' strategy, which results in the accumulation of inorganic salts (Na^+, H^+, and K^+) and biosynthesis of compatible

solutes like polyols, sugars, amino acids, betaines, and ectoine (Gunde-Cimerman et al., 2018).

Soda lakes are an example of extreme environments with a global distribution. However, most of the soda lakes are found along subtropical latitudes under arid and semi-arid conditions with underlying rocks contributing to CO_2-rich waters (Sorokin et al., 2014). When the rate of evaporation exceeds inflows, precipitation of Ca^{2+}/Mg^{2+} and an increase in the concentration of Na^+, Cl^-, and HCO_3/CO_3^{2-} occur (Jones et al., 1998). Consequently, the elevated sodium carbonate concentrations cause an increase in pH (Grant et al., 1999; Furian et al., 2013; Vavourakis et al., 2016; Andreote et al., 2018). These environmental conditions influence the diversity and structure of microbial communities in the soda lakes (Grant, 1999).

The adaptation of these microbes to extreme pH and salt conditions, alkalihalophilic bacteria has attracted interest as a source of biopolymer-degrading enzymes. This has further been supported by the recent resurgence of interest in natural products due to advancements in microbial whole genome sequencing, bioinformatics, and biosynthesis (Galanie et al., 2020). Genomic sequencing provides a repertoire of genetic resources for biopolymer degradation for potential use in industry (Ma et al., 2024). Such microorganisms secrete extracellular enzymes that convert complex biopolymers such as lignocellulose (Gupta et al., 2012; Berlemont & Martiny, 2016; Lovegrove et al., 2017) and lipids to simpler molecules such as glucose and fatty acids, which serve as energy and carbon sources (Azeri et al., 2010; Oren, 2010). Therefore, hydrolysis of these polymers is a critical step in carbon cycling in the surrounding environments (Stewart, 2012; Berlemont et al., 2013; Martin et al., 2015).

Industrial processes employing biopolymers such as biofuel production involve high salt concentrations, temperatures, pH values, and ionic strengths. Thus, biocatalysts that can function optimally under such conditions and/or robust enzymes, so-called extremozymes, superior to the commonly available mesophilic enzymes, are required (Elleuche et al., 2014). Enzymes from alkalihalophilic bacteria would be favorable

since they can be used for the conversion of recalcitrant molecules under industrial conditions (Karan et al., 2012).

The hyperalkaline saline lake Magadi is a unique ecosystem located at the southern end of the Kenyan Great Rift Valley. The near-saturation salt concentrations (200 - 350 g/l), and high pH values (9.0 - 12.5) make it the most saline and alkaline lake of the Rift Valley. The lake is, therefore, an interesting study site for the diversity and adaptation of microorganisms thriving in hypersaline systems (Duckworth et al 1996; Zhilina et al., 2005). Studies of the microbes in the lake using both culture-dependent and culture-independent approaches revealed the presence of strains belonging to *Bacillus*, *Clostridium*, *Halomonas*, *Alcaligenes*, *Verrumicrobium*, *Nitriliruptor*, *Methanococcus* and *Halobacterium Alkalibacterium*, *Arthrobacter* and *Micrococcus* (Tindall et al., 1980; Kambura et al., 2016; Nyakeri et al., 2018). Little information exists on the ability of these microbes to produce extremozymes. This study hypothesized that the hypersaline Lake Magadi is a reservoir of diverse bacteria with the potential to utilize a wide array of polysaccharides and potential utility in industrial processes involving high salt concentrations, temperatures, pH values, and ionic strengths. The isolated bacterial strains from the lake and evaluated their ability to produce extracellular polysaccharide-degrading enzymes by utilizing different test polysaccharides as substrates. Moreover, whole-genome sequencing and analysis of Carbohydrate-Active enZymes (CAZymes) were performed with selected isolates.

3.3 Materials and methods

3.3.1 Sampling site and sample collection

Samples were collected from Lake Magadi which is in the deepest depression of the Kenyan Rift Valley (1°43’-2’00’S and 36°13’-36°18’E). It has an altitude of 600 m above sea level and covers an area of approximately 90 km^2. Water is supplied by more than 200 geothermal springs along the perimeter of the lake and the two rainy seasons contribute about 500 mm of rainfall (Ruga, 2010). However, the lake lacks an outlet, and it experiences a high rate of evaporation due to high temperatures (>40ºC) during the day (Behr & Röhricht, 2000). The lake was sampled due to its high alkaline conditions (pH 9.0 – 12.5) that favor the growth of alkaliphilic prokaryotic communities (Grant et al., 1999; Tourova et al., 2014). Five sites from wet and dry sediments, lakeshore soils, brine, and microbial mats were sampled.

Three sub-samples of 50 ml each were collected from each site and pooled into a composite sample. They were put in a sterile conical centrifuge tube (Biologix, Shandong, China), immediately kept in sterile containers, and sealed. The sampling sites were designated: S1 (wet sediments), S2 (brine sediments), S3 (dry sediments), S4 (Magadi grass soils), and S5 (microbial mats). Samples and lake water were transported in cool boxes to the University of Embu microbiology laboratory.

3.3.2 Enrichment and isolation of polysaccharide utilizing bacteria

Enrichment cultures were prepared by adding 0.1 g of the sample from each sampling site to 9 ml of autoclaved lake water and serially diluted from 10^{-1} to 10^{-8}. An aliquot of 100 µl from dilutions 10^{-5} to 10^{-8} was dispensed into a 4 ml enrichment liquid basal medium containing (g/L of sterile lake water): $CaCl_2.7H_2O$, 0.05; K_2HPO_4, 1; and $MgSO_4.7H_2O$, 0.1. Each medium was supplemented with 1 g of either starch, carboxymethylcellulose, chitin, xanthan, cellulose, lignin, casein, olive oil, Tween 20, sawdust, bran, glycerol, or pectin. The cultures were incubated at 37ºC for 96 h while shaking at 150 rpm in an Innova 4200 incubator (New Brunswick, Edison, NJ, USA). An aliquot (100 µl) from each enrichment was spread on solid media (M1-M6) (Table 3.1).

Table 3.1. The enrichment media used in isolating bacteria from Lake Magadi.

Medium ID	Composition (g/L)
M1	Peptone, 2.0; yeast extract, 0.5; K_2HPO_4, 1.0; $CaCl_2.2H_2O$, 0.05; $MgSO_4.7H_2O$, 0.1; and agar, 14.0.
M2	Starch casein (SCA): soluble starch, 3.0; casein hydrolysate, 0.3; and agar, 14.0.
M3	Soil meat extract peptone (SMP): natural soil 1.0; meat extract 0.3; peptone, 0.5; and agar, 14.0.
M4	Humic vitamin agar (HAV): Humic acid, 1.0; 100X vitamin mix (Merck KGaA, Darmstadt, Germany); and agar, 14.0.
M5	Cellulose casein (CECA): cellulose, 3.0; casein, 0.3; and agar, 14.0.
M6	Calcium carbonate chitin (CCC): $CaCO_3$; 3, chitin, 3.0; and agar, 14.0.

All media were sterilized by autoclaving at 121ºC for 15 minutes and cooled before adding cycloheximide (50 µg/ml) to inhibit the growth of fungi. The plates were inverted and incubated at 37ºC for up to 11 days with continuous monitoring. The colonies were purified by re-streaking three times on trypticase soy agar (TSA) prepared by mixing the following ingredients (g/L): tryptone, 17.0; peptone from soy, 3.0; NaCl, 5.0; K_2HPO_4, 2.5; and 2.5 g of glucose monohydrate. TSA was supplemented with $NaCO_3$ (4% w/v) and NaCl (6% w/v) and incubated at 37ºC for 48 h.

3.3.3 Genomic DNA isolation and amplification of 16S rRNA gene

Bacterial isolates were grown in 4 ml trypticase soy broth (TSB) supplemented with $NaCO_3$ (1% w/v) and NaCl (4% w/v) and incubated in a rotary shaker (180 rpm; 37°C) for 12 h. DNA of 40 bacterial isolates was done using MasterPure complete DNA and RNA purification kit as recommended by the manufacturer (Epicentre, Madison, WI, USA). The quantity and quality of DNA were determined using a Nanodrop® ND 1000 V. 3.8.1 (Thermo Fisher Scientific, USA) and by electrophoresis using a 1% (w/v) agarose gel. Amplification of full-length 16S rRNA gene fragment was done using bacterial universal primers 27F (5′-AGAGTTTGATCMTGGCTCAG-3′) and 1492R (5′-TACGYTACCTTGTTTACGACTT-3′) (Satyapal et al., 2018). Each 50 µl reaction volume consisted of 32.5 µl of nuclease-free water, 0.5 µl of DNA template (100 ng/µl), 10 µl of 10 mM 5X HF Phusion buffer, 1 µl of 10 mM dNTPs, 2.5 µl of each primer (10 µM) and 1 µl of 2 U/µl Phusion high-fidelity DNA polymerase (Thermo Fischer Scientific, Waltham, MA, USA). The PCR amplification was performed under the following cycling parameters: Initial denaturation at 98°C for 2 min followed by 40 cycles of denaturation at 98ºC for 1 min, primer annealing at 53°C for 1 min, and extension at 72ºC for 1 min. A final extension was performed at 72ºC for 5 min. The amplified fragments were checked for correct size via gel electrophoresis using a 1% (w/v) agarose gel. Positive controls for PCR consisted of *Escherichia coli* DNA and negative controls contained nuclease-free water instead of DNA.

3.3.4 Purification of PCR products and sequencing and phylogenetic analysis

PCR products (30 µl) were purified using NucleoSpin® Gel and PCR Clean-up as described by the manufacturer (Macherey-Nagel GmbH & Co. KG. Düren,

Germany). The amplified full-length 16S rRNA genes were sequenced according to the manufacturer's protocol (Microsynth Seqlab GmbH, Göttingen, Germany). Chromatograms of the sense and antisense sequences were used to generate a contiguous sequence and edited using Clone Manager v9.0 (https://www.scied.com). Similarity searches of the 16S rRNA nucleotide sequences were done using the BLASTn search tool at the National Centre for Biotechnology Information (NCBI) (Zhang et al., 2000). Sequence alignments and phylogeny were assessed by employing MEGA X v10.0.5 (Kumar et al., 2018). The phylogenetic tree was generated using the Maximum Likelihood statistical method based on the Tamura 3-parameter model (Tamura, 1992). Type strain sequences of *Salipaludibacillus*, *Halomonas*, *Alkalibacterium*, *Alkalihalophilus*, *Evansella*, *Shouchella*, *Halalkalibacterium*, *Halalkalibacter Alkalibacterium*, *Salinicoccus* were retrieved from the GenBank and included in the phylogenetic analysis. The robustness of the tree was tested by bootstrap analysis employing 1,000 pseudo-replications. Data for 16S rRNA sequences have been deposited in the GenBank under accession numbers OM423173 - OM423212.

3.3.5 Screening for activity against polysaccharide and physiological characterization

Isolates were screened for polysaccharide utilization potential using a basal medium comprising (g/L): peptone, 2.0; yeast extract, 0.5; K_2HPO_4, 1.0; $CaClx2H_2O$, 0.05; $MgSO_4x7H_2O$; 0.1; NaCl, 36.0, $NaCO_3$, 10.0; and agar, 15 supplemented with 1 g of either starch, pectin (from apple), carboxymethylcellulose, xylan (from beech wood), xanthan gum (from *Xanthomonas* sp.) or 0.02 mm microcrystalline cellulose (SERVA GmbH, Heidelberg, Germany). Enzyme activity was assessed using the plate assay in triplicates whereby, a single colony was spotted on the agar plates followed by incubation at 37^{o}C for 48 h (Busch et al., 2017). Screening for cellulase, amylase, xanthanase, and pectinase activities was done by flooding the plates with Lugol's solution (Carl Roth GmbH, Karlsruhe, Germany) (Kasana et al., 2008) for 5 min. The formation of a light-yellow zone around the colony indicates enzyme activity (Dornelas et al., 2017). The enzyme activity index (EAI) was expressed as the ratio between the diameter of the clearance zone and the diameter of the colony (Huang et al., 2012; Soy et al., 2019). The dendrogram of activity indices with

selected substrates was generated as a heatmap using the *heatmap.2* function in R (R Development Core Team, 2020).

Additionally, the isolates were grown at varying pH values, temperature, and NaCl concentrations to check for optimum conditions. Isolates were grown in trypticase soy broth (TSB) with pH values ranging from 6 - 10, at temperatures of 35, 40, 45, and 50°C, and NaCl concentrations of 0, 5, 10, and 15% (w/v). Bacterial cultures were incubated for 24 hours. Cell density was determined by measuring the optical density at 600 nm (OD_{600}) using a spectrophotometer.

3.3.6 Isolation of genomic DNA for whole-genome sequencing

Based on substrate utilization screening, four isolates were selected for full genome sequencing (LMS6, LMS18, LMS25, and LMS39). LMS6 utilized one, LMS25 and LMS39 all, and LMS18 none of the tested substrates. Pure cultures were grown in 4ml of TSB medium supplemented with $NaCO_3$ (1% w/v) and NaCl (4% w/v) using a rotary shaker at 37°C and 180rpm for 12 h. Genomic DNA was isolated using the MasterPure complete DNA and RNA purification kit as recommended by the manufacturer (Lucigen, Middleton, WI, USA) with slight modifications. Briefly, overnight cell cultures were incubated on ice for 15 min and harvested by centrifugation (18,800 *g*, 2 min, 4°C). The resulting cell pellet was suspended in 600 µl Tissue & Cell Lysis solution, mixed with 4 µl Proteinase K (20 mg/µl; Qiagen, Hilden, Germany) and 1 µl Lysozyme (100 mg/ml; Sigma Aldrich, St. Louis, Mo, USA). The mixture was incubated at 65°C for 30 min and then kept on ice for 3 min. Subsequently, the cell lysate was treated with 10 µl RnaseA (5 µg/µl; Sigma-Aldrich Chemie GmbH) and incubated at 37°C for 1 h then chilled on ice for 3 min. Protein precipitation was done by mixing the lysate with 300 µl of pre-chilled MPC Protein Precipitation Agent and centrifuged Precipitation of DNA was done by adding 500 µl of ice-cold isopropanol to the supernatant followed by centrifugation (18,800 *g*, 10 min, 4°C). The DNA pellet was washed twice with 70% ethanol (stored at -20°C) air-dried for 15 min and eluted in 55 µl nuclease-free water. The DNA concentration and integrity were checked via the NanoDrop 1000 Spectrophotometer (Thermo Fischer, Waltham, MA, USA) and a 1% (w/v) agarose gel.

3.3.7 Sequencing, sequence assembly, annotation, and comparative genomics

Libraries for Illumina shotgun paired-end sequencing were prepared using the Nextera XT DNA Sample Preparation Kit as recommended by the manufacturer (Illumina, San Diego, CA, USA). Quality was controlled using the Agilent Bioanalyzer 2100 and the Agilent High Sensitivity DNA Kit as recommended by the manufacturer (Agilent Technologies, Waldbronn, Germany). Paired-end sequencing library concentrations were determined using the Qubit dsDNA HS Assay Kit as recommended by the manufacturer (Life Technologies GmbH, Darmstadt, Germany). Sequencing was performed using a MiSeq system and reagent kit v3 (600 cycles) according to the manufacturer's protocol (Illumina). Libraries for long-read Nanopore sequencing were generated using high-molecular-weight genomic DNA (1.5 µg). The 1D genomic DNA sequencing protocol for the MinION device was conducted using the Ligation Sequencing 1D Kit (SQK-LSK109) as recommended by the manufacturer (Oxford Nanopore Technologies, Oxford, UK). End-repair was done using the NEBNext FFPE RepairMix (New England Biolabs, Ipswich, MA, USA). Sequencing was done for 72 h using a SpotON flow cell Mk I (R9.4) with MinKNOW software v18.12.6 (Oxford Nanopore Technologies). Trimming and removal of adapter sequences were performed by fastp v0.20.1 (Chen et al., 2018) and PoreChop v0.2.4 (https://github.com/rrwick/Porechop) for Illumina and Nanopore reads, respectively. Potential PhiX contamination was removed from Illumina reads using Bowtie2 v2.3.5.1 (Langmead & Salzberg, 2013).

De novo assembly was performed by de Bruijn graph assembler Unicycler v.0.5.0 (Wick et al., 2017) where K21, K35, K55, K77, K99, and K27 k-mer lengths were assayed. In the end, K127 generated the highest rate of read utilization and the longest contiguous sequences (contig) length. The generated contigs were validated using Bandage v0.8.1 (Wick et al., 2015). Annotation of the assembled genomes was done using the Prokaryotic Genome Annotation Pipeline (PGAP) v6.0 (Tatusova et al., 2016). Briefly, contigs generated by Unicycler were loaded into PGAP using the command */pgap.py -r -o baccanu_anno ./input.yaml*. Structural annotation was done by scanning the ORFs against RefSef proteins using Hidden Markov Models (HMMs). GeneMarkS-2+ subsequently made *ab initio* predictions for coding regions for proteins that lack HMM. Prediction of the protein-coding genes (CDSs) was

performed by Prodigal v.2.6.3 (Hyatt et al., 2010). Functional assignments of predicted coding proteins were scanned against models such as HMMs, BlastRules, and domain architectures. Proteins were then assigned the name, clusters, and attributes of the highest hits to the protein family model. Prediction of putative lignocellulolytic enzymes was done by scanning the CAZy database with dbCAN2 meta server (Zhang et al., 2018) utilizing three models: HMMER v10.0 (E-value <1e-15, coverage >0.35) (Finn et al., 2011), eCAMI (important_k_mer_number >=5, k_mer size = 8) (Xu et al., 2020) and DIAMOND (E-value <1e-102) (Buchfink et al., 2014). Similarity searches of the deduced proteins were performed against RefSeq, a non-redundant protein database of the NCBI. Genes encoding for environmental adaptation were inspected manually.

Taxonomic assignment was performed using the pairwise sequence similarities (Meier-Kolthoff et al., 2013) for the 16S rRNA gene available in the genome-to-genome distance calculator (GGDC) web server hosted at Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ) (http://ggdc.dsmz.de/) (Meier-Kolthoff et al., 2022). To assess the relatedness between the sequenced and publicly available genomes, genomes of type strains available from the Type Strain Genome Server (TYGS) (Meier-Kolthoff & Göker, 2019) were downloaded from NCBI. Average nucleotide identity (ANI) was calculated using the ANIm algorithm of pyani v0.2.11 (http://widdowquinn.github.io/pyani/) (Arahal, 2014; Richter & Rosselló-Móra, 2009). Briefly, assembled contigs and the genomes of the related type strains previously downloaded from NCBI GenBank were loaded in pyani platform. Using the ANIm algorithm, each genome was compared to all other genomes, and the percentage identity was generated based on the aligned regions.

Whole-genome comparisons with reference-type strain genomes were done using BLASTn Ring Image Generator (BRIG) v0.95 (Alikhan et al., 2011) operated under NCBI blast v2.13.0. Protein sequences in this study and their respective type strains were pooled into the BRIG platform. Parameters were then set to differentiate the genomes including reference genome, legend color, and upper and lower identity threshold. The software then formatted Genbank files, ran BLAST, parsed the results, and generated the comparison images. The composition of orthologous gene clusters for whole genomes was performed in the OrthoVenn2 web server (Xu et al., 2019).

Comparative analysis of the Carbohydrate-Active enZymes (CAZymes) genes was done as described above.

The whole-genome sequencing projects for *Shouchella* LMS6, *Evansella* LMS18, *Salipaludibacillus* LMS25, and *Alkalihalobacterium* LMS39 have been submitted to GenBank under the accession numbers CP093302, CP093301, CP093299, and CP093300, and Bioproject accession numbers PRJNA813598, PRJNA813595, PRJNA813589, and PRJNA813594, respectively. The Illumina raw reads were deposited in the Sequence Read Archive (SRA) of NCBI under the accession numbers SRR18532844 (LMS6), SRR18516216 (LMS18), SRR18516323 (LMS25), and SRR18516662 (LMS39). Raw Nanopore reads were deposited under accession numbers SRR18532843 (LMS6), SRR18516215 (LMS18), SRR18516322 (LMS25), and SRR18516661 (LMS25).

3.4 Results

3.4.1 Genotypic and evolutionary relationship of polysaccharide utilizing bacteria

A total of 89 isolates were recovered and 40 of these isolates were selected for 16S rRNA gene analysis based on their ability to grow in the presence of at least one of the tested polysaccharides. The results revealed that 22 isolates were affiliated with the genera *Salipaludibacillus*, 5 with *Halomonas*, 3 with *Alkalibacterium*, 3 with *Alkalihalophilus*, 2 with *Evansella*, 1 with *Shouchella*, 1 with *Halalkalibacterium*, 1 with *Halalkalibacter*, 1 with *Alkalibacterium*, and 1 with *Salinicoccus* (Table 3.2; Appendix 1). The genus *Salipaludibacillus* comprised isolates associated with *Sb. Agaradhaerens* (98-99% identity). The genus *Halomonas* comprised of isolates related to *Hl. Mongoliensis* (99% identity), *Hl. Lactosivorans* (98-99% identity), and *Hl. Zhaodongensis* (99% identity). Isolates of the genus *Alkalibacterium* were affiliated with *Ak. pelagium* (99% identity) and *Ak. psychrotolerans* (99%). Members of the genus *Alkalihalophilus* were comprised of isolates associated with *Ap. pseudofirmus* (99% identity), with the *Evansella* group being represented by relatives of *Ev. Clarkii* (98-99% identity). The genus *Shouchella* comprised of a relative of *So. Lehensis* (98% identity), while *Halalkalibacterium* comprised a relative of *Hm. Halodurans* (99% identity). In addition, in each case, one isolate affiliated to *Hb.*

krulwichiae (99% identity), *Ab*. *bogoriense* (97% identity) and *Sc*. *Alkaliphilus* (99% identity) was obtained.

Salipaludibacillus affiliated isolates were distributed across all sites, except the microbial mats (Table 3.2; Appendix 2). Overall, the grass soils had the highest diversity of retrieved bacterial isolates, followed by wet sediments. A low number of isolates were retrieved from dry sediments (4), brine (2), and microbial mats (2). Only two isolates belonging to *Alkalibacterium* and *Alkalihalophilus* were retrieved from microbial mats.

Table 3.2. Taxonomic assignment of 16S rRNA gene sequences of 40 bacterial isolates isolated from Lake Magadi.

Isolate	**Closest relative**	**S***	**M****	**Similarity (%)**	**Accession#**
LMS6	*Shouchella lehensis* MLB2	S1	M3	98	NR036940
LMS9	*Halalkalibacterium halodurans* DSM497	S4	M5	99	NR025446
LMS16	*Halalkalibacter krulwichiae* NBRC102362	S4	M1	99	NR114066
LMS19	*Alkalihalophilus pseudofirmus* DSM8715	S1	M4	99	NR026139
LMS20	*Alkalihalophilus pseudofirmus* DSM8715	S4	M4	99	NR026139
LMS31	*Alkalihalophilus pseudofirmus* DSM8715	S5	M2	99	NR026139
LMS39	*Alkalihalobacterium bogoriense* LBB3	S4	M4	97	NR042894
LMS1	*Salipaludibacillus agaradhaerens* DSM8721	S3	M5	99	NR026142
LMS2	*Salipaludibacillus agaradhaerens* DSM8721	S4	M1	99	NR026142
LMS5	*Salipaludibacillus agaradhaerens* DSM8721	S3	M3	99	NR026142
LMS7	*Salipaludibacillus agaradhaerens* DSM8721	S1	M4	99	NR026142
LMS8	*Salipaludibacillus agaradhaerens* DSM8721	S1	M3	99	NR026142
LMS10	*Salipaludibacillus agaradhaerens* DSM8721	S2	M3	99	NR026142
LMS12	*Salipaludibacillus agaradhaerens* DSM8721	S4	M3	99	NR026142
LMS14	*Salipaludibacillus agaradhaerens* DSM8721	S4	M3	99	NR026142
LMS17	*Salipaludibacillus agaradhaerens* DSM8721	S1	M3	99	NR026142
LMS21	*Salipaludibacillus agaradhaerens* DSM8721	S1	M5	99	NR026142
LMS22	*Salipaludibacillus agaradhaerens* DSM8721	S1	M5	99	NR026142
LMS23	*Salipaludibacillus agaradhaerens* DSM8721	S1	M5	99	NR026142

LMS24	*Salipaludibacillus agaradhaerens* DSM8721	S1	M4	99	NR026142
LMS25	*Salipaludibacillus agaradhaerens* DSM8721	S1	M2	98	NR026142
LMS26	*Salipaludibacillus agaradhaerens* DSM8721	S1	M6	99	NR026142
LMS27	*Salipaludibacillus agaradhaerens* DSM8721	S4	M3	99	NR026142
LMS28	*Salipaludibacillus agaradhaerens* DSM8721	S4	M3	99	NR026142
LMS29	*Salipaludibacillus agaradhaerens* DSM8721	S4	M1	99	NR026142
LMS32	*Salipaludibacillus agaradhaerens* DSM8721	S2	M6	99	NR026142
LMS33	*Salipaludibacillus agaradhaerens* DSM8721	S3	M4	99	NR026142
LMS37	*Salipaludibacillus agaradhaerens* DSM8721	S4	M3	99	NR026142
LMS40	*Salipaludibacillus agaradhaerens* DSM8721	S4	M3	99	NR026142
LMS3	*Halomonas mongoliensis* Z-7009	S3	M1	99	NR043298
LMS4	*Halomonas mongoliensis* Z-7009	S1	M1	99	NR043298
LMS11	*Halomonas lactosivorans* CFH 90008	S4	M1	98	NR044100
LMS13	*Halomonas lactosivorans* CFH 90008	S4	M1	99	NR163654
LMS35	*Halomonas zhaodongensis* NEAU	S4	M2	99	NR125612
LMS15	*Alkalibacterium pelagium* NBRC103242	S4	M6	99	NR114241
LMS34	*Alkalibacterium pelagium* NBRC103242	S4	M4	99	NR114241
LMS38	*Alkalibacterium psychrotolerans*	S5	M6	99	NR112659
LMS18	*Evansella clarkii* YN-1	S1	M2	98	NR041571
LMS36	*Evansella clarkii* DSM8720	S4	M2	99	NR026141
LMS30	*Salinicoccus alkaliphilus* DSM16010	S4	M3	99	NR025108

Site*: S1 = wet sediments, S2 = brine sediments, S3 = dry sediments, S4 = Magadi grass soils, and S5 = microbial mats; Accession number#: Gene bank accession number. Media** represents the isolation media indicated in Table 3.1.

The phylogenetic analysis of the 16S rRNA gene sequences from the 40 isolates revealed that the isolates clustered together with related type strains from the GenBank database. The isolates clustered based on their genera into nine distinct clades with bootstrap support values of 63% for *Salipaludibacllus* and 100% for most other groups (Figure 3.1). Groups I to IX comprised the isolates of *Salipaludibacillus*, *Evansella*, *Shouchella Alkalihalophilus*, *Alkalihalobacterium*, *Halalkalibacter*, *Alkalibacterium*, *Salinicoccus*, and *Halomonas*, respectively.

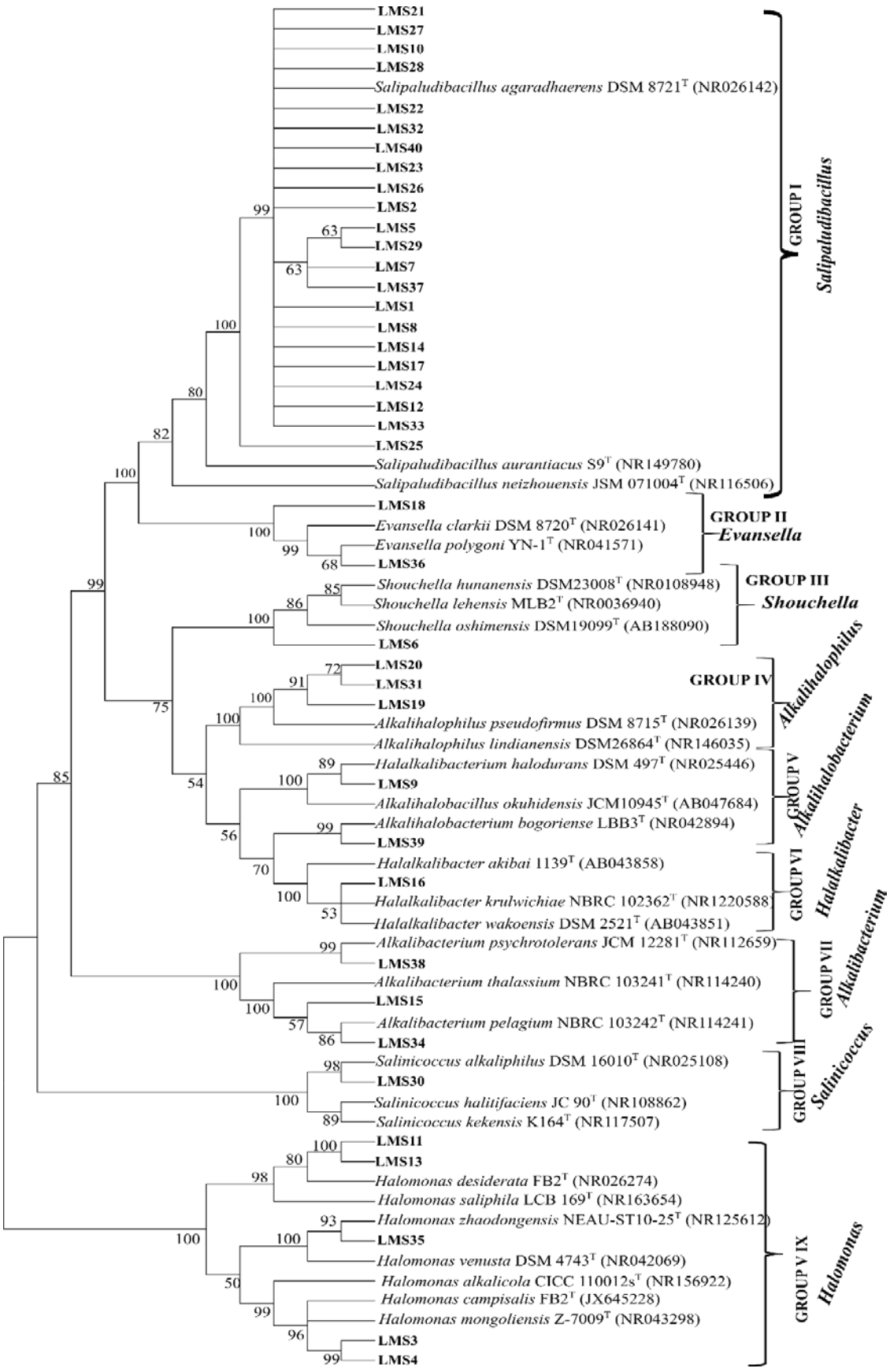


Figure 3.1: 16S rRNA gene sequence-based phylogenetic tree of the 40 bacterial isolates from Lake Magadi and related type strains.

The isolates from this study are indicated in bold. The evolutionary tree was inferred using the Maximum Likelihood method based on the Tamura-3 parameter model. The bootstrap values for 1,000 replicates are shown at the top of the nodes. Each node represents a common ancestor, whereas the branch length denotes the path of evolution from the shared ancestor. The scale bar (0.005) indicates the number of nucleotide substitutions per site.

3.4.2 Polysaccharide-degrading activity and physiological characterization

Most isolates exhibited notable hydrolytic activities (activity index of ≥ 2.0; Figures 3.2 and 3.3). The highest enzyme activity was recorded among the isolates belonging to the *Salipaludibacillus* while the least activity was recorded in isolates belonging to *Halomonas*. Isolates LMS18 and LMS36 affiliated to *Ev. Clarkii* (98 and 99%, respectively) LMS20 related to *Ap. pseudofirmus* (99%) and LMS30 affiliated to *Sc. alkaliphilus* (99%) showed no activity against all the tested substrates. Isolates LMS6 affiliated to *So. lehensis* (98%), and LMS34 affiliated to *Ak. pelagium* (99%) only utilized cellulose and CMC under the test conditions. Isolates affiliated with *Sb. Agaradhaerens* (98-99%) could utilize at least five of the six tested substrates. Members of the genus *Halomonas* demonstrated different abilities in the utilization of polysaccharides. LMS13, an isolate affiliated with *Hl. lactosivorans* demonstrated activity against all the selected substrates except pectin. Isolates related to *Hl. mongoliensis* (LMS3 and LMS4) (99%) showed no activity against any of the substrates, while LMS35 showed activity against starch only (Appendix 3). Growth tests indicated that the isolates had high OD_{600} values after 24 h at pH 7 – 9 (0.45 – 0.77), 40°C (0.92), and 5% NaCl concentrations (0.419) (Appendices 4 - 6).

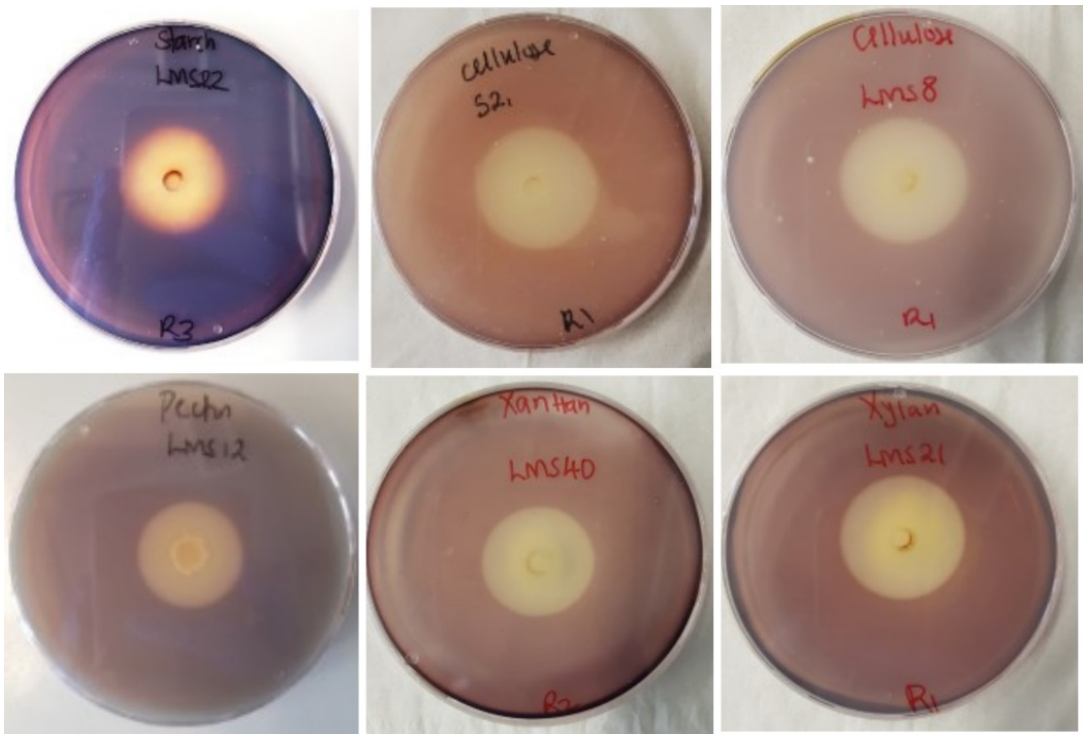


Figure 3.2: The enzyme activity of selected bacterial isolates with starch, cellulose, pectin, xanthan, and xylan, respectively. As an example, the activity of isolates LMS22, LMS2, LMS8, LMS12, LMS40, and LMS21 are shown. The halos indicate the activity of the bacterial isolates against the tested substrate. Halos were visualized by staining with Lugol's solution. The ratio of the hydrolysis zone and the colony zones was used to determine the enzyme activity indices (EAI).

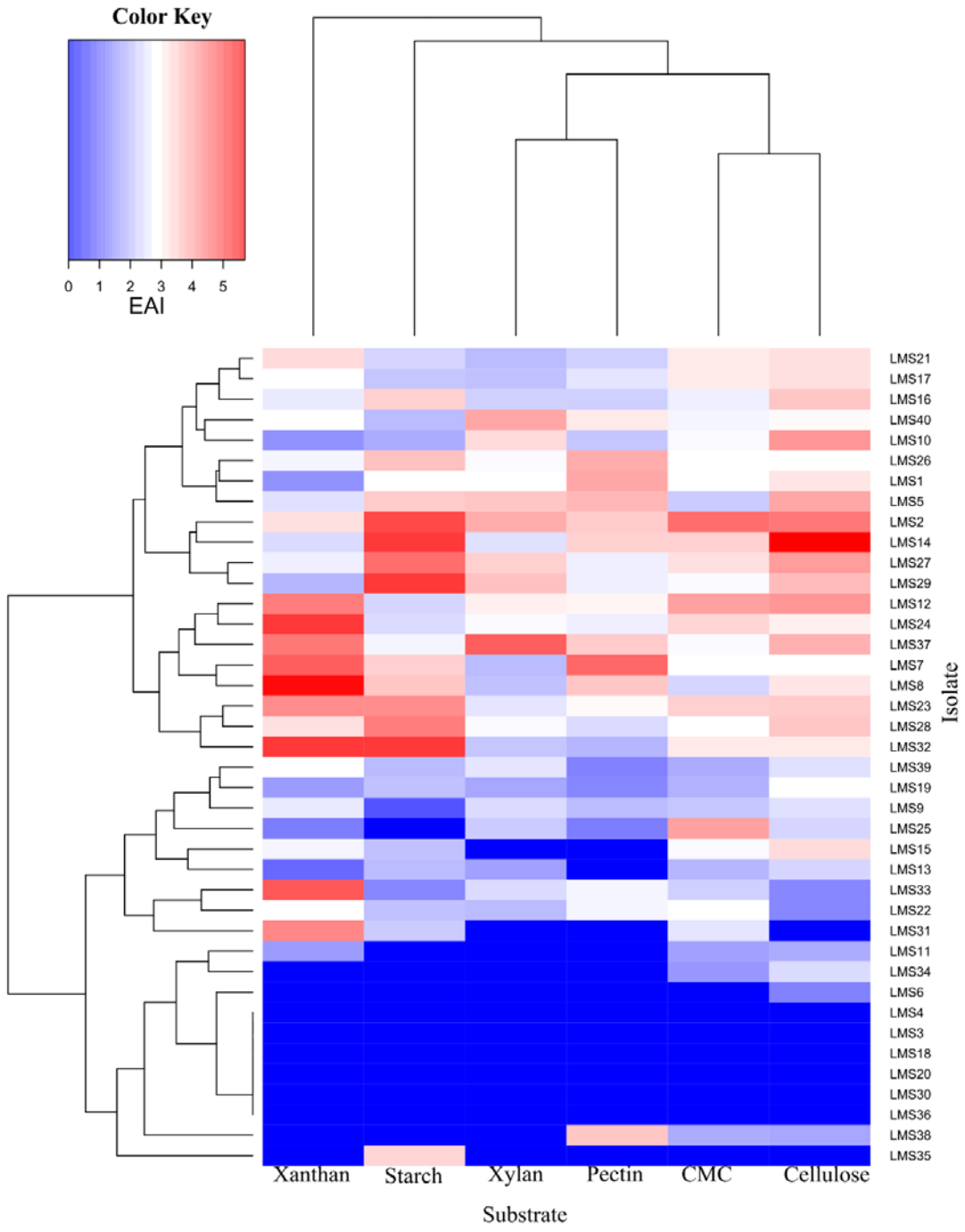


Figure 3.3: Substrate utilization of the bacterial isolates sampled from Lake Magadi. The generated dendrogram on the left side of the heatmap shows the clustering of the isolates relative to enzyme activity indices. The top dendrogram depicts clustering based on substrates. The scale (color key) indicates the average enzyme activity indices as depicted in the heatmap, where 1 denotes low enzyme activity index (EAI) while 5 denotes high EAI.

3.4.3 Genomic features and average nucleotide identity analyses

Whole-genome sequencing resulted in single chromosomes of isolates LMS6, LMS18, LMS25, and LMS39. Their genomic sizes were: LMS6 (3,886,007 bp), LMS18 (4,903,744, 4,525,667 bp), LMS25 (4,525,667 bp), and LMS39 (4,850,562 bp). Furthermore, the G+C content was lowest in LMS39 (37.0%) and highest in LMS18 (43.0%). CRISPR arrays were detected in the genomes of LMS6 and LMS25 (Table 3.3).

Table 3.3. Characteristic features of the four bacterial genomes derived from isolates LMS6, LMS18, LMS25, and LMS39.

Genome information	LMS6	LMS18	LMS25	LMS39
Sequencing coverage (Illumina)	120X	103X	126X	114X
sequencing coverage (Nanopore)	533X	147X	393X	771X
Size (bp)	3,886,007	4,903,744	4,525,667	4,850,562
Number of contigs	1	1	1	1
G+C content (%)	40.8	43.0	39.0	37.0
Number of putative genes	4,175	4,688	4,253	4,749
Protein coding sequences (CDSs)	4,040	4,500	4,020	4,571
Number of tRNAs	76	107	74	89
Number of rRNAs	25	30	25	30
CRISPR arrays	1	0	3	0

The analysis of the average nucleotide identity (ANI) (Figure 3.4) indicated that LMS6 was associated with the *Shouchella* group with 84.6% similarity to *So. gibsonii* DSM8722^T (ASM1679824v1 (Figure 3.4a). LMS18 was associated with the *Evansella* with 91.7% similarity to *Ev. clarkii* DSM 8720^T (NZ_MTIV00000000) (Figure 3.4b). The genome of LMS25 genome was associated with *Sb. agaradhaerens* DSM 8721^T (ASM201973v1) with an 88.9% ANI value (Figure 3.4c), while LMS39 was associated with *Ab. bogoriense* ATCC BAA-922^T (ASM62144v1) with 92.8% similarity (Figure 3.4d). All ANI values were below the species delineation boundary of 95 - 96% for reliable classification, indicating that these isolates belong to new species of their respective genera. Compared with their closest related type strain, the genomic size for LMS25 was larger whereas genomes of LMS6, LMS18, and LMS39 were comparatively smaller. Compared with their

closest type strain genomes, the genomic size for LMS25 was larger, whereas the genomes of LMS6, LMS18, and LMS39 were comparatively smaller (Appendix 7).

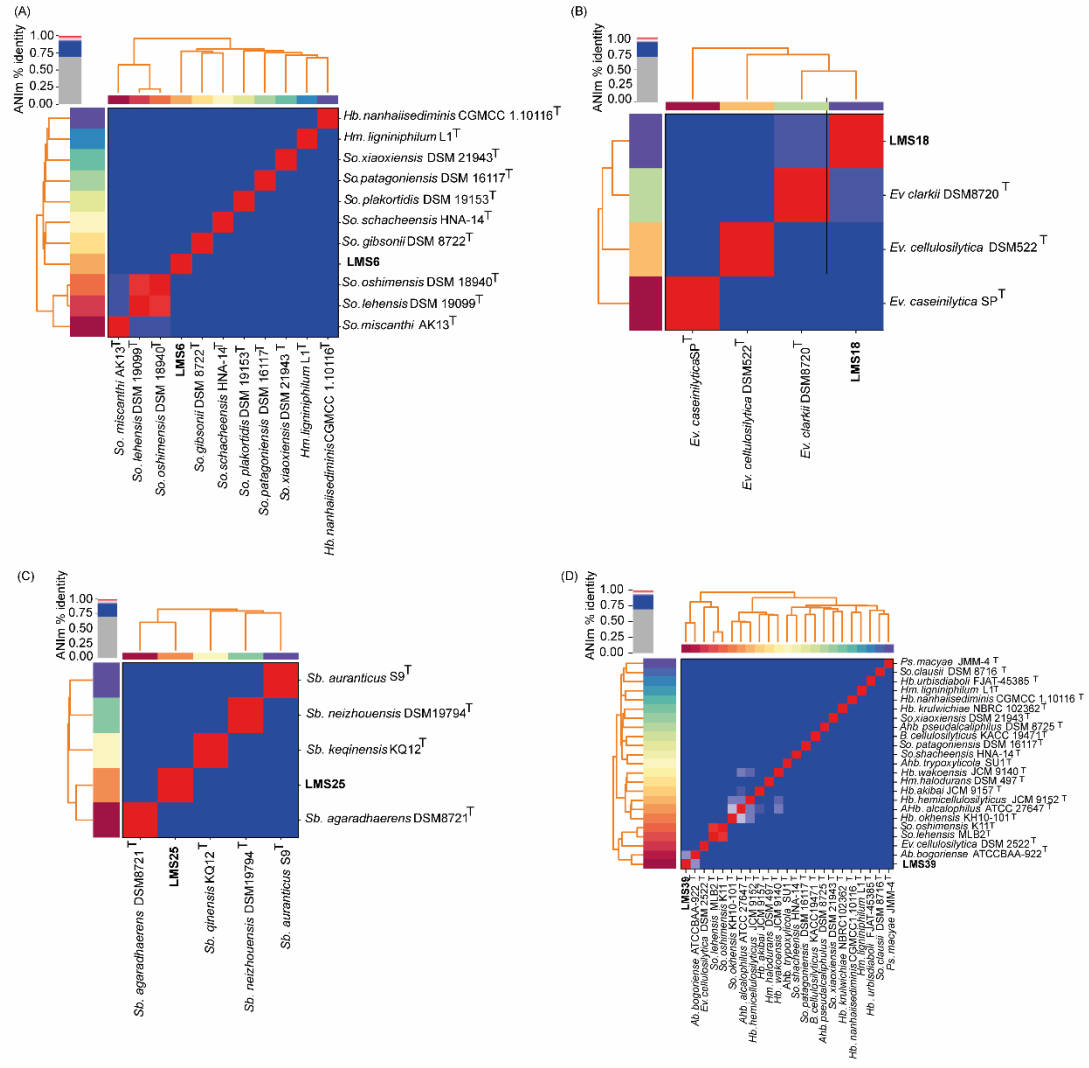


Figure 3.4: Analysis of the average nucleotide identity (ANI) for the whole genomes of the selected isolates. LMS6 (A), LMS18 (B), LMS25 (C), and LMS39 (D). All the available and sequenced type strains from the representative genera were included. Calculations were done with pyani software using the ANIm algorithm.

3.4.4.1 Comparative genomics of *Shouchella* sp. LMS6

In total, 3,214, 3,304, and 2,542 orthologous gene clusters were found in LMS6, *So. lehensis*, and *So. gibsonii*, respectively (Table 3.4).

Table 3.4: Analysis of protein-coding genes (CDSs), orthologous gene clusters, and singletons of the LMS6, LMS18, LMS25, LMS39, and their closest type strains.

Strain	CDSs	Clusters	Singletons
LMS6	4040	3214	748
So. lehensis (DSM19099^T)	3894	3304	534
So. gibsonii (DSM8722^T)	3904	2542	1258
LMS18	4500	4039	412
Ev. clarkii (DSM8720^T)	5109	4129	884
Ev. cellulosilytica (DSM522^T)	4266	2713	1378
LMS25	4020	3340	548
Sb. agaradhaerens (DSM8721^T)	3839	3386	414
Sb. keqinensis (KQ12^T)	3941	2643	1185
LMS39	4571	4182	344
Ab. bogoriense (ATCC BAA-922^T)	4752	4237	451
B. alkalicellulosilyticus (FJAT-44921^T)	4552	3094	1314

BRIG-BLAST analysis of LMS6 and its relatives suggested high genomic similarity between LMS6 and *So. lehensis* (Figure 3.5). A total of 2,877 clusters were common among the three genomes. In addition, 748, 534, and 1,258 singletons were found in LMS6, *So. lehensis*, and *So. gibsonii*, respectively. Furthermore, the highest number of orthologous clusters were found between LMS6 and *So. lehensis* DSM19099^T (901 common clusters, and 28 and 22 clusters specific for LMS6 and *So. lehensis*, respectively).

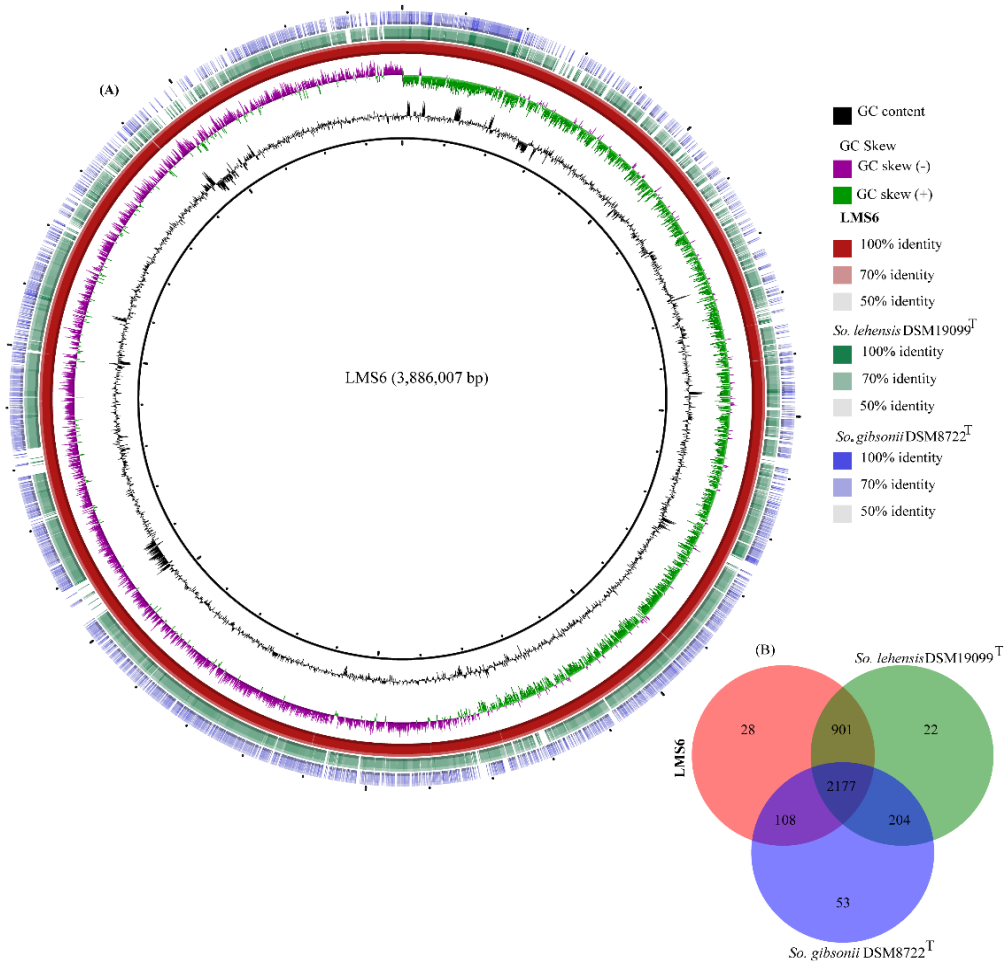


Figure 3.5: Comparative genomics of LMS6 and related type strains. **A**: Alignment of complete genomes of *Shouchella* strain LMS6 (reference) and two draft genomes of *So. lehensis* DSM19099^T and *So. gibsonii* DSM8722^T. BLASTn matches between the query and the reference genomes are shown as concentric colored rings on a sliding scale according to percentage identity (100%, 70%, or 50%). The rings indicate, from the center outwards: GC content (black ring); GC skew (green and purple ring); LMS6 (deep red ring); *So. lehensis* DSM19099^T (green ring), and *So. gibsonii* DSM8722^T (blue ring). **B**: Comparison of orthologous gene clusters among LMS6, *So. lehensis* DSM19099^T, and *So. gibsonii* DSM8722^T genomes.

3.4.4.2 Comparative genomics of *Evansella* sp. LMS18

A genomic comparison between LMS18 and its relatives is shown in (Figure 3.6). The results show that compared to *Ev. cellulosilytica* DSM522^T, LMS18 and *Ev. clarkii* DSM8720^T shares more genomic areas (Figure 3.6a). Furthermore, a total of

4,039, 4,129, and 2,713 orthologous gene clusters were found in the LMS18, *Ev. clarkii* DSM8720^T (ASM201969v1), and *Ev. cellulosilytica* DSM522^T (ASM17723v2) genomes, respectively. Overall, the three genomes shared 2,449 gene clusters (Figure 3.6b). A total of 1,568 gene clusters were solely found in LMS18 and *Ev. clarkii* (1,510 shared and 16 or 42 that are specific to each strain). In addition, 412, 884, and 1,378 singletons were found in LMS18, *Ev. clarkii*, and *Ev. cellulosilytica*, respectively (Table 3.4). 2,449 gene clusters were shared among the three genomes.

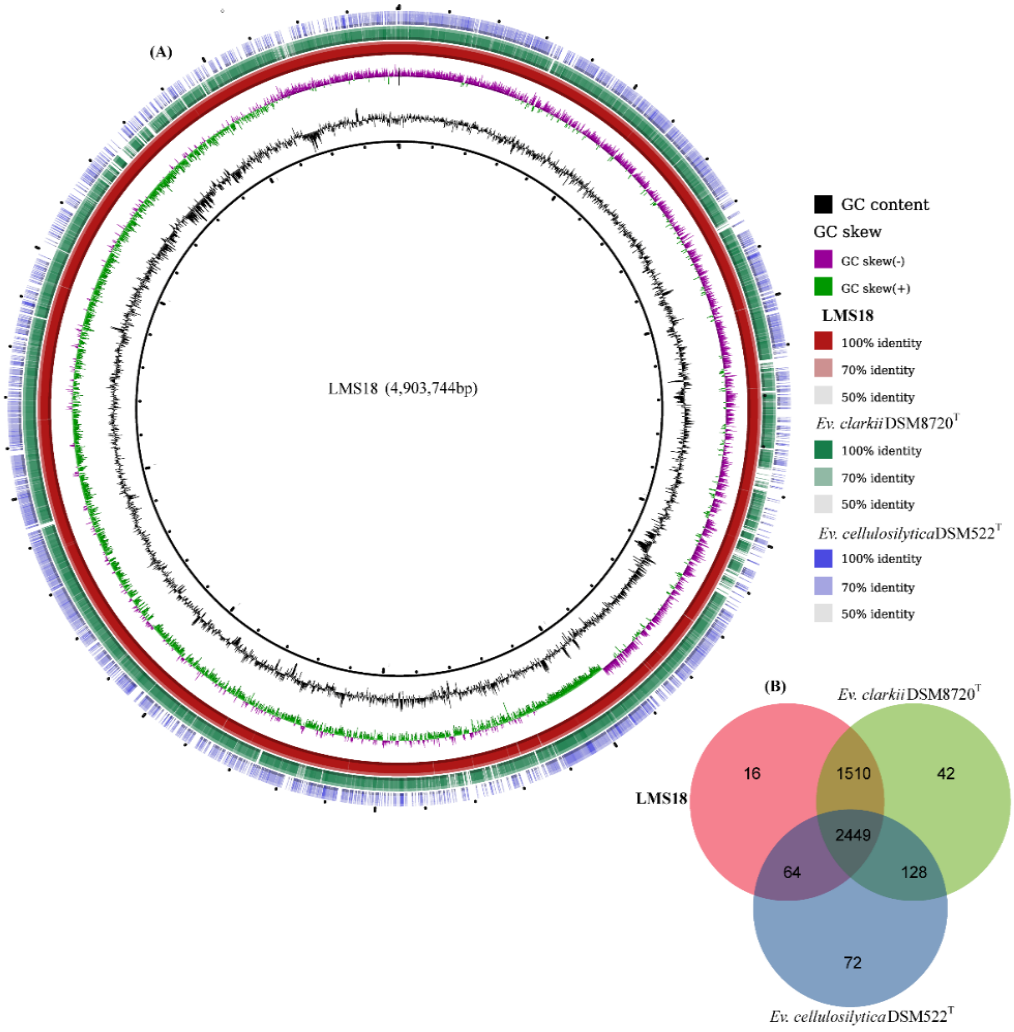


Figure 3.6: Comparative genomics of LMS18 and related type strains. **A**: Alignment of complete genomes of *Evansella* strain LMS18 and two draft genomes of *Ev. clarkii* DSM8720^T (reference) and *Ev. cellulosilytica* DSM522^T. BLASTn matches between the query and the reference genomes are shown as concentric colored rings on a sliding scale according to percentage similarity (100%, 70%, or 50%). The rings indicate, from the center outwards: GC content (black ring); GC skew (green and

purple ring); *Ev. clarkii* DSM8720^{T} (red ring); LMS18 (deep blue ring) and *Ev. cellulosilytica* DSM522^{T} (light blue ring). **B**: Comparison of the cluster of orthologous proteins among LMS18, *Ev. clarkii* DSM8720^{T,} and *Ev. cellulosilytica* DSM522^{T} genomes.

3.4.4.3 Comparative genomics of *Salipaludibacillus* sp. LMS25

Analysis of the genomes of LMS25, *Sb. agaradhaerens* DSM8721^{T} (ASM201973v1), and *Sb. keqinensis* KQ12^{T} (ASM322632v1) (Figure 3.7) revealed highly similar regions between the genomes of LMS25 and *Sb. agaradhaerens* DSM8721^{T} (Figure 3.7a). A total of 2,409 orthologous gene clusters were shared between them (Figure 3.7b). Furthermore, the highest number of gene clusters was found between LMS25 and *Sb. agaradhaerens* with 901 gene clusters (837 shared, and 43 and 21 genes solely found in LMS25 and *Sb. agaradhaerens* DSM8721^{T}, respectively). In addition, 548, 414, and 1,185 singletons were found in LMS25, *Sb. agaradhaerens* DSM8721^{T}, and *Sb. keqinensis* KQ12^{T}, respectively (Table 3.4).

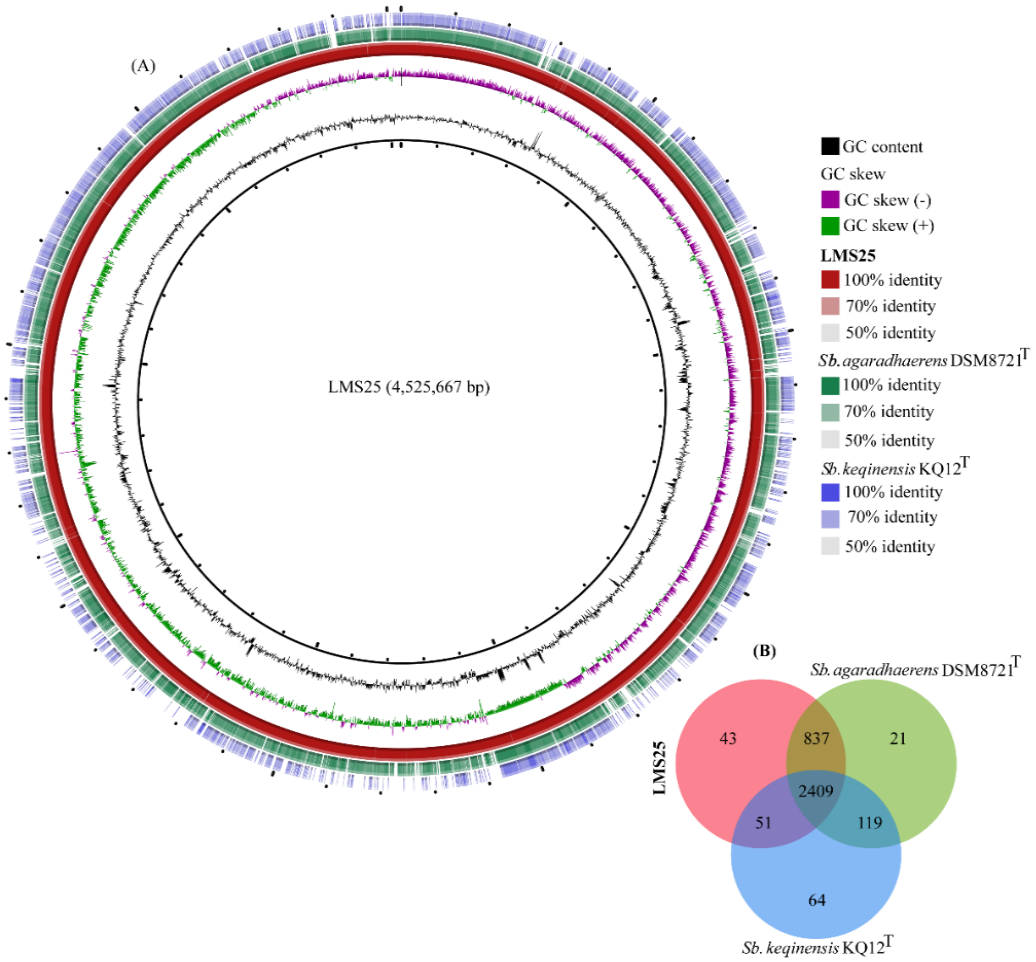


Figure 3.7: Comparative genomics of LMS25 and related type strains. Alignment of complete genomes of *Salipaludibacillus* strain LMS25 (reference) and two draft genomes of *Sb. agaradhaerens* DSM8721^T and *Sb. keqinensis* KQ12^T. BLASTn matches between the query and the reference genomes are shown as concentric colored rings on a sliding scale according to percentage identity (100%, 70%, or 50%). The rings indicate, from the center outwards: GC content (black ring); GC skew (green and purple ring); LMS25 (deep red ring); *Sb. agaradhaerens* DSM8721^T (deep green ring) and *Sb. keqinensis* KQ12^T (blue ring). **B**: Comparison of orthologous gene clusters among LMS25, *Sb. agaradhaerens* DSM8721^T and *Sb. keqinensis* KQ12^T genomes.

3.4.4.4 Comparative genomics of *Alkalihalobacterium* sp. LMS39

A comparison of the genome of LMS39 with its relatives (*Ab. bogoriense* ATCC BAA-922^T and *Bacillus alkalicellulosilyticus* FJAT-44921^T) (Figure 3.8) showed a high degree of similar regions between LMS39 and *Ab. bogoriense* ATCC BAA-922^T (Figure 3.8a). In total, 2,877 orthologous gene clusters were common among the three genomes (Figure 3.8b). In addition, 344, 451, and 1,314 singletons were found in

LMS39, *Ab. bogoriense*, and *B. alkalicellulosilyticus*, respectively (Table 3.4). Furthermore, 1,281 gene clusters were found between LMS39 and *Ab. bogoriense* (1,240 common, and 12 and 29 clusters unique to LMS39 and *Ab. bogoriense*, respectively).

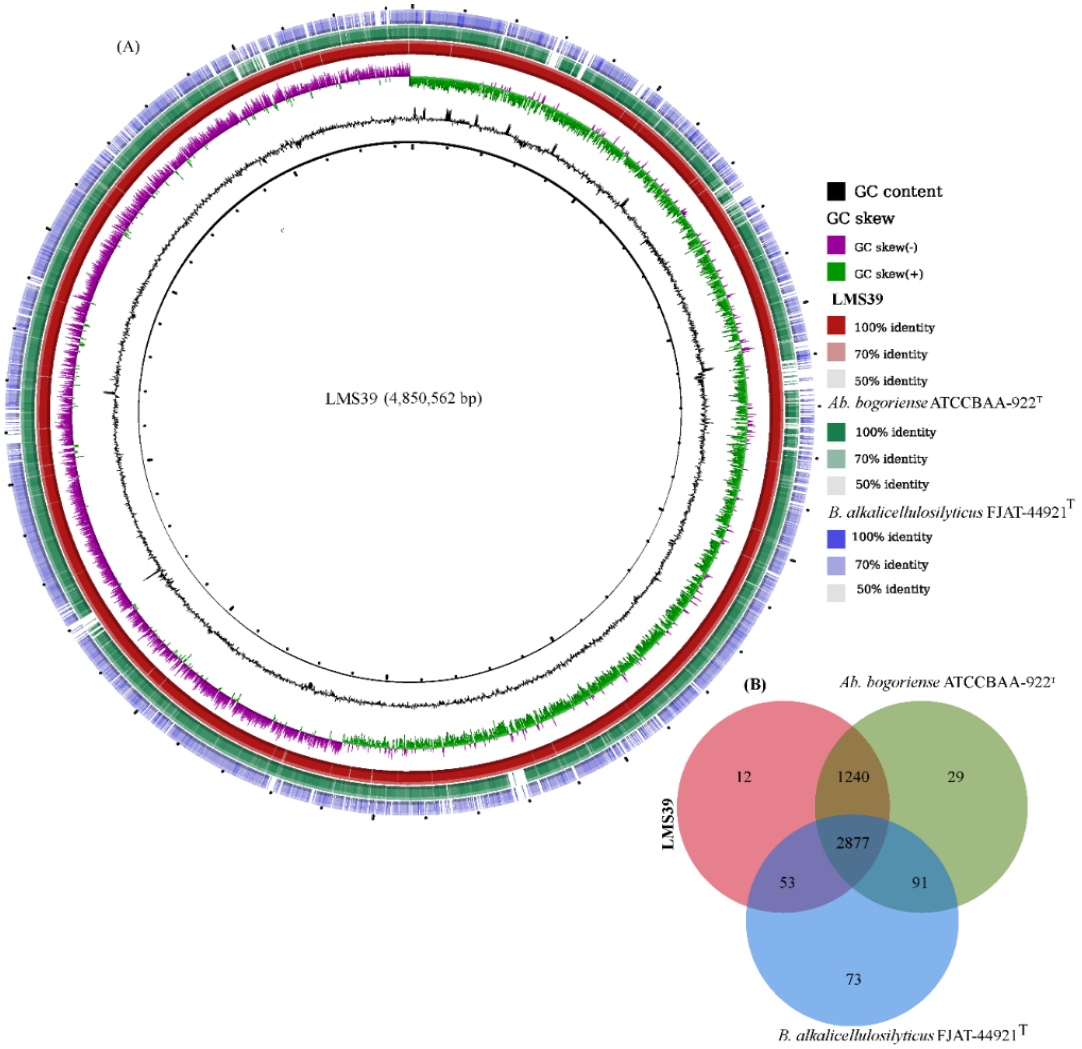


Figure 3.8: Comparative genomics of LMS39 and the reference type strain. **A:** comparison of complete genomes of *Alkalihalobaterium* strain LMS39 and a draft genome of *Ab. bogoriense* ATCC BAA-922^T and *B. alkalicellulosilyticus* FJAT-44921^T. BLASTn matches between the query and the reference genomes are shown as concentric colored rings on a sliding scale according to percentage identity (100%, 70%, or 50%). The rings indicate, from the center outwards: GC content (black ring); GC skew (green and purple ring); LMS39 (red ring); *Ab. bogoriense* ATCC BAA-922^T (green ring), and *B. alkalicellulosilyticus* (blue ring) **B**: Comparison of orthologous gene clusters among LMS39, *Ab. bogoriense*, and *B. alkalicellulosilyticus*.

3.4.5 Comparative analysis of putative CAZyme genes

All four genomes harbor putative CAZyme-encoding genes for glycoside hydrolases (GH), polysaccharide lyase (PL), carbohydrate esterases (CE), glycosyltransferases (GT), and carbohydrate-binding modules (CBM). Compared with the closest type strains, there was no notable divergence in the number of genes encoding a particular class of CAZymes (Figure 3.9). The dominant class across the genomes was the glycoside hydrolases group followed by glycosyltransferases, while polysaccharide lyase was the least recorded. The genes encoding glycoside hydrolases were highest in the LMS39 genome (69), followed by LMS25 (66), LMS6 (40), and LMS18 (19), respectively.

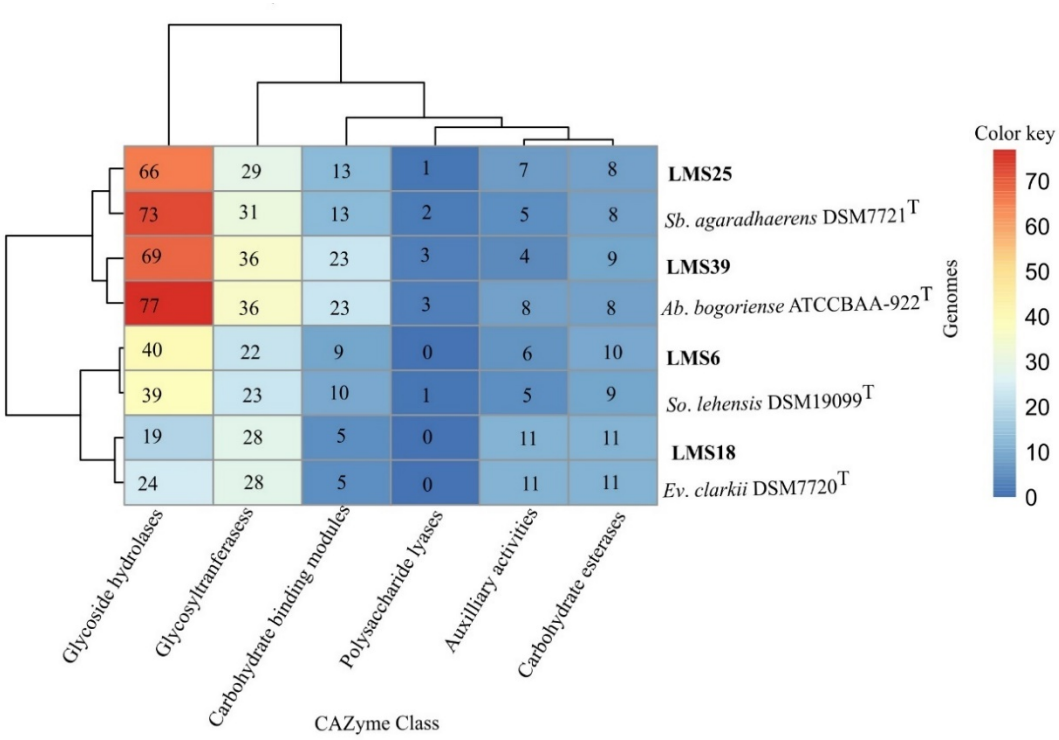


Figure 3.9: Comparison of CAZyme-encoding genes among the LMS6, LMS18, LMS25, and LMS39 genomes (bold) and their closest type strains. High representation (red) and low representation (blue) modules are depicted based on the number of CAZyme encoding genes in each genome.

The genes encoding polysaccharide depolymerizing enzymes included cellulases from genomes of isolates LMS6, LMS25, and LMS39. A chitinase-encoding gene was found in LMS39. Additionally, xylanase-encoding genes were established from LMS6, LMS25, and LMS39 genomes. Pectinase-encoding genes were found in LMS25 and LMS39 while those encoding amylases were found in LMS6 and LMS25 (Table 3.5).

Table 3.5: Products of the genes associated with carbohydrate metabolism and their amino acid sequence similarities in the genomes of LMS6, LMS18, LMS25, and LMS39.

GeneBank accession	Source	No. of amino acids	CAZyme class*	Closest identical CAZyme protein, accession no., no. of encoded amino acids, organism, E-value*	Blast identity (%)
UTR06916	LMS6	557	GH13	α-phosphotrehalase, WP_099302917.1, 557, *Bacillus* Marseille-P3800	100
UTR08188	LMS6	400	GH5	cellulase, WP_218014778.1, 400, *Bacillus* Marseille-P3800	97
UTR06965	LMS6	438	GH4	6-phospho-beta-glucosidase, WP_099302829.1, 438, *Bacillus* Marseille	94
UTR07232	LMS6	515	GH3	β-N-acetylhexosaminidase, WP_099302370.1, 515, *Bacillus* Marseille	99
UTR06199	LMS5	490	GH13	α-amylase, WP_176554439.1, 490, *Bacillus* Marseille-P3800	99
UTR07249	LMS6	581	GH13	GH13 family protein, WP_099302352.1,581, *Bacillus* Marseille-P3800	99
UTR05798	LMS6	433	GH4	α-glucosidase/α-galactosidase, WP_099303665.1, 433, *Bacillus* Marseille-	99
UTR06207	LMS6	579	GH2	GH2 family protein, WP_099304398.1, 579, *Bacillus* sp. Marseille-P3800	99
UTR11958	LMS18	388	GH10	endo-1,4-β-xylanase, WP_088032251.1, 388, *Ev. clarkii*	98
UTR11566	LMS18	554	GH13	α-glucosidase, WP_078594722.1, 554, *Ev. clarkii*	96
UTR10632	LMS18	706	GH13	type I pullulanase, WP_078598047.1, 706, *Ev. clarkii*	89
UTR13037	LMS18	699	GH3	GH family 3 protein, WP_088036931.1, 699, *Ev. clarkii*	95
UTR11836	LMS18	743	GH36	α-galactosidase, WP_078594153.1, 743, *Ev. clarkii*	96
UTR11835	LMS18	661	GH42	β-galactosidase, WP_078594155.1, 661, *Ev. clarkii*	97
UTR11903	LMS18	536	GH43	GH43 family, WP_088032287.1, 536, *Ev. clarkii*	97
UTR12973	LMS18	498	GH51	α-N-arabinofuranosidase, WP_078597734, 496, *Ev. clarkii*	94
UTR13087	LMS18	428	GH18	GH18 family WP_078595740.1, 428, *Ev. clarkii*	98
UTR15384	LMS25	712	GH13	α-amylase, WP_078579747.1, 712, *Sb. agaradhaerens*	95
UTR16423	LMS25	956	GH13	α-amylase, WP_078578793.1, 956, *Sb. agaradhaerens*	85
UTR15386	LMS25	587	GH13	α-glycosidase, WP_078579744.1, 587, *Sb. agaradhaerens*	91
UTR16833	LMS25	479	GH5	Cellulase, WP_078579508.1, 462, *Sb. agaradhaerens*	86
UTR13206	LMS25	528	GH13	α-amylase, WP_095995549.1, 527, *Sb. agaradhaerens*	98
UTR13210	LMS25	656	GH13	α-amylase, WP_078578233.1, 654, *Sb. agaradhaerens*	91
UTR16091	LMS25	720	GH36	α-galactosidase, WP_078579090.1, 720, *Sb. agaradhaerens*	90
UTR16092	LMS25	671	GH42	β-galactosidase, WP_245829098.1, 674, *Sb. agaradhaerens*	93

UTR16104	LMS25	438	GH4	α-glucosidase, WP_078579078.1, 433, *Sb. agaradhaerens*	92
UTR16961	LMS25	353	GH43	arabinan endo-1,5-α-L-arabinosidase, WP_078579005.1, 353, *Sb. agaradhaerens*	88
UTR16225	LMS25	493	GH51	α-N-arabinofuranosidase, WP_078578978.1, 493, *Sb. agaradhaerens*	92
UTR16965	LMS25	506	GH51	α-N-arabinofuranosidase, WP_078579970, 506, *Sb. agaradhaerens*	95
UTR13595	LMS25	774	GH31	α-xylosidase, WP_078577884.1, 774, *Sb. agaradhaerens*	88
UTR14054	LMS25	687	GH42	β-galactosidase, WP_078577494.1, 687, *Sb. agaradhaerens*	93
UOE94519	LMS39	405	GH10	endo-1,4-beta-xylanase, WP_026672600.1, 405, *Ab. bogoriense*	98
UOE96456	LMS39	529	GH18	GH18 family protein, WP_245613687.1, 513, *Ab. bogoriense*	96
UOE93572	LMS39	501	GH39	xylan 1,4-beta-xylosidase, WP_026671748.1, 501, *Ab. bogoriense*	92
UOE95130	LMS39	433	GH4	α-glucosidase/α-galactosidase, WP_250096557.1, 433, *Halalkalibacter* sp.	95
UOE96016	LMS39	448	GH1	β-glucosidase, WP_026674998.1, 448, *Ab. bogoriense*	99
UOE93553	LMS39	821	GH2	GH family 2 protein, WP_026671730.1, 821, *Ab. bogoriense*	90
UOE94950	LMS39	668	GH42	β-galactosidase, WP_035179913.1, 668, *Ab. bogoriense*	98
UOE93502	LMS39	711	GH13	type I pullulanase, WP_051556287.1, 711, *Ab. bogoriense*	96
UOE95362	LMS39	1145	GH13	pullulanase, WP_051556287.1, 1128, *Ab. bogoriense*	95
UOE92155	LMS39	1143	GH13	α-amylase, WP_051556325.1, 1143, *Ab. bogoriense*	93
UOE93565	LMS39	504	GH51	α-N-arabinofuranosidase, WP_026671741.1, 504, *Ab. bogoriense*	97
UOE96496	LMS39	496	GH51	α-N-arabinofuranosidase, WP_218971042.1, 496, *Ab. bogoriense*	96
UOE94753	LMS39	354	GH17	B-glucanase, WP_156323957.1, 353, *Bacillus* sp. JCM19034	79
UOE93577	LMS39	685	GH67	α-glucuronidase, WP_026671753.1, 686, *Ab. bogoriense*	90
UOE93610	LMS39	796	GH3	GH3 protein, WP_026671786.1, 796, *Ab. bogoriense*	92
UOE94060	LMS39	686	GH42	β-glucosidase, WP_026673793.1, 686, *Ab. bogoriense*	96

GH = glycoside hydrolase. The numerics indicate the protein family; E-value: The E-value for all the data was 0.0.

3.4.6 Putative genes encoding for soda lake adaptation

The analysis of the genes encoding for soda lake adaptation demonstrated that all genomes of the four isolates contained genes for survival in the soda lake ecosystem (Appendix 8). Genomic analyses revealed that all four genomes contained genes encoding for proteins for survival in hypersaline and alkaline ecosystems. They included Na+/H+ antiporter (NhaC), Glycine/betaine transporter (OpuD/betL), L-ectoine synthase (ectC), trehalose PTS system EIIBC or EIIBCA component (treB), K+ uptake system (TrkA), multicomponent Na+:H+ antiporter A-G (mnhA-mnhG), and Glycine betaine/carnitine transport binding protein (GbuC).

3.5 Discussion

3.5.1 Taxonomy of polysaccharides utilizing bacteria of Lake Magadi

Isolation efforts yielded isolates affiliated with the genus *Salipaludibacillus* as the dominant group. Moreover, *Salipaludibacillus* affiliates were isolated in all sites except microbial mats (S5). Members of this novel group are Gram-positive, rod-shaped, and endospore-forming (Sultanpuram and Mothe, 2016). They thrive and tolerate high pH, high temperature, moderate saline, hypersaline, and low oxygen environments such as soils and saline aquatic habitats (Kambura et al., 2016). Isolates of this genus have been obtained from other soda lakes (Sultanpuram and Mothe, 2016; Ibrahim et al., 2019; Wang et al., 2019), marine wetlands (Amoozegar et al., 2018), and an oilfield (Guo et al., 2023). Other isolates of the Gram-positive group included the affiliates of *Alkalihalophilus*, *Shouchella*, *Halalkalibacterium*, *Halalkalibacter*, *Alkalihalobacterium* (all previously named *Alkalihalobacillus*) (Patel & Gupta, 2020; Joshi et al., 2021), *Salinicoccus*, and *Alkalibacterium*. Relatives of *Alkalihalobacterium* have been frequently obtained from soda lake ecosystems (Vargas et al., 2005; Joshi et al., 2021). Affiliates of *Alkalihalophilus*, *Shouchella*, and *Halalkalibacterium* were originally described as halotolerant and alkaliphilic (Nielsen et al. 1995), while affiliates of *Halalkalibacter* have been isolated from soil (Yumoto et al., 2003) and Lake Lonar (Joshi et al., 2023). Members of the genus *Salinicoccus* have been found to inhabit solar salterns (Srinivas et al., 2016), salt mines, saline soils, fermented foods, and soda lakes (Hyun et al., 2013).

Non-spore-forming bacteria were solely represented by isolates of the genus *Halomonas*. *Halomonas* spp. have G+C content between 51-77%, and are

heterotrophic, moderately halophilic, and aerobic (Duckworth et al, 2000; Kim et al., 2013). Studies on the soda lakes of the Great Rift Valley, Kulundu Steppes (Russia), Mongolian soda soils and lakes (Mongolia), Wadi Natrun soda lake (Egypt), and hypersaline Playa (Iran), have shown the presence of members of this group, suggesting its adaptive ability to hypersaline and alkaline water and soil environments (Duckworth et al., 1996; Shapovalova et al., 2008; Babavalian et al., 2013). It is postulated that heterotrophic *Halomonas* sp. can utilize diverse substrates, which is the reason for their abundance across diverse environments (Naghoni et al., 2017; Boyadzhieva et al., 2018).

Grass soils (S4) showed the highest diversity of recovered polysaccharides utilizing bacteria belonging to *Salipaludibacillus*, *Shouchella*, *and Alkalibacterium*. Bacteria are highly active in nutrient recycling, mainly in the degradation of plant-derived polysaccharides such as lignocellulose and its phenolic derivatives (Soy et al., 2019). The proximity of the grass soils to the lake results in an exchange of nutrients and microbial guilds between the two environmental matrices, hence the rich diversity. On the other hand, microbial mats (S5) and brine (S2) yielded the lowest diversity of polysaccharide-utilizing bacteria in the recovered isolates with only *Salipaludibacillus* (phylum Firmicutes) being represented. However, a similar study using a culture-independent technique revealed that mats from Lake Magadi hot springs contained members of Firmicutes, Proteobacteria, and Verumicrobia (Kambura et al., 2016). The discordance could be attributed to the mechanism of isolation employed in this study. Microbial mats are sedimentary biofilm formed by a network of microbial filaments and extracellular slimes found mainly in hypersaline ponds, subtidal marine ponds, hot springs, lagoons, and freshwater lakes and rivers (Des Marais, 1990). Brine in soda lakes is predominated by soluble sodium salts formed through seepage of groundwater rich in sodium ions and carbon dioxide. This combination forms a buffered system and a rich habitat for alkalihalophilic bacteria (Sorokin et al., 2014; Vavourakis et al., 2016).

3.5.2 Polysaccharide degrading activity of bacterial isolates from Lake Magadi

Microbial degradation of complex polysaccharides and subsequent production of monosaccharides is crucial for the functioning of an ecosystem and the global carbon cycle (Berlemont et al., 2015). Most bacterial isolates exhibited high enzymatic

indices, suggesting proficiency in the utilization of the test polysaccharides. This study revealed that isolates associated with the genera *Alkalihalophilus*, *Shouchella*, *Halalkalibacterium*, *Halalkalibacter*, *Alkalihalobacterium*, *Salipaludibacillus*, and *Alkalibacterium* were identified as the most active against the starch, CMC, xylan xanthan, cellulose, and pectin. Particularly, isolates related to *Sb. agaradhaerens* and *Hm. halodurans* exhibited the highest enzymatic activity indices for cellulose, CMC, starch, xanthan, pectin, and xylan (Figure 3.3). According to earlier studies, members of these genera also exhibit catalytic activity against several polymers including lipids and potentially, long-chain fatty acids (Rattu et al., 2016; Kiplimo et al., 2019; Muazzam et al., 2019). Strains related to *Ap. pseudofirmus*, *So. lehensis*, *Ak. pelagium*, and *Ak. psychrotolerans* were shown to have low to mid-enzyme activity. This concurs with earlier reports which revealed that most of these organisms exhibited hydrolytic potential against xanthan, carboxymethylcellulose, and xylan (Mwirichia et al., 2010). Previous studies have reported that whole-genome sequencing of *Ap. pseudofirmus* GB1 revealed the presence of putative genes encoding carbohydrate-degrading enzymes (Manzo et al., 2011). *Ap. pseudofirmus* 703 demonstrated effective degradation of starch and amylopectin under alkaline pH conditions (Lu, et al., 2016). Furthermore, *Alkalibacterium* sp. SL3 is reported as a source of enzymes involved in the degradation of several complex compounds, including fatty acid esters, xylan, and starch (Lu et al., 2016; Wang et al., 2016; Wang et al., 2017).

Isolates affiliated with *Halomona*s exhibited discordant patterns in utilizing the tested polysaccharides. Only a few studies have documented the cellulose and starch degrading potential of *Halomonas* species, including *Halomonas* sp. PS47, *Halomona* sp. AAD21, and *Hl. meridiana* DSM 5425 from saline ecosystems (Coronado et al., 2000; Uzyol et al., 2012; Wang et al., 2019).

3.5.3 Physiological characterization of the bacterial isolates

Bacterial cell densities were high at pH 7 - 9, 40℃, and NaCl concentrations of 5-10% (w/v).The availability of carbon sources and the spatial-temporal variations are the main drivers of the diversity and abundance of bacteria in soda lakes (Hugoni et al., 2021; Omeroglu et al., 2021). Bacteria living in this ecosystem possess a negatively charged cell wall which reduces the charge density at the cell surface

hence stabilizing the cell membrane (Santos and Da Costa, 2002). Additionally, they possess Na^+/H^+ and K^+/H^+ antiporters which exchange cytoplasmic cations for protons outside of the cell to achieve lower cytoplasmic alkalinity than the external environment (Padan et al., 2005). Furthermore, haloalkaliphiles synthesize compatible solutes like betaine glycine, sugars (such as trehalose), polyols, amino acids, biosurfactants, and ectoines, which are involved in maintaining an isotonic environment (Bursy et al., 2008; Jayachandra et al., 2013).

The genomic analysis of LMS6, LMS18, LMS25, and LMS39 indicated that putative genes encoding for Na^+/H^+ and K^+/H^+ antiporters as well as K^+ uptake systems *trkA*, *trkG*, and *trkH* were present (Appendix 8). Genes for glycine betaine/carnitine transport binding protein *GbuAC*, glycine betaine transporters *OpuD/betL*, and glycine betaine permease have been identified as *GbuB* (Bursy et al., 2008). Trehalose plays a significant role as an osmotic stress regulator (Roberts, 2005; Reina-Bueno et al., 2012). The presence of trehalose is demonstrated in genes encoding trehalose ATP-binding protein *SugC*, trehalose transport system permease protein *SugB*, and trehalose PTS system EIIBC component. Furthermore, the ectoine expression cassette *ectABC* (Chen et al., 2021) was present in all four analyzed genomes, indicating intracellular biosynthesis of the osmolyte.

3.5.4 Genome-based taxonomic placement

Taxonomic relationships demonstrated that the genomes of isolates LMS6, LMS18, LMS25, and LMS39 have ANI values below the species delineation boundary of 95 – 96% for reliable classification, indicating that these isolates are new species of *Shouchella* sp., *Evansella* sp., *Salipaludibacillus* sp., and *Alkalihalobacterium* sp., respectively (Richter et al., 2009; Qin et al., 2014). Taxonomic assignments of prokaryotic communities have mainly relied on the 16S rRNA gene-based sequencing. However, owing to limitations such as high sequence similarity amongst closely related species, genome-based analysis provides an alternative method to infer phylogenetic relationships since it is a reliable, reproducible, and highly informative (Parks et al., 2018; Paul et al., 2020). Previous studies have indicated that members of the genus *Shouchella* have G+C values of 39.7–54 mol% (Joshi et al., 2021). However, ANI values indicated that members of this genus are highly heterogeneous and therefore need reclassification (Kim et al., 2023). Members of the

genus *Salipaludibacillus* are Gram-positive rods with catalase, and oxidase activities, but inactive against casein, gelatin, and starch. They have G+C content between 39 and 42 mol% (Sultanpuram and Mothe, 2016; Amoozegar et al., 2018). On the other hand, *Alkalihalobacterium* spp. are Gram-stain positive, long-rod shaped cells aerobic, and endospore-forming bacteria with DNA G + C content 35.1–37.5%. Their habitats are soda lakes, particularly in the soil and sediments (Joshi et al., 2021). The. Similarly, the distinction in the chromosomal organization of the studied isolates showed structurally similar regions with their respective type strains. However, several genomic regions within the genomes were not identical to their closest relatives (Figures 3.5 - 3.8).

3.5.5 Putative polysaccharide degrading genes in the genomes

The functional domains of CAZyme in the predicted genes were examined (Figure 3.9; Table 3.5). These enzymes are important in decomposing oligosaccharides, polysaccharides, and glycoconjugates (Cantarel et al., 2009). The results revealed that genes encoding enzymes involved in the breakdown of cellulose, starch, pectin, and xylan were present in the genomes (Table 3.5). Comparatively, there were notable differences in the number of genes encoding for a particular CAZyme class between the sequenced and the closest type strain genomes. Overall, the CAZyme repertoires were high in the genomes of *Salipaludibacillus* LMS25, *Alkalihalobacterium* LMS39, and their closest type strains. In particular, a high number of genes encoding for the GH class were recorded in these genomes (Figure 3.9. These genes included α-amylases, cellulases, α-glucosidase, β-glucosidase, endo-1,4-β-xyalanase, type II pullulanase, and β-glucanase. Together with CE and PL, the GH class plays a crucial role in the degradation of plant biomass to provide energy in an extreme environment (Zhao et al., 2014). Notably, a low number of PL genes (Sun et al., 2019) were recorded only in the genomes of LMS6, LMS25, and LMS39. Although several studies have established and characterized the CAZyme genes from alkalihalophilic bacteria (Hirasawa et al., 2006; Huang et al., 2010; Annamalai et al., 2013), none has comprehensively analyzed the composition of these genes in the chromosomes of *Evansella*, *Salipaludibacillus,* and *Alkalihalobacterium* groups.

The analyzed genomes revealed that cellulase- and xylanase-encoding genes were the most abundant in the GH CAZyme family. β-endoglucanase and β-xylanase convert

lignocellulosic biomass (LCB) to cello-oligosaccharides and xylooligosaccharides, respectively. Subsequently, β-glucosidases and β-xylosidases degrade these oligosaccharides into glucose and xylose (Patel et al., 2020). The genome of *Shouchella* sp. LMS6 contained an endoglucanase gene with 97% amino acid similarity to a cellulase family protein (WP218014778) derived from *Bacillus* sp. Marseille-P3800. Also contained within this genome is a α-amylase encoding gene with 99% similarity to an α-amylase (WP176554439) deduced from *Bacillus* sp. Marseille-P3800. Amylase and neopullanase genes (GH13) were also present in the genomes of LMS25 and LMS39 with 97% similarity to known genes. The genome of LMS18 (*Evansella* sp.) encodes 2 xylanase genes exhibiting 95% similarity to an endo-1,4-beta-xylanase (WP088034427) obtained from *Ev. clarkii*. Soley found in LMS25 and LMS39 were genes for pectate lyase and polygalacturonases, enzymes involved in pectin depolymerization (Tayi et al., 2016). Additionally, a chitinase (GH18) gene was encoded by the genome of LMS39. This gene had a similarity of 96% to a chitinase derived from *Ab. bogoriensis* (WP035177761). Several members of *Salipaludibacillus* and *Alkalihalobacterium* have attracted attention due to their ability to secrete alkalihalotolerant enzymes for depolymerizing several polymers (Gessesse & Gashe, 1997; Horikoshi, 1999). Particularly, the halotolerant and alkaliphilic bacterium, *Hm. ligininiphilus* contains genes encoding enzymes for hydrolysis of lignin (Zhu et al., 2017), while *Sb. agaradhaerens* has been found to contain a salt-activated cellulase gene (Hirasawa et al., 2006).

3.6 Conclusions

A considerable genetic diversity of bacteria possessing the potential to hydrolyze various polysaccharides were isolated from Lake Magadi. Employment of enrichment and long incubation periods yielded potentially new strains of *Shouchella, Evansella, Salipaludibacillus,* and *Alkalihalobacterium*. The genome analysis revealed the presence of genes encoding for different polysaccharide-degrading genes, further confirming the detected activities of the isolates with these substrates. Compared to *Evansella* and *Shouchella*, this study demonstrated that isolates belonging to *Alkalihalobacterium* sp. and *Salipaludibacillus* sp., contain more genes encoding for carbohydrate-active enzymes. This provides an opportunity for further screening by cloning and expression assays. This could be a rich source of

extremozymes with application potential for use in renewable energy, food, feed, pharmaceutical, agricultural, textile, and cosmetic industries.

CHAPTER FOUR

4.0 SPATIOTEMPORAL STRUCTURE AND COMPOSITION OF MICROBIAL COMMUNITIES IN HYPERSALINE LAKE MAGADI

4.1 Abstract

Soda lakes are habitats characterized by haloalkaline conditions also known to host unique microbial communities. The water chemistry changes with seasons due to evaporative concentration or floods from the surrounding grounds. However, the key ecological question is how the change in physiochemical parameters influences the spatiotemporal diversity and structure of microbial communities in these ecosystems. Using 16S rRNA gene amplicon sequencing, this study investigated the diversity and structure of microbial communities in water and brine samples taken from Lake Magadi between June and September 2018. Additionally, physicochemical parameters were also analyzed for every sampling site. A total of 4,837 operational taxonomic units (OTUs) were distributed in the domain Bacteria (3,802 OTUs) and Archaea (1,035 OTUs). The abundant bacterial phyla were Proteobacteria, Cyanobacteria, Bacteroidetes, Actinobacteria, Firmicutes, Verrumicrobia, Deinococcus-Thermus, Spirochaetes, and Chloroflexi. The Archaeal diversity was represented by phyla Euryachaeota, Crenarchaeota, and Thaumarchaeota. The dominant bacterial species were: *Euhalothece* sp. (10.3%), *Rhodobaca* sp. (9.6%), *Idiomarina* sp. (5.8%), *Rhodothermus* sp. (3.0%), *Roseinatronobacter* sp. (2.4%), *Nocardioides* sp. (2.3%), *Gracilimonas* sp. (2.2%), and *Halomonas* sp. (2.0%). The dominant archaeal species included *Halorubrum* sp. (18.3%), *Salinarchaeum* sp. (5.3%), and *Haloterrigena* sp. (1.3%). The composition of bacteria was higher than that of archaea, while their richness and diversity varied widely across the sampling seasons. The α-diversity indices showed that high diversity was recorded in August, followed by September, June, and July in that order. The findings demonstrated that temperature, pH, P+, K+, and total dissolved solids (TDS) contributed majorly to the diversity observed in the microbial community. Multivariate analysis revealed significant spatial and temporal effects on β-diversity and salinity and alkalinity were the major drivers of microbial distribution in Lake Magadi. The study provides insights into the relationships between microbial communities and geochemistry across time points and sampling sites in Lake Magadi.

4.2 Introduction

Most living organisms are adapted to habitats characterized by moderate temperature (10 – 37°C), pH (approximately 7), salinity (0.15 – 0.5 M NaCl), pressure (1 atm), and adequate supply of water (Aguilar et al., 1998; Antranikian et al., 2005). However, molecular techniques such as next-generation sequencing have revealed that diverse groups of organisms thrive even in biomes previously thought to be lifeless (Canganella & Wiegel, 2011; Rampelotto, 2013). Microbial communities in ecosystems such as the hypersaline lakes of the East African Rift Valley survive and

thrive under one or several extremes and are referred to as polyextremophiles (Sorokin et al., 2014; Urbieta et al., 2015).

The distribution and diversity of microbial communities in hypersaline lakes are mainly affected by physicochemical parameters (Tazi et al., 2014). Lake Magadi an extreme habitat with high concentrations of Na^+, K^+, CO_3^{2-}, Cl^-, HCO_3^-, and SiO_2, but low concentrations of Ca^{2+} and Mg^{2+} (Jones et al., 1998; Getenet et al., 2022). During the dry seasons, thermonatrite ($Na_2CO_3.H_2O$), and halite (NaCl) precipitate by evaporative concentration (Eugester, 1971; 1980). The lake is in a region with alternating wet and dry seasons. During the dry season, when ground temperatures exceed 40°C, there is extensive evaporation (Matagi, 2004; Muruga & Anyango, 2013). Furthermore, the lake is almost entirely covered by a white layer of soda, and flooding may occur when it rains due to feeding water from the surroundings.

Despite the extreme conditions existing in the lake, it is a highly productive ecosystem with diverse microbial communities driving active nitrogen, carbon, and sulfur cycles (Jones et al., 1998; Sorokin et al., 2007). The high productivity is mainly driven by *Arthrospira* sp. and other cyanobacteria (Melack & Kilham, 1974; Oduor & Schagerl, 2007). Cyanobacteria in lake lagoons only form algal mats in these lakes during rainy seasons (Jones et al., 1998; Muruga & Anyango, 2013; Krenitz & Shagerl, 2016). Reports indicate that *Ectothiorhodospira*, an anoxygenic phototrophic halophilic bacterium, and eukaryotes such as diatomic and green algae also play an essential part in primary production (Matagi, 2004; Grant, 2006).

Many bacterial species have been isolated from extreme environments, and they frequently exhibit adaptations to optimal growth under the prevailing conditions (Krulwich et al., 2011; De Maayer et al., 2014; Sorokin et al., 2014). Previously described isolates from Lake Magadi include the archaeal genera *Natronobacterium* and *Natronococcus* gen. nov. (Tindall et al., 1984) and *Natronobacterium magadii*, *Natrialba magadii* (Kamekura et al., 1997), bacterial species *Spirochaeta alkalica* sp. nov., *Spirochaeta Africana* (Zhilina et al., 1996), *Tindallia magadiensis* (Kevbrin et al., 1998), *Halomonas magadii* (Andrew W. Duckworth et al., 2000), *Amphibacillus fermentum* (renamed *Pelagirhabdus fermentum*) sp. nov., *Amphibacillus tropicus*, and *Halonatronum saccharophilum* (Zhilina et al., 2001), *Methylonatrum kenyense*

(Sorokin et al., 2007), *Euhalothece natronophila* (Mikhodyuk et al., 2008) and *Natranaerobaculum magadiense* (Zavarzina et al., 2013). Edwin et al. (2019) recovered 11 isolates affiliated with the cyanobacterial orders *Chroococcales*, *Oscillatoriales, Pleurocapsales,* and *Nostocales*. Recent studies have reported bacterial isolates affiliated with the genus *Bacillus, Alkalibacterium*, *Staphylococcus*, *Micrococcus, Halomonas,* and *Alkalilimnicola* (Kiplimo et al., 2019; Kipnyargis et al., 2022). A study by Orwa et al. (2020) recovered several fungal isolates affiliated with 18 different genera with *Aspergillus, Penicillium, Cladosporium*, *Phorma,* and *Acremonium* being dominant. Several studies have explored the microbial diversity in Lake Magadi using 16S rRNA amplicon analysis targeting groups such as fungi (Kambura et al., 2016; Salano et al., 2017; Mwirichia, 2022) or bacteria (Kambura et al., 2016).

A key ecological question is how microbial diversity changes with the fluctuating physicochemical conditions across seasons. This study hypothesized that microbial communities within the lake change in response to changes in the water chemistry over time. It also predicts that the communities in the brine and spring water are distinct from those in the open lake water. Therefore, the study explored the spatiotemporal variation in the microbial community over four months at different sites in Lake Magadi using 16S rRNA gene amplicon sequencing.

4.3 Materials and methods

4.3.1 Sampling sites and sampling criteria

Sampling was done in hypersaline Lake Magadi, Kenya. It is located 1°43 - 2°00 S and 36°13 - 36°18E in an enclosed basin with an annual precipitation of 500 mm (Behr & Röhricht, 2000). Lake Magadi is a relatively shallow water body fed by various hot springs distributed along the edges of the lake. The inflows have an influence on the lake volume and the water chemistry. Water samples were collected from different points in the lake including spring, brine, and open waters. Samples were collected from these sites in June, July, August, and September 2018 and the dry season lasted for all the sampling months, with June marking the start of the season and September being the driest month. The coordinates of the sampling sites were: S1 (S01°53.611' E036°18.158'), S2 (S01°52.856' E036°18.160'), S3 (S01°53.962' E036°18.080'), S4 (S01°54.088' E036°18.093'), S5 (S01°59.570' E036°15.579'), S6 (S01°59.410' E036°15.213') and BR1 (S01°52.877' E036°16.185')

(Figure 4.1). S1 was composed of spring water, S2–S6 were composed of open waters, and BR1 was brine. Three sub-samples of 50 ml each were collected from each site and pooled into a composite sample. In addition, water samples for physicochemical analysis were collected. All samples were collected in sterile Conical Centrifuge tubes (Biologix, Shandong, China, Cat. No. 430829) and transported in a cool box.

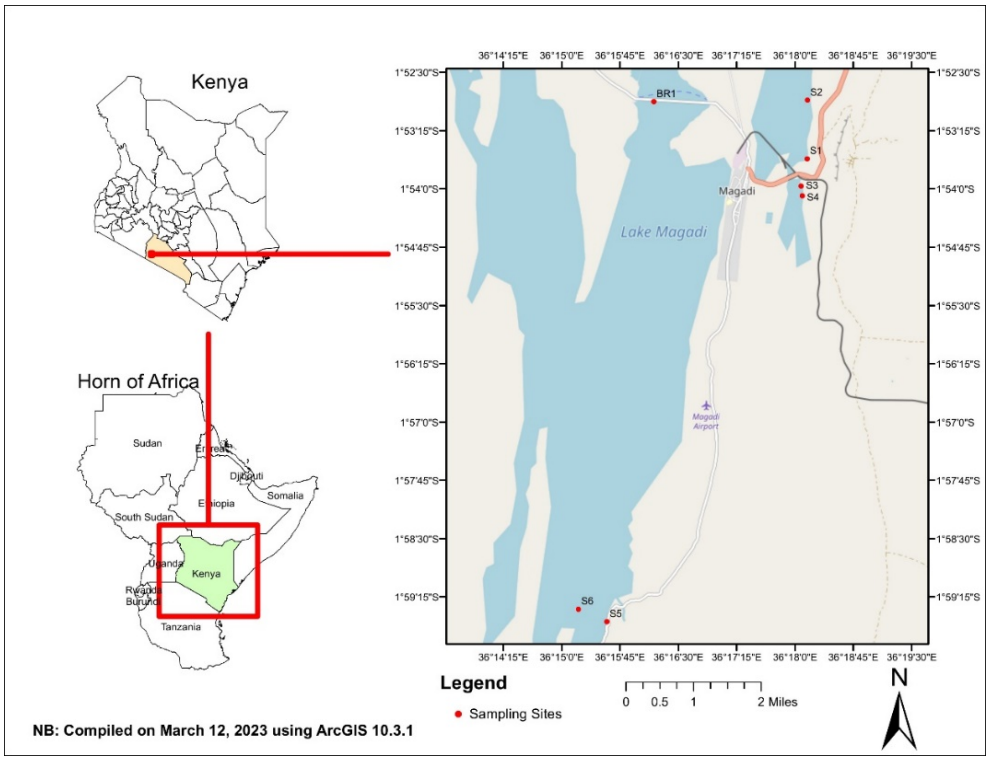


Figure 4.1: Map of Lake Magadi showing the sampling sites.

4.3.2 Analysis of physicochemical parameters

Water temperature, pH, total dissolved solids (TDS), and salinity measurements were determined *in situ*. Water temperature, TDS, and salinity were measured using a VWR phenomenal handheld Meter (VWR, Atlanta, GA, USA, Model CO 3100H), while pH was measured using a Hanna Combo pH meter (Hanna Instruments, Nusafalau, Romania, Model HI-98128). Water samples for dissolved P, K^{+}, NO_3^{-}, NH_4^{-}, Mg^{2+}, Na^{-}, Fe^{2-}, Ca^{2+}, SO_4^{2-}, Cl^{-}, and HCO_3^{-} measurements were collected in sterile 500ml bottles and stored in a cool box for transportation to Crop Nutrition Laboratory Services (CNLS), Nairobi where analysis was done. Cations such as Ca, Mg, K, Na, Mn, Fe, Cu, Mo, B, Zn, and S were analyzed using atomic absorption spectrometry (AAS), while anion analysis was carried out using mass spectrometry.

4.3.3 DNA extraction

Cell biomass for DNA extraction was obtained by centrifuging 50 ml of each water sample at 14,000 rpm for 20 minutes in an Eppendorf centrifuge (Eppendorf, Model 5415R, Cat. Z605212). The pellets were resuspended in 200 µl of a resuspension buffer (25% w/v sucrose (Sigma-Aldrich, Cat. No. S9378) in 50 mM Tris pH 8.5 (Sigma-Aldrich, Cat. No. 93352), and 50 mM EDTA; pH 8.0 (Sigma-Aldrich, Cat. No. 798681). To disrupt the cell wall of Gram positives, 2 µl of lysozyme (20 mg/ml) (Roche, Cat. No. 10837059001) and 10 µl of RNAse A (20 mg/ml) (Roche, Cat. No. 10109142001) were added and incubated at 37°C for 30 minutes. Cell lysis was achieved by the addition of 600 µl of a lysis buffer (1% SDS (Sigma-Aldrich, Cat. No. 8.17034) in 10 mM Tris pH 8.5 (Sigma-Aldrich) and 5 mM EDTA; pH 8.0 (Sigma-Aldrich). The samples were gently mixed with 10 µl of Proteinase K (20 mg/ml) (Sigma-Aldrich, Cat. No. 39450-01-6) and incubated at 65°C for 2 hours. DNA was recovered by adding an equal volume of chloroform (Sigma-Aldrich, Cat. No. C2432) followed by centrifugation at 13,200 rpm for 10 min at 4°C in an Eppendorf 5415R centrifuge. The aqueous layer was transferred into a new tube with 150 µl of sodium acetate (pH 5.2) (Sigma-Aldrich, Cat. No. S8750) and an equal volume of isopropyl alcohol (Sigma-Aldrich, Cat. No. 67-63-0). The contents were centrifuged at 13,200 rpm for 10 minutes and the DNA pellet was recovered by washing with 70% ethanol, air-dried, and dissolved in 30 µl of nuclease-free water (Sigma-Aldrich, Cat. No. 7732-18-5). DNA quality was checked by running an aliquot of 2 µl in 1% agarose (Sigma-Aldrich, Cat. No. A9918) gel electrophoresis (Orwa et al., 2020).

4.3.4 Sequencing of the 16S rRNA gene amplicons

The V4 hypervariable region of the 16S rRNA genes was amplified using the universal primers for bacteria and archaea 515F (5'-GTGCCAGCMGCCGCGGTAA-3') and 806R (5'-GGACTACHVGGGTWTCTAAT-3') (Caporaso et al., 2012). Amplification was done using HotStarTaq Plus Master Mix Kit (Qiagen, USA) under the following cycling conditions: initial denaturation at 94°C for 3 minutes, followed by 35 cycles of denaturation at 94°C for 30 seconds, annealing at 53°C for 40 seconds and elongation at 72°C for 1 minute, after which a final elongation step at 72°C for 5 minutes was performed. Three independent PCR reactions were performed per

sample and pooled in equimolar amounts. The PCR products were then checked in a 2% agarose gel. The sample was purified using calibrated Ampure XP beads (Beckman Coulter, Inc., IN, USA). DNA libraries were prepared using Illumina TruSeq DNA libraries (Illumina, Inc., San Diego, CA, United States) and sequencing was performed at MR DNA (Shallowater, TX, USA) on a MiSeq platform (2 × 300 bp) following the guidelines of the manufacturer (Illumina Inc.).

4.3.5 Sequence processing and taxonomic assignment

The Q25 sequence data derived from MiSeq sequencing was processed using the MR DNA ribosomal and functional gene analysis pipeline (MR DNA, Shallowater, TX). Sequences were depleted of primers, reads <250bp and ambiguous base calls were removed. The reads were quality-filtered using a maximum expected error threshold of 1.0. Sequences were further processed using VSEARCH v2.14 (Rognes et al., 2016). This included sorting and size-filtering of the paired reads to ≥300 bp (--sortbylength --minseqlength 300) and dereplication (--derep_fulllength). The sequences were then denoised and evaluated for potential chimeric sequences using UCHIME package v.11. (Edgar et al., 2011). Representative operational taxonomic units (OTUs) were picked *de novo* using VSEARCH v2.14 (Rognes et al., 2016), and assigned taxonomy using BLAST searches against the SILVA v132 rRNA reference database (Quast et al., 2012). A sequence identity cutoff of 97% was used to pick OTUs from the quality-filtered, denoised, non-chimeric sequences. Eukaryotic sequences were filtered from the dataset using the script *filter_otu_table.py*. in QIIME v1.90 (Caporaso et al., 2010). The Illumina raw reads for the 16S rRNA gene sequences were deposited in the Sequence Read Archive (SRA) of NCBI under the accession number PRJNA962270.

4.3.6 Microbial community analysis

Sequences that were assigned taxonomy were aligned using PyNast (Caporaso et al., 2010). The alpha diversity indices (Chao1, abundance-based coverage estimator (ACE), Simpson, Shannon, Fisher's alpha, Pielou's evenness, and Good's coverage) were calculated with QIIME v1.90 (Caporaso et al., 2010) using *alpha_rarefaction.py* employing the same level of surveying effort (37,000 per sample based on the lowest sample count). All subsequent steps were analyzed in R

software v4.2.0 (R Core Team, 2020) and RStudio v1.1.456 (RStudio Team, 2020). The results of all statistical tests were regarded as significant if $p \leq 0.05$. To compare the (dis)similarity of OTU compositions between communities the OTU abundance table was standardized using decostand (method = "hellinger"). Hierarchical cluster analysis was performed using hclust in R software v4.2.0 (R Core Team, 2020) (method = "average"). The heatmap was created using JColorGrid v1.86 (Joachimiak et al., 2006).

The OTU network generated in QIIME was filtered using an edge cut-off of 0.001 and visualized in Cytoscape v3.9.1 (Otasek et al., 2019) in an "edge-weighted spring-embedded layout". In this case, sampling sites were used as source nodes and bacterial families as target nodes. Redundancy analysis (RDA), based on Bray Curtis dissimilarity was used to test the correlation between the physicochemical parameters and the microbial community at the genus level. This was done using the *Microeco* package v0.15.0 (C. Liu et al., 2021) and plotted using the package *Pheatmap* in R.

To assess the beta diversity of microbial communities, principal component analysis (PCA) was performed using Bray-Curtis dissimilarities with the script compare_categories.py. test and weighted UniFrac distance matrix (Lozupone & Knight, 2005) as input using the Vegan package in R (Bray & Curtis, 1957; Oksanen, 2015).

4.4 Results

4.4.1 Physicochemical properties of the sampling sites

One of the objectives of this study was to investigate the change in water chemistry over time. It has been established that physicochemical factors play a critical role in shaping the structural composition of microbial communities in an ecosystem. Samples from site S1 exhibited lower concentrations of the various ions compared with the other samples. The water temperature ranged from 27°C to 38.7°C (average 33.7°C). The pH of the water was alkaline, ranging from 9.8 (S6_June) to 11.5 (BR1_June) recording the highest pH value of 11.5. The major water cations were Na^+ (10,300 - 160,000 ppm) and K^+ (131 - 4,280 ppm), and the major anions were HCO_3^- (15,400 - 277,000ppm) and Cl^- (4,050 -102,000 mg/L). Phosphorus levels

ranged from 2.38 -108 ppm, while magnesium and calcium levels were low, ranging from 0.02 - 16.1 and 0.05 - 127 ppm, respectively. The total dissolved solids (TDS) ranged from 27.1 to 153.5 ppm (Table 4.1).

Table 4.1: Physicochemical characteristics of the water samples collected from Lake Magadi.

	S1_ Jun	S1_ Sep	S2_ Jun	S2_ July	S2_Se p	S3_ Jun	S3_ Jul	S3_ Aug	S3_ Sep	S4_ Jun	S4_ Jul	S4_ Aug	S4_ Sep	S5_ Jun	S5_ Jul	S5_ Aug	S5_ Sep	S6- Jun	S6- Jul	S6- Aug	BR1- Jun	BR1- Sep
pH	10.5	10.4	10.6	10.2	10.7	10.5	10.5	11.3	10.4	10.5	10.5	11.1	10.9	10.2	9.9	10.3	10.2	10.3	9.8	10.4	11.5	11.2
Tem	35.6	35.6	35.1	32.8	37.2	38.7	27	34.6	35.6	37.3	27	33.6	37.8	34.2	32.7	29.8	32.8	32.5	30.2	29.2	38.3	34.8
TDS	27.1	27.1	137	144	145	114	135	139	153	111	135	140	143	46	46	46	43	105	118	83.9	136	134.8
P^{3-}	2.38	7.77	58.4	78	117	27.4	77.2	108	69.8	26.4	63.8	107	105	3.96	2.39	5.79	10.6	16.2	21.9	4.91	81.1	92.8
K^{+}	131	365	2,300	2,700	4,280	1,220	2,430	3,300	2,560	1,130	1,960	3,270	2,370	201	190	280	378	697	697	513	3,410	3,210
NO_3^{-}	0.89	9.03	1.81	0.01	5.98	0.01	0.01	0.2	4.96	0.01	0.01	0.01	6.99	4.56	0.01	0.19	8.5	32	0.01	0.01	0.01	7.44
NH_4^{+}	0.02	0.69	0.07	0.2	0.98	0.02	0.08	0.01	0.53	0.03	0.01	0.06	0.86	0.07	0.07	1.04	1.52	0.01	0.16	0.59	0.47	0.79
Mg^{2+}	0.03	5.86	2.49	6.58	2.63	2.82	8.2	0.02	4.76	2.41	6.31	0.3	2.31	0.81	4.46	0.02	2.57	3.89	6.02	0.02	1.44	16.1
Mn^{+}	0.06	0.01	0.05	0.07	0.01	0.08	0.23	0.01	0.02	0.15	0.13	0.11	0.01	0.01	0.01	0.01	0.01	0.24	0.74	0.15	0.028	0.53
S^{2-}	39.9	132	548	629	1010	324	708	958	710	304	585	973	949	97.2	90.7	71.4	182	213	271	178	845	974
Cu^{2+}	0.01	0.15	0.01	0.01	0.15	0.01	0.01	0.22	0.09	0.01	0.01	0.01	0.1	0.01	0.01	0.01	0.11	0.01	0.01	0.07	0.01	0.4
B^{3+}	6.8	14.6	103	126	197	55.7	126	184	116	52	108	184	169	13	12.4	11.2	20.4	37.8	43.6	28.2	147	160
Zn^{2+}	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.33	0.29
Na+	10,300	22,900	93,400	114,000	143,000	63,900	120,000	121,000	100,000	58,800	100,000	118,000	104,000	18,300	17,700	16,000	30,800	58,800	70,500	36,000	155,000	160,000
Fe^{2+}	2.79	0.53	0.01	1.34	0.01	2.13	9.63	0.91	0.92	2.42	3.17	1.59	0.17	0.25	2.5	0.01	0.25	4.1	14.7	3.82	0.86	0.49
Ca^{2+}	0.05	3.79	8.52	10.9	0.34	12	16.4	0.2	7.43	10.8	9.23	1.13	6.54	0.05	2.94	0.05	2.72	13.8	15.8	0.05	2.15	127
SO_4^{2-}	120	395	1,640	1,880	3,030	971	2,120	2,870	2,130	911	1,750	2,920	2,840	291	272	214	545	638	812	533	2,530	2,920
Mo^{+}	0.07	0.25	1.66	2.49	3.82	1.03	2.43	4.99	2.23	0.83	1.74	4.5	3.09	0.05	0.23	2.09	0.24	0.29	0.4	1.68	2.37	2.79
Cl^{-}	4,050	3,720	52,300	77,400	102,000	30,700	72,300	99,800	74,200	30,200	60,500	93,100	62,500	7,910	8,620	7,680	6,740	24,100	32,500	16,700	93,500	99,000
NO_3N	0.2	2.04	0.41	0.01	1.35	0.01	0.01	0.05	1.12	0.01	0.01	0.01	1.58	1.03	0.01	0.04	1.92	7.23	0.01	0.01	0.01	1.68
HCO_3^{-}	17,300	15,400	145,000	205,000	213,000	94,300	229,000	242,000	143,000	94,700	196,000	233,000	165,000	30,100	30,200	29,300	25,300	98,200	141,000	70,600	256,000	277,000
Si^{4+}	66.2	114	447	568	786	282	569	861	590	284	458	867	760	89.9	94.9	194	107	106	197	287	595	696
SiO_2	142	244	956	1,220	1,680	603	1,220	1,840	1,260	608	980	1,850	1,630	192	203	415	229	227	421	614	1,270	1,490
SAR	9,310	1,720	7,230	6,730	18,200	4,310	6,040	69,000	7,050	4,210	6,220	25,500	8,900	4,290	1,520	15,300	3,220	3,600	3,830	34,400	20,100	3,550
$CaCO_3$	0.23	33.5	31.5	54.2	11.6	41.6	74.6	0.58	38.1	36.9	48.9	4.05	25.8	3.45	25.6	0.21	17.3	50.4	64.2	0.21	11.3	384

*Except for temperature (in ℃) and pH, all other parameters were measured in parts per million (ppm).

4.4.2 Sequence analysis and diversity studies

After quality filtering, denoising, and removal of potential chimeras and non-bacterial sequences, approximately 3,197,447 high-quality sequences (Maximum 285,085; minimum 37,406) with an average read length of 325 bp were obtained from the entire dataset. The number of sequences per sample varied from 37,406 (sample S5_Jun) to 285,085 (sample BR1_Sep) with an average value of 121,603 sequences. The number of OTUs per sample varied from 852 (sample S3_July) to 2,024 (sample S5_Sep) (Table 3.2). All sequences were assigned taxonomy up to genus level and clustered into 4,837 OTUs (97% identity) distributed in the domain Bacteria (3,802 OTUs) and Archaea (1,035 OTUs). Overall, most OTUs (1868; 98%) were co-shared among all the sampling seasons. The highest unique OTUs were in September (450; 0.2%), while August had the least unique OTUs (25) (Figure 4.2).

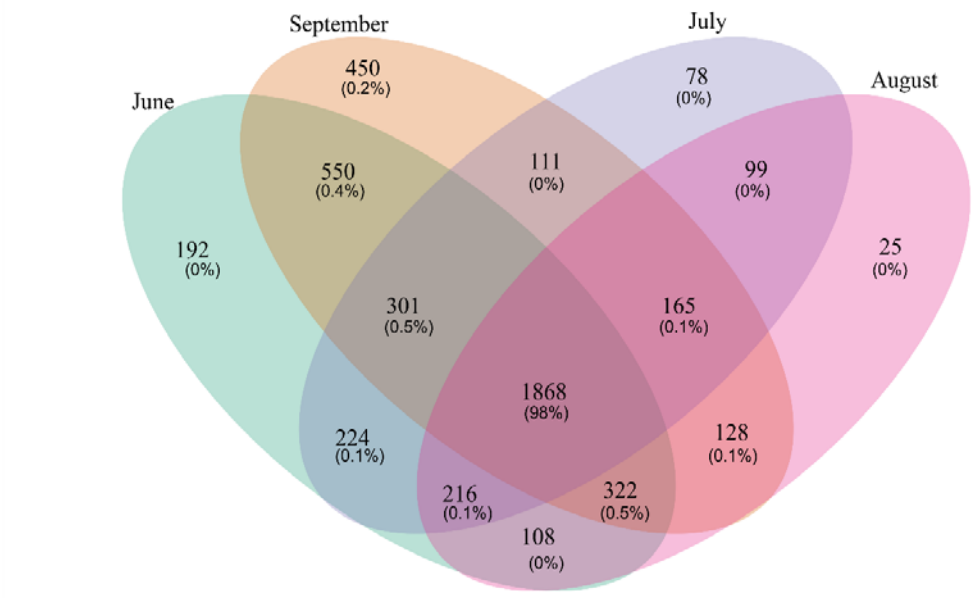


Figure 4.2: Distribution of Operational Taxonomic Units (OTUs) across sampling seasons in Lake Magadi.

4.4.3 Alpha diversity studies

The values of the good's coverage estimator ranged from 81% (S5_Sep) and 96% (S3_Aug) (Table 4.2), suggesting that the sequencing process captured a significant number of dominant communities. Within the open water samples (S2 – S6), S5 samples collected across the seasons had the highest alpha diversity indices suggesting that S5 had the highest species richness and diversity. S3_Aug samples (open waters) had the lowest alpha diversity indices. Within the hot spring samples (S1), samples collected in September had the highest species richness and diversity. Within the brine samples (BR1), Br1 samples collected in September had the highest species diversity and richness.

Table 4.2: Number of sequences generated, OTUs, and diversity indices of the sampling sites in Lake Magadi.

Sample	No. of Sequences	OTUs	Chao1	ACE	S*	O**	Sh#	Fisher alpha	Good coverage
S1_June	191,597	1,089	171.50	75.11	0.9	89	4.6	25.63	0.94
S2_June	174,894	1,267	180.77	181.63	0.9	94	4.3	27.66	0.94
S3_June	139,647	1,513	251.65	319.48	0.9	131	5.2	44.51	0.91
S4_June	130,649	1,514	306.14	397.09	0.9	140	5.0	49.13	0.89
S5_June	37,406	1,221	363.23	452.68	0.9	209	6.2	92.00	0.85
S6_June	116,475	1,665	350.70	385.77	0.9	179	5.9	71.63	0.87
Br1_Jun	149,051	1,416	357.30	414.77	0.9	148	5.3	53.40	0.89
S2_July	177,773	1,259	301.33	300.15	0.9	98.	4.5	29.32	0.92
S3_July	177,827	852	171.50	162.62	0.7	69	2.9	18.11	0.95
S4_July	156,660	1,167	166.62	189.02	1.0	87	4.3	24.84	0.94
S5_July	109,083	1,917	435.14	487.08	1.0	231	6.5	108.9	0.83
S6_July	103,156	1,149	181.32	192.37	1.0	106	4.2	32.76	0.93
S3_Aug	177,142	874	91.91	107.34	0.8	55	3.2	13.39	0.96
S4_Aug	172,666	934	175.86	143.57	0.8	70	3.7	18.46	0.95
S5_Aug	83,183	1,815	572.10	579.97	0.9	222	6.3	101.8	0.82
S6_Aug	64,022	1,732	497.95	457.95	1.0	200	6.6	85.59	0.86
S1_Sep	180,065	1,583	341.16	436.97	1.0	177	5.8	70.38	0.87
S2_Sep	170,143	1,114	184.23	213.10	1.0	90	4.3	26.03	0.94
S3_Sep	149,323	1,124	149.37	174.86	0.8	90	3.5	26.03	0.94
S4_Sep	151,058	956	185.17	135.22	0.8	68	3.3	17.76	0.95
S5_Sep	100,542	2,024	594.50	641.88	1.0	248	6.9	123.1	0.81
BR1_Se	285,085	1,710	323.50	356.33	1.0	178	5.8	71.00	0.88

OTU, operational taxonomic units; ACE, abundance-based coverage estimator; S* = Simpson diversity index; O** = observed species; Sh#= Shannon diversity index.

4.4.4 Beta diversity studies

Beta diversity ordination based on Bray-Curtis dissimilarity showed that samples (except hot spring and brine samples) did not cluster based on the sampling site or

month. Overall, all samples clustered together based on salinity and alkalinity, indicating the impact of these elements on the distribution of microbial communities (Figure 4.3; Table 4.1).

The principal component (PCA) analysis showed that the first (PC1) and second (PC2) axes described 38.7% and 17.9% of the variance in microbial communities, respectively. Accordingly, samples appeared to be clustered based on alkalinity and salinity into three distinct groups: Low alkalinity and salinity samples (pH 9.8 - 10.5; 10,300 ppm - 70,500 ppm) formed cluster I with nine samples (S1_06, S1_09, S5_06, S5_09, S5_08, S5_07, S6_08, S6_06, and S6_07). Moderately alkaline and saline samples (pH 10.5 – 10.6; 63,900 ppm – 100,000 ppm) formed cluster II with six samples (S3_06, S4_06, S2_06, S3_07, S2_07, and S4_07). Highly alkaline and saline samples (pH 10.7 – 11.5; >100,000 ppm) formed cluster III consisting of six samples (Br1_06, Br1_09, S4_09, S2_09, S3_08, and S4_08).

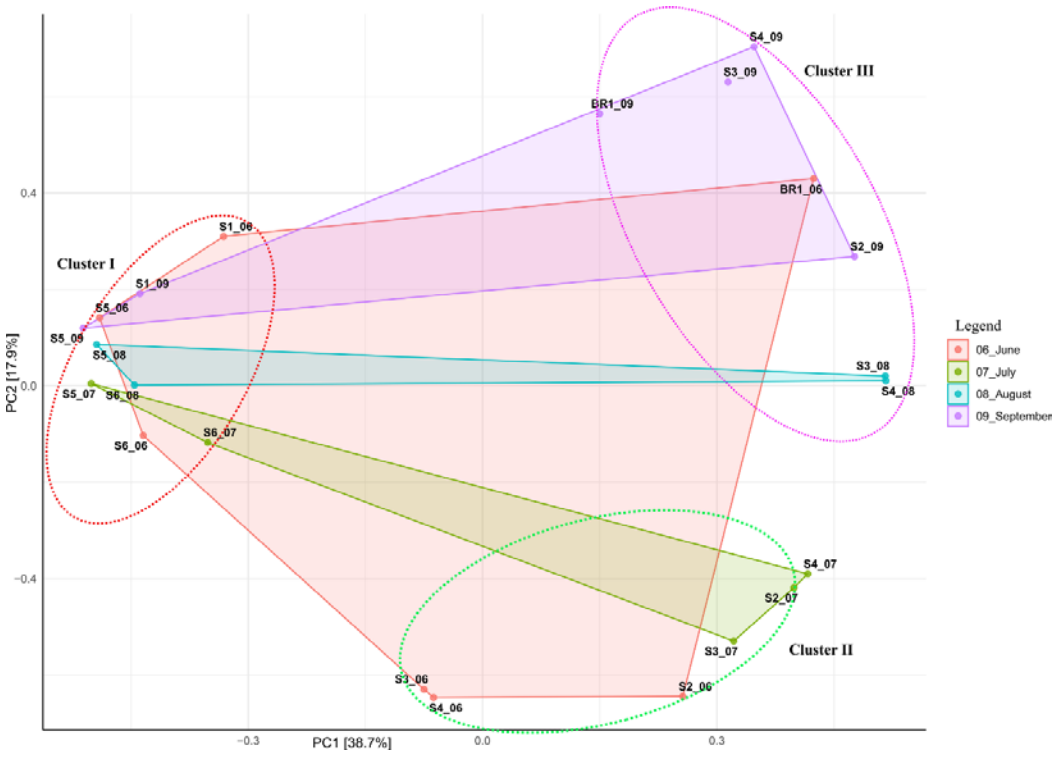


Figure 4.3: Ordination plots of OTU richness among different sampling seasons in Lake Magadi. Alpha diversity plots of OTU richness among different sampling months. Statistical significance was determined at $p < 0.05$. Individual sample values and outliers are shown in the form of dots.

4.4.5 Taxonomic composition and structure

The proportion of bacteria to archaea varied by season and sampling site. The results indicate that the archaeal population increased as the ion concentration increased while bacteria abundance was higher where the ion concentration was lower (sites 1, 4, and 5) (Table 4.1). In hot spring water (S1), archaea abundance was the lowest (0.3%) while bacterial abundance was the highest (99%). Within open water samples (S2 - S6), S4 had the highest abundance of archaea (49.5%), while S5 had the highest proportion of bacteria (99.3%). Within the brine (BR1), the archaea proportion was higher (68%) than the bacterial communities (31%) (Figure 4.4). The highest bacterial abundance was recorded in June with 39.5% followed by July (27%), September (16.8%), and August (16.2%) in that order. Archaeal abundance was highest in September with 53% followed by August (23%), June (12.7%), and July (11.4%) in that order.

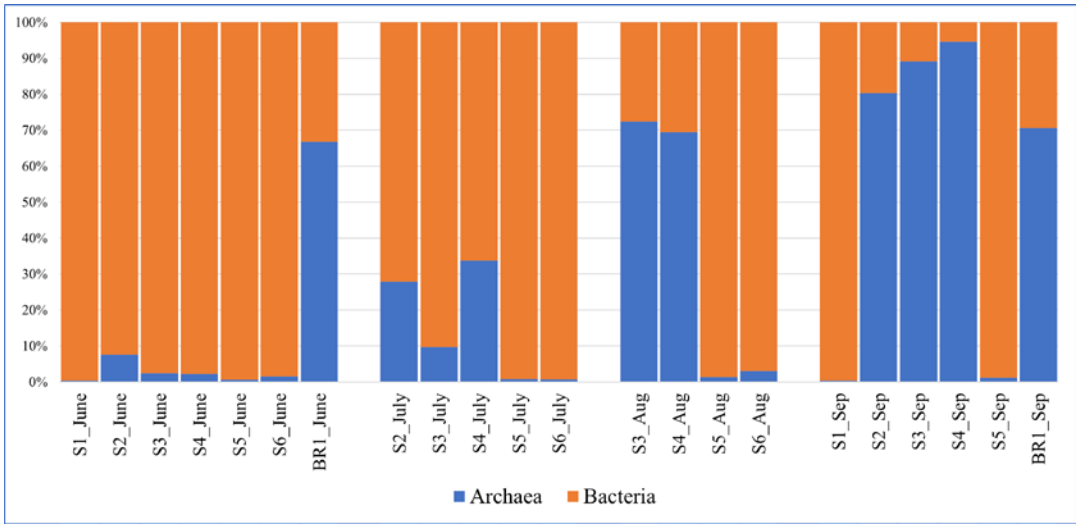


Figure 4.4: Composition and relative abundance of Domain Bacteria and Archaea across the sampling sites and months.

The bacterial reads were distributed across 25 phyla, 107 orders, 225 families, and 545 genera. The results revealed that the most abundant bacterial phyla across the sampling sites and the four months of sampling included Proteobacteria (35% of all the reads), Cyanobacteria (14.2%), Bacteroidetes (10.9%), Actinobacteria (5.2%), Firmicutes (2.7%), Verrumicrobia (1.1%). Deinococcus-Thermus (0.6%), Spirochaetes (0.4%), and Chloroflexi (0.1%) were detected in low abundance (Figure 4.5). The most dominant bacterial genera (> 1% of all sequences across all samples)

were *Euhalothece* (10.3%), *Rhodobaca* (9.6%), *Idiomarina* (5.8%), *Rhodothermus* (3.0%), *Roseinatronobacter* (2.4%), *Nocardioides* (2.3%), *Gracilimonas* (2.2%), *Halomonas* (2.0%), *Lewinella* (1.9%), *Synechococcus* (1.8%), *Aliidiomarina* (1.8%), *Nitriliruptor* (1.7%), *Thioalkalivibrio* (1.7%), *Salinibacter* (1.4%), *Alkalimonas* (1.25%), *Chelatococcus* (1.4%), and *Rhodovulum* (1.4%). Others included: *Cytophaga* (0.9%), *Natronocella* (0.9%), *Thiohalomonas* (0.9%), *Euzebya* (0.8%), *Paracoccus* (0.8%), and *Luteolibacter* (0.8%). The abundance of bacterial genera was higher in the sampling site S5 (25.3%) followed by S6 with 20.1%, S3 (14.5%), S1 and S4 with 12.8% each, and brine sample BR1 with 3.9% abundance in that order.

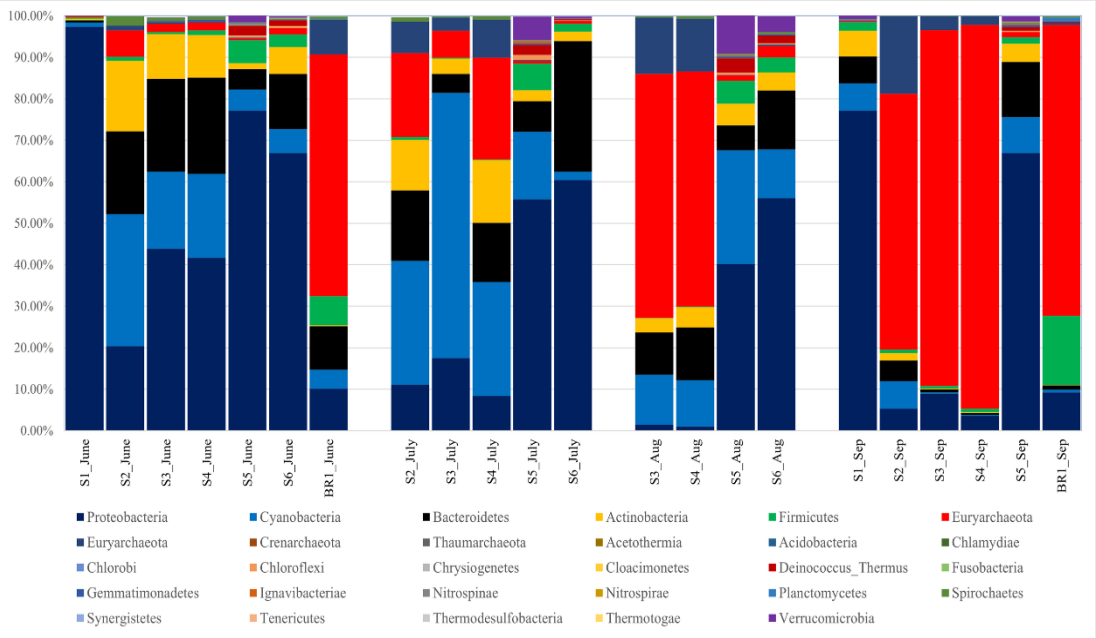


Figure 4.5: Relative abundance of the most popular bacterial and archaeal phyla across the sampling sites and months.

The archaeal reads were affiliated to three phyla (Euryachaeota, Crenarchaeota, and Thaumarchaeota), 14 orders, 20 families, and 62 genera. The dominant Phylum was Euryachaeota (87% of all Archaeal samples), with its dominant genera (> 1% of all sequences across all samples) being *Halorubrum* (18.3%), *Salinarchaeum* (5.4%) and *Haloterrigena* (1.3%). Other genera included *Methanomassiliicoccus* (0.6%), *Palaeococcus* (0.4%), *Halovenus* (0.3%), *Thermococcus* (0.3%), *Haladaptatus* (0.3%), *Halorientalis* (0.3%), *Methanobrevibacter* (0.2%), *Natronomonas* (0.2%), *Halohasta* (0.2%), *Haloquadratum* (0.1%), and *Methanobacterium* (0.1%). The

abundance of archaeal genera was higher in the S3 (27.2%) followed by brine site BR1 with 21.6% abundance. The phylum Crenarcheota was represented by the genus *Candidatus*, while Thaumarcheaota was represented by *Ntrososphaera* all below 1%. Overall, the least archaeal abundance was found in S1 and S6 samples with 0.08 and 0.8%, respectively.

The bacterial species composition (>1%) included *Euhalothece* sp. (10.3%), *Rhodobaca* sp. (9.6%), *Idiomarina* sp. (5.8%), *Rhodothermus* sp. (3.0%), *Roseinatronobacter* sp. (2.4%), *Nocardioides* spp. (2.3%), *Gracilimonas* spp. (2.2%), *Halomonas* sp. (2%), *Lewinella* (1.9%), *Synechococcus* sp. (1.8%), *Cyanobacterium* spp. (1.8%), *Aliidiomarina* sp. (1.7%), *Nitriliruptor* sp. (1,7%), *Thioalkalivibrio* sp. (1.7%), *Salinibacter* sp. (1.4%), *Alkalimonas* spp. (1.2%), *Chelatococcus* sp. (1.1%), and *Rhodovulum* sp. (1.1%). The *Euhalothece natrophila* species were abundant in June, July, and August, except in sites S5 and S6 across all seasons. *Idiomarina* spp. (5.8%) were largely concentrated in June, particularly in sites S1 (51%) and S5 (37.7), whereas *Rhodovulum* spp. was detected across all seasons. *Lewinella coherens* were sampled in June mostly in sites S3 and S4. On the other hand, the Domain Archaea was composed of *Halorubrum* spp. (18.3%), *Salinarchaeum* spp. (5.3%), *Haloterrigena* spp. (1.3%), *Methanomassiliicoccus* spp. (0.7%), and *Palaeococcus* spp. (0.5%). *Idiomarina vacuolatum* was sampled across all the sampling seasons but varied in abundance across the sampling sites. *Halorubrum vacuolatum* was mainly sampled in August (S1 and S2) and September (S3, S4 and S5). *Salinarchaeum* spp. was mainly detected in September, while *Haloterrigena* spp. was sampled across the seasons and sites, though in low proportions. The top 30 most abundant species of bacteria and archaea are shown in Figure 4.6. Overall, *Euhalothece* spp. (10.3%) was the most abundant species detected followed by *Halorubrum* spp. (9.8%) and *Rhodobaca* spp. (5.8%).

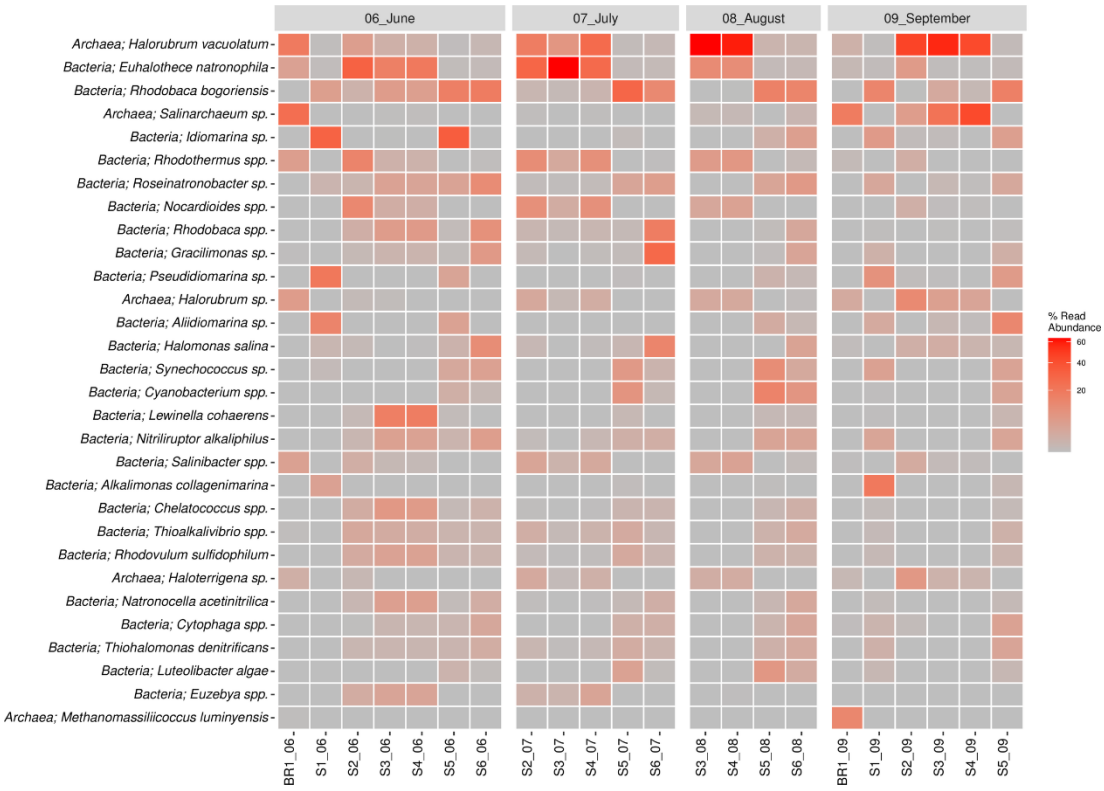


Figure 4.6: The percentage read abundance of the top 30 species across all samples.

At the Family level, Arcaheal Halobacteriaceae members were the most abundant with 27%, followed by Methanobacteriaceae (0.3%), and Archaeoglobaceae (0.1%). The bacterial families were dominated by Rhodobacteraceae (14.7%), followed by Chroococcales (10.4%), Idiomarinaceae (7.7%), Ectothiorhodospiraceae (4.7%), Rhodothermaceae (4.4%), Saprospiraceae (2.4%), Chitinophagaceae (2.3%), Halomonadaceae (2.2%), Synechococcaceae and Cyanobacteriaceae (1.8%), and Nitriliruptoraceae (1.7%). Other members were found in less than 1% abundance. The co-occurrence network revealed that the family Cyclobacteriaceae, Burkholderiaceae, and Alteromonadaceae were unique sampling site S1. Bacterial members of Halobacteroidaceae, Spirochaetaceae, halanaerobiaceae, and desulfohalobiaceae, as well as archaeal members of Archaeoglobaceae and Methanobacteriaceae, were found exclusively in the S2 sampling site. Phyllobacteriaceae and Nostocaceae were unique to S3, while Alcanivoracaceae, Carnobacteriaceae, and Marinilabiliaceae bacteria were found in S5 only. Unique to S6 were the bacterial Puniceicoccaceae family. The highest number of co-shared families was found between S5 and S1 co-sharing eight families, whereas Bacilaceae

and Natranaerobiaceae were found in S5 and S2, and S6 and S2 co-shared only one family (Pseudomonadaceae) (Figure 4.7).

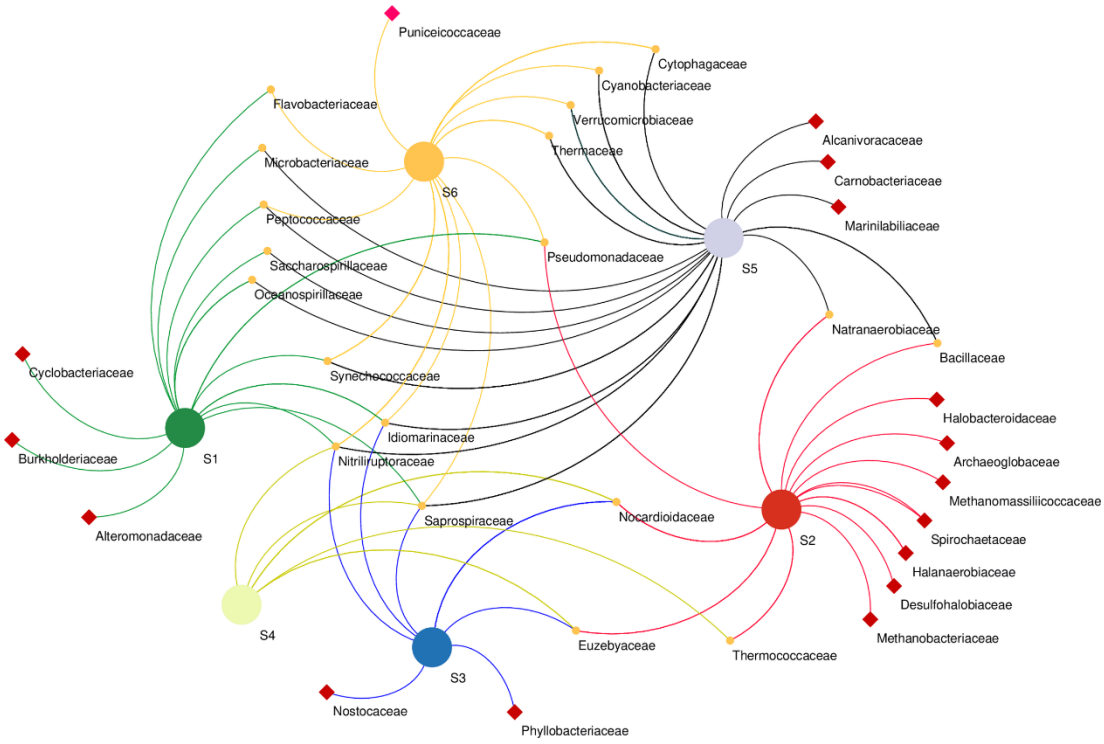


Figure 4.7: Network analysis of microbial communities at the Family level based on the sampling sites.

4.4.5 Physicochemical drivers of microbial community structure

Redundancy analysis (RDA) was used to assess the effect of water chemistry on microbial community structure in Lake Magadi. The results reveal that changes in the physicochemical parameters influenced the microbial communities in the lake. The RDA explained 62.2% and 17.2% of the variation in the first (RDA1) and the second (RDA2) axes, respectively (Figure 4.8).

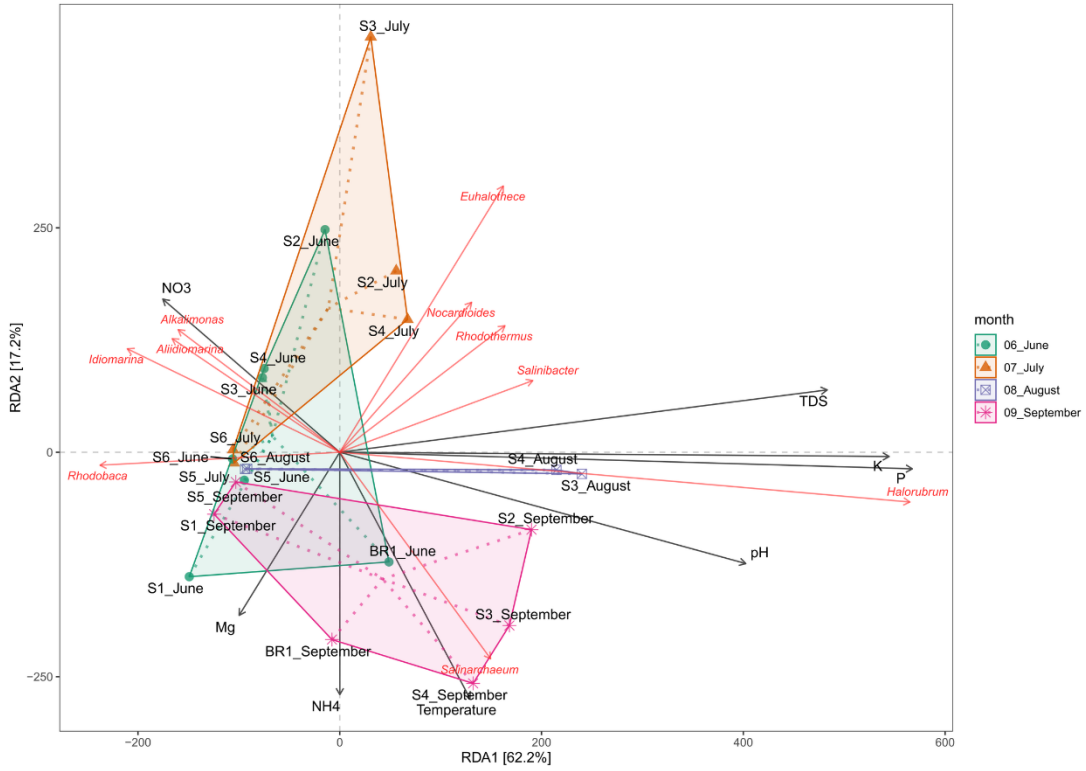


Figure 4.8: Redundancy analysis (RDA) ordination showing the relationships between the physicochemical factors and the dominant genera in Lake Magadi. Samples corresponding to their sampling month are indicated, while genera are written in red, and arrows indicate environmental variables. TDS corresponds to total dissolved solids.

The results indicated that temperature, pH, P, K, NO_3^-, and TDS influenced the microbial community structure. Generally, members of genera *Nocardiodes*, *Rhodothermus*, *Haloterrigena*, *Methanomasiliicoccus*, *Halorubrum*, *Palaeococcus*, *Nocardioides*, *Salinarcheum*, *Salinibacter*, and *Euhalothece* spp. had a wide range of adaptability. Conversely, *Synechococcus*, *Thioalkalivibrio*, *Cyanobacterium*, *Rhodovulum*, *Lewinella*, *Idiomarina*, *Pseudidiomarina*, *Chelatococcus*, *Aliidiomarina*, and *Alkalimonas* spp. were the least influenced by the tested physicochemical factors (Figure 4.9). Notably, *Halorubrum* and *Haloterrigena* spp. were positively correlated with P and K (R^2= 0.66, $p < 0.001$), but negatively correlated with Mn^+ and CO_3^{2-}. $NH4^+$ and pH positively correlate with the structure of the genus *Salinarcheum* appeared (R^2 0.245; $p < 0.004$), but negatively correlated with NO_3^-. Members of *Alkalimonas*, *Idiomarina*, and *Aliidiomarina* spp. were positively correlated with NO_3^- (R^2=0.049, $p < 0.210$), but negatively correlated with

all other tested parameters. Members of *Nocardiodes*, *Rhodothermus*, *Salinarcheum*, *Salinibacter*, and *Euhalothece* spp. were positively correlated with total dissolved solids (TDS), alkalinity, salinity, CO_3^{2+}, and NH_4^+ (R^2= 0.606, p < 0.001), but negatively correlated with Mg^{2+}, Mn^+, and NO_3^-. On the other hand, Mn^+, temperature $CO_3^{2-,}$ and NH_4^+ negatively affect the structure of *Rhodobaca*.

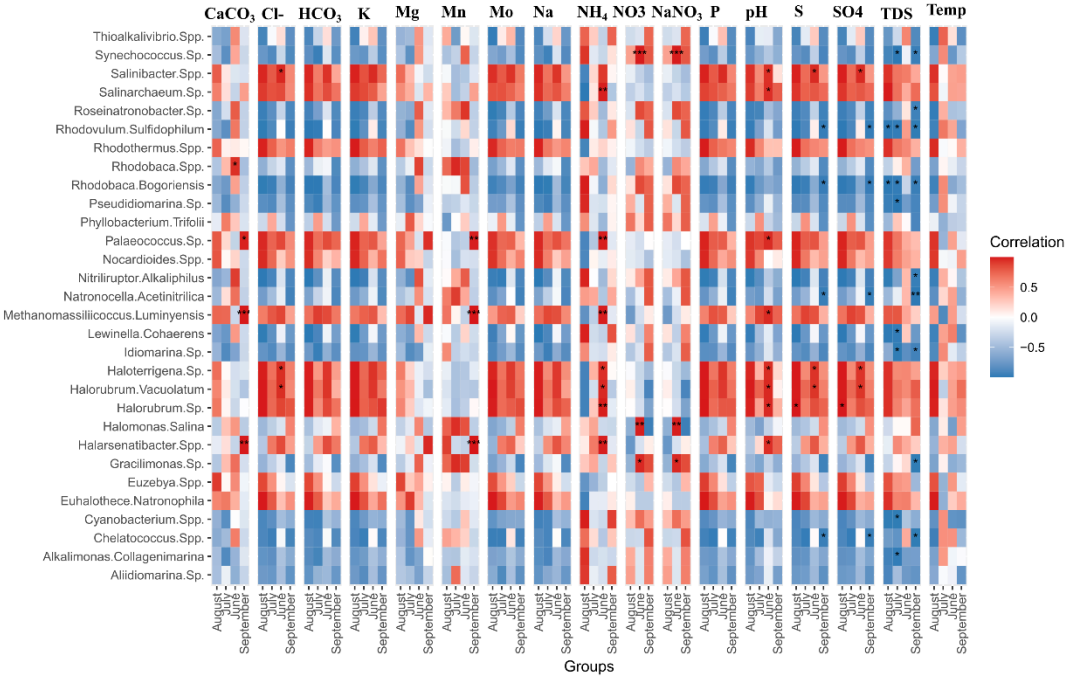


Figure 4.9: The effects of physicochemical factors on the structure of the microbial communities of Lake Magadi.

4.5 Discussion

This study explored the structure and composition of microbial communities based on the time points and physicochemical parameters of Lake Magadi. The physicochemical parameters revealed high concentrations of sodium salts, HCO_3^-, SO_4^{2-}, pH values of 9.8 - 11.5, temperatures of 27 - 38°C, and low concentrations of Ca^{2+}, Mg^{2+}, and Cu^{2+}. These findings were consistent with previous reports indicating that soda lakes are characterized by moderate to high temperatures, high concentrations of $HCO_3^-/CO_3^{2-,}$ and reduced concentrations of Ca^{2+} and Mg^{2+} (Sorokin et al., 2014; Vavourakis et al., 2018). However, total dissolved solids (TDS) ranged from 27 ppm (0.02g/L) to 143 ppm (0.143g/L), a situation that is lower than other soda lakes (Taher, 1999; Hosam et al., 2017; Pérez & Chebude, 2017). Sulfate concentration (120 - 3,030 ppm) concurred with trends in most soda lakes of East

Africa (Lameck et al., 2023) but was higher than the concentration reported in Lake Lancago in Qinghai-Tibet (Wang et al., 2022). Comparatively lower sulfate concentrations have been reported in lakes Sidi Ameur and Himalatt (Algeria) (Boutaiba et al., 2012) and Lake Hamra un Egypt (Mahmoud et al., 2024). The concentrations of the measured parameters (except pH) were variable from site to site. Still, they fluctuated with the sampling season, indicating that the lake chemistry is constantly changing in its constituent elements. A high and stable pH recorded in Lake Magadi is due to high amounts of carbonates that maintain constant pH in the soda lake ecosystems (Simachew et al., 2016). It is postulated that Ca^{2+} and Mg^{2+} precipitate as insoluble carbonates due to high evaporation rates in these ecosystems. As a result, an alkaline brine with Na^{+}, Cl^{-}, and HCO_3^{-}/CO_3^{2-} accumulates as main ions. The shift in $CO_2/HCO_3^{-}/CO_3^{2-}$ equilibrium towards CO_3^{2-}, leads to the formation of a soda (Na_2CO_3) lake with pH values of over 10.0 (Grant and Jones, 2016).

A high number of OTUs was detected for the Domain Bacteria with 3,802 OTUs while 1,035 OTUs were associated with archaea. Similar trends have been shown from soda ecosystems such as Solar saltern in Tunisia (Menasria et al., 2019), Lake Chott El Jerid (Abdallah et al., 2018), hot springs of Lake Magadi (Kambura et al., 2016), lakes Sonachi, Elmenteita, and Bogoria in Kenya (Mwirichia, 2022) and Lake Van (Omeroglu et al., 2021). Members of the phylum Proteobacteria were the most dominant group across all the sampling sites and time points. The role of the members of Proteobacteria such as *Burkholderiaceae* decompose recalcitrant organic matter, while *Beijerinckiaceae* fix atmospheric nitrogen (Li et al., 2012). The phylum Cyanobacteria was represented mainly by the *Euhalothece* spp. *Euhalothece* is a single-celled stenohaline cyanobacterium growing optimally at 7% (w/v) NaCl. They depict a morphological variability depending on the concentrations of NaCl and carbonates and the pH conditions (Mikhodiuk et al., 2008). This genus is the major contributor to nitrogen fixation in soda lakes (Sorokin & Kuenen, 2005). Thus, the presence of *Euhalothece* in Lake Magadi is supported by high salts and carbonates (Table 4.1; Figure 4.6). Contrastingly, Jones, et al (1998) reported that *Arthrospira* spp. are the main photosynthetic agents driving primary productivity in soda lakes, with the seasonal occurrence of *Cyanospira*, *Synechococcus*, and *Chroococcus* to

augment this process Phylum Bacteroidetes was majorly represented by the genera *Rhodothermus* (3%), *Roseinatrobacter* (2.4%), *Gracilimonas* (2.2%), *Lewinella* (1.9%), and *Cytophaga* (0.9%). Verrumicrobia was represented by *Verrucomicrobium* (0.1%), *Puniceicoccus* (0.1%), and *Coraliomargarita* (0.1%). Bacteroidetes and Verrucomicrobia thrive well in high-nutrient environments where they play a role in the degradation of biopolymers such as cellulose and chitin (Newton et al., 2011). Interestingly, the presence of Bacteroidetes and Verrumicrobia was often associated with the presence of Cyanobacteria across the sites and sampling seasons. Photosynthetic stages of cyanobacteria lead to a high rate of CO_2 and HCO_3^- consumption, and a consequent increase in pH (Almeida et al., 2011). As a result, nutrient release is enhanced from the sediments of the lake, hence acting as substrates for Bacteroidetes and Verrucomicrobia. Phylum Firmicutes (2.7% of all the reads) was represented by members of Class Clostridium (*Clostridium*, *Halanaerobium*, *Natranaerobius*, and *Moorella* spp.) and Bacilli (*Alkalibacterium* and *Anoxybacillus* spp.). The two Classes are common in soda lakes with the addition of Tenericutes in some instances (Omeroglu et al., 2021).

The archaeal community of the Lake Magadi microbiome was represented by Euryachaeota, Crenarchaeota, and Thaumarchaeota. Generally, archaea were more abundant in brine samples accounting for 24.5% of total archaea. Previous studies have indicated that archaea are more adapted to saline environments than bacteria (Mani et al., 2020). Euryachaeota was the most abundant group across the sites and the sampling seasons, with 87% of all archaeal communities. Euryachaeota has well-adapted inhabitants of hypersaline environments where members such as *Methanobrevibacter*, *Methanobacterium*, and *Methanocalculus* play a critical role in ecosystem services such as carbon cycling by functioning as methanogens (Jiang et al., 2007; Vavourakis et al., 2016). Euryachaeota was first characterized by Grant et al. (1999) in the alkaline saltern of Lake Magadi. The second most abundant group was Thaumarchaeota also known as ammonia-oxidizing agents (Andreote et al., 2018). In marine ecosystems, their distribution along a salinity gradient has been linked to changes in location, salinity, and sediment depth (Webster et al., 2015). A one-time sampling of the Lake Magadi hot springs depicted a similar picture with Euryachaeota accounting for up to 28% abundance while Thaumarchaeota and

Crenarchaeota were each 1% abundant (Kambura et al., 2016). Studies in other soda lakes have detected members of the phylum Euryachaeota and Crenarchaeota (Ghori et al., 2021), Crenarchaeota, Euryarchaeota, Woesearchaeota, and Pacearchaeota (Wang et al., 2022). However, explorations in Soda Lake in Salina Preta, Brazil, revealed the contrary where members of Thaumarchaeota were most abundant while Euryarchaeota was least abundant (Almeida et al., 2018). The most abundant archaeal genera belonged to *Halorubrum* (18.3%), *Salinarchaeum* (5.4%), and *Haloterrigena* (1.3%). Most of these microbes have their habitats in soda lakes and saline environments (Feng et al., 2005; Mwatha & Grant, 2016; Minegishi et al., 2017; Zhao et al., 2020). The study also revealed the presence of haloalkaliphilic archaea related to the genera *Natronomonas*, *Natrialba*, *Natrococcus*, *Natronobacterium*, *Natronolimnobius*, and *Halorubrum*. These groups have been isolated from brines of East African soda lakes and Inner Mongolian lakes where salinity values reach > 30%, and pH values of >10 (Grant and Sorokin, 2011). Overall, the results of this study reflect microbial composition in many soda lakes around the world (Sorokin et al., 2014; Kambura et al., 2016; Mani et al., 2020; Poyraz & Mutlu, 2020; Wang et al., 2022).

Co-occurrence network analysis demonstrates the interactions between microbial taxa, which can be symbiotic or competitive (He et al., 2019). At the family level the presence of heterogenous microbial communities that co-occur in different sampling sites along the lake and others that were unique to a particular site, suggesting mutual interactions of these communities across the sites. For instance, Desulfobacteriaceae were unique to S2. Correspondingly, S2 had the highest average concentration of sulfur and sulfate ions (Table 4.1). This family, particularly Desulfonatronum, Desulfonatronospira, Desulfonatronovibrio, and Desulfohalophilus have been shown to thrive in anoxic parts of soda lakes acting as sulfate-reducing bacteria (SRB) through the oxidation of hydrogen and formate or direct disproportionation of sulfite of thiosulfate (Sorokin et al., 2011). Noteworthy, *Thioalkalivibrio* sp. was not significantly influenced by the physicochemical properties investigated. This suggests that these strains have devised adaptive mechanisms to thrive under the prevailing salinity and alkalinity conditions of Lake Magadi. Indeed, *Thioalkalivibrio* sp., which are sulfur-oxidizing bacteria (SOB), have been shown to adapt well in soda

lake ecosystems (Li et al., 2022). Unique to the S6 site were the members of the family Puniceicoccaceae which have also been described in four soda lakes of the Cariboo Plateau in Canada (Zorz et al., 2019). Cyclobacteriaceae retrieved from the S1 site have established habitats in diverse ecosystems like cold marine regions, algal/microbial mats, haloalkaline soda lakes, Antarctica, freshwater bodies, marine waters, marine sediments, mangroves, hot springs, and mud volcanoes (Rosenberg et al., 2014). Members of families Rhodobacteraceae and Cyclobacteriaceae have sulfate–oxidizing properties, whereas Burkholderiaceae (unique to S1) have adapted to different ecological niches and are involved in processes such as catabolism of aromatic compounds as well as nitrogen fixation (Pérez-Pantoja et al., 2012).

Alpha diversity studies revealed that samples in the open waters had the highest species richness and diversity. However, open waters samples from S2 - S4 depicted varying degrees of microbial community diversity. While this could not be conclusively explained, it has been hypothesized that water waves affect the distribution and degradation of organic matter degradation and nutrient cycling, thereby influencing the composition diversity of microbes in water ecosystems (Zhu et al., 2018). Brine samples (BR1) collected in June and September had relatively high diversity and evenness indices. Despite high salinity and alkalinity in soda lake brine, the presence of high light intensity and dissolved CO_2 promotes the growth of photosynthetic microorganisms. Subsequently, these phototrophs generate large quantities of dissolved organic matter (DOM) which become substrates for sustaining the diverse microbial communities (Banda et al., 2020). Moreover, hot spring samples (S1) exhibited high Shannon and Simpson diversity indices, indicating high microbial diversity. Research on Soda Lake hot springs revealed a highly active and diverse microbial community, suggesting the high plasticity of these organisms toward extreme environments (Dadheech et al., 2013; Kambura et al., 2016).

The principal component analysis revealed that compared to other samples, hot spring (S1) and brine (BR1) samples clustered according to their sites (cluster I and II, respectively) (Figure 4.3). This suggests that hot spring and brine samples had distinct individual community similarities (Tao et al., 2019). However, sample clustering appeared to be driven by the changes in the salinity and alkalinity of the sampling sites from low to high salinity and alkalinity clusters. Therefore, alkalinity and

salinity rather than sampling season or sites influenced microbial community structure. Previous studies have established that salinity is the primary selective force driving the distribution of beta diversity, whereas alkalinity influences microbial richness (Antony et al., 2013; Boros & Kolpakova, 2018; Banda et al., 2020).

In terms of water chemistry (Figure 4.6), pH, temperature, PO_4^{3-}, K^+, NO_3^-, NH_4^+, Mn^+, Na^+, SO_4^{2-}, and TDS influenced the variation of microbial community composition in Lake Magadi. Specifically, members of genera *Nocardiodes*, *Rhodothermus*, *Haloterrigena*, *Methanomasiliicoccus*, *Halorubrum*, *Palaeococcus*, *Nocardioides*, *Salinarcheum*, *Salinibacter*, and *Euhalothece* had a wide range of physicochemical adaptability. Conversely, *Synechococcus*, *Thioalkalivibrio*, *Cyanobacterium* spp., *Rhodovulum*, *Lewinella*, *Idiomarina*, *Pseudidiomarina*, *Chelatococcus*, *Aliidiomarina*, and *Alkalimonas* were the least influenced by the physicochemical factors. The archaeal genera *Salinarchaeum* and *Halorubrum Halobellus*, *Halolamina*, *Methanobrevibacter*, and *Halorhabdus* are strongly influenced by salinity and pH, Mg^{2+}, Na^+, K^+, Ca^{2+}, and SO_4^{2-} (Han et al., 2017). Nitrate influenced the distribution of the members of the genera *Aliidiomarina*, *Idiomarina*, and *Alkalimonas*. Many strains of *Alkalimonas* have been isolated from Chahannor (China), Kulunda Steppe (Russia), and Elementaita (Kenya) soda lakes where they play a role in nitrate reduction and formation of H_2S (Ma et al., 2004; Vavourakis et al., 2016). However, in this study, sulfur was negatively correlated with *Alkalimonas*. The *Aliidiomarina* and *Idiomarina* belong to the family Idiomarinaceae and have also been described as nitrogen reducers but poor in carbohydrate utilization (Chiu et al., 2014).

4.6 Conclusion

Soda lake ecosystems are a unique habitat for specialized microbial communities that depict temporal changes and patterns of spatial distribution. The current study investigated the diversity and structure of the microbial communities in different sampling sites and time points in Lake Magadi and the influence of physicochemical parameters on the composition and diversity of these microorganisms. The results revealed that in comparison to archaea, bacteria are relatively abundant and diverse in this ecosystem. The most dominant bacterial communities were Proteobacteria,

Cyanobacteria, Bacteroidetes Actinobacteria, Firmicutes, and Verrumicrobia. On the other hand, archaea were majorly composed of Euryachaeota while Crenarchaeota, and Thaumarchaeota were found in the least abundance across the sampling seasons and sites. Species richness and diversity varied within open waters while samples from brine and hot springs revealed high diversity and evenness. Ordination studies indicated that samples clustered based on salinity and alkalinity rather than the sampling month. Temperature, pH, P, K, NO_3^-, and TDS positively influenced the microbial community distribution in Lake Magadi. These microbial communities are well adapted to the soda lake conditions where they contribute to the high productivity of soda lakes. The results advance theoretical frameworks in ecology and evolution while improving knowledge of extremophilic habitats. Future studies combining the functional profiles of samples during and after cyanobacterial blooms, including vertical profile stratification of Lake Magadi could elaborate more microbial communities in different depths.

CHAPTER FIVE

5.0 ARTIFICIAL METAGENOME FOR POLYSACCHARIDE-DEGRADING BACTERIA ISOLATED FROM LAKE MAGADI

5.1 Abstract

Shotgun metagenomic sequencing facilitates comprehensive analysis of diversity, metabolic pathways, and functional genes in a complex microbial sample. This study aimed to screen an artificial metagenome derived from combined genomic material from bacterial isolates from the hypersaline Lake Magadi. The targeted genes were those encoding polysaccharide hydrolyzing enzymes. DNA was extracted from forty morphologically distinct bacterial isolates, mixed in equal amounts, and sequenced on an Illumina platform. The findings revealed the presence of 47,817,430 paired-end reads and 5,162 assembled contigs with a total count of 46,641 putative genes. Among these genes, 45,907 were protein-encoding, 23 CRISPR, and 725 tRNAs. The taxonomic assignment showed that more than 99% of the sequences were affiliated to Firmicutes and Proteobacteria. Genes encoding putative Carbohydrate-Active enZymes (CAZymes) accounted for 10% of all the genes present and included genes encoding for chitinases, cellulases, amylases, and xylanases. Moreover, 3 endoglucanase-encoding genes with a potential signal peptide were detected. In addition, ten metagenome-assembled genomes (MAGs) of medium to high quality were generated. They were affiliated with the genera *Salipaludibacillus*, *Alkalihalophilus*, *Salinicoccus*, *Alkalibacterium*, and *Halomonas*. These findings demonstrate that alkalihalophilic bacteria from Lake Magadi are a rich source of genes encoding for polysaccharide degrading enzymes of industrial applications, such as bioremediation or the production of biofuels.

5.2 Introduction

Soda lakes are aquatic ecosystems characterized by high evaporation rates leading to an increase in soluble sodium carbonates and a consequent increase in pH values. These conditions cause low calcium and magnesium ion concentrations (Sorokin et al., 2014; Vavourakis et al., 2018). Lake Magadi in Kenya is characterized by salt concentrations of up to 30%, temperature ranges of 28 - 86^{o}C, and pH values of 9.0 - 12.5 (Duckworth et al., 1996; Nyakeri et al., 2018). Extremozymes derived from the inhabiting microorganisms are candidates for producing industrial products because the conditions in this extreme ecosystem are similar to those in industrial environments (Simon & Daniel, 2011; Sorokin et al., 2014). Biocatalysts such as salt and pH-stable cellulases can be used in the bio-hydrolysis of recalcitrant cellulose-based biomass derived from agricultural wastes including the generation of climate-friendly biofuels or important industrial building blocks (Saxena et al., 2017).

The largest portion of Earth's biomass, (~75%) are structural carbohydrates such as cellulose, chitin, starch, pectin, and xylan (Thomas et al., 2011; Berlemont & Martiny, 2015; Liu et al., 2019). In soda lake ecosystems, high temperatures, high light intensities, and abundance of CO_2 in the form of HCO_3^- and CO_3^{-2} are responsible for high primary productivity. Therefore, the enzymatic degradation of organic carbon plays a key role in carbon cycling in extreme environments. Heterotrophs produce various glycoside hydrolases (GHs) for the degradation of polysaccharides into monosaccharides which serve as energy and carbon sources for microbial growth (Thomas et al., 2011). However, little is known about the polysaccharide hydrolytic potential of bacteria in hypersaline environments such as Lake Magadi (Sorokin et al., 2015).

Several studies have explored the microbial diversity in Lake Magadi using 16S rRNA gene sequencing (Kambura et al., 2016; Nyakeri et al., 2018; Kiplimo et al., 2019). However, sequencing of 16S rRNA is a taxonomic identifier and does not provide direct insights into the functional capability of the corresponding bacterial strains. On the other hand, shotgun metagenomic sequencing provides insights into the genetic makeup of an environmental sample through random sequencing of DNA fragments. In this study, the genes encoding for Carbohydrate-Active enZymes (CAZymes) were investigated in the pooled chromosomal DNA of 40 bacterial isolates from Lake Magadi using shotgun metagenomic sequencing. This enriches the genetic information to the physiological screenings performed in Chapter 3. Additionally, metagenome-assembled genomes were generated and analyzed for CAZymes.

5.3 Materials and methods

Samples were collected from Lake Magadi, a tropical hypersaline soda lake located at the southern end of the Kenyan Rift Valley situated at 1°43´-2°00´S and 36°13´-36°18´E. It is a saturated hypersaline alkaline lake with a salinity of up to 30% (w/v), annual rainfall of 500 mm, and is fed by approximately 200 hot and warm springs (Behr & Röhricht, 2000; Ruga, 2010). Bacterial isolates were isolated as described in sections 3.3.1 to 3.3.2. Single bacterial colonies were sub-cultured in trypticase soy agar (TSA) supplemented with $NaCO_3$ (1% w/v) and NaCl (4% w/v) (section 3.3.2).

5.3.1 DNA isolation and sequencing of artificial metagenome

Bacterial isolates for DNA isolation were grown in a rotary (37°C; 12h; 180rpm) in trypticase soy broth (TSB) medium (Carl Roth, Karlsruhe, Germany) supplemented with $NaCO_3$ (1% w/v) and NaCl (3.3% w/v). Isolation of chromosomal DNA was done using MasterPureTM complete DNA and RNA purification kit as recommended by the manufacturer (Epicentre, Madison, WI, USA). The quality of DNA was determined in Nanodrop® ND 1000 V. 3.8.1 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA) and counterchecked by 1% agarose gel electrophoresis. A simulated bacterial community was generated by mixing 0.75 µl genomic DNA (100 ng/µl) from each of the 40 isolates. Illumina paired-end sequencing (2 x 250) libraries for the composite DNA were generated using a Nextera XT DNA sample preparation kit and sequenced using an Illumina HiSeq 2500 as recommended by the manufacturer (Illumina, San Diego, CA, USA).

5.3.2 Processing and annotation of metagenome reads

Metagenomic sequences were quality-controlled with the FastQC package (www.bioinformatics.babraham.ac.uk/projects/fastqc/). Sequences were depleted of adapter sequences, reads shorter than 35 bases, and reads with a Phred base quality score of < 33 using Trimmomatic (Bolger et al., 2014). Paired-end reads were merged using Flash v. 1.2.11 (Magoč & Salzberg, 2011) to produce longer reads with minimum and maximum overlap set at 10 and 65 read pairs, respectively. Sequence assembly was performed *de novo* using SPAdes v. 3.14.0 (Bankevich et al., 2012) with K21, K35, K55, K77, K99, and K27 k-mer lengths assayed. The quality of the assembled contigs was performed using Qualimap v. 2.2.2 (Okonechnikov et al., 2016). Open Reading Frames (ORFs) were predicted only for contigs longer than 1,000 bp using Prodigal v. 2.6.3 (Hyatt et al., 2010). Taxonomic assignment of the contigs was done using the Contig Annotation Tool (CAT) v. 5.0.3 (Von Meijenfeldt et al., 2019), and taxonomic profiles were summarized using a Krona chart (Ondov et al., 2011). Determination of proteins involved in various metabolic activities was done by annotations of putative functional ORFs with >100 amino acids in Prokka v. 1.14.5 (Seemann, 2014), while the predicted ORFs were functionally annotated using KEGG Ontology (KO) and GhostKOALA (Kanehisa et al., 2016). Carbohydrate active enzymes (CAZyme) were identified using dbCAN2 web servers (Yin et al., 2012). For this purpose, models DIAMOND (e-value 1e-102), Hotpep (Frequency

>2.6, Hits >6), and HMMER (e-Value <1e-15, coverage >0.35) were utilized identified various classes of polysaccharide degrading enzymes like glycoside hydrolases (GHs), polysaccharide lyases (PLs), glycosyltransferases (GTs), carbohydrate esterases (CEs), auxiliary activities (AA), and carbohydrate binding modules (CBM). To identify potential extracellular located enzymes, signal sequence prediction for classified GHs was performed using the SignalP version 5.0 program (Almagro Armenteros et al., 2019).

5.3.3 Generation of Metagenome assembled genomes (MAGs)

Contigs longer than 2.5 kb were binned based on tetranucleotide clusters and differential depth of coverage patterns using MetaBat2 (Kang et al., 2019). The generated metagenome-assembled genomes (MAGs) were assessed for completeness, and contamination using CheckM v.0.9.7 (Parks et al., 2015), following the Minimum Information about Metagenome-Assembled Genome (MIMAG) standards (Bowers et al., 2017). The non-gap alignment of MAGs to the reference genomes was performed using Bowtie2 (Langmead & Salzberg, 2013). Genome bins with contamination levels of ≤10% and completeness of ≥50% were considered for further analysis. Annotation of their ORFs was done using Prokka v. 1.14.5 (Seemann, 2014) and further functionally using the KEGG database (Kanehisa et al., 2016). Phylogenetic relationships of the MAGs were assessed by employing the Type (Strain) Genome Server (TYGS; https://tygs.dsmz.de), for a whole genome-based taxonomic analysis (Meier-Kolthoff & Göker, 2019). Pairwise comparisons between the MAGs and the closest genomes were done using the Genome Blast Distance Phylogeny approach (GBDP), whereas inter-genomic distances were inferred from Digital DNA-DNA hybridization (dDDH) values. The phylogenetic tree for each genus was inferred using minimum evolution under 100 pseudo-replicates in FastME version 2.1.4 (Lefort et al., 2015). All phylogenetic trees were merged, and a combined phylogenetic tree was viewed and annotated in the iTOL interactive tree of life (Letunic and Bork, 2019).

5.4 Results

5.4.1 Taxonomic assignments of the metagenomic dataset

Sequencing of the synthetic bacterial community resulted in 47,8172,430 assembled nucleotide bases which generated 6,513 assembled contigs ranging from 1001 to 757,980 bp (Table 5.1). The annotated contigs resulted in 46,641 putative genes, out of which 45,907 were protein-encoding, 23 CRISPR, and 725 tRNAs.

Table 5.1: Statistical summary of the artificial metagenome of bacterial isolates from Lake Magadi

Metagenome feature	Count
Number of bases	47,817,430
Number of contigs	5,162
Longest contig (bases)	757,980
Shortest contig (bases)	1,001
N50	37,616
Predicted genes	46,641
Protein coding sequences (CDSs)	45,907
CRISPR	23
RNA genes	953
tRNA	725
rRNA genes	84
5S rRNA	42
16S rRNA	28
23S rRNA	14

In the entire dataset, 99.20% of the contigs could be assigned to a taxonomic group with the dominant phyla being Firmicutes and Proteobacteria. Members of the phylum Firmicutes were the most abundant encompassing 92% of all reads. Most of them were affiliated to *Salipaludibacillus* (83% of all Firmicutes) while the rest could be conclusively classified. Phylum Proteobacteria (3% of all reads) was dominated by *Halomonas* (72% of all Proteobacteria). An interactive version of the Krona chart can be accessed here: https://doi.org/10.6084/m9.figshare.25330150.v1. Assignment to species level was revealed only for *Sb. agaradhaerens* (72%; Figure 5.1). Members of the phylum Proteobacteria were dominated by the class Gammaproteobacteria and the Order Oceanospirillales. Genus *Halomonas* accounted for 95% of this Order. Only *Halomonas* sp. TD01 accounting for 2% could be clarified, whereas 20% of Oceanospirillales was not classified.

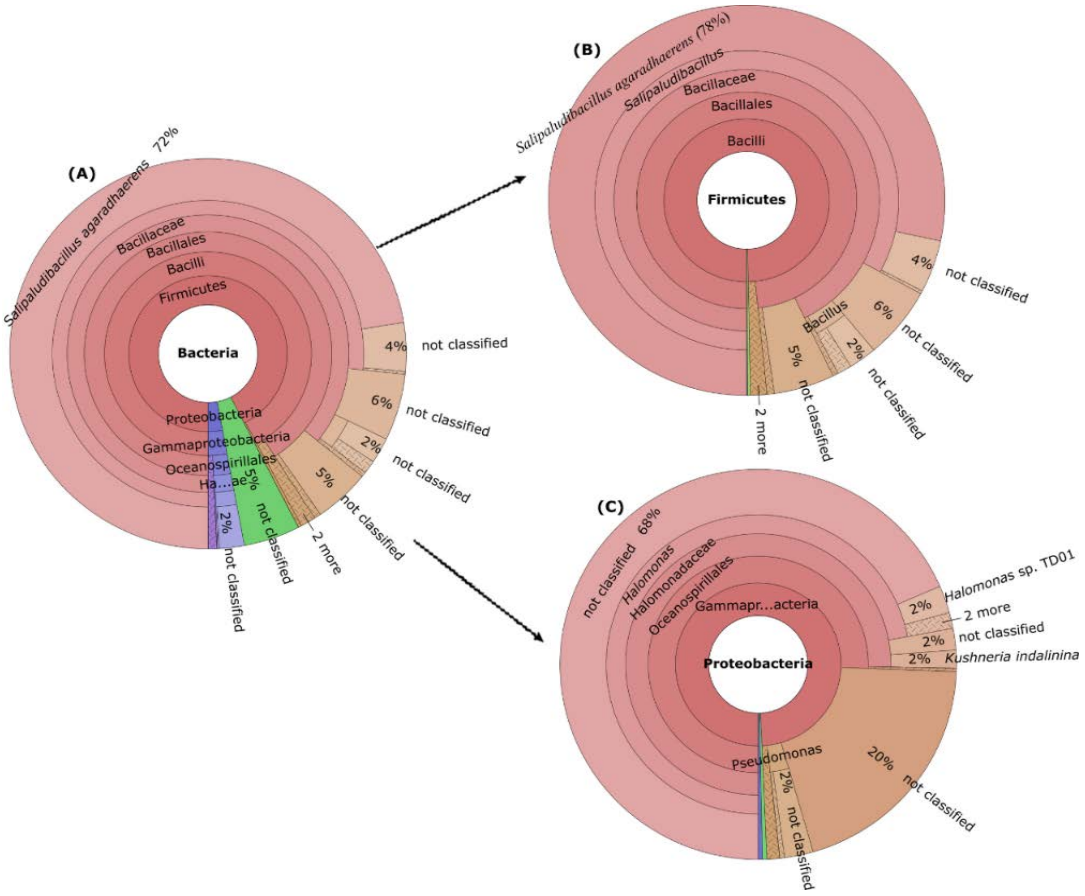


Figure 5.1: Taxonomic classification of two major phyla. **A:** Overall stratification of the entire artificial metagenomic dataset showing the assignment to phylum Firmicutes and phylum Proteobacteria. **B** and **C:** classification based on phylum Firmicutes and Proteobacteria, respectively.

5.4.2 Functional overview of gene categories

Analysis of prokka-generated ORFs using KEGG in the database revealed that 21,254 genes (46.3%) could be assigned to known protein categories. Protein groups involved in genetic information processing were the most abundant (22.6%) as shown in Figure 5.2. Other categories included genes coding for proteins related to signaling and cellular processes (11.6%), carbohydrate metabolism (10.9%), environmental information processing (9.8%), and amino acid metabolism (6.4%).

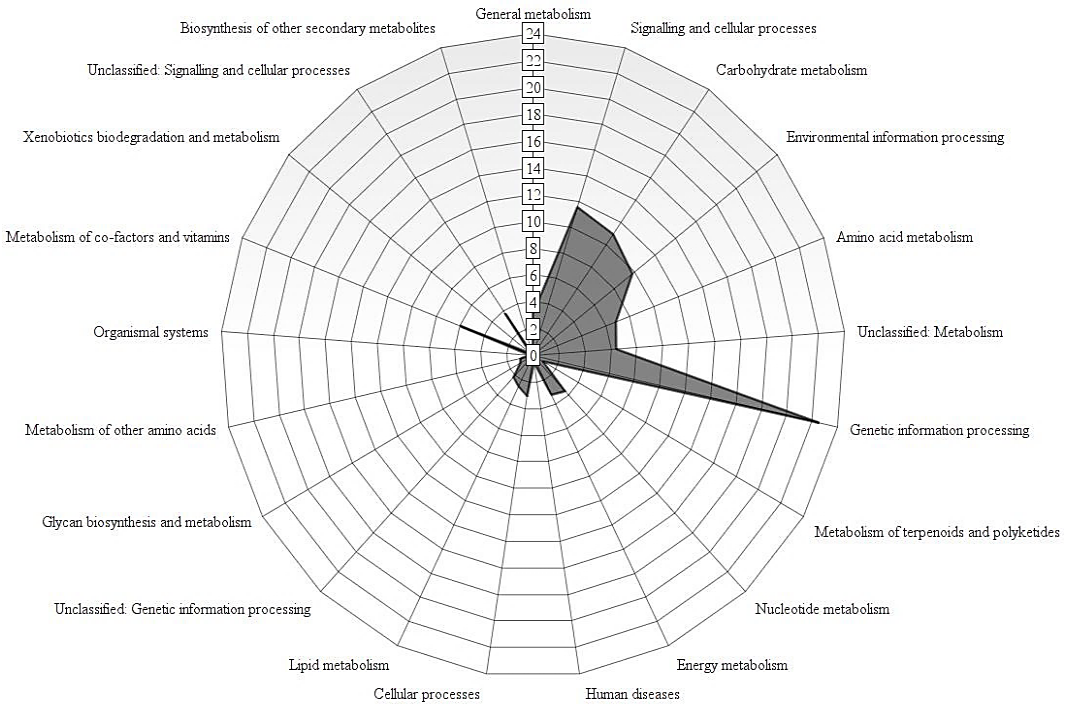


Figure 5.2: Percentage distribution of protein-encoding gene categories based on KEGG assignments.

5.4.3 Analysis and abundance of CAZymes-encoding genes

Sequence annotation with dbCAN2 revealed that 1,152 coding sequences (CDSs) could be assigned to putative CAZymes protein families, accounting for 2.5 % of the total annotated open reading frames. These enzyme-coding genes were distributed across the six classes of CAZyme. Out of these, 458 (39.7%) putative genes encoded glycoside hydrolases, 322 (27.9%) encoded glycosyltransferases, 148 (12.8%) encoded carbohydrate esterases, 161 (13.9%) encoded carbohydrate-binding modules, 54 (4.7%) encoded auxiliary activities and 7 (0.6%) encoded polysaccharide lyases (Appendix 9). Glycoside hydrolase encoding genes comprised of diverse families including those involved in the degradation of starch (GH13), cellulose (GH5), chitin (GH18), peptidoglycan hydrolases (GH23), beta-N-acetylhexosaminidase (GH73), pectin depolymerizing hydrolases (GH28), xylan depolymerizing enzymes (GH10 and GH11), arabinoxylan and pectin de-branching enzymes (GH43 and GH51) (Table 5.2).

Table 5.2: Distribution of the GH families in the artificial metagenome of bacterial isolates obtained from Lake Magadi

GH family	Activity	Firmicutes	Proteobacteria
Cellulases			
GH5	Endoglucanase	9	0
GH8	Exo-oligoxylanase	2	0
GH10	Endo-xylanase	10	0
GH11	Xylanases	2	0
GH28	Polygalacturanase; polygalacturonase	5	0
Disaccharide degrading enzyme			
GH88	chondroitin disaccharide hydrolase	1	0
GH105	Rhamnogalacturonyl hydrolase	8	0
Oligosaccharide degrading enzymes			
GH1	β-glucosidase	32	0
GH2	β-galactosidase	5	0
GH3	β-galactosidase; β-hexosaminidase	15	2
GH4	α-glucosidase; α-galactosidase; β-glucosidase	16	0
GH13	α-amylase; pullulanase	51	2
GH16	Xyloglucan; β-glucanase	4	0
GH18	Chitinases; peptidoglycan hydrolase	13	3
GH27	α-galactosidase	5	0
GH31	α-glucosidase; α-xylosidase	7	0
GH32	sucrose-6-phosphate hydrolase; levanase	15	0
GH35	β-galactosidase	3	0
GH36	α-galactosidase	8	0
GH38	mannosylglycerate hydrolase	5	0
GH39	α-xylosidase	3	0
GH42	α-galactosidase	10	0
GH43	β-xylosidase;α-L-arabinofuranosidase	29	1
GH52	α-xylosidase	1	0
GH53	Arabinogalactan-endo-beta-1,4-galactanase	1	0
GH57	α-amylase	2	0
GH59	β-galactosidase	1	0
GH68	Levansucrase	1	0
GH73	peptidoglycan hydrolase	19	0
GH95	α-L-fucosidase	3	0
GH113	β-mannanase	3	0
GH120	β-xylosidase	2	0
GH146	β-L-arabinofuranosidase	4	0
Debranching enzymes			
GH23	peptidoglycan lyase	11	9
GH33	Sialidase	2	0
GH51	L-arabinofuranosidase	11	0
GH67	α-glucuronidase	3	0
GH85	endo-β-N-acetylglucosaminidase	2	0
GH103	peptidoglycan lytic transglycosylase	0	3
GH125	exo-α-1,6-mannosidase	3	0
GH130	β-1,4-mannosylglucose phosphorylase	4	0
GH161	β-1,3-glucan phosphorylase	6	0

The abundance and distribution of GT, AA, PL, and CE families are shown in Figure 5.3. Most of the CAZyme families were derived from members of the phylum Firmicutes. Notably, the carbohydrate-binding module encoding genes was highly enriched by the CBM50, mainly found in association with GH18, GH23, and GH73. In general, CBMs were majorly found in association with the glycoside hydrolase encoding genes.

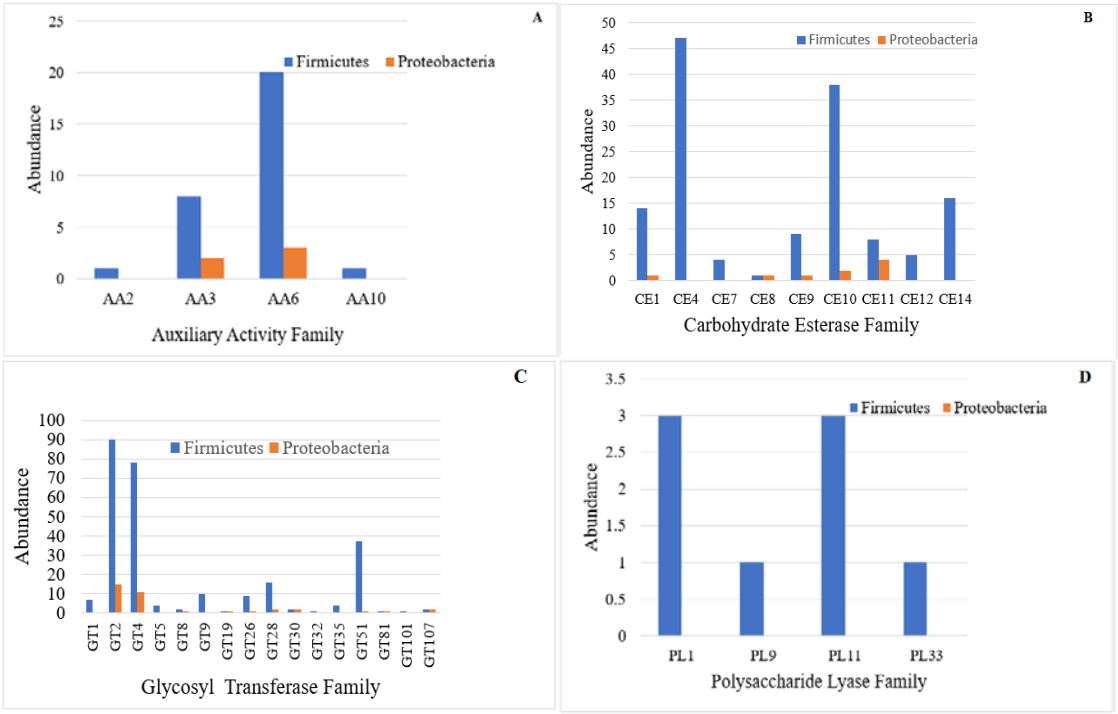


Figure 5.3: Abundance distribution of carbohydrate-active enzyme classes in the artificial metagenome. **A**. Auxiliary Activities; **B**. Carbohydrate esterases; **C**. Glycosyltransferases; **D**. Polysaccharide lyases.

During bacterial isolation (section 3.3.5), several polysaccharides such as starch, chitin, xanthan, pectin, cellulose, and carboxymethylcellulose were included as the only carbon source. Metagenome analysis revealed the presence of genes encoding enzymes responsible for the degradation of most of these polysaccharides. Several of these enzyme-coding genes revealed similarity to the reference genes in the GenBank database (65 – 100%). Furthermore, the SignalP program revealed that most of these genes harbored a putative signal peptide, indicating that they were destined for extracellular secretion (Table 5.3).

Table 5.3: Characteristics of polysaccharide degrading genes obtained from the synthetic metagenome dataset.

Locus Tag	Gene name	Source genus	Signal P value	Length (AA)	Similarity (%)
Cellulose degrading					
LMP_22770	*BglC*	*Halalkalibacterium*	0.94	568	98
LMP_37690	*BglC*	*Salipaludibacillus*	0.99	570	98
LMP_42667	*BglC*	*Salipaludibacillus*	0.98	571	94
Starch degrading					
LMP_11027	*AmyA*	*Salipaludibacillus*	0.99	957	86
LMP_28854	*AmyA*	*Halalkalibacterium*	0.97	615	99
LMP_45260	*AmyA*	*Salipaludibacillus*	0.001	152	97
LMP_01310	*AmyA*	*Halomonas*	0.56	587	98
LMP_09243	*AmyA*	*Alkalibacterium*	0.009	484	100
LMP_34447	*AmyA*	*Salipaludibacillus*	0.001	527	100
Xylan degrading					
LMP_35932	*XynA*	*Salipaludibacillus*	0.93	392	99
LMP_36000	*Xyn11E*	*Halalkalibacterium*	0.99	210	100
LMP_36829	*XynA*	*Salipaludibacillus*	0.99	248	100
LMP_40507	*XynA*	*Evansella*	0.98	388	98
LMP_41281	*XynA*	*Evansella*	0.99	388	98
LMP_09126	*XynA*	*Alkalibacterium*	0.97	1328	91
Chitin degrading					
LMP_33338	*ChiA*	*Halalkalibacterium*	0.98	599	99
Xanthan degrading					
LMP_30619	*gmuG*	*Salipaludibacillus*	0.01	353	93
LMP_05038	*MAN2C1*	*Alkalibacterium*	0.06	1039	65
LMP_37169	*MAN2C1*	*Salipaludibacillus*	0.01	1059	94
Pectin Degrading					
LMP_22844	*PelB*	*Halalkalibacter*	0.99	504	70
LMP_12747	*PelB*	*Halalkalibacterium*	0.99	333	100

5.4.4 Analysis of metagenome-assembled genomes

Binning of contigs >2.5 kb based on their coverage and tetranucleotide frequencies resulted in 10 metagenome-assembled genomes (MAGs) utilizing 1,088 contigs (21% of the contigs in the dataset) with contamination levels ≤3.45% and completeness ranging between 52 to 100%. In summary, there were four medium-quality (≥ 50% complete) and six high-quality (≥ 90% complete) MAGs. The classified MAGs were affiliated with the phylum Firmicutes (Genus: *Evansella*,

Halalkalibacterium, Alkalihalophilus, Alkalibacterium, Salinicoccus), and Proteobacteria dominated by the genus *Halomonas* (Table 5.4).

Table 5.4: Genus-level characterization of the metagenome-assembled genomes obtained from shotgun metagenomic sequencing.

MAG ID	Size (Mbp)	Related taxa (Genus)	Contigs	GC %	Completeness	Contamination
LMP_MAG9	2.36	*Alkalibacterium*	13	42.7	92.35	2.37
LMP_MAG4	3.15	*Alkalihalophilus*	218	40.1	74.14	1.72
LMP_MAG5	3.78	*Halalkalibacterium*	123	43.7	97.33	1.33
LMP_MAG11	4.88	*Evansella*	91	42.1	96.69	1.66
LMP_MAG2	2.59	*Salipaludibacillus*	269	38.5	55.17	0
LMP_MAG3	3.97	*Salipaludibacillus*	197	38.8	69.83	3.45
LMP_MAG1	3.62	*Halomonas*	33	67.9	98.71	1.15
LMP_MAG8	2.64	*Halomonas*	6	64	52.22	0
LMP_MAG10	3.06	*Halomonas*	34	53.1	100	0.43
LMP_MAG6	2.15	*Salinicoccus*	60	48.5	90.62	2.49

The phylogenetic relationship utilized 93 full genome sequences and 10 metagenome-assembled genomes (Figure 5.4). Briefly, clustering yielded 4 genus-specific clusters by their relative full genome sequences.

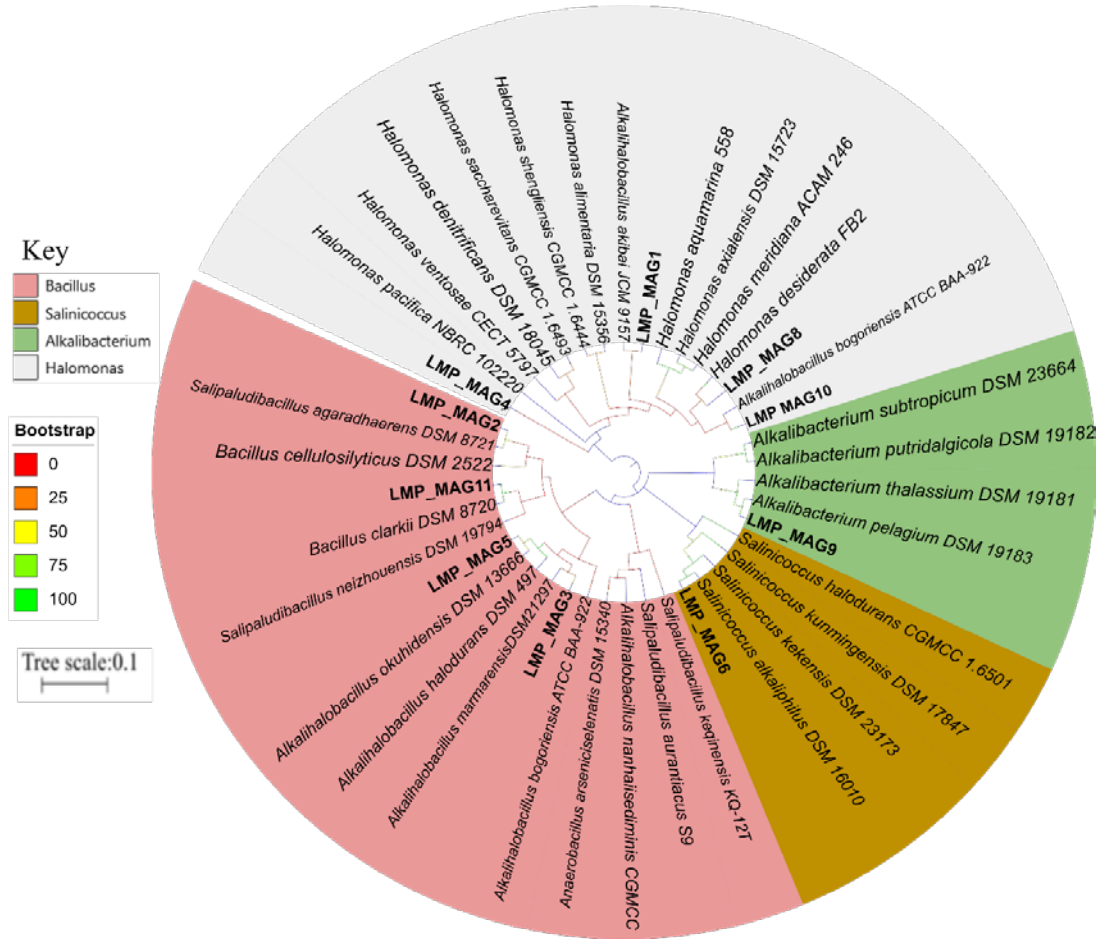


Figure 5.4: Minimum evolution genomic tree for the metagenome-assembled genomes (MAGs). The tree was generated with 100 pseudo-bootstrap replications. Query MAGs are indicated in bold. The colored squares indicate the bootstrap values (shown in the nodes in the innermost circle), and their representative genera are shown in the key. The tree was inferred using FastME and visualized in the iTOL tree.

Classification into protein families via KEGG orthology revealed a varied distribution of functional protein groups. The percentage distribution (Figure 5.5) shows four major protein clusters. The most abundant gene clusters were those involved in genetic information processing (16 - 23%). Genes involved in amino acid metabolism, metabolism-like, metabolism of cofactors and vitamins, carbohydrate metabolism, environmental information processing and signaling, unclassified, and cellular processing ranged between 5 - 12%. Furthermore, genes involved in energy metabolism, lipid metabolism, cellular processes, nucleotide metabolism, and signaling and cellular-like processes exhibited an abundance between 2 - 4%. The least abundant gene clusters were those involved in human diseases, xenobiotics

biodegradation and metabolism, metabolism of other amino acids, organismal systems, metabolism of terpenoids and polyketides, glycan biosynthesis and metabolism, and genetic information processing-like clusters (0 - 1%).

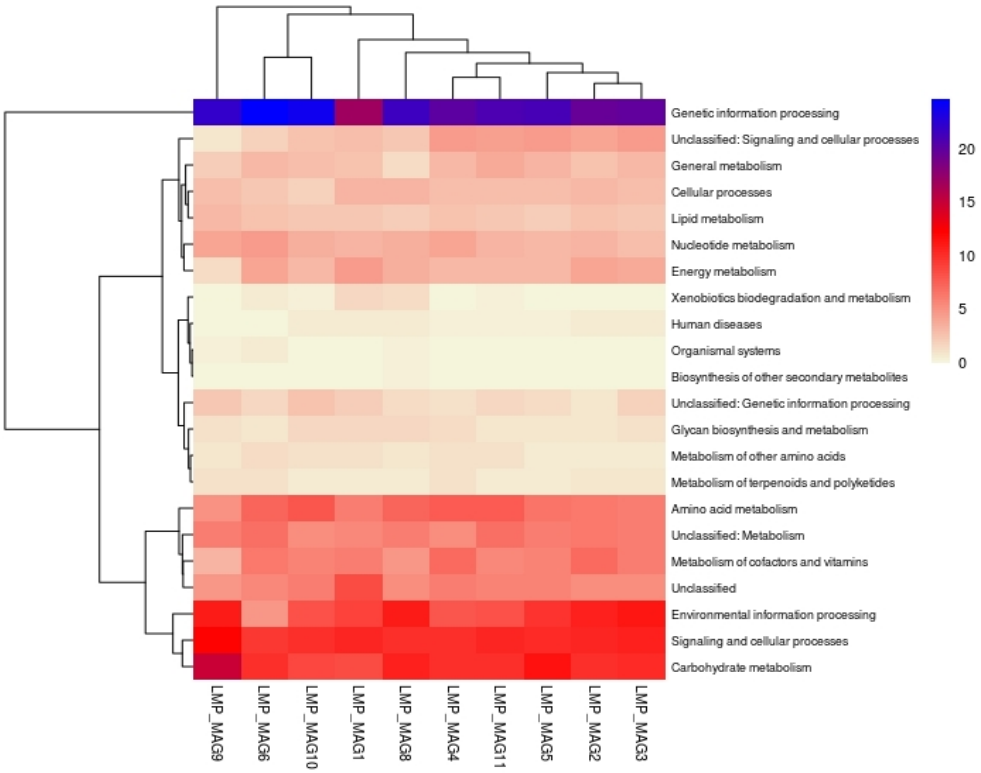


Figure 5.5: Heatmap of various functional protein groups in the individual metagenome-assembled genomes (MAGs) as assigned by the KEGG database. The left side dendrogram shows clustering based on the percentage distribution of protein clusters on individual MAGs. The dendrogram on the upper side represents clustering based on the MAGs.

Grouping into CAZyme classes revealed similar trends as observed in the artificial metagenome. The MAGs affiliated with Firmicutes except *Salinicoccus*-LMP_MAG6- harbored an overall higher abundance of genes encoding for carbohydrate-modifying enzymes than the ones derived from other phylogenetic groups (Figure 5.6). The *Halomonas*-related MAGs (LMP_MAGs 1, 8, and 10) exhibited the lowest number of CAZymes-encoding genes.

.

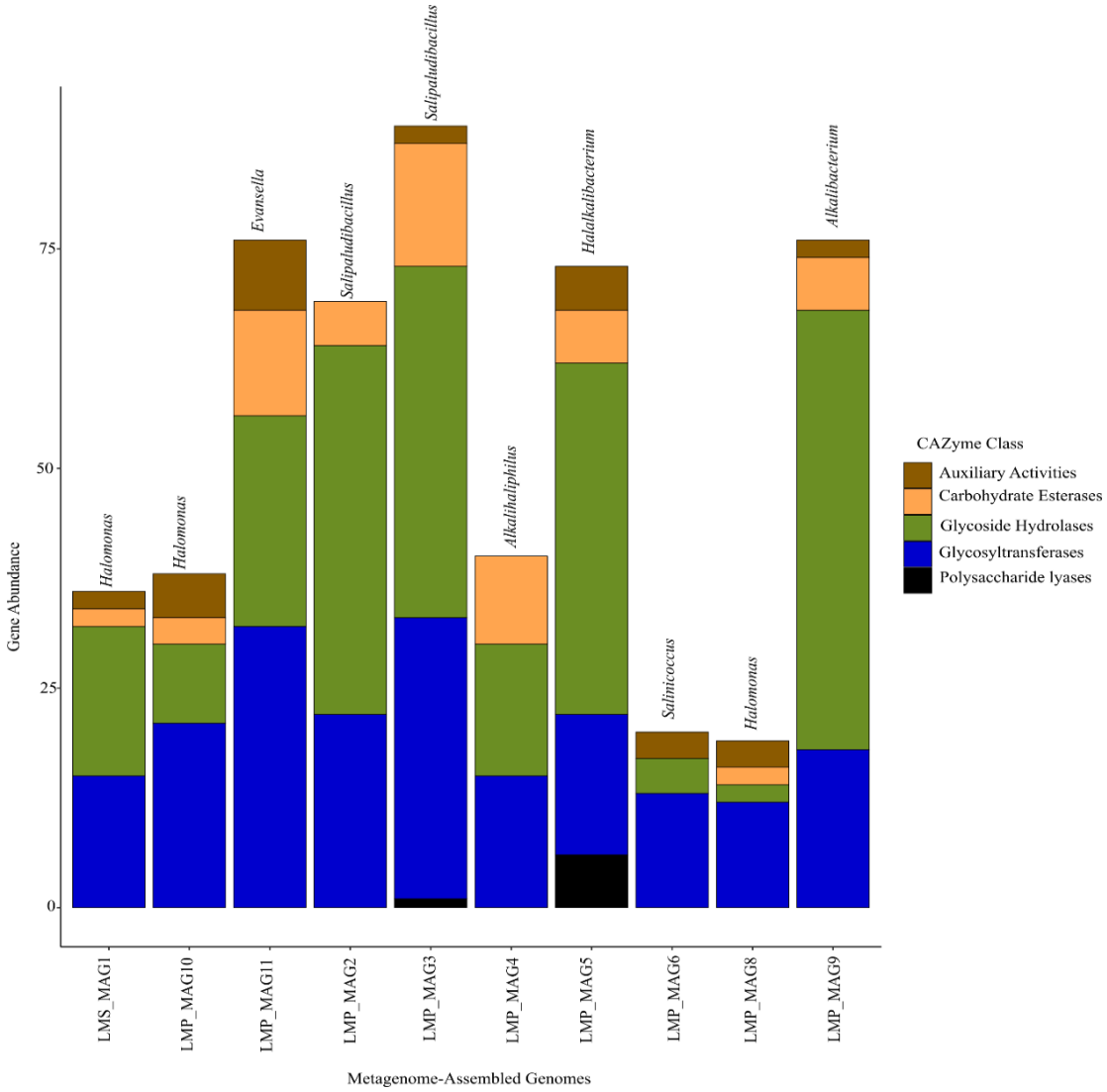


Figure 5.6: Distribution of CAZyme-encoding genes amongst the metagenome-assembled genomes (MAGs). The labels on top of the bars indicate the genus with which the MAG is associated.

5.5 Discussion

5.5.1 Taxonomic identity of the isolates in the artificial metagenome

Taxonomic classification of the metagenomic dataset generated Firmicutes and Proteobacteria as the only groups present. Classification of metagenome-assembled genomes revealed the Phylum Firmicutes was comprised of *Salipaludibacillus* (formerly *Bacillus agaradhaerens*), *Halalkalibacterium, Salinicoccus*, *Alkalibacterium* and *Evansella* spp. These findings are consistent with previous studies that have described the *Salipaludibacillus* group as haloalkaliphilic facultative anaerobes mainly associated with hypersaline soda lakes and possesses

degradative enzymatic activities against polymers such as proteins, starch, and fatty acids (Pandey & Singh, 2012; Sultanpuram & Mothe, 2016; Ibrahim et al., 2019). *Alkalibacterium* with lignocellulosic activity as well as *Salinicoccus* has also been associated with haloalkaline soda lakes (Zhang et al., 2002; Mwirichia et al., 2010; Wang et al., 2017). Phylum Proteobacteria was comprised exclusively of *Halomonas*, which is a phylogenetically heterogenous halophilic or halotolerant genus consisting of over 90 species (Erdyneeva et al., 2018). Their phylogenetic plasticity might explain why about 70% of *Halomonas*-affiliated metagenome sequences could not be classified at strain-level resolution (Berlemont & Martiny, 2013). They have been isolated in numerous environments such as soda lakes, terrestrial saline soils and sands, and solar salterns (Poli et al., 2013; Vamsi Bharadwaj et al., 2015).

Although a preselection criterion was used by employing different carbon sources in this study, the results followed a pattern depicted in previous studies of soda lake environments, which showed that Firmicutes and Proteobacteria were present in high proportions in Lonar Lake, a hypersaline alkaline in India (Wani et al., 2006; Paul et al., 2016); Lake Meyghan in Iran (Naghoni et al., 2017; Vavourakis et al., 2018) and haloalkaline Lake Elementaita in Kenya (Mwirichia et al., 2010). Soda lakes are habitats of various strains of *Alkalihalophilus*, and *Halalkalibacterium* (all formerly named *Alkalihalobacillus*) as shown by studies performed in three Ethiopian soda lakes. Here, starch hydrolyzing isolates were affiliated with *Alkalihalophilus*, *Halalkalibacterium*, and *Sb. agaradhaerens* (Martins et al., 2001). In Lake Magadi, culture-dependent isolation methods have shown that the soda lake ecosystem is rich in potentially different members of the formerly known as *Alkalihalobacillus* and *Halomonas* (Sahay et al., 2012; Nyakeri et al., 2018; Mulango et al., 2020).

5.5.2 Strains with novel functional potential

An overview of the total annotated protein clusters in KEGG reveals that proteins involved in genetic information processing were the most prevalent cluster of proteins in this dataset. These, according to KEGG, are genes participating in manipulation mechanisms of genetic material such as replication, transcription, translation, and DNA repair. Consequently, the results may suggest that these organisms have relatively high activity. It can be deduced that organisms living in extreme

environments such as soda lakes have a high energy demand to sustain life in this ecosystem (Aanderud et al., 2016; Gunde-Cimerman et al., 2018). Consequently, compared to other functional proteins, there were higher amounts of genes encoding the metabolism of various compounds like carbohydrates, amino acids, lipids, and vitamins to help bacteria cope with hypersaline alkaline conditions.

5.5.3 Carbohydrate-based metabolic genes

CAZyme assignment of genes showed the glycoside hydrolase (GH) class as the most abundant group (39.7%). In this class, the family GH13, found in both Proteobacteria and Firmicutes, accounted for 17% of the class. These enzymes consisted of α-amylases, pullulanases, α-1,6-glucosidases, branching enzymes, maltogenic amylases, neopullulanases, glucosylglycerate phosphorylases, and cyclodextrinases. Among them, amylases are the most crucial enzymes in the food and feed industry as they generate and reduce viscosity in maltose and glucose syrups, aid in juice clarification, are involved in product formation in the baking industry, and solubilize starch in the brewery processes (Labes et al., 2008). Production of amylases by members of Proteobacteria is a rare but not a new case. For instance, the isolation of microbes from alkaline Lake Bogoria revealed the presence of starch, amylopectin, and Arabic gum-depolymerizing *Hl. desiderata* strains (Vargas et al., 2004).

The presence of genes encoding the GH5 family, with exclusive affiliation to the Firmicutes in the metagenomic dataset, demonstrates that these bacterial communities of Lake Magadi have the potential to produce cellulases to degrade cellulosic materials. The family contains endo-β-1,4-glucanases and cellobiohydrolases, supplemented with hemicellulolytic β-glucosidase (GH1, 2, and 3). Together, they act in the simultaneous breakdown of crystalline cellulose (Berlemont and Martiny, 2013). Endo-β-1,4-glucanases cleave cellulose from the internal sides, whereas exoglucanases (cellobiohydrolases) generate cellobiose by attacking the reducing and non-reducing ends of cellulose. Cellobiose is then degraded by beta-glucosidases, an oligosaccharide degrader via hydrolysis of beta-linked dimers (Suurnäkki et al., 2000; Potprommanee et al., 2017). GH5 family enzymes are used in industrial processes that include baking, beer brewing, producing fermented foods and syrups flavoring, and enhancing the texture of substances

(Coughlan et al., 2015; Berini et al., 2017). Alternatively, GHs are used to eliminate exopolysaccharides (biofilms) excreted by disease-causing microorganisms (Costa et al., 2018). This study identified new enzyme members of the GH5 family, which might bear potential for industrial processes due to their extremophilic origin.

Notably, the families of GH involved in oligosaccharide degradation formed the largest part of the GH class. They consist of α-L-fucosidase, ß-galactosidase, ß-glucosidase, ß-glucuronidase, ß-mannosidase, ß-arabinofuranosidase, α-L-arabinofuranosidase, ß-hexosaminidase, mannosylglycerate hydrolases, and ß-xylosidase, enzymes cleaving side chains of hemicellulases and pectins as well as non-reducing carbohydrates in oligosaccharides (Harvey et al., 2000; Kimbrel et al., 2018). The presence of GH43 with 29 genes linked to Firmicutes suggests the availability of enzymes for debranching and hence degradation of hemicelluloses, like pectin and arabinoxylans (Mewis et al., 2016). Genes encoding the GH18 family derived from polygacturanases were found only amongst the Firmicutes. They participate in the de-esterification of the cortical fragment of the cell wall. Therefore, they play a crucial role in the depolymerization of peptidoglycan during spore formation and subsequent entry into dormancy (Moriyama et al., 1996; Heffron et al., 2011). Because the adaptive mechanism is essential for the survival of bacterial communities in hypersaline environments, microbes must enter an inactive state when exposed to high temperatures, minerals, and alkalinity. On the other hand, GH 18 harbors enzymes that depolymerize chitin, the second most abundant homopolymer after cellulose on Earth.

Certain marine bacteria degrade and utilize chitinous waste from crustaceans like lobsters, shrimps, and crabs as rich nitrogen and carbon sources. In the event, they contribute to C and N cycles (Hoster et al., 2005; Khandeparker et al., 2013; Asif et al., 2019). Furthermore, numerous isolates of *Bacillus*, *Aeromonas*, *Serratia*, *Pseudomonas*, and *Xanthomonas* also utilize chitinases to protect themselves against potential pathogenic fungi such as *Fusarium oxysporum*. Beneficially, this mechanism can now be harnessed to control fungal infections in agricultural fields (Gohel et al., 2006; Suryanto et al., 2010).

Glycosyltransferases (GT), a group of enzymes involved in the transfer of sugar molecules between the donor and acceptor molecules, were dominated by GT2, GT4, and GT51 (64.3%). GT2 and GT4 form a multifunctional enzyme family involved in bacterial cell-wall synthesis (DeAngelis et al., 2010). The class auxiliary activities (AA) comprises molecules used in the dissolution of non-carbohydrate structural components of the cell wall such as lignin (Park et al., 2018). The most abundant group was AA6 (benzoquinone reductases) (37%). They were linked to Firmicutes and play a role in the metabolism of xenobiotic compounds such as aromatic hydrocarbons (Kim, 2002). AA1 and AA2 were the second most abundant representatives of the AA group (33.3%). They are known for their activity in lignin degradation with laccase-like multicopper oxidase and lignin peroxidases (Levasseur et al., 2013). Others included AA3 involved in propagating the action of GHs in degrading lignocellulose. They include alcohol oxidase, aryl alcohol oxidase, glucose oxidase, and pyranose oxidase (Sützl et al., 2018). The results also showed one gene related to AA10, a family taking part in chitin depolymerization by the Firmicutes group. Also present and exclusively linked to Firmicutes were the genes encoding polysaccharide lyases (PLs). These comprised the PL1, PL9, PL11, and PL33 types (Figure 4.2) and were affiliated with pectin lyase, rhamnogalacturonan endolyase, and hyaluronate lyase. These are pectinolytic enzymes that cleave the α-1,4 bonds of the highly esterified pectin polymer (Alaña et al., 1991).

Mostly, polysaccharide hydrolysis is enhanced in the presence of non-catalytic, self-folding carbohydrate-binding modules (CMBs). They are known to invariably avail and anchor the catalytic modules of the enzyme to the surface of the substrate, hence increasing substrate turnover (Abbott & Boraston, 2012). Here, several CMBs were found in association with GH families and polysaccharide lyases degrading insoluble polysaccharides. They included CBMs involved in the hydrolysis of cellulose (CMB46 and CBM59), xylan (CMB6, CMB9, CBM22, and CBM60), starch (CBM25, CBM26, and GH41), bacterial cell wall and/or chitin (CBM50, CBM54), neopullulan (CBM34), glycogen (CBM48), sialic acid (CMB40), mannan (CBM35), and inulin (CBM38).

5.5.4 Metagenome-assembled genomes and their role in carbohydrate metabolism

Binning of the artificial metagenome of Lake Magadi samples resulted in 10 MAGs ranging from mid to high quality (≥ 55% completeness; <4% contamination). The MAGs presented a near-homogenous KEGG protein distribution as depicted in the artificial metagenome. In several instances (except LMP_MAG6 and LMP_MAG8), GH and GT classes predominated. Generally, most of them possessed genes that encode several carbohydrates metabolizing enzymes such as oligosaccharide degraders (GH1 and GH3), amylases (GH13), chitinases (GH18 and GH28), polysaccharide lyases, and endoglucanases (GH5). This reveals that some of the MAGs possess genes encoding the enzymes that decompose cellulose, starch, chitin, and pectin derivatives.

5.6 Conclusion

The synthetic shotgun metagenomic sequencing employed in this study yielded members of the Firmicutes and Proteobacteria as the major phyla derived from the enrichment method. These results agree with those of 16S rRNA gene sequencing albeit with limited clarification to species level. However, further classification into metagenome-assembled genomes (MAGs) classified the contigs into 10 mid and high-quality MAGs associated with *Salipaludibacillus*, *Halalkalibacterium*, *Salinicoccus*, *Alkalibacterium*, and *Evansella* spp. The technique also provided insights into the functional and metabolic capabilities of the members of Firmicutes and Proteobacteria. It revealed that the lake ecosystem is a potential source of polysaccharide-depolymerizing bacteria. Particularly, the phylum Firmicutes is enriched with the genes encoding for carbohydrate decomposition. The artificial metagenome also demonstrated that a large portion of the proteins (about 54%) are without assigned functions. This presents an opportunity for the assignment of full functionality in future studies. Furthermore, the diversity and enzyme activity across seasons would be revealed by metagenomic and metatranscriptomic analysis of the lake samples. Nevertheless, these findings underscore the potential of Lake Magadi to contribute to advancements of knowledge and enhance the understanding of metabolic reactions in bacterial communities in soda lake ecosystems.

CHAPTER SIX

6.0 HETEROLOGOUS EXPRESSION AND CHARACTERIZATION OF SALT-ACTIVATED ENDOGLUCANASE FROM AN ARTIFICIAL SODA LAKE METAGENOME

6.1 Abstract

Cellulases from alkalihalophilic bacteria are important enzymes due to their ability to withstand harsh conditions commonly associated with industrial settings. This study aimed to sequence, clone, and characterize a salt-activated endo-β-1,4-glucanase from an artificial shotgun metagenome of bacterial isolates obtained from a hypersaline Lake Magadi, Kenya. DNA from 40 bacterial isolates was mixed in equal proportions and sequenced using Illumina paired-end 2 x 250 bp runs. Sequences obtained were quality-filtered, assembled, and annotated using Trimmomatic, SPAdes, and Prokka packages, respectively. Four putative endoglucanase genes with locus tags LMP_22770, LMP_37690, LMP_42667, and LMP_45580 derived from the metagenomic dataset were selected for cloning. However, only gene *LMP_42667* could be cloned and expressed successfully. It had an open reading frame (ORF) of 1,716 bp, 571 amino acids, and a theoretical molecular weight of 65.7 kDa. Sequence homology searches showed that the gene had 99% similarity to an endo-β-1,4-glucanase from an uncultured bacterium. *LMP_42667* gene products belong to glycosyl hydrolase 5 subfamily 4 (GH5_4), with two carbohydrate-binding modules (CBM46 and CBM_X2). The gene was cloned into the pBAD18 vector and heterologous expression was performed in *E. coli* TOP10 cells. The purified enzyme (3.24 µg/ml) demonstrated optimum activity at a pH and temperature of 7.0 and 50°C, respectively. The gene product exhibited activity against carboxymethyl cellulose, hydroxyethylcellulose, microcrystalline cellulose, β-glucan, lichenan, and xylan. Enzyme activity was observed across a wide range of NaCl concentrations with optimum activity at 5M. The enzyme was stable when exposed to several metal ions, surfactants, protein inhibitors, and organic and inorganic solvents. Enzyme kinetics on CMC as a substrate resulted in V_{max} and K_m of 18.7 U/mL and 3.103 µmol/min/mg, respectively. Based on these results, the enzyme is characterized as an extreme halophile capable of hydrolyzing polymers in the presence of chemical reagents and, therefore, a potential candidate for industrial application.

6.2 Introduction

Cellulosic plant materials are the most abundant, renewable, and economical biomass in nature (Kucharska et al., 2018; Pang et al., 2019; Ariaeenejad et al., 2020). With the dwindling reserves of fossil reserves compounded by its contribution to climate change, enzymatic conversion of lignocellulose could be an alternative to the generation of biofuels such as bioethanol, butanol, and methane (Couturier et al., 2011; Isikgor & Becer, 2015). Polysaccharide conversion is successively carried out by a group of extracellular glycoside hydrolases secreted by fungi, bacteria, and

actinomycetes (Vadala et al., 2021). While fungi have been the main source of glycosyl hydrolases, bacterial sources have recently gained interest due to their ability to grow faster, high adaptability, and susceptibility to genetic engineering (Sadhu, 2013; Biswas & Paul, 2017).

Extreme environments such as hypersaline lakes are attractive biomes for research and the discovery of novel cellulases (Ogonda et al., 2021). These ecosystems are characterized by high temperatures, low oxygen, high salinity, and alkalinity (Ventosa et al., 2015). Therefore, it is postulated that extracellular enzymes obtained from bacteria living in hypersaline lakes can withstand a wide range of temperature, pH, and salt concentrations (Gomes & Steiner, 2004). In cellulose degradation, endo-β-1,4-glucanase (EC 3.2.1.4) is responsible for the cleavage of β-1,4-glycosidic linkages, resulting in the production of cello-oligosaccharides. Compared to their mesophilic counterparts, halophilic and halotolerant endoglucanases are active and stable over a wide range of salinities. Therefore, they could be utilized in industrial processes where the presence of high salt concentrations could hinder enzymatic activities (Ruginescu et al., 2020).

In the production of bioethanol, the high cost of biocatalysts accounts for approximately 40% of the total process (Mori et al., 2014; Pabbathi et al., 2021). Therefore, there is a need for continuous research on novel glycoside hydrolase capable of performing under elevated temperatures, pH, salinity, and surfactants commonly encountered in an industrial setup (Wohlgemuth et al., 2018; Li et al., 2020). In this work, a β-endoglucanase gene deduced from a shotgun metagenomic sequence of mixed isolates from soda Lake Magadi was cloned and expressed in *Escherichia coli* cells. The purified enzyme was shown to be an extreme halophile that is stable in moderate temperatures and the presence of metal ions and organic and inorganic chemicals. These features, therefore, qualify it as an ideal candidate for application in bioethanol production.

6.3 Materials and methods

6.3.1 Amplification of metagenome-based endoglucanase genes

DNA from 40 bacterial isolates was extracted from 12 h-grown cultures as described in chapter three (section 3.4.1). Four putative endoglucanase genes with locus tags LMP_22770, LMP_37690, LMP_42667, and LMP_45580 derived from the

metagenomic dataset were selected for cloning. Primers specific to these genes were designed using Clone Manager 9 (https://www.scied.com) with the inclusion of restriction endonuclease enzyme cutting sites. The amplification primers were synthesized by Merck (Darmstadt, Germany). Amplification of the genes was done on a Mastercycler gradient thermocycler (Eppendorf®) on all 40 isolates to establish the source of the genes. Amplification was done in a 50 µl reaction volume containing: 5X GC buffer, 10µl; 20µM of each primer, 0.5µl; 10 mM dNTPs, 1µl; Phusion high fidelity polymerase, 0.5µl; 25 mM $MgCl_2$, 0.20µl, and ddH_2O, 32µl. PCR conditions included initial denaturation for 1 minute at 98°C followed by 30 cycles of denaturation for 45s at 98°C, annealing (Table 6.1) for 30s, extension for 30s at 72°C, and a final extension of 5 min at 72°C. The PCR products were checked on a 1% (w/v) agarose gel.

6.3.2 Cloning and confirmation of gene orientation

Chemically competent *E. coli* TOP10 cells (genotype; *F–mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15ΔlacX74 recA1 araD139 Δ(ara-leu)7697 galUgalK λ– rpsL (StrR) endA1 nupG*) (Thermo Fisher Scientific, Waltham, MA, USA) were used as the host systems for cloning. Cloning was done using the pBAD18 vector expression system (Guzman et al., 1995). Endoglucanase genes and the vector were double digested with their respective restriction enzyme (Table 6.1). They were purified and ligated using T4 DNA ligase (Thermo Fisher Scientific, Waltham, MA, USA). Thereafter, *E. coli* cells were transformed with recombinant pBAD18 using a heat shock at 42°C for 60 seconds, and 600 µl of LB + ampicillin (100µg ml^{-1}) medium was added immediately. The cells were grown by shaking (180 rpm, 37°C, 30 min). Subsequently, 200 µl of the cells were spread on LB-agar media and grown overnight at 37°C.

A colony PCR was done using the previously described amplification conditions (section 6.3.1). Recombinant pBAD18 vectors were then isolated from positive colonies using Qiaprep® Spin kit (Qiagen GmbH, Hilden, Germany) according to the instructions of the manufacturer. To check for gene fidelity, primers flanking the insert on the vector's multiple cloning sites (MCS) were used to amplify the insert. The primers (FidelF and FidelR; Table 6.1) were used to set up a PCR mix as follows: 5X GC buffer, 10µl; 20µM of each primer, 0.5µl; 10 mM dNTPs, 1µl; Phusion^R high

fidelity polymerase, 0.5µl; 25 mM $MgCl_2$, 0.20µl, and ddH_2O, 32µl. PCR cycling conditions included an initial denaturation step of 1 min at 98°C followed by 25 cycles of denaturation for 45 s at 98°C, annealing (Table 6.1) for 30 s, extension for 30 s at 72°C and a final extension of 5 min at 72°C. The amplicons were sent for Sanger sequencing at Seqlab (Goettingen, Germany).

Table 6.1: Characteristic features of primers used to amplify putative endoglucanases.

Locus tag	Primer name	Primer sequence (5'- 3') *	Restriction enzyme	Annealing temp. (°C)	Source
LMP_22770	LMP22770F	**GAGCTC**CTGAGCTGCTATCCCAGTCAC	SacI	68.3	This study
	LMP22770R	**TCTAGA**ATGGCATGGTTGGCCGTTG	XbaI		
LMP_37690	LMP37690F	**GAGCTC**ATGGGTTATACCAAAGCGAAG	SacI	58.8	This study
	LMP37690R	**AAGCTT**TTACCGTCTACCTTGAACATAATTG	HindIII		
LMP_42667	LMP42667F	**AGCTCA**TGGGTTATACCCAAGCTAAGTG	SacI	65	This study
	LMP42667R	**TCTAGA**TTACCGTCTACCTTGAACATAATTTCC	XbaI		
LMP_45580	LMP45580F	**GAATTC**ATTTCATTTTTATGCAGGAACAC	EcoRI	57	This study
	LMP45580R	**AAGCTT**TTAATTGAGTGGTTCCCACG	HindIII		
pBAD18-R	FidelR	ATGCCATAGCATTTTTATCC	None	65	This study
pBAD18-R	FidelL	GATTTAATCTGTATCAGG	None	65	This study

*The bold and underlined sequences indicate the restriction enzyme identification and cutting sites.

6.3.3 Heterologous expression

For over-expression, BL21 (D3) *E. coli* cells were transformed with recombinant pBAD18 as previously described. The transformants were grown in 50 mL LB media supplemented with ampicillin (100 µg ml^{-1}). The preculture (OD_{600} = 0.1) was inoculated into a 600 mL LB medium and grown until OD_{600} between 0.6 - 0.8 was

achieved. Protein expression was induced by adding 1% L- (+)-arabinose (Sigma-Aldrich Chemie GmbH, Munich, Germany) and expression occurred for 12 hours.

Cells were harvested by centrifuging the broth cultures (6,000 rpm, 4°C, 10 min). The cell pellet was washed twice with sodium phosphate buffer (pH 7.0) and then re-suspended in LEW1 buffer included in the Protino-TED 2000 protein purification kit (Macherey-Nagel). DNase I (1 µg/ml) and lysozyme (10 µg/ml) (AppliChem, Darmstadt, Germany) were added to the cells and incubated on ice for 35 minutes. Cell disruption was performed by passing the cells three times through a French press at 6,894,757 pascals. The cell lysate was centrifuged (8,000x g, 4°C, 20 min) and the supernatant was filtered through a 0.45 µM filter to remove any remaining cell debris. The concentration of the crude protein was determined via Bradford assay with bovine serum albumin as a standard (Bradford, 1976).

6.3.4 Cloning and purification of His_6-tagged endoglucanase

Only one (*LMP_42667*) of the four endoglucanase genes was successfully expressed in *E. coli* cells. Therefore, the *LMP_42667* gene was re-amplified using the following set of primers: 42667F: (5'-CG**GAATTC**ATGGGTTATACCCAAGCTAAGTG-3') and

42667R:(5'T**TCTAGA**TTAATGGTGGTGGTGATGATGCCGTCTACCTTGAA CATAATTTCC-3'). Sequences encoding the His_6 tag were added to the C-terminal end (underlined) and *Xba*I and *Eco*RI restriction sites (bold). Cloning and expression were done as described previously, and the resultant recombinant product was designated pBAD18-42667Rec. $Hist_6$-tagged protein was purified by metal affinity chromatography using Protino® Ni-TED 2000 packed columns (Macherey Nagel, Düren, Germany). It was then de-salted by ultrafiltration in Vivaspin 20 (30,000 MWCO) ultrafilters (Sartorius-Stedim, Goettingen, Germany) according to the manufacturer's protocol. The purity of the enzyme was checked in a 12% (w/v) sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) as described by Laemmli (1970). The concentration of the protein was determined via Bradford assay. The activity of the pure enzyme was performed by spotting 10 µl of 3.02 mg/ml enzyme on LB-agar medium supplemented with 1% (w/v) low-viscosity CMC.

6.3.5 Bioinformatic analyses of the amino acid sequence

Amino acid sequences were checked for closest relatives in the GenBank using the BLASTp algorithm. Open reading frames (ORFs) of the protein-coding *LMP_42667* gene were predicted by the ORF finder (https://www.ncbi.nlm.nih.gov/orffinder/). The physicochemical features of the *LMP_42667* encoded protein were determined using the Protparam suite (Gasteiger et al., 2005). SignalP 5.0 (Juan et al., 2019) was used to predict the signal peptide sequence for cellular localization, while the assignment of glycoside hydrolase family, structural, and functional domains was predicted using the dbCAN2 web application (Zhang et al., 2018). Modeling of the 3-D molecular structure was done in SWISS-MODEL (Waterhouse et al., 2018) and counter-checked in Iterative Threading ASSEmbly Refinement (I-TASSER) web servers (Yang et al., 2015). From the Protein Databank (PDB) and NCBI databases, four putative cellulase genes were inferred to perform multiple sequence alignment. The genes were isolated from *Bacillus* sp. BG-CS10 (5e09.1.A; Zhang et al., 2015), *B. halodurans* (4V2X; Venditto et al., 2015), uncultured bacterium (CAD61243; Ghauri et al., 2003), and *Salipaludibacillus agaradhaerens* (WP_078577688; Davies et al., 1998). Multiple sequence alignment was then done using Muscle (https://www.ebi.ac.uk/Tools/msa/muscle/). The alignments were exported to ESPript 3.0 (Robert & Gouet, 2014) for annotation of the secondary structure. Phylogenetic analyses were done to compare the sequence and its homologs in the GenBank and multiple sequence alignment was performed in MAFFT (Katoh & Standley, 2013). The phylogenetic tree generated via the neighbor-joining algorithm was then viewed in iTOL tree v4 (Letunic & Bork, 2019).

6.3.6 Enzyme assay and substrate specificity

Unless otherwise mentioned, standard assays involving *LMP_42667*-catalyzed hydrolysis were conducted using dinitrosalicylic acid (DNS) assay with D-glucose as a standard (Miller, 1959). Substrate specificity tests were performed on low viscosity carboxymethyl cellulose (CMC) sodium salt (1%) (Sigma, St. Louis, MO, USA), 1% β-glucan from barley (Sigma-Aldrich), 1% starch from potato (Sigma-Aldrich), 0.5% mannan (Sigma-Aldrich), 1% hydroxyethyl cellulose (Sigma-Aldrich), 1% xylan from birchwood (Sigma-Aldrich), 0.5% chitin from crab shells (Sigma-Aldrich), and 1% cellulose microcrystalline (Serva, Heidelberg, Germany). Each substrate was

mixed in 50 mM sodium phosphate buffer (pH 7.0). A unit of the enzyme was defined as the amount of the enzyme required to release 1 μmol of a reducing sugar under standard assay conditions (Nacke et al., 2012). All experiments were performed in triplicates, and the results were plotted as mean values with standard errors of the mean.

6.3.7 Temperature and pH effects on enzyme activity

Enzyme activity for the *LMP_42667* gene product was assessed at pH values between 4 and 12. The buffer systems used comprised 50 mM sodium acetate buffer (pH 3.0 to 5.0), 50 mM sodium phosphate buffer (pH 6.0 to 8.0), and 50 mM glycine-NaOH buffer (pH 9.0-12.0). Residual activity was measured under the assay conditions. The optimum temperature for enzyme activity was assessed between 30 and 100°C and pH 7.0. To test its thermostability, the pure enzyme was pre-incubated in 50 mM sodium phosphate buffer with 1% CMC for 12 hours at 50, 60, and 70°C. Aliquots were assayed at 2-hour intervals and residual activity was measured under standard assay conditions. Relative activity was defined as the percentage of activity detected with respect to the enzyme activity under standard assay (Li et al., 2012). The influence of NaCl on enzyme activity was assessed with 0.5 – 6.0 M concentration of NaCl at temperatures between 35 and 100°C and pH between 3.0 and 12.0. The enzyme's relative activity was determined under standard assay conditions.

6.3.8 Effects of metal ions, inhibitors, surfactants, and organic solvents on enzyme activity

Effects of metal ions on enzyme activity were assayed at 5- and 10-mM concentrations. The metal ions were mixed with 3.02 mg/ml of purified enzyme in 50 mM sodium phosphate buffer pH 7.0 and supplemented with 1% CMC. The metals ions were: Cu^{2+}, Co^{2+}, Fe^{3+}, Mg^{2+}, Mn^{2+}, $Ni^{2+,}$ Ca^{2+}, and Zn^{2+}. Dithiothreitol (DTT), sodium dodecyl sulfoxide (SDS), ethylenediaminetetraacetic acid (EDTA), urea, ammonium chloride, guanidine hydrochloride, and phenylmethylsulfonyl fluoride (PMSF) were used to check the effect of protein inhibitors on enzyme activity. In addition, 1 and 5% (v/v) glycerol, Tween-20, Tween-80, and Triton-X 100 were added to establish the effect of surfactants on enzyme activity. Furthermore, the effects of both soluble and insoluble organic solvents on enzyme activity were tested

using 5 and 10% (v/v) methanol, ethanol, dimethyl sulfoxide (DMSO), acetone, isopropanol, ethyl acetate, chloroform, toluene, and diethyl ether. For all reactions, enzyme activity with 1% CMC and no chemical or metal ion was set at 100% activity.

6.3.9 Enzyme kinetic analysis

The rate of enzyme reaction was determined using 0.5 to 14.0 mg ml^{-1} CMC in 50 mM sodium phosphate buffer pH 7.0 as the substrate. The kinetic parameters, Michaelis constant (K_m), and maximum velocity (V_{max}) were determined by plotting the Lineweaver–Burk plot in GraphPad Prism 9.2.0 software (GraphPad Software, Inc.).

6.4 Results

6.4.1 Cloning and sequence analysis of the endoglucanase gene

Four endoglucanase protein sequences derived from artificial metagenome were cloned into the pBAD18 vector (Table 6.2).

Table 6.2: Characteristic features of endoglucanase genes obtained from artificial metagenome sample from Lake Magadi.

Gene locus Tag	Source organism	CAZyme Family	Length (Bp)	KO term*
LMP_22770	*So. lehensis*	GH5_4	1707	K01179
LMP_37690	*Sb. agaradhaerens*	GH5_4	1713	K01179
LMP_42667	*Sb. agaradhaerens*	GH5_4	1716	K01179
LMP_45580	*Sb. agaradhaerens*	GH5	540	K01179

*KO (KEGG orthology) term represents molecular representation of functional orthologs

However, only one (*LMP_42667*) gene could be expressed in *E. coli* cells. The gene originated from an isolate, which according to 16S rRNA gene sequencing, had 99% homology to *Salipaludibacillus agaradhaerens* str. DSM 8721. However, the phylogenetic relationship (Figure 6.1) depicted 99% similarity to an uncharacterized cellulase gene derived from an uncultured bacterium from a moderately saline L. Nakuru, Kenya (Ghauri et al., 2003) and 94% identity with cellulase from *Sb. agaradhaerens* (Davies et al., 1998).

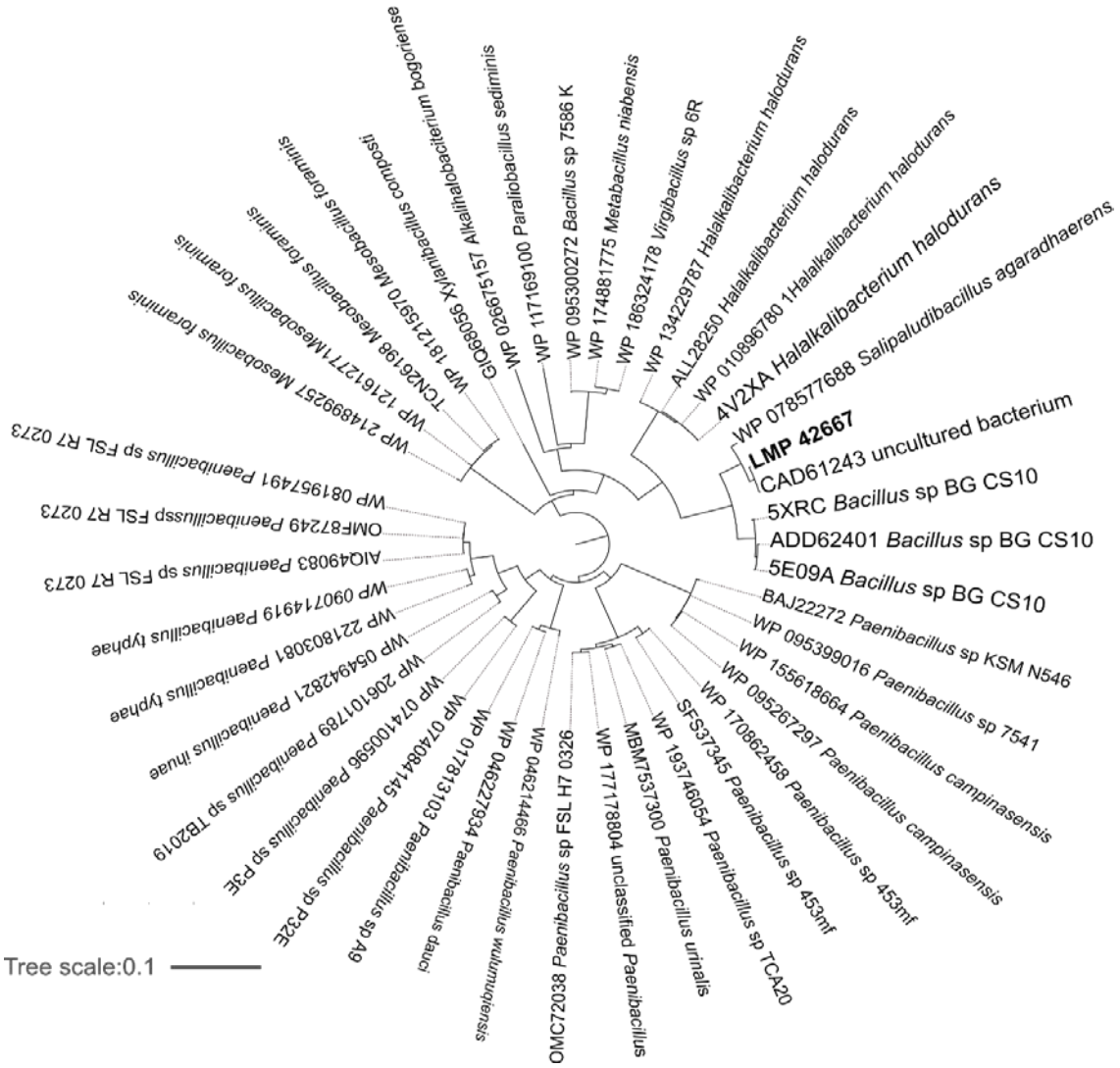


Figure 6.1: Phylogenetic relationship of *LMP_42667* gene product (in bold) with its endoglucanase homologs. From the nodes out, the accession number of the protein in the NCBI GenBank, then the organism from which it was derived.

In total, 43 samples belonging to the GH5 family with similarity ≥ 50% were deduced from the GenBank. Their accession numbers and source organisms are shown. Sequence alignment was performed using MAFFT (Katoh & Standley, 2013) whereas the tree was generated using the neighbor-joining algorithm in the NCBI database and viewed in the iTOL tree. The gene product *LMP_42667* gene product had a single open reading frame (ORF) of 1716 bases (Figure 6.2) and was classified into the GH_5 subfamily 4 of glycosyl hydrolases. The entire protein consists of 571 amino acids with a hypothetical molecular weight of 65.7 kDa; isoelectric point (*pI*) of 4.60; aliphatic index of 79.39; instability index of 34.78; half-life of 10 hours, and grand average of hydropathicity (GRAVY) index of -0.421. Cellular localization

demonstrated that the gene encoded a 29 amino acid signal peptide after the start codon and cleavage sites Ala33 and Glu34. The conserved domain consisted of the cellulase catalytic domain (Glu68- Gln348), and two carbohydrate-binding modules: CBM_2 (Glu378- Gln462) and CBM_46 (Asp467-Gly569).

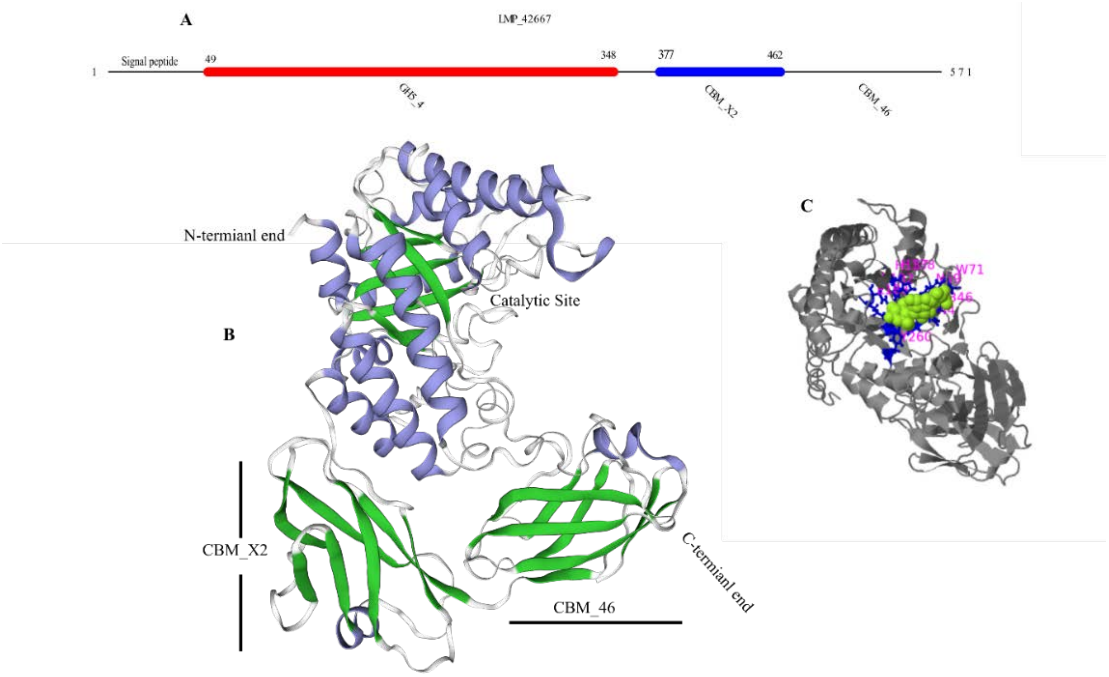


Figure 6.2: Modular structure of *LMP_42667* gene product. **A:** overall structure of the functional sections of the protein as annotated by dbCAN2 servers. **B:** Three-dimensional tertiary structure of the protein showing the functional sites as well as α (blue sheets) and β (green sheet) folding of the catalytic and the carbohydrate-binding modules. **C:** simulation of the substrate binding to the catalytic domain of the enzyme.

Modeling of *LMP_42667* resulted in four predicted models with only one having the highest C-score (confidence level) of -1.61 (2 is the highest). Based on the predicted protein structure (Figure 5.2), the cellulase depicted similarity to trimodular halophilic cellulase (*CelB*) isolated from *Bacillus* sp. BG-CS10. The protein catalytic domain structure common to the GH 5 family, the α/β TIM-barrel secondary structure was established for this enzyme (Figure 6.3). The structural similarity score between these two proteins (TM-score) was 0.924 (1 is the highest).

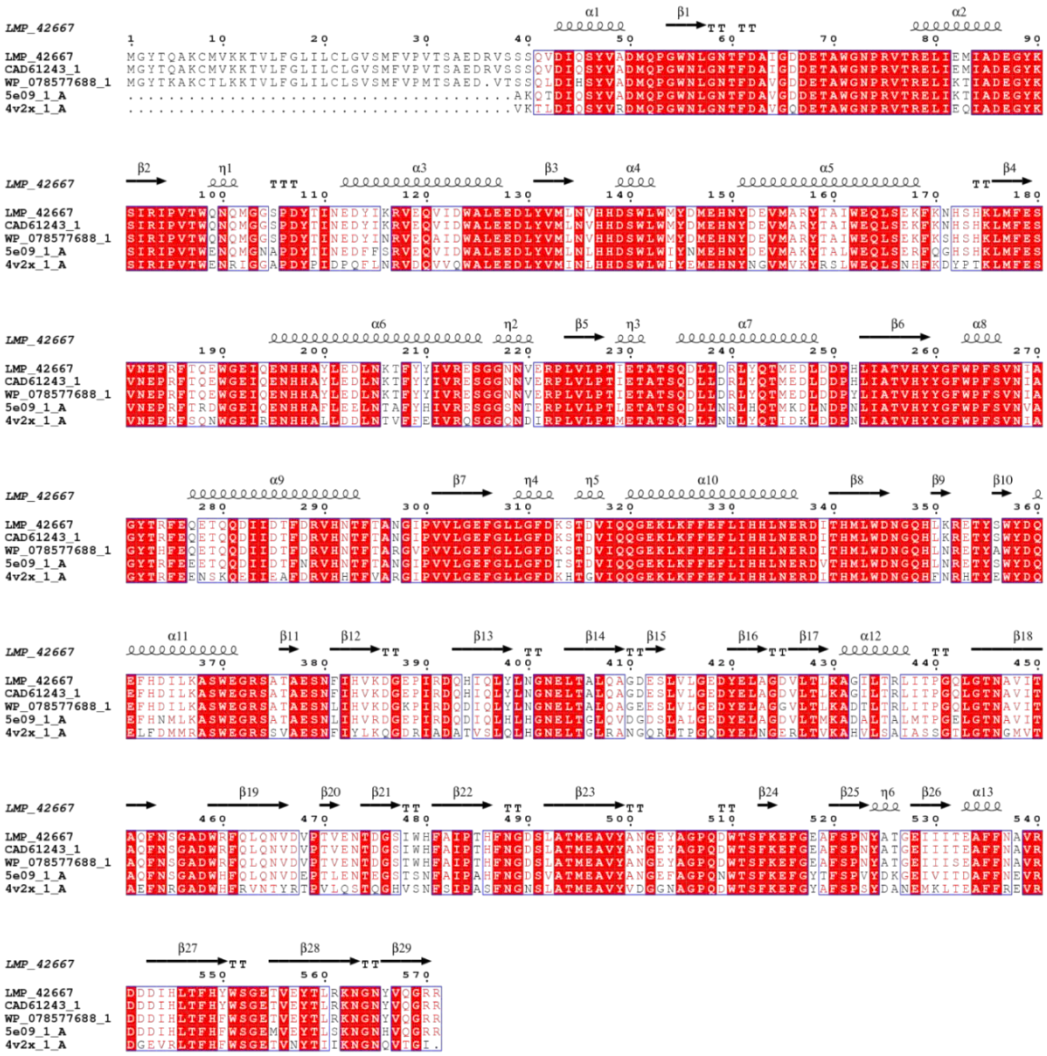


Figure 6.3: Multiple sequence alignment and secondary structure of the endoglucanase with its structural analogs obtained from protein databank and the GenBank. The secondary structures of the gene product are depicted by spring-like spirals for α helices and bold arrows for β sheets. The nearly conserved regions are enclosed in blue bars, the letters TT represent turns, while η represents the 3_{10}-helices. The amino acid sequence of the query protein was aligned against 5e09.1.A obtained from *Bacillus* sp. BG-CS10, 4V2X_1_A from *Hl. halodurans*, CAD61243 deduced from an uncultured bacterium, and WP_078577688 from *Sb. agaradhaerens*.

6.4.2 Purification and quantitative activity screening of *LMP_42667*

Purification of His6-tagged protein was done using Ni^{2+} affinity chromatography columns under native conditions (Figure 6.4). The purified enzyme showed a single band with a mass of 3.02 µg/ml and an estimated molecular of about 65 kDa (Figure 6.4a). Purification using this method led to a 29.04% recovery rate from the original

cell-free supernatant. Activity test screen of the purified enzyme using 1% CMC showed significant clearances after 48 hours (Figure 6.4b).

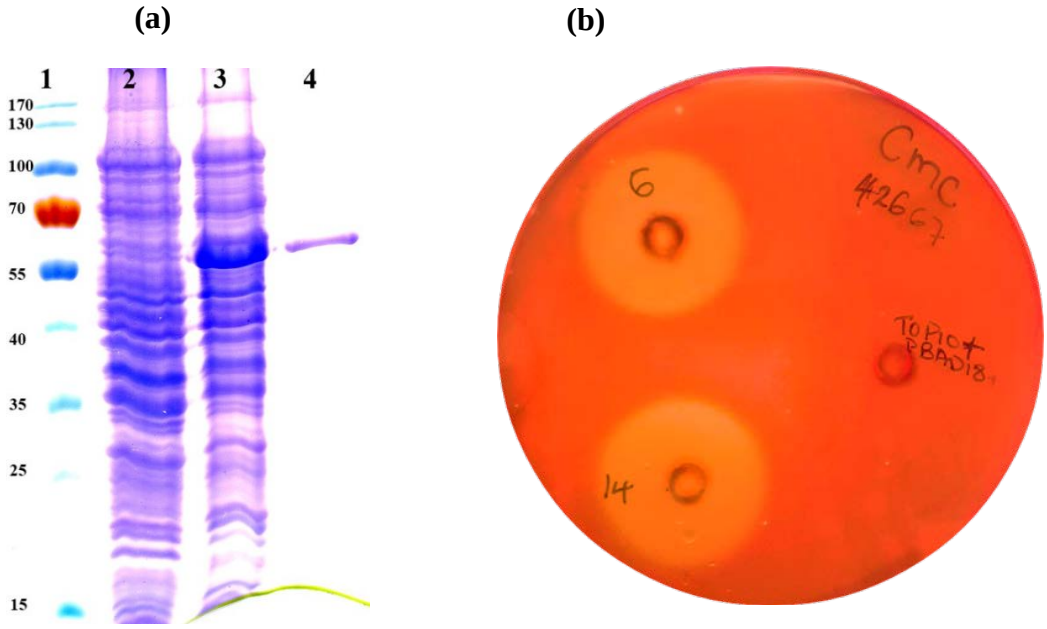


Figure 6.4: Purification and activity of endoglucanase gene *LMP_42667*. **(a)** SDS-PAGE analysis for the purified recombinant LMP_42667 protein. 1= PageRuler™ Prestained Protein Ladder in kDa, 2= induced *E. coli* cells with empty pBAD18 vector, 3: Cell-free supernatant with expressed protein, 4= purified endoglucanase. **(b)** CMCase activity of purified endoglucanase from two positive clones (labeled 6 and 14) of the *LMP_42667* gene. Negative control induced *E. coli* cells with empty pBAD18 vector (far right). The assay was performed by flooding inoculated CMC-agar plates with 1% Congo red stain and de-staining with 1.5 M NaCl.

6.4.3 Substrate specificity and enzyme kinetics

The endoglucanase showed positive activity on all the substrates possessing the β-linkages, indicating that it is endo-β-acting (Table 6.3). The highest activity was recorded for xylan with a nearly two-fold activity increase relative to carboxymethyl cellulose. The least relative activity was recorded for microcrystalline cellulose and lichenan (2.6 U/mg each), whereas no activity was detected against chitin, starch, and mannan. The kinetic reactions of the protein showed V_{max} and K_m values of 18.7 U/mL and 3.103 µmol/min/mg, respectively.

Table 6.3. Substrate specificity spectrum of the endoglucanase enzyme isolated from bacteria from Lake Magadi

Substrate and % (v/v)	Glycosidic linkage	Specificity activity (U/mg) #	Solubility in water
CMC-Na salt (1%)	β-1,4 glucan	4.4 ± 0.06	Soluble
HEC (1%)	β-1,4 glucan	3.03 ± 0.3	Soluble
Cellulose microcrystalline (1%)	β-1,4 glucan	2.6 ± 0.05	Insoluble
Xylan (1%)	β-1,4 xylose	9.8 ± 0.1	Soluble
Lichenan (1%)	β-1,3 and β-1,4 glucan	2.6 ± 0.1	Insoluble
β-glucan (1%)	β-1,3 glucan	6.4 ± 0.07	Soluble
Chitin (0.5%)	β-1,4 N-acetylglucosamine	ND*	Insoluble
Mannan (0.5%)	α-1,6 and α-1,3 glucan	ND*	Soluble
Starch (1%)	α-1,4 and α-1,6 glucan	ND*	Insoluble

*ND corresponds to not detected; #The values indicate mean specific activities and their standard errors of mean.

6.4.4 Effects of pH, temperature, and salinity on enzyme activity

The physiological characteristics of the endoglucanase are shown in Figure 6.5. The optimal pH for the purified enzyme was 7.0, with more than 50% activity observed between pH 5.0 - 8.0 (Fig. 6.5a). The temperature range varied, but the optimum temperature for the purified enzyme was 50^{o}C, after which activity declined (Figure 6.5b). The addition of NaCl, however, led to the activation of the enzyme along a concentration gradient, with optimum activity reported at a 5M concentration (Figure 6.5c). The pH and temperature activity ranges remained unchanged after adding 5M NaCl. However, there was a marked increase in activity with up to 2-fold and 1.5-fold at pH 7.0 and 50^{o}C, respectively (Figure 6.5a and 6.5b). The thermostability test indicated that the enzyme was relatively stable at 50^{o}C retaining over 50% activity for 12 hours. At 60 and 70^{o}C, the stability declined, but the enzyme still retained 50% of its activity at 8 and 6 hours, respectively (Figure 6.5d).

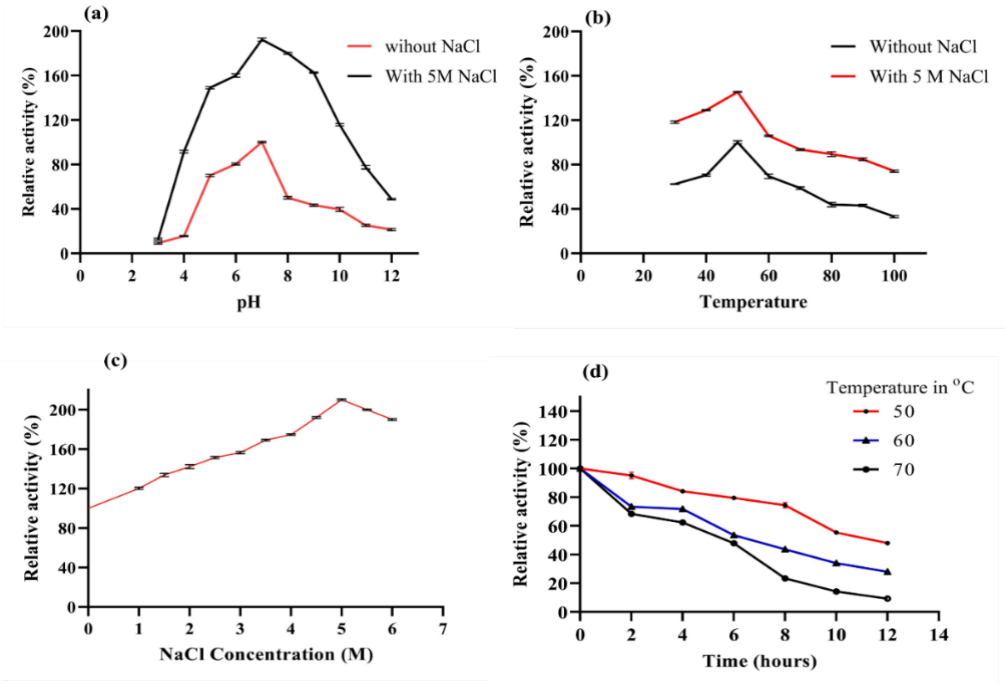


Figure 6.5: Effects of pH, temperature, salinity, and thermostability of the endoglucanase. (**a)** Effect of pH with 5 M NaCl and without NaCl on the endoglucanase activity. Buffers used were in 50 mM sodium acetate (3 – 5), sodium phosphate (pH 6 – 7), and glycine-NaOH (pH 8 – 12). (**b)** Effect of temperature on the activity of the cellulase. Assays for both with and without 5 M NaCl were measured. (**c)** Relative activity assessments on increasing concentrations of NaCl. Enzyme activity without NaCl was set at 100%. (**d)** Influence of three temperature conditions on the stability of the enzyme over time. Error bars represent standard error.

6.4.5 Effect of metal ions, inhibitors, surfactants, and organic solvents on enzyme activity

Enzyme activity in the presence of various chemical reagents was tested (Figure 6.6). Mg^{2+} and Mn^{2+} slightly increased relative activity to 123 and 112%, respectively, while Fe^{3+} (5%) and Cu^{2+} (47% relative activity) inactivated it at the concentrations tested. On the other hand, Co^{2+} (86%), Ni^{+} (70%), and Zn^{2+} (71%) slightly reduced the activity of the enzyme (Figure 6.6a). An increase in Mg^{2+} ion concentration from 5 to 10 mM led to an increase in enzyme activity, whereas the presence of Fe^{3+} ions at 5 and 10 mM rendered the enzyme to a near-inactive state (5%). Among the protein inhibitors, SDS produced the highest deleterious effect on enzyme activity with a

39% loss on the addition of 5 mM concentrations. Dithiothreitol (DTT) at 10 mM caused a 1.8-fold increase in enzyme activity, while 5 mM ethylenediaminetetraacetic acid (EDTA) (119%), urea (109%), guanidine HCl (114%), and phenylmethylsulfonyl fluoride (PMSF) (147%) had slight influence on enzyme activity. At 10 mM concentration all tested protein inhibitors except DTT caused a decrease in activity (Figure 6.6b). The presence of surfactants showed that Tween 20 decreased enzyme activity to 88% and 23% relative activity at 5 and 10 mM concentrations, respectively. Tween 80 and Triton X-100 led to an over 50% increase in relative activity. At 10% (v/v) concentration, all the surfactants caused a reduction in the activity of the enzyme (Figure 6.6c). All tested organic solvents except ethyl acetate led to an increase in the enzyme's relative activity. Addition of 10% (v/v) ethyl acetate led to an 85% loss in enzyme activity (Figure 6.6d). The presence of acetone (water-miscible), toluene, and diethyl ether (water-immiscible) led to a 1.5-fold increase in enzyme activity. Overall, there was no noticeable difference in the effects of both miscible and immiscible organic solvents.

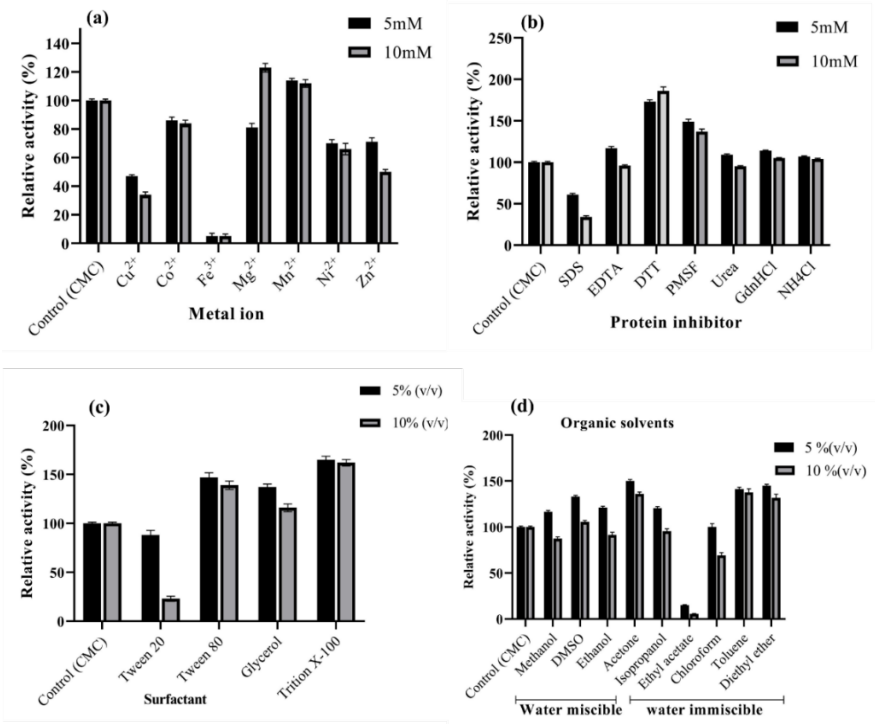


Figure 6.6: Effects of chemical reagents on the activity of LMP_42667 enzyme. **(a)** Effects of metal ions on the activity of the enzyme. **(b)** Influence of selected protein inhibitors on the enzyme action. **(c)** Surfactants' influence on the activity of the enzyme. **(d)** Enzyme activity in the presence of various organic solvents. Error bars indicate standard error.

6.5 Discussion

6.5.1 Bioinformatic analysis of the endoglucanase

Halophilic and halotolerant cellulases deduced from soda lake bacteria are valuable resources in the development of a bio-based economy due to their ability to tolerate extreme conditions in industrial setups (Li et al., 2020). This study cloned four artificial metagenome-derived endoglucanase encoding genes, but successfully expressed only one gene (*LMP_42667*). The gene was highly linked to *Salipaludibacillus agaradhaerens*, a halophilic bacterium that was previously isolated from moderate and hypersaline environments like soils and saline aquatic habitats (Sultanpuram & Mothe, 2016). Interestingly, the endoglucanase shared the highest sequence similarity (99%) with an uncharacterized cellulase gene from an uncultured bacterium isolated by Ghauri et al. (2003) from a moderately saline lake Nakuru, Kenya. Functionally, the endoglucanase (EC 3.2.1.4) was categorized into the GH5 class of CAZymes, a feature that was supported by a (β/α)8-barrel fold model associated with all members of the polyspecific GH5 (Wierzbicka-Woś et al., 2019). The hypothetical isoelectric point (pI) of 4.60 and a high proportion of acidic (15.6%) and negatively charged (89) amino acid residues indicate that it could be an acidic protein (Lu & Daniel, 2021). A low GRAVY index of -0.421 suggests a high likelihood of interaction between the protein and water (Sanjaya et al., 2021). In addition, the enzyme's catalytic domain was appended by two carbohydrate-binding modules (CBM_X2 and CBM46). In *Streptomyces*, *Bacillus licheniformis*, and *Hm. halodurans*, CM46 co-exists with the CBM_X2 module, which is an immunoglobulin (Ig)-like domain aiding in effective binding to both the cell wall of bacterial and cellulosic material (Venditto et al., 2015; Liberato et al., 2016; Sidar et al., 2020).

6.5.2 Physiological characterization of the enzyme

The pH and temperature optimum for the enzyme activity were recorded at 7.0 and 50^{o}C, respectively. Additionally, it functions under a narrow pH range, ranging from 5 to 7.0; whereas below pH 5.0, the enzyme becomes acid-labile. This implies that the endoglucanase hydrolyses best at moderate pH and temperature conditions. Between 50 and 60^{o}C the enzyme activity was retained at over 50% for up to 12 hours, while further increase in temperature drastically lowers activity. Therefore, *LMP_42667* gene product depicts a longer catalysis period when subjected to

moderate temperatures. This can be explained by the high aliphatic index (79.39) recorded for this enzyme due to increased aliphatic amino acids such as alanine, proline, valine, isoleucine, and leucine, known to aid the thermostability aspect of globular proteins (Yakimov et al., 2016). These results are similar to the halotolerant endocellulases previously reported (Aygan & Arikcan, 2008; Gao et al., 2010; Ogonda et al., 2021). However, a study by Ariaeenejad et al. (2020) reported superior alkali-thermostable endoglucanase capable of sustained optimum activity at 85^{o}C and pH 8.5 for 156 hours.

A notable property of *LMP_42667* was its optimal activation by 5 M NaCl concentration, indicating that it is an extreme haloenzyme (Didari et al., 2020). Compared with enzyme without NaCl, 5 M NaCl caused a significant increase in activity across various pH and temperature conditions. When 5M NaCl was added, enzyme retained over 50% activity at a pH range of 4 to 11, and temperatures between 40 – 100^{o}C. Several halophilic endo β-1,4-glucanases are activated by salt concentrations in the range of 1.5 to 4 M concentration (Hirasawa et al., 2006; Şafak et al., 2020; Li et al., 2021; Sar et al., 2021). However, as compared to non-halophiles, haloenzymes have more acidic amino acid residues (Garg et al., 2016; Lu et al., 2019; Sanjaya et al., 2021). Indeed, *in silico* analysis showed the presence of more acidic amino acid residues on *LMP_42667* than their basic counterparts. It is predicted that salts play a significant role in shielding enzymes against denaturants such as high temperatures, pH, and other chaotropic agents (Sinha & Khare, 2014). Pre-treatment of lignocellulosic materials usually takes place under alkali or acidic conditions. Resultant neutralization steps lead to the formation and accumulation of various salts whose removal consumes considerable water and energy (Cortes-Tolalpa et al., 2018; Kucharska et al., 2018). Therefore, mining for cellulolytic enzymes that can carry on with fermentation with minimal salt inhibition is an attractive venture (Gunny et al., 2014).

6.5.3 Influence of chemical reagents on the endoglucanase activity

The influence of metal ions on enzyme activity indicated that Mn^{2+} and Mg^{2+} ions slightly increased enzyme activity at 5 mM and 10 mM, respectively. Mg^{2+} is known as an essential co-factor in numerous enzyme activities. In most instances, Mn^{2+} and Mg^{2+} ions substitute each other in stimulating enzyme activity (Khrustalev et al.,

2016). This result agrees with that of halotolerant endo-β-1,4-glucanase from *Microbulbifer* sp. ALW1 (Li et al., 2021). On the contrary, the presence of Mn^{2+} ions has been found to inhibit cellulase activity in several instances (Zhu et al., 2011; Dong et al., 2016; Li et al., 2020). Furthermore, Co^{2+}, $Zn^{2+,}$ and Ni^{2+} ions slightly reduced activity, whereas Fe^{3+} and increasing Cu^{2+} ions significantly decreased the activity of the enzyme. Similar trends have been reported for previously studied endoglucanases (Jung et al., 2010; Wei et al., 2015; Huy et al., 2016). However, a study by Hirasawa et al. (2006) showed that the presence of Ni^{2+}, Mn^{2+}, Co^{2+}, Zn^{2+}, and Cu^{2+} did not cause any noticeable effect on the activity of endoglucanase Egl-AG. In general, heavy metals decrease enzyme activity by binding to the thiol group in the active site, hence hindering full activity (Murashima et al., 2002). In this study, the activity of the enzyme in the presence of heavy metals is dependent on the effect of the metal on the enzyme. Whereas Mn^{2+} slightly increased enzyme activity, Zn^{2+} (71%)$,$ and Co^{2+} (86%) slightly decreased activity, while Fe^{3+} (5%) and Cu^{2+} (47%) drastically decreased activity.

Amongst the protein inhibitors, an increase in the concentration of SDS showed an inhibitory effect on the enzyme. SDS is known to cause a conformational structure change at the active site by increasing the overall surface negative charge of the endoglucanase and the substrate, leading to a sustained increase in electrostatic repulsion between the substrate and the enzyme (Hou et al., 2020; Zhao et al., 2021). This could have been the case for *LMP_42667*. Instead, the presence of PMSF (serine modifier) led to a stimulatory effect on the enzyme, demonstrating that serine residues are not necessary for the catalytic site (Li & Yu, 2012). This contrasts with a previous report demonstrating that 5 mM PMSF led to a significant loss in the activity of haloalkaliphilic endoglucanase (Aygan & Arikcan, 2008; Li & Yu, 2012). Furthermore, 10 mM dithiothreitol (DTT), a di-sulfide reducer, caused a stimulatory effect on the activity of the enzyme (186%). During enzyme-substrate interaction, DTT ensures sustained conformational folding structure of the enzyme by hindering the formation of disulfide bonds, hence increasing activity (Zhu et al., 2011). Therefore, the results could indicate that increasing the concentration of DTT could have a consequent increase in activity. Increasing DTT from 1 – 10 mM led to a 30% increase in the activity of thermostable, halo-tolerant Cel5A and MaCel5A

endoglucanases (Liang et al., 2011; Li et al., 2021). Notably, EDTA (a metal chelator) increased the relative activity to 119%. The inability of EDTA to inhibit the endoglucanase activity indicates that the *LMP_42667* gene product is not a metalloenzyme (Li & Yu, 2012). On the other hand, the presence of denaturants such as urea and guanidine HCl did not produce any notable effect on the activity of the enzyme. Elsewhere, 5 mM urea was reported to cause a deleterious effect on halophilic endoglucanase (Aygan & Arikcan, 2008), whereas 10 mol/L led to complete inactivity of *MaCel5A* endoglucanases (Li et al., 2021).

The effect of surfactants tested demonstrated that Tween 80, Triton X-100 and glycerol had a stimulatory effect on the enzyme. Tween 20 on the other hand showed an inhibitory effect at 5 mM and a further decline in activity at 10 mM concentration. Non-ionic surfactants are known to enhance enzyme activity by stabilizing the cellulase, increasing enzyme-substrate desorption, and reducing non-productive substrate binding to the enzyme (Seo et al., 2011; Hsieh et al., 2015). Many surfactants apply to this principle, but the negative effect of Tween 20 reported here could not be explained.

Effects of both water-miscible and water-immiscible organic solvents on the enzyme were tested. In the present study, a 5 mM concentration of all the tested organic solvents but ethyl ether produced stimulatory effects on the endoglucanase activity. However, increasing the concentration to 10 mM in all cases led to a slight decrease in activity (Figure 6.6). This suggests that the enzyme would only function at lower concentrations of organic solvents. In terms of stability in organic solvents, *LMP_42667* performed better than a halotolerant cellulase from *Bacillus* sp. L1 (Li & Yu, 2012) and CKT3eng from *Haloarcula* sp. CKT3 (Şafak et al., 2020). Generally, the occurrence of halophilic and organic solvent-tolerant cellulases is scarce despite their usage in the bioremediation of wastewaters containing organic solvents from industry and carbohydrate-polluted salt marshes (Shafiei et al., 2011; Li & Yu, 2012). These results suggest that several metal ions, as well as chemical agents, could be strategically added to enhance the activity of the studied enzyme.

6.5.4 Substrate specificity and enzyme kinetics

The endoglucanase demonstrated activity against all substrates with β-glucan glycosidic linkages, confirming its endo-acting activity. The enzyme depicted the highest activity towards xylan (9.8±0.1 U/mg) followed by β-glucan (6.4±0.07 U/mg), while activity towards HEC (3.03±0.3 U/mg), lichenan (2.6 ±0.1 U/mg), and microcrystalline cellulose (2.6 ± 0.05 U/mg) was moderate. This indicates that the enzyme has a wide spectrum of specificity when primed with substrates containing β-1,3/4 and β-1,3/6-linkages (Table 6.3). On the other hand, no activity was recorded for chitin, mannan, and starch, all possessing α-glycosidic linkages. The xylan extracted from birchwood is known to be an acetylated hardwood. Studies have shown that the activity of some endo-acting enzymes could be enhanced when certain forms of a polysaccharide (such as xylan) are branched or have a substituent (Nacke et al., 2012). Enzyme kinetics were performed using CMC as the substrate. Michaelis Menten parameters revealed V_{max} and K_m values for the endo-β-1,4-glucanase of 18.7 mg/ mL and 3.103 μmol/min/mg, respectively. Generally, enzyme activity increased at higher concentrations of the substrate as indicated by high V_{max}. Usually, low K_m values are an indication of high substrate affinity for the enzyme (Khalid et al., 2019).

6.6 Conclusion

This study attempted to simultaneously clone four endoglucanase genes obtained from synthetic metagenomic sequences from the hypersaline lake Magadi. However, only one gene, *LMP_42667* was successfully cloned, expressed, and characterized. The gene product revealed an extremely halophilic β-1,4-endoglucanase, which was deduced from bacteria of the genus *Salipaludibacillus*. Bioinformatic analysis indicates that the *LMP_42667* gene product belongs to glycosyl hydrolase 5 sub-family 4 of hydrolytic enzymes. Substrate utilization assays revealed that the pure enzyme could catalyze the depolymerization of several polysaccharides, particularly those possessing β-1,3/4 and β-1,3/6-linkages. Enzyme assays showed an extreme halophilic enzyme, activated by a high concentration of salt (5M NaCl), and active across mild acidic and basic buffers. Its activity was increased in the presence of Mg^{2+} and Mn^{2+} ions, inhibitors (DTT, EDTA, and PMSF), and non-ionic surfactants (tween 80, glycerol, and TritonX-100). Its function was maintained above 50% for 12 hours

at temperatures between 50 - 60°C under the assay conditions. These properties, therefore, make the *LMP_42667* gene product a promising candidate for industrial and biotechnological studies.

CHAPTER SEVEN
7.0 CONCLUSION AND RECOMMENDATIONS

7.1 Conclusion

This study investigated the diversity, function, and genomic characterization of bacterial isolates derived from soda Lake Magadi. Isolation resulted in bacterial isolates affiliated with *Evansella*, *Alkalihalobacterium*, *Salinicoccus*, *Alkalibacterium*, *Saliplaludibacillus*, and *Halomonas*. The activity of individual isolates on selected polysaccharides indicated that isolates related to *Alkalihalobacterium* and *Salipaludibacillus* depicted extracellular enzyme activity against one of the substrates (starch, pectin, xanthan, cellulose, CMC, and xylan). These findings indicate that Lake Magadi is a rich reservoir for polysaccharide-degrading bacteria playing a critical role in biogeochemical nutrient cycling and decomposition of organic matter. Furthermore, whole-genome sequencing revealed that LMS6, LMS18, LMS25, and LMS39 were new species of *Shouchella* sp., *Evansella* sp., *Salipaludibacillus* sp., and *Alkalihalobcterium* sp., respectively. This provides an opportunity for phenotypic characterization and taxonomic placement of these isolates. It may also imply that many more strains of bacteria have not been fully classified in the Lake Magadi ecosystem. Genome analysis revealed the presence of carbohydrate-active enzymes (CAZymes) encoding for amylases, xanthanases, cellulases, pectinases, and xylanases. All these genes were found in all the genomes but predominantly domiciled in isolates of *Alkalihalobacterium* LMS39 and *Salipaludibacillus* LMS25 genomes.

Molecular typing using the V4 hypervariable region of the 16S rRNA gene revealed that the abundant bacterial phyla in the lake were Proteobacteria, Cyanobacteria, Bacteroidetes, Actinobacteria, Firmicutes, Verrumicrobia, Deinococcus-Thermus, Spirochaetes, and Chloroflexi. The Archaeal diversity was represented by phyla Euryachaeota, Crenarchaeota, Euryarchaeota, and Thaumarchaeota. This indicates that Lake Magadi is a rich ecosystem housing a rich and diverse group of microbial communities that symbiotically co-exist in an intricate relationship in driving biogeochemical cycling in the soda lake. Particularly, the presence of cyanobacteria across sampling seasons indicates an active primary productivity in Lake Magadi, thus survival of of microbial populations is assured (Sorokin and Kuenen, 2005).

Archaea on the other hand perform ecosystem functions such as carbon cycling and methanogenesis (Vavourakis et al., 2016). Water chemistry revealed that temperature, pH, P, K, NO_3^-, and TDS significantly influenced the microbial community distribution in Lake Magadi. It has been established that water chemistry affects the distribution of microbial communities by determining the environmental conditions by which the microbes must adapt and the availability of nutritional ingredients.

Direct sequencing of the environmental metagenome gives insight analysis of gene functions and facilitates the screening for potential CAZymes. The sequencing of synthetic metagenome generated results similar to the 16S rRNA and whole-genome sequencing. Ten metagenome-assembled genomes (MAGs) of medium and high quality affiliated with *Salipaludibacillus*, *Evansella*, *Alkalihalophilus*, *Halalkalibacterium*, *Salinicoccus*, *Alkalibacterium*, and *Halomonas* were generated in this study. The results indicated that MAGs have vast and largely understudied genetic resources that could depolymerize high molecular weight polymers such as cellulose, starch, xylan, chitin, pectin, and xanthan. Moreover, most MAGs could belong to novel species, which helps elucidate the so-called microbial dark matter. The generated MAGs demonstrated high CAZyme enrichment amongst the affiliates of *Salipaludibacillus* and *Alkhalihalobaterium*.

Moreover, endoglucanase genes derived from artificial metagenome sequencing were subjected to cloning and expression on *E. coli* hosts. Only one of the four cloned genes could be successfully cloned and expressed. The protein of gene LMP_42667, deduced from *Salipaludibacillus* sp. belongs to glycosyl hydrolase 5 sub-family 4 of hydrolytic enzymes. The enzyme is an extremely halophilic enzyme, activated by a high salt concentration, and active across mild acidic and basic buffers. Furthermore, the endoglucanase demonstrated activity against xylan, β-glucan, microcrystalline cellulose, CMC hydroxyethylcellulose (HEC), and lichenan. Its activity was increased in the presence of several metal ions, inhibitors, and non-ionic surfactants while having a relatively long half-life in moderate temperature conditions. Its activity at above-average temperatures, pH, surfactants, organic solvents, high salinity, and wide substrate specificity make it an ideal product for use in the industrial depolymerization of lignocellulosic biomass and bioremediation.

7.2 Recommendations

The study revealed the presence of a considerable diversity of polysaccharides utilizing bacteria from Lake Magadi. The selected isolates contained a high number of CAZyme encoding genes that would interest academic, biotechnological, and industrial applications.

1. Further studies utilizing a combination of water and sediment metagenomic and metatranscriptomic analysis could be conducted to survey more novel bacterial isolates within the lake.
2. Studies involving biochemical parameters of growth for isolates LMS6, LMS18, LMS25 and LMS39 should be performed. This will pave the way for naming the species identity of the isolates.
3. The artificial metagenome also demonstrated that a large portion of the proteins (about 54%) are without assigned functions. This presents an opportunity for the establishment of their full functionality in future studies.
4. Studies on the functional profiles of samples during and after cyanobacterial blooms, including vertical profile stratification of Lake Magadi could elaborate more microbial communities in different depths.

REFERENCES

Aanderud, Z. T., Vert, J. C., Lennon, J. T., Magnusson, T. W., Breakwell, D. P., & Harker, A. R. (2016). Bacterial dormancy is more prevalent in freshwater than in hypersaline lakes. *Frontiers in Microbiology*, *7*(6), 1–13. https://doi.org/10.3389/fmicb.2016.00853

Abbott, D. W., & Boraston, A. B. (2012). Quantitative approaches to the analysis of carbohydrate-binding module function. In *Methods in Enzymology* (1st ed., Vol. 510). Elsevier Inc. https://doi.org/10.1016/B978-0-12-415931-0.00011-2

Abdallah, M. Ben, Karray, F., Kallel, N., Armougom, F., & Mhiri, N. (2018). Abundance and diversity of prokaryotes in ephemeral hypersaline lake Chott El Jerid using Illumina Miseq sequencing, DGGE, and qPCR assays. *Extremophiles*. https://doi.org/10.1007/s00792-018-1040-9

Abdallah, M. Ben, Karray, F., Mhiri, N., Mei, N., Quéméneur, M., Cayol, J. L., Erauso, G., Tholozan, J. L., Alazard, D., & Sayadi, S. (2016). Prokaryotic diversity in a Tunisian hypersaline lake, Chott El Jerid. *Extremophiles*, *20*(2), 125–138. https://doi.org/10.1007/s00792-015-0805-7

Aguilar, A., Ingemansson, T., & Magnien, E. (1998). Extremophile microorganisms as cell factories: Support from the European Union. *Extremophiles*, *2*(3), 367–373. https://doi.org/10.1007/s007920050080

Ahn, A.-C., Jongepier, E., Schuurmans, J. M., Rijpstra, W. I. C., Sinninghe Damsté, J. S., Galinski, E. A., Roman, P., Sorokin, D., & Muyzer, G. (2021). Molecular and Physiological Adaptations to Low Temperature in *Thioalkalivibrio* Strains Isolated from Soda Lakes with Different Temperature Regimes. *American Society of Microbiology*, *6*(2), e01202-20. https://doi.org/10.1128/msystems.01202-20

Aktuganov, G. E., Galimzianova, N. F., Gilvanova, E. A., Kuzmina, L. Y., Boyko, T. F., Safina, V. R., & Melentiev, A. I. (2018). Characterization of Chitinase Produced by the Alkaliphilic *Bacillus mannanilyticus* IB-OR17 B1 Strain. *Applied Biochemistry and Microbiology*, *54*(5), 505–511. https://doi.org/10.1134/S0003683818050022

Alain, K., & Querellou, J. (2009). Cultivating the uncultured: Limits, advances, and future challenges. *Extremophiles*, *13*(4), 583–594.

Alaña, A., Llama, M. J., & Serra, J. L. (1991). Purification and some properties of the pectin lyase from *Penicillium italicum*. *FEBS Letters*, *280*(2), 335–340. https://doi.org/10.1016/0014-5793(91)80325-W

Alikhan, N. F., Petty, N. K., Ben Zakour, N. L., & Beatson, S. A. (2011). BLAST Ring Image Generator (BRIG): Simple prokaryote genome comparisons. *BMC Genomics*, *12*(1), 402. https://doi.org/10.1186/1471-2164-12-402

Allers, T. (2010). Overexpression and purification of halophilic proteins in Haloferax volcanii. *Bioengineered Bugs*, *1*(4), 290–292. https://doi.org/10.4161/bbug.1.4.11794

Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., von Heijne, G., & Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology*, *37*(4), 420–423. https://doi.org/10.1038/s41587-019-0036-z

Almeida, T. I. R., do Carmo Calijuri, M., Falco, P. B., Casali, S. P., Kupriyanova, E., Filho, A. C. P., Sigolo, J. B., & Bertolo, R. A. (2011). Biogeochemical processes and the diversity of Nhecolândia lakes, Brazil. *Anais Da Academia Brasileira de Ciencias*, *83*(2), 391–407. https://doi.org/10.1590/S0001-37652011000200004

Amoozegar, M. A., Siroosi, M., Atashgahi, S., Smidt, H., & Ventosa, A. (2017). Systematics of haloarchaea and biotechnological potential of their hydrolytic enzymes. *Microbiology (United Kingdom)*, *163*(5), 623–645. https://doi.org/10.1099/mic.0.000463

Andreote, A. P. D., Dini-Andreote, F., Rigonato, J., Machineski, G. S., Souza, B. C. E., Barbiero, L., Rezende-Filho, A. T., & Fiore, M. F. (2018). Contrasting the genetic patterns of microbial communities in soda lakes with and without cyanobacterial bloom. *Frontiers in Microbiology*, *9*(244), 1–13. https://doi.org/10.3389/fmicb.2018.00244

Annamalai, N., Rajeswari, M. V., Elayaraja, S., & Balasubramanian, T. (2013). Thermostable, haloalkaline cellulase from *Bacillus halodurans* CAS 1 by conversion of lignocellulosic wastes. *Carbohydrate Polymers*, *94*(1), 409–415. https://doi.org/10.1016/j.carbpol.2013.01.066

Antony, C. P., Kumaresan, D., Hunger, S., Drake, H. L., Murrell, J. C., & Shouche, Y. S. (2013). Microbiology of Lonar Lake and other soda lakes. *ISME Journal*, *7*(3), 468–476. https://doi.org/10.1038/ismej.2012.137

Antranikian, G., Vorgias, C. E., & Bertoldo, C. (2005). Extreme environments as a resource for microorganisms and novel biocatalysts. *Advances in Biochemical Engineering/Biotechnology*, *96*(8), 219–262. https://doi.org/10.1007/b135786

Aono, H., Sugimoto, E., Sadaoka, Y., Imanaka, N., & Adachi, G. (1990). Ionic conductivity of solid electrolytes based on lithium titanium phosphate. *Journal of the Electrochemical Society*, *137*(4), 1023–1027.

Arahal, D. R. (2014). Whole-Genome Analyses: Average Nucleotide Identity. *Methods in Microbiology*, *41*, 103–122. https://doi.org/10.1016/BS.MIM.2014.07.002

Arayes, M. A., Mabrouk, M. E. M., Sabry, S. A., & Abdella, B. (2021). Diversity and characterization of culturable haloalkaliphilic bacteria from two distinct hypersaline lakes in northern Egypt. *Biologia*, *76*(2), 751–761. https://doi.org/10.2478/s11756-020-00609-5

Ariaeenejad, S., Sheykh Abdollahzadeh Mamaghani, A., Maleki, M., Kavousi, K., Foroozandeh Shahraki, M., & Hosseini Salekdeh, G. (2020). A novel high-performance in-silico screened metagenome-derived alkali-thermostable endo-β-1,4-glucanase for lignocellulosic biomass hydrolysis in the harsh conditions.

BMC Biotechnology, *20*(1), 1–13. https://doi.org/10.1186/s12896-020-00647-6

Asha, R., Niyonzima, F. N., & Sunil, S. M. (2013). Purification and properties of pullulanase from Bacillus halodurans. *International Research Journal of Biological Sciences*, *2*(3), 35–43.

Ashabil, A., Sevtap, S., Sedat, K., & Huseyin, T. (2014). Production and characterization of alkaliphilic alpha-amylase from *Bacillus subtilis* A10 isolated from soils of Kahramanmaras, Turkey. *African Journal of Microbiology Research*, *8*(21), 2168–2173.

Asif, T., Javed, U., Zafar, S. B., Ansari, A., Ul Qader, S. A., & Aman, A. (2019). Bioconversion of Colloidal Chitin Using Novel Chitinase from *Glutamicibacter uratoxydans* Exhibiting Anti-fungal Potential by Hydrolyzing Chitin Within Fungal Cell Wall. *Waste and Biomass Valorization*, *11*(8), 4129–4143. https://doi.org/10.1007/s12649-019-00746-2

Aulitto, M., Strazzulli, A., Sansone, F., Cozzolino, F., Monti, M., Moracci, M., Fiorentino, G., Limauro, D., Bartolucci, S., & Contursi, P. (2021). Prebiotic properties of *Bacillus coagulans* MA-13: production of galactoside hydrolyzing enzymes and characterization of the transglycosylation properties of a GH42 β-galactosidase. *Microbial Cell Factories*, *20*(1), 1–17. https://doi.org/10.1186/s12934-021-01553-y

Aygan, A., & Arikcan, B. (2008). A new halo-alkaliphilic, thermostable endoglucanase from moderately halophilic *Bacillus* sp. C14 isolated from Van Soda Lake. *International Journal of Agriculture and Biology*, *10*(4), 369–374.

Azeri, C., Tamer, U., & Oskay, M. (2010). Thermoactive cellulase-free xylanase production from alkaliphilic *Bacillus* strains using various agro-residues and their potential in biobleaching of kraft pulp. *African Journal of Biotechnology*, *9*(1), 63–72.

Babavalian, H., Amoozegar, M. A., Pourbabaee, A. A., Moghaddam, M. M., & Shakeri, F. (2013). Isolation and identification of moderately halophilic bacteria producing hydrolytic enzymes from the largest hypersaline playa in Iran. *Microbiology*, *82*(4), 466–474. https://doi.org/10.1134/S0026261713040176

Babavalian, H., Amoozegar, M. A., Zahraei, S., Rohban, R., Shakeri, F., & Moghaddam, M. M. (2014). Comparison of Bacterial Biodiversity and Enzyme Production in Three Hypersaline Lakes; Urmia, Howz-Soltan and Aran-Bidgol. *Indian Journal of Microbiology*, *54*(4), 444–449. https://doi.org/10.1007/s12088-014-0481-9

Banciu, H. L., & Muntyan, M. S. (2015). Adaptive strategies in the double-extremophilic prokaryotes inhabiting soda lakes. *Current Opinion in Microbiology*, *25*, 73–79. https://doi.org/10.1016/j.mib.2015.05.003

Banda, J. F., Lu, Y., Hao, C., Pei, L., Du, Z., Zhang, Y., Wei, P., & Dong, H. (2020). The Effects of Salinity and pH on Microbial Community Diversity and Distribution Pattern in the Brines of Soda Lakes in Badain Jaran Desert, China. *Geomicrobiology Journal*, *37*(1), 1–12.

https://doi.org/10.1080/01490451.2019.1654568

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, *19*(5), 455–477. https://doi.org/10.1089/cmb.2012.0021

Bansod, S., Thigale, D., Dure, P., & Sohoni, H. (2012). Production , Purification and Enzyme Kinetics of Amylase From Bacteria Isolated From a Soil of Solapur , Maharashtra State , India. *International Journal of Bioassays*, *3*(2), 1785–1788.

Bar-On, Y. M., Phillips, R., & Milo, R. (2018). The biomass distribution on Earth. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(25), 6506–6511. https://doi.org/10.1073/pnas.1711842115

Baruah, J., Nath, B. K., Sharma, R., Kumar, S., Deka, R. C., Baruah, D. C., & Kalita, E. (2018). Recent trends in the pretreatment of lignocellulosic biomass for value-added products. *Frontiers in Energy Research*, *6*(12), 1–19. https://doi.org/10.3389/fenrg.2018.00141

Bates, L., M., Covington, A. D. , Allen, S. C. and Antunes, A. P. M. (2013). Methods of isolation and identification of pathogenic and potential pathogenic bacteria from skins and tannery effluents. *Journal of the American Leather Chemists Association*, *108*(2), 48–62.

Baumgarte, S. (2003). Microbial diversity of soda lake habitats. *Braunschweig: Carolo-Wilhelmina University; PhD Thesis*, 197.

Behera, B. C., Sethi, B. K., Mishra, R. R., Dutta, S. K., & Thatoi, H. N. (2017). Microbial cellulases – Diversity & biotechnology with reference to mangrove environment: A review. *Journal of Genetic Engineering and Biotechnology*, *15*(1), 197–210. https://doi.org/10.1016/j.jgeb.2016.12.001

Behr, H. J., & Röhricht, C. (2000). Record of seismotectonic events in siliceous cyanobacterial sediments (Magadi cherts), Lake Magadi, Kenya. *International Journal of Earth Sciences*, *89*(2), 268–283. isi:000089932800007

Berben, T., Overmars, L., Sorokin, D. Y., & Muyzer, G. (2019). Diversity and distribution of sulfur oxidation-related genes in *Thioalkalivibrio*, a genus of chemolithoautotrophic and haloalkaliphilic sulfur-oxidizing bacteria. *Frontiers in Microbiology*, *10*(2), 1–15. https://doi.org/10.3389/fmicb.2019.00160

Berini, F., Casciello, C., Marcone, G. L., & Marinelli, F. (2017). Metagenomics: Novel enzymes from non-culturable microbes. *FEMS Microbiology Letters*, *364*(21), 1–19. https://doi.org/10.1093/femsle/fnx211

Berlemont, R., & Martiny, A. C. (2013). Phylogenetic distribution of potential cellulases in bacteria. *Applied and Environmental Microbiology*, *79*(5), 1545–1554. https://doi.org/10.1128/AEM.03305-12

Berlemont, R., & Martiny, A. C. (2015). Genomic potential for polysaccharide deconstruction in bacteria. *Applied and Environmental Microbiology*, *81*(4), 1513–1519.

Berlemont, R., & Martiny, A. C. (2016). Glycoside Hydrolases across Environmental Microbial Communities. *PLoS Computational Biology*, *12*(12), 1–16. https://doi.org/10.1371/journal.pcbi.1005300

Bhattacharya, D., Nagpure, A., Gupta, R. K., & Bhattacharya, D. (2008). Bacterial chitinases : Properties and potential. *Critical Reviews in Biotechnology*, *27*(1), 21–28.

Biswas, J., & Paul, A. K. (2017). Diversity and Production of Extracellular Polysaccharide by Halophilic Microorganisms. *Biodiversity International Journal*, *1*(2), 32–39. https://doi.org/10.15406/bij.2017.01.00006

Bole, S., Maji, A., Dey, A., Acharya, A., & Dubey, S. (2013). Isolation , Purification and Characterization of Amylase From Airborne-Bacteria. *World Journal of Pharmacy and Pharmaceutical Sciences*, *2*(6), 2278 – 4357.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Boros, E., & Kolpakova, M. (2018). A review of the defining chemical properties of soda lakes and pans: An assessment on a large geographic scale of Eurasian inland saline surface waters. *PLoS ONE*, *13*(8). https://doi.org/10.1371/JOURNAL.PONE.0202205

Borzenko, S. V., & Shvartsev, S. L. (2019). Chemical composition of salt lakes in East Transbaikalia (Russia). *Applied Geochemistry*, *103*, 72–84. https://doi.org/10.1016/j.apgeochem.2019.02.014

Boutaiba S, Hacene H, Bidle KA, M.-F. J. (2012). Salt Lakes of the Algerian Sahara. *Journal of Arid Environment*, *75*(10), 909–916. https://doi.org/10.1016/j.jaridenv.2011.04.010.Microbial

Bowers, R. M., Kyrpides, N. C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T. B. K., Schulz, F., Jarett, J., Rivers, A. R., Eloe-Fadrosh, E. A., Tringe, S. G., Ivanova, N. N., Copeland, A., Clum, A., Becraft, E. D., Malmstrom, R. R., Birren, B., Podar, M., Bork, P. and Woyke, T. (2017). Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. In *Nature Biotechnology* (Vol. 35, Issue 8, pp. 725–731). Nature Publishing Group. https://doi.org/10.1038/nbt.3893

Boyadzhieva, I., Tomova, I., Radchenkova, N., Kambourova, M., Poli, A., & Vasileva-Tonkova, E. (2018). Diversity of Heterotrophic Halophilic Bacteria Isolated from Coastal Solar Salterns, Bulgaria and Their Ability to Synthesize Bioactive Molecules with Biotechnological Impact. *Microbiology (Russian Federation)*, *87*(4), 519–528. https://doi.org/10.1134/S0026261718040033

Bradford, M. M. (1976). A Rapid and Sensitive Method for the Quantitation Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. *Analytical Biochemistry*, *254*(72), 248–254.

Bray, J. R., & Curtis, J. T. (1957). An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs*, *27*(4), 325–349.

Britton, K. L., Baker, P. J., Fisher, M., Ruzheinikov, S., Gilmour, D. J., Bonete, M. J., Ferrer, J., Pire, C., Esclapez, J., & Rice, D. W. (2006). Analysis of protein solvent interactions in glucose dehydrogenase from the extreme halophile *Haloferax mediterranei*. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(13), 4846–4851. https://doi.org/10.1073/pnas.0508854103

Buchfink, B., Xie, C., & Huson, D. H. (2014). Fast and sensitive protein alignment using DIAMOND. In *Nature Methods* (Vol. 12, Issue 1, pp. 59–60). Nature Publishing Group. https://doi.org/10.1038/nmeth.3176

Bursy, J., Kuhlmann, A. U., Pittelkow, M., Hartmann, H., Jebbar, M., Pierik, A. J., & Bremer, E. (2008). Synthesis and uptake of the compatible solutes ectoine and 5-hydroxyectoine by *Streptomyces coelicolor* A3(2) in response to salt and heat stresses. *Applied and Environmental Microbiology*, *74*(23), 7286–7296.

Busch, A., Kunert, G., Heckel, D. G., & Pauchet, Y. (2017). Evolution and functional characterization of CAZymes belonging to subfamily 10 of glycoside hydrolase family 5 (GH5_10) in two species of phytophagous beetles. *PLoS ONE*, *12*(8), 1–23. https://doi.org/10.1371/journal.pone.0184305

Cabrera, M. Á., & Blamey, J. M. (2018). Biotechnological applications of archaeal enzymes from extreme environments. *Biological Research*, *51*(1), 1–15. https://doi.org/10.1186/s40659-018-0186-3

Canganella, F., & Wiegel, J. (2011). Extremophiles: From abyssal to terrestrial ecosystems and possibly beyond. *Naturwissenschaften*, *98*(4), 253–279. https://doi.org/10.1007/s00114-011-0775-2

Cantarel, B. I., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., & Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. *Nucleic Acids Research*, *37*(1), 233–238. https://doi.org/10.1093/nar/gkn663

Caporaso, J. G., Bittinger, K., Bushman, F. D., Desantis, T. Z., Andersen, G. L., & Knight, R. (2010). PyNAST: A flexible tool for aligning sequences to a template alignment. *Bioinformatics*, *26*(2), 266–267. https://doi.org/10.1093/bioinformatics/btp636

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Pẽa, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., … Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. In *Nature Methods* (Vol. 7, Issue 5, pp. 335–336). https://doi.org/10.1038/nmeth.f.303

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S. M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J. A., Smith, G., & Knight, R. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME Journal*, *6*(8), 1621–1624. https://doi.org/10.1038/ismej.2012.8

Cárdenas, J. P., Quatrini, R., & Holmes, D. S. (2016). Genomic and metagenomic challenges and opportunities for bioleaching: a mini-review. *Research in Microbiology*, *167*(7), 529–538. https://doi.org/10.1016/j.resmic.2016.06.007

Chandel, A. K., Garlapati, V. K., Singh, A. K., Antunes, F. A. F., & da Silva, S. S. (2018). The path forward for lignocellulose biorefineries: Bottlenecks, solutions, and perspective on commercialization. *Bioresource Technology*, *264*, 370–381. https://doi.org/10.1016/J.BIORTECH.2018.06.004

Chen, D. Di, Fang, B. Z., Manzoor, A., Liu, Y. H., Li, L., Mohamad, O. A. A., Shu, W. S., & Li, W. J. (2021). Revealing the salinity adaptation mechanism in halotolerant bacterium *Egicoccus halophilus* EGI 80432T by physiological analysis and comparative transcriptomics. *Applied Microbiology and Biotechnology*, *105*(6), 2497–2511. https://doi.org/10.1007/s00253-021-11190-5

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Chinnathambi, A. (2015). Industrial important enzymes from alkaliphiles-an overview. *Biosciences Biotechnology Research Asia*, *12*(3), 2007–2016. https://doi.org/10.13005/bbra/1868

Chiu, H. H., Rogozin, D. Y., Huang, S. P., Degermendzhy, A. G., Shieh, W. Y., & Tang, S. L. (2014). *Aliidiomarina shirensis* sp. nov., a halophilic bacterium isolated from Shira Lake in Khakasia, southern Siberia, and a proposal to transfer *Idiomarina maris* to the genus *Aliidiomarina*. *International Journal of Systematic and Evolutionary Microbiology*, *64*(4), 1334–1339. https://doi.org/10.1099/ijs.0.057851-0

Coronado, M. J., Vargas, C., Hofemeister, J., Ventosa, A., & Nieto, J. J. (2000). Production and biochemical characterization of an α-amylase from the moderate halophile *Halomonas meridiana*. *FEMS Microbiology Letters*, *183*(1), 67–71. https://doi.org/10.1016/S0378-1097(99)00628-X

Cortes-Tolalpa, L., Norder, J., van Elsas, J. D., & Falcao Salles, J. (2018). Halotolerant microbial consortia able to degrade highly recalcitrant plant biomass substrate. *Applied Microbiology and Biotechnology*, *102*(6), 2913–2927. https://doi.org/10.1007/s00253-017-8714-6

Costa, O. Y. A., Raaijmakers, J. M., & Kuramae, E. E. (2018). Microbial extracellular polymeric substances: Ecological function and impact on soil aggregation. *Frontiers in Microbiology*, *9*(7), 1–14. https://doi.org/10.3389/fmicb.2018.01636

Coughlan, L. M., Cotter, P. D., Hill, C., & Alvarez-Ordóñez, A. (2015). Biotechnological applications of functional metagenomics in the food and pharmaceutical industries. *Frontiers in Microbiology*, *6*(6), 1–22. https://doi.org/10.3389/fmicb.2015.00672

Couturier, M., Feliu, J., Haon, M., Navarro, D., Lesage-Meessen, L., Coutinho, P. M., & Berrin, J. G. (2011). A thermostable GH45 endoglucanase from yeast: Impact of its atypical multimodularity on activity. *Microbial Cell Factories*, *10*. https://doi.org/10.1186/1475-2859-10-103

Cycil, L. M., DasSarma, S., Pecher, W., McDonald, R., AbdulSalam, M., & Hasan, F. (2020). Metagenomic Insights Into the Diversity of Halophilic Microorganisms Indigenous to the Karak Salt Mine, Pakistan. *Frontiers in Microbiology*, *11*(7), 1567–1582. https://doi.org/10.3389/FMICB.2020.01567/BIBTEX

D'haeseleer, P., Lee, J. Z., Prufert-Bebout, L., Burow, L. C., Detweiler, A. M., Weber, P. K., Karaoz, U., Brodie, E. L., Glavina del Rio, T., Tringe, S. G., Bebout, B. M., & Pett-Ridge, J. (2017). Metagenomic analysis of intertidal hypersaline microbial mats from Elkhorn Slough, California, grown with and without molybdate. *Standards in Genomic Sciences*, *12*(1), 1–8. https://doi.org/10.1186/s40793-017-0279-6

Dadheech, P. K., Glöckner, G., Casper, P., Kotut, K., Mazzoni, C. J., Mbedi, S., & Krienitz, L. (2013). Cyanobacterial diversity in the hot spring, pelagic and benthic habitats of a tropical soda lake. *FEMS Microbiology Ecology*, *85*(2), 389–401. https://doi.org/10.1111/1574-6941.12128

Datta, S., Christena, L. R., & Rajaram, Y. R. S. (2013). Enzyme immobilization: an overview on techniques and support materials. *3 Biotech*, *3*(1), 1–9.

Davies, G. J., Dauter, M., Brzozowski, A. M., Bjørnvad, M. E., Andersen, K. V., & Schülein, M. (1998). Structure of the *Bacillus agaradherans* family 5 endoglucanase at 1.6 Å and its cellobiose complex at 2.0 Å resolution. *Biochemistry*, *37*(7), 1926–1932. https://doi.org/10.1021/bi972162m

De Maayer, P., Anderson, D., Cary, C., & Cowan, D. A. (2014). Some like it cold: Understanding the survival strategies of psychrophiles. *EMBO Reports*, *15*(5), 508–517. https://doi.org/10.1002/embr.201338170

DeAngelis, K. M., Gladden, J. M., Allgaier, M., D'haeseleer, P., Fortney, J. L., Reddy, A., Hugenholtz, P., Singer, S. W., Vander Gheynst, J. S., Silver, W. L., Simmons, B. A., & Hazen, T. C. (2010). Strategies for enhancing the effectiveness of metagenomic-based enzyme discovery in lignocellulolytic microbial communities. *Bioenergy Research*, *3*(2), 146–158. https://doi.org/10.1007/s12155-010-9089-z

Des Marais, D. J. (1990). Microbial mats and the early evolution of life. *Trends in Ecology and Evolution*, *5*(5), 140–144. https://doi.org/10.1016/0169-5347(90)90219-4

Di Donato, P., Buono, A., Poli, A., Finore, I., Abbamondi, G. R., Nicolaus, B., & Lama, L. (2018). Exploring marine environments for the identification of extremophiles and their enzymes for sustainable and green bioprocesses. *Sustainability (Switzerland)*, *11*(1), 149. https://doi.org/10.3390/su11010149

Didari, M., Bagheri, M., Amoozegar, M. A., Bouzari, S., Babavalian, H., Tebyanian, H., Hassanshahian, M., & Ventosa, A. (2020). Diversity of halophilic and halotolerant bacteria in the largest seasonal hypersaline lake (Aran-Bidgol-Iran). *Journal of Environmental Health Science and Engineering*, *18*(8), 961–971. https://doi.org/10.1007/s40201-020-00519-3

Dong, M., Yang, Y., Tang, X., Shen, J., Xu, B., Li, J., & Wu, Q. (2016). NaCl -, protease - tolerant and cold-active endoglucanase from *Paenibacillus* sp . YD236 isolated from the feces of Bos frontalis. *SpringerPlus*, *5*(746), 1–12. https://doi.org/10.1186/s40064-016-2360-9

Dornelas, J.C.M, Figueiredo, J.E.F., Abreu, C. S. De, Oliveira, C. A., & Marriel, I. E. (2017). Characterization and phylogenetic affiliation of Actinobacteria from tropical soils with potential uses for agro-industrial processes. *Genetics and Molecular Research*, *53*(3), 1–16.

Duckworth, a W., Grant, W. D., Jones, B. E., Meijer, D., Márquez, M. C., & Ventosa, A. (2000). *Halomonas magadii* sp. nov., a new member of the genus *Halomonas*, isolated from a soda lake of the East African Rift Valley. *Extremophiles*, *4*(1), 53–60. https://doi.org/10.1007/s007920050137

Duckworth, A. W., Grant, W. D., Jones, B. E., & Van Steenbergen, R. (1996). Phylogenetic diversity of soda lake alkaliphiles. *FEMS Microbiology Ecology*, *19*(3), 181–191. https://doi.org/10.1016/0168-6496(96)00003-7

Duckworth, A W, Grant, W. D., Jones, B. E., & Van Steenbergen, R. (1996). Phylogenetic diversity of soda lake alkaliphiles. *FEMS Microbiology Ecology*, *19*(3), 181–191. https://doi.org/10.1016/0168-6496(96)00003-7

Duckworth, Andrew W., Grant, W. D., Jones, B. E., Meijer, D., Márquez, M. C., & Ventosa, A. (2000). Halomonas magadii sp. nov., a new member of the genus *Halomonas*, isolated from a soda lake of the East African Rift Valley. *Extremophiles*, *4*(1), 53–60. https://doi.org/10.1007/s007920050007

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, *27*(16), 2194–2200. https://doi.org/10.1093/bioinformatics/btr381

Edwin, N., Akhwale, J. K., & Mwirichia, R. (2019). Phylogenetic and morphological diversity of culturable cyanobacteria from Lake Magadi in Kenya. *African Journal of Biological Sciences (South Africa)*, *1*(4), 24–31. https://doi.org/10.33472/AFJBS.1.4.2019.24-31

El-sersy, N. A., Abd-elnaby, H., Abou-elela, G. M., Ibrahim, H. A. H., & El-toukhy, N. M. K. (2010). Optimization, economization and characterization of cellulase produced by marine *Streptomyces ruber*. *African Journal of Biotechnology*, *9*(38), 6355–6364.

El Hidri, D., Guesmi, A., Najjari, A., Cherif, H., Ettoumi, B., Hamdi, C., Boudabous, A., & Cherif, A. (2013). Cultivation-dependant assessment, diversity, and ecology of haloalkaliphilic bacteria in arid saline systems of southern Tunisia. *BioMed Research International*, *2013*(648141). https://doi.org/10.1155/2013/648141

Elleuche, S., Schröder, C., Sahm, K., & Antranikian, G. (2014). Extremozymes-biocatalysts with unique properties from extremophilic microorganisms. *Current Opinion in Biotechnology*, *29*(1), 116–123. https://doi.org/10.1016/j.copbio.2014.04.003

Erdyneeva, E. B., Radnagurueva, A. A., Dunaevsky, Y. E., Belkova, N. L., Namsaraev, Z. B., & Lavrentieva, E. V. (2018). Aminopeptidase Activity of Haloalkalophilic Bacteria of the Genus *Halomonas* Isolated from the Soda-Saline Lakes in the Badain Jaran Desert. *Microbiology (Russian Federation)*, *87*(4), 538–548. https://doi.org/10.1134/S0026261718040069

Ersoy Omeroglu, E., Sudagidan, M., Yurt, M. N. Z., Tasbasi, B. B., Acar, E. E., & Ozalp, V. C. (2021). Microbial community of soda Lake Van as obtained from direct and enriched water, sediment and fish samples. *Scientific Reports*, *11*(1), 1–13. https://doi.org/10.1038/s41598-021-97980-3

Esclapez, J., Bonete, M., Camacho, M., Pire, C., Ferrer, J., Bautista, V., Ma Martínez-Espinosa, R., Zafrilla, B., Pérez-Pomares, F., & Díaz, S. (2006). An optimized method to produce halophilic proteins in *Escherichia coli*. *Microbial Cell Factories*, *5*(1), 1–2. https://doi.org/10.1186/1475-2859-5-S1-S22

Eugester, H. (1971). Origin and Deposition of Trona. *Rocky Mountain Geology*, *10*(1), 49–55.

Eugster, H. P. (1980). Lake magadi, kenya, and its precursors. *Developments in Sedimentology*, *28*(C), 195–232. https://doi.org/10.1016/S0070-4571(08)70239-5

Farhat, W., Venditti, R. A., Hubbe, M., Taha, M., Becquart, F., & Ayoub, A. (2017). A Review of Water-Resistant Hemicellulose-Based Materials: Processing and Applications. *ChemSusChem*, *10*(2), 305–323. https://doi.org/10.1002/cssc.201601047

Feng, J., Zhou, P., Zhou, Y. G., Liu, S. J., & Warren-Rhodes, K. (2005). *Halorubrum alkaliphilum* sp. nov., a novel haloalkaliphile isolated from a soda lake in Xinjiang, China. *International Journal of Systematic and Evolutionary Microbiology*, *55*(1), 149–152. https://doi.org/10.1099/ijs.0.63320-0

Finn, R. D., Clements, J., & Eddy, S. R. (2011). HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Research*, *39*(2), 29–37. https://doi.org/10.1093/nar/gkr367

Foti, M. J., Sorokin, D. Y., Zacharova, E. E., Pimenov, N. V., Kuenen, J. G., & Muyzer, G. (2008). Bacterial diversity and activity along a salinity gradient in soda lakes of the Kulunda Steppe (Altai, Russia). *Extremophiles*, *12*(1), 133–145. https://doi.org/10.1007/s00792-007-0117-7

Furian, S., Martins, E. R. C., Parizotto, T. M., Rezende-Filho, A. T., Victoria, R. L., & Barbiero, L. (2013). Chemical diversity and spatial variability in myriad lakes in Nhecolândia in the Pantanal wetlands of Brazil. *Limnology and Oceanography*, *58*(6), 2249–2261. https://doi.org/10.4319/lo.2013.58.6.2249

Galanie, S., Entwistle, D., Reports, J. L.-N. P., & 2020, U. (2020). Engineering biosynthetic enzymes for industrial natural product synthesis. *Natural Products Report*, *37*(8), 1122–1143. https://doi.org/10.1039/x0xx00000x

Garg, R., Srivastava, R., Brahma, V., Verma, L., Karthikeyan, S., & Sahni, G. (2016). Biochemical and structural characterization of a novel halotolerant cellulase from soil metagenome. *Scientific Reports*, *6*(June), 1–15. https://doi.org/10.1038/srep39634

Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., B. A. (2005). *Protein Identification and Analysis Tools on the ExPASy Server;(In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press* (pp. 571–607). https://web.expasy.org/protparam/protpar-ref.html

Gaur, R., & Tiwari, S. (2015). Isolation, production, purification and characterization of an organic-solvent-thermostable alkalophilic cellulase from *Bacillus vallismortis* RG-07. *BMC Biotechnology*, *54*(4), 424–429.

Gessesse, A., & Mamo, G. (1998). Purification and characterization of an alkaline xylanase from alkaliphilic *Micrococcus* sp AR-135. *Journal of Industrial Microbiology and Biotechnology*, *20*(3–4), 210–214. https://doi.org/10.1038/sj.jim.2900503

Gessesse, Amare, & Gashe, B. A. (1997). Production of alkaline protease by an alkaliphilic bacteria isolated from an alkaline soda lake. *Biotechnology Letters*, *19*(5), 479–481. https://doi.org/10.1023/A:1018308513853

Getenet, M., García-Ruiz, J. M., Otálora, F., Emmerling, F., Al-Sabbagh, D., & Verdugo-Escamilla, C. (2022). A Comprehensive Methodology for Monitoring Evaporitic Mineral Precipitation and Hydrochemical Evolution of Saline Lakes: The Case of Lake Magadi Soda Brine (East African Rift Valley, Kenya). *Crystal Growth and Design*, *22*(4), 2307–2317. https://doi.org/10.1021 /acs.cgd. 1c01391

Ghauri, M. A., Khalid, A. M., Grant, S., Heaphy, S., & Grant, W. D. (2003). Phylogenetic analysis of different isolates of *Sulfobacillus* spp. isolated from uranium-rich environments and recovery of genes using integron-specific primers. *Extremophiles*, *7*(5), 341–345. https://doi.org/10.1007/s00792-003-0354-3

Gohel, V., Singh, A., Vimal, M., Ashwini, P., & Chhatpar, H. S. (2006). Bioprospecting and antifungal potential of chitinolytic microorganisms. *African Journal of Biotechnology*, *5*(2), 54–72. https://doi.org/10.5897/AJB2006.000-5010

Gomes, J., & Steiner, W. (2004). The Biocatalytic Potential of Extremophiles and Extremozymes. *Food Technology and Biotechnology*, *42*(4), 223–235.

Grant, W.D. and Sorokin, D. . (2011). Distribution and Diversity of Soda Lake Alkaliphiles. In *Extremophiles Handbook* (Issue July 2015, pp. 28–54). https://doi.org/10.1007/978-4-431-53898-1

Grant, William D. Jones, B. E. (2016). Soda lakes of East Africa. In *Soda Lakes of East Africa* (pp. 1–408). https://doi.org/10.1007/978-3-319-28622-8

Grant, S., Grant, W. D., Jones, B. E., Kato, C., & Li, L. (1999). Novel archaeal phylotypes from an East African alkaline saltern. *Extremophiles*, *3*(2), 139–145. https://doi.org/10.1007/s007920050109

Grant, W D. (2006a). Alkaline environments and biodiversity. *Encyclopedia of Life Support Systems*, *3*(20), 1–7.

Grant, W D, Mwatha, W. E., & Jones, B. E. (1990). Alkaliphiles: ecology, diversity and applications. *FEMS Microbiology Reviews*, *6*(2–3), 255–269.

Grant, William D., & Heaphy, S. (2010). Metagenomics and recovery of enzyme genes from alkaline saline environments. *Environmental Technology*, *31*(10), 1135–1143. https://doi.org/10.1080/09593331003646661

Guan, T. W., Lin, Y. J., Ou, M. Y., & Chen, K. B. (2020). Isolation and diversity of sediment bacteria in the hypersaline aiding lake, China. *PLoS ONE*, *15*(7), 13–16. https://doi.org/10.1371/journal.pone.0236006

Gunde-Cimerman, N., Plemenitaš, A., & Oren, A. (2018). Strategies of adaptation of microorganisms of the three domains of life to high salt concentrations. *FEMS Microbiology Reviews*, *42*(3), 353–375. https://doi.org/10.1093/femsre/fuy009

Gunny, A. A. N., Arbain, D., Edwin Gumba, R., Jong, B. C., & Jamal, P. (2014). Potential halophilic cellulases for *in situ* enzymatic saccharification of ionic liquids pretreated lignocelluloses. *Bioresource Technology*, *155*, 177–181. https://doi.org/10.1016/j.biortech.2013.12.101

Guo, L., Liu, J., Hou, Q., Zhao, B., Zhang, X., & Wang, H. (2023). *Salipaludibacillus daqingensis* sp. nov., a moderately halophilic bacterium isolated from an oilfield. *International Journal of Systematic and Evolutionary Microbiology*, *73*(11), 006126. https://doi.org/10.1099/IJSEM.0.006126/CITE/REFWORKS

Gupta, P., Samant, K., & Sahu, A. (2012). Isolation of cellulose-degrading bacteria and determination of their cellulolytic potential. *International Journal of Microbiology*, *2012*. https://doi.org/10.1155/2012/578925

Guzman, L. M., Weiss, D. S., & Beckwith, J. (1995). Domain-swapping analysis of FtsI, FtsL, and FtsQ, bitopic membrane proteins essential for cell division in *Escherichia coli*. *Journal of Bacteriology*, *179*(16), 5094–5103. https://doi.org/10.1128/jb.179.16.5094-5103.1997

Han, R., Zhang, X., Liu, J., Long, Q., Chen, L., Liu, D., & Zhu, D. (2017). Microbial community structure and diversity within hypersaline Keke Salt Lake environments. *Canadian Journal of Microbiology*, *63*(11), 895–908. https://doi.org/10.1139/cjm-2016-0773

Handelsman, J. (2005). Metagenomics: Application of Genomics to Uncultured Microorganisms. *Microbiology and Molecular Biology Reviews*, *69*(1), 195–195. https://doi.org/10.1128/mmbr.69.1.195.2005

Harvey, A. J., Hrmova, M., De Gori, R., Varghese, J. N., & Fincher, G. B. (2000). Comparative modeling of the three-dimensional structures of family 3 glycoside hydrolases. *Proteins: Structure, Function and Genetics*, *41*(2), 257–269. https://doi.org/10.1002/1097-0134(20001101)41:2<257::AID-PROT100>3.0.CO;2-C

He, H., Fu, L., Liu, Q., Fu, L., Bi, N., Yang, Z., & Zhen, Y. (2019). Community Structure, Abundance and Potential Functions of Bacteria and Archaea in the Sansha Yongle Blue Hole, Xisha, South China Sea. *Frontiers in Microbiology*, *10*(2404), 1–16. https://doi.org/10.3389/fmicb.2019.02404

Heffron, J. D., Sherry, N., & Popham, D. L. (2011). In vitro studies of peptidoglycan binding and hydrolysis by the *Bacillus anthracis* germination-specific lytic enzyme SleB. *Journal of Bacteriology*, *193*(1), 125–131. https://doi.org/10.1128/JB.00869-10

Heyer, J., Berger, U., Hardt, M., & Dunfield, P. F. (2005). *Methylohalobius crimeensis* gen. nov., sp. nov., a moderately halophilic, methanotrophic bacterium isolated from hypersaline lakes of Crimea. *International Journal of Systematic and Evolutionary Microbiology*, *55*(5), 1817–1826.

Hirasawa, K., Uchimura, K., Kashiwa, M., Grant, W. D., Ito, S., Kobayashi, T., & Horikoshi, K. (2006). Salt-activated endoglucanase of a strain of alkaliphilic *Bacillus agaradhaerens*. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, *89*(2), 211–219. https://doi.org/10.1007/s10482-005-9023-0

Horikoshi, K. (1999). Alkaliphiles: some applications of their products for biotechnology. *Microbiology and Molecular Biology Reviews*, *63*(4), 735–750.

Hosam, E. E., Mohamad, S. A. E.-K., & Mai, A. W. (2017). Biodiversity of uncultured bacteria in hypersaline lakes, Siwa oasis, Egypt, as determined by polymerase chain reaction and denaturing gradient gel electrophoresis (PCR-DGGE) of 16S rRNA gene phylotypes. *International Journal of Genetics and Molecular Biology*, *9*(2), 8–15. https://doi.org/10.5897/ijgmb2016.0130

Hoster, F., Schmitz, J. E., & Daniel, R. (2005). Enrichment of chitinolytic microorganisms: Isolation and characterization of a chitinase exhibiting antifungal activity against phytopathogenic fungi from a novel *Streptomyces* strain. *Applied Microbiology and Biotechnology*, *66*(4), 434–442. https://doi.org/10.1007/s00253-004-1664-9

Hou, H., He, H., & Wang, Y. (2020). Effects of SDS on the activity and conformation of protein tyrosine phosphatase from *Thermus thermophilus* HB27. *Scientific Reports*, *10*(1), 3195–3202. https://doi.org/10.1038/s41598-020-60263-4

Hsieh, C. C., Cannella, D., Jørgensen, H., Felby, C., & Thygesen, L. G. (2015). Cellobiohydrolase and endoglucanase respond differently to surfactants during

the hydrolysis of cellulose. *Biotechnology for Biofuels*, *8*(1), 52–62. https://doi.org/10.1186/S13068-015-0242-Y

Huang, S., Sheng, P., & Zhang, H. (2012). Isolation and identification of cellulolytic bacteria from the gut of *Holotrichia parallela* larvae (Coleoptera: Scarabaeidae). *International Journal of Molecular Sciences*, *13*(3), 2563–2577.

Huang, Xiaoluo, Shao, Z., Hong, Y., Lin, L., Li, C., Huang, F., Wang, H., & Liu, Z. (2010). Cel8H, a novel endoglucanase from the halophilic bacterium *Halomonas* sp. S66-4: Molecular cloning, heterogonous expression, and biochemical characterization. *Journal of Microbiology*, *48*(3), 318–324. https://doi.org/10.1007/s12275-009-0188-5

Huang, Xiaoyun, Lin, J., Ye, X., & Wang, G. (2015). Molecular characterization of a thermophilic and salt- and alkaline-tolerant xylanase from *Planococcus* sp. SL4, a strain isolated from the sediment of a soda lake. *Journal of Microbiology and Biotechnology*, *25*(5), 662–671. https://doi.org/10.4014/jmb.1408.08062

Hugoni, M., Escalas, A., Bernard, C., Nicolas, S., Jézéquel, D., Vazzoler, F., Sarazin, G., Leboulanger, C., Bouvy, M., Got, P., Ader, M., Troussellier, M., & Agogué, H. (2018). Spatiotemporal variations in microbial diversity across the three domains of life in a tropical thalassohaline lake (Dziani Dzaha, Mayotte Island). *Molecular Ecology*, *27*(23), 4775–4786. https://doi.org/10.1111/MEC.14901

Huy, N. D., Nguyen, C. Le, Park, H. S., Loc, N. H., Choi, M. S., Kim, D. H., Seo, J. W., & Park, S. M. (2016). Characterization of a novel manganese dependent endoglucanase belongs in GH family 5 from *Phanerochaete chrysosporium*. *Journal of Bioscience and Bioengineering*, *121*(2), 154–159. https://doi.org/10.1016/J.JBIOSC.2015.06.009

Hyatt, D., Chen, G. L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, *11*, 119. https://doi.org/10.1186/1471-2105-11-119

Hyun, D. W., Whon, T. W., Cho, Y. J., Chun, J., Kim, M. S., Jung, M. J., Shin, N. R., Kim, J. Y., Kim, P. S., Yun, J. H., Lee, J., Oh, S. J., & Bae, J. W. (2013). Genome sequence of the moderately halophilic bacterium *Salinicoccus carnicancri* type strain CrmT (= DSM 23852T). *Standards in Genomic Sciences*, *8*(2), 255–263. https://doi.org/10.4056/sigs.3967649

Ibrahim, A. S. S., Elbadawi, Y. B., El-Tayeb, M. A., Al-maary, K. S., Maany, D. A. F., Ibrahim, S. S. S., & Elagib, A. A. (2019). Alkaline serine protease from the new halotolerant alkaliphilic *Salipaludibacillus agaradhaerens* strain AK-R: purification and properties. *3 Biotech*, *9*(11), 1–11. https://doi.org/10.1007/s13205-019-1928-9

Ishibashi, M., Oda, K., Arakawa, T., & Tokunaga, M. (2011). Cloning, expression, purification and activation by Na ion of halophilic alkaline phosphatase from moderate halophile *Halomonas* sp. 593. *Protein Expression and Purification*, *76*(1), 97–102. https://doi.org/10.1016/j.pep.2010.09.013

Isikgor, F. H., & Becer, C. R. (2015). Lignocellulosic biomass: a sustainable platform for the production of bio-based chemicals and polymers. *Polymer Chemistry*, *6*(25), 4497–4559. https://doi.org/10.1039/c5py00263j

Janaki, T. (2017). Enzymes From Actinomycetes – Review. *International Journal of ChemTech Research*, *10*(2), 176–182.

Jaralla, M. E., Al-Dabbagh, N. N., & Hameed, N. and Abdul-Hussein, N. (2014). Screening for enzymatic production ability and antimicrobial activity of Actinomycetes isolated from soil in Hillah/Iraq. *Journal of Pharmacy and Biological Sciences*, *9*(5), 42–47.

Jayachandra, S., Kumar, A., Shouche, S. &, & Mulochana, M. (2013). Culturable Diversity of Extremely Halophilic Bacteria From. *International Journal of Biology, Pharmacy and Allied Sciences*, *2*(2), 391–405.

Jayani, R. S., Saxena, S., & Gupta, R. (2005). Microbial pectinolytic enzymes: A review. *Process Biochemistry*, *40*(9), 2931–2944. https://doi.org/10.1016/j.procbio.2005.03.026

Jenifer, J. S. C. A., Donio, M. T. B. S., Viji, V. T., Velmurugan, S., Babu, M. M., dhas Albin, S., & Citarasu, T. (2015). Haloalkaliphilic actinomycetes from solar salt works in India: Diversity and antimicrobial activity. *Heavy Metals, Algae and Microbial Activities in Marine Systems*, *2*(1), 375–389.

Jiang, H., Dong, H., Yu, B., Liu, X., Li, Y., Ji, S., & Zhang, C. L. (2007). Microbial response to salinity change in Lake Chaka, a hypersaline lake on Tibetan plateau. *Environmental Microbiology*, *9*(10), 2603–2621. https://doi.org/10.1111/j.1462-2920.2007.01377.x

Joachimiak, M. P., Weisman, J. L., & May, B. C. H. (2006). JColorGrid: Software for the visualization of biological measurements. *BMC Bioinformatics*, *7*, 1–5. https://doi.org/10.1186/1471-2105-7-225

Jones, B. E., Grant, W. D., Duckworth, A. W., & Owenson, G. G. (1998). Microbial diversity of soda lakes. *Extremophiles*, *2*(3), 191–200. https://doi.org/ 10.1007 /s007920050060

Jones, B. E., Grant, W. D., Duckworth, A. W., Owenson, G. G., Horikoshi, K., Jones, B. E., Grant, W. D., Duckworth, · A W, & Owenson, · G G. (1998). Microbial diversity of soda lakes. *Extremophiles*, *2*, 191–200.

Jones, B. F., Eugster, H. P., & Rettig, S. L. (1977). Hydrochemistry of the Lake Magadi basin, Kenya. *Geochimica et Cosmochimica Acta*, *41*(1), 53–72. https://doi.org/10.1016/0016-7037(77)90186-7

Joshi, A., Thite, S., Karodi, P., Joseph, N., & Lodha, T. (2021). *Alkalihalobacterium elongatum* gen. nov. sp. nov.: An Antibiotic-Producing Bacterium Isolated From Lonar Lake and Reclassification of the Genus *Alkalihalobacillus* Into Seven Novel Genera. *Frontiers in Microbiology*, *12*(10), 1–22. https://doi.org/10.3389/fmicb.2021.722369

Joshi, A., Thite, S., Lodha, T., Joseph, N., & Mengade, P. (2023). Genomic insights

of an alkaliphilic bacterium *Halalkalibacter alkaliphilus* sp. nov. isolated from an Indian Soda Lake. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, *116*(5), 435–445. https://doi.org/10.1007/S10482-023-01816-1/METRICS

Juan, J., Armenteros, A., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., Von Heijne, G., & Nielsen, H. (2019). Brief Communication SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology*, *37*, 420–423. https://doi.org/10.1038/s41587-019-0036-z

Jung, Y., Lee, Y., Park, I., Chandra, M. S., Kim, K., & Choi, Y. (2010). glucanase from *Bacillus subtilis* A8-8 Molecular cloning , purification and characterization of thermostable β -1 , 3-1 , 4 glucanase from *Bacillus subtilis* A8-8. *Indian Journal of Biochemistry & Biophysics*, *7*(August), 203–210.

Kaeberlein, T. and Epstein, S. S. (2002). Isolating "Uncultivable" Microorganisms in Pure Culture in a Simulated Natural Environment. *Science*, *296*(5570), 1127–1129.

Kamagata, Y., & Tamaki, H. (2005). Cultivation of Uncultured Fastidious Microbes. *Microbes and Environments*, *20*(2), 85–91.

Kamble, R. D., & Jadhav, A. R. (2012). Isolation, purification, and characterization of xylanase produced by a new species of *Bacillus* in solid state fermentation. *International Journal of Microbiology*, *2012*.

Kambura, A. K. (2016). Metagenomic and metatranscriptomic analysis of bacterial, Archaeal and Fungal communities within the hot springs of Lake Magadi in Kenya. *PhD Thesis, Jomo Kenyatta University of Agrioculture and Technology, Kenya*.

Kambura, A. K., Mwirichia, R. K., Kasili, R. W., Karanja, E. N., Makonde, H. M., & Boga, H. I. (2016). Bacteria and Archaea diversity within the hot springs of Lake Magadi and Little Magadi in Kenya. *BMC Microbiology*, *16*(1), 1–12.

Kamekura, M., Dyall-smith, M. L., Upasani, V., Ventosa, A., & Kates, M. (1997). Diversity of Alkaliphilic Halobacteria : Proposals for Transfer of *Natronobacterium vacuolatum*, *Natronobacterium magadii*, and *Natronobacterium pharaonis* to *Halorubrum*, *Natrialba*, and *Natronomonas* gen . nov ., Respectively, as *Halorubrum*. *International Journal of Systematic Bacteriology*, *47*(7), 853–857.

Kanehisa, M., Sato, Y., & Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology*, *428*(4), 726–731. https://doi.org/10.1016/j.jmb.2015.11.006

Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., & Wang, Z. (2019). MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*, *2019*(7), 1–13. https://doi.org/10.7717/peerj.7359

Karan, R., Capes, M. D., & DasSarma, S. (2012). Function and biotechnology of extremophilic enzymes in low water activity. *Aquatic Biosystems*, *8*(1), 4.

Karray, F., Ben Abdallah, M., Kallel, N., Hamza, M., Fakhfakh, M., & Sayadi, S. (2018). Extracellular hydrolytic enzymes produced by halophilic bacteria and archaea isolated from hypersaline lake. *Molecular Biology Reports*, *45*(5), 1297–1309. https://doi.org/10.1007/s11033-018-4286-5

Kasana, R. C., Salwan, R., Dhar, H., Dutt, S., & Gulati, A. (2008). A rapid and easy method for the detection of microbial cellulases on agar plates using Gram's iodine. *Current Microbiology*, *57*(5), 503–507. https://doi.org/10.1007/s00284-008-9276-8

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kempe, S., & Kazmierczak, J. (2011). Soda Lakes. In *Encyclopedia of Earth Sciences Series* (Issue 978, pp. 824–829). Springer, Dordrecht. https://doi.org/10.1007/978-1-4020-9212-1_191

Kevbrin, V. V., Zhilina, T. N., Rainey, F. A., & Zavarzin, G. A. (1998). *Tindallia magadii* gen. nov., sp. nov.: An alkaliphilic anaerobic ammonifier from Soda Lake Deposits. *Current Microbiology*, *37*(2), 94–100.

Khalid, A., Tayyab, M., Shakoori, A. R., Hashmi, A. S., Yaqub, T., Awan, A. R., Wasim, M., Firyal, S., Hussain, Z., & Ahmad, M. (2019). Cloning, Expression and Characterization of Highly Active Recombinant Thermostable Cellulase from Thermotoga naphthophila. *Pakistan Journal of Zoology*, *51*(3), 925–934. https://doi.org/10.17582/journal.pjz/2019.51.3.925.934

Khandeparker, L., Gaonkar, C. C., & Desai, D. V. (2013). Degradation of barnacle nauplii: Implications to chitin regulation in the marine environment. *Biologia (Poland)*, *68*(4), 696–706. https://doi.org/10.2478/s11756-013-0202-6

Khrustalev, V. V., Barkovsky, E. V., & Khrustaleva, T. A. (2016). Magnesium and manganese binding sites on proteins have the same predominant motif of secondary structure. *Journal of Theoretical Biology*, *395*, 174–185. https://doi.org/10.1016/J.JTBI.2016.02.006

Kim, K. (2002). Characterization of 1,4-benzoquinone reductase from bovine liver. *Biotechnology and Bioprocess Engineering*, *7*(4), 216–220. https://doi.org/10.1007/BF02932973

Kim, K. H., Han, D. M., Lee, J. K., & Jeon, C. O. (2023). *Alkalicoccobacillus porphyridii* sp. nov., isolated from a marine red alga, reclassification of *Shouchella plakortidis* and *Shouchella gibsonii* as *Alkalicoccobacillus plakortidis* comb. nov. and *Alkalicoccobacillus gibsonii* comb. nov., and emended descriptio. *International Journal of Systematic and Evolutionary Microbiology*, *73*(8), 1–10. https://doi.org/10.1099/ijsem.0.006019

Kim, K. K., Lee, J. S., & Stevens, D. A. (2013). Microbiology and epidemiology of Halomonas species. *Future Microbiology*, *8*(12), 1559–1573.

https://doi.org/10.2217/fmb.13.108

Kimbrel, J. A., Ballor, N., Wu, Y. W., David, M. M., Hazen, T. C., Simmons, B. A., Singer, S. W., & Jansson, J. K. (2018). Microbial community structure and functional potential along a hypersaline gradient. *Frontiers in Microbiology*, *9*(7), 1492. https://doi.org/10.3389/fmicb.2018.01492

Kiplimo, D., Mugweru, J., & Kituyi, Sarah, Kipnyargis, Alex, Mwirichia, R. (2019). Diversity of esterase and lipase producing haloalkaliphilic bacteria from Lake Magadi in Kenya. *Journal of Basic Microbiology*, *59*(12), 1173–1184. https://doi.org/10.1002/jobm.201900353

Kipnyargis, A., Mwirichia, R., Pfeiffer, B., & Daniel, R. (2022). Complete Genome Sequence of *Alkalihalobacillus* sp. Strain LMS39, a Haloalkaliphilic Bacterium Isolated from a Hypersaline Lake. *Microbiology Resource Announcements*, *11*(7). https://doi.org/10.1128/MRA.00325-22

Kobayashi, T., Kanai, H., Hayashi, T., Akiba, T., Akaboshi, R., Horikoshi, K. (1992). Haloalkaliphilic Maltotriose-Forming ox-Amylase from the Archaebacterium Natronococcus sp. Strain Ah-36. *Journal of Bacteriology*, *174*(11), 3439–3444.

Kopejtka, K., Tomasch, J., Bunk, B., Spröer, C., Wagner-Döbler, I., & Koblížek, M. (2018). The complete genome sequence of *Rhodobaca barguzinensis* alga05 (DSM 19920) documents its adaptation for life in soda lakes. *Extremophiles*, *22*(6), 839–849. https://doi.org/10.1007/s00792-018-1041-8

Krenitz, L., & Shagerl, M. (2016). Tiny and Tough: Microphytes of East African Soda Lakes. In M. Schagerl (Ed.), *Soda Lakes of East Africa* (pp. 149–177). Springer.

Krienitz, L., Dadheech, P. K., & Kotut, K. (2013). Mass developments of the cyanobacteria *Anabaenopsis* and *Cyanospira* (Nostocales) in the soda lakes of Kenya: Ecological and systematic implications. *Hydrobiologia*, *703*(1), 79–93. https://doi.org/10.1007/s10750-012-1346-z

Krulwich, T. A., Sachs, G., & Padan, E. (2011). Molecular aspects of bacterial pH sensing and homeostasis. *Nature Reviews Microbiology*, *9*(5), 330.

Kucharska, K., Rybarczyk, P., Hołowacz, I., Łukajtis, R., Glinka, M., & Kamiński, M. (2018). Pretreatment of lignocellulosic materials as substrates for fermentation processes. *Molecules*, *23*(11), 1–32. https://doi.org/10.3390/molecules23112937

Kumar, G, P., & Suneetha, V. (2015). Pectinases from Actinomycetes : A Thorough Study. *International Journal of ChemTech Research*, *8*(7), 345–350.

Kumar, Sudhir, Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549. https://doi.org/10.1093/molbev/msy096

Kumar, Sumit, Karan, R., Kapoor, S., Singh, S. P., & Khare, S. K. (2012). Screening and isolation of halophilic bacteria producing industrially important enzymes.

Brazilian Journal of Microbiology, *43*(4), 1595–1603. https://doi.org/10.1590/S1517-838220120004000044

Labes, A., Karlsson, E. N., Fridjonsson, O. H., Turner, P., Hreggvidson, G. O., Kristjansson, J. K., Holst, O., & Schönheit, P. (2008). Novel members of glycoside hydrolase family 13 derived from environmental DNA. *Applied and Environmental Microbiology*, *74*(6), 1914–1921. https://doi.org/10.1128/AEM.02102-07

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, *227*(5259), 680–685. https://doi.org/10.1038/227680a0

Lameck, A. S., Skutai, J., & Boros, E. (2023). Review of chemical properties of inland soda and saline waters in East Africa (rift valley region). *Journal of Hydrology: Regional Studies*, *46*(4), 101323. https://doi.org/10.1016/J.EJRH.2023.101323

Langmead, B., & Salzberg, S. (2013). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923.Fast

Lanzén, A., Simachew, A., Gessesse, A., Chmolowska, D., Jonassen, I., & Øvreås, L. (2013). Surprising Prokaryotic and Eukaryotic Diversity, Community Structure and Biogeography of Ethiopian Soda Lakes. *PLoS ONE*, *8*(8). https://doi.org/10.1371/journal.pone.0072577

Le, B., Chung, G., & Yang, S. H. (2018). Chitinase-producing Salinivibrio bacteria isolated from salt-fermented shrimp with antimicrobial and safety assessments. *Journal of Applied Biological Chemistry*, *61*(3), 233–238. https://doi.org/10.3839/jabc.2018.034

Lefort, V., Desper, R., & Gascuel, O. (2015). FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program. *Molecular Biology and Evolution*, *32*(10), 2798–2800. https://doi.org/10.1093/molbev/msv150

Letunic, I., & Bork, P. (2019). Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Research*, *47*(W1), 256–259. https://doi.org/10.1093/nar/gkz239

Levasseur, A., Drula, E., Lombard, V., Coutinho, P. M., & Henrissat, B. (2013). Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnology for Biofuels*, *6*(1), 1. https://doi.org/10.1186/1754-6834-6-41

Li, F., Dong, J., Lv, X., Wen, Y., & Chen, S. (2020). Recombinant expression and characterization of two glycoside hydrolases from extreme alklinphilic bacterium *Cellulomonas bogoriensis* 69B4T. *AMB Express*, *10*(1), 1–10. https://doi.org/10.1186/s13568-020-00979-8

Li, H., Hu, Q., Hong, X., Jiang, Z., Ni, H., Li, Q., & Zhu, Y. (2021). Molecular cloning and characterization of a thermostable and halotolerant endo-β-1,4-glucanase from *Microbulbifer* sp. ALW1. *3 Biotech*, *11*(5), 1–11.

https://doi.org/10.1007/s13205-021-02801-z

Li, R., Khafipour, E., Krause, D. O., Entz, M. H., de Kievit, T. R., & Fernando, W. G. D. (2012). Pyrosequencing Reveals the Influence of Organic and Conventional Farming Systems on Bacterial Communities. *PLoS ONE*, *7*(12). https://doi.org/10.1371/journal.pone.0051897

Li, Xiangyuan, Yang, M., Mu, T., Miao, D., Liu, J., & Xing, J. (2022). Composition and key-influencing factors of bacterial communities active in sulfur cycling of soda lake sediments. *Archives of Microbiology*, *204*(6), 1–19. https://doi.org/10.1007/s00203-022-02925-7

Li, Xin, Wang, H. L., Li, T., & Yu, H. Y. (2012). Purification and characterization of an organic solvent-tolerant alkaline cellulase from a halophilic isolate of Thalassobacillus. *Biotechnology Letters*, *34*(8), 1531–1536. https://doi.org/10.1007/s10529-012-0938-z

Li, Xin, & Yu, H. Y. (2012). Purification and characterization of an organic-solvent-tolerant cellulase from a halotolerant isolate, *Bacillus* sp. L1. *Journal of Industrial Microbiology and Biotechnology*, *39*(8), 1117–1124. https://doi.org/10.1007/s10295-012-1120-2

Liang, C., Xue, Y., Fioroni, M., Rodríguez-Ropero, F., Zhou, C., Schwaneberg, U., & Ma, Y. (2011). Cloning and characterization of a thermostable and halo-tolerant endoglucanase from *Thermoanaerobacter tengcongensis* MB4. *Applied Microbiology and Biotechnology*, *89*(2), 315–326. https://doi.org/10.1007/s00253-010-2842-6

Liao, Z., Holtzapple, M., Yan, Y., Wang, H., Li, J., & Zhao, B. (2019). Insights into xylan degradation and haloalkaline adaptation through whole-genome analysis of *Alkalitalea saponilacus*, an anaerobic haloalkaliphilic bacterium capable of secreting novel halostable xylanase. *Genes*, *10*(1), 1–11. https://doi.org/10.3390/genes10010001

Liberato, M. V., Silveira, R. L., Prates, E. T., De Araujo, E. A., Pellegrini, V. O. A., Camilo, C. M., Kadowaki, M. A., Neto, M. D. O., Popov, A., Skaf, M. S., & Polikarpov, I. (2016). Molecular characterization of a family 5 glycoside hydrolase suggests an induced-fit enzymatic mechanism. *Scientific Reports*, *6*(February), 1–16. https://doi.org/10.1038/srep23473

Lin, L., Liu, X., Zhou, Y., Guan, L., He, J., & Huang, W. (2016). A novel pH-stable, endoglucanase (JqCel5A) isolated from a salt-lake microorganism, *Jonesia quinghaiensis*. *Electronic Journal of Biotechnology*, *24*, 56–62. https://doi.org/10.1016/j.ejbt.2016.09.004

Lindahl, B. D., & Finlay, R. D. (2006). Activities of chitinolytic enzymes during primary and secondary colonization of wood by basidiomycetous fungi. *New Phytologist*, *169*(2), 389–397. https://doi.org/10.1111/j.1469-8137.2005.01581.x

Linhoff, B. S., Bennett, P. C., Puntsag, T., & Gerel, O. (2011). Geochemical evolution of uraniferous soda lakes in Eastern Mongolia. *Environmental Earth Sciences*,

62(1), 171–183. https://doi.org/10.1007/s12665-010-0512-8

Liu, C., Cui, Y., Li, X., & Yao, M. (2021). Microeco: An R package for data mining in microbial community ecology. *FEMS Microbiology Ecology*, *97*(2), 1–9. https://doi.org/10.1093/femsec/fiaa255

Liu, N., Li, H., Chevrette, M. G., Zhang, L., Cao, L., Zhou, H., Zhou, X., Zhou, Z., Pope, P. B., Currie, C. R., Huang, Y., & Wang, Q. (2019). Functional metagenomics reveals abundant polysaccharide-degrading gene clusters and cellobiose utilization pathways within gut microbiota of a wood-feeding higher termite. *ISME Journal*, *13*(1), 104–117. https://doi.org/10.1038/s41396-018-0255-1

Lovegrove, A., Edwards, C.H., Noni, I.D., Patel, H., El, S.N., Grassby, T., Zielke, C., Ulmius, M., Nilssone, L., Butterworth, P.J. , Ellis, P.R. and Shewry, P. R. (2017). Role of polysaccharides in food, digestion, and health. *Critical Reviews in Food Science and Nutrition*, *57*(2), 237–253.

Lozupone, C., & Knight, R. (2005). UniFrac: A new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology*, *71*(12), 8228–8235. https://doi.org/10.1128/AEM.71.12.8228-8235.2005

Lu, Zhengui Wang, Qinhong Jiang, Sijing Zhang, Guimin Ma, Y. (2016). Truncation of the unique N-terminal domain improved the thermos-stability and specific activity of alkaline α-amylase Amy703. *Scientific Reports*, *6*(1), 1–10. https://doi.org/10.1038/srep22465

Lu, M., & Daniel, R. (2021). A novel carboxylesterase derived from a compost metagenome exhibiting high stability and activity towards high salinity. *Genes*, *12*(1), 1–18. https://doi.org/10.3390/genes12010122

Lu, M., Dukunde, A., & Daniel, R. (2019). Biochemical profiles of two thermostable and organic solvent–tolerant esterases derived from a compost metagenome. *Applied Microbiology and Biotechnology*, *103*(8), 3421–3437. https://doi.org/10.1007/s00253-019-09695-1

Ma, D., Chen, H., Liu, D., Feng, C., Hua, Y., Gu, T., Guo, X., Zhou, Y., Wang, H., Tong, G., Li, H., & Zhang, K. (2024). Soil-derived cellulose-degrading bacteria: screening, identification, the optimization of fermentation conditions, and their whole genome sequencing. *Frontiers in Microbiology*, *15*(7), 1–14. https://doi.org/10.3389/fmicb.2024.1409697

Ma, Y., Xue, Y., Grant, W. D., Collins, N. C., Duckworth, A. W., Van Steenbergen, R. P., & Jones, B. E. (2004). *Alkalimonas amylolytica* gen. nov., sp. nov., and *Alkalimonas delamerensis* gen. nov., sp. nov., novel alkaliphilic bacteria from soda lakes in China and East Africa. *Extremophiles*, *8*(3), 193–200. https://doi.org/10.1007/s00792-004-0377-4

Magoč, Tanja & Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, *27*(21), 2957–2963. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3198573/

Mahmoud, A. M. A., Ali, M. H. H., Abdelkarim, M. S., & Al-Afify, A. D. G. (2024). Chemical, biochemical, and bioactivity studies on some soda lakes, Wadi El-Natrun, Egypt. *Environmental Monitoring and Assessment*, *196*(5), 436. https://doi.org/10.1007/S10661-024-12573-7/TABLES/5

Maki, M., Leung, K. T., & Qin, W. (2009). The prospects of cellulase-producing bacteria for the bioconversion of lignocellulosic biomass. *International Journal of Biological Sciences*, *5*(5), 500–516. https://doi.org/10.7150/ijbs.5.500

Mani, K., Taib, N., Hugoni, M., Bronner, G., Bragança, J. M., & Debroas, D. (2020). Transient dynamics of archaea and bacteria in sediments and brine across a salinity gradient in a solar saltern of Goa, India. *Frontiers in Microbiology*, *11*(8), 1–19. https://doi.org/10.3389/fmicb.2020.01891

Manzo, N., D'Apuzzo, E., Coutinho, P. M., Cutting, S. M., Henrissat, B., & Ricca, E. (2011). Carbohydrate-active enzymes from pigmented Bacilli: A genomic approach to assess carbohydrate utilization and degradation. *BMC Microbiology*, *11*. https://doi.org/10.1186/1471-2180-11-198

Martin, M., Barbeyron, T., Martin, R., Portetelle, D., Michel, G., & Vandenbol, M. (2015). The cultivable surface microbiota of the brown alga *Ascophyllum nodosum* is enriched in macroalgal-polysaccharide-degrading bacteria. *Frontiers in Microbiology*, *6*(12), 1–14.https://doi.org/10.3389/fmicb.2015.01487

Martins, R. F., Davids, W., Abu Al-Soud, W., Levander, F., Rådström, P., & Hatti-Kaul, R. (2001). Starch-hydrolyzing bacteria from Ethiopian soda lakes. *Extremophiles*, *5*(2), 135–144. https://doi.org/10.1007/s007920100183

Matagi, S. V. (2004). A biodiversity assessment of the Flamingo Lakes of eastern Africa. *Journal of Limnology*, *5*(1), 13–26. https://doi.org/10.1080/14888386.2004.9712715

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H. P., & Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics*, *14*(1), 60. https://doi.org/10.1186/1471-2105-14-60

Meier-Kolthoff, J. P., Carbasse, J. S., Peinado-Olarte, R. L., & Göker, M. (2022). TYGS and LPSN: A database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. *Nucleic Acids Research*, *50*(D1), D801–D807. https://doi.org/10.1093/nar/gkab902

Meier-Kolthoff, J. P., & Göker, M. (2019). TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-10210-3

Melack, J. M., & Kilham, P. (1974). Photosynthetic rates of phytoplankton in East African alkaline, saline lakes1. *Limnology and Oceanography*, *19*(5), 743–755. https://doi.org/10.4319/LO.1974.19.5.0743

Menasria, T., Monteoliva-Sánchez, M., Benammar, L., Benhadj, M., Ayachi, A., Hacène, H., Gonzalez-Paredes, A., & Aguilera, M. (2019). Culturable halophilic

bacteria inhabiting Algerian saline ecosystems: A source of promising features and potentialities. *World Journal of Microbiology and Biotechnology*, *35*(9), 132. https://doi.org/10.1007/s11274-019-2705-y

Menon, G., Mody, K., Keshri, J., & Jha, B. (2010). Isolation, purification, and characterization of haloalkaline xylanase from a marine *Bacillus pumilus* strain, GESF-1. *Biotechnology and Bioprocess Engineering*, *15*(6), 998–1005.

Mesbah, N. M., & Wiegel, J. (2017). A Halophilic, Alkalithermostable, Ionic Liquid-Tolerant Cellulase and Its Application in In Situ Saccharification of Rice Straw. *Bioenergy Research*, *10*(2), 583–591. https://doi.org/10.1007/s12155-017-9825-8

Mewis, K., Lenfant, N., Lombard, V., & Henrissat, B. (2016). Dividing the large glycoside hydrolase family 43 into subfamilies: A motivation for detailed enzyme characterization. *Applied and Environmental Microbiology*, *82*(6), 1686–1692. https://doi.org/10.1128/AEM.03453-15

Mikhodyuk, O. S., Gerasimenko, L. M., Akimov, V. N., Ivanovsky, R. N., & Zavarzin, G. A. (2008). Ecophysiology and polymorphism of the unicellular extremely natronophilic cyanobacterium *Euhalothece* sp. Z-M001 from Lake Magadi. *Microbiology*, *77*(6), 717–725. https://doi.org/10.1134/S0026261708060106

Miller, G. L. (1959). Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. *Analytical Chemistry*, *31*(3), 426–428. https://doi.org/10.1021/ac60147a030

Minegishi, H., Enomoto, S., Echigo, A., Shimane, Y., Kondo, Y., Inoma, A., Kamekura, M., Takai, K., Itoh, T., Ohkuma, M., Ihara, K., Takahashi-Ando, N., Fukushima, Y., Ishii, S., Yoshida, Y., & Usami, R. (2017). *Salinarchaeum chitinilyticum* sp. Nov., a chitin-degrading haloarchaeon isolated from commercial salt. *International Journal of Systematic and Evolutionary Microbiology*, *67*(7), 2274–2278. https://doi.org/10.1099/ijsem.0.001941

Møller, M. F., Kjeldsen, K. U., & Ingvorsen, K. (2010). *Marinimicrobium haloxylanilyticum* sp. nov., a new moderately halophilic, polysaccharide-degrading bacterium isolated from Great Salt Lake, Utah. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, *98*(4), 553–565. https://doi.org/10.1007/s10482-010-9472-y

Moore, L. R., Coe, A., Zinser, E. R., Saito, M. A., Sullivan, M. B., Lindell, D., Frois-Moniz, K., Waterbury, J., & Chisholm, S. W. (2007). Culturing the marine cyanobacterium Prochlorococcus. *Limnology and Oceanography: Methods*, *5*(10), 353–362.

Mori, T., Kamei, I., Hirai, H., & Kondo, R. (2014). Identification of novel glycosyl hydrolases with cellulolytic activity against crystalline cellulose from metagenomic libraries constructed from bacterial enrichment cultures. *SpringerPlus*, *3*(1), 1–7. https://doi.org/10.1186/2193-1801-3-365

Moriyama, R., Kudoh, S., Miyata, S., Nonobe, S., Hattori, A., & Makino, S. (1996). A germination-specific spore cortex-lytic enzyme from *Bacillus cereus* spores: Cloning and sequencing of the gene and molecular characterization of the enzyme. *Journal of Bacteriology*, *178*(17), 5330–5332. https://doi.org/10.1128/jb.178.17.5330-5332.1996

Morris, G. and Harding, S. (2009). Polysaccharides, Microbial. *Encyclopaedia of Micrbiology, 3*, 482–494.

Motlagh, A. M., Bhattacharjee, A. S., Coutinho, F. H., Dutilh, B. E., Casjens, S. R., & Goel, R. K. (2017). Insights of phage-host interaction in hypersaline ecosystem through metagenomics analyses. *Frontiers in Microbiology*, *8*(3), 1–18. https://doi.org/10.3389/FMICB.2017.00352/BIBTEX

Muazzam, A., Malik, B., Rashid, N., Irshad, S., & Fatima, M. (2019). Soluble expression of *Bacillus licheniformis* atcc 27811 α-amylase and charachterization of purified recombinant enzyme. *Journal of Animal and Plant Sciences*, *29*(1), 99–108.

Mukhtar, S., Abdulla Malik, K., & Mehnaz, S. (2018). Isolation and Characterization of Haloalkaliphilic Bacteria Isolated from the Rhizosphere of Dichanthium annulatum. *Journal of Advanced Research in Biotechnology*, *3*(1), 1–9. https://doi.org/10.15226/2475-4714/3/1/00133

Mukhtar, S., Zaheer, A., Aiysha, D., Malik, K. A., & Mehnaz, S. (2017). Actinomycetes : A source of industrially important enzymes. *Journal of Proteomics and Bioinformatics*, *10*(12), 316–319.

Mulango, E., Kasili, R., Mwirichia, R., Kambura, A., & Muhonja, C. (2020). Isolation and characterization of haloalkaliphilic bacteria from the hot springs of Lake Magadi. *African Journal of Microbiology Research*, *14*(7), 294–302. https://doi.org/10.5897/AJMR2018.8953

Murashima, K., Nishimura, T., Nakamura, Y., & Koga, J. (2002). Purification and characterization of new endo-1 , 4- NL - D -glucanases from Rhizopus oryzae. *Enzyme and Microbial Technology*, *30*(1), 319–326.

Muruga, B., & Anyango, B. (2013). A Survey of Extremophilic Bacteria in Lake Magadi, Kenya. *American Journal of Molecular and Cellular Biology*, *2*(1), 14–26.

Mwatha, W. E., & Grant, W. D. (2016). *Natronobacterium vacuolata* sp. nov., a Haloalkaliphilic Archaeon Isolated from Lake Magadi, Kenya. *International Journal of Systematic Bacteriology*, *43*(3), 401–404.

Mwirichia, R. (2022). Amplicon-Based Analysis of the Fungal Diversity across Four Kenyan Soda Lakes. *Scientifica*, *2022*. https://doi.org/10.1155/2022/9182034

Mwirichia, R., Muigai, A. W., Tindall, B., Boga, H. I., & Stackebrandt, E. (2010). Isolation and characterisation of bacteria from the haloalkaline Lake Elmenteita, Kenya. *Extremophiles*, *14*(4), 339–348. https://doi.org/10.1007/s00792-010-0311-x

Nacke, H., Engelhaupt, M., Brady, S., Fischer, C., Tautzt, J., & Daniel, R. (2012). Identification and characterization of novel cellulolytic and hemicellulolytic genes and enzymes derived from German grassland soil metagenomes. *Biotechnology Letters*, *34*(4), 663–675. https://doi.org/10.1007/s10529-011-0830-2

Naghoni, A., Emtiazi, G., Amoozegar, M. A., Cretoiu, M. S., Stal, L. J., Etemadifar, Z., Shahzadeh Fazeli, S. A., & Bolhuis, H. (2017). Microbial diversity in the hypersaline Lake Meyghan, Iran. *Scientific Reports*, *7*(1), 1–13. https://doi.org/10.1038/s41598-017-11585-3

Newton, R. J., Jones, S. E., Eiler, A., McMahon, K. D., & Bertilsson, S. (2011). A Guide to the Natural History of Freshwater Lake Bacteria. In *Microbiology and Molecular Biology Reviews* (Vol. 75, Issue 1). https://doi.org/10.1128/mmbr.00028-10

Nielsen, P., Fritze, D., & Priest, F. G. (1995). Phenetic diversity of alkaliphilic Bacillus strains: Proposal for nine new species. *Microbiology*, *141*(7), 1745–1761. https://doi.org/10.1099/13500872-141-7-1745

Nwodo, U. U., Green, E., & Okoh, A. I. (2012). Bacterial exopolysaccharides: Functionality and prospects. *International Journal of Molecular Sciences*, *13*(11), 14002–14015.

Nyakeri, E. M., Mwirichia, R., & Boga, H. (2018). Isolation and characterization of enzyme producing bacteria from Lake Magadi , an extreme soda lake in. *Journal of Microbiology and Experimentation*, *6*(2), 57–68.

Oduor, S. O., & Schagerl, M. (2007). Temporal trends of ion contents and nutrients in three Kenyan Rift Valley saline-alkaline lakes and their influence on phytoplankton biomass. *Hydrobiologia*, *584*(1), 59–68. https://doi.org/10.1007/s10750-007-0605-x

Ogonda, L. A., Saumonneau, A., Dion, M., Muge, E. K., Wamalwa, B. M., Mulaa, F. J., & Tellier, C. (2021). Characterization and engineering of two new GH9 and GH48 cellulases from a *Bacillus pumilus* isolated from Lake Bogoria. *Biotechnology Letters*, *43*(3), 691–700. https://doi.org/10.1007/s10529-020-03056-z

Okonechnikov, K., Conesa, A., & García-Alcalde, F. (2016). Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*, *32*(2), 292–294. https://doi.org/10.1093/bioinformatics/btv566

Oksanen, J. (2015). Vegan: an introduction to ordination. *Management*, *1*, 1–10. http://doi.acm.org/10.1145/2037556.2037605%5Cnftp://ftp3.ie.freebsd.org/pub/cran.r-project.org/web/packages/vegan/vignettes/intro-vegan.pdf

Ondov, B. D., Bergman, N. H., & Phillippy, A. M. (2011). Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics*, *12*(1), 385. https://doi.org/10.1186/1471-2105-12-385

Oren, A. (2002). Diversity of halophilic microorganisms: Environments, phylogeny, physiology, and applications. *Journal of Industrial Microbiology and*

Biotechnology, *28*(1), 56–63. https://doi.org/10.1038/sj/jim/7000176

Oren, Aharon. (2010). Industrial and environmental applications of halophilic microorganisms. *Environmental Technology*, *31*(8–9), 825–834. https://doi.org/10.1080/09593330903370026

Orwa, P., Mugambi, G., Wekesa, V., & Mwirichia, R. (2020). Isolation of haloalkaliphilic fungi from Lake Magadi in Kenya. *Heliyon*, *6*(1). https://doi.org/10.1016/j.heliyon.2019.e02823

Otasek, D., Morris, J. H., Bouças, J., Pico, A. R., & Demchak, B. (2019). Cytoscape Automation: Empowering workflow-based network analysis. *Genome Biology*, *20*(1), 1–15. https://doi.org/10.1186/s13059-019-1758-4

Pabbathi, N. P. P., Velidandi, A., Tavarna, T., Gupta, S., Raj, R. S., Gandam, P. K., & Baadhe, R. R. (2021). Role of metagenomics in prospecting novel endoglucanases, accentuating functional metagenomics approach in second-generation biofuel production: a review. *Biomass Conversion and Biorefinery*. https://doi.org/10.1007/s13399-020-01186-y

Padan, E., Bibi, E., Ito, M., & Krulwich, T. A. (2005). Alkaline pH homeostasis in bacteria: New insights. *Biochimica et Biophysica Acta - Biomembranes*, *1717*(2), 67–88. https://doi.org/10.1016/j.bbamem.2005.09.010

Pandey, S., & Singh, S. P. (2012). Organic solvent tolerance of an α-Amylase from haloalkaliphilic bacteria as a function of pH, temperature, and salt concentrations. In *Applied Biochemistry and Biotechnology* (Vol. 166, Issue 7, pp. 1747–1757). https://doi.org/10.1007/s12010-012-9580-4

Pang, J., Wang, J., Liu, Z., Zhang, Q., & Qi, Q. (2019). Identification and characterization of an Endo-glucanase secreted from cellulolytic *Escherichia coli* ZH-4. *BMC Biotechnology*, *19*(1), 1–9. https://doi.org/10.1186/s12896-019-0556-0

Park, Y. J., Jeong, Y. U., & Kong, W. S. (2018). Genome sequencing and carbohydrate-active enzyme (CAZyme) repertoire of the white rot fungus *Flammulina elastica*. *International Journal of Molecular Sciences*, *19*(8), 2379. https://doi.org/10.3390/ijms19082379

Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P. A., & Hugenholtz, P. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature Biotechnology*, *36*(10), 996. https://doi.org/10.1038/nbt.4229

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, *25*(7), 1043–1055. https://doi.org/10.1101/gr.186072.114

Patel, M., Patel, H. M., Vohra, N., & Dave, S. (2020). Complete genome sequencing and comparative genome characterization of the lignocellulosic biomass degrading bacterium *Pseudomonas stutzeri* MP4687 from cattle rumen. *Biotechnology Reports*, *28*, e00530. https://doi.org/10.1016/j.btre.2020.e00530

Patel, S., & Gupta, R. S. (2020). A phylogenomic and comparative genomic framework for resolving the polyphyly of the genus *Bacillus*: Proposal for six new genera of bacillus species, *Peribacillus* gen. nov., cytobacillus gen. nov., *Mesobacillus* gen. nov., *Neobacillus* gen. nov., metabacillu. *International Journal of Systematic and Evolutionary Microbiology*, *70*(1), 406–438. https://doi.org/10.1099/ijsem.0.003775

Pathak, A. P., & Jadhav, S. R. (2019). Production and characterization of a thermo-ph stable pectinase from *Bacillus licheniformis* unp-1: A novel strain isolated from unapdev hot spring. *Indian Journal of Geo-Marine Sciences*, *48*(5), 670–677.

Paul, B., Kavia Raj, K., Murali, T. S., & Satyamoorthy, K. (2020). Species-specific genomic sequences for classification of bacteria. *Computers in Biology and Medicine*, *123*, 103874. https://doi.org/10.1016/J.COMPBIOMED.2020.103874

Paul, D., Kumbhare, S. V., Mhatre, S. S., Chowdhury, S. P., Shetty, S. A., Marathe, N. P., Bhute, S., & Shouche, Y. S. (2016). Exploration of microbial diversity and community structure of Lonar Lake: The only hypersaline meteorite crater lake within basalt rock. *Frontiers in Microbiology*, *6*(1), 1–12. https://doi.org/10.3389/fmicb.2015.01553

Pawaskar, G. M., Raval, K., Rohit, P., Shenoy, R. P., & Raval, R. (2021). Cloning, expression, purification and characterization of chitin deacetylase extremozyme from halophilic *Bacillus aryabhattai* B8W22. *3 Biotech*, *11*(12), 1–13. https://doi.org/10.1007/s13205-021-03073-3

Pérez-Pantoja, D., Donoso, R., Agulló, L., Córdova, M., Seeger, M., Pieper, D. H., & González, B. (2012). Genomic analysis of the potential for aromatic compounds biodegradation in Burkholderiales. *Environmental Microbiology*, *14*(5), 1091–1117. https://doi.org/10.1111/j.1462-2920.2011.02613.x

Pérez, E., & Chebude, Y. (2017). Chemical Analysis of Gaet'ale, a Hypersaline Pond in Danakil Depression (Ethiopia): New Record for the Most Saline Water Body on Earth. *Aquatic Geochemistry*, *23*(2), 109–117. https://doi.org/10.1007/s10498-017-9312-z

Pham, V. H. T., & Kim, J. (2012). Cultivation of unculturable soil bacteria. *Trends in Biotechnology*, *30*(9), 475–484.

Podar, M., & Reysenbach, A. L. (2006). New opportunities revealed by biotechnological explorations of extremophiles. *Current Opinion in Biotechnology*, *17*(3), 250–255. https://doi.org/10.1016/j.copbio.2006.05.002

Poli, A., Nicolaus, B., Denizci, A. A., Yavuzturk, B., & Kazan, D. (2013). *Halomonas smyrnensis* sp. nov., a moderately halophilic, exopolysaccharide-producing bacterium. *International Journal of Systematic and Evolutionary Microbiology*, *63*(1), 10–18. https://doi.org/10.1099/ijs.0.037036-0

Pollegioni, L., Tonin, F., & Rosini, E. (2015). Lignin-degrading enzymes. *The FEBS Journal*, *282*(7), 1190–1213. https://doi.org/10.1111/FEBS.13224

Potprommanee, L., Wang, X. Q., Han, Y. J., Nyobe, D., Peng, Y. P., Huang, Q., Liu, J. Y., Liao, Y. L., & Chang, K. L. (2017). Characterization of a thermophilic cellulase from *Geobacillus* sp. HTA426, an efficient cellulase-producer on alkali pretreated of lignocellulosic biomass. *PLoS ONE*, *12*(4), 1–16. https://doi.org/10.1371/journal.pone.0175004

Poyraz, N., & Mutlu, M. (2020). Characterization of Microbial Populations of Lake Van By 16S Metagenomics Study. *Eskişehir Teknik Üniversitesi Bilim ve Teknoloji Dergisi - C Yaşam Bilimleri Ve Biyoteknoloji*, *9*(1), 80–88. https://doi.org/10.18036/estubtdc.680795

Prakash, B., Vidyasagar, M., Madhukumar, M. S., Muralikrishna, G., & Sreeramulu, K. (2009). Production, purification, and characterization of two extremely halotolerant, thermostable, and alkali-stable α-amylases from *Chromohalobacter* sp. TVSP 101. *Process Biochemistry*, *44*(2), 210–215. https://doi.org/10.1016/J.PROCBIO.2008.10.013

Priya, B. S., Stalin, T., & Selvam, K. (2012). Efficient utilization of xylanase and lipase producing thermophilic marine actinomycetes (*Streptomyces albus* and *Streptomyces hygroscopicus*) in the production of ecofriendly alternative energy from waste. *African Journal of Biotechnology*, *11*(78), 14320–14325.

Qin, Q. L., Xie, B. Bin, Zhang, X. Y., Chen, X. L., Zhou, B. C., Zhou, J., Oren, A., & Zhang, Y. Z. (2014). A proposed genus boundary for the prokaryotes based on genomic insights. *Journal of Bacteriology*, *196*(12), 2210–2215. https://doi.org/10.1128/JB.01688-14

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2012). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, *41*(D1), 590–596. https://doi.org/10.1093/nar/gks1219

Raddadi, N., Cherif, A., Daffonchio, D., Neifar, M., & Fava, F. (2015). Biotechnological applications of extremophiles, extremozymes and extremolytes. *Applied Microbiology and Biotechnology*, *99*(19), 7907–7913. https://doi.org/10.1007/s00253-015-6874-9

Rafael, L. e O., Camila, B. A. B., Sergio, D. J., Patricia, M. A., Norma, B. de G., Edmar, V. de A., & Leonor, A. de O. da S. (2016). Production and characterization of endoglucanase secreted by *Streptomyces capoamus* isolated from Caatinga. *African Journal of Biotechnology*, *15*(42), 2394–2401.

Rampelotto, P. H. (2013). Extremophiles and extreme environments. *Life*, *3*(3), 482–485. https://doi.org/10.3390/life3030482

Rattu, G., Joshi, S., & Satyanarayana, T. (2016). Bifunctional recombinant cellulase–xylanase (rBhcell-xyl) from the polyextremophilic bacterium *Bacillus halodurans* TSLV1 and its utility in valorization of renewable agro-residues. *Extremophiles*, *20*(6), 831–842. https://doi.org/10.1007/s00792-016-0870-6

Ravindran, R., & Jaiswal, A. (2016). Microbial Enzyme Production Using Lignocellulosic Food Industry Wastes as Feedstock: A Review. *Bioengineering*,

3(4), 30.

Rees, H. C., Grant, W. D., Jones, B. E., & Heaphy, S. (2004). Diversity of Kenyan soda lake alkaliphiles assessed by molecular methods. *Extremophiles*, *8*(1), 63–71. https://doi.org/10.1007/S00792-003-0361-4/FIGURES/2

Reina-Bueno, M., Argandoña, M., Salvador, M., Rodríguez-Moya, J., Iglesias-Guerra, F., Csonka, L. N., Nieto, J. J., & Vargas, C. (2012). Role of trehalose in salinity and temperature tolerance in the model halophilic bacterium chromohalobacter salexigens. *PLoS ONE*, *7*(3). https://doi.org/10.1371/journal.pone.0033587

Richter, M., & Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(45), 19126–19131. https://doi.org/10.1073/pnas.0906412106

Robert, X., & Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research*, *42*(W1), 320–324. https://doi.org/10.1093/nar/gku316

Roberts, M. F. (2005). Organic compatible solutes of halotolerant and halophilic microorganisms. *Saline Systems*, *1*(1), 1–30. https://doi.org/10.1186/1746-1448-1-5

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: A versatile open source tool for metagenomics. *PeerJ*, *2016*(10), 1–22. https://doi.org/10.7717/peerj.2584

Rohban, R., Amoozegar, M. A., & Ventosa, A. (2009). Screening and isolation of halophilic bacteria producing extracellular hydrolyses from Howz Soltan Lake, Iran. *Journal of Industrial Microbiology and Biotechnology*, *36*(3), 333–340. https://doi.org/10.1007/s10295-008-0500-0

Romano, N., Gioffre, A., Sede, S. M., Campos, E., Cataldi, A., & Talia, P. (2013). Characterization of cellulolytic activities of environmental bacterial consortia from an argentinian native forest. *Current Microbiology*, *67*(2), 138–147.

Ronoh, R. C., Budambula, N. L. M., Mwirichia, R. K., & Boga, H. I. (2013). Isolation and characterization of actinobacteria from Lake Magadi, Kenya. *African Journal of Microbiology Research*, *7*(32), 4200–4206.

Rosenberg, E., De Long, E. F., Lory, S., Stackebrandt, E., & Thompson, F. (2014). The prokaryotes: Other major lineages of bacteria and the archaea. *The Prokaryotes: Other Major Lineages of Bacteria and The Archaea*, *43*, 1–1028. https://doi.org/10.1007/978-3-642-38954-2

Ruga, B. N. M. U. (2010). The diversity of extremophilic bacteria in Lake Magadi, Kenya. In *University of Nairobi, MSc. Thesis*.

Ruginescu, R., Gomoiu, I., Popescu, O., Cojoc, R., Neagu, S., Lucaci, I., Batrinescu-Moteau, C., & Enache, M. (2018). Bioprospecting for novel halophilic and halotolerant sources of hydrolytic enzymes in brackish, saline and hypersaline

lakes of romania. *Journal of Environmental Protection*, *8*(12), 1066–1081. https://doi.org/10.3390/microorganisms8121903

Ruginescu, R., Gomoiu, I., Popescu, O., Cojoc, R., Neagu, S., Lucaci, I., Batrinescu-Moteau, C., & Enache, M. (2020). Bioprospecting for novel halophilic and halotolerant sources of hydrolytic enzymes in brackish, saline and hypersaline lakes of romania. *Microorganisms*, *8*(12), 1–16. https://doi.org/10.3390/microorganisms8121903

Sadhu, S. and M. T. K. (2013). Cellulase Production by Bacteria: A Review. *British Microbiology Research Journal*, *3*(3), 235–258. https://doi.org/10.9734/bmrj/2013/2367

Şafak, H., Otur, Ç., & Kurt-Kızıldoğan, A. (2020). Molecular and biochemical characterization of a recombinant endoglucanase rCKT3eng, from an extreme halophilic *Haloarcula* sp. strain CKT3. *International Journal of Biological Macromolecules*, *151*, 1173–1180. https://doi.org/10.1016/j.ijbiomac.2019.10.161

Sahay, H., Mahfooz, S., Singh, A. K., Singh, S., Kaushik, R., Saxena, A. K., & Arora, D. K. (2012). Exploration and characterization of agriculturally and industrially important haloalkaliphilic bacteria from environmental samples of hypersaline Sambhar lake, India. *World Journal of Microbiology and Biotechnology*, *28*(11), 3207–3217. https://doi.org/10.1007/s11274-012-1131-1

Salano, O. A, Makonde, H. M., Kasili, R. W., Wangai, L. N., Nawiri, M. P., & Boga, H. I. (2017). African Journal of Microbiology Research Diversity and distribution of fungal communities within the hot springs of soda lakes in the Kenyan rift valley. *African Journal of Microbiology Research*, *11*(19), 764–775. https://doi.org/10.5897/AJMR2017.8510

Salano, Odilia Atamba, Makonde, M., Kasili, R. W., Wangai, L. N., Nawiri, M. P., & Boga, H. I. (2017). Diversity and distribution of fungal communities within the hot springs of soda lakes in the Kenyan rift valley. *Academicjournals.Org*, *11*(19), 764–775. https://doi.org/10.5897/AJMR2017.8510

Sánchez-Porro, C., Martín, S., Mellado, E., & Ventosa, A. (2003). Diversity of moderately halophilic bacteria producing extracellular hydrolytic enzymes. *Journal of Applied Microbiology*, *94*(2), 295–300. https://doi.org/10.1046/j.1365-2672.2003.01834.x

Sanjaya, R. E., Putri, K. D. A., Kurniati, A., Rohman, A., & Puspaningsih, N. N. T. (2021). In silico characterization of the GH5-cellulase family from uncultured microorganisms: physicochemical and structural studies. *Journal of Genetic Engineering and Biotechnology*, *19*(1), 143. https://doi.org/10.1186/s43141-021-00236-w

Santos, H., & Da Costa, M. S. (2002). Compatible solutes of organisms that live in hot saline environments. *Environmental Microbiology*, *4*(9), 501–509.

Sar, A., Pal, S., Islam, S., Mukherjee, P., & Dam, B. (2021). An Alkali-Halostable Endoglucanase Produced Constitutively by a Bacterium Isolated from Sambhar

Lake in India with Biotechnological Potential. *Proceedings of the National Academy of Sciences India Section B - Biological Sciences*, *91*(2), 319–326. https://doi.org/10.1007/s40011-021-01230-5

Satari, B., Karimi, K., & Kumar, R. (2019). Cellulose solvent-based pretreatment for enhanced second-generation biofuel production: A review. In *Sustainable Energy and Fuels* (Vol. 3, Issue 1). Royal Society of Chemistry. https://doi.org/10.1039/c8se00287h

Satyapal, G. K., Mishra, S. K., Srivastava, A., Ranjan, R. K., Prakash, K., Haque, R., & Kumar, N. (2018). Possible bioremediation of arsenic toxicity by isolating indigenous bacteria from the middle Gangetic plain of Bihar, India. *Biotechnology Reports*, *17*, 117–125. https://doi.org/10.1016/j.btre.2018.02.002

Saxena, R., Dhakan, D. B., Mittal, P., Waiker, P., Chowdhury, A., Ghatak, A., & Sharma, V. K. (2017). Metagenomic analysis of hot springs in central india reveals hydrocarbon degrading thermophiles and pathways essential for survival in extreme environments. *Frontiers in Microbiology*, *7*(2017), 2123. https://doi.org/10.3389/fmicb.2016.02123

Schagerl, M. (2016). Soda lakes of East Africa. In *Soda Lakes of East Africa* (pp. 77–93). https://doi.org/10.1007/978-3-319-28622-8

Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*, *30*(14), 2068–2069. https://doi.org/10.1093/bioinformatics/btu153

Seo, D. J., Fujita, H., & Sakoda, A. (2011). Effects of a non-ionic surfactant, Tween 20, on adsorption/desorption of saccharification enzymes onto/from lignocelluloses and saccharification rate. *Adsorption*, *17*(5), 813–822. https://doi.org/10.1007/s10450-011-9340-8

Shafiei, M., Ziaee, A. A., & Amoozegar, M. A. (2011). Purification and characterization of an organic-solvent-tolerant halophilic α-amylase from the moderately halophilic *Nesterenkonia* sp. strain F. *Journal of Industrial Microbiology and Biotechnology*, *38*(2), 275–281. https://doi.org/10.1007/s10295-010-0770-1

Shapovalova, A. A., Khijniak, T. V., Tourova, T. P., Muyzer, G., & Sorokin, D. Y. (2008). Heterotrophic denitrification at extremely high salt and pH by haloalkaliphilic Gammaproteobacteria from hypersaline soda lakes. *Extremophiles*, *12*(5), 619–625. https://doi.org/10.1007/s00792-008-0166-6

Sharma, R., Prakash, O., Sonawane, M. S., Nimonkar, Y., Golellu, P. B., & Sharma, R. (2016). Diversity and distribution of phenol oxidase producing fungi from soda lake and description of *Curvularia lonarensis* sp. nov. *Frontiers in Microbiology*, *7*(11), 1–19. https://doi.org/10.3389/FMICB.2016.01847/BIBTEX

Shweta, A. (2014). Cellulases of Bacterial Origin and their Applications : A Review. *International Journal of Science and Research*, *3*(10), 1652–1655.

Sidar, A., Albuquerque, E. D., Voshol, G. P., Ram, A. F. J., Vijgenboom, E., & Punt, P. J. (2020). Carbohydrate Binding Modules: Diversity of Domain Architecture

in Amylases and Cellulases From Filamentous Microorganisms. *Frontiers in Bioengineering and Biotechnology*, *8*(7), 1–15. https://doi.org/10.3389/fbioe.2020.00871

Simachew, A., Lanzén, A., Gessesse, A., & Øvreås, L. (2016a). Prokaryotic Community Diversity Along an Increasing Salt Gradient in a Soda Ash Concentration Pond. *Microbial Ecology*, *71*(2), 326–338. https://doi.org/10.1007/s00248-015-0675-7

Simachew, A., Lanzén, A., Gessesse, A., & Øvreås, L. (2016b). Prokaryotic Community Diversity Along an Increasing Salt Gradient in a Soda Ash Concentration Pond. *Microbial Ecology*, *71*(2), 326–338. https://doi.org/10.1007/s00248-015-0675-7

Simon, C., & Daniel, R. (2011). Metagenomic analyses: Past and future trends. *Applied and Environmental Microbiology*, *77*(4), 1153–1161. https://doi.org/10.1128/AEM.02345-10

Singh, G., Patel, A. K., Gupta, A., Gupta, D., & Mishra, V. K. (2019). Current Advancements in Recombinant Technology for Industrial Production of Cellulases: Part-II. *Approaches to Enhance Industrial Production of Fungal Cellulases*, 177–201. https://doi.org/10.1007/978-3-030-14726-6_11

Sinha, R., & Khare, S. K. (2014). Protective role of salt in catalysis and maintaining structure of halophilic proteins against denaturation. *Frontiers in Microbiology*, *5*(165), 1–6. https://doi.org/10.3389/fmicb.2014.00165

Siroosi, M., Amoozegar, M. A., & Khajeh, K. (2016). Purification and characterization of an alkaline chloride-tolerant laccase from a halotolerant bacterium, *Bacillus* sp. strain WT. *Journal of Molecular Catalysis B: Enzymatic*, *134*, 89–97. https://doi.org/10.1016/j.molcatb.2016.10.001

Solat, N., & Shafiei, M. (2021). A novel pH and thermo-tolerant halophilic alpha-amylase from moderate halophile *Nesterenkonia* sp. strain F: gene analysis, molecular cloning, heterologous expression and biochemical characterization. *Archives of Microbiology*, *203*(6), 3641–3655. https://doi.org/10.1007/s00203-021-02359-7

Somogyi, B., Felföldi, T., Dinka, M., & Vörös, L. (2010). Periodic picophytoplankton predominance in a large, shallow alkaline lake (Lake Ferto″, Neusiedlersee). *Annales de Limnologie*, *46*(1), 9–19. https://doi.org/10.1051/limn/2010001

Sorokin, D. Y., & Kolganova, T. V. (2014). Bacterial chitin utilization at halophilic conditions. *Extremophiles*, *18*(2), 243–248. https://doi.org/10.1007/s00792-013-0611-z

Sorokin, D. Y., Panteleeva, A. N., Tourova, T. P., Kaparullina, E. N., & Muyzer, G. (2011). *Natronoflexus pectinivorans* gen. nov. sp. nov., an obligately anaerobic and alkaliphilic fermentative member of Bacteroidetes from soda lakes. *Extremophiles*, *15*(6), 691–696.

Sorokin, D. Y., Tourova, T. P., Panteleeva, A. N., Kaparullina, E. N., & Muyzer, G. (2012). Anaerobic utilization of pectinous substrates at extremely haloalkaline

conditions by *Natranaerovirga pectinivora* gen. nov., sp. nov., and *Natranaerovirga hydrolytica* sp. nov., isolated from hypersaline soda lakes. *Extremophiles*, *16*(2), 307–315. https://doi.org/10.1007/s00792-012-0431-6

Sorokin, D. Y., Tourova, T. P., Sukhacheva, M. V., Mardanov, A. V., & Ravin, N. V. (2012). Bacterial chitin utilisation at extremely haloalkaline conditions. *Extremophiles*, *16*(6), 883–894. https://doi.org/10.1007/s00792-012-0484-6

Sorokin, Dimitry Y., Banciu, H. L., & Muyzer, G. (2015). Functional microbiology of soda lakes. *Current Opinion in Microbiology*, *25*, 88–96. https://doi.org/10.1016/j.mib.2015.05.004

Sorokin, Dimitry Y., Berben, T., Melton, E. D., Overmars, L., Vavourakis, C. D., & Muyzer, G. (2014). Microbial diversity and biogeochemical cycling in soda lakes. *Extremophiles*, *18*(5), 791–809.

Sorokin, Dimitry Y., Gumerov, V. M., Rakitin, A. L., Beletsky, A. V., Damsté, J. S. S., Muyzer, G., Mardanov, A. V., & Ravin, N. V. (2014). Genome analysis of *Chitinivibrio alkaliphilus* gen. nov., sp. nov., a novel extremely haloalkaliphilic anaerobic chitinolytic bacterium from the candidate phylum Termite Group 3. *Environmental Microbiology*, *16*(6), 1549–1565. https://doi.org/10.1111/1462-2920.12284

Sorokin, Dimitry Yu, & Kuenen, J. G. (2005). Chemolithotrophic haloalkaliphiles from soda lakes. *FEMS Microbiology Ecology*, *52*(3), 287–295. https://doi.org/10.1016/j.femsec.2005.02.012

Sorokin, Dimitry Yu, Trotsenko, Y. A., Doronina, N. V., Tourova, T. P., Galinski, E. A., Kolganova, T. V., & Muyzer, G. (2007). *Methylohalomonas lacus* gen. nov., sp. nov. and *Methylonatrum kenyense* gen. nov., sp. nov., methylotrophic gammaproteobacteria from hypersaline lakes. *International Journal of Systematic and Evolutionary Microbiology*, *57*(12), 2762–2769. https://doi.org/10.1099/ijs.0.64955-0

Soy, S., Nigam, V. K., & Sharma, S. R. (2019). Cellulolytic, amylolytic and xylanolytic potential of thermophilic isolates of Surajkund hot spring. *Journal of Biosciences*, *44*(5). https://doi.org/10.1007/s12038-019-9938-7

Srinivas, A., Divyasree, B., Tushar, L., Suresh, G., Sasikala, C., & Ramana, C. V. (2016). *Salinicoccus amylolyticus* sp. nov., isolated from a saltern. *International Journal of Systematic and Evolutionary Microbiology*, *66*(10), 3814–3820. https://doi.org/10.1099/IJSEM.0.001270/CITE/REFWORKS

Srinivas, A., Rahul, K., Sasikala, C., Subhash, Y., Ramaprasad, E. V. V., & Ramana, C. V. (2012). *Georgenia satyanarayanai* sp. nov., an alkaliphilic and thermotolerant amylase-producing actinobacterium isolated from a soda lake. *International Journal of Systematic and Evolutionary Microbiology*, *62*(10), 2405–2409.

Stewart, E. J. (2012). Growing unculturable bacteria. *Journal of Bacteriology*, *194*(16), 4151–4160.

Streit, W. R., & Daniel, R. (2010). Metagenomics: Methods and Protocols. In

Methods in Molecular Biology (Vol. 668, Issue 3). https://doi.org/10.1007/978-1-60761-823-2

Sultanpuram, V. R., & Mothe, T. (2016). Salipaludibacillus aurantiacus gen. Nov., sp. nov. a novel alkali tolerant bacterium, reclassification of *Bacillus agaradhaerens* as *Salipaludibacillus agaradhaerens* comb. nov. and *Bacillus neizhouensis* as *Salipaludibacillus neizhouensis* comb. nov. *International Journal of Systematic and Evolutionary Microbiology*, 66(7), 2747–2753. https://doi.org/10.1099/ijsem.0.001117

Sun, Z, Liu, H, Wang, X, Yang, F and Li, X. (2019). Proteomic Analysis of the Xanthan-Degrading Pathway of Microbacterium sp. XT11. *ACS Omega*, *4*, 19096−19105. https://doi.org/10.1021/acsomega.9b02313

Sun, S., Sun, S., Cao, X., & Sun, R. (2016). The role of pretreatment in improving the enzymatic hydrolysis of lignocellulosic materials. *Bioresource Technology*, *199*, 49–58. https://doi.org/10.1016/j.biortech.2015.08.061

Surakasi, V. P., Antony, C. P., Sharma, S., Patole, M. S., & Shouche, Y. S. (2010). Temporal bacterial diversity and detection of putative methanotrophs in surface mats of Lonar crater lake. *Journal of Basic Microbiology*, *50*(5), 465–474. https://doi.org/10.1002/jobm.201000001

Suryanto, D. W. I., Patonah, S. I. T. I., & Munir, E. (2010). Control of Fusarium Wilt of Chili With Chitinolytic Bacteria. *HAYATI Journal of Biosciences*, *17*(1), 5–8. https://doi.org/10.4308/hjb.17.1.5

Sützl, L., Laurent, C. V. F. P., Abrera, A. T., Schütz, G., Ludwig, R., & Haltrich, D. (2018). Multiplicity of enzymatic functions in the CAZy AA3 family. *Applied Microbiology and Biotechnology*, *102*(6), 2477–2492. https://doi.org/10.1007/s00253-018-8784-0

Suurnäkki, A., Tenkanen, M., Siika-Aho, M., Niku-Paavola, M. L., Viikari, L., & Buchert, J. (2000). *Trichoderma reesei* cellulases and their core domains in the hydrolysis and modification of chemical pulp. *Cellulose*, *7*(2), 189–209. https://doi.org/10.1023/A:1009280109519

Taher, A. G. (1999). Inland saline lakes of Wadi El Natrun depression, Egypt. *International Journal of Salt Lake Research*, *8*(2), 149–170. https://doi.org/10.1023/A:1009027813903

Tamura, K. (1992). Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. *Molecular Biology and Evolution*, *9*(4), 678–687. https://doi.org/10.1093/oxfordjournals.molbev.a040752

Tandon, K., Baatar, B., Chiang, P. W., Dashdondog, N., Oyuntsetseg, B., & Tang, S. L. (2020). A large-scale survey of the bacterial communities in lakes of western mongolia with varying salinity regimes. *Microorganisms*, *8*(11), 1–16. https://doi.org/10.3390/microorganisms8111729

Tao, X., Guo, F., Zhou, Q., Hu, F., Xiang, H., Xiao, G. G., & Shang, D. (2019). Bacterial community mapping of the intestinal tract in acute pancreatitis rats

based on 16S rDNA gene sequence analysis. *RSC Advances*, *9*(9), 5025–5036. https://doi.org/10.1039/c8ra09547g

Tatusova, T., Dicuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E. P., Zaslavsky, L., Lomsadze, A., Pruitt, K. D., Borodovsky, M., & Ostell, J. (2016). NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Research*, *44*(14), 6614–6624. https://doi.org/10.1093/nar/gkw569

Tayi, L., Maku, R. V., Patel, H. K., & Sonti, R. V. (2016). Identification of pectin degrading enzymes secreted by *Xanthomonas oryzae* pv. Oryzae and determination of their role in virulence on rice. *PLoS ONE*, *11*(12), 1–15. https://doi.org/10.1371/journal.pone.0166396

Tazi, L., Breakwell, D. P., Harker, A. R., & Crandall, K. A. (2014). Life in extreme environments: Microbial diversity in Great Salt Lake, Utah. *Extremophiles*, *18*(3), 525–535. https://doi.org/10.1007/s00792-014-0637-x

Thirumala, M., & Reddy, S. V. (2017). Isolation of a cellulolytic bacterium from the Lonar Soda Lake and genomic analysis of it. *Journal of Pure and Applied Microbiology*, *11*(2), 1105–1112. https://doi.org/10.22207/JPAM.11.2.54

Thomas, F., Hehemann, J. H., Rebuffet, E., Czjzek, M., & Michel, G. (2011). Environmental and gut Bacteroidetes: The food connection. *Frontiers in Microbiology*, *2*(MAY), 93. https://doi.org/10.3389/fmicb.2011.00093

Tindall, B. J., Ross, H. N. M., & Grant, W. D. (1984). *Natronobacterium* gen. nov. and *Natronococcus* gen. nov., Two New Genera of Haloalkaliphilic Archaebacteria. *Systematic and Applied Microbiology*, *5*(1), 41–57. https://doi.org/10.1016/S0723-2020(84)80050-8

Tindall, Brian J, Mills, A. A., & Grant, W. D. (1980). An alkalophilic red halophilic bacterium with a low magnesium requirement from a Kenyan soda lake. *Microbiology*, *116*(1), 257–260.

Tourova, T. P., Grechnikova, M. A., Kuznetsov, B. B., & Sorokin, D. Y. (2014). Phylogenetic diversity of bacteria in soda lake stratified sediments. *Microbiology*, *83*(6), 869–879.

Travers, R. S., Martin, P. A. W., Charles, F., & Russell, S. (1987). Selective Process for Efficient Isolation of Soil *Bacillus spp*. 53(6), 1263–1266.

Tutuncu, H. E., Balci, N., Tuter, M., & Karaguler, N. G. (2019). Recombinant production and characterization of a novel esterase from a hypersaline lake, Acıgöl, by metagenomic approach. *Extremophiles*, *23*(5), 507–520. https://doi.org/10.1007/s00792-019-01103-w

Ulrih, N. P., Gmajner, D., & Raspor, P. (2009). Structural and physicochemical properties of polar lipids from thermophilic archaea. *Applied Microbiology and Biotechnology*, *84*(2), 249–260.

Urbieta, M. S., Donati, E. R., Chan, K. G., Shahar, S., Sin, L. L., & Goh, K. M. (2015). Thermophiles in the genomic era: Biodiversity, science, and applications. *Biotechnology Advances*, *33*(6), 633–647.

https://doi.org/10.1016/J.BIOTECHADV.2015.04.007

Uzyol, K. S., Sariyar Akbulut, B., Denizci, A. A., & Kazan, D. (2012). Thermostable α-amylase from moderately halophilic *Halomonas* sp. AAD21. *Turkish Journal of Biology*, *36*(3), 327–338. https://doi.org/10.3906/biy-1106-7

Vadala, B. S., Deshpande, S., & Apte-Deshpande, A. (2021). Soluble expression of recombinant active cellulase in E.coli using *B. subtilis* (natto strain) cellulase gene. *Journal of Genetic Engineering and Biotechnology*, *19*(7), 1–7. https://doi.org/10.1186/s43141-020-00103-0

Vaijayanthi, G., & Vijayakumar, Ramasamy, D. D. (2016). Actinobacteria — A biofactory of novel enzymes. *Intech*, 330–352.

Vamsi Bharadwaj, S. V., Shrivastav, A., Dubey, S., Ghosh, T., Paliwal, C., Maurya, R., & Mishra, S. (2015). Draft genome sequence of *Halomonas hydrothermalis* MTCC 5445, isolated from the west coast of India. *Genome Announcements*, *3*(1), 3–4. https://doi.org/10.1128/genomeA.01419-14

Van Zyl, L. J., Nemavhulani, S., Cass, J., Cowan, D. A., & Trindade, M. (2016). Three novel bacteriophages isolated from the East African Rift Valley soda lakes. *Virology Journal*, *13*(1), 1–14. https://doi.org/10.1186/S12985-016-0656-6

Vargas, V. A., Delgado, O. D., Hatti-Kaul, R., & Mattiasson, B. (2004). Lipase-producing microorganisms from a Kenyan alkaline soda lake. *Biotechnology Letters*, *26*(2), 81–86. https://doi.org/10.1023/B:BILE.0000012898.50608.12

Vargas, V. A., Delgado, O. D., Hatti-Kaul, R., & Mattiasson, B. (2005). *Bacillus bogoriensis* sp, nov., a novel alkaliphilic, halotolerant bacterium isolated from a Kenyan soda lake. *International Journal of Systematic and Evolutionary Microbiology*, *55*(2), 899–902. https://doi.org/10.1099/IJS.0.63318-0/CITE/REFWORKS

Vartoukian, S. R., Palmer, R. M., & Wade, W. G. (2010). Strategies for culture of ‘unculturable’ bacteria. *FEMS Microbiology Letters*, *309*(1), 1–7. https://doi.org/10.1111/J.1574-6968.2010.02000.X

Vavourakis, C. D., Andrei, A. S., Mehrshad, M., Ghai, R., Sorokin, D. Y., & Muyzer, G. (2018a). A metagenomics roadmap to the uncultured genome diversity in hypersaline soda lake sediments. *Microbiome*, *6*(1), 1–18. https://doi.org/10.1186/s40168-018-0548-7

Vavourakis, C. D., Andrei, A. S., Mehrshad, M., Ghai, R., Sorokin, D. Y., & Muyzer, G. (2018b). A metagenomics roadmap to the uncultured genome diversity in hypersaline soda lake sediments 06 Biological Sciences 0605 Microbiology 06 Biological Sciences 0604 Genetics. *Microbiome*, *6*(1), 1–18. https://doi.org/10.1186/s40168-018-0548-7

Vavourakis, C. D., Ghai, R., Rodriguez-Valera, F., Sorokin, D. Y., Tringe, S. G., Hugenholtz, P., & Muyzer, G. (2016a). Metagenomic insights into the uncultured diversity and physiology of microbes in four hypersaline soda lake brines. *Frontiers in Microbiology*, *7*(FEB), 1–18.

https://doi.org/10.3389/fmicb.2016.00211

Vavourakis, C. D., Ghai, R., Rodriguez-Valera, F., Sorokin, D. Y., Tringe, S. G., Hugenholtz, P., & Muyzer, G. (2016b). Metagenomic insights into the uncultured diversity and physiology of microbes in four hypersaline soda lake brines. *Frontiers in Microbiology*, *7*(211), 1–18.

Vavourakis, C. D., Ghai, R., Rodriguez-Valera, F., Sorokin, D. Y., Tringe, S. G., Hugenholtz, P., & Muyzer, G. (2016c). Metagenomic insights into the uncultured diversity and physiology of microbes in four hypersaline soda lake brines. *Frontiers in Microbiology*, *7*(211), 1–18. https://doi.org/10.3389/fmicb.2016.00211

Vaz-Moreira, I., Egas, C., Nunes, O. C., & Manaia, C. M. (2013). Bacterial diversity from the source to the tap: A comparative study based on 16S rRNA gene-DGGE and culture-dependent methods. *FEMS Microbiology Ecology*, *83*(2), 361–374. https://doi.org/10.1111/1574-6941.12002

Venditto, I., Najmudin, S., Luís, A. S., Ferreira, L. M. A., Sakka, K., Knox, J. P., Gilbert, H. J., & Fontes, C. M. G. A. (2015). Family 46 carbohydrate-binding modules contribute to the enzymatic hydrolysis of xyloglucan and β-1,3-1,4-glucans through distinct mechanisms. *Journal of Biological Chemistry*, *290*(17), 10572–10586. https://doi.org/10.1074/jbc.M115.637827

Ventosa, A., de la Haba, R. R., Sánchez-Porro, C., & Papke, R. T. (2015). Microbial diversity of hypersaline environments: A metagenomic approach. *Current Opinion in Microbiology*, *25*, 80–87. https://doi.org/10.1016/j.mib.2015.05.002

Vera-Gargallo, B., & Ventosa, A. (2018). Metagenomic insights into the phylogenetic and metabolic diversity of the prokaryotic community dwelling in hypersaline soils from the odiel saltmarshes (SW Spain). *Genes*, *9*(152), 1–22. https://doi.org/10.3390/genes9030152

Von Meijenfeldt, F. A. B., Arkhipova, K., Cambuy, D. D., Coutinho, F. H., & Dutilh, B. E. (2019). Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. *Genome Biology*, *20*(1), 217. https://doi.org/10.1186/s13059-019-1817-x

Wang, G., Ren, Y., Ng, T. B., Streit, W. R., & Ye, X. (2019). High-throughput amplicon sequencing demonstrates extensive diversity of xylanase genes in the sediment of soda lake Dabusu. *Biotechnology Letters*, *41*(3), 409–418. https://doi.org/10.1007/s10529-019-02646-w

Wang, G., Wang, Q., Lin, X., Ng, T. B., Yan, R., Lin, J., & Ye, X. (2016). A novel cold-adapted and highly salt-tolerant esterase from *Alkalibacterium* sp. SL3 from the sediment of a soda lake. *Scientific Reports*, *6*(1w), 1–10. https://doi.org/10.1038/srep19494

Wang, G., Wu, J., Yan, R., Lin, J., & Ye, X. (2017). A novel multi-domain high molecular, salt-stable alkaline xylanase from *Alkalibacterium* sp. SL3. *Frontiers in Microbiology*, *7*(JAN), 1–11. https://doi.org/10.3389/fmicb.2016.02120

Wang, M., Zhang, X., Shu, Z., Wang, Z., Tao, Y., Lv, C., Zhu, D., & Shen, G. (2022).

Bacterial and archaeal communities within the alkaline soda Langaco Lake in the Qinghai-Tibet Plateau. *Annals of Microbiology*, *72*(1), 1–15. https://doi.org/10.1186/s13213-022-01691-7

Wani, A. A., Surakasi, V. P., Siddharth, J., Raghavan, R. G., Patole, M. S., Ranade, D., & Shouche, Y. S. (2006). Molecular analyses of microbial diversity associated with the Lonar soda lake in India: An impact crater in a basalt area. *Research in Microbiology*, *157*(10), 928–937. https://doi.org/10.1016/j.resmic.2006.08.005

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., De Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research*, *46*(W1), 296–303. https://doi.org/10.1093/nar/gky427

Webster, G., O'Sullivan, L. A., Meng, Y., Williams, A. S., Sass, A. M., Watkins, A. J., Parkes, R. J., & Weightman, A. J. (2015). Archaeal community diversity and abundance changes along a natural salinity gradient in estuarine sediments. *FEMS Microbiology Ecology*, *91*(2), 1–18. https://doi.org/10.1093/femsec/fiu025

Wei, K. S. C., Teoh, T. C., Koshy, P., Salmah, I., & Zainudin, A. (2015). Cloning, expression and characterization of the endoglucanase gene from *Bacillus subtilis* UMC7 isolated from the gut of the indigenous termite macrotermes malaccensis in escherichia coli. *Electronic Journal of Biotechnology*, *18*(2), 103–109. https://doi.org/10.1016/j.ejbt.2014.12.007

Wi, S. G., Cho, E. J., Lee, D. S., Lee, S. J., Lee, Y. J., & Bae, H. J. (2015). Lignocellulose conversion for biofuel: A new pretreatment greatly improves downstream biocatalytic hydrolysis of various lignocellulosic materials. *Biotechnology for Biofuels*, *8*(1), 1–11. https://doi.org/10.1186/s13068-015-0419-4

Wick, R. R., Judd, L. M., Gorrie, C. L., & Holt, K. E. (2017). Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Computational Biology*, *13*(6), 1–22. https://doi.org/10.1371/journal.pcbi.1005595

Wick, R. R., Schultz, M. B., Zobel, J., & Holt, K. E. (2015). Bandage: Interactive visualization of de novo genome assemblies. *Bioinformatics*, *31*(20), 3350–3352. https://doi.org/10.1093/bioinformatics/btv383

Wierzbicka-Woś, A., Henneberger, R., Batista-García, R. A., Martínez-Ávila, L., Jackson, S. A., Kennedy, J., & Dobson, A. D. W. (2019). Biochemical characterization of a novel monospecific endo-β-1,4-glucanase belonging to GH family 5 from a rhizosphere metagenomic library. *Frontiers in Microbiology*, *10*(6), 1–19. https://doi.org/10.3389/fmicb.2019.01342

Wohlgemuth, R., Littlechild, J., Monti, D., Schnorr, K., van Rossum, T., Siebers, B., Menzel, P., Kublanov, I. V., Rike, A. G., Skretas, G., Szabo, Z., Peng, X., & Young, M. J. (2018). Discovering novel hydrolases from hot environments.

Biotechnology Advances, *36*(8), 2077–2100. https://doi.org/10.1016/j. biotechadv. 2018.09.004

Xu, J., Zhang, H., Zheng, J., Dovoedo, P., & Yin, Y. (2020). ECAMI: Simultaneous classification and motif identification for enzyme annotation. *Bioinformatics*, *36*(7), 2068–2075. https://doi.org/10.1093/bioinformatics/btz908

Xu, L., Dong, Z., Fang, L., Luo, Y., Wei, Z., Guo, H., Zhang, G., Gu, Y. Q., Coleman-Derr, D., Xia, Q., & Wang, Y. (2019). OrthoVenn2: A web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Research*, *47*(W1), W52–W58. https://doi.org/10.1093/nar/gkz333

Yakimov, A. P., Afanaseva, A. S., Khodorkovskiy, M. A., & Petukhov, M. G. (2016). Design of Stable a-Helical Peptides and Thermostable Proteins in Biotechnology and Biomedicine. *Acta Naturae*, *8*(4), 70–81. https://doi.org/10.32607/20758251-2016-8-4-70-81

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., & Zhang, Y. (2015). The I-TASSER suite: Protein structure and function prediction. *Nature Methods*, *12*(1), 7–8. https://doi.org/10.1038/nmeth.3213

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., & Xu, Y. (2012). DbCAN: A web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research*, *40*(W1), W445-Q451. https://doi.org/10.1093/nar/gks479

Yumoto, I., Yamaga, S., Sogabe, Y., Nodasaka, Y., Matsuyama, H., Nakajima, K., & Suemori, A. (2003). *Bacillus krulwichiae* sp. nov., a halotolerant obligate alkaliphile that utilizes benzoate and m-hydroxybenzoate. *International Journal of Systematic and Evolutionary Microbiology*, *53*(5), 1531–1536. https://doi.org/10.1099/IJS.0.02596-0/CITE/REFWORKS

Zahran, H. H. (1997). Diversity, adaptation and activity of the bacterial flora in saline environments. *Biology and Fertility of Soils*, *25*(3), 211–223. https://doi.org/10.1007/s003740050306

Zavarzina, D. G., Zhilina, T. N., Kuznetsov, B. B., Kolganova, T. V., Osipov, G. A., Kotelev, M. S., & Zavarzin, G. A. (2013). *Natranaerobaculum magadiense* gen. nov., sp. nov., an anaerobic, alkalithermophilic bacterium from soda lake sediment. *International Journal of Systematic and Evolutionary Microbiology*, *63*(PART 12), 4456–4461. https://doi.org/10.1099/ijs.0.054536-0

Zhang, G., Li, S., Xue, Y., Mao, L., & Ma, Y. (2012). Effects of salts on activity of halophilic cellulase with glucomannanase activity isolated from alkaliphilic and halophilic *Bacillus* sp. BG-CS10. *Extremophiles*, *16*(1), 35–43. https://doi.org/10.1007/s00792-011-0403-2

Zhang, G., Mao, L., Zhao, Y., Xue, Y., & Ma, Y. (2010). Characterization of a thermostable xylanase from an alkaliphilic *Bacillus* sp. *Biotechnology Letters*, *32*(12), 1915–1920. https://doi.org/10.1007/s10529-010-0372-z

Zhang, Han, Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., Busk, P. K., Xu, Y., & Yin, Y. (2018). DbCAN2: A meta server for automated carbohydrate-

active enzyme annotation. *Nucleic Acids Research*, *46*(W1), W95–W101. https://doi.org/10.1093/nar/gky418

Zhang, Huaidong, Zhang, G., Yao, C., Junaid, M., Lu, Z., Zhang, H., & Ma, Y. (2015). Structural Insight of a Trimodular Halophilic Cellulase with a Family 46 Carbohydrate-Binding Module. *PloS One*, *10*(11), e0142107. https://doi.org/10.1371/journal.pone.0142107

Zhang, W., Xue, Y., Ma, Y., Zhou, P., Ventosa, A., & Grant, W. D. (2002). *Salinicoccus alkaliphilus* sp. nov., a novel alkaliphile and moderate halophile from Baer Soda Lake in Inner Mongolia Autonomous Region, China. *International Journal of Systematic and Evolutionary Microbiology*, *52*(3), 789–793. https://doi.org/10.1099/ijs.0.01912-0

Zhang, Z., Schwartz, S., Wagner, L., & Miller, W. (2000). A Greedy algorithm for aligning DNA sequences. *Journal of Computational Biology*, *7*(1–2), 203–214.

Zhao, B., Yan, Y., & Chen, S. (2014). How could haloalkaliphilic microorganisms contribute to biotechnology? *Canadian Journal of Microbiology*, *60*(11), 717–727.

Zhao, D., Zhang, S., Xue, Q., Chen, J., Zhou, J., Cheng, F., Li, M., Zhu, Y., Yu, H., Hu, S., Zheng, Y., Liu, S., & Xiang, H. (2020). Abundant Taxa and Favorable Pathways in the Microbiome of Soda-Saline Lakes in Inner Mongolia. *Frontiers in Microbiology*, *11*(7), 1–17. https://doi.org/10.3389/fmicb.2020.01740

Zhao, X., Liu, L., Deng, Z., Liu, S., Yun, J., Xiao, X., & Li, H. (2021). Screening, cloning, enzymatic properties of a novel thermostable cellulase enzyme, and its potential application on water hyacinth utilization. *International Microbiology*, *24*(3), 337–349. https://doi.org/10.1007/s10123-021-00170-4

Zhao, Z., Liu, H., Wang, C., & Xu, J. R. (2014). Correction to Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi [BMC Genomics 14(2013) 274]. *BMC Genomics*, *15*(1). https://doi.org/10.1186/1471-2164-15-6

Zhilina, T. N., Garnova, E. S., Tourova, T. P., Kostrikina, N. A., & Zavarzin, G. A. (2001). Amphibacillus fermentum sp. nov. and *Amphibacillus tropicus* sp. nov., new alkaliphilic, facultatively anaerobic, saccharolytic bacilli from Lake Magadi. *Microbiology*, *70*(6), 711–722. https://doi.org/10.1023/A:1013196017556

Zhilina, T. N., Kevbrin, V. V., Tourova, T. P., Lysenko, A. M., Kostrikina, N. A., & Zavarzin, G. A. (2005). *Clostridium alkalicellum* sp. nov., an obligately alkaliphilic cellulolytic bacterium from a soda lake in the baikal region. *Mikrobiologiya*, *74*(5), 642–653.

Zhilina, T. N., Zavarzin, G. A., Rainey, F., Kevbrin, V. V., Kostrikina, N. A., & Lysenko, A. M. (1996). *Spirochaeta alkalica* sp. nov., *Spirochaeta africana* sp. nov., and *Spirochaeta asiatica* sp. nov., alkaliphilic anaerobes from the Continental Soda Lakes in Central Asia and the East African rift. *International Journal of Systematic Bacteriology*, *46*(1), 305–312.

https://doi.org/10.1099/00207713-46-1-305

Zhou, C., Xue, Y., & Ma, Y. (2017). Cloning, evaluation, and high-level expression of a thermo-alkaline pectate lyase from alkaliphilic *Bacillus clausii* with potential in ramie degumming. *Applied Microbiology and Biotechnology*, *101*(9), 3663–3676. https://doi.org/10.1007/s00253-017-8110-2

Zhou, S., Raouche, S., Grisel, S., Sigoillot, J.-C., & Herpoël-Gimbert, I. (2017). Efficient biomass pretreatment using the White-rot Fungus Polyporus Brumalis. *Fungal Genomics & Biology*, *7*(1), 1–6. https://doi.org/10.4172/2165-8056.1000150

Zhu, C., Xu, Z., & Song, R. (2011). The endoglucanase from *Bacillus subtilis* BEC-1 bears halo-tolerant, acidophilic and dithiothreitol-stimulated enzyme activity. *World Journal of Microbiology and Biotechnology*, *27*(12), 2863–2871. https://doi.org/10.1007/s11274-011-0767-6

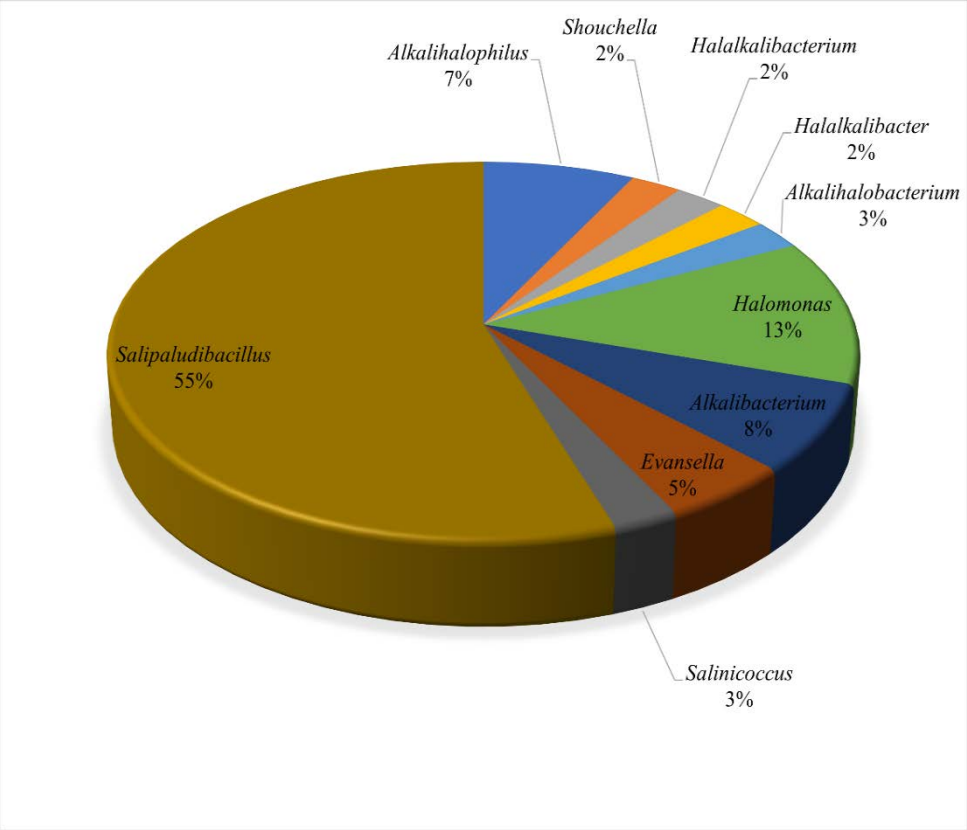
Zhu, D., Zhang, P., Xie, C., Zhang, W., Sun, J., Qian, W. J., & Yang, B. (2017). Biodegradation of alkaline lignin by *Bacillus ligniniphilus* L1. *Biotechnology for Biofuels*, *10*(1), 1–14. https://doi.org/10.1186/s13068-017-0735-y

Zhu, P., Wang, Y., Shi, T., Zhang, X., Huang, G., & Gong, J. (2018). Intertidal zonation affects diversity and functional potentials of bacteria in surface sediments: A case study of the Golden Bay mangrove, China. *Applied Soil Ecology*, *130*(5), 159–168. https://doi.org/10.1016/j.apsoil.2018.06.003

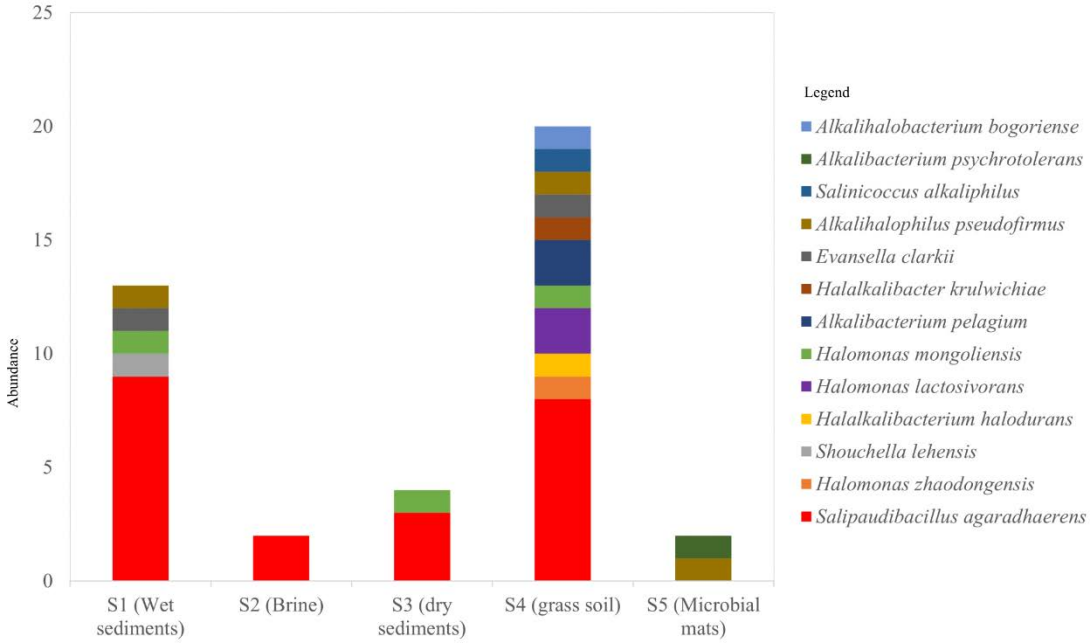
Zorz, J. K., Sharp, C., Kleiner, M., Gordon, P. M. K., Pon, R. T., Dong, X., & Strous, M. (2019). A shared core microbiome in soda lakes separated by large distances. *Nature Communications*, *10*(1), 1–10. https://doi.org/10.1038/s41467-019-12195-5

APPENDICES

Appendix 1: Percentage distribution at genus level of all the bacterial isolates retrieved from Lake Magadi.



Appendix 2: Taxonomic distribution and abundance (in numbers) of bacterial isolates across the sampled sites of Lake Magadi.



Appendix 3: Enzyme activity indices for the bacterial 40 isolates obtained from Lake Magadi.

	EAI ± standard error of mean					
Isolate	**Cellulose**	**CMC**	**Starch**	**Xanthan**	**Pectin**	**Xylan**
LMS1	3.2±0.6	2.9±0.6	2.9 ± 0.0	1.6±1.7	3.8±1.2	2.9±1.0
LMS2	4.4±0.6	4.5±1.0	4.9 ± 0.6	3.0±1.0	3.4±1.5	3.8±0.6
LMS3	0	0	0	0	0	0
LMS4	0	0	0	0	0	0
LMS5	3.8±1.2	2.3±0.6	3.4±0.1	2.5±0.6	3.7±0.6	3.5±0.6
LMS6	1.5±0.6	0	0	0	0	0
LMS7	2.9±0.0	2.9±0.6	3.4 ± 1.0	4.7±1.5	4.5±0.6	2.1±0.0
LMS8	3.2±0.6	2.4±1.2	3.5 ± 1.0	5.6±1.0	3.5±0.0	2.1±0.0
LMS9	2.5±0.0	2.2±0.6	0.9 ± 0.6	2.6±1.5	2.1±1.2	2.4±0.6
LMS10	4.1±1.2	2.7±0.6	1.9 ± 0.6	1.6±1.2	2.2±1.0	2.7±0.6
LMS11	1.9±0.6	1.8±1.2	0	1.4±1.5	0	0
LMS12	4.0±0.6	3.9±0.6	2.4 ± 0.6	4.3±1.2	3.0±0.0	3.0±0.6
LMS13	2.2±0.6	2.0±1.2	2.1 ± 0.6	1.2±0.6	0	1.5±0.6
LMS14	5.7±0.0	3.4±0.6	5.1 ± 0.6	2.4±1.0	3.4±1.2	2.5±0.6
LMS15	3.3±0.6	2.8±0.6	2.2 ± 0.6	2.7±0.6	0	0
LMS16	3.5±1.5	2.7±0.6	3.4 ± 0.6	2.6±2.3	2.3±1.0	2.9±1.5
LMS17	3.2±1.0	3.1±0.0	2.2 ± 0.0	2.8±0.6	2.6±0.6	2.1±0.0
LMS18	0	0	0	0	0	0
LMS19	2.8±0.6	2.0±0.6	2.2 ± 1.5	1.7±1.0	1.5±1.0	1.7±0.6
LMS20	0	0	0	0	0	0
LMS21	3.2±0.6	3.1±1.0	2.4 ± 2.1	3.3±2.7	2.3±1.2	2.1±1.0
LMS22	1.5±1.2	2.9±0.6	2.1±0.6	2.8±1.5	2.7±0.6	2.1±1.0
LMS23	3.5±0.6	3.4±0.6	4.1±0.6	4.1±1.2	2.9±0.6	2.3±0.6
LMS24	3.1±0.6	3.4±0.6	2.3±1.5	5.1±0.6	2.7±0.6	2.8±0.6
LMS25	2.4±0.6	3.9±0.6	1.1±0.6	1.4±0.6	1.4±0.6	2.3±0.6
LMS26	2.8±0.0	2.8±1.0	3.6±0.0	2.7±0.6	3.8±0.6	2.8±0.6
LMS27	4.0±0.6	3.2±0.6	4.5±1.0	2.7±0.6	2.7±0.6	3.4±1.0
LMS28	3.5±0.6	2.9±0.6	4.3±1.0	3.2±1.0	2.3±0.6	2.8±0.6
LMS29	3.6±1.0	2.7±0.6	5.1±0.6	2.0±0.6	2.7±0.6	3.6±1.0
LMS30	0	0	0	0	0	0
LMS31	0	2.6±0.6	2.3±1.5	4.2±0.0	0	0
LMS32	3.1±0.6	3.1±0.6	2.1±0.6	5.0±0.6	2.0±0.6	2.6±0.6
LMS33	1.5±0.6	2.3±0.6	1.5±0.6	4.7±1.5	2.7±0.6	2.4±0.6
LMS34	3.4±0.6	1.7±0.6	0	0	0	0
LMS35	0	0	3.3±2.0	0	0	0
LMS36	0	0	0	0	0	0
LMS37	3.8±0.0	2.8±0.0	2.7±0.6	4.4±0.6	3.3±1.5	4.7±0.6
LMS38	1.9±0.0	1.9±0.6	0	0	3.5±0.6	0
LMS39	2.5±1.0	1.9±0.6	2.1±0.6	2.8±0.6	1.4±0.6	2.6±0.6
LMS40	3.0±0.6	2.7±1.5	2.1±1.2	2.8±1.0	3.1±1.0	3.8±0.6

Appendix 4: OD_{600} for cell densities at various NaCl concentrations.

Isolate	0% NaCl		5% NaCl		10% NaCl	
	Mean	Std dev	Mean	Std dev	Mean	Std dev
LMS1	0.347	0.001	0.287	0.002	0.19	0.003
LMS2	0.240	0.013	0.291	0.004	0.177	0.009
LMS3	1.167	0.002	1.246	0.141	0.735	0.001
LMS4	0.353	0.255	1.283	0.004	1.033	0.002
LMS5	0.237	0.008	0.270	0.001	0.206	0.004
LMS6	0.251	0.007	0.739	0.046	0.554	0.046
LMS7	0.340	0.004	0.274	0.021	0.230	0.032
LMS8	0.280	0.063	0.291	0.001	0.197	0.008
LMS9	0.329	0.001	0.279	0.006	0.187	0.001
LMS10	0.382	0.000	0.283	0.001	0.208	0.002
LMS11	0.600	0.006	0.710	0.011	0.314	0.008
LMS12	0.229	0.019	0.239	0.011	0.255	0.001
LMS13	0.816	0.009	0.740	0.021	0.266	0.011
LMS14	0.378	0.001	0.315	0.006	0.212	0.004
LMS15	0.220	0.003	0.242	0.002	0.173	0.001
LMS16	0.200	0.004	0.232	0.009	0.012	0.002
LMS17	0.370	0.001	0.326	0.001	0.220	0.009
LMS18	0.232	0.013	0.760	0.013	0.744	0.026
LMS19	0.435	0.008	0.564	0.008	0.369	0.009
LMS20	0.002	0.001	0.002	0.001	0.002	0.001
LMS21	0.348	0.016	0.246	0.006	0.208	0.006
LMS22	0.323	0.003	0.389	0.006	0.268	0.009
LMS23	0.353	0.001	0.405	0.019	0.326	0.002
LMS24	0.296	0.006	0.265	0.004	0.187	0.001
LMS25	0.140	0.047	0.342	0.001	0.181	0.004
LMS26	0.293	0.000	0.300	0.001	0.194	0.008
LMS27	0.429	0.004	0.458	0.005	0.309	0.001
LMS28	0.377	0.013	0.304	0.011	0.168	0.005
LMS29	0.265	0.009	0.337	0.038	0.270	0.004
LMS30	0.554	0.014	0.417	0.017	0.430	0.035
LMS31	0.003	0.002	0.009	0.000	0.005	0.001
LMS32	0.272	0.003	0.235	0.004	0.175	0.036
LMS33	0.314	0.006	0.383	0.007	0.275	0.016
LMS34	0.084	0.096	0.163	0.000	0.048	0.003
LMS35	0.903	0.045	1.574	0.026	0.750	0.051
LMS36	0.241	0.009	0.492	0.014	0.445	0.011
LMS37	0.248	0.003	0.217	0.001	0.220	0.063
LMS38	0.185	0.001	0.173	0.009	0.100	0.010
LMS39	0.537	0.011	0.370	0.001	0.090	0.000
LMS40	0.221	0.010	0.334	0.015	0.253	0.001

Appendix 5: OD_{600} for cell densities at temperatures between 35 to 50 °C.

	35°C		40°C		45°C		50°C	
Isolate	**Mean**	**Std dev**	**Mean**	**Std dev**	**Mean**	**Std dev**	**Mean**	**Std dev**
LMS1	0.352	0.004	0.494	0.003	0.603	0.009	0.246	0.004
LMS2	0.797	0.002	0.505	0.003	0.250	0.028	0.099	0.105
LMS3	1.211	0.011	2.156	0.012	2.108	0.006	0.555	0.013
LMS4	0.767	0.013	2.251	0.018	0.092	0.003	0.033	0.001
LMS5	0.533	0.001	0.270	0.003	0.138	0.004	0.002	0.001
LMS6	0.519	0.020	1.689	0.023	0.014	0.003	0.002	0.001
LMS7	0.355	0.020	0.402	0.011	0.009	0.001	0.005	0.001
LMS8	0.365	0.011	0.434	0.025	0.126	0.003	0.041	0.003
LMS9	0.359	0.010	0.434	0.009	0.151	0.006	0.002	0.001
LMS10	0.488	0.009	0.679	0.001	0.152	0.003	0.003	0.001
LMS11	0.498	0.105	2.296	0.008	2.318	0.001	0.004	0.001
LMS12	0.262	0.003	0.643	0.014	0.201	0.001	0.015	0.004
LMS13	0.263	0.002	2.211	0.006	2.230	0.006	0.004	0.001
LMS14	0.343	0.000	0.354	0.001	0.104	0.006	0.001	0.000
LMS15	0.268	0.007	0.410	0.001	0.243	0.001	0.003	0.001
LMS16	0.296	0.006	0.416	0.003	0.056	0.001	0.003	0.001
LMS17	0.634	0.006	0.809	0.006	0.282	0.001	0.023	0.002
LMS18	0.814	0.071	1.865	0.006	1.858	0.003	1.816	0.004
LMS19	0.370	0.080	2.031	0.009	1.847	0.009	0.008	0.002
LMS20	0.244	0.010	0.010	0.010	0.000	0.000	0.000	0.000
LMS21	0.438	0.005	1.109	0.023	0.066	0.002	0.013	0.003
LMS22	0.415	0.008	0.421	0.006	0.163	0.006	0.002	0.001
LMS23	0.390	0.002	0.710	0.001	0.149	0.001	0.001	0.000
LMS24	0.340	0.008	0.394	0.002	0.150	0.004	0.002	0.000
LMS25	0.360	0.028	0.430	0.006	0.179	0.004	0.003	0.001
LMS26	0.440	0.016	0.399	0.001	0.109	0.008	0.004	0.001
LMS27	0.429	0.007	1.562	0.017	0.002	0.001	0.000	0.000
LMS28	0.363	0.004	1.770	0.007	0.020	0.003	0.002	0.001
LMS29	0.501	0.005	0.514	0.011	0.734	0.001	0.023	0.002
LMS30	0.319	0.028	2.174	0.153	0.106	0.004	0.007	0.001
LMS31	0.156	0.001	0.089	0.002	0.103	0.004	0.004	0.001
LMS32	0.379	0.005	0.313	0.002	0.349	0.005	0.153	0.011
LMS33	0.347	0.004	1.001	0.023	0.033	0.005	0.002	0.000
LMS34	0.758	0.002	0.309	0.001	0.019	0.002	0.003	0.001
LMS35	0.944	0.028	1.913	0.001	0.031	0.001	0.010	0.001
LMS36	0.463	0.010	1.511	0.029	0.555	0.000	0.448	0.020
LMS37	0.328	0.005	0.226	0.004	0.187	0.026	0.002	0.001
LMS38	0.364	0.006	0.291	0.005	0.004	0.003	0.003	0.001
LMS39	0.360	0.004	1.265	0.040	2.003	0.018	0.002	0.001
LMS40	0.336	0.000	0.386	0.012	0.148	0.003	0.002	0.001

Appendix 6: OD_{600} for cell densities at pH values between 6.0 and 10.0.

	pH 6.0		pH 7.0		pH 8.0		pH 9.0		pH 10.0	
Isolate	**Mean**	**Std dev**	**Mean**	**Std dev**	**Mean**	**Std dev**	**Mean**	**Std dev**	**Mean**	**Std dev**
LMS1	0.21	0.09	0.522	0.016	0.57	0.001	0.57	0.052	0.58	0.004
LMS2	0.01	0.001	0.273	0.013	0.42	0.018	0.51	0.057	0.17	0.049
LMS3	0.01	0.001	0.577	0.018	0.9	0.023	1.57	0.025	1.60	0.016
LMS4	0.16	0.008	1.032	0.004	1.70	0.018	1.89	0.017	1.66	0.006
LMS5	1.56	0.004	0.515	0.005	0.79	0.037	0.65	0.004	0.61	0.049
LMS6	0.01	0.001	0.025	0.005	0.21	0.015	0.20	0.013	0.19	0.004
LMS7	0.01	0.001	0.496	0.020	0.57	0.012	1.33	0.035	1.23	0.106
LMS8	0.24	0.001	0.444	0.020	0.60	0.012	0.58	0.035	0.74	0.106
LMS9	0.01	0.001	0.276	0.008	0.41	0.021	0.40	0.001	0.33	0.016
LMS10	0.24	0.004	0.473	0.019	0.72	0.020	0.82	0.015	0.57	0.002
LMS11	0.02	0.003	0.245	0.008	1.43	0.006	1.62	0.124	1.68	0.021
LMS12	0.11	0.001	0.334	0.007	0.50	0.015	0.33	0.016	0.30	0.021
LMS13	0.01	0.003	0.107	0.013	1.07	0.095	1.35	0.162	1.43	0.158
LMS14	0.17	0.002	0.397	0.001	0.57	0.004	0.41	0.044	0.45	0.004
LMS15	0.01	0.001	0.002	0.000	0.04	0.004	0.04	0.001	0.13	0.074
LMS16	0.01	0.000	0.023	0.007	0.14	0.008	0.32	0.039	0.22	0.003
LMS17	0.15	0.001	0.484	0.004	0.66	0.010	0.44	0.007	0.24	0.148
LMS18	0.01	0.001	0.257	0.019	1.11	0.078	1.18	0.129	0.69	0.045
LMS19	0.01	0.000	0.828	0.093	1.09	0.011	1.19	0.055	1.05	0.084
LMS20	0.13	0.002	0.087	0.000	0.08	0.000	0.07	0.001	0.08	0.004
LMS21	0.17	0.001	0.367	0.008	0.69	0.028	0.82	0.006	0.79	0.052
LMS22	0.11	0.001	0.401	0.003	0.67	0.098	0.55	0.021	0.47	0.011
LMS23	0.45	0.012	0.382	0.020	1.03	0.026	1.03	0.002	0.88	0.026

LMS24	0.32	0.001	0.393	0.001	0.53	0.001	0.49	0.001	0.36	0.005
LMS25	0.3	0.001	0.558	0.124	0.52	0.100	0.42	0.086	0.31	0.025
LMS26	0.22	0.003	0.380	0.004	0.69	0.006	0.75	0.051	0.76	0.028
LMS27	0.16	0.002	0.533	0.048	1.07	0.065	1.25	0.001	0.76	0.006
LMS28	0.21	0.000	0.550	0.001	0.77	0.000	0.76	0.024	0.47	0.006
LMS29	0.09	0.001	0.495	0.011	0.78	0.004	0.90	0.042	0.89	0.025
LMS30	0.22	0.002	0.215	0.012	0.66	0.040	0.85	0.123	0.49	0.022
LMS31	0.2	0.004	0.212	0.001	0.23	0.004	0.29	0.003	0.30	0.001
LMS32	0.25	0.001	0.358	0.009	0.51	0.010	0.39	0.015	0.41	0.011
LMS33	0.32	0.001	0.323	0.013	0.67	0.010	0.88	0.005	0.99	0.012
LMS34	0.25	0.015	0.086	0.001	0.10	0.010	0.13	0.026	0.19	0.006
LMS35	0.26	0.032	2.286	0.006	2.74	0.014	2.75	0.005	2.65	0.049
LMS36	0.12	0.001	0.108	0.002	1.36	0.023	1.09	0.004	0.4	0.011
LMS37	0.09	0.001	0.352	0.001	0.52	0.039	0.44	0.002	0.33	0.001
LMS38	0.23	0.001	0.914	0.083	0.22	0.006	0.24	0.004	0.29	0.002
LMS39	0.24	0.004	0.727	0.001	0.85	0.005	1.09	0.057	1.33	0.083
LMS40	0.22	0.001	0.377	0.003	0.62	0.013	0.22	0.025	0.28	0.009

Appendix 7: Comparative analysis of genomic features for the sequenced genomes with their type strain.

Isolate and type strain	**Genome size (bp)**	**No. of contigs**	**G+C content (%)**	**No. of RNAs**
LMS6	3,886,007	1	43	138
Shouchella lehensis DSM19099^T	3,994,080	8	40	101
LMS18	4,903,744	1	40.8	102
Evansella clarkii DSM8720^T	5,476,388	27	42.1	142
LMS25	4,850,562	1	39	120
Salipaludibacillus agaradhaerens DSM 8721^T	4,319,439	2	38.9	96
LMS39	4,525,667	1	37	97
Ab. bogoriense ATCC BAA-922^T	5,002,403	51	36.7	81

Appendix 8: Genes involved in the soda lake environmental adaptation of four bacterial isolates from Lake Magadi.

Gene component	**LMS6**	**LMS18**	**LMS25**	**LMS39**
Glycine betaine transporter *OpuD/betL*	✓	✓	✓	✓
glycine betaine/carnitine transport permease protein *GbuB*	✓	✓	✓	✓
Glycine betaine/carnitine transport *GbuAC*	✓	✓	✓	✓
Ectoine binding periplasmic protein *TeaA*	✓	-	✓	✓
K+/H+ antiporter *Nhap2*	✓	-		✓
K+ uptake systems *trkA*, *trkH* and *trkG*	✓	✓	✓	✓
Na+/H+ antiporter subunit A-G	✓	✓	✓	✓
Trehalose synthase	✓	-	-	✓
PTS system trehalose specific *EIIBE* component	✓	✓	✓	✓
ectoine expression cassette *ectABC*	✓	✓	✓	✓

Appendix 9: Identity analysis of the carbohydrate-active enzymes derived from genomes of four bacterial isolates.

Gene locus tag	Source	Closest relative and source	similarity (%)	Accession No.
EJIFPJLI_01880	LMS6	cellulase family GH (*Bacillus* sp. Marseille-P3800)	97	WP218014778
EJIFPJLI_00557	LMS6	6-phospho-beta-glucosidase [*Bacillus* sp. Marseille-P3800]	94	WP099302829
EJIFPJLI_03917	LMS6	GH_2 [*Bacillus* sp. Marseille-P3800]	92	WP099304398
EJIFPJLI_03908	LMS6	α-amylase [*Bacillus* sp. Marseille-P3800]	99	WP176554439
EJIFPJLI_00866	LMS6	GH13 protein [*Bacillus* sp. Marseille-P3800]	99	WP099302352
PIHNMHKA_04277	LMS18	endo-1,4-beta-xylanase (*E. clarkii*)	95	WP088034427
FBGOGOLJ_02004	LMS25	Endoglucanase_5A [*Salipaludibacillus agaradhaerens*]	85	WP078579508
FBGOGOLJ_02205	LMS25	α-amylase [*Sb. agaradhaerens*]	97	WP095995549
FBGOGOLJ_02212	LMS25	α-glycosidase [*Sb. agaradhaerens*]	88	WP078578230
PCDFDMCF_00462	LMS39	alkaline thermostable endoxylanase [*Bacillus* sp. NG-27]	97	AAB70918
PCDFDMCF_03001	LMS39	chitinase (*Alkalihalobacterium bogoriensis*)	96	WP035177761
PCDFDMCF_04153	LMS39	1,4-β-xylosidase [*Ab. bogoriense*]	91	WP026671748
PCDFDMCF_03570	LMS39	GH28 protein [*Ab. bogoriense*]	89	WP026674505
PCDFDMCF_03228	LMS39	Rhamnogalacturonan lyase (*Ahb. hemicellulosilyticus* JCM 9152)	72	GAE30456