scientific reports

Check for updates

OPEN

A long-read sequencing approach to high-resolution profiling of bacterioplankton diversity in a shallow freshwater lake

Stephanie O. Castro Márquez[1,2], Viktor R. Tóth[3], Sofiia Kolchanova[2], Walter W. Wolfsberger[1] & Taras K. Oleksyk[1]✉

Lake Balaton, a large shallow freshwater lake in Hungary, exhibits diverse bacterioplankton communities influenced by various environmental factors. This study aims to evaluate the bacterial diversity in Lake Balaton using the long-read approach to 16 S rRNA gene sequencing. Water samples were collected from a wide network of 33 locations across the lake's four basins and analyzed for bacterial community composition. Sequencing results revealed a high taxonomic diversity with significant zonal variations. Dominant families included Comamonadaceae, Burkholderiaceae, and Methylophilaceae. Environmental parameters such as temperature, pH, and CDOM were found to significantly correlate with bacterial abundance and diversity. The study underscores the utility and portability of using the long-read sequencing technology in assessing microbial diversity and provides insights into the ecological dynamics of bacterioplankton in freshwater lakes.

Keywords Lake Balaton, Littoral, Pelagic, Microbial communities, Oxford nanopore, Long read sequencing, Next generation sequencing

One of the least understood and the most underestimated components of aquatic ecosystems is bacterioplankton. In freshwater lakes, bacterioplankton communities are highly variable in both space and time and are influenced by a wide range of factors[1–4]. Studying bacterioplankton has always relied heavily on technological advancements due to their microscopic nature. Over the years, methods have evolved from microscopy to culture-based techniques, each with its own limitations, such as the inability to cultivate a significant portion of microbial communities[5]. PCR allowed for the amplification of specific DNA sequences but introduced biases and contamination risks. The recent advent of next-generation sequencing (NGS) has revolutionized bacterioplankton research, enabling comprehensive analysis of microbial communities directly from environmental samples without the need for cultivation. This technology provides a more accurate depiction of microbial diversity but comes with challenges, including data processing and interpretation, biases in sequencing, and the need for substantial computational resources[6]. Despite these challenges, NGS continues to enhance our understanding of bacterioplankton and their ecological roles, which is crucial for the conservation and management of freshwater ecosystems.

Europe is particularly dependent on its shallow lakes for essential resources such as water and tourism due to their limited number across the continent. These shallow lakes are critical to the sustainability of local ecosystems and economies, and provide essential services, including water supply for drinking, agriculture, and industry, especially in regions where other water sources are scarce. In addition, lakes are vital for tourism, offering unique recreational opportunities that attract millions of visitors each year. The economic impact of tourism associated with shallow lakes can be substantial, contributing to the livelihood of local communities and supporting conservation efforts. Since Hungary adopted the EU Water Framework Directive in 2006, biological variables have gained increased importance in water quality assessments[7].

Lake Balaton, with a large surface area of 596 km^2 and a length of 78 km, exhibits a remarkable diversity of environments ranging from micro- to mesoscale variations[8]. This is the largest freshwater lake in Central Europe and is a renowned tourist destination, characterized by its shallow depth, averaging 3.0–3.6 m (Fig. 1A, B). The lake's water is continuously mixed by wind action, which leads to the resuspension of sediments[9,10]. The lake's water quality has been monitored regularly both mostly for algological studies[11–13].

[1]Department of Biological Sciences, Oakland University, Rochester, MI, USA. [2]Department of Biology, University of Puerto Rico – Mayagüez, Mayagüez, Puerto Rico. [3]HUN-REN Balaton Limnological Research Institute, Tihany, Hungary. ✉email: oleksyk@oakland.edu

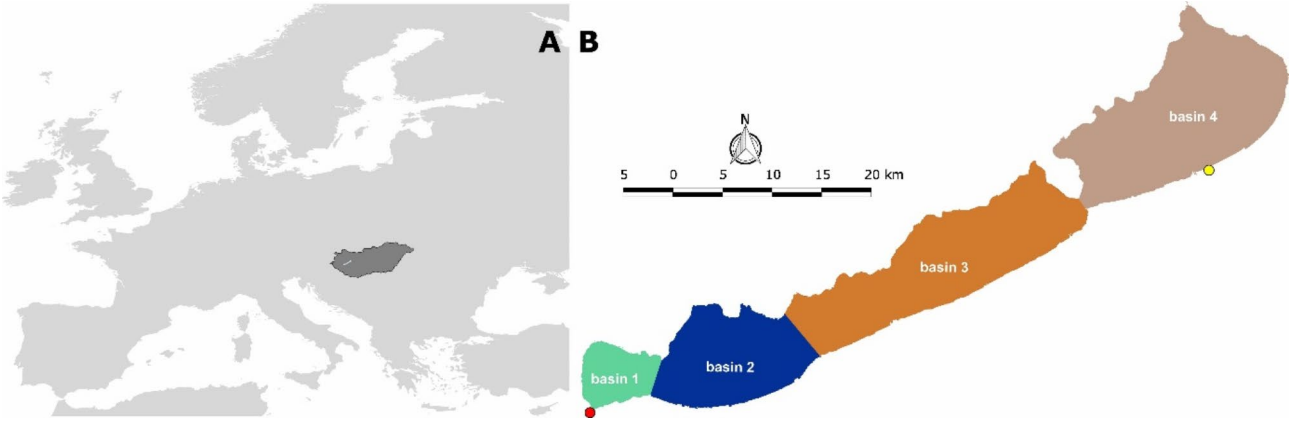

Fig. 1. (**A**) Map of Europe showing Hungary (dark grey) and Lake Balaton (blue polygon). (**B**) Map of Lake Balaton and its division into basins. Red dot in the westernmost basin (1) shows the inflow (River Zala), yellow dot in the easternmost basin (4) shows the outflow (Sió canal).

The combination of ecological factors creates a highly diverse habitat mosaic within Lake Balaton, which provides an ideal environment for bacteria to thrive. Bacteria, including bacterioplankton, play an important role in nutrient cycling, organic matter decomposition, and overall ecosystem functioning[14,15]. The abundance and diversity of bacterioplankton in Lake Balaton are influenced by the complex interactions between these environmental factors, shaping the microbial communities and their ecological functions[16,17]. Several features determine variability, of which the west-east trophic gradient is the most prominent. The eutrophic conditions in the western part of Lake Balaton are attributed to the largest inflow, the River Zala that historically was bringing nutrients into the lake[13,18](Fig. 1B). This inflow also brings water rich in colored dissolved organic matter (CDOM) to the lake, contributing to the unique environmental conditions[17,19]. The prevailing north-westerly winds create specific physical and chemical environments that contribute to the overall diversity of the lake's environment[18,20,21].

Picoeukaryotes dominate the phytoplankton in Lake Balaton during the winter, even under ice-covered conditions, growing at temperatures as low as 1–2 °C[22]. Along the shores of Lake Balaton, there are large stands of macrophytes, both emergent and submerged[21,23]. These macrophytes serve as important habitats, providing shelter, nutrients, and structural complexity. The macrophyte communities further enhance the ecological heterogeneity of the lake. Conversely, it has been reported in the literature that throughout the spring-autumn period, the phycocyanin and phycoerythrin-rich Cyanobacteria are the predominant phytoplankton with eukaryotes becoming negligible[24,25].

Monitoring lake water quality at the global scale is increasingly important for understanding the ecological consequences of climate change, eutrophication, and anthropogenic pressures[26]. Understanding the interplay between different environmental factors and the dynamics of bacterioplankton in a lake is crucial in our understanding of microbial ecology and contributes to a broader understanding of the ecological processes and resilience of shallow lakes[27,28] such as Lake Balaton. Most of these studies focused on specific taxa and locations or regions of Lake Balaton, which may not fully represent the microbial diversity across the entire lake. In contrast, the influence of environmental factors on microbial communities were mostly understudied. With this limited study design, the variations in environmental conditions and microbial communities in other parts of the lake would be overlooked.

Previous microbiological studies determined regional specificities and temporal dynamics of bacteria in Lake Balaton[29,30]. The most recent studies used a combination of short-read NGS, targeted PCR assays, and microscopy cell counting to explore the prokaryotic community structure in the littoral and pelagic regions of Lake Balaton[30–33] while other studies assessed the association of bacterial communities with fish and zooplankton communities[34,35]. However, there has never been a survey of the microbial profile across the entire lake basin.

The diverse environmental conditions across the basins of Lake Balaton, present an optimal setting for investigating the feasibility of a long-read sequencing approach to monitor bacterioplankton communities in a large freshwater lake. In this study, we address the limitations of the previous studies by including water from 33 locations distributed across littoral and pelagic zones in all four basins, to capture the variability in microbial populations influenced by different trophic gradients, inflows, and physical-chemical conditions.

Recent developments in long-read sequencing technology have significantly improved our ability to characterize microbial communities in complex environments. While short-read platforms (e.g., Illumina-based 16 S rRNA sequencing) have been widely used, their limited read length (~ 250–300 bp) often constrains taxonomic resolution, frequently resulting in ambiguous classifications or an inability to discern closely related taxa[36,37]. By contrast, long-read sequencing platforms, such as Oxford Nanopore and PacBio, enable nearly full-length 16 S rRNA gene sequencing (~ 1,500 bp), facilitating more robust classification at the species level[36,38]. This extended read length also helps mitigate PCR biases and allows for better detection of rare or novel taxa, providing a more comprehensive view of microbial diversity[39,40]. Moreover, the portability, real-time data

streaming, and cost-effectiveness of certain long-read platforms (e.g., Oxford Nanopore MinION) have distinct advantages for field-based studies, enabling on-site sequencing and reducing sample handling biases[41–43].

Long-read sequencing technologies, such as those developed by Oxford Nanopore Technologies, offer significant advantages over traditional short-read sequencing methods, particularly in monitoring microbial communities in complex environments like freshwater lakes[6]. In general, using long-read sequencing technologies to monitor microbial communities in freshwater lakes represents a significant advancement in microbial ecology. By providing a deeper and more accurate understanding of microbial diversity, functional potential, and ecological interactions, long-read sequencing is cost-effective method can enhance our ability to protect and manage freshwater ecosystems[44]. Therefore, we selected a long-read sequencing approach to achieve a higher-resolution analysis of bacterioplankton communities and to capture microbial diversity that might otherwise be underrepresented with short-read methods.

This study covered a range of spatial scales of bacterioplankton communities in a single large freshwater lake: the mesoscale (basin level) and local variability (littoral and pelagic zones, northern and southern shores). Utilizing Oxford Nanopore long-read sequencing technology, we intended to conduct a comprehensive analysis of bacterioplankton communities, enabling accurate identification and classification of microorganisms, including those closely related species that are often indistinguishable from short-read sequences. The following hypotheses were tested: (i) the bacterioplankton community in Lake Balaton exhibits high taxonomic diversity; (ii) there are distinct zonal variations in the composition and abundance of bacterioplankton in different regions of Lake Balaton; (iii) major environmental factors significantly influence the structure and diversity of bacterioplankton communities in Lake Balaton. Finally, by setting up a mobile lab on-site, we wanted to demonstrate the utility of the portability of long-read sequencers in a field setting and its ability to generate and analyze data in real-time, providing rapid and detailed insights into the lake's microbial diversity.

Results

Water transparency of Lake Balaton is low (average Secchi depth at the time of sampling was ~ 33 cm, Table 1); thus, high light attenuation results in low euphotic depths. The lake has a great number of inflows, with the main tributary, the Zala River, flowing into the westernmost, basin 1, while the outflow of the lake is situated on the southern shore of basin 4 (Fig. 1B).

The long-read sequencing analysis revealed a substantial number of bacterioplankton taxa, with an average of 147 families in samples across the lake, which was supported by the Chao and ACE indices (Tables 2, 3). The lake had moderate variability ($H = 2.47 \pm 0.03$) and low dominance ($D = 0.16 \pm 0.01$, Tables 2, 3). Figure 2 shows the distribution of bacterioplankton families in Lake Balaton, showing the average percentages of different families. Three families (Comamonadaceae, Burkholderiaceae, and Methylophilaceae) constituted the majority (62.48%) of the families (Fig. 2). In addition, there were 136 families representing 17.77% of the total bacterioplankton community.

Bacterial taxa were detected at the family level and could be interpreted at in the context of ecological and health related conditions. The bacterioplankton community was dominated by three families, which together constituted 62.48% of the total families. In addition, Ilumatobacteraceae, a family of Actinomycetota (or Actinobacteria), and Cryomorphaceae, a family of bacteria in the order Flavobacteriaceae, phylum Bacteroidota are also abundant. Overall, 136 families represented 17.77% of the total community, indicating a rich microbial diversity (Fig. 2). While some bacterial species are associated with health risk, the presence of Bacillota and Spirochaetota in basin 1 (Fig. 2) can also indicate nutrient-rich conditions as they are often found in environments with high organic matter, as they play crucial roles in the decomposition of organic material and nutrient cycling within aquatic ecosystems.

The long-read sequencing results from this study also indicate that, in addition to the clear differentiation in the bacterioplankton abundance on the basin basis, Lake Balaton shows some differences in diversity indices as well, as the difference between the basin with the lowest (basin 4) and highest (basin 3) Shannon indices were merely 13% apart (Tables 2, 3; Fig. 3). Basins 1, 2, and 4 showed relatively similar levels of family abundance and species diversity (Fig. 3; Tables 2, 3).

	Balaton	Basin 1	Basin 2	Basin 3	Basin 4
Surface (km^2)	596	38	144	186	228
Average depth (m)	3.7	2.7	3.3	3.7	4.1
Volume (km^3)	1.95	0.09	0.42	0.62	0.82
Temperature (°C)	14.9 ± 0.9	14.6 ± 0.4	13.6 ± 0.3	15.7 ± 0.2	14.5 ± 0.8
pH	8.6 ± 0.1	8.6 ± 0.1	8.6 ± 0.0	8.6 ± 0.0	8.7 ± 0.0
Conductance (mV)	729 ± 9	736 ± 7	731 ± 4	730 ± 7	723 ± 11
Chlorophyl-a ($\mu g\ l^{-1}$)	5.1 ± 2.3	6.0 ± 1.5	7.6 ± 2.8	5.3 ± 2.1	2.8 ± 0.7
CDOM ($mg\ l^{-1}$)	9.2 ± 5.5	17.8 ± 7.5	9.9 ± 1.4	7.7 ± 1.7	5.0 ± 1.2
Secchi (cm)	32.7 ± 10.8	33.7 ± 9.5	17.8 ± 6	30.2 ± 7.8	43.9 ± 6.2
TSM ($mg\ l^{-1}$)	38.8 ± 37.7	26.4 ± 8.5	96.2 ± 91.7	37.0 ± 12.8	22.8 ± 4.3

Table 1. Major Geomorphological parameters and physical Limnological characteristics (at the time of sampling, average ± sd, $n = 16–64$) of lake Balaton and its basins.

		Basins				Shores		Zones	
	Balaton	1	2	3	4	Southern	Northern	Littoral	Pelagic
n	33	6	4	14	9	16	17	17	16
Taxa (S)	147 ± 8	111 ± 13	114 ± 29	188 ± 5	151 ± 12	153 ± 11	142 ± 11	152 ± 11	142 ± 11
Dominance (D)	0.16 ± 0.01	0.17 ± 0.02	0.15 ± 0.01	0.15 ± 0.01	0.18 ± 0.01	0.16 ± 0.01	0.17 ± 0.01	0.17 ± 0.01	0.15 ± 0.01
Shannon (H)	2.47 ± 0.03	2.44 ± 0.06	2.52 ± 0.10	2.57 ± 0.05	2.37 ± 0.05	2.49 ± 0.04	2.46 ± 0.05	2.43 ± 0.05	2.52 ± 0.04
Brillouin	2.45 ± 0.03	2.4 ± 0.06	2.45 ± 0.13	2.57 ± 0.05	2.36 ± 0.05	2.47 ± 0.05	2.42 ± 0.05	2.4 ± 0.05	2.49 ± 0.04
Menhinick	0.88 ± 0.06	1.07 ± 0.16	1.14 ± 0.15	0.77 ± 0.02	0.71 ± 0.08	0.80 ± 0.06	0.96 ± 0.09	0.86 ± 0.09	0.91 ± 0.08
Margalef	14 ± 1	11 ± 1	12 ± 2	17 ± 0	14 ± 1	14 ± 1	14 ± 1	14 ± 1	14 ± 1
Fisher (a)	20 ± 1	16 ± 1	17 ± 3	24 ± 1	19 ± 1	20 ± 1	19 ± 1	20 ± 1	19 ± 1
Chao-1	177 ± 8	142 ± 14	138 ± 30	217 ± 6	182 ± 11	183 ± 11	170 ± 11	182 ± 11	171 ± 12
ACE	175 ± 8	141 ± 13	136 ± 30	217 ± 6	181 ± 11	182 ± 11	169 ± 11	180 ± 11	170 ± 12

Table 2. Bacterial abundance and α-diversity mean and standard deviation (SD) at lake Balaton, the four basins of lake Balaton, the Southern and Northern parts, and at the pelagic and Littoral zone of lake Balaton.

	Basins	Shore	Zone
Taxa (S)	15.50**	1.02	0.39
Dominance (D)	3.84	1.07	2.05
Shannon (H)	6.56	0.21	2.53
Brillouin	6.60	0.59	1.83
Menhinick	8.93*	3.91*	0.38
Margalef	19.22***	0.89	0.15
Fisher (a)	21.00***	0.68	0.05
Chao-1	15.84**	0.93	0.46
ACE	18.66***	1.43	0.40

Table 3. Kruskal-Wallis (Kruskal-Wallis one way analysis of variance on Ranks) comparison of the studied parameters based on the 4 basins of lake Balaton, the Southern and Northern parts, and the pelagic and Littoral zones of lake Balaton. * - shows significance $P < 0.05$, ** - shows significance $P < 0.01$, *** - shows significance $P < 0.001$.

The spatial distribution of families captured by the long-read sequencing analysis across the basin shows a remarkable variability (Fig. 3; Tables 2, 3). Not only there is a more than 3x difference between the most diverse and the least diverse sampling sites (central pelagic part of the basin 3 vs. western pelagic part of the basin 2: 224 vs. 63 families, Fig. 3), but there was a 70% difference between the least diverse and most diverse basin of the lake (basin 1 vs. basin 3; 111 ± 13 and 188 ± 5 families respectively, Tables 2, 3). The captured genetic variability of bacterioplankton (Fig. 3) suggests complex interaction of ecological dynamics and environmental factors that influence the distribution of microbial communities in the lake.

Differences in bacterioplankton abundance were observed across the lake's basins. Basins 3 and 4 showed the highest and lowest Shannon indices, respectively, with a 13% difference. Basins 1, 2, and 4 exhibited similar levels of family abundance and species diversity (Tables 2, 3; Fig. 3). The most diverse sampling site (central pelagic part of basin 3) had 224 families, while the least diverse (western pelagic part of basin 2) had 63 families, showing significant spatial variability (Fig. 3).

The microbiota of Lake Balaton featured 253 microbial families overall and could be divided into four separate basins (Figs. 1 and 4). The long read sequencing approach has identified different numbers of taxa for each basin: Basin 1 contained 230 taxa, Basin 2–222, Basin 3–248, and Basin 4 had 244. A Venn diagram was generated to illustrate the similarities and differences in family compositions found among the basins (Fig. 4). Most of these (209 taxa – 83%) were shared across all basins, suggesting a high degree homogeneity and a core microbiome essential for shallow lakes functioning, nutrient cycling and organic matter decomposition. Only Basin 3 and 4 contained unique families, with three (Desulfurobacteriaceae, Hafniaceae and Parvularculaceae) and two (Bernardetiaceae and Rubritaleaceae) families, respectively. There were also taxa shared only by two basins. For example, basins 3 and 4 shared 17 taxa. Specific overlaps in composition of the microbiota between the basins may indicate potential similarities in environmental conditions or microbial exchange due to hydrological connectivity.

The β-diversity of bacterioplankton in Lake Balaton in the whole lake was moderate to high (for example, Whittaker: 0.71 and Wilson-Shmida: 11.3, Table 4). This suggests that there were differences in bacterioplankton composition between sites within Lake Balaton, with some areas exhibiting a higher degree of turnover in species composition compared to others. For example, the westernmost basin 1 had the highest β-diversity indices (Whittaker: 1.07 and Wilson-Shmida: 3.8), indicating that the bacterioplankton composition showed the most pronounced differences between sites within basin 1, while basin 3 exhibited the lowest indices (Whittaker: 0.32

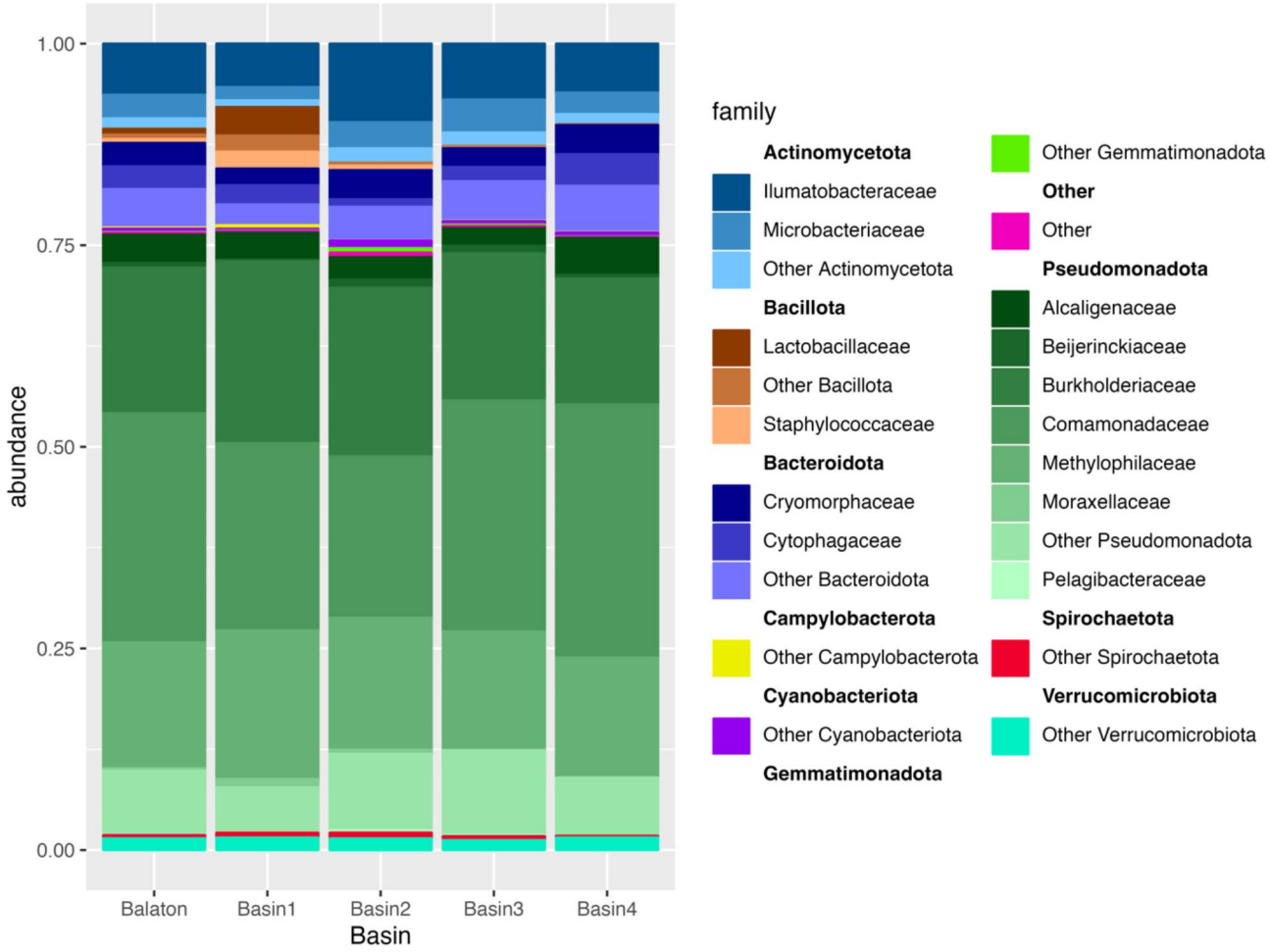

Fig. 2. Relative frequency of bacterioplankton community of Lake Balaton and its basins, showing only families exceeding 1%.

and Wilson-Shmida: 1.93), thus this basin was the most homogeneous, showing the least variation in species composition between sites (Table 4).

To fully understand the dynamics of bacterioplankton communities in Lake Balaton, it is crucial to examine the spatial distribution and variability of these microorganisms across different regions of the lake. Spatial variability can reveal how environmental factors, and localized conditions influence the composition and abundance of microbial populations. This variability can be visualized using principal component analysis (PC, Fig. 5), where PC-1 and PC-2 combined explain over 50% of the total variation. Regional differences are likely attributed to local environmental conditions and territorial influences. Although the PCA shows substantial differences between the basins, these differences were not statistically significant ($P > 0.05$), indicating together with the content of Fig. 4 that the bacterioplankton community of Lake Balaton is largely uniform across the basins, with only minor local variations, that this consistency, along with the observed spatial differences, suggests a stable microbial community that is resilient to environmental fluctuations, and that more detailed studies are needed to understand the underlying causes of spatial distribution patterns. The basins have been designated in the past based on the limnological data. Since there may not be a strict congruence between the limnological basins and the representative microbiota we identified with the new approach. In this context, the distribution of representation of six most widespread families across the entire basin using the percentage of sequencing reads representing each taxon (Fig. 6) illustrates the significant spatial heterogeneity of bacterioplankton within Lake Balaton. While some bacterial families, such as Methylophilaceae and Cryomorphaceae, are evenly distributed throughout the lake, others like Comamonadaceae and Burkholderiaceae show localized abundance, being prevalent in certain areas and scarce in others.

In addition to the basins, the contrast between the studied areas (northern and southern shores as well as littoral zone and pelagic zone) of Lake Balaton showed moderate to high β-diversity, indicating the higher variability of bacterioplankton communities in these limnological areas than in some basins (Table 4). In particular, the Wilson-Shmida index showed that the β-diversity of these mesoscale limnological areas (the littoral zone of Lake Balaton is 240 km long) could be higher than the basin-level biodiversity (the most diverse basin 1 is merely 8 km wide) (Table 4). Although some of the families may be associated with specific niches,

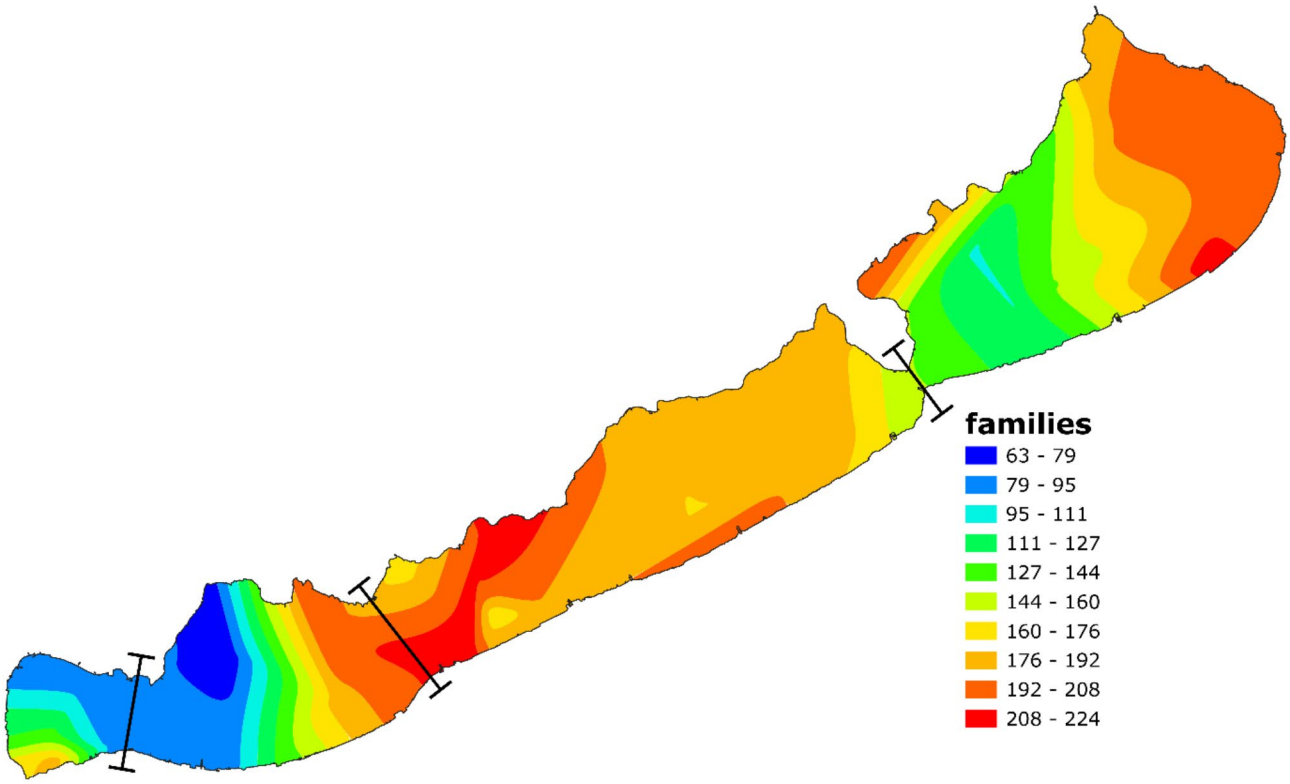

Fig. 3. The spatial variability of bacterioplankton abundance in Lake Balaton in May 2019. The black straight lines delineate the borders of the basins.

they were observed in Lake Balaton's littoral and pelagic zones, resulting in a nearly identical bacterioplankton footprint of the studied zones (Fig. 7A).

In addition, some minor differences were also observed between the bacterioplankton communities of the southern and northern shores of Lake Balaton (Fig. 7B). The similarity of the community composition was also supported by the very similar α- and β-diversity parameters (Tables 2, 3 and 4). These results suggest quite a homogenous bacterioplankton community within the whole of Lake Balaton.

The in situ determined limnological parameters in Lake Balaton, including temperature, pH, conductivity, chlorophyll-a, CDOM, Secchi depth, TSM, showed varying degrees of correlation with the abundance of different bacterioplankton families (Table 5). The observed levels of correlations provide valuable insights into the level of influence of environmental parameters of the lake on the distribution of the bacterioplankton communities within Lake Balaton. For example, the amount of dissolved organic matter (CDOM) was significantly correlated with the abundance of 46% of the dominant bacterioplankton families and was the only factor significantly (negatively) correlated with total bacterioplankton abundance. In contrast, pH, conductivity, and suspended matter were only marginally correlated with family abundance (Table 5).

Discussion

The study of bacterioplankton communities in Lake Balaton using long-read sequencing has provided significant insights into the diversity and distribution of microbial life within the lake, bringing attention to the taxonomic diversity and spatial variability with this lake. The findings indicate a high taxonomic richness as 253 microbial families were identified, ruling our first hypothesis supported. This high diversity underscores the complexity of ecological interactions and the dynamic nature of microbial ecosystems in large freshwater bodies as well as advantage of the use of long-read sequencing technologies in examining bacterioplankton communities that led to significant improvements in taxonomic resolution compared to previous studies[30,31] the ability to detect rare species, and the acquisition of valuable functional insights on microbial communities. These advancements are essential for comprehending the dynamics of microbial populations in freshwater environments.

The α-diversity indices revealed a moderate diversity of bacterioplankton across the lake and a relatively even distribution of species, with no single taxon overwhelmingly dominating the community. This balance may be crucial for the ecological stability of Lake Balaton, allowing for effective nutrient cycling and organic matter decomposition[45,46]. The presence of three predominant families (Comamonadaceae, Burkholderiaceae, and Methylophilaceae, comprising 62.5% of the community), indicates that these groups may play key roles in the lake's microbial processes. These data suggest that the specificity of Lake Balaton, dominated by mesoscale environmental factors (temperature, light availability, geomorphology, etc.), plays the dominant shaping role. However, significant local variations were observed at the basin level. For example, basin 3 had the highest

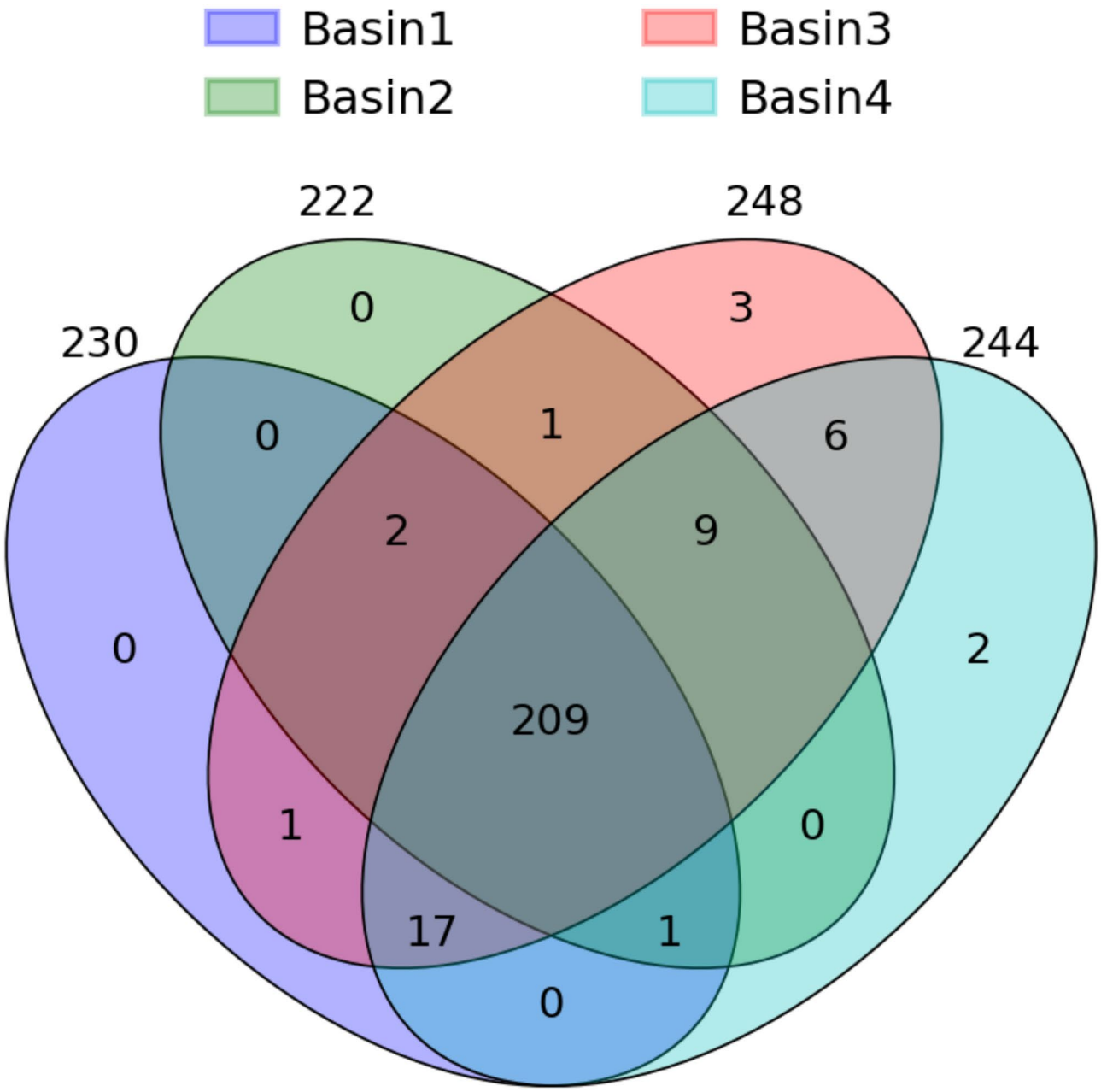

Fig. 4. A Venn diagram illustrating the numbers of shared and unique microbial families among 4 basins of Lake Balaton (a verbose explanation of intersections is included in Figure S2).

α-diversity, with a greater number of taxa and a more even distribution of species, indicative of stable, oligotrophic conditions, whereas the westernmost basin 1 had the lowest α-diversity, indicating a less diverse community, consistent with eutrophic conditions of the basin[8,21,47]. These differences in α-diversity suggest that local environmental factors specific to each basin (trophic states, sediment characteristics, etc.) also influence the composition of their bacterioplankton communities.

The moderate to high β-diversity indicates high community turnover across the lake. Basin 1, exhibiting the highest β-diversity, suggests significant variability in species composition across its sites, a high degree of species turnover connected to great spatial environmental heterogeneity within this basin. This outstanding variability is related to the inflow from Kis-Balaton Water Protection System, as its wetland provides a significant amount of organic matter (CDOM), nutrients (phosphorus, nitrogen), which are consumed in the westernmost basin of Lake Balaton and probably specific bacterial communities from the wetlands. In contrast the meso-oligotrophic basin 3 is more homogeneous despite being nearly 5 times larger, with a more consistent bacterioplankton composition across its sampling sites. This diversity in community composition could be influenced by factors such as water movement, lack of significant nutrient influx, and habitat homogeneity. Moreover, the meso-scale limnological areas, such as the littoral or pelagic zone had lower β-diversity compared to the β-diversity of basin

		Basins				Shores		Zones	
	Balaton	1	2	3	4	Southern	Northern	Littoral	Pelagic
n	33	6	4	14	9	16	17	17	16
Whittaker	0.71	1.07	0.95	0.32	0.61	0.62	0.75	0.64	0.73
Harrison	0.016	0.097	0.189	0.025	0.051	0.030	0.034	0.029	0.035
Cody	1657	420.5	293	356	481	857	782.5	828	826
Routledge	0.141	0.186	0.156	0.081	0.130	0.130	0.146	0.134	0.144
Wilson-Shmida	11.3	3.8	2.6	1.9	3.2	5.6	5.5	5.4	5.8
Mourelle	0.26	0.34	0.51	0.15	0.26	0.27	0.25	0.25	0.28
Harrison 2	0.003	0.011	0.019	0.008	0.009	0.005	0.006	0.005	0.008
Williams	0.108	0.109	0.086	0.097	0.098	0.097	0.109	0.100	0.138

Table 4. Bacterioplankton β-diversity at lake Balaton, the four basins of lake Balaton, the Southern and Northern Shores, and at the pelagic and Littoral zone of lake Balaton.

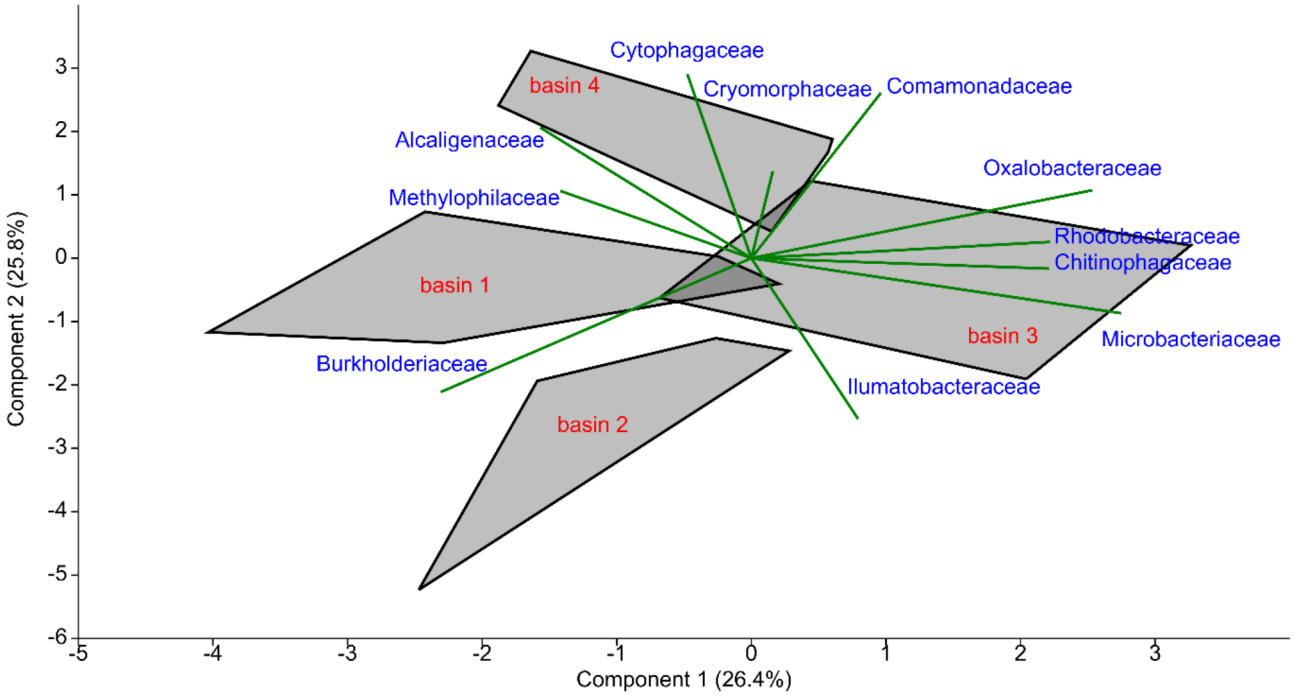

Fig. 5. PCA biplot with families' arrows (green lines) of Lake Balaton bacterioplankton based on the 11 most frequent families (over 1%). The detailed distribution of bacterioplankton families in the basins 1 and 4 of Lake Balaton are described in Figures S3 and S4.

1. Moreover, while basin 3 was the most homogenous, it also emerged as the most taxonomically rich, hosting 188 ± 5 families, indicating that while the basin harbors a high number of taxa, the environmental conditions allow these taxa to be evenly distributed, leading to a consistent, stable community structure, showing that on the bacterioplankton community basis the basin 3 is a strong and resilient ecosystem. Contrary to this, basin 1 had the lowest taxa abundance (111 families), implying a more unbalanced ecosystem in the basin.

Differences between southern and northern shores were not pronounced despite the specificities of the southern (more intensive wave impact, sandy shoreline, lower organic content) and northern (deeper waters, clayey sediments, higher nutrient content) shore[23]. More importantly, quite similar diversities were also observed between the pelagic and littoral zones, although the pelagic zone showed slightly higher diversity indices. The littoral zone showed reduced diversity, probably due to physical peculiarities and selective pressure from human recreational use. Although principal component analysis revealed clustering trends among lake basins, complementary non-parametric comparisons did not reveal statistically significant differences in community structure. This suggests that the bacterioplankton community remains largely stable across the lake, and future research should focus on potential functional differentiation using metagenomic or transcriptomic approaches.

The lack of a statistically significant difference in bacterioplankton composition between the basins of Lake Balaton suggests a high degree of connectivity and homogenization across the lake. Lake Balaton is a relatively shallow, wind-driven system with strong water circulation that facilitates the mixing of bacterioplankton

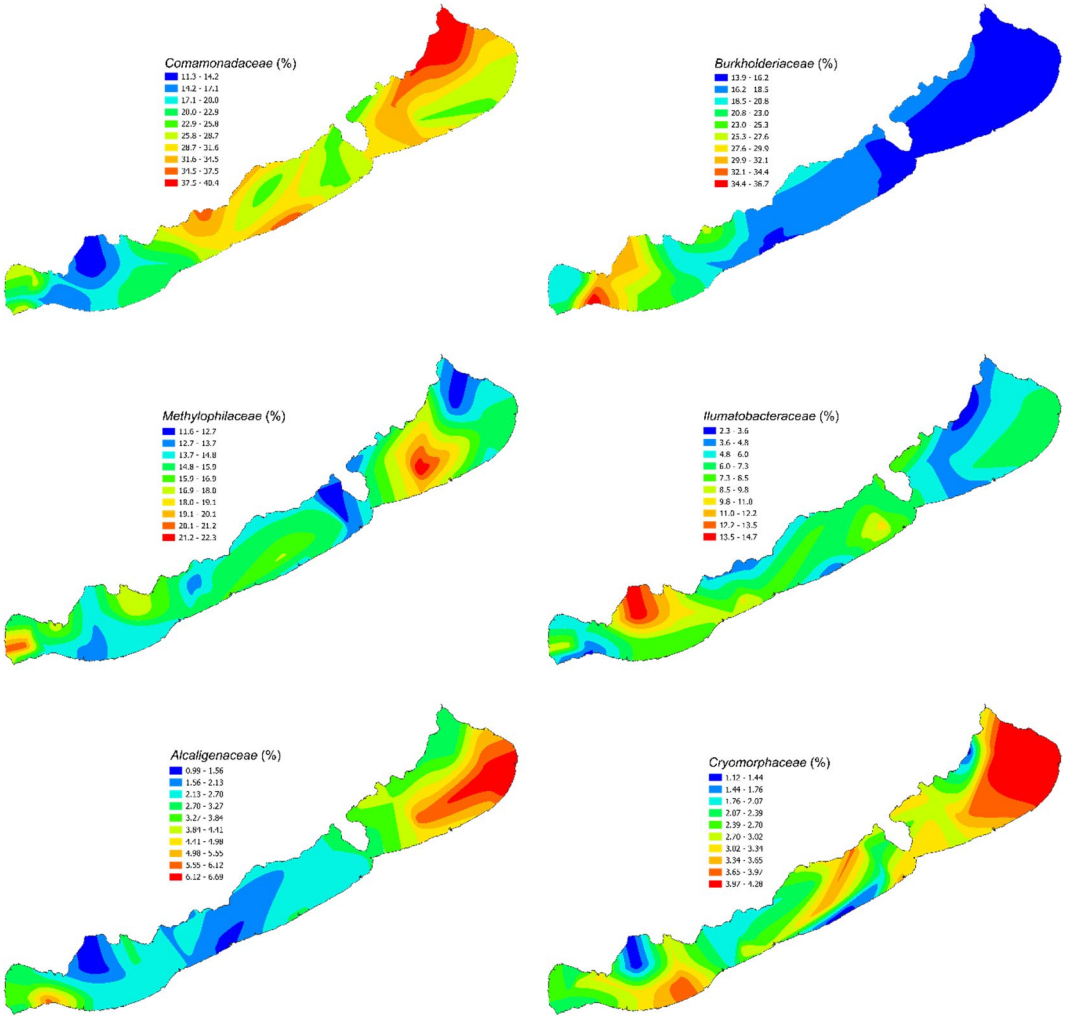

Fig. 6. Spatial distribution of bacterioplankton families in Lake Balaton. distribution of representation of six most widespread families (Comamonadaceae, Burkholderiaceae, Methylophilaceae, Ilumatobacteraceae, Alcaligenaceae, and Cryomorphaceae) across the entire basin using the percentage of sequencing reads representing each taxa.

communities across basins. This hydrological connectivity is likely to reduce spatial differentiation by dispersing microorganisms across the lake. Furthermore, despite some local variations, key limnological parameters driven by the uniform environmental background of the lake (geomorphology, temperature, light availability) are broadly similar across the basins, creating the idiosyncrasy of Lake Balaton. The presence of a stable, functionally important core bacterioplankton community that thrives in all basins suggests that environmental filtering selects for taxa that are well adapted to the overall conditions of the lake. This common core community may contribute to the observed similarity in diversity indices and underlies the bacterial resilience of the lake.

The observed bacterioplankton variability could be contextualized within the ecological history and management of Lake Balaton. The lake's west-east trophic gradient, historically shaped by differential nutrient loading from the Zala River, remained a key driver of algal community structure[8,21,47]. Successful restoration efforts, such as the Kis-Balaton Water Protection System, reduced external phosphorus inputs and shifted the lake from hypertrophic to meso-eutrophic, but affected the westernmost Basin 1 with additional organic matter (CDOM) loading and bacterial colonization from its wetlands. In addition, climate change, with its projected warming and increased evaporation, could intensify thermal stratification and alter oxygen and nutrient dynamics in the pelagic zone, particularly in the westernmost basin 1[48]. The study provided valuable insights into the ecological factors influencing the distribution of microbial communities by correlating key limnological parameters with bacterioplankton abundance in Lake Balaton. It was shown that changes in temperature and CDOM correlated with the distribution of major families, suggesting their possible role in shaping microbial communities, possibly by influencing growth conditions. Chlorophyll-a concentration showed a significant correlation with several families. These correlations may indicate direct or indirect relationships between phytoplankton and bacterioplankton, where higher chlorophyll-a may favor some bacterial groups while others may decline due to competition or a predator-prey relationship. Other measured limnological parameters (conductivity, pH, and total suspended solids) showed only marginal correlations with the abundance of

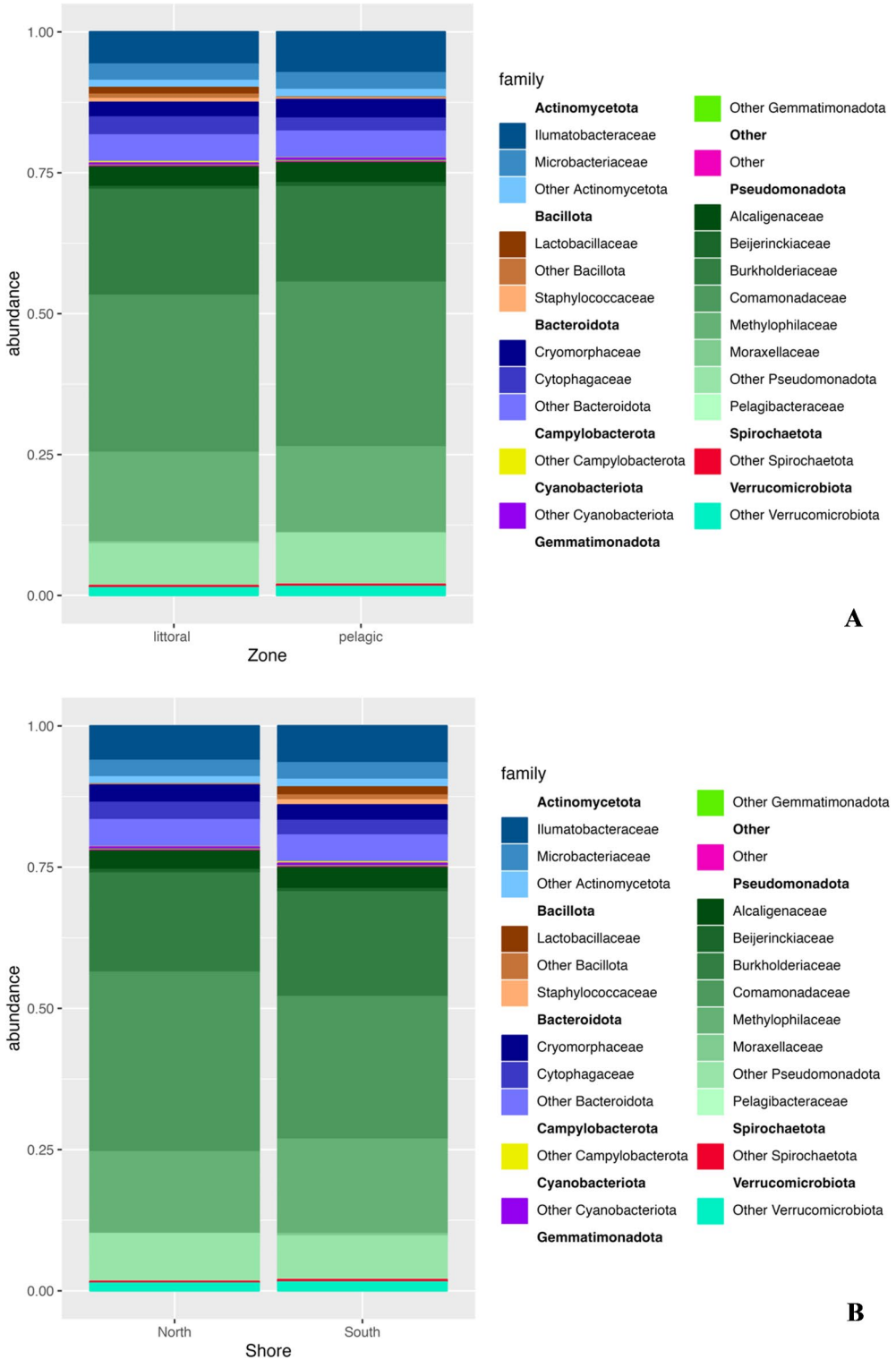

Fig. 7. Family level variability in relative abundance between the bacterioplankton communities in the littoral and pelagic zone (**A**), and the northern and southern shores (**B**) of Lake Balaton.

bacterial families, probably due to the spatial uniformity of these conditions in Lake Balaton. These results highlight the importance of considering these interactions when assessing the ecological status and functioning of the lake and when formulating strategies for water quality management.

The high degree of shared taxa across all basins (209 taxa, 83%) suggests the presence of a core microbiome essential for the lake's ecosystem functioning. This core microbiome likely plays a vital role in maintaining ecological processes such as nutrient cycling and organic matter decomposition, which are crucial for the lake's

	Temp.	pH	Conduct	Chl_a	CDOM	Secchi	TSM
Comamonadaceae	**0.475****	0.175	−0.297	−0.321	**−0.466****	0.256	−0.081
Burkholderiaceae	−0.184	−0.156	**0.390***	**0.665*****	**0.782*****	**−0.557*****	0.300
Methylophilaceae	0.062	0.046	−0.123	−0.076	0.021	0.292	−0.269
Ilumatobacteraceae	−0.083	−0.045	0.066	0.218	0.204	−0.278	0.294
Alcaligenaceae	**−0.455****	**0.439***	−0.214	**−0.473****	−0.290	**0.566*****	**−0.628*****
Cryomorphaceae	**−0.424***	0.261	−0.190	**−0.393*****	−0.291	0.250	−0.260
Microbacteriaceae	**0.427***	−0.117	−0.141	0.026	−0.103	−0.161	0.338
Cytophagaceae	0.063	0.103	**−0.378***	**−0.644*****	**−0.554*****	**0.787*****	**−0.640*****
Oxalobacteraceae	**0.559*****	0.070	−0.268	−0.151	**−0.432****	−0.168	0.349
Rhodobacteraceae	**0.519****	0.021	0.042	**−0.430****	**−0.441****	0.179	−0.091
Chitinophagaceae	0.304	−0.129	0.101	−0.040	−0.102	−0.070	0.215
Total runs	0.302	−0.025	−0.270	−0.138	**−0.353***	0.223	0.006

Table 5. Spearman rank order correlation coefficients and significance between the measure Limnological parameters and bacterioplankton abundance in lake Balaton in May of 2019. * - shows significance $P < 0.05$, ** - shows significance $P < 0.01$, *** - shows significance $P < 0.001$.

health and stability. At the same time, the different groups identified between basins in lake water suggest a complex microbial ecosystem and highlight the need for ongoing monitoring to assess potential health risks and ecological changes[26].

This study reveals Lake Balaton as a dynamic microbial ecosystem characterized by high diversity and complex spatial variability ruling our second hypothesis supported. The observed patterns of bacterioplankton diversity and distribution are influenced by a combination of environmental factors and hydrological connectivity. Understanding these dynamics is critical for managing and preserving the ecological health of large freshwater shallow lakes, like Lake Balaton. Future research could focus on the functional roles of dominant and unique taxa and their responses to environmental changes to gain deeper insights into microbial ecosystem functioning.

The use of Oxford Nanopore MinION sequencing technology (ONT) in this study of Lake Balaton has provided unprecedented insights into the genetic diversity and distribution of bacterioplankton across the lake. The application of this sequencing method has enabled the detailed characterization of microbial communities, revealing a total of 253 microbial families distributed among the lake's various areas. This technological advancement has allowed for a more comprehensive understanding of the bacterioplankton diversity than previous studies, highlighting the potential of long-read sequencing to uncover complex ecological patterns in aquatic environments.

While short-read next-generation sequencing (NGS) methods—particularly Illumina-based platforms—have been foundational in microbial ecology, several studies have demonstrated critical limitations in their application to environmental microbiome analysis. However, the major drawback of the NGS for analyzing microbiota is their restricted read length (typically ~ 250–300 bp), which limits phylogenetic resolution, especially at the species or strain level[36,42]. This often results in ambiguous or incorrect taxonomic assignments, particularly among closely related taxa. Moreover, short-read 16 S rRNA sequencing relies on the amplification of hypervariable regions (e.g., V3–V4), which are not uniformly informative across all bacterial clades and can introduce biases due to primer mismatches[6,49]. In highly diverse and novel ecosystems such as freshwater lakes, this leads to the underrepresentation of rare or uncultured taxa and a skewed community profile[38]. Short-read sequencing also struggles to resolve full-length 16 S gene structures, thereby limiting downstream applications such as genome-resolved metagenomics or strain tracking. These limitations have motivated the adoption of long-read platforms like Oxford Nanopore and PacBio, which can sequence the entire 16 S rRNA gene (~ 1,500 bp) or even full operons, greatly improving taxonomic resolution and minimizing PCR biases[37,40]. The improved accuracy of long-read technologies makes them a viable and often superior option for characterizing complex and poorly studied environmental microbiomes.

Recent advancements in long-read sequencing have improved its feasibility for microbial community analysis despite historical concerns regarding higher error rates and costs. Short-read sequencing technologies, such as Illumina, exhibit error rates as low as 0.24% ± 0.06%, making them highly reliable for high-throughput applications[50]. However, short-read methods often struggle to resolve closely related taxa beyond the genus level due to read length limitations. In contrast, PacBio HiFi sequencing has reached reported accuracy levels of up to 99.9%, offering species- and strain-level taxonomic resolution crucial for microbiome studies[36,37,51]. While Oxford Nanopore Technology (ONT) initially exhibited error rates between 15% and 40%, continual improvements in flowcell chemistry (e.g., R10.4.1) and base-calling algorithms have substantially increased its accuracy, now approaching 99% for species-level microbial profiling[52,53]. Although long-read sequencing is still more expensive per-gigabase basis, its ability to capture full-length 16 S rRNA gene sequences reduces the need for complex computational assembly. It helps minimize biases introduced by primer selection[40]. Comparative studies have further demonstrated that full-length 16 S sequencing, coupled with well-curated reference databases, significantly enhances microbial community profiling—especially in complex environmental and clinical settings[38,39]. Thus, while short-read platforms remain cost-effective and precise for many large-scale

projects, the increasing accuracy and resolution of long-read platforms make them an attractive complement for achieving higher taxonomic resolution in microbiome research.

One remarkable aspect of the ONT long-read sequencers is their portability, enabling their use in diverse settings. Unlike typical, large, benchtop sequencing devices, these small long-read sequencing devices can be directly plugged into a laptop and have been employed in unconventional environments[42], from the International Space Station[54] to remote field sites during disease outbreaks[41], showcasing their adaptability and potential for rapid response in monitoring microbial populations. The design of our study took advantage of this portability by setting up a mobile lab with a minimal setup that moved a portable sequencer to the research site at Lake Balaton. This allowed us to generate and analyze sequencing data for each sample in real time, providing results much more quickly than typical short-read sequencing technologies. While short-read sequencing technologies offer higher throughput due to their ability to multiplex many more samples on one sequencing run, the rapid response and detailed insights provided by long-read sequencing were invaluable for our study.

Several challenges remain before the long read sequences can be used to routinely monitor the lake microbiota. One of these is in the use of PVDF membranes for primary genetic analysis, which focused on bacterial profiles smaller than 0.1 μm. This approach inherently limits our analysis to smaller bacterial species, potentially excluding larger bacteria from comprehensive genetic profiling. To account for this, we employed additional pre-filtering with GF-C membranes (Millipore Sigma, Massachusetts, USA), which were stored at −20 °C for comparison with the 0.1 μm PVDF membranes. As anticipated, this fraction was overwhelmingly dominated by Cyanobacteria (Table S2). However, it is important to note that Cyanobacteria are large enough to be effectively counted and analyzed using alternative methods such as microscopy, which raises questions about the necessity and efficiency of using genetic analysis for such large organisms. The other major factor is the comparatively higher cost of long-read vs. short-read sequencing which can be a barrier for some research programs. In addition, there is an issue of accuracy - the long-read technologies historically have higher error rates compared to short-read sequencers[55]. Finally, there is an issue of advanced analysis of long-read sequencing, because this type of data requires specialized bioinformatics tools and expertise. Developing user-friendly software and pipelines will be essential for integrating long-read sequencing into routine environmental monitoring. However, long-read sequencing technologies are constantly improving and gradually closing the gap in terms of accuracy and costs[56], and continuous improvements in base-calling algorithms and error-correction methods are mitigating these issues. Efforts to reduce costs and increase the accessibility of these technologies are crucial for broader adoption in ecological studies.

We recognize that long-read sequencing can theoretically provide genus- or species-level taxonomic resolution[36,37]. However, environmental microbiomes frequently include novel or uncultured taxa with highly similar 16 S rRNA gene sequences, making species-level assignments challenging even with long-read platforms[38,39]. Moreover, our study utilized older Oxford Nanopore flow cells (R9.4) and a previous base-caller (Guppy), which present higher error rates than the latest chemistries[40]. Consequently, to maximize reliability and reduce false positives, we focused on reporting family-level identifications in our primary analysis. Nonetheless, for transparency and potential further exploration, we have provided the full taxonomic dataset (family, genus, and species assignments) in Table S3. This supplementary material enables researchers to get into the higher-resolution classifications where databases and methodological improvements allow.

The genetic analysis has uncovered significant differences in bacterioplankton diversity and distribution among the basins of Lake Balaton, with basin 3 identified as the most taxonomically rich and homogeneous, indicating an existing stable ecosystem that support a wide array of bacterioplankton families. In contrast, basin 1 was found to have the lowest bacterial abundance and the highest β-diversity, suggesting an unbalanced ecosystem characterized by considerable variability in community composition. These findings underscore the importance of spatial heterogeneity in shaping microbial communities and highlight the need for targeted management strategies tailored to the unique conditions of each basin. The stable and diverse bacterioplankton communities observed in Basin 3 suggest that existing environmental conditions should be preserved to maintain ecological resilience. Management practices should prioritize protecting water quality and preventing disturbances that could disrupt the balanced ecosystem. Conversely, Basin 1 requires attention to address its unbalanced microbial community and high variability. Management efforts should focus on stabilizing environmental conditions by reducing nutrient, dissolved organic matter inputs and mitigating potential stressors, such as pollution and habitat disturbance.

In contrast to previous studies of Lake Balaton that used short-read sequencing (e.g., Illumina) or T-RFLP, which generally reported findings at the family or higher taxonomic levels (Farkas et al., 2020; Tóth et al., 2020), our application of full-length 16 S rRNA gene sequencing enabled higher-resolution taxonomic assignments. Specifically, we identified over 120 distinct genera and 200 species across all samples. In contrast, earlier investigations typically reported fewer bacterial taxa due to the constraints of shorter read lengths and less comprehensive reference alignments. This expanded resolution is critical for elucidating the ecological roles of closely related or low-abundance taxa within Lake Balaton's bacterioplankton community. By capturing this finer-scale diversity, long-read sequencing facilitates more accurate assessments of community structure, potential functional capabilities, and the detection of rare organisms that may substantially influence nutrient cycling and other lake-wide processes.

Beyond taxonomic resolution, the observed microbial diversity in Lake Balaton plays a fundamental role in shaping ecosystem function. The high Shannon diversity in some basins suggests greater functional redundancy, which enhances resilience against environmental fluctuations. In contrast, distinct beta diversity patterns across basins indicate potential niche specialization, where taxa are differentially adapted to varying CDOM concentrations, hydrological inputs, and nutrient availability. A substantial body of research has demonstrated that CDOM (or its derivatives, such as humic substances or dissolved organic matter, amongst others) exerts a significant influence on the abundance and production of bacterioplankton and its community composition[57–60].

Lake Balaton	Basin 1	Basin 2	Basin 3	Basin 4
Area (km^2)	38	144	186	228
Volume (million m^3)	82	413	600	802
Average depth (m)	2.7	3.4	3.7	4.2
Length (km)	8	14	32	24
Width (km)	6	8	8	10

Table 6. Geomorphological characteristics of the individual basins of lake Balaton. The parameters shown are long-term (> 20 years) averages.

Such spatial heterogeneity underscores the role of bacterioplankton in sustaining key biogeochemical processes, including organic matter degradation and nitrogen cycling, which are crucial for maintaining freshwater ecosystem balance.

From a limnobacterial perspective, the study demonstrates that bacterioplankton diversity is influenced by environmental factors, though our study highlighted only the dissolved organic matter content of the water. The high degree of taxonomic richness and the presence of a core microbiome across the lake suggest that Lake Balaton still harbors a resilient microbial ecosystem capable of maintaining essential ecological functions. However, the variability observed in certain basins indicates that some areas may be more susceptible to environmental disturbances and may require focused management efforts to enhance ecological balance and stability.

Methods

Study area

Lake Balaton is a large (596 km^2) lake in Central Europe located in the western part of Hungary (N 46.83°, E17.71°, Fig. 1A). Lake Balaton is 78 km long, with a width of between 1.4 and 15.3 km; its average water depth is 3.7 m, with the deepest point of 12 m. Lake Balaton is traditionally divided into 4 basins due to the dominant geomorphological and limnological features (Fig. 1B; Table 1). The shoreline of the lake is 240 km long, with 110 km covered with emergent macrophytes (mostly common reed, ~ 14 km^2), while nearly all of its shoreline has submerged macrovegetation.

The 5765 km^2 catchment has a moderately developed industry, while the dominant land use is agriculture (arable and mixed cropping, orchards and vineyards), with tourism prevailing near Lake Balaton. Lake Balaton has one main (61% of the inflows) tributary, the Zala River, and a large number of smaller inflows (Fig. 1B). While Lake Balaton is mainly fed by the Zala River in the westernmost basin 1, and its only outlet is the Sió Canal (easternmost basin 4), resulting in a water residence time of about 4–5 years (Table 6).

Water sampling

Water samples were collected from Lake Balaton at 45 locations that were distributed over all four basins of the lake (Table S1, Figure S5), at the end of May and beginning of June 2019. This period was specifically chosen because of the lower abundance of cyanobacteria in the lake. Samples were categorized by basin, nearest shoreline (north and south) and distance from shore: samples collected less than 500 m from the shoreline were considered littoral, while samples collected at least 2 km from the shoreline were considered pelagic. The water samples were collected as depth-integrated over the upper 2 m of the water column into previously sterilized one-liter bottles, one bottle for genetic analysis, and one bottle for limnological analysis. The bottles were kept in a cooling box to maintain their original temperature and brought to the laboratory for further analysis within 2 h.

For genetic analysis, 400 mL of water were filtered under low vacuum pressure in the laboratory first through a Whatman GF-C coarse filter (Sigma Aldrich, Missouri, USA) and then through a Durapore PVDF membranes fine filter (MilliporeSigma, Massachusetts, USA). The PVDF membranes were used for primary genetic analysis focusing on the bacterial profile > 0.1 µm. In addition, pre-filtering GF-C (Millipore Sigma, Massachusetts, USA) membranes were stored at −20 °C for comparison with the 0.1 µm PVDF membranes.

Total suspended matter (TSM) was measured gravimetrically after filtering the water through a previously dried and pre-weighed GF-5 fiberglass filter (nominal pore size = 0.4 µm), drying for 2 h at 105 °C, then weighing again and subtracting the filter weight from the total weight. TSM concentration was calculated based on the predetermined volume of the filtered sample. The concentration of colored dissolved organic matter (CDOM) was measured spectrophotometrically at 440 nm (Hitachi, U-2900) after filtering the water through a cellulose acetate filter (pore size 0.45 µm) and was expressed as Pt (platina) units (mg Pt l^{-1})[61]. Chlorophyll-*a* of the water column was determined from freshly collected samples. 1 L of sampled water was filtered through Whatman GF-5 filter papers, and pigments were analyzed following extraction in 60 °C methanol and clarification by centrifugation (10,000 rpm for 10 min)[62].

Extraction of metagenomic DNA

For each location, the total DNA was extracted from a 0.1-micron Durapore™ PVDF membranes filter (MilliporeSigma, Massachusetts, USA) using DNeasy PowerWater Kit from QIAGEN (#14900, QIAGEN LLC, Maryland, USA) optimized to increase genomic DNA yields from low biomass samples like the filtered water samples, including turbid water. The extraction protocol was modified by cutting the membranes, adding 20 µL of 20 mg/mL Proteinase K (AM2546, ThermoFisher, Massachusetts, USA), adding either 500 µL or 1000

µL of Tris-buffered saline (TBS, 10x diluted into 1x, pH 7.4) (BP2471-100, Fisher Scientific, Massachusetts, USA) to the extraction beaded vials, incubating samples in a water bath for at least 8 h (overnight) at 65 °C and diluting the DNA in 80 µL solution of elution buffer (EB) (Supplemental Table 1). DNA was extracted from all 0.1-micron Durapore™ PVDF membrane filters. The DNA quality and concentrations were assessed by an IMPLEN Nanophotometer (IMPLEN, München, Germany).

Targeted 16S rRNA sequencing

The extracted DNA samples, *E. coli* DNA (positive control) and negative control, were divided into groups of twelve and subjected to targeted long-read sequencing on the MinION platform (Oxford Nanopore Technologies, Oxford, UK). A total of 44 samples were sequenced (Table S1). The library preparation, which includes PCR amplification of 16 S rRNA genes, was carried out by allocating each sample to a barcoded from twelve available in the Oxford Nanopore 16 S Barcoding DNA kits (SQK-RAB204) as per manufacturer's instructions using NEB LongAmp Taq 2X master mix (New England, Biolabs, USA). The protocol was modified by extending the PCR program from 25 to 30 cycles. Cycling conditions were the following: an initial denaturation at 95 °C for 1 min, followed by 30 cycles of 95 °C denaturation for 20 s, 55 °C annealing for 30 s, and 65 °C extension for 2 min, with a 65 °C final extension for 5 min. The 1.5k bp amplicons were purified using Agencourt AMPure X beads (Beckman Coulter, Brea, CA), pooled together, and normalized to a total of ~ 200 ng (16.7 ng/sample). Long read 1D sequencing was directly performed using R9.4 flow cells (Oxford Nanopore Technologies, Ltd) by loading 75 µL of sequencing mix to the SpotON port of the MinION Mk1B device run using MinKNOW 3.1.19 (Oxford Nanopore Technologies, Ltd., 2019).

Sequence quality control

In ONT-based sequencing, reads frequently display lower Phred scores relative to short-read technologies. To balance accuracy with read retention, we evaluated multiple Phred score cutoffs ($Q > 7$, $Q > 10$, $Q > 15$, and $Q > 20$) by examining their impact on total read counts and resulting taxonomic profiles. We ultimately chose $Q > 7$ (≈ 80% base accuracy) as our primary filter, following recommended practice in several environmental microbiome studies using older ONT chemistries[40,42].

Data processing and analysis

Sequencing was performed for 48 h on a MinION platform using version R9.4 flow cell (FLO-MIN106). Read processing and base calling of nanopore signals were performed locally on a MiniIT portable PC (Oxford Nanopore Technologies, Ltd. 2019, ARM processor 6 cores, 256 Core GPU, 8 GB RAM, 512 GB SSD) via MinKnow GUI, using base caller Guppy v2.3.5 pipeline[63]. At the time of sequencing, the Phred quality score was applied to the raw data. Reads with a Phred score of $Q > 7$ (representing a base accuracy of > 80%) were separated from the rest and labeled as passed the initial Quality Control Step.

The fastq files that passed QC were uploaded into EPI2ME[64] using the "FASTQ 16s" taxonomic classifier workflow, which is based on Centrifuge software[65,66] against the NCBI Bacterial 16 S taxonomic database (accession No.: PRJNA33175). Output data from the EPI2ME 16 S pipeline was analyzed using the CSVNanopore-master python script, which filters by accuracy > 80%, QC score > 7, e- value and read length > 1400 and < 1700 bp and retrieves the phylum and family name from the NCBI taxonomy IDs[67]. Frequencies of all taxonomic units (families and phyla) used in the analysis (Table S2). We have provided the full taxonomic dataset (family, genus, and species assignments) accessible in Table S3. The .csv files from the script were used to generate a taxonomic abundance table for further statistical analyses. Nanopore sequencing data from all available sequences has been deposited at the European Genome-Phenome Archive (ENA, PRJEB80228).

In this study, bacterioplankton abundance was defined as the number of sequencing reads assigned to bacterial taxa. This approach provides a relative measure of bacterioplankton abundance and community composition, reflecting the proportion of different taxa within the sampled microbial assemblage. Although read counts do not directly correspond to absolute cell numbers due to potential biases in DNA extraction and sequencing efficiency, they provide a robust comparative metric for assessing spatial and taxonomic variation in bacterioplankton communities across Lake Balaton.

Statistical analysis

Bacterial community dissimilarities between basins, zones, and shores were calculated by comparing abundance and α-diversity indexes obtained from the vegan package in R-studio[68]. Data transformation and visualization were performed using R Studio packages and Python programming language's data processing and visualization libraries. Statistical analyses, α- and β- diversity calculations for bacteria at the family level were performed with the 'vegan' package in R using R-Studio Version 1.2.1335[68]. To assess differences in α-diversity between the basins of Lake Balaton, the Kruskal-Wallis one-way analysis of variance on ranks was performed using R version 4.3.2[69]. This non-parametric test was chosen due to the non-normal distribution of the α-diversity metrics, as determined by the Shapiro-Wilk normality test. Families with less than 5 reads per basin were removed, and samples were rarefied to assess whether the sequencing effort was sufficient to capture the diversity present in the sample ('ggiNEXT' package R) (Figure S1)[70].

Data availability

All available sequences has been deposited at the European Genome-Phenome Archive (ENA, PRJEB80228).

Received: 29 October 2024; Accepted: 28 March 2025
Published online: 10 April 2025

References

1. Lindström, E. S., Forslund, M., Algesten, G. & Bergström, A. K. External control of bacterial community structure in lakes. *Limnol. Oceanogr.* **51**, 339–342 (2006).
2. Wu, Q. L. & Hahn, M. W. Differences in structure and dynamics of Polynucleobacter communities in a temperate and a subtropical lake, revealed at three phylogenetic levels. *FEMS Microbiol. Ecol.* **57**, 67–79 (2006).
3. Yannarell, A. C. & Triplett, E. W. Within- and between-Lake variability in the composition of bacterioplankton communities: Investigations using multiple Spatial scales. *Appl. Environ. Microbiol.* **70**, 214–223 (2004).
4. Schauer, M., Kamenik, C. & Hahn, M. W. Ecological differentiation within a cosmopolitan group of planktonic freshwater bacteria (SOL cluster, *Saprospiraceae*, *Bacteroidetes*). *Appl. Environ. Microbiol.* **71**, 5900–5907 (2005).
5. Amann, R. I., Ludwig, W. & Schleifer, K. H. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev.* **59**, 143–169 (1995).
6. Kim, C., Pongpanich, M. & Porntaveetus, T. Unraveling metagenomics through long-read sequencing: A comprehensive review. *J. Transl Med.* **22**, 111 (2024).
7. Padisák, J., Borics, G. & Grigorszky, I. Soróczki-Pintér, É. Use of phytoplankton assemblages for monitoring ecological status of lakes within the water framework directive: The assemblage index. *Hydrobiologia* **553**, 1–14 (2006).
8. Tóth, V. R. Exploring the functional connectivity between the Kis-Balaton water protection system and lake Balaton using satellite data. *Aquat. Sci.* **85**, 106 (2023).
9. Luettich, R. A., Harleman, D. R. F. & SomlyÓdy, L. Dynamic behavior of suspended sediment concentrations in a shallow lake perturbed by episodic wind events. *Limnol. Oceanogr.* **35**, 1050–1067 (1990).
10. Istvánovics, V., Honti, M., Kovács, Á. & Osztoics, A. Distribution of submerged macrophytes along environmental gradients in large, shallow lake Balaton (Hungary). *Aquat. Bot.* **88**, 317–330 (2008).
11. Herodek, S. The eutrophication of lake Balaton: Measurements, modeling and management. *SIL Proc. 1922–2010.* **22**, 1087–1091 (1984).
12. Vörös, L. & Göde, P. N. Long term changes of phytoplankton in Lake Balaton (Hungary). *SIL Proceedings, 1922–2010* **25**, 682–686 (1993).
13. Padisák, J. & Reynolds, C. S. Selection of phytoplankton associations in lake Balaton, Hungary, in response to eutrophication and restoration measures, with special reference to the cyanoprokaryotes. *Hydrobiologia* **384**, 41–53 (1998).
14. Newton, R. J., Jones, S. E., Eiler, A., McMahon, K. D. & Bertilsson, S. A guide to the natural history of freshwater lake bacteria. *Microbiol. Mol. Biol. Rev.* **75**, 14–49 (2011).
15. Garner, R. E. et al. A genome catalogue of lake bacterial diversity and its drivers at continental scale. *Nat. Microbiol.* **8**, 1920–1934 (2023).
16. Bullerjahn, G. S. et al. Community dynamics and function of algae and bacteria during winter in central European great lakes. *J. Great Lakes Res.* **46**, 732–740 (2020).
17. Vörös, L., Balogh, K. V. & Herodek, S. Microbial food web in a large shallow lake (Lake Balaton, Hungary). *Hydrobiologia* **339**, 57–65 (1996).
18. Istvánovics, V. & somlyódy, L. Factors influencing lake recovery from eutrophication — the case of basin 1 of lake Balaton. *Water Res.* **35**, 729–735 (2001).
19. Aulló-Maestro, M. E. et al. Spatio-seasonal variability of chromophoric dissolved organic matter absorption and responses to photobleaching in a large shallow temperate lake. *Biogeosciences* **14**, 1215–1233 (2017).
20. Padisak, J., -Toth, G., Rajczy, M. & L. & Stir-up effect of wind on a more-or-less stratified shallow lake phytoplankton community, lake Balaton, Hungary. *Hydrobiologia* **191**, 249–254 (1990).
21. Tóth, V. R. Making Temporal dynamics relevant: The role of macrophytes in the phenological complexity of a shallow freshwater lake. *Ecohydrology* **17**, 2663 (2024).
22. Vörös, L., Mózes, A. & Somogyi, B. A five-year study of autotrophic winter picoplankton in lake Balaton, Hungary. *Aquat. Ecol.* **43**, 727–734 (2009).
23. Tóth, V. R. Reed stands during different water level periods: Physico-chemical properties of the sediment and growth of phragmites australis of lake Balaton. *Hydrobiologia* **778**, 193–207 (2016).
24. Mózes, A., Présing, M. & Vörös, L. Seasonal dynamics of Picocyanobacteria and Picoeukaryotes in a large shallow lake (Lake Balaton, Hungary). *Int. Rev. Hydrobiol.* **91**, 38–50 (2006).
25. Kovács, A. W., Tóth, V. R. & Vörös, L. Light-dependent germination and subsequent proliferation of N_2 -fixing cyanobacteria in a large shallow lake. *Ann. De Limnologie - Int. J. Limnol.* **48**, 177–185 (2012).
26. Naderian, D. et al. A water quality database for global lakes. *Resour. Conserv. Recycl.* **202**, 107401 (2024).
27. Naderian, D. et al. Environmental controls on the conversion of nutrients to chlorophyll in lakes. *Water Res.* **274**, 123094 (2025).
28. Saravani, M. J. et al. Predicting Chlorophyll- *a* concentrations in the world's largest lakes using Kolmogorov-Arnold networks. *Environ. Sci. Technol.* **59**, 1801–1810 (2025).
29. Langó, Z. A Balaton Északi parti Régiója Bacillus Populációjának számítógépes analízise. *Az. MTA Biológiai Tudományok Osztályának Közleményei*. **25**, 313–326 (1982).
30. Farkas, M. et al. Planktonic and benthic bacterial communities of the largest central European shallow lake, lake Balaton and its main inflow Zala river. *Curr. Microbiol.* **77**, 4016–4028 (2020).
31. Tóth, E. et al. Insight into the hidden bacterial diversity of lake Balaton, Hungary. *Biol. Futur.* **71**, 383–391 (2020).
32. Szabó-Tugyi, N. et al. Aerobic anoxygenic phototrophs are highly abundant in hypertrophic and polyhumic waters. *FEMS Microbiol. Ecol.* **95**, fiz104 (2019).
33. Farkas, M. et al. Bacterial community dynamics along a River-Wetland-Lake system. *Water* **14**, 3519 (2022).
34. Homonnay, Z. G., Kéki, Z., Márialigeti, K. & Tóth, E. M. Bacterial communities in the gut of the freshwater copepod *Eudiaptomus gracilis*. *J. Basic. Microbiol.* **52**, 86–90 (2012).
35. Borsodi, A. K. et al. Gut content microbiota of introduced bigheaded carps (Hypophthalmichthys spp.) inhabiting the largest shallow lake in central Europe. *Microbiol. Res.* **195**, 40–50 (2017).
36. Callahan, B. J. et al. High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. *Nucleic Acids Res.* **47**, e103–e103 (2019).
37. Buetas, E. et al. Full-length 16S rRNA gene sequencing by PacBio improves taxonomic resolution in human Microbiome samples. *BMC Genom.* **25**, 310 (2024).
38. Jin, J., Liu, X. & Shiroguchi, K. Long journey of 16S rRNA-amplicon sequencing toward cell-based functional bacterial microbiota characterization. *iMetaOmics* **1**, e9 (2024).
39. Pascoal, F., Duarte, P., Assmy, P., Costa, R. & Magalhães, C. Full-length 16S rRNA gene sequencing combined with adequate database selection improves the description of Arctic marine prokaryotic communities. *Ann. Microbiol.* **74**, 29 (2024).
40. Waechter, C. et al. Comparative analysis of full-length 16s ribosomal RNA genome sequencing in human fecal samples using primer sets with different degrees of degeneracy. *Front. Genet.* **14**, 1213829 (2023).
41. Hoenen, T. et al. Nanopore sequencing as a rapidly deployable Ebola outbreak tool. *Emerg. Infect. Dis.* **22**, 331 (2016).
42. Johnson, S. S., Zaikova, E., Goerlitz, D. S., Bai, Y. & Tighe, S. W. Real-Time DNA sequencing in the Antarctic dry valleys using the Oxford nanopore sequencer. *J. Biomol. Tech.* **28**, 2–7 (2017).
43. Liem, M., Regensburg-Tuïnk, T., Henkel, C., Jansen, H. & Spaink, H. Microbial diversity characterization of seawater in a pilot study using Oxford nanopore technologies long-read sequencing. *BMC Res. Notes*. **14**, 42 (2021).

44. Menegon, M. et al. On site DNA barcoding by nanopore sequencing. *PLoS One*. **12**, e0184741 (2017).
45. Haukka, K. et al. Effect of nutrient loading on bacterioplankton community composition in lake mesocosms. *Microb. Ecol.* **51**, 137–146 (2006).
46. Vrede, K. Nutrient and temperature limitation of bacterioplankton growth in temperate lakes. *Microb. Ecol.* **49**, 245–256 (2005).
47. Somogyi, B., Felföldi, T., Tóth, L. G., Bernát, G. & Vörös, L. Photoautotrophic picoplankton – a review on their occurrence, role and diversity in lake Balaton. *Biol. Futur.* **71**, 371–382 (2020).
48. Li, H. et al. 150-year daily data (1870–2021) in lakes and rivers reveals intensifying surface water warming and heatwaves in the Pannonian ecoregion (Hungary). *J. Hydrol. Reg. Stud.* **56**, 101985 (2024).
49. Guo, F., Ju, F., Cai, L. & Zhang, T. Taxonomic precision of different hypervariable regions of 16S rRNA gene and annotation methods for functional bacterial groups in biological wastewater treatment. *PLoS One*. **8**, e76185 (2013).
50. Oehler, J. B., Wright, H., Stark, Z., Mallett, A. J. & Schmitz, U. The application of long-read sequencing in clinical settings. *Hum. Genomics*. **17**, 73 (2023).
51. Callahan, B. J., Grinevich, D., Thakur, S., Balamotis, M. A. & Yehezkel, T. Ben. Ultra-accurate microbial amplicon sequencing with synthetic long reads. *Microbiome* **9**, 130 (2021).
52. Johnson, J. S. et al. Evaluation of 16S rRNA gene sequencing for species and strain-level Microbiome analysis. *Nat. Commun.* **10**, 5029 (2019).
53. Zhang, T. et al. The newest Oxford nanopore R10.4.1 full-length 16S rRNA sequencing enables the accurate resolution of species-level microbial community profiling. *Appl. Environ. Microbiol.* **89**, e00605-23 (2023).
54. Castro-Wallace, S. L. et al. Nanopore DNA sequencing and genome assembly on the international space station. *Sci. Rep.* **7**, 18022 (2017).
55. Espinosa, E., Bautista, R., Larrosa, R. & Plata, O. Advancements in long-read genome sequencing technologies and algorithms. *Genomics* **116**, 110842 (2024).
56. Adewale, B. A. Will long-read sequencing technologies replace short-read sequencing technologies in the next 10 years? *Afr. J. Lab. Med.* **9**, 1–5 (2020).
57. Gasol, J. M. et al. A transplant experiment to identify the factors controlling bacterial abundance, activity, production, and community composition in a eutrophic canyon-shaped reservoir. *Limnol. Oceanogr.* **47**, 62–77 (2002).
58. Piccini, C., Conde, D., Pernthaler, J. & Sommaruga, R. Alteration of chromophoric dissolved organic matter by solar UV radiation causes rapid changes in bacterial community composition. *Photochem. Photobiol. Sci.* **8**, 1321–1328 (2009).
59. Figueroa, D., Rowe, O. F., Paczkowska, J., Legrand, C. & Andersson, A. Allochthonous Carbon—A major driver of bacterioplankton production in the Subarctic Northern Baltic sea. *Microb. Ecol.* **71**, 789–801 (2016).
60. Traving, S. J. et al. The effect of increased loads of dissolved organic matter on estuarine microbial community composition and function. *Front. Microbiol.* **8**, 351 (2017).
61. Cuthbert, I. D. & del Giorgio, P. Toward a standard method of measuring color in freshwater. *Limnol. Oceanogr.* **37**, 1319–1326 (1992).
62. Iwamura, T., Nagai, H. & Ichimura, S. Improved methods for determining contents of chlorophyll, protein, ribonucleic acid, and deoxyribonucleic acid in planktonic populations. *Int. Revue Der Gesamten Hydrobiol. Und Hydrographie*. **55**, 131–147 (1970).
63. ONT Development Team. Guppy basecalling software. Preprint at (2018). https://nanoporetech.com/document/Guppy-protocol#faq
64. Development Team, O. N. T. EPI2ME. Preprint at (2021). https://epi2me.nanoporetech.com/epi2me-docs
65. Kim, D., Song, L., Breitwieser, F. P. & Salzberg, S. L. Centrifuge: Rapid and sensitive classification of metagenomic sequences. *Genome Res.* **26**, 1721–1729 (2016).
66. Kim, D. et al. Preprint at (2018). https://www.ccb.jhu.edu/software/centrifuge/
67. Heikema, A. P. et al. Comparison of illumina versus nanopore 16S rRNA gene sequencing of the human nasal microbiota. *Genes* **11**, 1105 (2020).
68. Team, R-C. RStudio: Integrated Development for R. Preprint at (2020). http://www.rstudio.com/
69. R Core Team. R: A language and environment for statistical computing. Preprint at (2012). https://www.R-project.org/
70. Hsieh, T. C., Ma, K. H. & Chao, A. iNEXT: An R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods Ecol. Evol.* **7**, 1451–1456 (2016).

Acknowledgements

This research was part of the MS Thesis in Biological Sciences research supported by the Oakland University Startup Fund and the Visiting Fellowship at the Balaton Limnological Institute of the Centre for Ecological Research of the Hungarian Academy of Sciences.

Author contributions

SOCM, VRT and TKO had the original idea; SCOM, VRT and SK performed sampling and lab analysis, SOCM, TKO and WWW performed the analysis; SOCM wrote the original draft; SOCM, VRT TKO, WWW and SKO contributed to editing and final draft.

Funding

VRT was supported by the RRF-2.3.1-21-2022-00008 project and the Fulbright 2022 Scholarship. SOCM, WWW, and TKO were supported through the Oakland University Startup Fund, and TKO was supported through a Visiting Fellowship at the Balaton Limnological Institute of the Centre for Ecological Research of the Hungarian Academy of Sciences.

Declarations

Competing interests

The authors declare no competing interests.

Ethics approval

This research did not contain any studies involving animal or human participants, nor did it occur in any private or protected areas. No specific permissions were required for corresponding locations.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-96558-7.

Correspondence and requests for materials should be addressed to T.K.O.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.