



Check for updates

RESEARCH ARTICLE

REVISED Spatiotemporal structure and composition of the microbial communities in hypersaline Lake Magadi, Kenya

[version 2; peer review: 1 approved, 2 approved with reservations, 1 not approved]

Alex Kipnyargis ¹, Eucharia Kenya¹, Fathiya Khamis², Romano Mwirichia ¹¹Department of Biological Sciences, University of Embu, Embu, Kenya²Arthropod Pathology, International Centre of Insect Physiology and Ecology, Nairobi, Nairobi County, Kenya

V2 First published: 03 Jan 2024, 13:11
<https://doi.org/10.12688/f1000research.134465.1>
Latest published: 17 May 2024, 13:11
<https://doi.org/10.12688/f1000research.134465.2>

Abstract**Background**

Soda lakes are habitats characterized by haloalkaline conditions also known to host unique microbial communities. The water chemistry changes with seasons due to evaporative concentration or floods from the surrounding grounds. However, it is not yet clear if the change in physiochemical changes influences the spatiotemporal diversity and structure of microbial communities in these ecosystems.







Methods



Using 16S rRNA gene amplicon sequencing, we investigated the diversity and structure of microbial communities in water and brine samples taken from Lake Magadi between June and September 2018. Additionally, physicochemical parameters were also analyzed for every sampling site. Additionally, physicochemical parameters were also analyzed for every sampling site.

Results

The abundant bacterial phyla were Proteobacteria, Cyanobacteria, Bacteroidetes, Actinobacteria, Firmicutes, Verrumicrobia, Deinococcus-Thermus, Spirochaetes, and Chloroflexi. The Archaeal diversity was represented by phyla Euryarchaeota, Crenarchaeota, Euryarchaeota, and Thaumarchaeota. The dominant bacterial species were: *Eubalthece* sp. (10.3%), *Rhodobaca* sp. (9.6%), *Idiomarina* sp. (5.8%), *Rhodothermus* sp. (3.0%), *Roseinatronobacter* sp. (2.4%), *Nocardioidea* sp. (2.3%), *Gracilimonas* sp. (2.2%), and *Halomonas* sp. (2%). The dominant archaeal species included *Halorubrum* sp. (18.3%),

Open Peer Review**Approval Status**    

	1	2	3	4
version 2 (revision) 17 May 2024	 view		 view	 view
				
version 1 03 Jan 2024	 view	 view		

1. **Marianne Haines**, University of Calgary, Calgary, Canada
2. **Srijak Bhatnagar** , Athabasca University, Athabasca, Canada
3. **Cristiano Pedroso-Roussado** , Universidade de Lisboa Instituto Superior Tecnico, Lisboa, Portugal
4. **Addis Simachew**, Addis Ababa University, Addis Ababa, Ethiopia

Any reports and responses or comments on the article can be found at the end of the article.

Salinarchaeum sp. (5.3%), and *Haloterrigena* sp. (1.3%). The composition of bacteria was higher than that of archaea, while their richness and diversity varied widely across the sampling seasons. The α -diversity indices showed that high diversity was recorded in August, followed by September, June, and July in that order. The findings demonstrated that temperature, pH, P+, K+, NO₃⁻, and total dissolved solids (TDS) contributed majorly to the diversity observed in the microbial community. Multivariate analysis revealed significant spatial and temporal effects on β -diversity and salinity and alkalinity were the major drivers of microbial composition in Lake Magadi.

Conclusions

We provide insights into the relationships between microbial structure and geochemistry across various sampling sites in Lake Magadi.

Keywords

Soda lake, spatiotemporal, archaea, bacteria, physicochemical parameters, α -diversity, β -diversity



This article is included in the **Bioinformatics** gateway.

Corresponding author: Alex Kipnyargis (kipnyargis.alex@embuni.ac.ke)

Author roles: **Kipnyargis A:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Kenya E:** Conceptualization, Formal Analysis, Investigation, Methodology, Resources, Supervision, Validation, Visualization, Writing – Review & Editing; **Khamis F:** Conceptualization, Investigation, Methodology, Resources, Supervision, Visualization, Writing – Review & Editing; **Mwirichia R:** Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This research was funded by the National Research Fund, Kenya (grant number NRF/1/MMC/470) to Romano Mwirichia.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2024 Kipnyargis A *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Kipnyargis A, Kenya E, Khamis F and Mwirichia R. **Spatiotemporal structure and composition of the microbial communities in hypersaline Lake Magadi, Kenya [version 2; peer review: 1 approved, 2 approved with reservations, 1 not approved]** F1000Research 2024, 13:11 <https://doi.org/10.12688/f1000research.134465.2>

First published: 03 Jan 2024, 13:11 <https://doi.org/10.12688/f1000research.134465.1>

REVISED Amendments from Version 1

This version contains the revisions that were suggested by the first reviewer.

Abstract: Multivariate analysis was used to explain the clustering of microbial communities. In this case, the results showed that concentrations of salts and alkalinity levels were responsible for the clustering of samples rather than the sampling time point.

Rather than using correlation analysis, the effects of water chemistry on the microbial structure were paraphrased to emphasize their influence.

Results: There was a reordering of data and figures to show results in chronological order. A few misleading sentences were also paraphrased as suggested by the reviewer.

Discussion: This section has been revised to depict the relationship of our findings with preexisting information on soda lakes. Those paragraphs that were seen as irrelevant to our results have been deleted.

Conclusion: The entire conclusion has been revised by me to better convey the important points about variety among sample locations and seasons as well as the influence of environmental conditions on the dynamics of microorganism populations in Lake Magadi. Your remarks were very beneficial.

Any further responses from the reviewers can be found at the end of the article

Introduction

Most living organisms are adapted to habitats characterized by moderate temperature (10–37°C), pH (of approximately 7), salinity (0.15–0.5 M NaCl), pressure (1 atm), and adequate supply of water (Aguilar *et al.*, 1998; Antranikian *et al.*, 2005). However, molecular techniques such as next-generation sequencing have revealed that diverse groups of organisms thrive even in biomes previously thought to be lifeless (Canganella & Wiegel, 2011; Rampelotto, 2013). Microbial communities in ecosystems such as the hypersaline lakes of the East African Rift Valley survive and thrive under one or several extreme conditions and are referred to as polyextremophiles (Sorokin *et al.*, 2014a; Urbietta *et al.*, 2015).

The distribution and diversity of microbial communities in hypersaline lakes are mainly affected by physicochemical parameters (Tazi *et al.*, 2014). Lake Magadi is an example of an extreme habitat characterized by high concentrations of Na⁺, K⁺, CO₃²⁻, Cl⁻, HCO₃⁻, and SiO₂, but low concentrations of Ca²⁺ and Mg²⁺ (Jones *et al.*, 1998; Getenet *et al.*, 2022). During the dry seasons, thermonatrite (Na₂CO₃·H₂O), and halite (NaCl) precipitate by evaporative concentration (Eugester, 1971, 1980). The lake is in a region with alternating wet and dry seasons. During the dry season, when ground temperatures exceed 40°C, there is extensive evaporation (Matagi, 2004; Muruga & Anyango, 2013). In addition, a white layer of soda covers the lake almost completely, and rainy days may result in flooding from the surrounding water entering the lake.

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

Many bacterial species have been isolated from extreme environments, and they frequently exhibit adaptations to optimal growth under the prevailing conditions (Krulwich *et al.*, 2011; De Maayer *et al.*, 2014; Sorokin *et al.*, 2014b). Previously described isolates from Lake Magadi include the archaeal genera *Natronobacterium* and *Natronococcus* gen. nov. (Tindall *et al.*, 1984) and *Natronobacterium magadii*, *Natrialba magadii* (Kamekura *et al.*, 1997), bacterial species *Spirochaeta alkalica* sp. nov., *Spirochaeta Africana* (Zhilina *et al.*, 1996), *Tindallia magadiensis* (Kevbrin *et al.*, 1998), *Halomonas magadii* (Duckworth *et al.*, 2000), *Amphibacillus fermentum* (renamed *Pelagirhabdus fermentum*) sp. nov., *Amphibacillus tropicus*, and *Halonatronum saccharophilum* (Zhilina *et al.*, 2001), *Methylnatronum kenyense* (Sorokin *et al.*, 2007), *Euhalothece natronophila* (Mikhodyuk *et al.*, 2008) and *Natranaerobaculum magadiense* (Zavarzina *et al.*, 2013). Edwin *et al.* (2019) recovered 11 isolates affiliated with the cyanobacterial orders *Chroococcales*, *Oscillatoriales*, *Pleurocapsales*, and *Nostocales*. Recent studies have reported isolates affiliated with the genus *Bacillus*, *Alkalibacterium*, *Staphylococcus*, *Micrococcus*, *Halomonas*, and *Alkalilimnicola* (Kiplimo *et al.*, 2019; Kipnyargis *et al.*, 2022). Orwa *et al.* (2020) recovered several fungal isolates affiliated with 18 different genera with *Aspergillus*, *Penicillium*,

Cladosporium, *Phorma*, and *Acremonium* being dominant. Several studies have explored the microbial diversity in Lake Magadi using amplicons analysis targeting groups such as fungi (Kambura *et al.*, 2016; Salano *et al.*, 2017; Mwirichia, 2022) or bacteria (Kambura *et al.*, 2016).

A key ecological question is how microbial diversity changes with the fluctuating physicochemical conditions with seasons. We hypothesized that microbial communities within the lake shift in response to changes in the water chemistry over time. We predict that the communities in the brines are different from those in the open lake water. In this study, we explored the spatiotemporal variation in the microbial community over four months at different sites in Lake Magadi using 16S rRNA gene sequencing.

Methods

Sampling site and sampling criteria

Sampling was done in hypersaline Lake Magadi, Kenya. It is located $1^{\circ}43'2''\text{S}$ and $36^{\circ}13'36''\text{E}$ in an enclosed basin with an annual precipitation of 500 mm (Behr & Röhrich, 2000). Lake Magadi is a relatively shallow water body that is fed by various hot springs distributed along the edges of the lake. The inflows have an influence on the lake volume and the water chemistry. Water samples were collected from different points in the lake including spring, brine, and open waters. Samples were collected from these sites in June, July, August, and September 2018 and the dry season lasted for all the sampling months, with June marking the start of the season and September being the driest month. The coordinates of the sampling sites were: S1 (1.891380°S ; $36.302632^{\circ}\text{E}$), S2 (1.895020°S ; $36.299372^{\circ}\text{E}$), S3 (1.900988°S ; $36.301307^{\circ}\text{E}$), S4 (1.908460°S ; $36.301996^{\circ}\text{E}$), S5 (1.991601°S ; $36.258904^{\circ}\text{E}$), S6 (1.975517°S ; $36.236564^{\circ}\text{E}$) and BR1 (1.887908°S ; $36.300855^{\circ}\text{E}$) (Figure 1). S1 was composed of hot spring water, S2–S6 were composed of open waters, and BR1 was brine. Three sub-samples of 50 ml each were collected from each site and pooled into a composite sample. In addition, water samples for physicochemical analysis were collected. All samples were collected in sterile Conical Centrifuge tubes (Biologix, Shandong, China, Cat. No. 430829) and transported in a cool box.

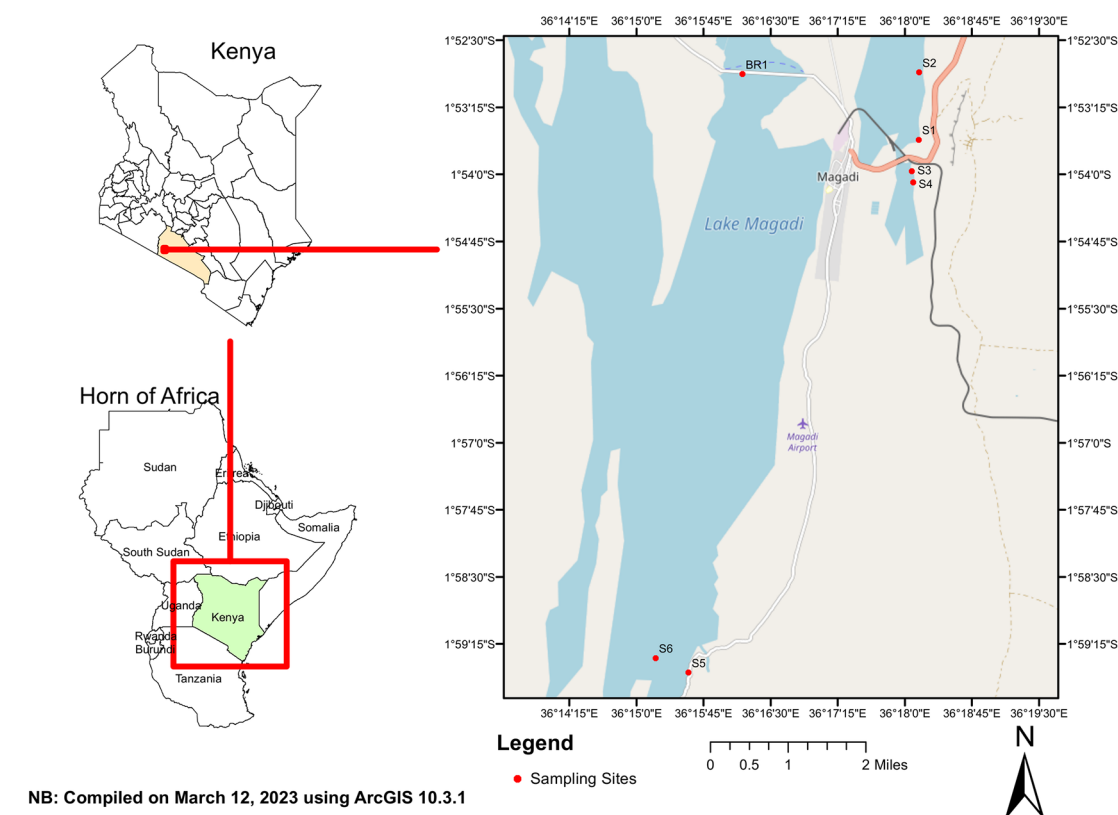


Figure 1. Map of Lake Magadi showing the sampling sites.

Analysis of physicochemical parameters

Water temperature, pH, total dissolved solids (TDS), and salinity measurements were recorded *in situ*. Water temperature, TDS, and salinity were measured using a VWR phenomenal handheld Meter (VWR, Atlanta, GA, USA, Model CO 3100H), while pH was measured using a Hanna Combo pH meter (Hanna Instruments, Nusafalau, Romania, Model HI-98128). In this case, about 100 ml of sample water was put in a sterile 400 ml glass beaker (Marienfeld, Germany, Cat. No. BR91236). A pre-calibrated meter was dipped in the sample and the readings were recorded. Water samples for dissolved P, K⁺, NO₃⁻, NH₄⁺, Mg²⁺, Na⁺, Fe²⁺, Ca²⁺, SO₄²⁻, Cl⁻, and HCO₃⁻ measurements were collected in sterile 500ml bottles and stored in a cool box for transportation to Crop Nutrition Laboratory Services (CNLS), Nairobi where analysis was done. Cations such as Ca, Mg, K, Na, Mn, Fe, Cu, Mo, B, Zn, and S were analyzed using atomic absorption spectrometry (AAS). At the same time, anion analysis was carried out using mass spectrometry.

DNA extraction

Cell biomass for DNA extraction was obtained by centrifuging 50 ml of each water sample at 14,000 rpm for 20 minutes in an Eppendorf centrifuge (Eppendorf, Model 5415R, Cat. Z605212). The pellets were resuspended in 200 µl of a resuspension buffer (25% w/v sucrose (Sigma-Aldrich, Cat. No. S9378) in 50 mM Tris pH 8.5 (Sigma-Aldrich, Cat. No. 93352), and 50 mM EDTA; pH 8.0 (Sigma-Aldrich, Cat. No. 798681). To disrupt the cell wall of Gram positives, 2 µl of lysozyme (20 mg/ml) (Roche, Cat. No. 10837059001) and 10 µl of RNase A (20 mg/ml) (Roche, Cat. No. 10109142001) were added and incubated at 37°C for 30 minutes. Cell lysis was achieved by the addition of 600 µl of a lysis buffer (1% SDS (Sigma-Aldrich, Cat. No. 8.17034) in 10 mM Tris pH 8.5 (Sigma-Aldrich) and 5 mM EDTA; pH 8.0 (Sigma-Aldrich). The samples were gently mixed with 10 µl of Proteinase K (20 mg/ml) (Sigma-Aldrich, Cat. No. 39450-01-6) and incubated at 65°C for 2 hours. DNA was recovered by adding an equal volume of chloroform (Sigma-Aldrich, Cat. No. C2432) followed by centrifugation at 13,200 rpm for 10 min at 4°C in an Eppendorf 5415R centrifuge. The aqueous layer was transferred into a new tube with 150 µl of sodium acetate (pH 5.2) (Sigma-Aldrich, Cat. No. S8750) and an equal volume of isopropyl alcohol (Sigma-Aldrich, Cat. No. 67-63-0). The contents were centrifuged at 13,200 rpm for 10 minutes and the DNA pellet was recovered by washing with 70% ethanol, air-dried for 15 minutes, and dissolved in 30 µl of nuclease-free water (Sigma-Aldrich, Cat. No. 7732-18-5). DNA quality was checked by running an aliquot of 2 µl in 1% agarose (Sigma-Aldrich, Cat. No. A9918) gel electrophoresis (Orwa *et al.*, 2020).

Sequencing of the 16S rRNA amplicons

The V4 hypervariable region of the 16S rRNA genes was amplified using the universal primers for bacterial and archaeal primers 515F (5'-GTGCCAGCMGCCGCGGTAA-3') and 806R (5'-GGACTACHVGGGTWTCTAAT-3') (Caporaso *et al.*, 2012). Amplification was done using HotStarTaq Plus Master Mix Kit (Qiagen, USA) under the following cycling conditions: initial denaturation at 94°C for 3 minutes, followed by 35 cycles of denaturation at 94°C for 30 seconds, annealing at 53°C for 40 seconds and elongation at 72°C for 1 minute, after which a final elongation step at 72°C for 5 minutes was performed. Three independent PCR reactions were performed per sample and pooled in equimolar amounts. The PCR products were then checked in a 2% agarose gel. The sample was purified using calibrated Ampure XP beads (Beckman Coulter, Inc., IN, USA). DNA libraries were prepared using Illumina TruSeq DNA libraries (Illumina, Inc., San Diego, CA, United States) and sequencing was performed at MR DNA (Shallowater, TX, USA) on a MiSeq platform (2 × 300 bp) following the guidelines of the manufacturer (Illumina Inc.).

Sequence processing and taxonomic assignment

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

The Illumina raw reads for the 16S rRNA gene sequences were deposited in the Sequence Read Archive (SRA) of NCBI under the accession numbers PRJNA962270 (Kipnyargis *et al.*, 2023b).

Table 1. Physicochemical characteristics of the water samples collected from Lake Magadi. TDS total dissolved solids, SAR sodium absorption ratio. The samples are denoted as S1 to S6, while BR1 represents the brine sample.

Sample	S1_June	S1_Sep	S2_June	S2_July	S2_Sep	S3_June	S3_July	S3_Aug	S3_Sep	S4_June	S4_July	S4_Aug	S4_Sep	S5_June	S5_July	S5_Aug	S5_Sep	S6_June	S6_July	S6_Aug	BR1_June	BR1_Sep
pH	10.5	10.4	10.6	10.2	10.7	10.5	10.5	11.3	10.4	10.5	10.5	11.1	10.9	10.2	9.9	10.3	10.2	10.3	9.8	10.4	11.5	11.2
Temp. (°C)	35.6	35.6	35.1	32.8	37.2	38.7	27	34.6	35.6	37.3	27	33.6	37.8	34.2	32.7	29.8	32.8	32.5	30.2	29.2	38.3	34.8
TDS (ppm)	27.1	27.1	137.6	143.9	145	114.1	135.4	139.2	153.5	110.7	134.6	139.8	143.3	46	45.7	45.7	42.9	104.9	118.1	83.9	136	134.8
P (ppm)	2.38	7.77	58.4	78	117	27.4	77.2	108	69.8	26.4	63.8	107	105	3.96	2.39	5.79	10.6	16.2	21.9	4.91	81.1	92.8
K (ppm)	131	365	2,300	2,700	4,280	1,220	2,430	3,300	2,560	1,130	1,960	3,270	2,370	201	190	280	378	697	697	513	3,410	3,210
NO ₃ (ppm)	0.89	9.03	1.81	0.01	5.98	0.01	0.01	0.2	4.96	0.01	0.01	0.01	6.99	4.56	0.01	0.19	8.5	32	0.01	0.01	0.01	7.44
NH ₄ (ppm)	0.02	0.69	0.072	0.2	0.98	0.017	0.082	0.013	0.53	0.033	0.01	0.056	0.86	0.076	0.076	1.04	1.52	0.01	0.16	0.59	0.47	0.79
Mg (ppm)	0.026	5.86	2.49	6.58	2.63	2.82	8.2	0.02	4.76	2.41	6.31	0.3	2.31	0.81	4.46	0.02	2.57	3.89	6.02	0.02	1.44	16.1
Mn (ppm)	0.06	0.01	0.045	0.066	0.01	0.076	0.23	0.01	0.019	0.15	0.13	0.11	0.01	0.01	0.01	0.01	0.01	0.24	0.74	0.15	0.028	0.53
S (ppm)	39.9	132	548	629	1010	324	708	958	710	304	585	973	949	97.2	90.7	71.4	182	213	271	178	845	974
Cu (ppm)	0.01	0.15	0.01	0.01	0.15	0.01	0.01	0.22	0.086	0.01	0.01	0.01	0.098	0.01	0.01	0.01	0.11	0.01	0.01	0.068	0.01	0.4
B (ppm)	6.8	14.6	103	126	197	55.7	126	184	116	52	108	184	169	13	12.4	11.2	20.4	37.8	43.6	28.2	147	160
Zn (ppm)	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.33	0.29
Na (ppm)	10,300	22,900	93,400	114,000	143,000	63,900	120,000	121,000	100,000	58,800	100,000	118,000	104,000	18,300	17,700	16,000	30,800	58,800	70,500	36,000	155,000	160,000
Fe (ppm)	2.79	0.53	0.01	1.34	0.01	2.13	9.63	0.91	0.92	2.42	3.17	1.59	0.17	0.25	2.5	0.01	0.25	4.1	14.7	3.82	0.86	0.49
Ca (ppm)	0.05	3.79	8.52	10.9	0.34	12	16.4	0.2	7.43	10.8	9.23	1.13	6.54	0.05	2.94	0.05	2.72	13.8	15.8	0.05	2.15	127
SO ₄ (ppm)	120	395	1,640	1,880	3,030	971	2,120	2,870	2,130	911	1,750	2,920	2,840	291	272	214	545	638	812	533	2,530	2,920
Mo (ppm)	0.067	0.25	1.66	2.49	3.82	1.03	2.43	4.99	2.23	0.83	1.74	4.5	3.09	0.048	0.23	2.09	0.24	0.29	0.4	1.68	2.37	2.79
Cl (ppm)	4,050	3,720	52,300	77,400	102,000	30,700	72,300	99,800	74,200	30,200	60,500	93,100	62,500	7,910	8,620	7,680	6,740	24,100	32,500	16,700	93,500	99,000
NO ₃ -N (ppm)	0.2	2.04	0.41	0.01	1.35	0.01	0.01	0.045	1.12	0.01	0.01	0.01	1.58	1.03	0.01	0.043	1.92	7.23	0.01	0.01	0.01	1.68
HCO ₃ (ppm)	17,300	15,400	145,000	205,000	213,000	94,300	229,000	242,000	143,000	94,700	196,000	233,000	165,000	30,100	30,200	29,300	25,300	98,200	141,000	70,600	256,000	277,000
Si (ppm)	66.2	114	447	568	786	282	569	861	590	284	458	867	760	89.9	94.9	194	107	106	197	287	595	696
SiO ₂ (ppm)	142	244	956	1,220	1,680	603	1,220	1,840	1,260	608	980	1,850	1,630	192	203	415	229	227	421	614	1,270	1,490
SAR (ppm)	9,310	1,720	7,230	6,730	18,200	4,310	6,040	69,000	7,050	4,210	6,220	25,500	8,900	4,290	1,520	15,300	3,220	3,600	3,830	34,400	20,100	3,550
CaCO ₃ (ppm)	0.23	33.5	31.5	54.2	11.6	41.6	74.6	0.58	38.1	36.9	48.9	4.05	25.8	3.45	25.6	0.21	17.3	50.4	64.2	0.21	11.3	384

Microbial community analysis

Sequences with assigned taxonomy were aligned using PyNast (Caporaso *et al.*, 2010a), and a phylogenetic tree was constructed using FastTree v2.1.7 (Price *et al.*, 2010). The alpha diversity indices (Chao1, abundance-based coverage estimator (ACE), Simpson, Shannon, Fisher's alpha, Pielou's evenness, and Good's coverage) were calculated with QIIME v1.90 (Caporaso *et al.*, 2010) using *alpha_rarefaction.py* employing the same level of surveying effort (37,000 per sample based on the lowest sample count). All subsequent steps were analyzed in R software v4.2.0 (R Core Team, 2020) and RStudio v1.1.456 (RStudio Team, 2020). The results of all statistical tests were regarded as significant if $p \leq 0.05$. To compare the (dis) similarity of OTU compositions between communities the OTU abundance table was standardized using decostand (method = "hellinger"). Hierarchical cluster analysis was performed using hclust in R software v4.2.0 (R Core Team, 2020) (method = "average"). The heatmap was created using JColorGrid v1.86 (Joachimiak *et al.*, 2006).

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

To assess the beta diversity of microbial communities, a non-metric multidimensional scaling (NMDS) was performed using Bray-Curtis dissimilarities with the script *compare_categories.py*. test and weighted UniFrac distance matrix (Lozupone & Knight, 2005) as input using the Vegan package in R (Bray & Curtis, 1957; Oksanen, 2015).

Results

Physicochemical properties of the sampling sites

One of the objectives of this study was to investigate the change in water chemistry over time. It has been established that physicochemical factors play a critical role in shaping the structural composition of microbial communities in an ecosystem. Samples from site S1 (hot spring water) exhibited lower concentrations of the various ions and cations as compared to the other samples. The water temperature ranged from 27°C to 38.7°C (average 33.7°C). The pH of the water was alkaline, ranging from 9.8 (S6_June) to 11.5 (BR1_June) recording the highest pH value of 11.5. The major water cations were Na^+ (10,300–160,000 ppm) and K^+ (131–4,280 ppm), and the major anions were HCO_3^- (15,400–277,000ppm) and Cl^- (4,050–102,000 mg/L). Phosphorus levels ranged from 2.38–108 ppm, while magnesium and calcium levels were low, ranging from 0.02–16.1 and 0.05–127 ppm, respectively. The total dissolved solids (TDS) ranged from 27.1–153.5 ppm (Table 1).

Sequence analysis and diversity studies

After quality filtering, denoising, and removal of potential chimeras and non-bacterial sequences, approximately 3,197,447 high-quality sequences with an average read length of 525 bp were obtained from the entire dataset. The number of sequences per sample varied from 37,406 (sample S5_Jun) to 285,085 (sample BR1_Sep) with an average value of 121,603 sequences. The number of OTUs per sample ranged from 852 (sample S3_July) to 2,024 (sample S5_Sep) (Table 2). Sequencing generated a total of 4,837 OTUs distributed in the domain Bacteria (3,802 OTUs) and Archaea (1,035 OTUs). Overall, most OTUs were found in the open waters samples of S5, while S4 had the least number of OTUs. The distribution of shared OTUs based on the month of sampling is shown in *Extended data*, Supplementary Figure 1 (Kipnyargis *et al.*, 2023a).

Alpha diversity studies

The values of the good's coverage estimator ranged from 81% (S5_Sep) and 96% (S3_Aug) suggesting that the sequencing process captured a significant number of dominant communities. Within the open water samples (S2–S6), site 5 samples collected across the seasons had the highest alpha diversity indices suggesting that S5 had the highest species richness and diversity. S3_Aug samples (open waters) had the lowest alpha diversity indices. Within the hot spring samples (S1), S1 samples collected in September had the highest species richness and diversity. Within the brine samples (BR1), Br1 samples collected in September had the highest species diversity and richness (Table 2).

The alpha diversity indices showed that high microbial diversity was recorded in August, followed by September, June, and July in that order (Figure 2).

Beta diversity studies

Beta diversity ordination based on Bray-Curtis dissimilarity showed that samples (except hot spring and brine samples) did not cluster based on the sampling site. Overall, all samples clustered together based on salinity and

Table 2. Number of sequences generated, OTUs, and diversity indices of 22 sampling sites in Lake Magadi.
OTUs, operational taxonomic units; ACE, abundance-based coverage estimator.

Sample	Number of Sequences	OTUS	Chao1	Ace	Simpson	Observed species	Shannon	Fisher alpha	Goods coverage
BR1_June	149,051	1,416	357.30	414.77	0.93	148	5.28	53.40	0.89
S1_June	191,597	1,089	171.50	175.11	0.93	89	4.63	25.63	0.94
S2_June	174,894	1,267	180.77	181.63	0.88	94	4.32	27.66	0.94
S3_June	139,647	1,513	251.65	319.48	0.94	131	5.15	44.51	0.91
S4_June	130,649	1,514	306.14	397.09	0.92	140	5.00	49.13	0.89
S5_June	37,406	1,221	363.23	452.68	0.96	209	6.16	92.00	0.85
S6_June	116,475	1,665	350.70	385.77	0.96	179	5.90	71.63	0.87
S2_July	177,773	1,259	301.33	300.15	0.91	98	4.49	29.32	0.92
S3_July	177,827	852	171.50	162.62	0.66	69	2.86	18.11	0.95
S4_July	156,660	1,167	166.62	189.02	0.90	87	4.31	24.84	0.94
S5_July	109,083	1,917	435.14	487.08	0.97	231	6.54	108.85	0.83
S6_July	103,156	1,149	181.32	192.37	0.86	106	4.19	32.76	0.93
S3_Aug	177,142	874	91.91	107.34	0.79	55	3.20	13.39	0.96
S4_Aug	172,666	934	175.86	143.57	0.84	70	3.70	18.46	0.95
S5_Aug	83,183	1,815	572.10	579.97	0.97	222	6.28	101.75	0.82
S6_Aug	64,022	1,732	497.95	457.95	0.98	200	6.57	85.59	0.86
BR1_Sep	285,085	1,710	323.50	356.33	0.95	178	5.75	71.00	0.88
S1_Sep	180,065	1,583	341.16	436.97	0.95	177	5.77	70.38	0.87
S2_Sep	170,143	1,114	184.23	213.10	0.89	90	4.34	26.03	0.94
S3_Sep	149,323	1,124	149.37	174.86	0.75	90	3.50	26.03	0.94
S4_Sep	151,058	956	185.17	135.22	0.80	68	3.26	17.76	0.95
S5_Sep	100,542	2,024	594.50	641.88	0.98	248	6.85	123.09	0.81

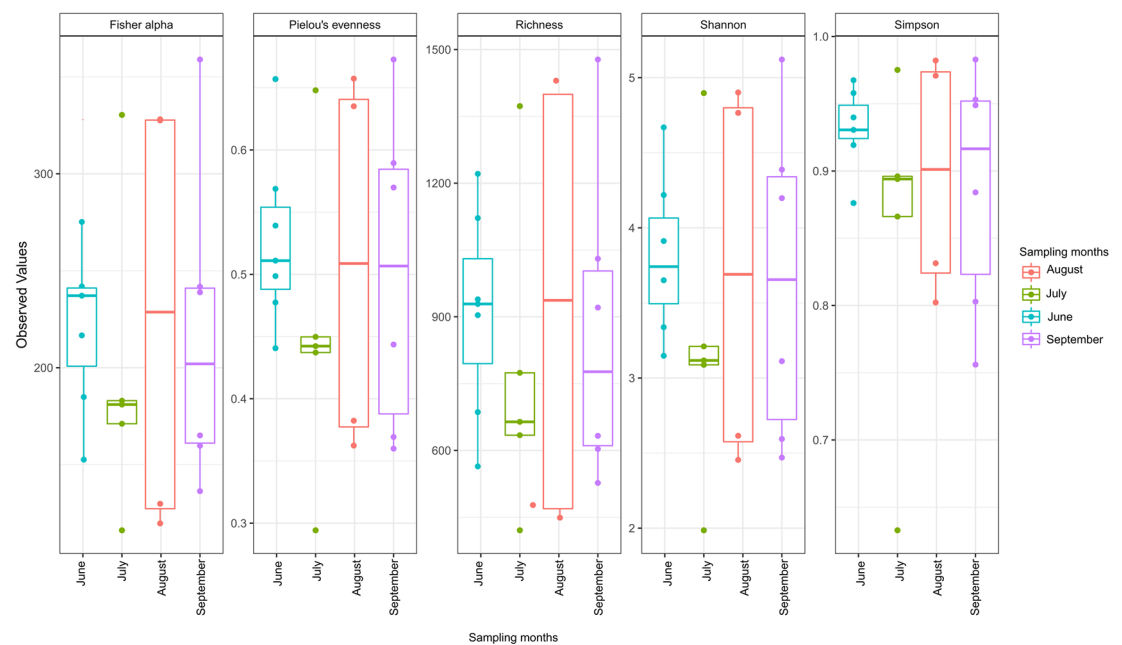


Figure 2. Alpha diversity plots of OTU richness among different sampling months.

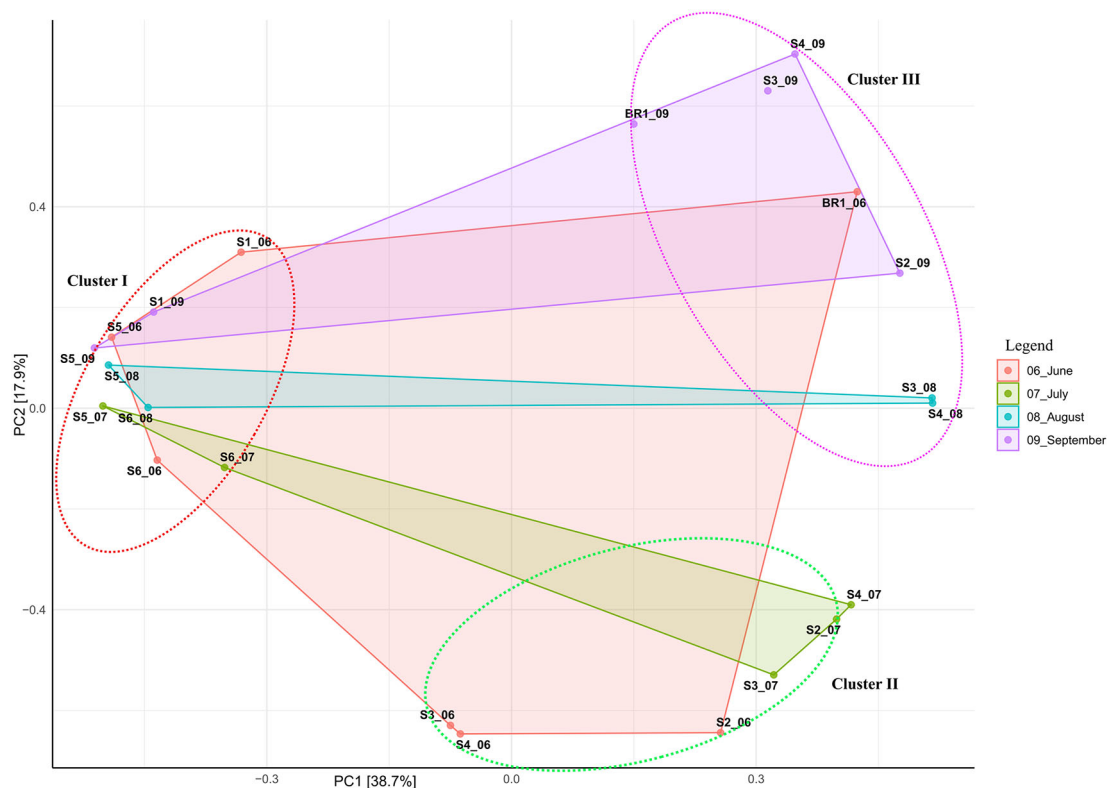


Figure 3. Principal component analysis (PCA) ordination of differences in microbial communities from sampling sites and sampling months.

alkalinity, indicating the impact of these elements on the structure of the bacterial and archaeal communities (Figure 3; Table 1). The principal component (PCA) analysis showed that the first (PC1) and second (PC2) axes described 38.7% and 17.9% of the variance in microbial communities, respectively. Accordingly, samples were clustered into three distinct groups based on alkalinity and salinity. Low alkalinity and salinity samples (pH 9.8 – 10.5; 10,300 ppm – 70,500 ppm) formed cluster I with nine samples (S1_06, S1_09, S5_06, S5_09, S5_08, S5_07, S6_08, S6_06, and S6_07). Moderately alkaline and saline samples (pH 10.5 – 10.6; 63,900 ppm – 100,000 ppm) formed cluster II with six samples (S3_06, S4_06, S2_06, S3_07, S2_07, and S4_07). Highly alkaline and saline samples (pH 10.7–11.5; >100,000 ppm) formed cluster III consisting of six samples (Br1_06; 155,000 ppm, Br1_09; 160,000 ppm, S4_09; 104,000 ppm, S2_09; 143,000 ppm, S3_08; 121,000 ppm, and S4_08; 118,000 ppm).

Taxonomic composition and structure

The proportion of bacteria to archaea varied by season and sampling month (Figure 4). The results indicate that the archaeal population increased as the ion concentration increased while bacteria abundance was higher where the ion concentration was lower (sites 1, 4, and 5) (Table 1). In hot spring water (S1), archaea abundance was the lowest while bacterial abundance was the highest. Within open water samples (S2–S6), S4 had the highest abundance of archaea, while S5 had the highest proportion of bacteria. Within the brine (BR1), the archaea proportion was relatively higher than the bacterial communities. From June to September 2018, bacterial abundance decreased while archaeal abundance increased (Figure 4A).

The bacterial reads were distributed across 25 phyla, 107 orders, 225 families, and 545 genera. The results revealed that the most abundant bacterial phyla across the sampling sites and the four months of sampling included Proteobacteria (35% of all the reads), Cyanobacteria (14.2%), Bacteroidetes (10.9%), Actinobacteria (5.2%), Firmicutes (2.7%), Verrucomicrobia (1.1%), Deinococcus-Thermus (0.6%), Spirochaetes (0.4%), and Chloroflexi (0.1%) were detected in lower abundances (Figure 4B). At the genus level, the dominant bacterial genera (> 1% of all sequences across all samples) were *Euihalotheca* (10.3%), *Rhodobaca* (9.6%), *Idiomarina* (5.8%), *Rhodothermus* (3.0%), *Roseinatronobacter* (2.4%), *Nocardioides* (2.3%), *Gracilimonas* (2.2%), *Halomonas* (2.0%), *Lewinella* (1.9%), *Synechococcus* (1.8%), *Aliidimarina* (1.8%), *Nitriliruptor* (1.7%), *Thioalkalivibrio* (1.7%), *Salinibacter* (1.4%), *Alkalimonas* (1.25%), *Che-latococcus* (1.4%), and *Rhodovulum* (1.4%). Others included: *Cytophaga* (0.9%), *Natronocella* (0.9%), *Thiohalomonas*

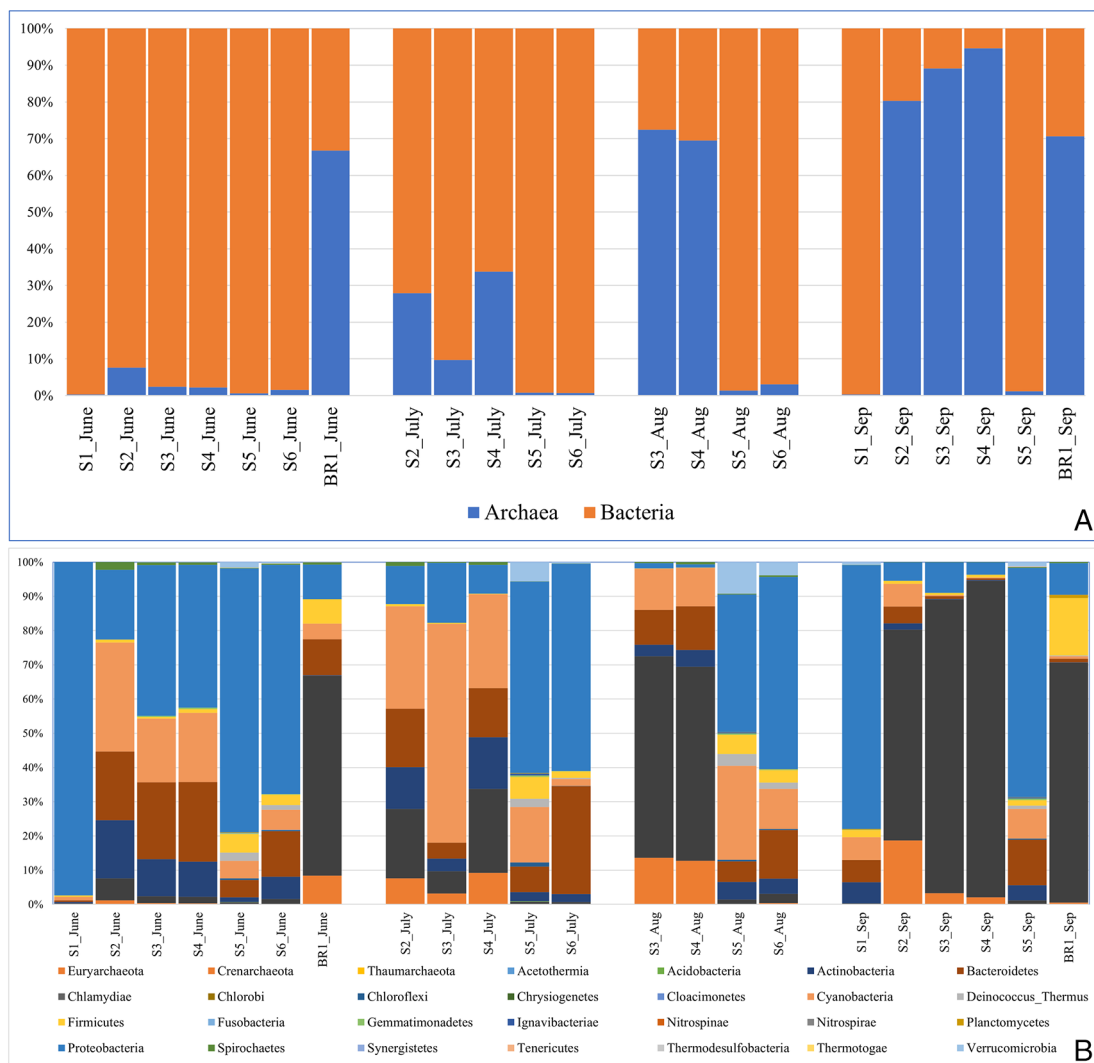


Figure 4. Relative abundance and composition of the microbial community taxa based on sampling sites and month of sampling. (A) The proportion of Domains bacteria and archaea across the sampling sites and months. (B) Relative abundance of the most popular bacterial and archaeal phyla across the sampling sites and the sampling months.

(0.9%), *Euzebya* (0.8%), *Paracoccus* (0.8%), and *Luteolibacter* (0.8%). The abundance of bacterial genera was higher in the sampling site S5 (25.3%) followed by S6 with 20.1%, S3 (14.5%), S1 and S4 with 12.8% each, and brine sample BR1 with 3.9% abundance in that order. In terms of the sampling month, June had the highest bacterial abundance with 39.5% followed by July (27%), September (16.8%), and August (16.2%) in that order.

The archaeal reads were affiliated to three phyla (Euryarchaeota, Crenarchaeota, and Thaumarchaeota) (Figure 4B), 14 orders, 20 families, and 62 genera. The dominant Phylum was Euryarchaeota (87% of all Archaeal samples), with its dominant genera (>1% of all sequences across all samples) being *Halorubrum* (18.3%), *Salinarchaeum* (5.4%), and *Haloterrigena* (1.3%). Other genera included *Methanomassiliicoccus* (0.6%), *Palaeococcus* (0.4%), *Halovenus* (0.3%), *Thermococcus* (0.3%), *Haladaptatus* (0.3%), *Halorientalis* (0.3%), *Methanobrevibacter* (0.2%), *Natronomonas* (0.2%), *Halohasta* (0.2%), *Haloquadratum* (0.1%), and *Methanobacterium* (0.1%). The abundance of archaeal genera was higher in S3 (27.2%) followed by brine site BR1 with 21.6% abundance. S1 and S6 had the least archaeal abundance with 0.08 and 0.8%, respectively. In terms of the sampling month, September had the highest archaeal abundance with 53% followed by August (23%), June (12.7%), and July (11.4%).

The bacterial species composition (>1%) included *Euhalothece* spp. (10.3%), *Rhodobaca* spp. (9.6%), *Idiomarina* spp. (5.8%), *Rhodothermus* spp. (3.0%), *Roseinatronobacter* spp. (2.4%), *Nocardioides* spp. (2.3%), *Gracilimonas* spp. (2.2%), *Halomonas* sp. (2%), *Lewinella* (1.9%), *Synechococcus* spp. (1.8%), *Cyanobacterium* spp. (1.8%), *Aliidiomarina* spp. (1.7%), *Nitriliruptor* spp. (1.7%), *Thioalkalivibrio* spp. (1.7%), *Salinibacter* spp. (1.4%), *Alkalimonas* spp. (1.2%), *Chelatococcus* spp. (1.1%), and *Rhodovulum* spp. (1.1%). The *Euhalothece natrophila* species were abundant in June, July, and August, except in sites S5 and S6 across all seasons. *Rhodobaca bogoriensis* was largely sampled in June and site S6 in July and August 2018. *Idiomarina* spp. were largely concentrated in June, particularly in sites S1 and S5, whereas *Rhodovulum* spp. were sampled across all seasons. *Lewinella coherens* were sampled in June mostly in sites S3 and S4. On the other hand, *Halorubrum* spp. (18.3%), *Salinarchaeum* spp. (5.3%), *Haloterrigena* spp. (1.3%), *Methanomassiliicoccus* spp. (0.7%), and *Palaeococcus* spp. (0.5%) were the major species in the Archaeal Domain. *Idiomarina vacuolatum* was sampled across all the sampling seasons but its abundance varied across the sampling sites. *Halorubrum vacuolatum* was mainly sampled in August (S1 and S2) and September (S3, S4, and S5). *Salinarchaeum* sp. was mainly sampled in September, while *Haloterrigena* spp. was sampled across the seasons and sites, though in low proportions. The top 30 most abundant species of bacteria and archaea are shown in [Figure 5](#). Overall, *Halorubrum* spp. was the most abundant species sampled followed by *Euhalothece* spp. and *Rhodobaca* spp. (Extended data, Supplementary Figure 2 ([Kipnyargis et al., 2023a](#))).

Physicochemical drivers of bacterial and archaeal community structure

Page 11 of 41

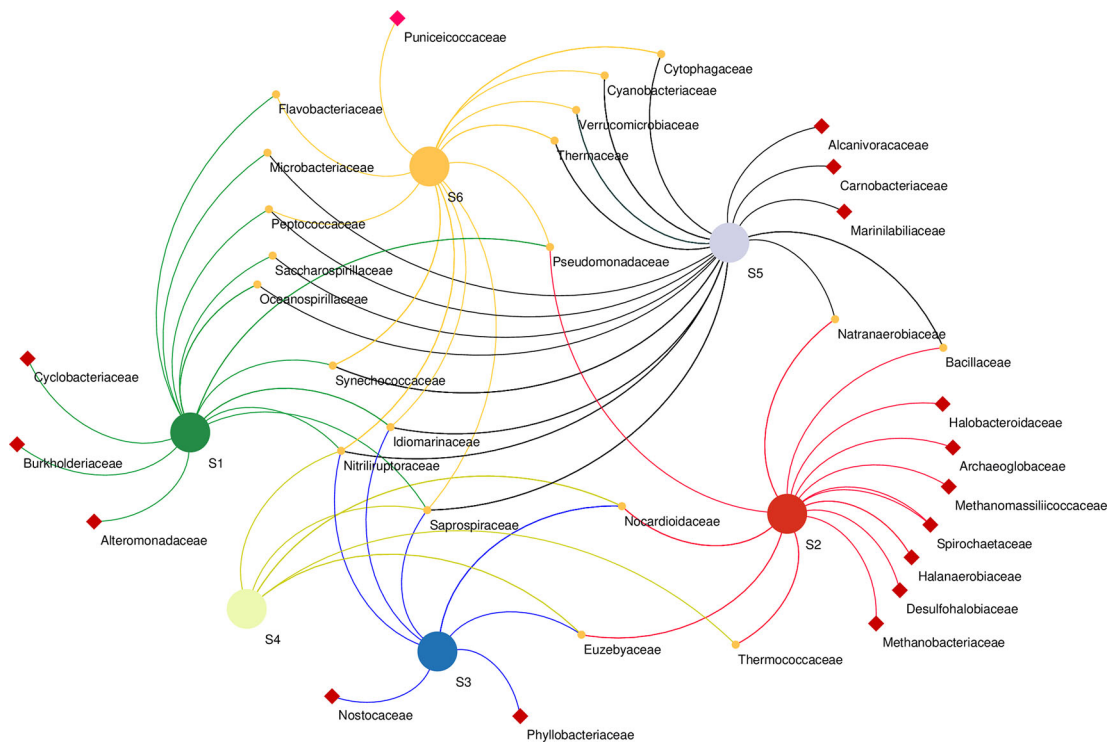


Figure 6. Network analysis of microbial community at Family level based on sampling sites. Samples marked with red squares indicate their exclusive site of isolation.

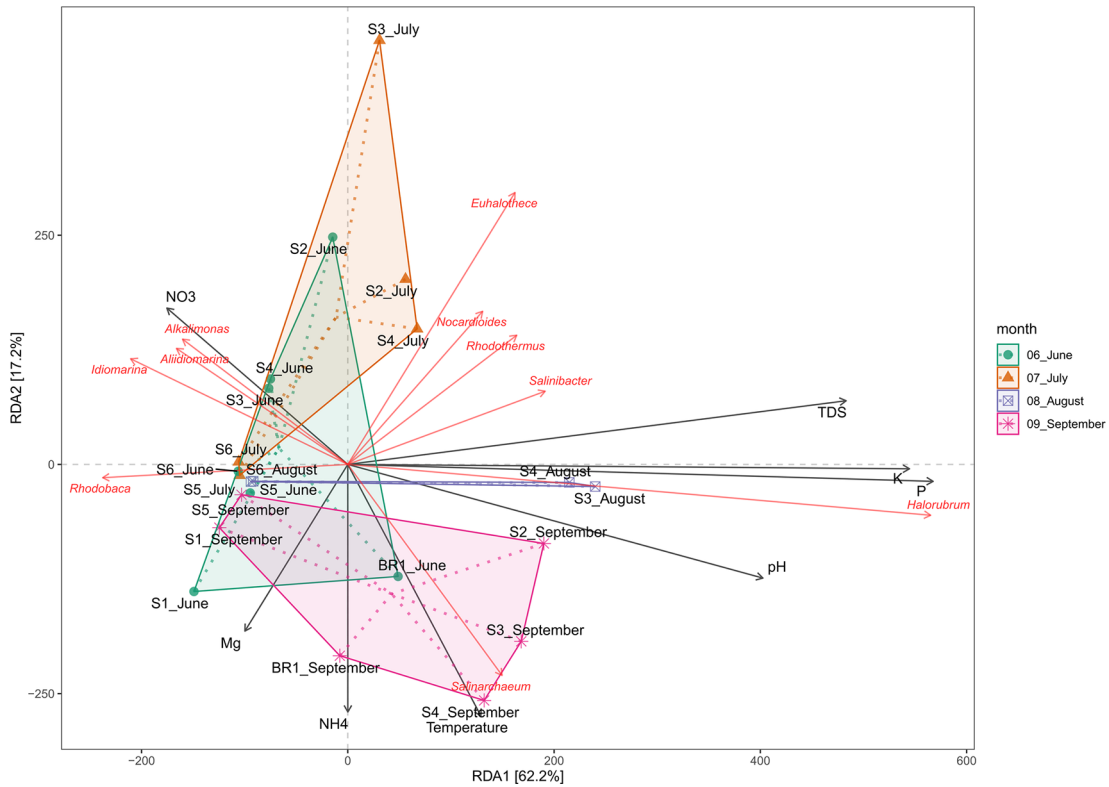


Figure 7. Redundancy Analysis (RDA) ordination showing the relationships between the physicochemical factors and the dominant genera in Lake Magadi. Samples corresponding to their sampling month are indicated, while genera are written in red, and environmental variables are indicated by arrows. TDS corresponds to total dissolved solids.

lake (Extended data, Supplementary Table 2 (Kipnyargis *et al.*, 2023a)). The RDA explained 62.2% and 17.2% of the variation in the first (RDA1) and the second (RDA2) axes, respectively (Figure 7).

The results demonstrated that temperature, pH, P, K, NO_3^- , and TDS significantly influenced the microbial community structure. Generally, members of genera *Nocardioides*, *Rhodothermus*, *Haloterrigena*, *Methanomasiliicoccus*, *Halorubrum*, *Palaeococcus*, *Nocardioides*, *Salinarcheum*, *Salinibacter*, and *Euhalothece* spp. had a wide range of adaptability. Conversely, *Synechococcus*, *Thioalkalivibrio*, *Cyanobacterium*, *Rhodovulum*, *Lewinella*, *Idiomarina*, *Pseudidiomarina*, *Chelatococcus*, *Aliidiomarina*, and *Alkalimonas* spp. were least influenced by the tested physicochemical factors (Extended data, Supplementary Figure 3 (Kipnyargis *et al.*, 2023a)). Notably, *Halorubrum* and *Haloterrigena* spp. were positively correlated with P and K ($R^2 = 0.66$, $p < 0.001$), but negatively correlated with Mn and CO_3^{2-} . pH and NH_4^+ appear to positively correlate with the structure of the members of the genus *Salinarcheum* ($R^2 = 0.245$; $p < 0.004$), but negatively correlated with NO_3^- . Members of *Alkalimonas*, *Idiomarina*, and *Aliidiomarina* spp. were positively correlated with NO_3^- ($R^2 = 0.049$, $p < 0.210$), but negatively correlated with all other tested parameters. Members of *Nocardioides*, *Rhodothermus*, *Salinarcheum*, *Salinibacter*, and *Euhalothece* spp. were positively correlated with total dissolved solids (TDS), alkalinity, salinity, CO_3^{2+} , and NH_4^+ ($R^2 = 0.606$, $p < 0.001$), but negatively correlated with Mg^{2+} , Mn, and NO_3^- . On the other hand, Mn, temperature CO_3^{2-} , and NH_4^+ negatively affect the structure of *Rhodobaca*.

Discussion

We explored the structure and composition of microbial communities based on the seasonality and physicochemical parameters of Lake Magadi. The physicochemical parameters revealed high concentrations of sodium salts, HCO_3^- , SO_4^{2-} , pH values of 9.8–11.5, temperatures of 27–38°C, and low concentrations of Ca^{2+} , Mg^{2+} , and Cu^{2+} . These findings were consistent with previous reports indicating that soda lakes are characterized by moderate to high temperatures, high concentrations of $\text{HCO}_3^-/\text{CO}_3^{2-}$, and reduced concentrations of Ca^{2+} and Mg^{2+} (Sorokin *et al.*, 2014a; Vavourakis *et al.*, 2018). However, total dissolved solids (TDS) ranged from 27 ppm (0.02g/L) to 143 ppm (0.143g/L), a situation that is lower than other soda lakes (Taher, 1999; Hosam *et al.*, 2017; Pérez & Chebude, 2017). Sulfate concentration (120–3,030 ppm) was similar to most soda lakes in East Africa (Lameck *et al.*, 2023), but higher than the concentration reported in Lake Lancago in Qinghai-Tibet (Wang *et al.*, 2022). Comparatively lower sulfate concentrations have been reported from lakes Sidi Ameur and Himalatt (Algeria) (Boutaiba *et al.*, 2012) and Lake Hamra in Egypt (Mahmoud *et al.*, 2024). The concentrations of the measured elements (except pH) were variable from site to site and fluctuated with time, indicating that the lake chemistry is constantly changing in its constituent elements. A high and stable pH recorded in Lake Magadi is due to high amounts of carbonates that maintain a constant pH in soda lake ecosystems (Simachew *et al.*, 2016). It is postulated that Ca^{2+} and Mg^{2+} precipitate as insoluble carbonates due to high evaporation rates in these ecosystems. As a result, an alkaline brine with Na^+ , Cl^- , and $\text{HCO}_3^-/\text{CO}_3^{2-}$ accumulates as main ions. The shift in $\text{CO}_2/\text{HCO}_3^-/\text{CO}_3^{2-}$ equilibrium towards CO_3^{2-} , leads to the formation of a soda (Na_2CO_3) lake with pH values of over 10.0 (Grant and Jones, 2016).

A high number of OTUs was detected for the domain Bacteria with 3,802 OTUs while Archaea had 1,035 OTUs. Bacterial diversity was dominated by the phyla Proteobacteria (35% of all the reads), Cyanobacteria (14.2%), Bacteroidetes (10.9%), Actinobacteria (5.2%), Firmicutes (2.7%), Verrucomicrobia (1.1%). On the other hand, Deinococcus-Thermus (0.6%), Spirochaetes (0.4%), and Chloroflexi (0.1%) were detected in lower abundances in Lake Magadi. Similar results have been shown from soda ecosystems such as Solar saltern in Tunisia (Menasria *et al.*, 2019), lake Chott El Jerid (Abdallah *et al.*, 2018), hot springs of Lake Magadi (Kambura *et al.*, 2016), and lakes Sonachi, Magadi, Elmenteita, and Bogoria in Kenya (Mwirichia, 2022). Members of the phylum Proteobacteria were the most dominant group across all the time points and sampling sites. The role of members of Proteobacteria such as *Burkholderiaceae* is to decompose recalcitrant organic matter while others like *Beijerinckiaceae* fix atmospheric nitrogen (Li *et al.*, 2012). The phylum Cyanobacteria, the major contributor to nitrogen fixation in soda lakes (Sorokin & Kuenen, 2005), was represented mainly by the *Euhalothece* spp. *Euhalothece* is a single-celled stenohaline cyanobacterium growing optimally at 7% (w/v) NaCl. They depict a morphological variability depending on the concentrations of NaCl and carbonates as well as the pH conditions (Mikhodyuk *et al.*, 2008). The presence of *Euhalothece* in Lake Magadi is thus supported by high salts and carbonates (Table 1). Contrastingly, the results in other soda lakes have shown that *Arthrospira* spp. are the main photosynthetic agents driving primary productivity, with the seasonal occurrence of *Cyanospira*, *Synechococcus*, and *Chroococcus* to augment this process (Jones *et al.*, 1998). Thirdly, members of the phylum Bacteroidetes were majorly represented by the genera *Rhodothermus*, *Roseinatronbacter*, *Gracilimonas*, *Lewinella*, and *Cytophaga*. Verrucomicrobia was represented by *Verrucomicrobium*, *Puniceicoccus*, and *Coralimargarita*. Bacteroidetes and Verrucomicrobia are known to thrive in high-nutrient environments where they play a role in the degradation of biopolymers such as cellulose and chitin (Newton *et al.*, 2011). Interestingly, the presence of Bacteroidetes and Verrucomicrobia was often associated with the availability of Cyanobacteria across the sites and the sampling periods (Extended data, Supplementary Table 1 (Kipnyargis *et al.*, 2023a)). Photosynthetic stages of cyanobacteria lead to a high

rate of CO_2 and HCO_3^- consumption, and a consequent increase in pH (Almeida *et al.*, 2011). As a result, nutrient release is enhanced from the sediments of the lake, hence acting as substrates for Bacteroidetes and Verrucomicrobia. Phylum Firmicutes (2.7% of all the reads) was represented by members of Class Clostridium (*Clostridium*, *Halanaerobium*, *Natranaerobius*, and *Moorella* sp.) and Bacilli (*Alkalibacterium* and *Anoxybacillus* sp.).

The archaeal community of the Lake Magadi microbiome was represented by three phyla, the Euryarchaeota (25% of all the reads), Crenarchaeota (0.01%), and Thaumarchaeota (0.01%). Generally, archaea were more abundant in brine samples with Br1_June and Br1_Sept accounting for 24.5% of total archaea. Previous studies have indicated that archaea are more adapted to saline environments than bacteria (Mani *et al.*, 2020). Euryarchaeota was the most abundant phylum across the sites and the sampling seasons with 87% of all archaeal communities. Euryarchaeota has well-adapted inhabitants of hypersaline environments where they play a critical role in ecosystem services such as carbon cycling by functioning as methanogens (Jiang *et al.*, 2007; Vavourakis *et al.*, 2016). Grant *et al.* (1999) first characterized this phylum in the alkaline saltern of Lake Magadi. The second most abundant group was Thaumarchaeota also known as ammonia-oxidizing agents (Andreote *et al.*, 2018). In marine environments, their distribution along a salinity gradient has been linked to changes in location, salinity, and sediment depth (Webster *et al.*, 2015). A one-time sampling of the Lake Magadi hot springs depicted a similar picture with Euryarchaeota accounting for up to 28% abundance while Thaumarchaeota and Crenarchaeota were each 1% abundant (Kambura *et al.*, 2016). Studies in other soda lakes that have detected members of the phylum Euryarchaeota and Crenarchaeota (Ghori *et al.*, 2021), Crenarchaeota, Euryarchaeota, Woesearchaeota, and Pacearchaeota (Wang *et al.*, 2022). However, explorations in Soda Lake in Salina Preta, Brazil, revealed the contrary where members of Thaumarchaeota were most abundant while Euryarchaeota was least abundant (Andreote *et al.*, 2018). The most abundant archaeal genera belonged to *Halorubrum* (18.3%), *Salinarchaeum* (5.4%), and *Haloterrigena* (1.3%). Most of these microbes have their habitats in soda lakes and neutral saline environments (Feng *et al.*, 2005; Mwatha & Grant, 2016; Minegishi *et al.*, 2017; Zhao *et al.*, 2020). Other haloalkaliphilic archaea related to genera *Natronomonas*, *Natrialba*, *Natrococcus*, *Natronobacterium*, *Natronolimnobi*, and *Halorubrum*, all of whom were detected in this study, have previously been isolated from brines of East African soda lakes and Inner Mongolian lakes where salinity values reach > 30%, and pH values of > 10 (Grant and Sorokin, 2011). Overall, the results of this study reflect bacterial composition in many soda lakes around the world (Sorokin *et al.*, 2014a; Kambura *et al.*, 2016; Mani *et al.*, 2020; Poyraz & Mutlu, 2020; Wang *et al.*, 2022).

Co-occurrence network analysis demonstrates the interactions between microbial taxa, which can be symbiotic or competitive (He *et al.*, 2019). At the family level revealed the presence of heterogeneous microbial communities that co-occur in different sampling sites along the lake as well as others that were unique to a particular site, suggesting mutual interactions of these communities across the sites. For instance, Desulfobacteriaceae were unique to S2. Correspondingly, S2 had the highest concentration of both sulfur and sulfate ions (Table 1). This family, particularly Desulfonatronum, Desulfonatronospira, Desulfonatronovibrio, and Desulfobacteriaceae have been shown to thrive in anoxic parts of soda lakes acting as sulfate-reducing bacteria (SRB) through the oxidation of hydrogen and formate or direct disproportionation of sulfite of thiosulfate (Sorokin *et al.*, 2011). Noteworthy, *Thioalkalivibrio* sp. was not significantly influenced by the physicochemical properties investigated. This suggests that these strains have devised adaptive mechanisms to thrive under the prevailing harsh conditions of Lake Magadi. Indeed, *Thioalkalivibrio* sp., which are sulfur-oxidizing bacteria (SOB), have been found to adapt well in soda lake ecosystems (Li *et al.*, 2022). Unique to the S6 site were the members of the family Puniceicoccaceae which have also been described in four soda lakes of the Cariboo Plateau in Canada (Zorz *et al.*, 2019). Cyclobacteriaceae retrieved from the S1 site have established habitats in diverse ecosystems like cold marine regions like algal/microbial mats, haloalkaline soda lakes, Antarctica, freshwater bodies, marine waters, marine sediments, mangroves, hot springs, and mud volcanoes (Rosenberg *et al.*, 2014). Members of families Rhodobacteraceae and Cyclobacteriaceae have sulfate-oxidizing properties, whereas Burkholderiaceae (unique to S1) have adapted to different ecological niches and are involved in processes such as catabolism of aromatic compounds as well as nitrogen fixation (Pérez-Pantoja *et al.*, 2012).

Alpha diversity studies revealed that samples in the open waters, particularly S5, had the highest species richness and diversity. However, open waters samples from S2–S4 depicted varying degrees of microbial diversity. While this could not be conclusively explained, it has been suggested that water waves affect the distribution and degradation of organic matter degradation and nutrient cycling, thereby influencing the composition diversity of microbes in water ecosystems (Zhu *et al.*, 2018). Brine samples (BR1) in June and September had relatively high diversity and evenness indices. Despite high salinity and alkalinity in soda lake brine, the presence of high light intensity and dissolved CO_2 promotes the growth of photosynthetic microorganisms. Subsequently, these phototrophs generate large quantities of dissolved organic matter (DOM) which become substrates for sustaining the diverse microbial communities (Banda *et al.*, 2020). Moreover, hot spring samples (S1) exhibited high Shannon and Simpson diversity indices, an indication of high microbial diversity. Research on Soda Lake hot springs revealed a highly active and diverse microbial community, suggesting the high plasticity of these organisms toward extreme environments (Dadheech *et al.*, 2013; Kambura *et al.*, 2016).

The principal component analysis revealed that compared with other samples, hot spring (S1) and brine (BR1) samples clustered according to their sites (cluster I and II, respectively) (Figure 3). This indicates that hot spring and brine samples had distinct community similarities (Tao *et al.*, 2019). However, samples appeared to cluster based on the changes in the salinity and alkalinity of the sampling sites from low to high salinity and alkalinity clusters. This is an indication that the impact of time point sampling on the community structure was limited. Previous studies have established that salinity is the primary selective force driving the distribution of beta diversity, whereas alkalinity influences microbial richness (Antony *et al.*, 2013; Boros & Kolpakova, 2018; Banda *et al.*, 2020). Moreover, extreme salinity and alkalinity confine the microbial communities to a few taxa highly adapted to the prevailing conditions (Oren, 2011).

In terms of water chemistry (Figure 7; Extended data, Supplementary Figure 3 (Kipnyargis *et al.*, 2023a)), pH, temperature, PO_4^{3-} , K, and NO_3^- , NH_4^+ , Mn, Na^+ , SO_4^{2-} , and TDS influenced the variation of microbial community composition in Lake Magadi. Salinity and alkalinity tend to influence the distribution of the microbial communities in the Soda Lake ecosystem. Specifically, members of genera *Nocardiodes*, *Rhodothermus*, *Haloterrigena*, *Methanomasiliicoccus*, *Halorubrum*, *Palaeococcus*, *Nocardioides*, *Salinarchaeum*, *Salinibacter*, and *Euhalothece* had a wide range of physicochemical adaptability. Conversely, *Synechococcus*, *Thioalkalivibrio*, *Cyanobacterium* spp., *Rhodovulum*, *Lewinella*, *Idiomarina*, *Pseudidiomarina*, *Chelatococcus*, *Aliidiomarina*, and *Alkalimonas* were least influenced by the tested physicochemical factors (Extended data, Supplementary Figure 3 (Kipnyargis *et al.*, 2023a)). The archaeal genera *Salinarchaeum* and *Halorubrum* *Halobellus*, *Halolamina*, *Methanobrevibacter*, and *Halorhabdus* have been strongly associated with salinity and factors such as pH, Mg^{2+} , Na^+ , K^+ , Ca^{2+} , and SO_4^{2-} (Han *et al.*, 2017). Nitrate appears to drive the structure of the members of the genera *Aliidiomarina*, *Idiomarina*, and *Alkalimonas* (Figure 4, Extended data Supplementary Figure 3 (Kipnyargis *et al.*, 2023a)). Many strains of *Alkalimonas* have been isolated from Chahannor (China), Kulunda Steppe (Russia), and Elementaita (Kenya) soda lakes where they play a role in nitrate reduction and formation of H_2S (Ma *et al.*, 2004; Vavourakis *et al.*, 2016). However, in this study, sulfur was negatively correlated with *Alkalimonas*. The *Aliidiomarina* and *Idiomarina* belong to the family Idiomarinaceae and have also been described as nitrogen reducers but poor in carbohydrate utilization (Chiu *et al.*, 2014).

Conclusion

Studies combining ecological, physiological, and taxonomic aspects have shown the remarkable diversity of halocaliophiles in numerous saline and alkaline lakes. The current study determined the structure of the microbial communities in different sampling sites and time points in Lake Magadi and explored the influence of physicochemical parameters on the composition and diversity of these microorganisms. Our findings indicated that results revealed that in comparison to archaea, bacteria are much more diverse in this ecosystem. The bacterial community was dominated by the phyla Proteobacteria, Cyanobacteria, Bacteroidetes Actinobacteria, Firmicutes, and Verrucomicrobia. On the other hand, archaea were majorly composed of Euryarchaeota while Crenarchaeota, and Thaumarchaeota were found in the least abundance across the sampling seasons and sites. Species richness and diversity varied within open waters while samples from brine (Br_1) and hot spring (S1) revealed high diversity and evenness was recorded. Ordination studies indicated that samples clustered based on salinity and alkalinity rather than the sampling season. Water chemistry depicted high concentrations of sodium salts, HCO_3^- , SO_4^{2-} , pH values of 9.8–11.5, temperatures of 27–38°C, and low concentrations of Ca^{2+} , Mg^{2+} , and Cu^{2+} . Temperature, pH, P, K, NO_3^- , and TDS significantly influenced the microbial community distribution in Lake Magadi. Our research highlights the composition and diversity of microbial communities across four successive dry months of soda Lake Magadi and the key physicochemical factors influencing them. We recommend future research focus on the functional profiles of samples during and after cyanobacterial blooms, including vertical profile stratification of Lake Magadi. The inclusion of sediment samples will also elucidate the taxonomic and functional profile of anoxic microbial communities.

Data availability

Underlying data

NCBI BioProject: Prokaryotic diversity within the hypersaline Lake Magadi in Kenya. Accession number: PRJNA962270. <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA962270/> (Kipnyargis *et al.*, 2023b).

Extended data

Figshare: Structure and composition of the microbial communities in hypersaline Lake Magadi: Additional Materials. <https://doi.org/10.6084/m9.figshare.22699456> (Kipnyargis *et al.*, 2023a).

This project contains the following extended data:

- Supplementary Figure 1.png (The distribution of shared operational taxonomic units (OTUs) based on the month of sampling in Lake Magadi.)

- Supplementary Figure 2.png (Percentage read abundance of the top 20 species of the microbial communities collected from Lake Magadi.)
- Supplementary Figure 3.png (The influence of physicochemical parameters on the structure of microbial communities in Lake Magadi.)
- Supplementary Table 1.xlsx (Percentage abundance of bacterial and archaeal communities in Lake Magadi across the sampling months, broken down by phylum.)
- Supplementary Table 2.docx (Mantel test results of the effects of physicochemical factors on microbial structure and composition in Lake Magadi.)

Data are available under the terms of the [Creative Commons Attribution 4.0 International license](#) (CC-BY 4.0).

References

- Abdallah MB, Karray F, Kallel N, *et al.*: **Abundance and diversity of prokaryotes in ephemeral hypersaline lake Chott El Jerid using Illumina Miseq sequencing, DGGE and qPCR assays.** *Extremophiles*. 2018; **22**: 811–823.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Aguilar A, Ingemansson T, Magnien E: **Extremophile microorganisms as cell factories: Support from the European Union.** *Extremophiles*. 1998; **2**(3): 367–373.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Almeida TIR, Calijuri MC, Falco PB, *et al.*: **Biogeochemical process and the diversity of Nhecolândia lakes, Brazil.** *An. Acad. Bras. Ciênc.* 2011; **83**: 391–407.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Andreote APD, Dini-Andreote F, Rigonato J, *et al.*: **Contrasting the genetic patterns of microbial communities in soda lakes with and without cyanobacterial bloom.** *Front. Microbiol.* 2018; **9**(244): 1–13.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Antony CP, Kumaresan D, Hunger S, *et al.*: **Microbiology of Lonar Lake and other soda lakes.** *ISME J.* 2013; **7**(3): 468–476.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Antranikian G, Vorgias CE, Bertoldo C: **Extreme environments as a resource for microorganisms and novel biocatalysts.** *Adv. Biochem. Eng. Biotechnol.* 2005; **96**(8): 219–262.
[Publisher Full Text](#)
- Banda JF, Lu Y, Hao C, *et al.*: **The Effects of Salinity and pH on Microbial Community Diversity and Distribution Pattern in the Brines of Soda Lakes in Badain Jaran Desert, China.** *Geomicrobiol. J.* 2020; **37**(1): 1–12.
[Publisher Full Text](#)
- Behr HJ, Röhrlich C: **Record of seismotectonic events in siliceous cyanobacterial sediments (Magadi cherts), Lake Magadi, Kenya.** *Int. J. Earth Sci.* 2000; **89**(2): 268–283. [doi:000089932800007](#).
[Publisher Full Text](#)
- Boros E, Kolpakova M: **A review of the defining chemical properties of soda lakes and pans: An assessment on a large geographic scale of Eurasian inland saline surface waters.** *PLoS One*. 2018; **13**(8): e0202205.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Boutaiba S, Hacene H, Bidle KA, *et al.*: **Salt Lakes of the Algerian Sahara.** *J. Arid Environ.* 2012; **75**(10): 909–916.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bray JR, Curtis JT: **An ordination of the upland forest communities of southern Wisconsin.** *Ecol. Monogr.* 1957; **27**(4): 325–349.
[Publisher Full Text](#)
- Canganella F, Wiegand J: **Extremophiles: From abyssal to terrestrial ecosystems and possibly beyond.** *Naturwissenschaften*. 2011; **98**(4): 253–279.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Caporaso JG, Bittinger K, Bushman FD, *et al.*: **PyNAST: A flexible tool for aligning sequences to a template alignment.** *Bioinformatics*. 2010a; **26**(2): 266–267.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Caporaso JG, Kuczynski J, Stombaugh J, *et al.*: **QIIME allows analysis of high-throughput community sequencing data.** *Nat. Methods*. 2010b; **7**(5): 335–336.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Caporaso JG, Lauber CL, Walters WA, *et al.*: **Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms.** *ISME J.* 2012; **6**(8): 1621–1624.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Chiu HH, Rogozin DY, Huang SP, *et al.*: **Aliidiomarina shirensis sp. nov., a halophilic bacterium isolated from Shira Lake in Khakasia, southern Siberia, and a proposal to transfer Idiomarina maris to the genus Aliidiomarina.** *Int. J. Syst. Evol. Microbiol.* 2014; **64**(4): 1334–1339.
[Publisher Full Text](#)
- Dadheech PK, Glöckner G, Casper P, *et al.*: **Cyanobacterial diversity in the hot spring, pelagic, and benthic habitats of a tropical soda lake.** *FEMS Microbiol. Ecol.* 2013; **85**(2): 389–401.
[PubMed Abstract](#) | [Publisher Full Text](#)
- De Maayer P, Anderson D, Cary C, *et al.*: **Some like it cold: Understanding the survival strategies of psychrophiles.** *EMBO Rep.* 2014; **15**(5): 508–517.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Duckworth AW, Grant WD, Jones BE, *et al.*: **Halomonas magadii sp. nov., a new member of the genus Halomonas, isolated from a soda lake of the East African Rift Valley.** *Extremophiles*. 2000; **4**(1): 53–60.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Edgar RC, Haas BJ, Clemente JC, *et al.*: **UCHIME improves sensitivity and speed of chimera detection.** *Bioinformatics*. 2011; **27**(16): 2194–2200.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Edwin N, Akhwale JK, Mwirichia R: **Phylogenetic and morphological diversity of culturable cyanobacteria from Lake Magadi in Kenya.** *Afr. J. Biol. Sci.* 2019; **01**(4): 24–31.
[Publisher Full Text](#)
- Eugster H: **Origin and Deposition of Trona.** *Rocky Mt. Geol.* 1971; **10**(1): 49–55.
- Eugster HP: **Lake Magadi, Kenya, and its precursors.** *Dev. Sedimentol.* 1980; **28**(C): 195–232.
[Publisher Full Text](#)
- Feng J, Zhou P, Zhou YG, *et al.*: **Haloarubrum alkaliphilum sp. nov., a novel haloalkaliphile isolated from a soda lake in Xinjiang, China.** *Int. J. Syst. Evol. Microbiol.* 2005; **55**(1): 149–152.
[Publisher Full Text](#)
- Getenet M, García-Ruiz JM, Otálora F, *et al.*: **A Comprehensive Methodology for Monitoring Evaporitic Mineral Precipitation and Hydrochemical Evolution of Saline Lakes: The Case of Lake Magadi Soda Brine (East African Rift Valley, Kenya).** *Cryst. Growth Des.* 2022; **22**(4): 2307–2317.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ghori N, Michael JW, Andrew SW: **Temporal microbial community dynamics within a unique acid saline lake.** *Front. Microbiol.* 2021; **12**: 649594.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Grant WD, Sorokin D: **Distribution and Diversity of Soda Lake Alkaliphiles.** *Extremophiles Handbook*. 2011; (July 2015): 28–54.
[Publisher Full Text](#)
- Grant WD, Jones BE: **Soda lakes of East Africa.** *Soda Lakes of East Africa*. 2016; 1–408.
[Publisher Full Text](#)
- Grant S, Grant WD, Jones BE, *et al.*: **Novel archaeal phylotypes from an East African alkaline saltern.** *Extremophiles*. 1999; **3**(2): 139–145.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Grant WD: **Alkaline environments and biodiversity.** *Encyclopedia of Life Support Systems*. 2006; **20**.

- Han R, Zhang X, Liu J, *et al.*: **Microbial community structure and diversity within hypersaline Keke Salt Lake environments.** *Can. J. Microbiol.* 2017; **63**(11): 895–908.
[Publisher Full Text](#)
- He H, Fu L, Liu Q, *et al.*: **Community Structure, Abundance and Potential Functions of Bacteria and Archaea in the Sansha Yongle Blue Hole, Xisha, South China Sea.** *Front. Microbiol.* 2019; **10**(2404): 1–16.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hosam EE, Mohamad SAE-K, Mai AW: **Biodiversity of uncultured bacteria in hypersaline lakes, Siwa oasis, Egypt, as determined by polymerase chain reaction and denaturing gradient gel electrophoresis (PCR-DGGE) of 16S rRNA gene phylotypes.** *Int. J. Genet. Mol. Biol.* 2017; **9**(2): 8–15.
[Publisher Full Text](#)
- Jiang H, Dong H, Yu B, *et al.*: **Microbial response to salinity change in Lake Chaka, a hypersaline lake on Tibetan plateau.** *Environ. Microbiol.* 2007; **9**(10): 2603–2621.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Joachimski MP, Weisman JL, May BCH: **JColorGrid: Software for the visualization of biological measurements.** *BMC Bioinformatics.* 2006; **7**: 1–5.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Jones BE, Grant WD, Duckworth AW, *et al.*: **Microbial diversity of soda lakes.** *Extremophiles.* 1998; **2**: 191–200.
[Publisher Full Text](#)
- Kambura AK, Mwirichia RK, Kasili RW, *et al.*: **Bacteria and Archaea diversity within the hot springs of Lake Magadi and Little Magadi in Kenya.** *BMC Microbiol.* 2016; **16**(1): 1–12.
[Publisher Full Text](#)
- Kamekura M, Dyall-smith ML, Upasani V, *et al.*: **Diversity of Alkaliphilic Halobacteria: Proposals for Transfer of Na trono bacterium vacuola tum, Na trono ba cterium magadii, and Natronobacterium pharaonis to Halorubrum, Natrialba, and Natronomonas gen. nov., Respectively, as Halorubrum.** *Int. J. Syst. Bacteriol.* 1997; **47**(7): 853–857.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kevbrin VV, Zhilina TN, Rainey FA, *et al.*: **Tindallia magadii gen. nov., sp. nov.: An alkaliphilic anaerobic ammonifier from Soda Lake Deposits.** *Curr. Microbiol.* 1998; **37**(2): 94–100.
[Publisher Full Text](#)
- Kiplimo D, Mugwero J, Kituyi S, *et al.*: **Diversity of esterase and lipase producing haloalkaliphilic bacteria from Lake Magadi in Kenya.** *J. Basic Microbiol.* 2019; **59**(12): 1173–1184.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kipnyargis A, Eucharia K, Fathiya K, *et al.*: **Spatiotemporal structure and composition of the microbial communities in hypersaline Lake Magadi, Kenya: Extended data.** *Figshare.* 2023a.
[Reference Source](#)
- Kipnyargis A, Mwirichia R, Eucharia K, *et al.*: **ID 962270 - BioProject - NCBI.** 2023b.
[Reference Source](#)
- Kipnyargis A, Mwirichia R, Pfeiffer B, *et al.*: **Complete Genome Sequence of Alkalihalobacillus sp. Strain LMS39, a Haloalkaliphilic Bacterium Isolated from a Hypersaline Lake.** *Microbiol. Resour. Announc.* 2022; **11**(7): e0032522.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Krenitz L, Shagerl M: **Tiny and Tough: Microphytes of East African Soda Lakes.** Schagerl M, editor. *Soda Lakes of East Africa.* Springer; 2016; pp. 149–177.
- Krulwich TA, Sachs G, Padan E: **Molecular aspects of bacterial pH sensing and homeostasis.** *Nat. Rev. Microbiol.* 2011; **9**(5): 330–343.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Lameck AS, Skutai J, Boros E: **Review of chemical properties of inland soda and saline waters in East Africa (rift valley region).** *J. Hydrol.: Reg. Stud.* 2023; **46**: 101323.
[Publisher Full Text](#)
- Li R, Khafipour E, Krause DO, *et al.*: **Pyrosequencing Reveals the Influence of Organic and Conventional Farming Systems on Bacterial Communities.** *PLoS One.* 2012; **7**(12): e51897.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Li X, Yang M, Mu T, *et al.*: **Composition and key-influencing factors of bacterial communities active in sulfur cycling of soda lake sediments.** *Arch. Microbiol.* 2022; **204**(6): 317–319.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Liu C, Cui Y, Li X, *et al.*: **Microeco: An R package for data mining in microbial community ecology.** *FEMS Microbiol. Ecol.* 2021; **97**(2): 1–9.
[Publisher Full Text](#)
- Lozupone C, Knight R: **UniFrac: A new phylogenetic method for comparing microbial communities.** *Appl. Environ. Microbiol.* 2005; **71**(12): 8228–8235.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ma Y, Xue Y, Grant WD, *et al.*: **Alkalimonas amylolytica gen. nov., sp. nov., and Alkalimonas delamerensis gen. nov., sp. nov., novel alkaliphilic bacteria from soda lakes in China and East Africa.** *Extremophiles.* 2004; **8**(3): 193–200.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Mahmoud AMA, Ali MHH, Abdelkarim MS, *et al.*: **Chemical, biochemical, and bioactivity studies on some soda lakes, Wadi El-Natron, Egypt.** *Environ. Monit. Assess.* 2024; **196**: 436.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Mani K, Taib N, Hugoni M, *et al.*: **Transient dynamics of archaea and bacteria in sediments and brine across a salinity gradient in a solar saltern of Goa, India.** *Front. Microbiol.* 2020; **11**(August): 1–19.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Matagi SV: **A biodiversity assessment of the Flamingo Lakes of eastern Africa.** *J. Limnol.* 2004; **5**(1): 13–26.
[Publisher Full Text](#)
- Melack JM, Kilham P: **Photosynthetic rates of phytoplankton in East African alkaline, saline lakes1.** *Limnol. Oceanogr.* 1974; **19**(5): 743–755.
[Publisher Full Text](#)
- Menasria T, Monteoliva-Sánchez M, Benammar L, *et al.*: **Culturable halophilic bacteria inhabiting Algerian saline ecosystems: A source of promising features and potentialities.** *World J. Microbiol. Biotechnol.* 2019; **35**(9): 132.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Mikhodyuk OS, Gerasimenko LM, Akimov VN, *et al.*: **Ecophysiology and polymorphism of the unicellular extremely natronophilic cyanobacterium Euhalothece sp. Z-M001 from Lake Magadi.** *Microbiology.* 2008; **77**(6): 717–725.
[Publisher Full Text](#)
- Minegishi H, Enomoto S, Echigo A, *et al.*: **Salinarchaeum chitinilyticum sp. Nov., a chitin-degrading haloarchaeon isolated from commercial salt.** *Int. J. Syst. Evol. Microbiol.* 2017; **67**(7): 2274–2278.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Muruga B, Anyango B: **A Survey of Extremophilic Bacteria in Lake Magadi, Kenya.** *Am. J. Mol. Cell. Biol.* 2013; **2**(1): 14–26.
- Mwatha WE, Grant WD: **Natronobacterium vacuolata sp. nov., a Haloalkaliphilic Archaeon Isolated from Lake Magadi, Kenya.** *Int. J. Syst. Bacteriol.* 2016; **43**(3): 401–404.
- Mwirichia R: **Amplicon-Based Analysis of the Fungal Diversity across Four Kenyan Soda Lakes.** *Scientifica.* 2022; **2022**: 1–9.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Newton RJ, Jones SE, Eiler A, *et al.*: **A Guide to the Natural History of Freshwater Lake Bacteria.** *Microbiol. Mol. Biol. Rev.* 2011; **75**(1): 14–49.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Oduor SO, Schagerl M: **Temporal trends of ion contents and nutrients in three Kenyan Rift Valley saline-alkaline lakes and their influence on phytoplankton biomass.** *Hydrobiologia.* 2007; **584**(1): 59–68.
[Publisher Full Text](#)
- Oksanen J: **Vegan: an introduction to ordination.** *Management.* 2015; **1**: 1–10.
[Reference Source](#)
- Oren A: **Thermodynamic limits to microbial life at high salt concentrations.** *Environ. Microbiol.* 2011; **13**(8): 1908–1923.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Orwa P, Mugambi G, Wekesa V, *et al.*: **Isolation of haloalkaliphilic fungi from Lake Magadi in Kenya.** *Heliyon.* 2020; **6**(1): e02823.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Otasek D, Morris JH, Bouças J, *et al.*: **Cytoscape Automation: Empowering workflow-based network analysis.** *Genome Biol.* 2019; **20**(1): 115–185.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Pérez-Pantoja D, Donoso R, Agulló L, *et al.*: **Genomic analysis of the potential for aromatic compounds biodegradation in Burkholderiales.** *Environ. Microbiol.* 2012; **14**(5): 1091–1117.
[Publisher Full Text](#)
- Pérez E, Chebude Y: **Chemical Analysis of Gaet'ale, a Hypersaline Pond in Danakil Depression (Ethiopia): New Record for the Most Saline Water Body on Earth.** *Aquat. Geochem.* 2017; **23**(2): 109–117.
[Publisher Full Text](#)
- Poyraz N, Mutlu M: **Characterization of Microbial Populations of Lake Van By 16S Metagenomics Study.** *Eskişehir Teknik Üniversitesi Bilim ve Teknoloji Dergisi - C Yaşam Bilimleri Ve Biyoteknoloji.* 2020; **9**(1): 80–88.
[Publisher Full Text](#)
- Price MN, Dehal PS, Arkin AP: **FastTree 2 - Approximately maximum-likelihood trees for large alignments.** *PLoS One.* 2010; **5**(3): e9490.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Quast C, Pruesse E, Yilmaz P, *et al.*: **The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools.** *Nucleic Acids Res.* 2012; **41**(D1): D590–D596.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

R Core Team: *R: A language and environment for statistical computing* Vienna, Austria: R Foundation for Statistical Computing; 2020.

[Reference Source](#)

RStudio Team: *RStudio: integrated development for R*. PBC, Boston, MA: RStudio Team; 2020.

[Reference Source](#)

Rampelotto PH: **Extremophiles and extreme environments**. *Life*. 2013; **3**(3): 482–485.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Rognes T, Flouri T, Nichols B, *et al.*: **VSEARCH: A versatile open source tool for metagenomics**. *PeerJ*. 2016; **4**(10): e2584–e2522.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Rosenberg E, De Long EF, Lory S, *et al.*: **The prokaryotes: Other major lineages of bacteria and the archaea**. *The Prokaryotes: Other Major Lineages of Bacteria and The Archaea*. 2014; **43**: 1–1028.

[Publisher Full Text](#)

Salano OA, Makonde M, Kasili RW, *et al.*: **Diversity and distribution of fungal communities within the hot springs of soda lakes in the Kenyan rift valley**. *Academicjournals. Org*. 2017; **11**(19): 764–775.

[Publisher Full Text](#)

Simachew A, Lanzén A, Gessesse A, *et al.*: **Prokaryotic Community Diversity Along an Increasing Salt Gradient in a Soda Ash Concentration Pond**. *Microb. Ecol*. 2016; **71**(2): 326–338.

[PubMed Abstract](#) | [Publisher Full Text](#)

Sorokin DY, Panteleeva AN, Tourova TP, *et al.*: **Natronoflexus pectinivorans gen. nov. sp. nov., an obligately anaerobic and alkaliphilic fermentative member of Bacteroidetes from soda lakes**. *Extremophiles*. 2011; **15**(6): 691–696.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Sorokin DY, Berben T, Melton ED, *et al.*: **Microbial diversity and biogeochemical cycling in soda lakes**. *Extremophiles*. 2014a; **18**(5): 791–809.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Sorokin DY, Gumerov VM, Rakitin AL, *et al.*: **Genome analysis of Chitinivibrio alkaliphilus gen. nov., sp. nov., a novel extremely haloalkaliphilic anaerobic chitinolytic bacterium from the candidate phylum Termite Group 3**. *Environ. Microbiol*. 2014b; **16**(6): 1549–1565.

[PubMed Abstract](#) | [Publisher Full Text](#)

Sorokin DY, Kuenen JG: **Chemolithotrophic haloalkaliphiles from soda lakes**. *FEMS Microbiol. Ecol*. 2005; **52**(3): 287–295.

[PubMed Abstract](#) | [Publisher Full Text](#)

Sorokin DY, Trotsenko YA, Doronina NV, *et al.*: **Methylohalomonas lacus gen. nov., sp. nov. and Methylohalobium kenyaense gen. nov., sp. nov., methylophilic gammaproteobacteria from hypersaline lakes**. *Int. J. Syst. Evol. Microbiol*. 2007; **57**(12): 2762–2769.

[PubMed Abstract](#) | [Publisher Full Text](#)

Taher AG: **Inland saline lakes of Wadi El Natrun depression, Egypt**. *Int. J. Salt Lake Res*. 1999; **8**(2): 149–170.

[Publisher Full Text](#)

Tao X, Guo F, Zhou Q, *et al.*: **Bacterial community mapping of the intestinal tract in acute pancreatitis rats based on 16S rDNA gene sequence analysis**. *RSC Adv*. 2019; **9**(9): 5025–5036.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Tazi L, Breakwell DP, Harker AR, *et al.*: **Life in extreme environments: Microbial diversity in Great Salt Lake, Utah**. *Extremophiles*. 2014; **18**(3): 525–535.

[PubMed Abstract](#) | [Publisher Full Text](#)

Tindall BJ, Ross HNM, Grant WD: **Natronobacterium gen. nov. and Natronococcus gen. nov., Two New Genera of Haloalkaliphilic Archaeobacteria**. *Syst. Appl. Microbiol*. 1984; **5**(1): 41–57.

[Publisher Full Text](#)

Urbietta MS, Donati ER, Chan KG, *et al.*: **Thermophiles in the genomic era: Biodiversity, science, and applications**. *Biotechnol. Adv*. 2015; **33**(6): 633–647.

[Publisher Full Text](#)

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): 118–168.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

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): 1–18.

[Publisher Full Text](#)

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**(1): 1–15.

[Publisher Full Text](#)

Webster G, O'Sullivan LA, Meng Y, *et al.*: **Archaeal community diversity and abundance changes along a natural salinity gradient in estuarine sediments**. *FEMS Microbiol. Ecol*. 2015; **91**(2): 1–18.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Zavarzina DG, Zhilina TN, Kuznetsov BB, *et al.*: **Natronaerobaculum magadiense gen. nov., sp. nov., an anaerobic, alkalithermophilic bacterium from soda lake sediment**. *Int. J. Syst. Evol. Microbiol*. 2013; **63**(PART 12): 4456–4461.

[PubMed Abstract](#) | [Publisher Full Text](#)

Zhao D, Zhang S, Xue Q, *et al.*: **Abundant Taxa and Favorable Pathways in the Microbiome of Soda-Saline Lakes in Inner Mongolia**. *Front. Microbiol*. 2020; **11**(July): 1–17.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Zhilina TN, Garnova ES, Tourova TP, *et al.*: **Amphibacillus fermentum sp. nov. and Amphibacillus tropicus sp. nov., new alkaliphilic, facultatively anaerobic, saccharolytic bacilli from Lake Magadi**. *Microbiology*. 2001; **70**(6): 711–722.

[Publisher Full Text](#)

Zhilina TN, Zavarzin GA, Rainey F, *et al.*: **Spirochaeta alkalica sp. nov., Spirochaeta africana sp. nov., and Spirochaeta asiatica sp. nov., alkaliphilic anaerobes from the Continental Soda Lakes in Central Asia and the East African rift**. *Int. J. Syst. Bacteriol*. 1996; **46**(1): 305–312.

[PubMed Abstract](#) | [Publisher Full Text](#)

Zhu P, Wang Y, Shi T, *et al.*: **Intertidal zonation affects diversity and functional potentials of bacteria in surface sediments: A case study of the Golden Bay mangrove, China**. *Appl. Soil Ecol*. 2018; **130**(5): 159–168.

[Publisher Full Text](#)

Zorz JK, Sharp C, Kleiner M, *et al.*: **A shared core microbiome in soda lakes separated by large distances**. *Nat. Commun*. 2019; **10**(1): 4210–4230.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status:    

Version 2

Reviewer Report 11 November 2024

<https://doi.org/10.5256/f1000research.166004.r330095>

© 2024 Simachew A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Addis Simachew

Institute of Biotechnology Associate Professor of Microbial Ecology and Biotechnology, Addis Ababa University, Addis Ababa, Ethiopia

I would like to thank the authors for conducting this nice work which fills a gap by assessing microbial communities structure in response to seasonality in Lake Magadi, Kenya, one of the soda lakes in the East African Rift Valley System.

General comments

The authors have attempted to assess the structure and composition of microbial communities in response to the seasonality and physicochemical parameters of Lake Magadi. Accordingly, the authors assessed the microbial communities structure and composition using three types of samples (hot spring water, open lake water and brine) of selected sampling sites in lake Magadi; sampled at four consecutive months (June to September of 2018). The key research question as pointed out by the authors was "how microbial diversity changes with the fluctuating physicochemical conditions with seasons, specifically wet/rainy and dry seasons?"

In general, the manuscript needs major revision for acceptance due to the following major issues:

1.The sampling months were successive dry months, as described by the authors, it contradicts with the objective and hypothesis of the study. What does the authors mean by seasonality considering only dry months of the year?. The study was designed as stated by the authors, to investigate the effect of seasonality, i.e., the effect of dry and wet/rainy seasons on the microbial communities structure and composition. However, the study design did not include the rainy season of the area to answer the research question and test the research hypothesis. To this end, in the authors concluded that "our research highlights the composition and diversity of microbial communities across four successive dry months of soda Lake Magadi and the key physicochemical factors influencing them" in contradiction to the research objective and hypothesis.

2.The number of samples taken from each sample type were not proportional. Only a single composite sample was taken from the hot spring (S1) and brine (Br1), while five composite samples (S2-S6) were taken from the open lake water. The authors did not justify why such sampling scheme was employed. Unless justified the representativeness and the reliability of the data is questionable.

3.Three sample sites (S2, S3 and S4) were on the northern tip of the lake while the other two sample sites (S4 and S5) were on the southern tip of the lake (Figure 1). These sample sites were

also around the shore of the lake while the large portion of the lake in the middle was not covered. The key question is that, are the sample sites representative to assess the lake microbial community? What was/were the reason/s to consider such sampling scheme?

4.The data for Br1-June, Br1-August, S1-June, S1-August and S2-August are not available. Then, how comparison is possible without having the full set of data for each sample site?

5.The temperature of the sample sites were in the range of 27°C (S1-June and S4-July) to 38.7°C (S3-June) while the maximum temperature the sample site considered as hot spring was 35.6 °C both in June and September which is less than to some of the other sample sites. What is/are the reason/s for considering the sampling site with low temperature value as hot spring?

6.There is a new bacterial classification system. In this new classification system, some phylum names are changed, Proteobacteria to Pseudomonata; Firmicutes to Bacillota etc. Hence, it would be better if the new classification systems is used.

Specific comments

Taxonomic composition and structure

- Paragraph 1

1.The proportion of bacteria to archaea varied by season and sampling month (Figure 4)...

The sampling month represents the season. Then what do you mean by season and sampling month?

2. In hot spring water (S1), archaea abundance was the lowest while bacterial abundance was the highest

3.What was the proportion of archaeal abundance in these months? It would be clear if the percentage values are presented. What about for July and August?(sites 1, 4, and 5)

4.No such description of sites either in the method nor the result. Please make it consistent for the sake of understanding.

5. From June to September 2018, bacterial abundance decreased while archaeal abundance increased (Figure 4A)

6.It would be more informative if the proportion is described in percentage or in numerical. The data in Figure 4A shows the microbial abundance at domain level along sampling sites. It would be more informative if the data were presented along months for a specific month. It is very difficult to conclude if the variations was due to month or site specific variations. Moreover, the data for some sampling sites is not complete. For instance, No data for S1 for July, August and the same is true for Br1.

- Paragraph 3

7.S1 and S6 had the least archaeal abundance with 0.08 and 0.8%, respectively. In terms of the sampling month, September had the highest archaeal abundance with 53% followed by August (23%), June (12.7%), and July (11.4%)...

8.It would be appropriate if this information is included in paragraph 1.

- Paragraph 4

9.The word “sampled” is used several times. It is better if other better word is used.

10.. (Kipnyargis et al., 2023a))...Why this citation?

- Paragraph 5

11. Halobacteroidaceae, Spirochaetaceae, halanaerobiaceae, and desulfohalobiaceae...
halanaerobiaceae, and desulfohalobiaceae: start with capital letters

- Paragraph 5

12.Generally, members of genera Nocardiodes, Rhodothermus, Haloterrigena,

Methanomasiliicoccus, Halorubrum, Palaeococcus, Nocardioidea, Salinarchaeum, Salinibacter, and Eubacterium spp. had a wide range of adaptability---

13. Conversely, Synechococcus, Thioalkalivibrio, Cyanobacterium, Rhodovulum, Lewinella, Idiomarina, Pseudidiomarina, Chelatococcus, Aliidiomarina, and Alkalimonas spp. were least influenced by the tested physicochemical factors---The above statements are contradictory.

14. The first statement describes that members of the genera mentioned had a wide range of adaptability which implies that they are less influenced by the physicochemical parameters of the environments. The second statement also indicates the same for members of the microbial genera listed out. It needs attention of the authors.

...Mn and CO₃²⁻. pH and NH₄⁺ ...

The dot next to should be corrected

... correlated with Mg²⁺, Mn, and NO₃⁻ ...

Mn should be written in the ionic form as the other species

...On the other hand, Mn, temperature CO₃²⁻, and NH₄⁺ ...

The same comment as above

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

I cannot comment. A qualified statistician is required.

Are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Haloalkaliphilic Microbial Ecology and Biotechnology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 23 October 2024

<https://doi.org/10.5256/f1000research.166004.r330100>

© 2024 Pedroso-Roussado C. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Cristiano Pedroso-Roussado 

Universidade de Lisboa Instituto Superior Tecnico, Lisboa, Portugal

I want to thank the authors for performing this work which represents a relevant microbial profile study for bacteria and archaea in Lake Magadi, Kenya.

General:

The work needs a whole revision for refine its text, and allow a more fluid narrative. The methods' section does not match with the results section. The discussion needs to have a more critical take on the findings. Key words does not reveal a clear picture of the work. Hence, I believe this work needs another round of revision before being fully accepted.

In detail:

INTRO

- Alpha-diversity is an index that speaks about diversity within samples and this seems not to be fully understood by the authors. The same for beta-diversity, that speaks about diversity between samples. A more critical take on these indexes is necessary.
- . It is not clear the relevance of measuring TDS.
- . Annotations are not correct: replace "ml" for "mL" and give a space for "X °C".
- Is there any reference for the taxonomical update of *Pelagirhabdus fermentum*?
- What is a lake lagoon?
- What is the relevance of the isolates that Edwin et al (2019) and Orwa et al (2020) collected with the work you here introduce?

METHODS

- It is not clear how many water samples were collected nor at which depth. More details on the methods' section is necessary.
- Why do you justify the collection of 50 mL per sample instead of >5 L?
- How long did it take for traveling with samples? And at what temperature?

RESULTS

- You should not state the objective of the study in the results' section.
- Table 2 is showing all the results which is not a conventional way to show different variables (taxonomy vs diversity). I would prefer to see the sequencing QC as well.
- . Does the figure 2 show the three different collection points combined (spring, brine, open waters)? It is needed more clarity for the space and time importance on the results.
- What does the proportion of bacteria to archaea tell us?
- It is not mentioned how did you perform the co-occurrence and redundancy analyses.
- Your study relates to the microbiota and not to the microbiome (Berg et al., (2020)(Ref-1))

References

1. Berg G, Rybakova D, Fischer D, Cernava T, et al.: Microbiome definition re-visited: old concepts and new challenges. *Microbiome*. 2020; **8** (1). [Publisher Full Text](#)

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

No

Are sufficient details of methods and analysis provided to allow replication by others?

No

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Applied microbiology, microbiome analysis, next-generation sequencing, third-generation sequencing.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 17 June 2024

<https://doi.org/10.5256/f1000research.166004.r279828>

© 2024 Haines M. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Marianne Haines

University of Calgary, Calgary, Alberta, Canada

Well done getting this all updated. The manuscript is in much better shape. I can see all the hard work you have put into it. Well done. There are a few minor things I would suggest amending and then that's all from me. Good luck with all your future science – I hope it brings you much joy and remains inspiring. See below:

Abstract

Abstract: background

- “However, it is not yet clear if the change in physiochemical changes influences the spatiotemporal diversity and structure of microbial communities in these ecosystems.” VS “The distribution and diversity of microbial communities in hypersaline lakes are mainly affected by physicochemical parameters (Tazi et al., 2014)” in the 2nd paragraph of the Introduction. These two statements are contradictory. Perhaps you are trying to say that the influence of physiochemical parameters *within a single lake* have not been explored.

Abstract: methods

- You have repeated a sentence twice

Abstract: results

- The “composition” of Archaea cannot be higher. I think you mean the total abundance of Archaea was higher than that of Bacteria.
- What is P+ ? Phosphorous? Why does it have a “+”?

Methods:

- All these updates make this much better

Results

- I still think Table 2 needs to define what an “observed species” is. Please refer to review 1 for more on this. Is it the number of unique species that were assigned?
- “Low alkalinity and salinity samples (pH 9.8 – 10.5; 10,300 ppm – 70,500 ppm) formed cluster I with nine samples (S1_06, S1_09, S5_06, S5_09, S5_08, S5_07, S6_08, S6_06, and S6_07). Moderately alkaline and saline samples (pH 10.5 – 10.6; **63**, 900 ppm – 100,000 ppm)” – just need to remove the spacing between 63, and 900 to reader doesn’t get confused and think these are two separate numbers. My interpretation is that this is actually 63,900 ppm?
- Still need to be more specific with this one “pH and NH₄⁺ appear to positively correlate with the structure of the members of the genus *Salinarchaeum* ($R^2 = 0.245$; $p < 0.004$), but negatively correlated with NO₃⁻.” How does something positively correlate with the “structure of the members” – do you mean at increasing pH and ammonium concentrations more members of that genus were present? The subject of this sentence is “pH and NH₄⁺” so as it is this sentence implies that pH and NH₄⁺ were negatively correlated the nitrate. I don’t think that is what you mean, I think you mean that with increases in NO₃⁻ the presence and/or abundance of *Salinarchaeum* was reduced. Is that it?

Discussion

- This has come a long way, I can really see that you have now compared your results to others. Well done.

Conclusion

- “halocaliaphiles” spelling
- Much better conclusion

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: alkaline soda lake communities, alkaliphilic cyanobacteria

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 11 May 2024

<https://doi.org/10.5256/f1000research.147523.r251335>

© 2024 Bhatnagar S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Srijak Bhatnagar 

¹ Athabasca University, Athabasca, Alberta, Canada

² Athabasca University, Athabasca, Alberta, Canada

Major Feedback:

1. Please run us statistical test on the alpha diversity values as it appears the August values could be very similar to June values statistical test is required to ensure the difference mentioned by the authors
2. Sequence analysis subsection under result, line 2: average read length is listed as 525bp The Amplicon expected from the primer mentioned should be around 300 and the length of my seek is also around 300bp. How is read length possible?
3. In various places site, S1 is described as a hot spring, but the temperature is shown as roughly around 35°C. How is this a hot spring?
4. Subsection taxonomic, composition, and structure: "Population increased due to

evaporative iron concentration". This is a causal statement, and no evidence of causation is shown in the study.

5. Figure 4B: the results mention Euryarcheota as the most abundant phylum. This is not evident from the figure. Either the legend is wrong, or the result is misinterpreted
6. Subsection physicochemical drivers: "TDS has a positive influence". This is a subjective statement with no evidence provided. Unsure what positive influence here means.
7. Use of homonyms: Certain terms have a well-defined and established meaning and could cause confusion, please avoid them. For eg.,
 - Figure 6:" exclusive site of isolation". Isolation is a specific microbial technique, which was not employed in this study.
 - DOM -> Dissolved Organic Matter, but in the manuscript used for Dead Organic Materials
8. The authors missed the opportunity to guide the discussion by focusing on the temporal changes in physicochemical factors driving the microbial community changes and then discussing possible reasons. The way the discussion is currently framed, it is a mix of results retelling and describing various taxa that were observed by other publications.

Minor Issues:

- There is no mention of Phix filtering and sequence processing under methods
- Typo in paragraph two of microbial community analysis, subsection Bray Curtis dissimilarity, written as Bray dissimilarity
- Table 1: add the charge to the measured ions
- Results first paragraph line 3: various ions and cations-> various ions or various anions and cations
- Please mention the medium library size. It is more informative than the average size.
- Incomplete sentence. beta diversity studies, subsection, line 2: samples better individual community similarity
- Figure 3: Label the clusters as described in the results section
- Figure 5: plastering of samples and taxa that here would be helpful and visualising the heat map while it's understandable that samples are shown by month class based upon composition, should be able to substantiate some of the observations made by the authors.
- Discussion: "The concentration of these elements" - >"The concentration of these ions"

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Bioinformatics, microbiomes, and environmental microbiology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Reviewer Report 01 April 2024

<https://doi.org/10.5256/f1000research.147523.r251341>

© 2024 Haines M. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Marianne Haines**

¹ University of Calgary, Calgary, Alberta, Canada

² University of Calgary, Calgary, Alberta, Canada

Summary

The authors sampled 7 sites representing spring, brine, and open waters in Lake Magadi, Kenya, over a 4-month period. Each sample underwent physiochemical testing and DNA extraction for amplicon sequencing and subsequent microbial community analysis. The authors thoroughly explore the composition of the microbial communities at each site and timepoint. Community composition is explored through standard taxon abundance measures and multiple diversity measures. Physiochemical drivers of the success (abundance) of different taxa are also evaluated. This work is valuable to microbial ecologists interested in extreme environments, notably, alkaline soda lakes.

Comments

Firstly, thank-you so much for your commitment to science. Well done! It takes a lot of work to get a manuscript to this stage. You have put together a very comprehensive analysis of your community data. Your results are stated very well and are nicely presented in figures. I really enjoyed the in-depth exploration of the community data. Overall, the manuscript is in good shape, although there are some places, particularly in the discussion and conclusion, where I think it requires some work.

Abstract

- Add in that you did 16S *AND* analysed physiochemical properties in the methods
- "Furthermore, salinity and alkalinity affect β -diversity rather than the sampling site or seasonality." Probably seasonality and site also impact diversity but they are masked by the stronger drivers of salinity and alkalinity. You could rephrase this to say salinity and

alkalinity had a *stronger* or the *most notable* impact on beta diversity. Indeed you actually directly contradict this point in the last sentence: "Multivariate analysis revealed significant spatial and temporal effects on β -diversity and salinity and alkalinity were the major drivers of microbial composition in Lake Magadi."

- "The effects of physicochemical parameters on the microbial community structure showed that temperature, pH, P+, K+, NO₃ -, and total dissolved solids (TDS) had a positive correlation with the microbial community structure" I don't know what you mean by positive correlation on microbial community structure. Explain.

Introduction

- What do you mean by "validly described", be more specific what your definition of valid descriptions are.
- This is stated really nicely, "A key ecological question is how microbial diversity changes with the fluctuating physicochemical conditions with seasons."
- "In this study, we explored the spatiotemporal variation in the microbial community AND physiochemical properties of the water over four months at different sites in Lake Magadi using Illumina 16S sequencing." 16S is more specific than the type of sequencing.

Methods

- The naming of samples makes it easy to distinguish brine from the other sample types but it is hard to know that S1 is the only spring.
- Figure 1 – it is hard to see which sites are S3 and S4
- Bit unnecessary to have the type of cool box.
- I haven't done extensive diversity analysis of 16S data so I can't comment on these methods.
- Should include some text that states that the primer is designed to capture both Bacteria and Archaea.
- It would be helpful to state early on in the piece which seasons these months fall in to in Kenya. Are these warm or cool months, rainy or dry? Not all readers will know this immediately.

Results

There was an uneven distribution in sampling times, that needs to be mentioned. Why did this happen? How could it have influenced/biased your results?

Physicochemical properties of the sampling sites

- Table 1 needs units beside each metric in column 1 even though I see they are stated in the body of the text, one should be able to look at the Table without needing to seek clarification from the text.
- Why are there months missing for some of the samples? If data wasn't collected for those months, how did that impact RDA which was used to test the correlation between physiochemical parameters and genera? Why these samples are missing and how the analysis overcame this needs to be explained.

Sequence analysis and diversity studies

- Table 2: What are "observed species", you say that taxa were assigned to genus level, not to species level. Is this column actually talking about the number of different genera? This column should have the 00s removed.
- Table 2 rows should be ordered by sample and then chronologically.

Alpha diversity studies

- This is the first time in the piece you have mentioned the spring is a hot spring. That needs to be updated in the methods.

- Figure 2 has the x axis categories in alphabetical order but I think from the perspective of the reader this should be presented chronologically. Seasonality usually has a profound impact on community structure, it would be nice to visualise easily if there are (or are not) diversity trends based on seasonality.
- In Figure 2 legend, you state “Statistical significance was determined at $p < 0.05$ ” but I don’t see any indication of statistical significance conveyed on the graph so I don’t think this sentence is necessary.
- I like that you have presented the results from so many different measures of species diversity here. It is not necessary to do it for this work, but one way to simplify this type of graph (especially because all tests show similar trends) would be to normalise all values and then plot them on a single graph with box plot colours indicating the test type. At present, your coloured key here is unnecessary because you already have the months as categories in the x-axis.
- “The alpha diversity indices showed that high microbial diversity was recorded in the month of August, followed by September, June, and July in that order (Figure 2).” I don’t think these results clearly show this, they are too variable and overlapping.

Beta diversity studies

- “Low alkalinity and salinity samples (pH 9.8 – 10.5; 10 ***what is this number?***, 300 ppm – 70,500 ppm).” Same thing with the 63 in the medium cluster, and then that number is missing in the high cluster.
- Figure 3 needs to tell me which clusters are which, preferably on top of the graph are in the legend. That’s cool that they cluster like that.

Taxonomic composition and structure

- “The results indicate that the archaeal population increased due to evaporative ion concentration while bacteria abundance was higher where the ion concentration was lower (sites 1, 4, and 5) (Table 1).” I don’t think you can conclude this from your results. You might say there is a trend and it could be because of ion concentration. This is a discussion point and shouldn’t be in the results section.
- Figure 4: I am a bit confused with the alignment of different samples, why are there gaps between bars in some cases but not in others? I think 4A and 4B should have their samples aligned vertically according to month and sample site. If there are gaps where one sample was completed for Bacteria and not for Archaea that’s fine, but the figure needs to simply explain why. Gaps between sample types makes in 4A but things go astray in 4B.
- Is the relative abundance on the y-axis of figure 4A based on read number? How was that calculated?
- Figure 6: What a cool plot. This is a cool way to visualise data. I like it!
- Figure 7: In the legend, “Samples corresponding to their sampling month are indicated as indicated”. This is a bit of a clumsy sentence. Good news is the key is completely self-explanatory so you just delete it. You don’t need to explain TDS again, its common enough in the field and you already have described it in the methods.
- “Notably, members of the phylum Proteobacteria were the most dominant group across all the sampling sites, accounting for 35% abundance.” Is this abundance based on all sample site and timepoints pooled together? That should be clarified.
- “Idiomarina vacuolatum was sampled across all the sampling seasons but varied in structure across the sampling sites.” Do you mean varied in abundance? Be as specific as you can when stating results.

Physicochemical drivers of bacterial and archaeal community structure

- “Overall, temperature, pH, P+, K+, NO₃-, and TDS had a positive influence on the microbial

community structure.” – what does this mean? What do you define as a positive influence?

- “were adapted to fewer physiochemical factors” bold statement. Keep it more objective here? You could say that these species were not found in high concentrations of X, Y, Z.
- “pH appears to positively correlate with the structure of the members of the genus *Salinarcheum*”, what does structure of members mean?

Discussion

Where this manuscript needs attention is in the discussion. In general, discussion **clarity and reasoning** needs improvement. This is particularly evident when citing other literature. There were a few places I wasn't sure what the authors meant and there are some repeated / missing words.

- “Furthermore, sulfate concentration (39–958 ppm) was lower than that of lakes Sidi Ameur and Himalatt (Algeria) (Boutaiba et al., 2012).” This isn't a meaningful reference to use here. I don't understand why this single comparison (of 2 lakes in Algeria) is important to the discussion. Sulfate concentration varies quite a lot throughout soda lakes. Is this sulfate concentration low or high relative to soda lakes in general?
- “The concentrations of these elements (except pH) were varied across the sampling months and sites, suggesting that the lake chemistry is constantly changing in its constituent elements” This isn't a suggestion ... you measured it. Instead you could say “measured elements were variable from site to site and fluctuated with time.
- “Hypersaline lakes are characterized by ...” The remainder of this paragraph is not relevant to the discussion. It is either introduction material or should be deleted. It does not “discuss” your results.
- “Under the extremes of salinity and alkalinity, microorganisms in soda ...” This entire paragraph is also introduction material or should be deleted. It does not discuss the results.
- “Notably, the phylum Proteobacteria was the most dominant group across all the sampling sites, accounting for 35% abundance.” Be specific about what the abundance means, it is 35% of all reads from all timepoints and samples sites were attributed to Proteobacteria?
- “They were followed by the phylum Cyanobacteria (14.2%) represented mainly by the *Eubacter* spp. *Eubacter* is a single- celled stenohaline cyanobacterium that grow optimally at 7% (w/v) NaCl.” This is where you can compare to the salt concentration in your system – is it above or below this?
- It seems that Eulohaleaceae are the main bacterial phototrophs in your system. So the following discussion about *Spirulina* doesn't make too much sense. You could say this is *in contrast* to many other soda lakes systems where *Arthrospira* dominates. Does Lake Magadi have more NaCl than other lakes, is that why Eulohaleaceae and not *Arthrospira* dominate? Have either of these two taxa been found to be common in the lake in other studies?
- “Our findings showed a shift in bacterial composition throughout the sampling seasons where a high abundance of Actinobacteria and Proteobacteria was accompanied by a lower abundance of Actinobacteria.” How is Actinobacteria in high abundance and low abundance?
- “Interestingly, the presence of Bacteroidetes was often associated with the availability of Cyanobacteria across the sampling periods” Keep going here... why could this be the case?
- “However, the abundance of Firmicutes, which possess diverse metabolic capacities and are resistant to oxygen limitations (Martiny et al., 2006) depicted an uneven behavior in relation to Actinobacterial composition.” What do you mean depicted an uneven behaviour?
- “Grant et al. (1999) first characterized this phylum IN the alkaline saltern of Lake Magadi.”
- “Their distribution along a salinity gradient in estuarine sediments may be linked to changes

in location and/or salinity as well as gradient sediment depth (Webster et al., 2015).” I don’t understand what you mean here.

- “Crenarchaeota (Ghori et al., 2021), Crenarchaeota, Euryarchaeota, Woesearchaeota, and Pacearchaeota, Euryarchaeota and Woesearchaeota (Wang et al., 2022).” Crenarchaeota and Woesearchaeota are repeated in this.
- “Remarkably, Thioalkalivibrio sp. was not significantly affected by the physicochemical properties investigated.” Yes! Good observation. But why? Tell me why it is remarkable or unexpected. E.g., “We had expected that Thioalkalivibrio would have responded positively to XYZ...”
- “However, open waters samples from S2–S4 depicted varying degrees of community structure and this could be due to variations in the intertidal water zones (Zhu et al., 2018).” Lacking specificity, what do you mean by community structure here, diversity / abundance / species type? What do you mean by variations in intertidal water zones? What type of variations?
- “This phenomenon can be attributed to the maintenance of homogenous abundances by microbial communities during brine formation, hence resulting in higher biodiversity (Banda et al., 2020).” Homogenous abundances of each microbe as evaporation occurs? Why would this happen?
- “Hot spring samples collected in September showed high species diversity indices.” Microbial samples from Lake Magadi hot springs were established to be stable and active (Kambura et al., 2016).” Relevance? Are high species diversity indices unusual in hot springs? Is it unusual to have a stable, active community? What does stable mean here? Similarly structured over time? Is that what is seen at S1? There are only two samples from S1 so that conclusion cannot be drawn anyway.

Conclusion

The conclusion needs some work. The clarity needs to be improved by being much more specific.

- “The results depict a great deal of diversity in bacterial diversity as compared to archaea.” Consider rephrasing this.
- “Salinity and alkalinity are the main drivers of the microbial community in the lake. Overall, members of Nocardiodetes, Rhodothermus, Haloterrigena, Methanomasiliicoccus, Halorubrum, Palaeococcus, Nocardiodetes, Salinarcheum, Salinibacter, and Eubacterium had a wide range of adaptability.” Does this mean these members have a wide range of adaptability to salinity and alkalinity or are present in many sites?
- “Conversely, Synechococcus, Thioalkalivibrio, Cyanobacterium spp., Rhodovulum, Lewinella, Idiomarina, Pseudidiomarina, Chelatococcus, Aliidiomarina, and Alkalimonas were affected by fewer physicochemical factors.” Does this also mean they had a wide range of adaptability? Or does this mean they weren’t positively correlated with particular nutrients?
- “...have a positive correlation with microbial community structure.” What is a positive correlation with microbial community structure? How do they influence the community structure?

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

I cannot comment. A qualified statistician is required.

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: alkaline soda lake communities, alkaliphilic cyanobacteria

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 30 Apr 2024

Alex Kipnyargis

- Add in that you did 16S *AND* analysed physiochemical properties in the methods
- **Response: Methods (paraphrased):** Using 16S rRNA gene amplicon sequencing, we investigated the diversity and structure of microbial communities in water and brine samples taken from Lake Magadi between June and September 2018. Physicochemical parameters were also analyzed for every sampling site.
- "Furthermore, salinity and alkalinity affect β -diversity rather than the sampling site or seasonality." Probably seasonality and site also impact diversity but they are masked by the stronger drivers of salinity and alkalinity. You could rephrase this to say salinity and alkalinity had a *stronger* or the *most notable* impact on beta diversity. Indeed you actually directly contradict this point in the last sentence: "Multivariate analysis revealed significant spatial and temporal effects on β -diversity and salinity and alkalinity were the major drivers of microbial composition in Lake Magadi."
- **Response: The sentence** "Furthermore, salinity and alkalinity affect β -diversity rather than the sampling site or seasonality." **is deleted completely.**
The statement should now read: "Multivariate analysis revealed that salinity and alkalinity were the major drivers of changes in the structure of microbial communities in Lake Magadi."
- "The effects of physicochemical parameters on the microbial community structure showed that temperature, pH, P+, K+, NO₃ -, and total dissolved solids (TDS) had a

positive correlation with the microbial community structure" I don't know what you mean by positive correlation on microbial community structure. Explain.

- **Response: The statement now reads** "The findings demonstrated that temperature, pH, P+, K+, NO₃⁻, and total dissolved solids (TDS) contributed majorly to diversity observed in the microbial community."

Introduction

- What do you mean by "validly described", be more specific what your definition of valid descriptions are.
- **Response:** "validly described" is replaced with "previously described"
- This is stated really nicely, "A key ecological question is how microbial diversity changes with the fluctuating physicochemical conditions with seasons."
- **Response:** Thank you.
- "In this study, we explored the spatiotemporal variation in the microbial community AND physiochemical properties of the water over four months at different sites in Lake Magadi using Illumina 16S sequencing." 16S is more specific than the type of sequencing.
- **Response:** "Illumina" **will be deleted. It will now read:** "In this study, we explored the spatiotemporal variation in the microbial community and physiochemical properties of the water over four months at different sites in Lake Magadi using 16S rRNA gene sequencing."

Methods

- The naming of samples makes it easy to distinguish brine from the other sample types but it is hard to know that S1 is the only spring.

Response: The sentence after the GPRS coordinates describes the features of the sampling sites

- Figure 1 – it is hard to see which sites are S3 and S4

Response: This has been corrected and a new map will be submitted

- Bit unnecessary to have the type of cool box.

Response: *This has been removed*

- I haven't done extensive diversity analysis of 16S data so I can't comment on these methods.

Response: No response since it is a comment.

- Should include some text that states that the primer is designed to capture both Bacteria and Archaea.

Response: That sentence will now read: “The V4 hypervariable region of the 16S rRNA genes was amplified using the universal primers for bacteria and archaea 515F (5'-GTGCCAGCMGCCGCGGTAA-3') and 806R (5'-GGACTACHVGGGTWTCTAAT-3')”

- It would be helpful to state early on in the piece which seasons these months fall in to in Kenya. Are these warm or cool months, rainy or dry? Not all readers will know this immediately.

Response: The dry season lasted for all of the sampling months, with June marking the start of the season and September being the driest month. This has been effected in the manuscript.

Results

There was an uneven distribution in sampling times, that needs to be mentioned. Why did this happen? How could it have influenced/biased your results?

Response: You are right. Uneven distribution in sampling sites was a result of:

1. S1 and Br_1 (hot spring and brine, respectively were only available in June and September).
2. The DNA extraction protocol did not yield sufficient DNA for sequencing in sample S2 for August 2018.

Physicochemical properties of the sampling sites

- Table 1 needs units beside each metric in column 1 even though I see they are stated in the body of the text, one should be able to look at the Table without needing to seek clarification from the text.

Response: This has been done as suggested.

- Why are there months missing for some of the samples? If data wasn't collected for those months, how did that impact RDA which was used to test the correlation between physiochemical parameters and genera? Why these samples are missing and how the analysis overcame this needs to be explained.

Response:

1. S1 and Br_1 (hot spring and brine, respectively were only available in June and September).
 2. The DNA extraction protocol did not yield sufficient DNA for sequencing in sample S2 for August 2018.
- However, the Good's coverage (Table 2) ranged between 0.81–0.96, indicating that the

sequencing depth was sufficient to cover most microbial species information and the sample size was sufficient to reflect the diversity differences among different communities.

Sequence analysis and diversity studies

- **Table 2:** What are “observed species”, you say that taxa were assigned to genus level, not to species level. Is this column actually talking about the number of different genera? This column should have the 00s removed.

Response: The assignment to genus level is a more confident way of taxonomic assignment while using 16S rRNA gene. Qiime would pick out the species identified in the OTUs, but we would not be confident to pick out species to difficulties surrounding 16S in resolving closely related species. We would probably remove that statement and recommend that future studies should consider the use of Amplified Sequence Variants (ASVs) to confidently assign species. The 00s in the “observed species” column have been removed.

- Table 2 rows should be ordered by sample and then chronologically.

Response: This has been ordered accordingly.

Alpha diversity studies

- This is the first time in the piece you have mentioned the spring is a hot spring. That needs to be updated in the methods.

Response: The word hot spring will be used consistently.

- Figure 2 has the x axis categories in alphabetical order but I think from the perspective of the reader this should be presented chronologically. Seasonality usually has a profound impact on community structure, it would be nice to visualise easily if there are (or are not) diversity trends based on seasonality.

Response: Figure 2 has been presented in chronological order.

- In Figure 2 legend, you state “Statistical significance was determined at $p < 0.05$ ” but I don’t see any indication of statistical significance conveyed on the graph so I don’t think this sentence is necessary.

Response: This has been removed

- I like that you have presented the results from so many different measures of species diversity here. It is not necessary to do it for this work, but one way to simplify this type of graph (especially because all tests show similar trends) would be to normalise all values and then plot them on a single graph with box plot colours indicating the test type. At present, your coloured key here is unnecessary because you already have the months as categories in the x-axis.

Response: I have removed the key

- “The alpha diversity indices showed that high microbial diversity was recorded in the month of August, followed by September, June, and July in that order (Figure 2).” I don’t think these results clearly show this, they are too variable and overlapping.

Response: You are right, I will rephrase this sentence.

Beta diversity studies

- “Low alkalinity and salinity samples (pH 9.8 – 10.5; 10 ***what is this number?***, 300 ppm – 70,500 ppm).” Same thing with the 63 in the medium cluster, and then that number is missing in the high cluster.

Response: The spaces between the numbers have been removed and the numbers read appropriately. Numbers for high cluster have been inserted.

- Figure 3 needs to tell me which clusters are which, preferably on top of the graph are in the legend. That’s cool that they cluster like that.

Response: A brief explanation of the clusters has been inserted in the figure caption.

Taxonomic composition and structure

- “The results indicate that the archaeal population increased due to evaporative ion concentration while bacteria abundance was higher where the ion concentration was lower (sites 1, 4, and 5) (Table 1).” I don’t think you can conclude this from your results. You might say there is a trend and it could be because of ion concentration. This is a discussion point and shouldn’t be in the results section.

Response: Good suggestion. I have paraphrased the sentence

- Figure 4: I am a bit confused with the alignment of different samples, why are there gaps between bars in some cases but not in others? I think 4A and 4B should have their samples aligned vertically according to the month and sample site. If there are gaps where one sample was completed for Bacteria and not for Archaea that’s fine, but the figure needs to simply explain why. Gaps between sample types makes in 4A but things go astray in 4B.

Response: The bars have been ordered according to sample and month.

- Is the relative abundance on the y-axis of figure 4A based on read number? How was that calculated?
- Figure 6: What a cool plot. This is a cool way to visualise data. I like it!

Response: Thank you for the comment.

- Figure 7: In the legend, “Samples corresponding to their sampling month are indicated as indicated”. This is a bit of a clumsy sentence. Good news is the key is

completely self-explanatory so you just delete it. You don't need to explain TDS again, its common enough in the field and you already have described it in the methods.

Response: This has been corrected

- "Notably, members of the phylum Proteobacteria were the most dominant group across all the sampling sites, accounting for 35% abundance." Is this abundance based on all sample site and time points pooled together? That should be clarified.

Response: This is a cumulative abundance based on all the time points and sites. The clarification has been effected in the manuscript.

- "Idiomarina vacuolatum was sampled across all the sampling seasons but varied in structure across the sampling sites." Do you mean varied in abundance? Be as specific as you can when stating results.

Response: Yes, it is the abundance. I have changed this in the manuscript.

Physicochemical drivers of bacterial and archaeal community structure

- "Overall, temperature, pH, P⁺, K⁺, NO₃⁻, and TDS had a positive influence on the microbial community structure." – what does this mean? What do you define as a positive influence?

Response: You are right. The sentence has been changed to read "The results demonstrated that temperature, CaCO₃, pH, P⁺, K⁺, NO₃⁻, and TDS influenced the microbial community structure."

- "were adapted to fewer physiochemical factors" bold statement. Keep it more objective here? You could say that these species were not found in high concentrations of X, Y, Z.

Response: This has been corrected to capture the aspect of objectivity.

- "pH appears to positively correlate with the structure of the members of the genus Salinarchaeum", what does the structure of members mean?

Response: Structure means how members of a particular microbial community distribute across seasons and sites.

Discussion

Where this manuscript needs attention is in the discussion. In general, discussion **clarity and reasoning** needs improvement. This is particularly evident when citing other literature. There were a few places I wasn't sure what the authors meant and there are some repeated / missing words.

- "Furthermore, sulfate concentration (39–958 ppm) was lower than that of lakes Sidi Ameur and Himalatt (Algeria) (Boutaiba et al., 2012)." This isn't a meaningful reference to use here. I don't understand why this single comparison (of 2 lakes in Algeria) is important to the discussion. Sulfate concentration varies quite a lot

throughout soda lakes. Is this sulfate concentration low or high relative to soda lakes in general?

Response: More comparisons have been introduced in the manuscript to capture similar, higher, and lower concentrations of sulfates in soda lakes.

- “The concentrations of these elements (except pH) were varied across the sampling months and sites, suggesting that the lake chemistry is constantly changing in its constituent elements” This isn’t a suggestion ... you measured it. Instead you could say “measured elements were variable from site to site and fluctuated with time.

Response: Thank you. I have borrowed your words verbatim and the sentence reads well now.

- “Hypersaline lakes are characterized by ...” The remainder of this paragraph is not relevant to the discussion. It is either introduction material or should be deleted. It does not “discuss” your results.

Response: That part has been paraphrased to explain the reason for high and stable pH. I felt that since pH was relatively stable across seasons and site, it is important to explain how this happens.

- “Under the extremes of salinity and alkalinity, microorganisms in soda ...” This entire paragraph is also introduction material or should be deleted. It does not discuss the results.

Response: I agree. This paragraph has been entirely deleted.

- “Notably, the phylum Proteobacteria was the most dominant group across all the sampling sites, accounting for 35% abundance.” Be specific about what the abundance means, it is 35% of all reads from all timepoints and samples sites were attributed to Proteobacteria?

Response: Proteobacteria accounted for 35% of all the reads.

- “They were followed by the phylum Cyanobacteria (14.2%) represented mainly by the *Eubacter* spp. *Eubacter* is a single-celled stenohaline cyanobacterium that grow optimally at 7% (w/v) NaCl.” This is where you can compare to the salt concentration in your system – is it above or below this?

Response: I have supported this by a sentence reading: “The existence of *Eubacter* in Lake Magadi is thus supported by high salts and carbonates (Table 1).”

- It seems that Eubacteriaceae are the main bacterial phototrophs in your system. So the following discussion about *Spirulina* doesn’t make too much sense. You could say this is *in contrast* to many other soda lakes systems where *Arthrospira* dominates. Does Lake Magadi have more NaCl than other lakes, is that why Eubacteriaceae and not *Arthrospira* dominate? Have either of these two taxa been found to be common in the lake in other studies?

Response: The sentence has been paraphrased to capture the contrasting aspect of Lake Magadi with other soda lakes.

- “Our findings showed a shift in bacterial composition throughout the sampling seasons where a high abundance of Actinobacteria and Proteobacteria was accompanied by a lower abundance of Actinobacteria.” How is Actinobacteria in high abundance and low abundance?

Response: The entire paragraph does not discuss the results adequately and has been deleted

- “Interestingly, the presence of Bacteroidetes was often associated with the availability of Cyanobacteria across the sampling periods” Keep going here... why could this be the case?

Response: The sentence has been enhanced to explain how photosynthetic cyanobacteria contribute towards the generation of substrates for Bacteroidetes and Verrumicrobia.

- “However, the abundance of Firmicutes, which possess diverse metabolic capacities and are resistant to oxygen limitations (Martiny et al., 2006) depicted an uneven behavior in relation to Actinobacterial composition.” What do you mean depicted an uneven behaviour?

Response: The entire paragraph does not discuss the results adequately and has been deleted.

- “Grant et al. (1999) first characterized this phylum IN the alkaline saltern of Lake Magadi.”

Response: The word “in” has been inserted appropriately. Thank you!

- “Their distribution along a salinity gradient in estuarine sediments may be linked to changes in location and/or salinity as well as gradient sediment depth (Webster et al., 2015).” I don’t understand what you mean here.

Response: This sentence has been paraphrased

- “Crenarchaeota (Ghori et al., 2021), Crenarchaeota, Euryarchaeota, Woesearchaeota, and Pacearchaeota, Euryarchaeota and Woesearchaeota (Wang et al., 2022).” Crenarchaeota and Woesearchaeota are repeated in this.

Response: True. This has been deleted

- “Remarkably, Thioalkalivibrio sp. was not significantly affected by the physicochemical properties investigated.” Yes! Good observation. But why? Tell me why it is remarkable or unexpected. E.g., “We had expected that Thioalkalivibrio would have responded positively to XYZ...”

Response: This has been changed to explain the reason for the effect.

- “However, open waters samples from S2–S4 depicted varying degrees of community structure and this could be due to variations in the intertidal water zones (Zhu et al., 2018).” Lacking specificity, what do you mean by community structure here, diversity / abundance / species type? What do you mean by variations in intertidal water zones? What type of variations?

Response: Structure has been replaced with diversity. The sentence has been paraphrased

to capture the aspect of water waves and their effect on the distribution of nutrients. This could explain the variation in the diversity of microbes in water ecosystems.

- “This phenomenon can be attributed to the maintenance of homogenous abundances by microbial communities during brine formation, hence resulting in higher biodiversity (Banda et al., 2020).” Homogenous abundances of each microbe as evaporation occurs? Why would this happen?

Response: The sentence does not justify the diversity enough. I have introduced another theorem. “Despite high salinity and alkalinity in soda lake brine, the presence of high light intensity and dissolved CO₂ promotes the growth of photosynthetic microorganisms. Subsequently, these phototrophs generate large quantities of dissolved organic matter (DOM) which become substrates for sustaining the diverse microbial communities (Banda 2020).”

- “Hot spring samples collected in September showed high species diversity indices.” Microbial samples from Lake Magadi hot springs were established to be stable and active (Kambura et al., 2016).” Relevance? Are high species diversity indices unusual in hot springs? Is it unusual to have a stable, active community? What does stable mean here? Similarly structured over time? Is that what is seen at S1? There are only two samples from S1 so that conclusion cannot be drawn anyway.

Response: I have rewritten the statement to capture the relationship of our findings with other studies on hot springs and the reason for such high diversity studies. Research on Soda Lake hot springs revealed a highly active and diverse microbial community, suggesting the high plasticity of these organisms toward extreme environments.

Conclusion

The conclusion needs some work. The clarity needs to be improved by being much more specific.

- “The results depict a great deal of diversity in bacterial diversity as compared to archaea.” Consider rephrasing this.
- “Salinity and alkalinity are the main drivers of the microbial community in the lake. Overall, members of Nocardiodetes, Rhodothermus, Haloterrigena, Methanomasiliicoccus, Halorubrum, Palaeococcus, Nocardiodetes, Salinarchaeum, Salinibacter, and Eubacterium had a wide range of adaptability.” Does this mean these members have a wide range of adaptability to salinity and alkalinity or are present in many sites?
- “Conversely, Synechococcus, Thioalkalivibrio, Cyanobacterium spp., Rhodovulum, Lewinella, Idiomarina, Pseudidiomarina, Chelatococcus, Aliidiomarina, and Alkalimonas were affected by fewer physicochemical factors.” Does this also mean they had a wide range of adaptability? Or does this mean they weren’t positively correlated with particular nutrients?

- "...have a positive correlation with microbial community structure." What is a positive correlation with microbial community structure? How do they influence the community structure?

Response to conclusion section: I have rewritten the entire conclusion in a way that captures the conclusive elements of both diversity across seasons and sampling sites and the effects of the environmental factors on the population dynamics of microorganisms in Lake Magadi. Your comments were truly helpful.

Competing Interests: No competing interests

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research