MedCrave
Step into the World of Research

Journal of Microbiology & Experimentation

Research Article

Open Access

CrossMark

Metagenomic analysis of bacterial communities inhabiting Lake Natron waters in Arusha, Tanzania

Abstract

Metagenomics studies which involve extraction of genetic material directly from environmental samples and sequencing in order to gain insight into the microbial world and the relationships that exist between microbes and their surroundings are very few especially in extreme environments such as soda lakes. Lake Natron is one of the soda lakes found in the eastern branch of East African Gregory Rift system, in Arusha region, Tanzania and is among the paramount extreme environments characterized with high alkaline in nature. The present study used full length 16S rRNA reads sequenced through PacBio system to reveal the first metagenomic snapshots of bacterial communities from the water samples of Lake Natron. The results showed an extremity of physicochemical parameters (high salinity and pH) and high number and diversity of bacterial species. Bacterial community was dominated by Proteobacteria, Firmicutes, Bacteroidota and Actinomycetota at Phylum level; Alphaproteobacteria, Gammaproteobacteria, Bacteroidia and Bacilli at Class level; Oceanibaculaceae, Devosiaceae, Rhizobiaceae, Izemoplasmataceae and ML635J at Family level; *Oceanibaculum, Devosia, Allorhizobium* and *Izimaplasma* at Genus level. High abundances of unknown and uncultured species were also recorded suggesting presence of new taxa. Consequently, there is high species diversity and variations with a notable dominance and close variation of some bacteria species among the sampling points of Lake Natron. Therefore, due to the exceptional bacterial diversity in Lake Natron, the present study highlights the need for continued exploration to uncover new taxa, understand the ecological significance of these bacteria, and harness their biotechnological potential.

Keywords: metagenomics, pacBio, 16S rRNA gene, soda lake, Lake Natron, Tanzania

Volume 12 Issue 2 - 2024

Sadikiel E Kaale,[1,2] Robert S Machang'u,[3] Thomas J Lyimo[1]

[1]Department of Molecular Biology and Biotechnology, University of Dar es Salaam, Tanzania
[2]School of Life Sciences and Biomedical Engineering, Nelson Mandela Institute of Science and Technology, Tanzania
[3]Department of Microbiology, Saint Francis University College of Health and Allied Sciences, Tanzania

Correspondence: Sadikiel E. Kaale, Department of Molecular Biology and Biotechnology, University of Dar es Salaam, Dar es Salaam, Tanzania. P.O. Box 35179 - Dar es Salaam, Tanzania, Tel +255655238856, Email sadikiel.kaal@nm-aist.ac.tz

Received: March 29, 2024 | **Published:** April 15, 2024

Abbreviations: NGS, Next generation sequencing; EC, electrical conductivity; TDS, total dissolved salts; gDNAs, genomic DNAs; QIMME2, Quantitative Insights Into Microbial Ecology; CDI, community diversity indices

Introduction

Metagenomics is the study involving the extraction of DNA or RNA directly from environmental organisms. It is frequently used for studying particular communities of microorganisms in various habitats. To directly access the genetic information of vast populations of species in their entire communities, metagenomics employs an array of next generation sequencing (NGS) technologies and analysis through bioinformatics tools.[1] One of the NGS technologies is PacBio system which has been used recently in the sequencing of full-length 16S rRNA reads. The technique refined the classification of bacterial communities up to the species level.[2] These technologies have enhanced the understanding of microorganisms, particularly bacteria, in their natural habitats especially in extreme and/or unexplored environments. A habitat is considered to be an extreme environment if it has unusual and difficult-to-survive conditions, such as extremely high or low salinity, pH, and temperatures.[3]

Extreme environments in Tanzania include the soda lakes along the Gregory Rift Valley which possess alkaline characteristics, difficult for different life forms to survive. Their natural settings maintain a pH of 9, or above, consistently (compared to usual pH values of 6 or 7), frequently, or for extended periods of time.[4] Lake Natron is an alkaline or salt lake found in the eastern branch of the East African Gregory rift in Tanzania. The lake is the major breeding site for Africa's lesser flamingoes, and the alkalinity of the water is mainly caused by leaching through the nearby Mount Ol- Doinyo Lengai volcanic materials.[5] It contains extreme conditions as soda ashes which can be seen at the shoreline of the lake. Also, the water pH can rise up to 12 during the dry seasons.[4] Strikingly, Lake Natron hosts a variety of bacteria with potential biotechnological importance as discussed in our previous work in the lake.[6]

There are many hindrances to bacterial studies in the soda lakes of Tanzania, such as limited genomic information of the microorganisms and their characterization. This is because most of the lakes are found in hard to reach areas that limit sampling.[7] Also, the bias of culture-dependent techniques is another hindrance as the diversity is not entirely captured by conservative and conventional culture-dependent approaches, which for this case it requires the setting of special culture conditions.[8] Considering the limited explorations done and relatively narrow scope of metagenomic studies on the bacterial communities of Tanzania's Lake Natron and other water bodies in general dos Santos et al.,[9] this study explored, for the first time, the diversity of bacteria species inhabiting the waters of Lake Natron by using full-length 16S rRNA reads sequenced by a PacBio Sequel IIe system.

Materials and Methods

Samplings and measurement of environmental parameters

Sampling was done in the western side of Lake Natron; a soda lake is located in Arusha region, northern of Ngorongoro and Loliondo Districts, Tanzania (Figure 1). High levels of evaporation have left behind natron sodium carbonate decahydrate (natron) and sodium sesquicarbonate dihydrate (trona). The caustic nature of the lake's water with a pH over 10 (greater than 12 especially in dry conditions) can burn the eyes and skin of those animals not adapted by them. Typically the lake unique alkaline ecosystem supports only flamingos and wetland birds; in fact it is the crucial breeding site in the east

Africa region for Africa's lesser flamingoes.[4] These characteristics stipulate Lake Natron as one of the most extreme soda lake and this calls for more research of the discovery of its inhabitants as it may be beneficial to the scientific community.

Water samples were aseptically taken from 10 random points in the shoreline of west coast of the Lake Natron as shown in Figure 1 and coordinates in Table 1. Triplicate samples of water (200 mL) were collected at the surface in sterile 250 mL bottles. The bottles were tightly capped and transported in a cool box (packed with ice cubes) to the Molecular Biology Laboratory of the University of Dar es Salaam. Immediately upon arrival in the laboratory (approximately 1 day after sampling), the samples were processed for the DNA extraction. Multi-Parameter meter (Hach 40HQD) was used on-site to measure the physicochemical parameters and average measurements, in triplicates. The parameters measured were Temperature, pH, salinity, electrical conductivity (EC) and total dissolved salts (TDS). (Table 1)

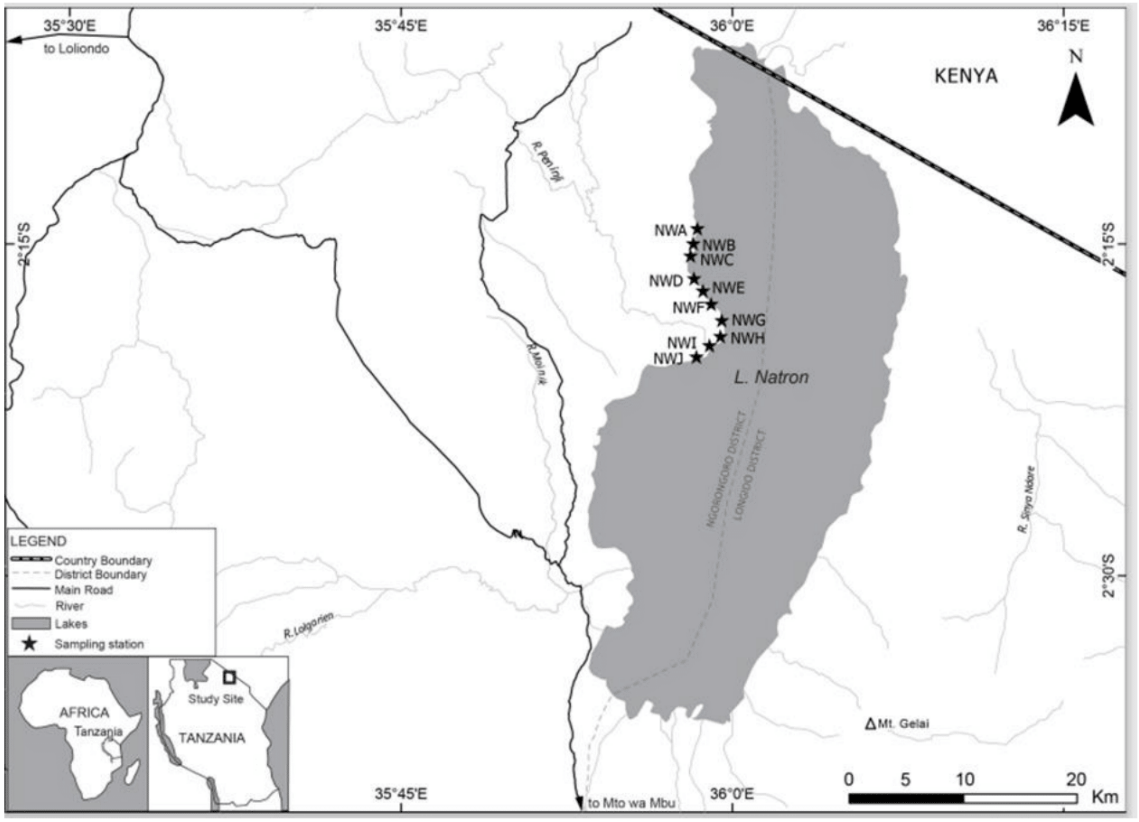

Figure 1 Map of Lake Natron (shaded) in Arusha, Tanzania, showing sampling points (area with asterices) at the shoreline of the lake.

Table 1 Lake Natron waters' physico-chemical parameters (mean ± standard deviation) measured *in situ* (October, 2023) from random ten sampling sites along the coast of the lake

	Parameters					Coordinates (in decimal degrees)	
Sampling Point	**Temperature (°C)**	**pH**	**Salinity (psu)**	**EC (mS/cm)**	**TDS (g/L)**	**Latitude**	**Longitude**
NWA	36.40 ± 0.63	9.58 ± 0.10	15.62 ± 0.15	5.45 ± 0.73	3.41 ± 0.14	-2.268427	35.968552
NWB	36.02 ± 0.34	9.63 ± 0.41	15.47 ± 0.22	5.09 ± 0.32	3.25 ± 0.12	-2.279748	35.974731
NWC	36.50 ± 0.45	9.71 ± 0.21	15.42 ± 0.23	5.53 ± 0.16	3.51 ± 0.04	-2.290039	35.982285
NWD	37.19 ± 0.71	9.51 ± 0.23	15.63 ± 0.12	5.32 ± 0.43	3.32 ± 0.23	-2.294842	35.985718
NEW	36.20 ± 0.15	9.84 ± 0.21	15.47 ± 1.15	5.13 ± 1.51	3.17 ± 0.84	-2.302389	35.990524
NWF	36.30 ± 0.19	9.61 ± 0.38	15.43 ± 1.01	5.21 ± 0.22	3.12 ± 0.25	-2.308907	35.993271
NWG	37.15 ± 0.24	9.56 ± 0.15	16.01 ± 0.36	6.85 ± 0.32	3.64 ± 0.22	-2.316111	35.992584
NWH	36.35 ± 0.38	9.87 ± 0.31	15.37 ± 1.62	4.92 ± 0.02	3.06 ± 0.22	-2.323315	35.986061
NWI	35.30 ± 0.39	9.65 ± 0.12	15.86 ± 0.32	5.63 ± 0.03	3.55 ± 0.26	-2.330519	35.978165
NWJ	36.20 ± 0.29	9.61 ± 0.05	15.02 ± 0.29	5.14 ± 0.04	2.95 ± 0.24	-2.337379	35.964775

Extraction of genomic DNA

The filtration of water samples, DNA extraction and quantity/quality check of eluted DNAs followed the protocol depicted by Msaki et al.,[10] with the modifications, by using ZymoBIOMICS™ DNA Miniprep kit (Cat No. D4304, Zymoresearch Corp. USA), following the manufacture's protocol for DNA purification. The isolated genomic DNAs (gDNAs) were kept in storage at -80 °C freezer (Forma 900 series, Thermo Scientific).

Citation: Kaale SE, Machang'u RS, Lyimo TJ. Metagenomic analysis of bacterial communities inhabiting Lake Natron waters in Arusha, Tanzania. *J Microbial Exp.* 2024;12(2):32–40. DOI: 10.15406/jmen.2024.12.00413

PCR amplification of bacterial 16S rRNA gene

Prior to PCR amplification, the gDNAs from different sampling points were pooled together in equal concentration as follows; from sample NWA, NWB, and NWC were pooled and named NW1; sample from NWD, NWE and NWF were pooled and named NW2; sample from NWG, NWH, NWI and NWJ were pooled and named NW3. The now pooled DNAs (NW1, NW2 and NW3) were sent to INQABA®, genomics facility, Pretoria, South Africa, for PacBio sequencing. Briefly, the pooled three samples of gDNAs were PCR amplified using the bacterial universal primer pair 27F (5′-AGAGTTTGATCMTGGCTCAG-3′) as a forward primer and 1492R (5′-GGTTACCTTGTTACGACTT-3′) as a reverse primer, targeting the full-length of 16S rRNA gene.[11] These primers were marked with PacBio M13 adapter sequences on both 5' and 3' ends to enable barcoding of each amplicon so as to multiplex the resultant amplicons using limited cycle PCR.

PacBio full-length 16S rRNA gene sequencing and bioinformatic analysis

The resulting barcoded amplicons were quantified and pooled equimolar, followed by ampure PB bead-based purification step. The PacBio SMRTbell library was made using the pooled amplicons in accordance with the Manufacturer's instructions (https://www.pacb.com/wp-content/uploads/Procedure-checklist-Preparing-SMRTbell-libraries-using-PacBio-barcoded-M13-primers-for-multiplex-SMRT-sequencing.pdf). Sequencing, primer annealing and polymerase binding were done following SMRT link software protocol to prepare the library for sequencing on PacBio Sequel IIe system. The SMRTlink (v11.0) was used to process the raw sub-reads obtained after full-length 16S rRNA gene amplicon sequencing. Circular consensus sequences (CCS) technique was utilized to get accurate readings (>QV40), which were processed through Vsearch version 2.23.0 (https://github.com/torognes/vsearch). Taxonomic information was ascertained based on the Quantitative Insights Into Microbial Ecology (QIMME2) platform, a NGS microbiome bioinformatics tool for microorganism diversity and statistics.[12]

The sequences of pooled samples were then submitted to the NCBI BioProject database, three accession numbers namely NW1 (SAMN38511730), NW2 (SAMN38511731) and NW3 (SAMN38511732) were retrieved. Moreover, to demonstrate the alpha bacterial community diversity, species richness and evenness, the following diversity indices were computed using R-studio software viz., Shannon-Wiener diversity, Simpson, Chao1, Goods coverage and dominance.[13]

Ethics approval

Neither endangered nor protected species found at Lake Natron were included in the study. The research permits was given by the Tanzania National Parks (TANAPA), Tanzania Wildlife Research Institute (TAWIRI), Tanzania Wildlife Management Authority (TAWA), and Tanzania Commission for Science and Technology (COSTECH) that issued research permit number 2022-428-NA-2022-165 for field sampling.

Results and discussion

Physicochemical parameters of Lake Natron's water

The results for physicochemical parameters viz. temperature, pH, salinity, EC, and TDS were as shown in Table 1. The values recorded had minor variations in all sampling points along the lake shore. The temperature ranged from 35.30±0.39 to 37.19±0.71, which is within the range previously reported by Nonga et al.,[14] in Lake Natron. The temperature values were high considering October is within the dry season, although, this temperature range is still favorable for the proliferation of mesophilic bacteria, it being a vital requirement for their metabolic activities.[5] The pH and salinity ranged from 9.51±0.23 to 9.87±0.31 and 15.02±0.29 to 16.01±0.36, respectively. The high values of pH and salinity resemble those reported by Yona et al.,[5] which are attributed to the chemical composition of the lake's water which contains different salts which raises the pH and salinity levels.[15] Moreover, the two parameters are crucial for the distribution of bacteria species in soda lakes and the main reason why the waters of the lake are not suitable for use by humans and most of animals.[16] The EC and TDS recorded moderate values, ranging from 4.92±0.02 to 6.85±0.32 and 2.95±0.24 to 3.64±0.22, respectively. In Lake Natron water, EC and TDS levels are attributed by the presence of carbonate, bicarbonate, chloride, sulfate, phosphate and nitrate salts.[17] Therefore, high water temperatures, pH and salinity indicate that the bacteria communities inhabiting waters of Lake Natron are extremophiles (Archaea) which have adapted in this environment.

Metagenomic analysis of full length 16S gene amplicons

This study revealed the first metagenomic snapshot of bacteria inhabiting Lake Natron from water samples. The results shows, all three pooled samples namely NW1 (SAMN38511730), NW2 (SAMN38511731) and NW3 (SAMN38511732) produced about 13979.0, 3906.0, and 1328.0 numbers of reads, respectively, all belonging to Domain Bacteria by 100%. These results agrees with other metagenomic studies done in other soda lakes.[18] This study has investigated the diversity of bacteria living in the waters of Lake Natron for the first time using the PacBio IIe system sequencing full-length 16S rRNA reads as NGS technology. This NGS high-throughput amplicon sequencing analysis is effective as it covers substantial volumes of genomic data in a particular niche. Moreover, the results of previous studies which incorporated the same technique in soda lake environments have been precise and reliable because culture-independent DNA-based studies generate large-scale data sets that describe a full picture of microbial composition of a community.[19] Further taxonomical classification at phylum, class, family, genus and species level is well detailed below.

Microbial diversity at the phylum level

The results revealed a total of 29 different Phyla in the waters of Lake Natron of which 14 phyla in NW1 (SAMN38511730), 20 phyla in NW2 (SAMN38511731) and 25 phyla in NW3 (SAMN38511732) samples. As shown in Figure 2 and **Supplementary material Tables 1a and 1b**, Phylum Proteobacteria was dominant in NW1 (SAMN38511730) and NW2 (SAMN38511731) as it had the highest relative abundances of 98.46% and 70.46%, respectively. In NW3 (SAMN38511732), Firmicutes was the dominant Phylum with a relative abundance of 35.32% followed by Bacteroidota and Proteobacteria. The dominance of Proteobacteria was also observed by Vavourakis et al.,[20] who explored the uncultured bacteria of four hypersaline soda lake brines by metagenomic analysis. The main reason for the dominance of Proteobacteria in direct water samples microbiome is attributed to their adaptability in extreme habitats. Apart from the fact that they are versatile, they also can use a variety of energy and carbon sources, including organic and inorganic compounds, are capable of anaerobic respiration, and perform a variety of highly metabolic activities. The dominance of Firmicutes has also been found in five Siberian soda lakes by Vavourakis et al.,[21] Phylum Bacteroidota was the third dominant species in all sampling

points, results similar to Rojas et al.,[22] shows the dominance of these particular species in very alkaliphic soda lakes. In spite of the extremely high salinity and alkaline pH of soda lakes, Firmicutes and Bacteroidota contribute to their ecosystem stability through biogeochemical cycling processes.[8] Various metagenomic soda lake studies have reported the concurrently dominance of Proteobacteria, Firmicutes and Bacteroidota Phyla over others as they compete for sources of energy and produce antimicrobial compounds inhibiting the growth of other microorganisms.[16,23]

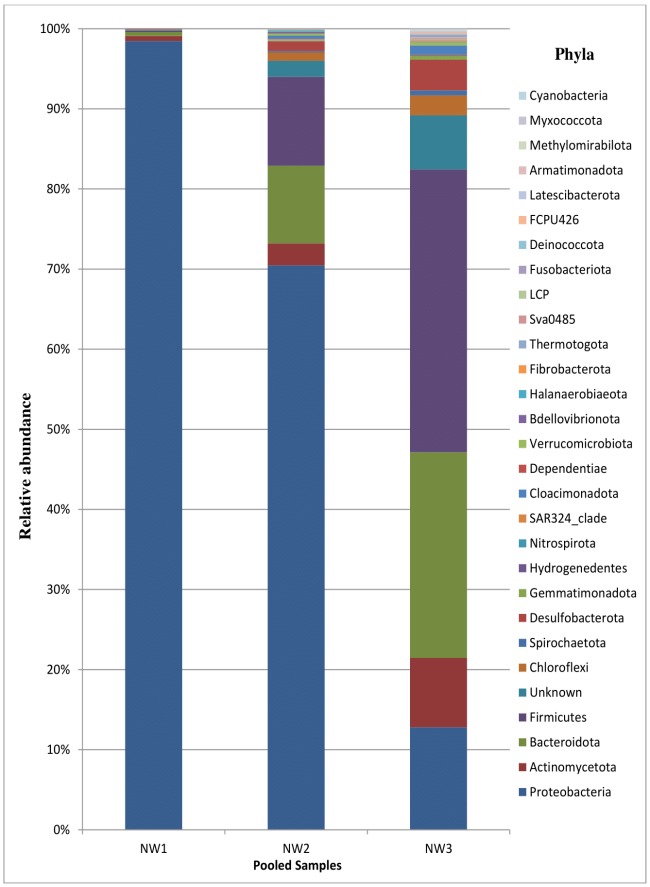

Figure 2 Taxonomic classification at Phylum level of different bacteria from the waters of Lake Natron in Arusha, Tanzania.

Other subdominant Phyla found in Lake Natron waters consisted of Actinomycetota, Desulfobacterota, Chloroflexi, and Cloacimonadota. While all species of bacteria recovered from the waters of Natron Soda Lake were examined in this study, the Phylum Actinomycetota in particular was of interest. Actinomycetota, previously known as Actinobacteria,[24] have been abundant in Tanzanian soda lakes; they have been found from Momela soda lakes (Small momela, Big momela, Rishateni, Lekandiro and Tulusia) located in Arusha national park,[25] and as recently isolated from Lake Natron waters.[6] Due to the fact that the Actinomycetota species discovered in Tanzania's soda lakes have been shown to have various antibacterial and biotechnological potentials,[26] it is imperative that all other unexplored soda lakes be further investigated.

The presence of Desulfobacterota and Desulfobacterota in this study supports the fact that these Phyla are the dominant communities in soda lake habitats, especially from direct water samples. Furthermore, they have adapted to the harsh physicochemical parameters posed by Lake Natron waters, which have high pH, salinity and rich in carbonate and bicarbonates. Previously, these bacteria have been reported by Chen et al.,[27] from Siberian soda lakes; their key role in soda lake waters is that they are integral part of sulfur cycle. The results showed a significant presence of Phyla Chloroflexi and Cloacimonadota in soda lakes as previously reported by Luo et al.,[28] from Oloidien Soda Lake in Kenya and Cabello-Yeves et al.,[29] from Lake El Tobar in Spain. Ecologically, members of Phylum Chloroflexi partake in nitrogen removal, biofilm aggregation and organic matter degradation[30] while Cloacimonadota have been reported to be involved in carbon cycling.[31] The rest of the other Phyla were represented by a relative abundance of less than 1% (**Supplementary material Tables 1a− 1c**), although they have been identified in such small quantities, their presence in Lake Natron water groups narrates them as halophiles and potentially ecologically and biotechnologically.

Several earlier studies done in Lake Natron focused mainly on flamingos and other non microbial life diversity in the Lake[4,5] with little attention on bacterial species.[6,12,32] This first metagenomic study to be conducted in Lake Natron, revealed that few bacteria match with previously culture based studies conducted. Only Phylum Actinomycetota species isolated by Kaale et al.,[6] and Phylum Cyanobacteria species isolated by Nonga et al.,[12] in Lake Natron have also been found in this study, indicating that they are abundantly distributed. On the other hand, Yakimov et al.,[30] isolated *Alcalilimnicola halodurans* belonging to phylum Pseudomonadota in Lake Natron but have not been found in this study. This may connote that Pseudomonadota are found in such low abundances that metagenomic sequencing could not detect them.[33]

Microbial diversity at the class level

This study has shown a total of 58 different Classes of which 22 from NW1 (SAMN38511730), 39 from NW2 (SAMN38511731) and 47 from NW3 (SAMN38511732) samples. The most dominant class was Alphaproteobacteria with a relative abundance of 93.59% in NW1 (SAMN38511730) and 67.72% in NW2 (SAMN38511731) as shown in Figure 3. Class Bacteroidia and Bacilli were the most dominant in NW3 (SAMN38511732), with relative abundance of 23.80% and 23.19%, respectively (**Supplementary material Tables 2a−2c**). The dominance of Alphaproteobacteria, Bacteroidia and Bacilli from the soda lakes water samples has been reported previously by Sorokin et al.,[8] and these Phyla have been deployed in different biogeochemical cycling of nitrogen, carbon and sulfur. The results also reveal the four classes of Actinomycetota species namely Acidimicrobiia, Coriobacteriia, Actinobacteria and Thermoleophilia were abundant in Lake Natron waters, this further insinuates the ubiquitous diversity of the species due to their adaptive mechanisms.

Moreover, Gammaproteobacteria, Gammaproteobacteria, Dethiobacteria, Clostridia, Anaerolineae, Desulfobulbia, Cloacimonadia, Rhodothermia, Ignavibacteria, Desulfobacteria, Desulfuromonadia, Spirochaetia and Desulfovibrionia classes were also highly represented in all samples with the relative abundance of more than 0.5% as shown in Figure 3; other Classes were presented in relative abundances of lower than 0.5% (**Supplementary material Tables 2a−2c**). Basically, the functions of these bacteria are diverse in soda lakes, however, in general, they are involved in nutrient cycling, sulfur metabolism, and organic compound degradation.[8]

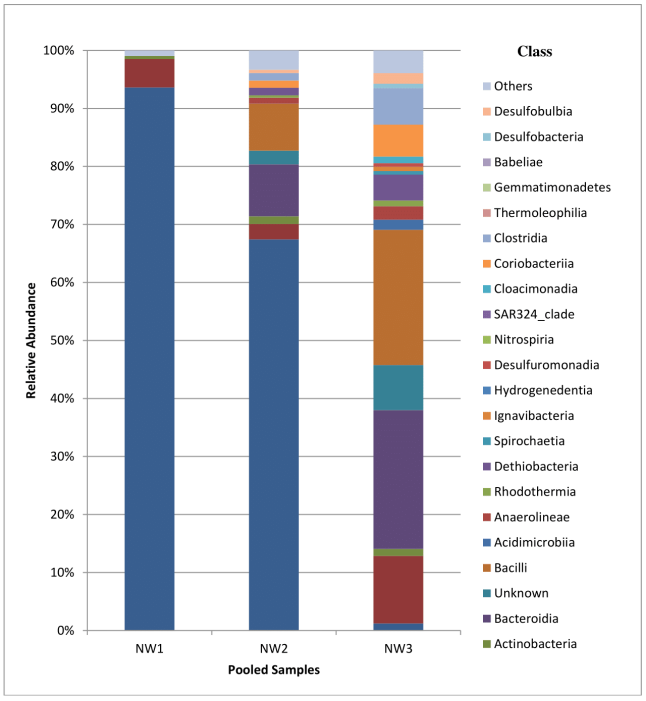

Figure 3 Taxonomic classification at Class level of different bacteria from the waters of Lake Natron, Arusha, Tanzania. The Figure shows the classes with the relative abundances of 0.5%≥.

Microbial diversity at the family level

The results revealed the number of bacterial Families inhabiting Lake Natron water as 35 in NW1 (SAMN38511730), 99 in NW2 (SAMN38511731) and 109 in NW3 (SAMN38511732) samples. In total, 139 different bacterial families were found in direct water samples of Lake Natron. The most dominant families were *Oceanibaculaceae, Rhizobiaceae* and *Izemoplasmataceae* in NW1 (SAMN38511730), NW2 (SAMN38511731) and NW3 (SAMN38511732) respectively, with relative abundances of 52.43%, 66.62% and 12.50%, respectively (Figure 4). Based on literature pertaining to metagenomics, this is the first study showing the dominance of family *Oceanibaculaceae, Rhizobiaceae* and *Izemoplasmataceae* in a soda lake environment. Basically, Family Oceanibaculaceae has been proposed recently[34] and are able to use carbon dioxide as a carbon source and sulfur as an energy source due to their capacity for sulfur-dependent lithoautotrophy. Being a recently proposed Family, there is little known of the distribution and characteristics of Oceanibaculaceae. Similarly, the data on the ecological roles of Rhizobiaceae is limited as this Family has not been found in soda lakes but mainly in plants. It's possible to infer that just like in other habitats, Rhizobiaceae is involved in sulfur-dependent lithoautotrophy and nitrogen fixation.[35]

In family level the Actinomycetota species were chiefly represented by Family *Nocardioidaceae*, species found in this family are characterized by producing antimicrobial products and enzymes.[26] Other notable Families included: *Devosiaceae, Oxalobacteraceae, Rhizobiaceae, Izemoplasmataceae*, ML635J, *Bacteroidetes_*BD2, *Dethiobacteraceae, Acholeplasmataceae*, OPB41, *Peptostreptococcales, Rhodocyclaceae*, RF39, *Anaerolineaceae, Erysipelotrichaceae*, *Desulfobulbaceae, Comamonadaceae, Nitrincolaceae* and *Woeseiaceae*. Collectively, members of these Families are ecologically involved in different nutrient cycling in soda lakes and produce a variety of products of biotechnological and industrial importance. Other Families represented had relative abundances of less than 0.5%. (**Supplementary material Tables 3a–3c**)

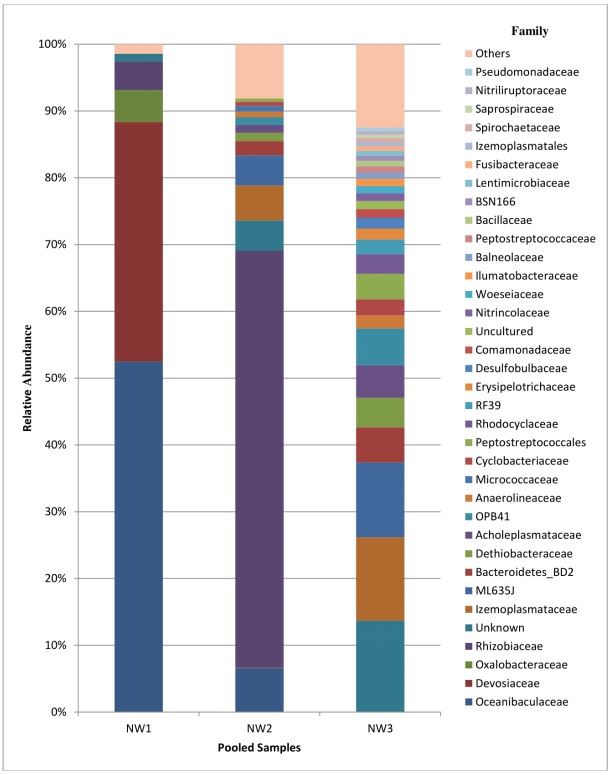

Figure 4 Classification at Family level of bacteria from the waters of Lake Natron in Arusha, Tanzania. The figure shows only the families with the relative abundances of 0.5%≥.

Microbial diversity at the genus level

Lake Natron waters were shown to contain a number of diverse bacteria at the Genus level. The results showed 28 different Genera in NW1 (SAMN38511730), 104 in NW2 (SAMN38511731) and 110 in NW1 (SAMN38511730) samples. In total, 162 different Genera of bacteria were found in the Lake Natron waters. The genus *Oceanibaculum* was the most dominant of the bacteria recovered in NW1 (SAMN38511730), with the relative abundance of 52.44% (Figure 5). This particular genus was first proposed in 2009 by Lai et al.,[36] it contains the bacteria which have been solely recovered from oceanic environments with the capability of reducing nitrate to nitrites. This is the first study revealing the presence of *Oceanibaculum* bacteria in a soda lake environment.

Genus *Allorhizobium* was the most dominant in NW2 (SAMN38511731). The members of the genus are gram negative bacteria exclusively found in soil and are responsible for nitrogen fixation in legumes and crown gall.[37,38] The presence of *Allorhizobium* bacteria in Lake Natron suggests that there is mix between the Lake's soil and water at the shoreline facilitated by the low depth of the lake. The results show *Izimaplasma* as the dominant genus in NW3 (SAMN38511732) by a relative abundance of 12.15%. Genus *Izimaplasma* is found in diverse environments and are typical commensals or parasites in their eukaryotic hosts.[39]

Other subdominant genera included *Devosia* and ML635J. Genus *Devosia* includes motile bacteria with features to adapt in diverse set of habitats. Their characteristics make them potential in agriculture as competent biofertilizers.[40] The presence of genus

ML635J continues to illustrate its dominance in soda lakes and agrees with[41] who identified this genus in Siberian soda lakes. Other genera found were notably: *Massilia*, *Aminobacter*, *Bacteroidetes*_BD2, *Acholeplasma*, OPB41, *Anoxynatronum*, RF39, ADurb, *Dethiobacter*, *Erysipelothrix* and *Azoarcus*. Generally, the members of these genera are potential in the production of different biotechnological products and, ecologically, they are involved in various cycling processes.[8] The representation of other genera was low as they had relative abundances of less than 0.5% (**Supplementary material Tables 4a−4c**). The Actinomycetota species were massively found at genus level as it included by *Alkaliphilus*, *Streptomyces*, *Nocardioides*, *Alkalitalea*, *Micromonospora*, *Agrococcus*, *Kocuria* and *Nitriliruptora* solely capable of surviving the alkaline environments such as Lake Natron waters with high pH and salinity.

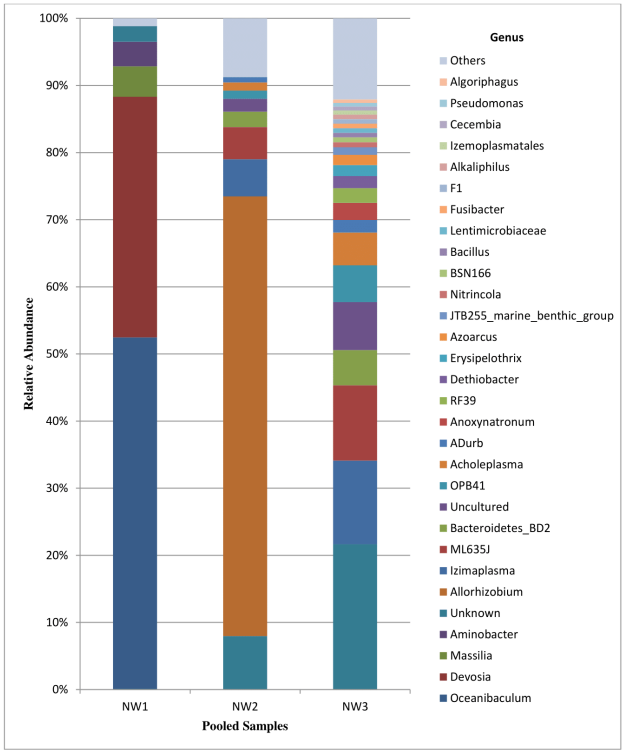

Figure 5 Classification at Genus level of bacterial isolates from the waters of Lake Natron, Arusha, Tanzania. The Figure shows the genera with the relative abundances of 0.5%≥.

Presence of vast unknown and uncultured species in Lake Natron waters

Most of metagenomic studies often characterize the microbiome at least up to the genus level due to the high number of microbial unknown species (poorly characterized) at the species level within the environmental samples. The presence of abundances of unknown groups of bacteria at Family level in Lake Natron waters proves further that only a small fraction of bacteria have been explored, still leaving a large number unknown.[19] Thus, further investigation of the unknown groups through NGS technologies is of importance as it gives more insight of soda lake microbial communities.[10] The results of this study present notable abundances of unknown species of bacteria communities inhabiting Lake Natron waters in nearly all taxonomic ranks particularly at species level.

At Phylum level, there was a considerable presence of members assigned to unknown Phyla, especially in NW2 (SAMN38511731) as represented by a relative abundance of 2% and NW3 (SAMN38511732) by 6.78% as shown in Figure 2 (**Supplementary material Tables 1b & 1c**). These results resembles other metagenomic studies that reported the presence of a large set of unidentified groups in other soda lake[21] and non soda lake environments.[10] Similarly at class, to species level, the numbers of unknown bacteria were high as shown in Figure 3-6, in all taxa. It was observed that, there is a trend that the lower the taxa level goes the percentages of unknown bacteria were much higher as it is evidently seen in all samples, NW1 (SAMN38511730), NW2 (SAMN38511731) and NW3 (SAMN38511732) were the most dominant species recovered from Lake Natron waters with the relative abundances of 98.80%, 79.13% and 42.55%, respectively. This trend was also observed in uncultured bacteria, all three samples contained nearly only uncultured and unknown bacteria at species level as seen in **Supplementary material Tables 5a−5c**.

The presence of such huge data of sequence sets of unknown and uncultured groups at all taxa levels could be due to their DNA sequences not being well represented in the reference databases used for bioinformatics analysis, and as a result these species are described as unknown or uncultured.[10] Also, such results call for bioinformatics development to create more tools to decode and characterize the unknown groups, which will give novel insights of the ecology and evolution of microbes in soda lakes.[42] Considering the technique used in this study, full length sequencing of 16S rRNA gene directly from water samples through PacBio Sequel IIe system and identification of such significant presence of unknown group of bacteria infers the possibility of putative new taxa at the genus level from Lake Natron waters that have not been described previously. This goes in parallel with the indication that enormous microbial life diversity is yet to be untapped. Considering that the better understanding of the microbial diversity will potentially contribute to the development of new biotechnological applications Handelsman,[43] hence, further studies on new bacteria species are important to know their taxonomy and ascertain their values in biotechnology and other fields.

Other subdominant species found in Lake Natron waters included the uncultured bacteria of *Izimaplasma*, L635J, *Erysipelothrix*, *Dethiobacter, Bacteroidetes*_BD2, OPB41, ADurb and *Firmicutes* (Figure 6). All of the species with the relative abundance of equal or greater than 0.5% of their communities were uncultured bacteria except *Cecembia lonarensis, Nitrincola* sp. and *Acholeplasma tenericutes.* Furthermore, the high numbers of uncultured bacteria continue to support the notion that, the use of high throughput sequencing technologies, directly from environmental samples, is superior to classical microbiology techniques in discovering unique extremophilic microbial communities in soda lakes.[18] Conversely, presence of *Cecembia lonarensis* in Lake Natron waters makes this the second study to report it in soda lake habitat after Anil et al.,[42] reported it from haloalkaline water samples of Lonar Lake of Buldhana district, India. The ecological and biotechnological importance of this haloalkalitolerant bacterium is not well understood, thus further studies are needed in order to assess their role in soda lake water habitats. Similarly, *Nitrincola* spp. has been previously isolated from soda lakes and other alkaline habitats.[44,45] Ecologically, the *Nitrincola* spp. are known to convert gaseous nitrogen into nitrates or nitrites. Other species were less presented as most of them had relative abundances of less than 0.5%. (**Supplementary material Tables 5a−5c**)

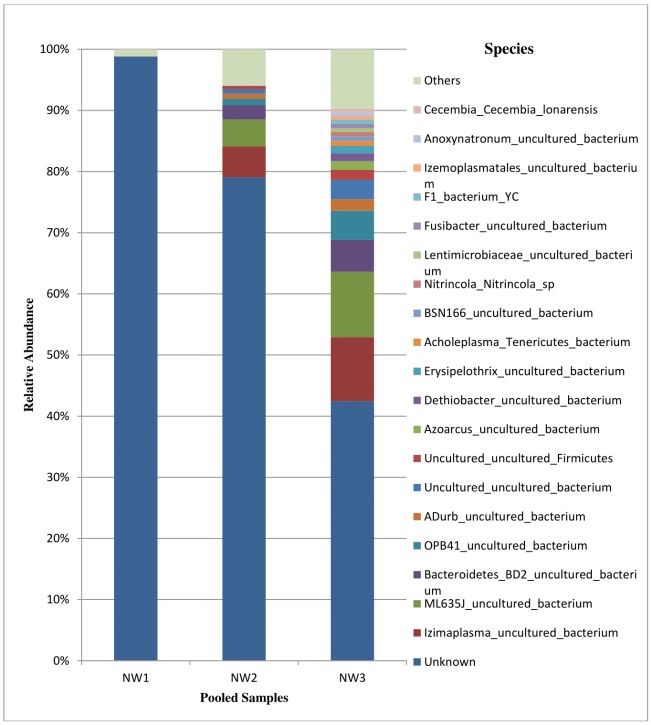

Figure 6 Speciation of different bacterial species of the waters of Lake Natron, Arusha, Tanzania. The figure shows the species with the relative abundances of 0.5%≥.

Bacterial community diversity indices

In microbial ecology studies, the community diversity indices (CDI) are the measures of diversity, richness and evenness of biological distribution in a given community. So far there is no universal agreement on specific CDI to use when conducting microbiome studies on soda lakes, or any other water body, although the ones used in this study are commonly used.[46] The results show that there were 18 species in NW1 (SAMN38511730), 75 species in NW2 (SAMN38511731), 91 species in NW3 (SAMN38511732) and in total 133 different species in all samples. Chao1 is the species richness estimator which expresses the total number of species present in a community by using the frequency of occurrence of rare operational taxonomic units (OTUs). As shown in Table 2, Chao1 values are higher than OUT values in all pooled samples and remarkably in sample NW2 (SAMN38511731). This means species richness is higher than the number of species actually observed in the samples OTU. It should be noted that as an alpha diversity index, Chao1 calculates not only the observed OUTs but also rare species not commonly observed in the sample. Hence, these results suggest that there might be more undetected species in the Lake Natron water samples, as the index accounts also for the rare species.[13]

The Shannon Diversity Index (SDI) quantitatively measures the number of different species within a sample concurrently, taking into consideration of how evenly a particular species is distributed among the others.[46] The high value of SDI in NW3 (SAMN38511732) indicates a high diversity of species in this community and lower in NW1 (SAMN38511730). This goes in tandem with high value of Simpson Diversity Index at NW3 (SAMN38511732) indicating a high number of species present and their abundances in their community (**Supplementary material Table 5c**). The dominance index value of 0.39 in NW1 (SAMN38511730) shows that there are more dominant species in that particular community than in the others (**Supplementary material Table 5a**). The Good's Coverage Index (GCI) is used in combination with other alpha diversity metrics and its value was high in NW1 (SAMN38511730), estimating the highest percentage of total species represented in a sample as higher GCI indicates a more complete sampling of the microbial community.[47] Generally, all these alpha diversity metrics involved had small variations throughout the samples indicating small variation of species along the sampled points in the shoreline of Lake Natron.[48]

Table 2 Summary of community diversity indices

Pooled Sample Name/Accession number	OTUs	Chao1	Shannon Index	Simpson Index	Good's coverage	Dominance
NW1 (SAMN38511730)	636	3093.02	2.52	0.64	0.96	0.39
NW2 (SAMN38511731)	918	4906.68	4.29	0.61	0.8	0.36
NW3 (SAMN38511732)	828	3097.24	8.98	0.99	0.5	0.01

Conclusion

This study has described the bacterial diversity harbored in the harsh environments of Lake Natron waters using NGS. To the best of our knowledge, this is the first full length 16S rRNA gene sequencing metagenomic research on soda lakes of Tanzania which correspond with soda lake metagenomic studies done elsewhere. In summary, the physicochemical parameters of the lake's water, especially the high pH and salinity insinuated the uniqueness and extremity of the habitat considering they are determining factor for bacterial species diversity. The results show the dominance of Proteobacteria, Firmicutes, Bacteroidota and Actinomycetota at phylum level; *Alphaproteobacteria, Gammaproteobacteria, Bacteroidia* and Bacilli at class level; *Oceanibaculaceae, Devosiaceae, Rhizobiaceae, Izemoplasmataceae* and ML635J at family level and *Oceanibaculum, Devosia, Allorhizobium, Izimaplasma* and ML635J at the genus level. The high relative abundances of unknown and uncultured species suggest putative new taxa and that only a fraction of the bacteria has been discovered. The alpha diversity indices indicated a high number and diversity of species, dominance and close variation of species among the lake's shoreline sampling points of this study. The similarity between Actinomycetota species found in this study with those previously isolated from Lake Natron and Momella soda lakes suggests the ubiquity of the phylum in Tanzania soda lakes, considering the biotechnological importance of its species further studies are recommended. The fact that some of the previous culture-based studies done in Lake Natron isolated bacteria that have not been demonstrated in this study indicates that for the purpose of getting the full picture of a particular community it is necessary to use both culture dependent and independent techniques. Furthermore, the study of bacteria functional metagenomics is of importance as it complements the information about their roles in ecology.

Funding

This research was partially funded by The Nelson Mandela Institute of Science and Technology (Arusha, Tanzania) through the World Bank's Higher Education for economic Transformation Project (HEET) as part of Mr. SE Kaale's PhD studies and Inqaba

Biotechnological Company, Tanzania Branch. The funding bodies had no role in the study design, analysis and interpretation of data or in the writing of the manuscript.

Acknowledgments

The authors are grateful to Dr. Prosper Mosha, Ms. Witness Lema and Ms. Winnie Kimaro for their assistance during laboratory analysis. Also we extend our gratitude to Mr. Abdillahi Kiula, Mr. William Faraji and Ms. Angela Siima for their assistance during the metagenomic analysis.

Conflicts of interests

The authors have no relevant conflict of interests to disclose.

References

1. Thomas T, Gilbert J, Meyer F. Metagenomics - a guide from sampling to data analysis. *Microb Inform Exp*. 2012;2(1):3.
2. Hur M, Park S. Identification of Microbial profiles in heavy-metal-contaminated soil from full-length 16S rRNA reads sequenced by a pacbio system. *Microorganisms*. 2019;7(9):357.
3. Ando N, Barquera B, Bartlett DH, et al. The molecular basis for life in extreme environments. *Ann Rev Biophys*. 2021;50:343−372.
4. Mgimwa EF, John JR, Lugomela CV. The influence of physical–chemical variables on phytoplankton and lesser flamingo (*Phoeniconaias minor)* abundances in Lake Natron, Tanzania. *Afr J Ecol*. 2021;59:667–675.
5. Yona C, Makange M, Moshiro E, et al. Water pollution at Lake Natron Ramsar site in Tanzania: A threat to aquatic life. *Ecohydrology and Hydrobiology*. 2023;23:98–108.
6. Kaale SE, Machangu RS, Lyimo TJ. Molecular characterization and phylogenetic diversity of actinomycetota species isolated from Lake Natron sediments at Arusha, Tanzania. *Microbiol Res*. 2024;278:127543.
7. Zorz JK, Sharp C, Kleiner M, et al. A shared core microbiome in soda lakes separated by large distances. *Nat commun*. 2019;10(1):4230.
8. Sorokin DY, Berben T, Melton ED, et al. Microbial diversity and biogeochemical cycling in soda lakes. *Extremophiles*. 2014;18:791–809.
9. Dos Santos RS, Babiloni-Chust I, Marroqui L, et al. Screening of metabolism-disrupting chemicals on pancreatic α-cells using in vitro methods. *Int J Mol Sci*. 2022;24(1):231
10. Msaki GL, Kaale SE, Njau KN, et al. Bacterial communities structure in constructed wetlands for municipal and industrial wastewater treatment in Tanzania. *Water Practice & Technology*. 2023;20:23155.
11. Palkova L, Tomova A, Repiska G, et al. Evaluation of 16S rRNA primer sets for characterisation of microbiota in paediatric patients with autism spectrum disorder. *Sci Rep*. 2021;11(1):6781.
12. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. 2010;7(5):335−336.
13. Kim BR., Shin J, Guevarra RB, et al. Deciphering diversity indices for a better understanding of microbial communities. *J Microbiol Biotechnol*. 2017;27(12):2089−2093.
14. Nonga HE, Mdegela RH, Sandvik M, et al. Cyanobacteria and cyanobacterial toxins in the alkaline-saline Lakes Natron and Momela, Tanzania. *Tanzania Veterinary Journal*. 2017;32:108–116.
15. Clarisse L, Van Damme M, Gardner W, et al. Atmospheric ammonia (NH_3) emanations from Lake Natron's saline mudflats. *Sci Rep.* 2019;9(1):1–12.
16. Chavan V, Mulaje S, Mohalkar R. A review on actinomycetes and their biotechnological application. *International Journal of Pharmaceutical Sciences and ResearcH*. 2013;4:1730–1742.
17. Philip N, Mosha S. Salt Lakes of the African Rift System: A valuable research opportunity for insight into nature's concenrtated multi-electrolyte science. *Tanzania Journal of Science* 2012;38:1–13.
18. Omeroglu E, Sudagidan M, Yurt NZ et al. Microbial community of soda Lake Van as obtained from direct and enriched water, sediment and fish samples. *Sci Rep*. 2021;11(1):18364.
19. Jeilu O, Gessesse A, Simachew A, et al. Prokaryotic and eukaryotic microbial diversity from three soda lakes in the East African Rift Valley determined by amplicon sequencing. *Front Microbiol*. 2022;13:999876.
20. Vavourakis CD, Ghai R, Rodriguez-Valera F, et al. Metagenomic insights into the uncultured diversity and physiology of microbes in four hypersaline soda lake brines. *Front Microbiol*. 2016;7:211.
21. Vavourakis CD, Andrei AS, Mehrshad M, et al. A metagenomics roadmap to the uncultured genome diversity in hypersaline soda lake sediments. *Microbiome*. 2018;6:1–18.
22. Rojas P, Rodríguez N, de la Fuente V, et al. Microbial diversity associated with the anaerobic sediments of a soda lake (Mono Lake, California, USA). *Can J Microbiol*. 2018;64(6):385–392.
23. Wang M, Zhang X, Shu, Z, et al. Bacterial and archaeal communities within the alkaline soda Langaco Lake in the Qinghai-Tibet Plateau. *Ann Microbiol*. 2022;72:33.
24. Oren A, Garrity GM. Valid publication of the names of forty-two phyla of prokaryotes. *Int J Syst Evol Microbiol.* 2021;71:005056.
25. Kaale SE, Mahadhy A, Damas M, et al. Phylogenetic diversity of actinobacteria from momela soda lakes, Arusha National Park, Tanzania. *African Journal of Aquatic Science*. 2022;47(2):149−162
26. Lema WS, Mahadhy A, Damas M, et al. Characterisation and Antimicrobial Potential of Actinobacteria Isolated from Momela Soda Lakes, Tanzania. *Tanzania Journal of Science*. 2022;48:607−622.
27. Chen YH, Chiang PW, Rogozin DY, et al. Salvaging high-quality genomes of microbial species from a meromictic lake using a hybrid sequencing approach. *Commun Biol*. 2021;4(1):996.
28. Luo W, Li H, Kotut K, et al. Molecular diversity of plankton in a tropical crater lake switching from hyposaline to subsaline conditions: Lake Oloidien, Kenya. *Hydrobiologia*. 2017;788:205–229
29. Cabello-Yeves PJ, Picazo A, Roda-Garcia JJ, et al. Vertical niche occupation and potential metabolic interplay of microbial consortia in a deeply stratified meromictic model lake. *Limnology and Oceanography*. 2023;68(11):2492–2511
30. Bovio-Winkler P, Guerrero LD, Erijman L, et al. Genome-centric metagenomic insights into the role of Chloroflexi in anammox, activated sludge and methanogenic reactors. *BMC Microbiol*. 2023;23(1):45.
31. Johnson LA, Hug LA. Cloacimonadota metabolisms include adaptations in engineered environments that are reflected in the evolutionary history of the phylum. *Environ Microbiol Rep*. 2022;14(4):520−529.
32. Yakimov MM, Giuliano L, Chernikova TN, et al. *Alcalilimnicola halodurans* gen. nov., sp. nov., an alkaliphilic, moderately halophilic and extremely halotolerant bacterium, isolated from sediments of soda-depositing Lake Natron, East Africa Rift Valley. *Int J Syst Evol Microbiol*. 2001;51:2133−2143.
33. Hilton SK, Castro-Nallar E, Pérez-Losada M, et al. Metataxonomic and metagenomic approaches vs. culture-based techniques for clinical pathology. *Front Microbiol*. 2016;7:484.
34. Koziaeva VV, Sorokin DY, Kolganova TV, et al. *Magnetospirillum sulfuroxidans sp*. nov., capable of sulfur-dependent lithoautotrophy and a taxonomic reevaluation of the order Rhodospirillales. *Syst Appl Microbiol*. 2023;46(3):126406.
35. Gopalakrishnan S, Sathya A, Vijayabharathi R, et al. Plant growth promoting rhizobia: challenges and opportunities. *3 Biotech*. 2015;5:355−377.

36. Lai Q, Yuan J, Wu C, et al. Oceanibaculum indicum gen. nov., sp. nov., isolated from deep seawater of the Indian Ocean. *Int J Syst Evol Microbiol*. 2009;59(7):1733−1737.

37. de Lajudie P, Laurent-Fulele E, Willems A, et al. Allorhizobium undicola gen. nov., sp. nov., nitrogen-fixing bacteria that efficiently nodulate Neptunia natans in Senegal. *Int J Syst Evol Microbiol.*1998;48(4):1277−1290.

38. Mousavi SA, Willems A, Nesme X, et al. Revised phylogeny of Rhizobiaceae: proposal of the delineation of Pararhizobium gen. nov., and 13 new species combinations. *Syst Appl Microbiol*. 2015;38(2):84–90.

39. Skennerton C, Haroon M, Briegel A et al. Phylogenomic analysis of *Candidatus* 'Izimaplasma' species: free-living representatives from a *Tenericutes* clade found in methane seeps. *ISME J.* 2016;10(11):2679–2692.

40. Chhetri G, Kim I, Kang M, et al. *Devosia rhizoryzae* sp. nov., and *Devosia oryziradicis* sp. nov., novel plant growth promoting members of the genus *Devosia*, isolated from the rhizosphere of rice plants. *J Microbiol.* 2022;60(1):1–10.

41. Vavourakis CD, Mehrshad M, Balkema C, et al. Metagenomes and metatranscriptomes shed new light on the microbial-mediated sulfur cycle in a Siberian soda lake. *BMC Biol.* 2019;17(1):69.

42. Vanni C, Schechter MS, Acinas SG, et al. Unifying the known and unknown microbial coding sequence space. *Elife*. 2022;11:e67667.

43. Handelsman J. Metagenomics: application of genomics to uncultured microorganisms. *Microbial Mol Biol Rev*. 2004;68(4):669−685.

44. Anil Kumar P, Srinivas TNR, Madhu S, et al. *Cecembia lonarensis* gen. nov., sp. nov., a haloalkalitolerant bacterium of the family Cyclobacteriaceae, isolated from a haloalkaline lake and emended descriptions of the genera Indibacter, Nitritalea and Belliella. *Int J Syst Evol Microbiol*. 2012;62(9):2252−2258.

45. Borsodi AK, Korponai K, Schumann P, et al. *Nitrincola alkalilacustris* sp. nov. and *Nitrincola schmidtii* sp. nov., alkaliphilic bacteria isolated from soda pans, and emended description of the genus *Nitrincola*. *Int J Syst Evol Microbiol*. 2017;67(12):5159−5164.

46. Dimitriu PA, Shukla SK, Conradt J, et al. *Nitrincola lacisaponensis* gen. nov., sp. nov., a novel alkaliphilic bacterium isolated from an alkaline, saline lake. *Int J Syst Evol Microbiol*. 2005;55(6):2273−2278.

47. Morris EK, Caruso T, Buscot F, et al. Choosing and using diversity indices: insights for ecological applications from the German Biodiversity Exploratories. *Ecol Evol*. 2014;4(18):3514–3524.

48. Bardenhorst SK, Vital M, Karch A, et al. Richness estimation in microbiome data obtained from denoising pipelines. *Comput Struct Biotechnol J*. 2022;20:508−520.

49. Hou T, Liu F, Liu Y, et al. Classification of metagenomics data at lower taxonomic level using a robust supervised classifier. *Evolu Bioinform Online*. 2015;11:S20523.