Electronic Journal of Biotechnology 77 (2025) 48–58

Contents lists available at ScienceDirect

Electronic Journal of Biotechnology

journal homepage: www.elsevier.com/locate/ejbt

Research article

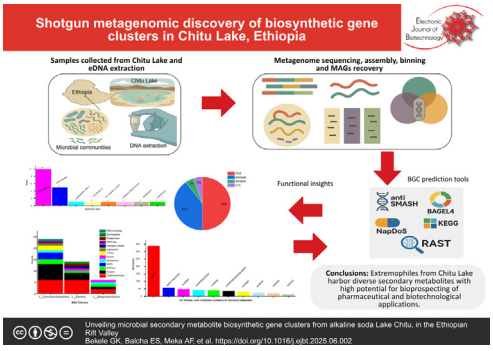
Unveiling microbial secondary metabolite biosynthetic gene clusters from alkaline soda Lake Chitu, in the Ethiopian Rift Valley ☆

Gessesse Kebede Bekele [a,b,*], Ermias Sissay Balcha [c], Abu Feyisa Meka [d], Eskedar Getachew Assefa [b,e], Ebrahim M. Abda [a,b], Fassil Assefa Tuji [f], Mesfin Tafesse Gemeda [a,b,*]

[a] Department of Biotechnology, College of Natural and Applied Sciences, Addis Ababa Science and Technology University, Addis Ababa, Ethiopia
[b] Biotechnology and Bioprocess Center of Excellence, Addis Ababa Science and Technology University, Addis Ababa, Ethiopia
[c] School of Medical Laboratory Science, College of Medicine and Health Sciences, Hawassa University, Hawassa, Ethiopia
[d] Department of Biology, Bule Hora University, Bule Hora, Ethiopia
[e] Department of Food Science and Applied Nutrition, College of Natural and Applied Sciences, Addis Ababa Science and Technology University, Addis Ababa, Ethiopia
[f] Department of Cellular, Microbial and Molecular Biology, College of Natural and Computational Sciences, Addis Ababa University, Addis Ababa, Ethiopia

GRAPHICAL ABSTRACT

Unveiling microbial secondary metabolite biosynthetic gene clusters from alkaline soda Lake Chitu, in the Ethiopian Rift Valley.



ARTICLE INFO

Article history:
Received 21 April 2025
Accepted 11 June 2025
Available online 8 August 2025

Keywords:
antiSMASH
BAGLE4
Biosynthetic gene clusters
KEGG
Metagenome-assembled genomes
NaPDoS
Secondary metabolites

ABSTRACT

Background: Microorganisms inhabiting alkalihalo-soda lakes are known for producing diverse secondary metabolites with potential biotechnological and pharmaceutical applications. This study explored the biosynthetic capabilities of microbial communities from Ethiopia's Chitu Lake through shotgun metagenomic sequencing and metagenome-assembled genome (MAG) analyses using various bioinformatics tools.

Results: Analysis of MAGs using the Antibiotics and Secondary Metabolite Analysis Shell (antiSMASH) revealed 13 major types of biosynthetic gene clusters. The most abundant were terpene-precursors (32%) and terpene clusters (25%), followed by ribosomally synthesized and post-translationally modified peptides (9%) and nonribosomal peptide synthetases (7%). Other less common BGCs (5% each) included betalactone, ectoine, and Type I polyketide synthase, while rare types (2% each) comprised arylpolyene, hydrogen cyanide, phosphonate, ranthipeptide, and others. The Natural Product Domain Seeker (NaPDoS) detected ketosynthase domains linked to pharmaceutically important such as various fatty acid

☆ Audio abstract available in Supplementary material.

Peer review under responsibility of Pontificia Universidad Católica de Valparaíso.

* Corresponding authors.
E-mail addresses: gessesse.kebede@aastu.edu.et (G.K. Bekele), mesfin.tafesse@aastu.edu.et (M.T. Gemeda).

https://doi.org/10.1016/j.ejbt.2025.06.002

Soda Lakes
Terpene clusters
Terpene-precursors

synthesis, modular and iterative domain classes, and condensation domain which is associated with L-amino acid coupling (LCL) domain class, such as those involved in syringomycin biosynthesis. In addition, bacteriocin analysis identified sactipeptides (56%) and lasso peptides (28%) as dominant types. Kyoto Encyclopedia of Genes and Genomes pathway analysis uncovered several secondary metabolite pathways including those for penicillin, cephalosporins, alkaloids, and phenazines. Rapid Annotation using Subsystem Technology further highlighted secondary metabolism pathways vital for microbial survival in Chitu Lake's extreme environment.

Conclusions: The discovery of diverse biosynthetic gene cluster positions Chitu Lake as a valuable source of secondary metabolites, highlighting the biotechnological, industrial, pharmaceutical, agricultural and environmental potential of its extremophilic microbes and supporting further bioprospecting efforts.

How to cite: Bekele GK, Balcha ES, Meka AF, et al. Unveiling microbial secondary metabolite biosynthetic gene clusters from alkaline soda Lake Chitu, in the Ethiopian Rift Valley. Electron J Biotechnol 2025;77. https://doi.org/10.1016/j.ejbt.2025.06.002.

1. Introduction

Haloalkaliphiles, a unique member of extremophilic microorganisms, thrive in dual-stress environments characterized by elevated salinity and alkalinity, such as soda lakes and hypersaline soils [1]. These organisms span the domains Archaea, Bacteria, *Viruses* and Eukarya, showcasing evolutionary ingenuity in colonizing some of Earth's most inhospitable ecological niches with multiple adaptive mechanisms [2]. To endure osmotic stress, desiccation, and oxidative damage inherent to these habitats, haloalkaliphiles employ a triad of survival strategies: (i) intracellular accumulation of osmoprotectants (e.g., glycerol, trehalose, betaine) to stabilize cellular hydration and turgor [3]; (ii) synthesis of extremozymes and stress-optimized metabolic networks that maintain catalytic efficiency under alkaline and hypersaline conditions [4]; and (iii) production of secondary metabolites, including radical-scavenging antioxidants (e.g., carotenoids) that neutralize reactive oxygen species [5,6].

Their biochemical versatility positions haloalkaliphiles as microbial powerhouses for industrial and medical innovation. Through targeted bioprospecting, these organisms yield high-value metabolites such as antimicrobial compounds, thermostable enzymes, and carotenoid pigments [7,8]. Such molecules serve as foundational components for biosensors, biodegradable polymers, photoelectric materials, and nutraceutical additives [5,9,10,11]. Notably, their bioactive compounds—including antitumor agents and next-generation antibiotics—hold transformative potential for pharmaceutical development. Central to this metabolic diversity are biosynthetic gene clusters (BGCs), genomic modules encoding coordinated pathways for specialized metabolite synthesis [12]. Advances in bioinformatics now enable systematic mining of microbial genomes to identify and characterize cryptic BGCs, accelerating the discovery of novel biomolecules. This computational-to-experimental pipeline not only deciphers the evolutionary logic of extremophile adaptation but also unlocks scalable routes for biomanufacturing.

The secondary metabolite BGCs embrace a wide variety of biological and chemical components, including NRPs, RiPPs, polyketides (PKs), bacteriocins, and terpenes and different biosynthetic pathways for novel antibiotic production, food preservation, microbial ecology, and plant biocontrol [13,14,15,16]. This, therefore, necessitates the focus on BGCs to discover novel microbial products from both culturable and non-cultivable microorganisms. For many years now, culture-dependent techniques have been used to discover bioactive compounds from Lake Chitu [17,18]; however, the microbial secondary metabolite BGCs have not been investigated by previous studies. Quite recently, Balcha et al. [19] reported microbial diversity and BGC potential of Lake Afdera, one of the Ethiopian Rift Valley Lakes, and identified the microbial profile and various distinct BGCs that encode secondary metabolites with several categories of functions. Therefore, in this study, a more comprehensive exploration of microbial secondary metabolite BGCs of Lake Chitu has been studied using a shotgun metagenomics sequencing approach. Lake Chitu is a halo-alkaline in the Rift Valley region of Ethiopia and is a crater lake with 6% salinity and around 10.5 pH value [1]. The lakeside was prioritized due to its logistical and safety challenges (depth, boats, Ekman grab etc.) during this study and yet requires specialized equipment and dynamic environmental conditions—such as fluctuating pH, salinity, and oxygen levels—which are known to drive microbial diversity. These transitional zones often host unique aerobic extremophilic microbes to periodic desiccation and nutrient influx. Such adaptations make them ideal for investigating biosynthetic gene clusters linked to secondary metabolite production. Despite its harsh physicochemical environment, Lake Chitu represents a promising biome for discovering novel secondary metabolites and specialized metabolic pathways, shaped by the lake's unique, poly-extreme conditions. For this purpose, bioinformatics pipelines such as antiSMASH [11], BActeriocin GEnome mining tooL version 4 (BAGEL4) [13] and NaPDoS [7,20] have employed to detect the secondary metabolite BGCs and the analysis of MAGs. Additionally, both the Kyoto Encyclopedia of Genes and Genomes (KEGG) annotation database [21] and the RAST server [19] were employed to identify genes and proteins involved in secondary metabolite biosynthesis and their associated metabolic pathways. Thus, the study identified microbial secondary metabolite BGCs, which likely hold potential for various industrial, pharmaceutical and biotechnological applications.

2. Materials and methods

2.1. Study and sampling sites

Samples of soil and mud were taken from Chitu Lake, which is situated in the central region of Ethiopia, 287 km south of Addis Ababa, the country's capital, at an elevation of 1540 m above sea level and with coordinates of 7°24′0′N and 38°25′0′E. Except for a few hot springs that rise on the beach and run into the lake, it is a crater lake with no input or outflow [1] (Fig. 1).

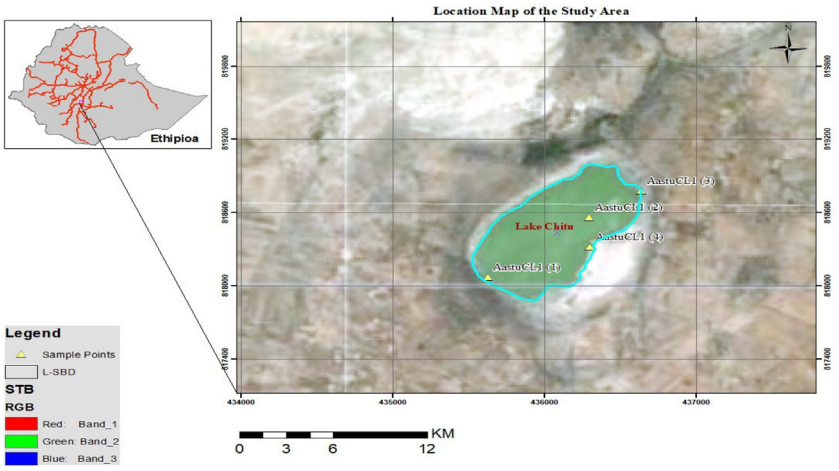


Fig. 1. The map shows the sampling site and points at Lake Chitu. The sampling point labeled with AastuCL1 refers to Addis Ababa Science and Technology University for Chitu Lake's sampling point.

2.2. The sampling and sample designation

The samples were collected in triplicate under strict aseptic settings at four different sampling points (7°24′02′′N 38°25′0′′E, 7°24′18′′N 38°26′27′′E, 7°24′25′′N 38°25′33′′E, and 7°24′10′′N 38°25′22′′E) (Fig. 1), each georeferenced with a Garmin® handheld GPSMAP64 from Chitu Lake and were pooled and labeled with AastuCL1 (Addis Ababa Science and Technology University for Chitu Lake). The soil and mud samples were then placed in sterile polyethylene bags and transported to the laboratory in a cold room maintained at –20°C for further analysis.

2.3. The determination of physicochemical parameters and elemental composition

The on-site measurements of physico-chemical parameters, including temperature, pH, salinity, electrical conductivity, dissolved oxygen, turbidity, ammonium, chloride, and nitrate, were conducted using multi-parameter devices (HANNA HI929). The elemental makeup of the lake was also analyzed using the novAA 800 Atomic Absorption Spectrometer at the Environmental Protection Agency Laboratory in Addis Ababa City, Ethiopia.

2.4. Metagenomic DNA extraction and sequencing

Metagenomic DNA was extracted from soil and mud samples using the procedure outlined by Verma et al. [22]. The extracted DNA was later pooled prior to shotgun metagenome sequencing. Following that, a NanoDrop 3300 spectrophotometer (Thermo Fisher Scientific, Washington, DE, USA) and 1% agarose gel electrophoresis were used to evaluate the quantity and quality of the extracted DNA (Fig. S1a). The library construction process was subsequently made easier by randomly shearing the purified DNA to create small-sized fragments of about 350 bp. Then, the bioanalyzer (Agilent 2100, Agilent, Santa Clara, CA, USA) was used to examine the size distribution of the library before PCR amplification. To prepare a paired-end library, an Illumina Ultra DNA Library Prep Kit (New England Biolabs, Beverly, MA, USA) was used. A NovaSeq PE150 device (Illumina, Tsim Sha Tsui, Hong Kong) was used to sequence the pooled libraries. The results were demultiplexed to produce distinct paired-end raw reads. The quality of these reads was evaluated using FastQC in preparation for downstream analysis. Then, clean reads were obtained using Trimmomatic v0.36 [23], which involved trimming low-quality reads (Q) <38, removing reads with unknown "N" bases (N > 10 bp), eliminating adapter overlaps (≥15 bp), and discarding scaffolds shorter than 500 base pairs [24,25].

2.5. Metagenomic assembly, binning, and annotation

Scaffolds from high-quality reads were obtained by the use of MEGAHIT software (v1.2.9; -presets meta-large -m 180,000,000,000 -t 2) for metagenome assembly [25,26]. Using MetaGeneMark (version 3.05; -a-d-f G -p 1), open reading frame (ORF) prediction was performed on scaffolds that were at least 500 base pairs long, and CD-HIT (v.4.5.8; -T 6 -G 0 -aS 0.9 -g 1 -d 0 -c 0.95 -n 5 -M 8000) was used to create nonredundant gene catalogs [26]. Following this, Bowtie2 (v.2.2.4; –end-to-end –sensitive –no-hd –no-sq –I 200 –X 400 –threads 8) was used to map clean reads to the gene catalog (≥95% identity and 90% coverage) [25]. Then, the contigs were binned into MAGs using MetaBAT2 (v1.7; –assembly-ref –min-contig-length 1500), CONCOCT (v1.1; –read-mapping-tool Bowtie2 –min-contig-length 2500 –contig-split-size 10,000 –contig-split-overlap 0 –kmer-length 4 –max-clusters-vgmm 400 –max-iterations-vgmm 500 –percent-pca 90), and MaxBin2 (v2.2.4; –assembly-object –probability-threshold 0.8 –marker-set 40 –min-contig-length 1000 –plot-marker –threads 8) [27,28]. Then, after pooling the recovered MAGs using the dereplication, aggregation, and scoring tool (DAS Tool (v1.1.2; –gene-identification-tool Diamond –score-threshold 0.5 –duplicate-penalty 0.6 –megabin-penalty 0.5) [27], CheckM (v1.0.18) (≥70% completeness and <10% contamination) [29] was used to evaluate the qualified MAGs for BGCs prediction.

2.6. Taxonomic assessment of the assembled contigs

The microbial taxonomic and functional diversity was assessed using DIAMOND software (v2.1.6; –p 4 –e 1e-5 –k 50 –id 30 –sensitive) with the National Center for Biotechnology Information (NCBI) NR database (blastp, *e* value ≤1e-5) [26,30]. Taxonomic classification of the microbial metagenome from the sample was visualized using Krona plots and stack bar charts, providing a detailed overview of the microbial diversity.

2.7. Detection of biosynthetic gene clusters

The MAGs were examined with antiSMASH version 7.0 in order to detect BGCs [11]. Further predictions of possible biosynthetic pathways and chemical scaffolds were also made possible by the MIBiG (Minimum Information about a Biosynthesis-related Gene Cluster) database [31] and the recognized cluster blast feature of

antiSMASH [32]. KEGG pathway annotation (v2.1.6; –p 4 –e 1e-5 –k 50 –id 30 –sensitive) was also applied to profile genes associated with secondary metabolite biosynthesis [21]. The RAST-web tool was further used to identify genes and proteins linked to adaptation to harsh environmental conditions [19]. Visualization of the identified BGCs was done using the ggplot2 packages in R Studio.

2.8. The detection of Bacteriocins, RiPPs, KS and C-domains

In the MAGs, bacteriocin or RiPPs were identified using the BAGEL4 tool, and known natural product biosynthetic domains such as Ketosynthase (KS) and Condensation (C) were predicted using the NaPDoS tool [33].

3. Results

3.1. The physicochemical measurement and elemental composition of Chitu Lake

Lake Chitu's chemical and physical characteristics indicate that its water is both highly alkaline and salty (Table S1). The lake exhibits a pH of 10.56, turbidity of 33.35 NTU (Nephelometric turbidity units), and an unusually high salinity of 46,500 mg/L. It also has a high mineral content, with an electrical conductivity of 65,780 μS/cm and a temperature of 27.2°C. The total suspended solids are 81,540 mg/L, and dissolved oxygen is 6.5 mg/L. Nutrient levels (nitrate, nitrite, and phosphate) are elevated, and the lake contains trace essential metals (iron, zinc, manganese, cobalt, and copper) and heavy metals (including lead, chromium, cadmium, mercury and nickel).

3.2. The metagenome data

A total of 88,141,232 cleaned reads were generated, accounting for 99.73% of the data, with a GC content of 63.88% (Table 1, Fig. S1b). These reads were assembled into 1,008,197 scaftings containing 956, 295, 942 bp. The Bins from the metagenomic assembly generated were 61 using CONCOCT, 55 using MaxBin2, and 48 using MetaBAT2. Optimization with the DAS tool refined 20 MAGs with their quality and taxonomic identification, of which 10 candidate MAGs qualified for further BGC investigations (Table S2, Fig. S2).

3.3. Microbial community composition analysis

The data analysis obtained from Chitu Lake's metagenome using shotgun metagenomic sequencing revealed that the microbial community is predominantly composed of members from the kingdom Bacteria (89%), followed by Archaea (4%), Viruses (0.2%), and Eukaryota (0.04%). Furthermore, unclassified taxa (0.3%), unknown taxa (i.e., sequences that could not be matched to any known taxonomic group; Fig. 2a), and others (representing a combination of low-abundance, unassigned, or unidentified taxa; Fig. 2b) collectively accounted for more than 6% of the total reads.

From the analysis, the most dominant phyla were Pseudomonadota (40%) and Actinomycetota (21%) followed by Gemmatimonadota (6%), Chloroflexota (5%), Euryarchaeota (5%), Acidobacteriota (3%), Balneolota (3%), Planctomycetota (1%), and Bacillota (2%) (Fig. 3a). At the genera level, Nitriliruptor and Halomonas were found to be the most dominant, followed by Wenzhouxiangella, Thioalkalivibrio, Egicoccus and others (Fig. 3b).

Table 1
The shotgun metagenomic sequence and assembly data sets statistics for Chitu Lake.

Data Sets	Quantified Values
Raw data (GB)	6.6 GB
Raw data (bp)	13,257,189,900
Raw reads (bp)	88,381,266
Clean reads (bp)	88,141,232
Library insert size (bp)	350
% bases >Q20	97.4%
Length of single reads (bp)	150
Effectiveness (%)	99.73
GC content (%)	63.89
Scaftings number	1,008,197
Average scaftings length (bp)	948.53
Maximum scafting length (bp)	143,282
Total scafting length (bp)	956,295,942
N50 length (bp)	936
N90 len(bp)	549

3.4. Secondary metabolite BGCs identified by antiSMASH

The antiSMASH analysis of BGCs across 10 MAGs from Chitu Lake uncovered a diverse array of secondary metabolite pathways. No BGCs were detected in MAG04 and MAG09, whereas the remaining eight MAGs collectively contained 353 biosynthetic domains spanning 13 distinct BGC types (Table S3). Among the identified BGCs, terpene-precursor clusters were the most abundant, comprising 32% (14 counts) of the total, followed by terpene clusters at 25% (11 counts). RiPP-like clusters represented 9% (4 counts), and NRPS clusters accounted for 7% (3 counts). Less prevalent BGC types included betalactone, ectoine, and Type I polyketide synthase (T1PKS), each constituting 5% (2 counts). Other categories—arylpolyne, hydrogen cyanide (HCN), NRPS-like, phosphonate, ranthipeptide, and RRE-containing—were each observed once (2%) (Fig. 4a). Taxonomic analysis revealed that MAGs affiliated with *c__Gammaproteobacteria* (MAG03, MAG08, MAG011, and MAG013) harbored the highest diversity of BGCs (10 types), including NRPS, Terpene-precursor, Terpene, RiPP-like, Phosphonate, Betalactone, Arylpolyne, T1PKS, and NRPS-like clusters (Fig. 4b). This was followed by *k__Bacteria* (7 types) and *c__Betaproteobacteria* (4 types). Notably, MAG08 and MAG011 exhibited the highest number of BGC types (six each), followed by MAG06 with five types, while the remaining MAGs contained three to four BGC types (Fig. S3).

Taxonomic analysis identified that MAGs classified under c__Gammaproteobacteria (MAG03, MAG08, MAG011, and MAG013) exhibited the highest diversity of biosynthetic gene clusters (BGCs), encompassing 10 distinct types, including nonribosomal peptide synthetase (NRPS), terpene-precursor, terpene, RiPP-like, phosphonate, betalactone, arylpolyne, type 1 polyketide synthase (T1PKS), and NRPS-like clusters (Fig. 4b, Fig. S3). This was followed by MAGs within k__Bacteria (MAG06, MAG015, and MAG020) comprise 7 types of BGC and c__Betaproteobacteria (MAG07) comprises 4 types of BGCs. Notably, MAG08 demonstrated a particularly high number of hits for BGCs.

Further analysis of the microbiome revealed 17 distinct classes of biosynthetic gene clusters (BGCs), which were systematically categorized into three major groups based on their biochemical characteristics and biosynthetic mechanisms (Fig. 5, Table S3). The classification comprised six main BGC classes (including terpenes, polyketides, NRPs, alkaloids, RiPPs, and Saccharides), five hybrid classes (combining multiple biosynthetic systems such as NRP-polyketide, polyketide-terpens, NRP-polyketide saccharides, NRP-saccharides and NRP-terpene) and others (other specified and unspecified). The distribution of major BGC classes showed remarkable specialization, with terpene biosynthetic pathways dominating the profile at 36% (126 counts), reflecting their crucial role in microbial adaptation to environmental stresses. NRPs and polyketides represented significant secondary components at 14% to 16% indicating substantial potential for bioactive compound

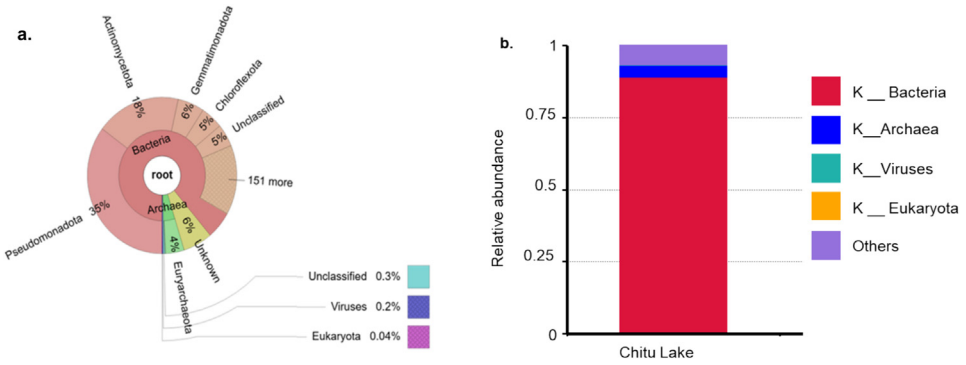


Fig. 2. The microbial community composition in Chitu Lake. The taxonomic composition of the microbial community in Chitu Lake was analyzed using shotgun metagenomic sequencing followed by alignment against the Micro-NR database. A Krona plot (Fig. 2a) illustrates the hierarchical distribution of identified taxa, including Bacteria, Archaea, Viruses, and Eukaryota, across all taxonomic levels from kingdom to phylum. The radial visualization depicts relative abundances based on spectral counts, with concentric circles representing progressively finer taxonomic classifications (inner to outer: kingdom, phylum, etc.). Additionally, a bar plot (Fig. 2b) summarizes the relative abundances of the four major microbial kingdoms within the Chitu Lake ecosystem, providing a quantitative comparison of their distribution.

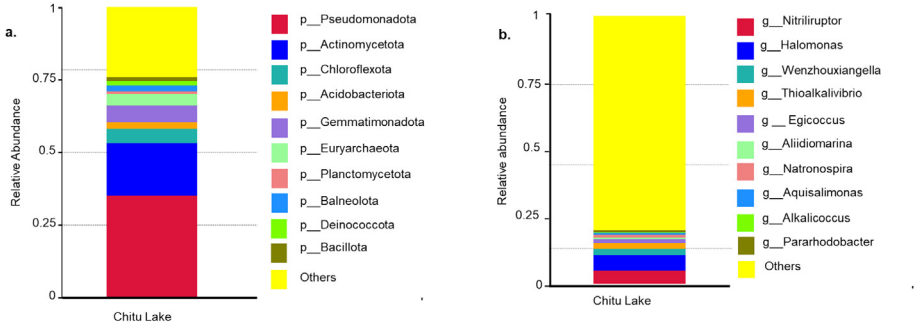


Fig. 3. The bar graph using the micro-NR database showed the relative abundance of microbes at phyla level (a) and genera level (b) in Chitu Lake. It is showed that bacteria are the most dominant microbial community followed by the archaeal groups.

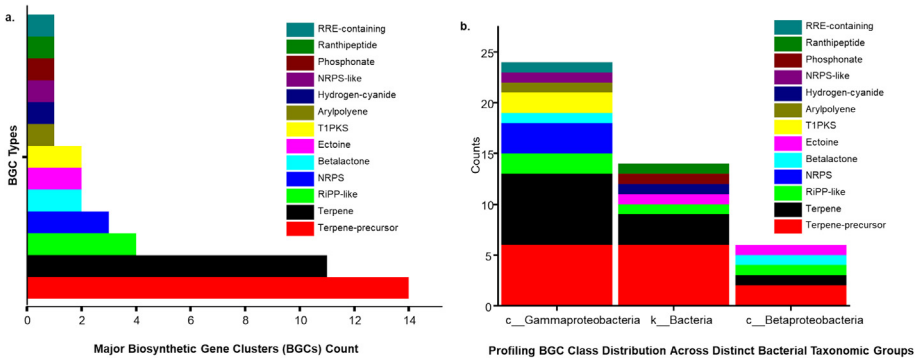


Fig. 4. Composition and taxonomic distribution of BGCs in Chitu Lake microbiome. (a) Analysis of BGCs diversity revealed that terpene-precursors represented the most abundant class, accounting for 57% of all identified BGCs. This was followed by RiPP-like and beta-lactone clusters, indicating a strong biosynthetic potential for specialized metabolites within this ecosystem. (b) Taxonomic profiling of BGCs identified distinct phylogenetic patterns, with MAGs affiliated with Gammaproteobacteria exhibiting the greatest diversity and abundance of BGCs. This suggests that Gammaproteobacteria are likely key contributors to secondary metabolite production in Chitu Lake. Other taxonomic groups showed narrower, but taxon-specific, BGC repertoires, reflecting patterns of niche-specific metabolic specialization.

production. On the other hand, hybrid (Mixed) BGCs accounted for a small but notable fraction of the total, with NRP–polyketide combinations being the most frequent (18 counts, 5%). Other mixed types such as Polyketide–Terpene, NRP–Polyketide–Saccharide and others occurred less frequently, each contributing below 1.2%. In addition, six other BGC classes were identified, representing specialized biosynthetic pathways grouped under 'others'. These include both "Others unspecified" represent 16% (54 counts), while "Others specified" comprising rarer pathways such as aminocoumarin, shikimate-derived, phosphonate, ectoine, and nucleoside-related types collectively contribute 3% (12 counts).

3.5. Region-to-region analysis of MIBiG analysis

The MIBiG comparison analysis also provided an in-depth examination of distinct region-to-region properties of BGCs of 181 types of compounds. Among these, the analysis revealed three most dominant compounds (carotenoids, ectoines, and isorenieratene) that were significantly more abundant than other metabolites in the microbial community (Table S3). Carotenoids emerged as the most prevalent compounds, with 53 identified instances, primarily produced *by Enterobacteriaceae bacterium* DC404, *Myxococcus xanthus* and *Brevundimonas* spp. These pig-

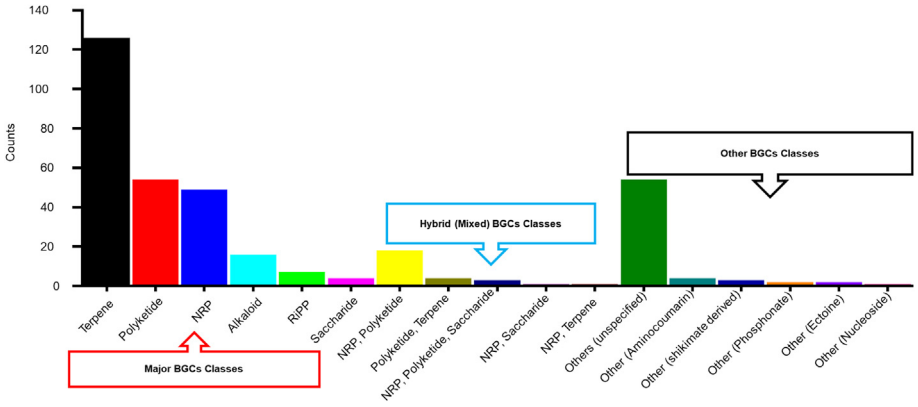


Fig. 5. Classification and abundance of biosynthetic gene cluster (BGC) classes as identified by AntiSMASH analysis. Terpenes represent the most dominant class followed by polyketides and nonribosomal peptide synthetases (NRPs). Hybrid clusters, particularly NRP–Polyketide combinations dominant followed by Polyketide–Terpene, and NRP–Polyketide–Saccharide. Additionally, BGCs grouped under "Others unspecified" are dominant followed by "Others specified" comprising rarer pathways. This distribution highlights the diversity of secondary metabolite biosynthetic potential among the analyzed microbial genomes.

ments likely play crucial roles in photoprotection and oxidative stress resistance in the high-light environment. Ectoine, detected 22 times, was the second most abundant compound, predominantly synthesized by halotolerant genera such as *Methylomicrobium* spp., *Methylophaga* spp. and *Streptomyces* spp., reflecting its importance as an osmoprotectant in saline conditions. Isorenieratene, a specialized carotenoid variant, occurred 11 times and was almost exclusively associated with *Streptomyces* species, suggesting genus-specific adaptations for membrane stability and pigmentation. In addition, fischerindole L, caprutriene, ent-quiannuatene, boleracene, astallatene and (2R,3s,3′s)-2-hydroxyas taxanthin, appearing in 6 instances, represented the most abundant compound and were uniquely produced by different microbes, highlighting their essential roles in microbial survival, ecological interactions, and competitive dynamics within the ecosystem.

3.6. Ketosynthase (KS) and Condensation (C) domains analysis

Analysis of ketosynthase (KS) and Condensation (C) domains using the NaPDoS tool from Chitu Lake MAGs revealed distinct metabolic pathway signatures. Fatty acid synthase (FAS)-associated domains dominated the observed profiles, with FabF_Bacillus_FAS, and FabF_Ecoli_FAS recurring across MAG03, MAG06, MAG07, MAG008, MAG011, and MAG015 (Table S4), indicating conserved roles in essential lipid biosynthesis. Quantitatively, FAS domains were the most abundant (10 occurrences), followed by modular domains (8 hits), while iterative and LCL (linear condensation-like) domains classes exhibited markedly lower frequencies (1 hit each) (Fig. 6a). Notably, several MAGs encoded specialized biosynthetic clusters, with MAG08 displaying exceptional diversity. Its genome harbored two FAS KS domains, six modular KS domains linked to pharmaceutically relevant polyketides (including epothilone (4 hits), nystatin (1 hit), 5-alkenyl-3,3(2H)-furanone (1 hit)), an iterative KS domain associated with the antifungal HSAF (heat-stable antifungal factor) (1 hit) and C domain classified as LCL-type (L-amino acid to L-amino acid condensation) (1hit), implicated in syringomycin biosynthesis, a phytotoxic and antimicrobial lipopeptide (Fig. 6b). In addition, MAG03 and MAG011 contained modular KS domains potentially involved in candicidin and tetronomycin biosynthesis, respectively.

3.7. The identification of putative BGCs using BAGEL4

The BAGEL4 analysis revealed five types of bacteriocin BGCs with varying abundances at their respective genomic locations and functional annotations (Table S5). Among these, sactipeptides were frequently identified and accounted for 56% of the clusters, followed by lasso peptides (27.5%), lanthipeptide class I (5.5%), Zoocin A (5.5%), and patellin 3 TruE1 (5.5%) (Fig. 7). The MAGs assigned to *k_Bacteria*, *c_Gammaproteobacteria*, and *c_Betaproteobacteria* contained a significant number of genes counted 8, 5, 3, and 2, respectively, encoding diverse bacteriocins.

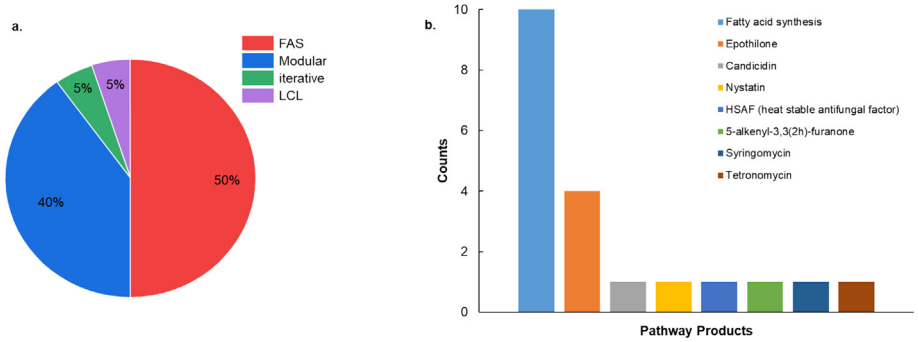


Fig. 6. The pie chart and bar plot demonstrate the distribution of BGCs matches across domain classes and pathway products, comparing NaPDoS database classifications for KS and C domains. Comparative analysis revealed that KS domains are primarily associated with FAS followed by modular domain classes (a), while BGC matches align predominantly with both FAS and epothilone biosynthesis (b). This highlights a divergence between core metabolic functions (e.g., lipid synthesis) and specialized secondary metabolite production (e.g., polyketide-derived compounds like epothilones) within the analyzed datasets.

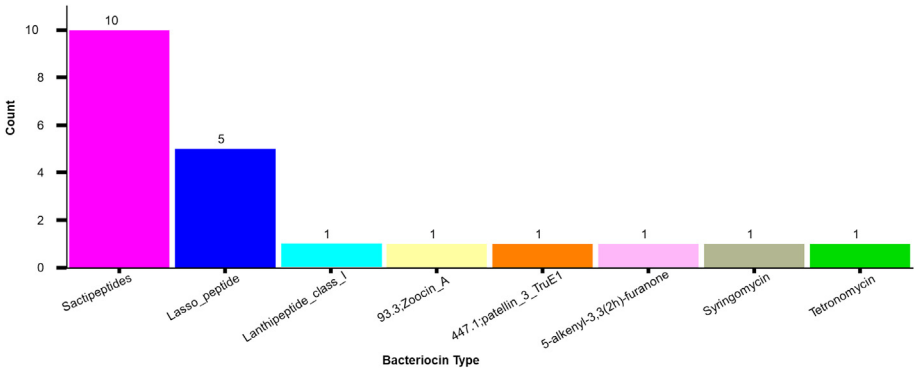


Fig. 7. The bar plot displayed the abundance of the type of bacteriocin cluster recovered from various MAGs using the BAGEL4 tool analysis. The most prevalent anticipated bacteriocin class in Chitu Lake was found to be Sactipeptides followed by Lasso-peptides.

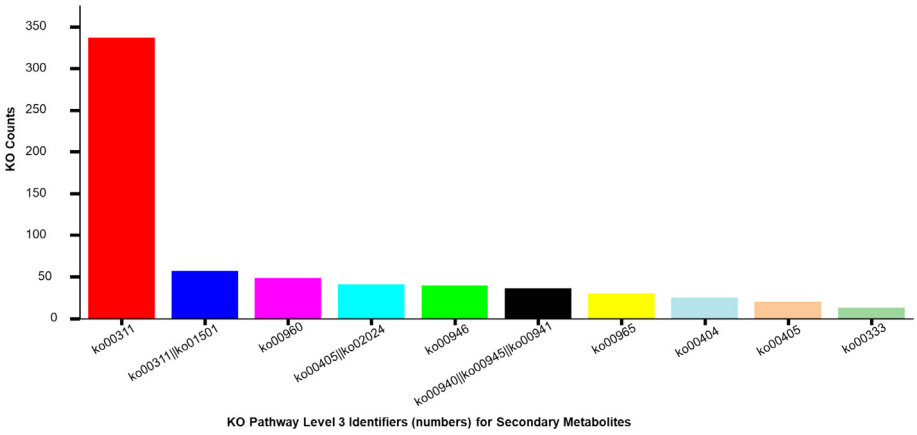


Fig. 8. Top ten dominant secondary metabolite pathways in Chitu Lake based on KEGG Orthology (KO) identifiers at KO_pathway_level 3. Each KO number represents a secondary metabolite produced by a unique biosynthetic pathway indicative of the microbial diversity and metabolic potential within this extreme environment.

3.8. KEGG annotation and analysis

A total of 674 unigenes linked to secondary metabolites were identified using KEGG level 3 (Table S6). The most prominent pathways included penicillin biosynthesis (ko00311) followed by cephalosporin biosynthesis (ko00311||ko01501), tropane, piperidine, and pyridine alkaloid biosynthesis (ko00960), phenazine biosynthesis (ko00405||ko02024), degradation of flavonoids (ko00946), phenylpropanoid, flavonoid, stilbenoid, diarylheptanoid, and gingerol biosynthesis (ko00940||ko00945||ko00941), betalain biosynthesis (ko00965), staurosporine biosynthesis (ko00404), and others (Fig. 8).

3.9. RASTk-driven discovery of secondary metabolite diversity in metagenome-assembled genomes

The RASTk annotation of MAGs from Chitu Lake identified nearly 181 secondary metabolite biosynthesis pathways across ten bins (MAG03, MAG04, MAG06, MAG07, MAG08, MAG09, MAG011, MAG013, MAG015, and MAG020) (Table S7). The most prevalent pathways included 3-amino-5-hydroxybenzoic acid (AHBA) synthesis, a key precursor for ansamycin antibiotics. In addition, chorismate synthesis (linked to aromatic amino acid metabolism) and anaerobic benzoyl-CoA-mediated aromatic degradation. Additional pathways such as alkane biosynthesis, colicin V and bacteriocin production, fatty acid synthesis, fatty acid biosynthesis FASII, 2-aminophenol metabolism, Benzoyl-CoA anaerobic degradation pathway and phenylacetyl-CoA catabolism were identified. In addition, enzymes critical to these processes, including 3-dehydroquinate dehydratase II, acetyl-CoA carboxyl transferase, and shikimate 5-dehydrogenase I alpha, were annotated across MAGs. Functional categorization highlighted roles in secondary metabolism, amino acid derivatives, membrane transport, and fatty acid biosynthesis, with genomic features varying in contig counts (134–550), gene totals (1,253–2,716), and GC content (49.87–74.05%), suggesting metabolic and genomic diversity among microbial populations.

4. Discussion

The microbial community analysis of Lake Chitu, based on shotgun metagenomic sequencing, indicated a predominant presence of bacteria (89%), followed by archaea (4%), viruses (0.2%), and eukaryota (0.04%). All the microbial communities in extreme biomes exhibit remarkable biochemical adaptations that enable survival in extreme conditions, including high salinity, alkalinity, and oxidative stress [7,9]. These adaptations are largely driven by biosynthetic gene clusters (BGCs), which encode secondary metabolites with crucial ecological, biotechnological, and industrial roles [20]. This study designed a metagenomic pipeline to explore secondary metabolite BGCs from the underexplored microbiomes of Lake Chitu, shedding lights into their chemical potential as well as their adaptation mechanisms within an extreme habitat. The physical and chemical characteristics of the Lake indicate an

extremely saline and alkaline environment. It has high turbidity (33.35 NTU), a strongly basic pH (10.56), and remarkably high salinity (46,500 mg/L). Additionally, nutrient levels, including nitrate, nitrite, and phosphate, are elevated, potentially supporting the growth of specialized microbial communities adapted to this extreme environment [34].

Thus, shotgun metagenomic sequencing revealed a diverse range of microorganisms, with bacteria overrepresented by *Pseudomonadota, Actinomycetota,* and *Gammatimonadota*, while archaea were primarily dominated by *Euryarchaeota.* This aligns with findings from other soda lake ecosystems [1,35,36], reinforcing the concept of convergent evolution in extreme habitats and the production of natural products [37]. Genomic mining, refined MAGs were analyzed using antiSMASH 7.0, which revealed a wide range of BGCs, including terpenes, ranthipeptides, RiPPs, ectoine, lanthipeptides, NRPS-like, aryl polyenes, and phosphonates. These BGCs were linked to various bacterial taxa, including *c_Gammaproteobacteria*, *c_Betaproteobacteria*, and *k_bacteria* and highlighting their potential for producing diverse bioactive compounds [38]. Notably, *c_Gammaproteobacteria* exhibited high BGC diversity, likely due to its metabolic versatility and adaptations to extreme conditions. These secondary metabolites are vital for microbial defense, UV protection, stress response, and symbiotic interactions [39].

The presence of FAS-related KS domains across multiple MAGs with modular and iterative KS domains indicates potential secondary metabolite production including antifungal (candicidin, nystatin, HSAF), anticancer (epothilone), and antibacterial (tetronomycin, syringomycin) compounds. In addition, the high diversity in *Gammaproteobacteria* taxa suggests a rich secondary metabolite biosynthetic potential. Therefore, the investigation of microbial secondary metabolite biosynthetic gene clusters from Chitu Lake revealed a wealth of genetic resources with biotechnological, pharmaceutical, industrial, and environmental applications. Notably, terpene biosynthesis emerged as a hallmark of Chitu Lake's BGC repertoire, a finding consistent with their putative role in mitigating membrane destabilization under osmotic stress [40,41]. In addition, the prevalence of these compounds may reflect selective pressures favoring structural simplicity and functional versatility, traits that confer ecological advantage in hypersaline niches (including antimicrobial properties, signaling functions, and oxidative stress protection [42]. RiPPs (another significant class identified) exhibit remarkable structural and functional diversity, with subclasses such as lasso peptides, lanthipeptides, thiopeptides, sactipeptides, and linear thiazole/oxazole-containing peptides, and are prominently represented, echoing observations in Lake Afdera's bacteriocin-rich microbiome [19]. Such peptides are postulated to mediate microbial antagonism and quorum sensing [43,44], suggesting their dual utility in ecological competition and synthetic biology applications. Of particular interest is the identification of ranthipeptide clusters reliant on radical S-adenosylmethionine (rSAM) enzymes [45], a class of biocatalysts distinguished by their capacity to initiate radical-based modifications. These enzymes, which are underrepresented in conventional bioprospecting pipelines, may constitute a reservoir of novel catalytic mechanisms with applications in antibiotic development [46]. The coexistence of such systems with stress-adaptive BGCs—including ectoine synthases in bacteria which survive under extreme conditions of salinity, drought, irradiation, pH, and temperature [47] and hydrogen cyanide (HCN) biosynthetic clusters (encoded by *hcnABC* gene cluster) in *Pseudomonadota* (*Pseudomonas* spp., *P. vancouverensis)* and *Bacillota* (*Bacillus* spp.*)* [48,49] suggests a metabolic nexus wherein secondary metabolite production is inextricably linked to environmental resilience. This synergy is further exemplified by aryl polyenes (APEs), which contribute to biofilm formation and oxidative stress tolerance [50], and homoserine lactones, signal molecules, which regulate quorum-sensing cascades in Gram-negative taxa [51].

KEGG pathway analysis revealed key biosynthetic pathways for secondary metabolites, including phenylpropanoids, flavonoids, β-lactams, phenazines, and antibiotics such as streptomycin, penicillin, cephalosporin, and monobactam, highlighting their pharmaceutical relevance [19,21,52,53]. The coexistence of antibiotic synthesis genes (e.g., phzS, EC 1.14.13.218) with resistance markers (e.g., β-lactamases) suggests that microbial competition drives both metabolic innovation and resistance mechanisms. Beyond antibiotics, the presence of flavonoids, phenylpropanoids, and stilbenoids indicates potential applications in medicine, cosmetics, and environmental management, while pathways for tropane, piperidine, and pyridine alkaloids, along with acarbose, validamycin, and staurosporine, suggest promising sources of novel antidiabetic and anticancer compounds [9].

RAST annotation also provided the prevalence of AHBA synthesis genes originating from the shikimate-type biosynthetic pathway across all bins and suggests that Chitu Lake's microbial communities may produce ansamycin-like antibiotics, which could influence microbial interactions and competitive dynamics [54]. In addition, the widespread colicin V and Bacteriocin clusters reinforce the ecological importance of antimicrobial strategies in shaping community structure. Furthermore, the detection of alkane biosynthesis genes suggests microbial adaptation to environmental stressors, possibly through hydrocarbon-mediated membrane stabilization. Pathways such as benzoyl-CoA anaerobic degradation and phenylacetyl-CoA catabolism highlight the community's capacity for aromatic compound breakdown, critical for nutrient cycling in lake ecosystems, particularly in low-oxygen sediments. The unique presence of gentisate 1,2-dioxygenase in MAG07 expands the functional repertoire for aromatic degradation, indicating niche-specific metabolic specialization. The integration of secondary metabolism with subsystems like fatty acid synthesis and cofactor production reflects metabolic versatility, enabling survival in fluctuating conditions [55]. These findings position Chitu Lake's microbiota as a reservoir of biosynthetic potential, with implications for understanding microbial ecology, bioremediation, and biotechnological applications in secondary metabolite discovery.

The biotechnological potential of Chitu Lake's microbial diversity is vast. Terpenes and aryl polyenes, with their roles in UV protection and biofilm stabilization [50], offer promising applications in cosmetics and pharmaceuticals [41,47]. RiPPs, including sactipeptides and lasso peptides, feature structural motifs well-suited for antimicrobial and antitumor therapies [44,56]. Phosphonates and β-lactones, which are relatively rare in conventional screening libraries, represent untapped sources of enzyme inhibitors [57,58]. However, unlocking this biosynthetic potential requires overcoming challenges related to cryptic gene expression. Strategies such as heterologous production in model organisms (e.g., *Streptomyces* spp. or *Escherichia coli*) [15,46] and activity-guided fractionation are crucial for harnessing these metabolites.

Beyond pharmaceuticals, the identification of HCN biosynthetic clusters presents ecological and industrial relevance. While toxic, HCN-producing microbes could serve as bio-based pesticides and fungicides [48,49]. Additionally, the presence of RiPP Recognition Element-containing (RRE-containing) BGCs suggests a promising avenue for novel bioactive compound discovery, including antimicrobial agents [59,60]. Similarly, Phosphonate natural products (PNPs), known for their potent bioactivity, have broad-spectrum antimicrobial and agricultural applications [57,58]. Many phos-

phonates originate from *Streptomyces* spp., *Bacillus* spp., *Actinobacteria* spp., and *Pseudomonas* spp., further demonstrating their industrial and medical importance [61].

Overall, Chitu Lake harbors a highly specialized microbial community shaped by extreme environmental conditions, with potential genes crucial for the survival and adaptation of extremophiles. However, to substantiate this claim, further studies are needed to determine whether these genes are significantly enriched or over-represented in the genomes compared to other microorganisms. Moreover, the wide range of BGCs identified in this study suggests promising biotechnological applications, particularly in the development of antibiotics, enzyme inhibitors, and other bioactive compounds. Nevertheless, further research is required to optimize expression strategies for cryptic BGCs and leverage synthetic biology approaches to maximize their potential. In addition, while this study focuses on microbial communities in lakeside soil and mud samples, it does not encompass variations across different lake strata including the lake bottom. This represents a limitation in capturing the full microbial diversity within the lake system. Future studies should consider comparative analyses across various sediment depths to provide a more comprehensive understanding of microbial adaptations and biosynthetic potential in alkaline soda lakes.

5. Conclusions

This shotgun metagenomic study provides a comprehensive analysis of the biosynthetic potential within the microbial community of Chitu Lake, a high-salinity and alkaline environment. The microbiome was dominated by bacterial phyla such as *Pseudomonadota*, *Actinomycetota*, and *Gemmatimonadota,* with Archaea mainly represented by the *Euryarchaeota* phylum. Genomic mining uncovered a diverse array of BGCs, underscoring the microbial community's capacity to produce bioactive secondary metabolites. Key BGCs identified include those encoding terpenes, PKS/NRPS hybrids, RiPP-like compounds, RRE-containing peptides, ectoine, lanthipeptides, and beta-lactones. The identification of these BGCs underscores the potential of Chitu Lake's microbial community for bioprospecting secondary metabolites across various sectors like pharmaceutical, agricultural, and industrial applications. These findings also provide valuable insights into the adaptive mechanisms of extremophilic microorganisms in Chitu Lake, enabling their survival under extreme conditions such as high salinity, alkalinity, heavy metal exposure, radiation, and desiccation. Future research should focus on the functional characterization of these biosynthetic gene clusters, with an emphasis on elucidating their biosynthetic pathways and underlying molecular mechanisms. Additionally, targeted metabolomic approaches, such as LC-MS, are recommended for the isolation and functional validation of bioactive compounds produced by the promising microbial taxa identified in this study. Moreover, investigating the ecological roles of these metabolites and their production dynamics will further enhance our understanding of microbial diversity and its biotechnological applications in Chitu Lake.

CRediT authorship contribution statement

Gessesse Kebede Bekele: Visualization, Conceptualization, Methodology, Investigation, Writing – original draft, Formal analysis. **Ermias Sissay Balcha:** Writing – review & editing, Formal analysis. **Abu Feyisa Meka:** Data curation, Writing – review & editing. **Eskedar Getachew Assefa:** Writing – review & editing, Visualization. **Ebrahim M. Abda:** Validation, Data curation, Supervision, Resources, Writing – review & editing, Project administration. **Fassil Assefa Tuji:** Writing – review & editing, Supervision. **Mesfin Tafesse Gemeda:** Project administration, Writing – review & editing, Supervision.

Financial support

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability

The raw shotgun metagenomic sequencing data for the Chitu Lake sample (AastuCL1) have been deposited in the National Center for Biotechnology Information (NCBI) database under GenBank accession number **PRJNA1081624** and are accessible at: https://dataview.ncbi.nlm.nih.gov/object/PRJNA1081624?reviewer=7o82me3g5j8027i3lloighgkj6.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank the Addis Ababa Science and Technology University (AASTU) for its financial and technical assistance. Additionally, the authors would like to express their sincere gratitude to Novogen metagenomic sequencing facilities (HK) in Singapore for their assistance in processing shotgun metagenomic sequencing data and delivering the resulting information. We also extend our gratitude to the ICT Directorate and its staff at AASTU for providing the servers used for data storage and analysis.

Supplementary material

https://doi.org/10.1016/j.ejbt.2025.06.002.

References

[1] Ali Y, Simachew A, Gessesse A. Diversity of culturable alkaliphilic nitrogen-fixing bacteria from a Soda Lake in the East African Rift Valley. Microorganisms 2022;10(9):1760. https://doi.org/10.3390/microorganisms10091760. PMid: 36144362.

[2] Cavicchioli R, Amils R, Wagner D, et al. Life and applications of extremophiles. Environ Microbiol 2011;13(8):1903–7. https://doi.org/10.1111/j.1462-2920.2011.02512.x. PMid: 22236328.

[3] Gunde-Cimerman N, Plemenitaš A, Oren A. Strategies of adaptation of microorganisms of the three domains of life to high salt concentrations. FEMS Microbiol Rev 2018;42(3):353–75. https://doi.org/10.1093/femsre/fuy009. PMid: 29529204.

[4] Edbeib MF, Wahab RA, Huyop F. Halophiles: Biology, adaptation, and their role in decontamination of hypersaline environments. World J Microb Biot 2016;32:135. https://doi.org/10.1007/s11274-016-2081-9. PMid: 27344438.

[5] Corral P, Amoozegar MA, Ventosa A. Halophiles and their biomolecules: Recent advances and future applications in biomedicine. Mar Drugs 2019;18(1):33. https://doi.org/10.3390/md18010033. PMid: 31906001.

[6] Ramawat KG, Goyal S. Co-evolution of secondary metabolites during biological competition for survival and advantage: An overview. In: Mérillon JM, Ramawat K, editors. Co-evolution of secondary metabolites. Reference Series in Phytochemistry. Cham: Springer; 2020. p. 3–17. https://doi.org/10.1007/978-3-319-96397-6_45.

[7] Rutledge PJ, Challis GL. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. Nat Rev Microbiol 2015;13(8):509–23. https://doi.org/10.1038/nrmicro3496. PMid: 26119570.

[8] Chen R, Wong HL, Kindler GS, et al. Discovery of an abundance of biosynthetic gene clusters in Shark Bay microbial mats. Front Microbiol 2020;11:1950. https://doi.org/10.3389/fmicb.2020.01950. PMid: 32973707.

[9] Vaishnav P, Demain AL. Unexpected applications of secondary metabolites. Biotechnol Adv 2011;29(2):223–9. https://doi.org/10.1016/j.biotechadv.2010.11.006. PMid: 21130862.

[10] Yadav AN, Kour D, Rana KL, et al. Metabolic engineering to synthetic biology of secondary metabolites production. In: Gupta VK, Pandey A, editors. New and

future developments in microbial biotechnology and bioengineering. Elsevier; 2019. p. 279–320. https://doi.org/10.1016/B978-0-444-63504-4.00020-7.
[11] Blin K, Shaw S, Augustijn HE, et al. antiSMASH 7.0: New and improved predictions for detection, regulation, chemical structures and visualisation. Nucleic Acids Res 2023;51(W1):W46–50. https://doi.org/10.1093/nar/gkad344. PMid: 37140036.
[12] Doroghazi JR, Albright JC, Goering AW, et al. A roadmap for natural product discovery based on large-scale genomics and metabolomics. Nat Chem Biol 2014;10:963–8. https://doi.org/10.1038/nchembio.1659. PMid: 25262415.
[13] van Heel AJ, de Jong A, Song C, et al. BAGEL4: A user-friendly web server to thoroughly mine RiPPs and bacteriocins. Nucleic Acids Res 2018;46(W1): W278–81. https://doi.org/10.1093/nar/gky383. PMid: 29788290.
[14] Helfrich EJ, Lin GM, Voigt CA, et al. Bacterial terpene biosynthesis: Challenges and opportunities for pathway engineering. Beilstein J Org Chem 2019;15 (1):2889–906. https://doi.org/10.3762/bjoc.15.283. PMid: 31839835.
[15] Shi J, Xu X, Liu PY, et al. Discovery and biosynthesis of guanipiperazine from a NRPS-like pathway. Chem Sci 2021;12(8):2925–30. https://doi.org/10.1039/D0SC06135B. PMid: 34164059.
[16] Han SW, Won HS. Advancements in the application of ribosomally synthesized and post-translationally modified peptides (RiPPs). Biomolecules 2024;14 (4):479. https://doi.org/10.3390/biom14040479. PMid: 38672495.
[17] Chali K, Belay Z, Bacha K. *In vitro* bioassay of antibacterial and antifungal activity studies of Actinomycetes from Soda Lakes of Ethiopia. Sys Rev Pharm 2022;13(8):798–807.
[18] Mulugeta K, Kamaraj M, Tafesse M, et al. Biomolecules from *Serratia* sp. CS1 indigenous to Ethiopian natural alkaline lakes: Biosurfactant characteristics and assessment of compatibility in a laundry detergent. Environ Monit Assess 2022;194(12):873. https://doi.org/10.1007/s10661-022-10533-7. PMid: 36227369.
[19] Balcha ES, Macey MC, Gemeda MT, et al. Mining the microbiome of Lake Afdera to gain insights into microbial diversity and biosynthetic potential. FEMS Microbes 2024;5:xtae008. https://doi.org/10.1093/femsmc/xtae008. PMid: 385606258.
[20] Mantri SS, Negri T, Sales-Ortells H, et al. Metagenomic sequencing of multiple soil horizons and sites in close vicinity revealed novel secondary metabolite diversity. mSystems 2021;6(5):e01018–e1021. https://doi.org/10.1128/msystems.01018-21. PMid: 34636675.
[21] Loke KK, Rahnamaie-Tajadod R, Yeoh CC, et al. Transcriptome analysis of *Polygonum minus* reveals candidate genes involved in important secondary metabolic pathways of phenylpropanoids and flavonoids. PeerJ 2017;5:e2938. https://doi.org/10.7717/peerj.2938. PMid: 28265493.
[22] Verma SK, Singh H, Sharma PC. An improved method suitable for isolation of high-quality metagenomic DNA from diverse soils. 3 Biotech 2017;7:171. https://doi.org/10.1007/s13205-017-0847-x. PMid: 28660454.
[23] Perez-Mon C, Qi W, Vikram S, et al. Shotgun metagenomics reveals distinct functional diversity and metabolic capabilities between 12 000-year-old permafrost and active layers on Muot da Barba Peider (Swiss Alps). Microb Genom 2021;7(4):000558. https://doi.org/10.1099/mgen.0.000558. PMid: 33848236.
[24] Qin N, Yang F, Li A, et al. Alterations of the human gut microbiome in liver cirrhosis. Nature 2014;513(7516):59–64. https://doi.org/10.1038/nature13568. PMid: 25079328.
[25] Wang B, Wang X, Wang Z, et al. Comparative metagenomic analysis reveals rhizosphere microbial community composition and functions help protect grapevines against salt stress. Front Microbiol 2023;14:1102547. https://doi.org/10.3389/fmicb.2023.1102547. PMid: 36891384.
[26] Yang J, Li W, Teng D, et al. Metagenomic insights into microbial community structure, function, and salt adaptation in saline soils of arid land, China. Microorganisms 2022;10(11):2183. https://doi.org/10.3390/microorganisms10112183. PMid: 36363774.
[27] Sieber CM, Probst AJ, Sharrar A, et al. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nat Microbiol 2018;3 (7):836–43. https://doi.org/10.1038/s41564-018-0171-1. PMid: 29807988.
[28] Keller-Costa T, Kozma L, Silva SG, et al. Metagenomics-resolved genomics provides novel insights into chitin turnover, metabolic specialization, and niche partitioning in the octocoral microbiome. Microbiome 2022;10(1):151. https://doi.org/10.1186/s40168-022-01343-7. PMid: 36138466.
[29] Parks DH, Imelfort M, Skennerton CT, et al. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 2015;25(7):1043–55. https://doi.org/10.1101/gr.186072.114. PMid: 25977477.
[30] Hao DC, Zhang CR, Xiao PG. The first *Taxus* rhizosphere microbiome revealed by shotgun metagenomic sequencing. J Basic Microbiol 2018;58(6):501–12. https://doi.org/10.1002/jobm.201700663. PMid: 29676472.
[31] Terlouw BR, Blin K, Navarro-Munoz JC, et al. MIBiG 3.0: A community-driven effort to annotate experimentally validated biosynthetic gene clusters. Nucleic Acids Res 2023;51(D1):D603–10. https://doi.org/10.1093/nar/gkac1049. PMid: 36399496.
[32] Walker AS, Clardy J. A machine learning bioinformatics method to predict biological activity from biosynthetic gene clusters. J Chem Inf Model 2021;61 (6):2560–71. https://doi.org/10.1021/acs.jcim.0c01304. PMid: 34042443.
[33] Klau LJ, Podell S, Creamer KE, et al. The Natural Product Domain Seeker version 2 (NaPDoS2) webtool relates ketosynthase phylogeny to biosynthetic function. J Biol Chem 2022;298(10):102480. https://doi.org/10.1016/j.jbc.2022.102480. PMid: 36108739.
[34] Martínez-Espinosa RM. Microorganisms and their metabolic capabilities in the context of the biogeochemical nitrogen cycle at extreme environments. Int J Mol Sci 2020;21(12):4228. https://doi.org/10.3390/ijms21124228. PMid: 32545812.
[35] Vavourakis CD, Andrei AS, Mehrshad M, et al. A metagenomics roadmap to the uncultured genome diversity in hypersaline soda lake sediments. Microbiome 2018;6:168. https://doi.org/10.1186/s40168-018-0548-7. PMid: 30231921.
[36] Paul Antony C, Kumaresan D, Hunger S, et al. Microbiology of Lonar Lake and other soda lakes. ISME J 2013;7(3):468–76. https://doi.org/10.1038/ismej.2012.137. PMid: 23178675.
[37] Valentine DL. Adaptations to energy stress dictate the ecology and evolution of the Archaea. Nat Rev Microbiol 2007;5(4):316–23. https://doi.org/10.1038/nrmicro1619. PMid: 17334387.
[38] Parrilli E, Tedesco P, Fondi M, et al. The art of adapting to extreme environments: The model system *Pseudoalteromonas*. Phys Life Rev 2021;36:137–61. https://doi.org/10.1016/j.plrev.2019.04.003. PMid: 31072789.
[39] Cho SH, Lee E, Ko SR, et al. Elucidation of the biosynthetic pathway of vitamin B groups and potential secondary metabolite gene clusters via genome analysis of a marine bacterium *Pseudoruegeria* sp. M32A2M. J Microbiol Biotechnol 2020;30(4):505–14. https://doi.org/10.4014/jmb.1911.11006. PMid: 31986560.
[40] Costa JAV, Cassuriaga APA, Moraes L, et al. Biosynthesis and potential applications of terpenes produced from microalgae. Bioresour Technol Rep 2022;19:101166. https://doi.org/10.1016/j.biteb.2022.101166.
[41] Fan M, Yuan S, Li L, et al. Application of terpenoid compounds in food and pharmaceutical products. Fermentation 2023;9(2):119. https://doi.org/10.3390/fermentation9020119.
[42] Avalos M, Garbeva P, Vader L, et al. Biosynthesis, evolution and ecology of microbial terpenoids. Nat Prod Rep 2022;39(2):249–72. https://doi.org/10.1039/D1NP00047K. PMid: 34612321.
[43] Zhong Z, He B, Li J, et al. Challenges and advances in genome mining of ribosomally synthesized and post-translationally modified peptides (RiPPs). Synth Syst Biotechnol 2020;5(3):155–72. https://doi.org/10.1016/j.synbio.2020.06.002. PMid: 32637669.
[44] Li C, Alam K, Zhao Y, et al. Mining and biosynthesis of bioactive lanthipeptides from microorganisms. Front Bioeng Biotechnol 2021;9:692466. https://doi.org/10.3389/fbioe.2021.692466. PMid: 34395400.
[45] Chen Y, Yang Y, Ji X, et al. The SCIFF-derived lanthipeptides participate in quorum sensing in solventogenic clostridia. J Biotechnol 2020;200:136–44. https://doi.org/10.1002/biot.202000136. PMid: 32713052.
[46] Precord TW, Mahanta N, Mitchell DA. Reconstitution and substrate specificity of the thioether-forming radical *S*-adenosylmethionine enzyme in freyrasin biosynthesis. ACS Chem Biol 2019;14(9):1981–9. https://doi.org/10.1021/acschembio.9b00457. PMid: 31449382.
[47] Bilstein A, Heinrich A, Rybachuk A, et al. Ectoine in the treatment of irritations and inflammations of the eye surface. Biomed Res Int 2021;2021(1):8885032. https://doi.org/10.1155/2021/8885032. PMid: 33628826.
[48] Sehrawat A, Sindhu SS, Glick BR. Hydrogen cyanide production by soil bacteria: Biological control of pests and promotion of plant growth in sustainable agriculture. Pedosphere 2022;32(1):15–38. https://doi.org/10.1016/S1002-0160(21)60058-9.
[49] Tamang P, Upadhaya A, Paudel P, et al. Mining biosynthetic gene clusters of *Pseudomonas vancouverensis* utilizing whole genome sequencing. Microorganisms 2024;12(3):548. https://doi.org/10.3390/microorganisms12030548. PMid: 38543599.
[50] Johnston I, Osborn LJ, Markley RL, et al. Identification of essential genes for *Escherichia coli* aryl polyene biosynthesis and function in biofilm formation. NPJ Biofilms Microbe 2021;7(1):56. https://doi.org/10.1038/s41522-021-00226-3. PMid: 34215744.
[51] Ye YQ, Hao ZP, Yue YY, et al. Characterization of *Kordiimonas marina* sp. nov. and *Kordiimonas laminariae* sp. nov. and comparative genomic analysis of the genus *Kordiimonas*, a marine-adapted taxon. Front Mar Sci 2022;9:919253. https://doi.org/10.3389/fmars.2022.919253.
[52] Reshi ZA, Ahmad W, Lukatkin AS, et al. From nature to lab: A review of secondary metabolite biosynthetic pathways, environmental influences, and *in vitro* approaches. Metabolites 2023;13(8):895. https://doi.org/10.3390/metabo13080895. PMid: 37623839.
[53] Alam K, Islam MM, Li C, et al. Genome mining of *Pseudomonas* species: Diversity and evolution of metabolic and biosynthetic potential. Molecules 2021;26(24):7524. https://doi.org/10.3390/molecules26247524. PMid: 34946606.
[54] Taft F, Brünjes M, Floss HG, et al. Highly active ansamitocin derivatives: Mutasynthesis using an AHBA-blocked mutant. Chembiochem 2008;9 (7):1057–60. https://doi.org/10.1002/cbic.200700742. PMid: 18381586.
[55] Fernández-Martínez LT, Hoskisson PA. Expanding, integrating, sensing and responding: The role of primary metabolism in specialised metabolite production. Curr Opin Microbiol 2019;51:16–21. https://doi.org/10.1016/j.mib.2019.03.006. PMid: 31005749.
[56] Bruna P, Núñez-Montero K, Contreras MJ, et al. Biosynthetic gene clusters with biotechnological applications in novel Antarctic isolates from *Actinomycetota*.

Appl Microbiol Biot 2024;108(1):325. https://doi.org/10.1007/s00253-024-13154-x. PMid: 38717668.
[57] Kayrouz CM, Zhang Y, Pham TM, et al. Genome mining reveals the phosphonoalamide natural products and a new route in phosphonic acid biosynthesis. ACS Chem Biol 2020;15(7):1921–9. https://doi.org/10.1021/acschembio.0c00256. PMid: 32484327.
[58] Zhang Y, Chen L, Wilson JA, et al. Valinophos reveals a new route in microbial phosphonate biosynthesis that is broadly conserved in nature. J Am Chem Soc 2022;144(22):9938–48. https://doi.org/10.1021/jacs.2c02854. PMid: 35617676.
[59] Li H, Ding W, Zhang Q. Discovery and engineering of ribosomally synthesized and post-translationally modified peptide (RiPP) natural products. RSC Chem Biol 2024;5:90–108. https://doi.org/10.1039/D3CB00172F. PMid: 38333193.
[60] Ren H, Dommaraju SR, Huang C, et al. Genome mining unveils a class of ribosomal peptides with two amino termini. Nat Commun 2023;14(1):1624. https://doi.org/10.1038/s41467-023-37287-1. PMid: 36959188.
[61] Wilson J, Cui J, Nakao T, et al. Discovery of antimicrobial phosphonopeptide natural products from *Bacillus velezensis* by genome mining. Appl Environ Microbiol 2023;89(6):e00338–e423. https://doi.org/10.1128/aem.00338-23. PMid: 37377428.