BIODIVERSITAS
Volume 24, Number 8, August 2023
Pages: 4636-4644

ISSN: 1412-033X
E-ISSN: 2085-4722
DOI: 10.13057/biodiv/d240845

Microalgal diversity assessment of Laguna Lake (Philippines) tributaries using Next-Generation Sequencing for water quality

JOY ANN SANTOS[1,2,♥] , JULIUS AARON MEJIA[2], M.A. JOWINA GALARION[3], PIERANGELI VITAL[1]

[1]Biological Research and Services Laboratory, Natural Sciences Research Institute, University of the Philippines. Diliman, Quezon City, Metro Manila 1101, Philippines. Tel.: +632-89252963, ♥email: jpsantos8@up.edu.ph, pgvital@up.edu.ph

[2]Research and Biotechnology Department, Manila HealthTek, Inc.. Marikina City, Metro Manila 1800, Philippines

[3]Institute of Molecular Biology and Biotechnology, University of the Philippines National Institutes of Health. Manila, Metro Manila 1000, Philippines

Manuscript received: 10 March 2023. Revision accepted: 28 August 2023.

Abstract. *Santos JA, Mejia JA, Galarion MAJ, Vital P. 2023. Microalgal diversity assessment of Laguna Lake (Philippines) tributaries using Next-Generation Sequencing for water quality. Biodiversitas 24: 4636-4644.* The species composition and density of microalgae have been successfully used as bioindicators of water quality as they respond quickly to any change in the water chemistry brought about by pollution. Laguna Lake in the Philippines, one of the country's most important hotspots for aquaculture, continues to be threatened by eutrophication, yet its microalgal profile has been understudied. The presence of microalgae, especially those identified to become causative agents of harmful algal blooms in lakes, can cause corresponding health effects in humans and the environment. This study determined microalgal diversity in samples from tributaries of Laguna Lake, Philippines, through 18S rRNA metagenomic sequencing or Next-Generation Sequencing (NGS). Out of the five sampling sites, only two sites passed the NGS quality control pipeline. *Chlorella* M.Beijerinck, dominated the eukaryotic microbial population in the Alabang-Cupang River sample, which is 95% of the total population. While in the Biñan River sample, relevant species found were *Cyclotella scaldensis* Muylaert & Sabbe, *Teleaulax acuta* (Butcher) Hill, *Cryptomonas curvata* Ehrenb., and *Peridiniopsis penardii* (Lemmerm.) Bourr. This is the first study to use NGS to detect microalgae in Laguna Lake and warrants further elucidation of the microalgal population through increased sampling efforts in other recreational and livelihood hotspots. Furthermore, the study results open the importance of implementing new schemes for water quality assessments focused on molecular-based approaches.

Keywords: Laguna Lake, metagenomic sequencing, microalgal diversity, Next-Generation Sequencing, Philippines, water quality

INTRODUCTION

Water has been conveyed as a feasible reservoir for pathogens and a transmission route. Human and animal pathogens and even potentially pathogenic bacteria and protozoa are released with wastewater into aquatic environments that cause water pollution. Water bodies in high-density urban areas, such as lakes, rivers, and even artificial fountains, play essential roles in providing recreational opportunities for people. Humans may be repeatedly exposed to water during recreational activities, and microbial pathogens in water might threaten public health (Cui et al. 2016).

Laguna Lake is the largest inland body of water in the Philippines, with an aggregate surface area of 900 km^2 and around 76,000 hectares, with an average highest elevation of 12.50 meters and an average lowest elevation of 10.50 meters. Around 100 rivers and streams drain into the lake, of which 22 are significant river systems. These tributary rivers are the Sta. Cruz River, the Balanak River, the Pagsanjan River, the Marikina River, the Mangangate River, the Tunasan River, the San Pedro River, the Molawin, Dampalit and Pele Rivers in Los Baños, the Tanay River, the Morong River, the Cabuyao River, the San Cristobal River, the San Juan River, the Pangil River, the Bay, Calo, and Maitem Rivers in Bay, the Siniloan River, and the Sapang Baho River, Angono, Manggahan, Calauan, Sta.Maria, Jala-jala, Pililia, Baras, and Pila. The Napindan Channel is the only outlet through the Pasig River that drains lake waters to Manila Bay (Hernandez et al. 2012; Laguna Lake Development Authority (LLDA) 2015).

Laguna Lake is considered a multi-use resource body of water. It was selected to supply the expanding capital and mega-city of Metro Manila District, Philippines, with its fish and water needs while serving as a sink for urban effluents and floodwater (Saguin 2014). It is the largest lake in the Philippines and one of the significant aquaculture sites in the country, being one of the primary sources of freshwater fish. However, the lake is continuously being threatened by the growing population and harmful practices in its vicinity, posing a risk to the ecosystem of Laguna de Bay. Based on the 2015 Ecosystem Health Report Card of LLDA, Laguna de Bay scored *C*- on water quality and *F* in fisheries. This indicates that the lake, although it passed the Department of Environment and Natural Resources (DENR) guidelines (2014) in water quality, fared poorly because of its high population, introduction of invasive species, and industrialization of its vicinity and tributaries. The low scores translate into the lake being highly eutrophic, harming the primary source of livelihood (DENR – Environment Management Bureau, 2014). In the same year as the samples were collected (2019), LLDA reported that sampling stations near the sampling sites of this study

failed the Quality Guidelines for Classes A (public water supply), B (recreational water), C (Fishery Water) and D (navigation) in terms of the following parameters: biochemical oxygen demand, dissolved oxygen, ammonia, and fecal coliform (LLDA 2019).

Due to its highly eutrophic situation, Laguna Lake is prone to harmful algal blooms, which generally cause issues in various sectors, especially in the local aquaculture climate (Caballero and Navarro 2021). Although LLDA has both the developmental and regulatory mandate in Laguna Lake, the office cannot assume its developmental mandate to provide environmental management and control, preserve ecological systems, and prevent ecological deterioration due to its limited resources, with some of its policies not yet exercised. Since Laguna Lake has many practical uses, including fish cultivation and source for potable water production, among others (Ana and Espino 2020), the lake's water quality must be maintained as clean as possible and conform to monitoring guidelines, including monitoring harmful algal blooms. Hence, understanding the composition of the microbial communities in this kind of water is beneficial for detecting pathogens and improving our understanding of their ecological niches, tracking changes in the abundance of organisms responsible for adverse effects, such as corrosion or biofouling, and characterizing the assemblages of microbiota responsible for degradation of contaminants and microbial substrates in treatment processes (Garner et al. 2021).

Microalgae have been recognized as significant water quality bioindicators because of their sensitivity to alterations in the chemistry of the water brought on by pollution (Khalil et al. 2021). Palmer (1969) presented a pollution index based on the algae genus and species found in aquatic environments. Numerous studies have utilized this pollution index to analyze the water quality for organic pollution (Khalil et al. 2021; Villaruel and Camacho 2022). Hence, knowledge of the microalgal profile of Laguna Lake will not only help in providing data for the algal genus pollution index but also help design measures to control the growth of specific algal species better to prevent harmful algal blooms and implement more effective general water quality management strategies. However, methods used in the study of microalgae include mostly direct isolation from water samples and culture methods (Tayaban et al. 2018; Goss et al. 2020), which are time-consuming and labor-intensive processes.

Next-Generation Sequencing (NGS) is applied to explore the diversity of microorganisms for water quality assessment (Vierheilig et al. 2015; Shrestha et al. 2017). Past studies only dealt with the culture and morphology of algae; hence, they do not provide an accurate situation and analysis. The current study is the first to assess the microalgal diversity in samples from tributaries of Laguna Lake by using 18S metagenomic sequencing to determine the water quality. This will help establish baseline data on what species are in the lake and the potential benefit or contribution to adverse effects for better management and development process.

MATERIALS AND METHODS

Study area

Laguna Lake, Philippines, is fed by its 21 significant tributaries and 45,000 km^2 (17,000 sq mi) of catchment areas. Among these are the Pagsanjan River, which is the source of 35% of the lake's water; the Santa Cruz River with 15%; the Marikina River (through the Manggahan Floodway), the Tunasan River, the Mangangate River, the San Pedro River, the Cabuyao River, the San Juan River, the San Cristobal River, the Bay, Dampalit and Pele rivers in Los Baños, Calo and Maitem rivers in Bay, the Molawin, the Pangil River, the Siniloan River, the Tanay River, the Morong River, and the Sapang Baho River (Figure 1).

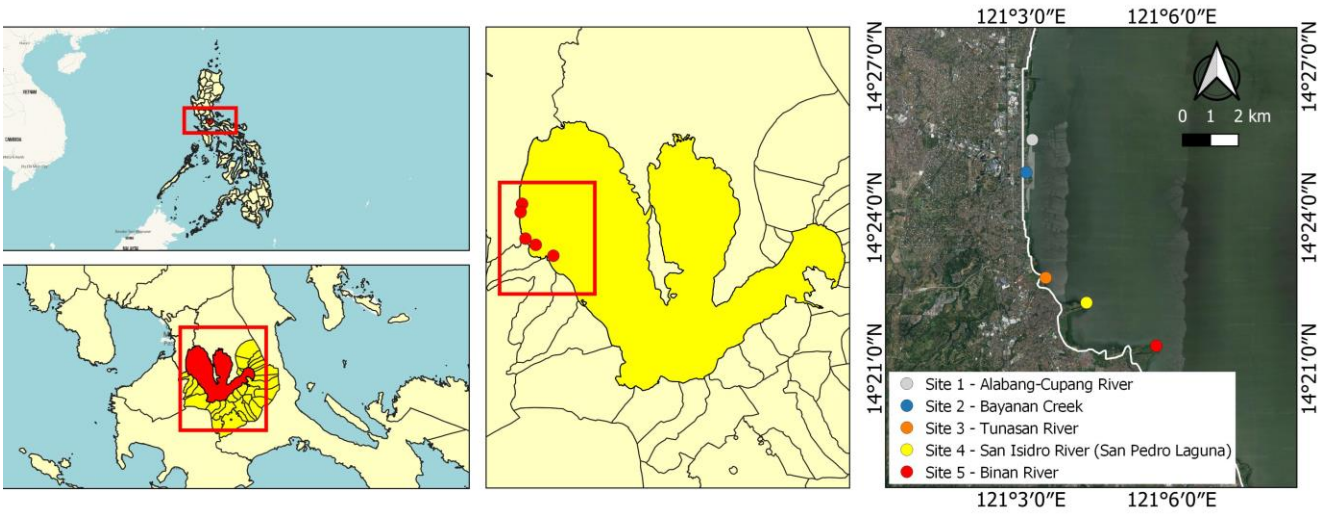

Figure 1. Location Laguna Lake, Philippines, indicating the sampling sites

Figure 1 shows the map of Laguna Lake and its surrounding land area, along with the tributaries that served as sampling sites for this study. Numbers and symbols indicated the locations. The identified sampling sites were the Alabang-Cupang River (Site 1), Bayanan Creek (Site 2), Tunasan River (Site 3), San Isidro River (Site 4), and Biñan River (Site 5), as they connect high population density areas in Metro Manila and Laguna to Laguna Lake. Coordinate of each site as follow: Site 1: N 14°25'33.323" E 121°3'13.741"; Site 2: N 14°24'54.421" E 121°3'6.623"; Site 3: N 14°22'48.871" E 121°3'29.869"; Site 4: N 14°22'19.196" E 121°4'18.507" and Site 5: N 14°21'27.76" E 121°5'41.834".

Sample collection and DNA extraction

Water samples were obtained from selected sites (Site 1 to 5) during the dry period (November 2019). At each river site, 2 L of water was collected a few meters from where the river meets the lake, 10 cm below the water surface. Samples were placed in sterile containers and transported in an ice box to the laboratory within 4 hours of collection. For nucleic acid extraction, 300 to 500 mL of water of each sample were filtered on 0.45 µm mixed cellulose membranes (GSWP29325, Merck Millipore, USA) and were subsequently filtered using 0.2 µm nitrocellulose membranes filters before adding DNA/RNA Stabilization reagent (Zymo Research, USA) and subsequently stored at -20ºC until nucleic acid extraction. Following the manufacturer's protocol, DNA was extracted using ZymoBiomics DNA/RNA Miniprep (Zymo Research, USA). The extracted nucleic acid quality was checked using Qubit™ 4 Fluorometer, based on the recommended protocol of the Qubit™ dsDNA HS Assay Kit (Invitrogen, USA) before being sent to Macrogen Korea for 18srRNA metagenomic sequencing using the MiSeq (Illumina) platform.

NGS analysis for characterization of microalgal community

The extracted DNA's quality was further validated using Victor 3 Fluorometry, where all samples for further processing should pass the minimum concentration of >0.1 ng/µL. Library construction was performed using the Herculase II Fusion DNA Polymerase (Agilent, USA) – Nextera XT Index Kit V2 (Illumina, USA) following Macrogen Korea's protocol for library preparation.

The sequence data for 18S NGS were analyzed using the MG-RAST version 4.0.3 software (Meyer et al. 2019). The application is an automated platform to perform a metagenomic analysis pipeline of quality control, annotation, comparative analysis, and archiving service of metagenomic and amplicon sequences using various bioinformatics tools. After upload, low-quality regions of the sequence data were removed by SolexaQA (Cox et al. 2010). A dereplication step was performed using a K-mer approach to remove artificial duplicate reads. The sequences were then screened to remove near-exact matches of the human host using Bowtie (Langmead et al. 2009).

Further screening through rRNA detection was performed to identify and cluster rRNA-similar reads at 99% identity using the cd-hit, and the longest sequence was selected as the cluster representative (Fu et al. 2012). The longest sequence of each cluster was picked as a representative and was used for a BLAST search against the M5rna database integrated with a reduced version of the SILVA database (Quast et al. 2013). A gene-calling step was performed to predict coding regions of the sequences through the FragGeneScan machine-learning approach (Rho et al. 2010). The data were integrated into various data products, such as abundance profiles, which were loaded into the database for analysis. Taxonomic labels were assigned to each contig and were visualized using a Krona pie chart. The sample has a corresponding Krona pie, indicating the organisms present and their abundance within the sample. To closely look at the more abundant organism in each sample, the contigs were filtered to include only those with at least 301 base pairs in length (corresponding to the expected size of the 18S library amplicon used in library preparation) and with those at least 5X depth of coverage (Ondov et al. 2011). Below is the typical workflow for each NGS approach used in this study (Figure 2).

RESULTS AND DISCUSSION

Quality control of water samples for the 18S metagenomic sequencing revealed that only sites 1, 2, 4, and 5 passed and were subsequently prepared for library construction. Initially, a more targeted amplicon NGS approach to detect protozoan parasites, *Cryptosporidium* spp. Tyzzer, 1907, *Giardia* sp., *Blastocystis* sp. A.G.Alexeieff, *Entamoeba* spp. and free-living amoebae were done using the primer pairs EUKAF (5' - GCC GCG GTA ATT CCA GCT C - 3') and EUKAR (5' - CYT TCG YYC TTG ATT RA - 3') (Moreno et al. 2018). However, during library construction, the primers did not work based on the failed results of the library quantity checking with a minimum concentration requirement of 10 ng/µL. This could be explained by the scarce target species or empty oocysts and cysts devoid of DNA, as previously described (Moreno et al. 2018). Hence, the universal 18S rRNA primers NS1 (5' - GTA GTC ATA TGC TTG TCT C - 3') and NS8 (5' - TCC GCA GGT TCA CCT ACG GA - 3') available from Macrogen Korea were then used of which only samples from Sites 1 and 5 passed the library quantity checking and were subsequently processed for library construction. These failed samples did not pass the QC criteria of a single run which may be due to the shortage of sample volume or amount during the library construction or possibly, the primers used in the library preparation did not work on the samples even though the primer pair EUKAF-EUKAR used has been reported to amplify a wide variety of eukaryotes, including algae (White et al. 1990; Huo et al. 2017; Jo et al. 2020); however, it is more commonly used for fungal identification (Symonová et al. 2013; Aslam et al. 2017; Banos et al. 2018). In addition, it has been described that studies reporting on the performance of commonly utilized universal 18S rRNA are lacking (Khaw et al. 2020). The study of Khaw et al. (2020) found that

using the ss5 and ss3 primer pair is suitable for microalgae identification and was subsequently used in the study. However, only two samples produced sufficient library metagenome amplicon. Because of budget constraints, sending additional samples to try another set of primers that can amplify all samples from the five sites was impossible. Nevertheless, the data generated for Sites 1 and 5 samples using NS1 and NS8 primer pairs were subjected to analysis.

Sequence breakdown and distribution of taxa by domain

The data set of Site 1 (Alabang-Cupang River) contains 112,530 sequences totaling 44,418,975 base pairs with an average length of 395 bps. Zero (0) sequences failed to pass the QC pipeline, with 15,545 (13.81%) unknown and 96,985 (86.19%) predicted features. Distribution of taxa using a contigLCA algorithm shows that 99.90% were eukaryotes and 0.10% were bacteria (Figure 3.A and B). Meanwhile, the data set of Site 5 (Biñan River) contains 139,362 sequences totaling 57,272 base pairs with an average length of 411 bps. One-hundred ten (110) sequences (0.08%) failed to pass the QC pipeline, 10,834 (7.77%) are unknown, and 128,418 (92.15%) are with predicted features. Distribution of taxa using a contigLCA algorithm shows that 127,261 (99.96%) are eukaryotes, and 46 (0.04%) are bacteria (Figure 3.C and 3.D).

Taxonomic distribution of microbial population

Further analysis of the sequences revealed the taxonomic hits distribution of the microbial population up to the genus level on Site 1 (Figure 4) and Site 5 (Figure 5).

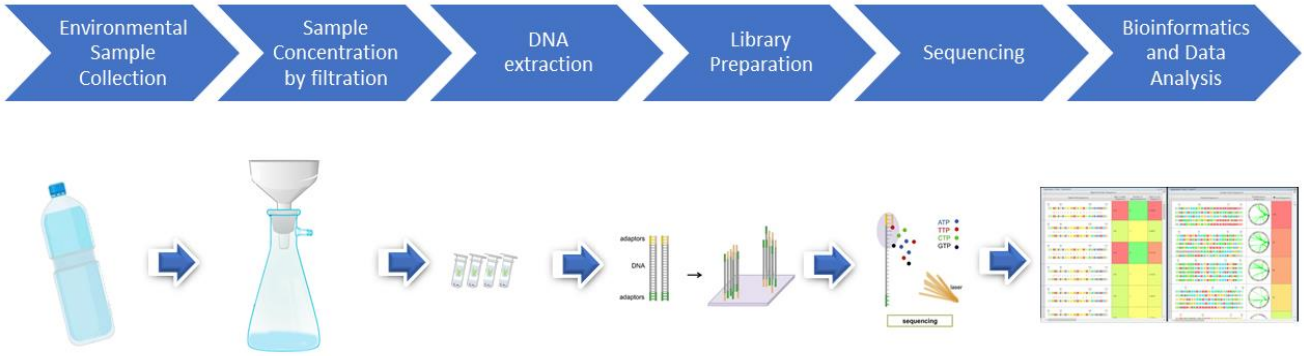

Figure 2. Typical workflow for each NGS approach used in this study

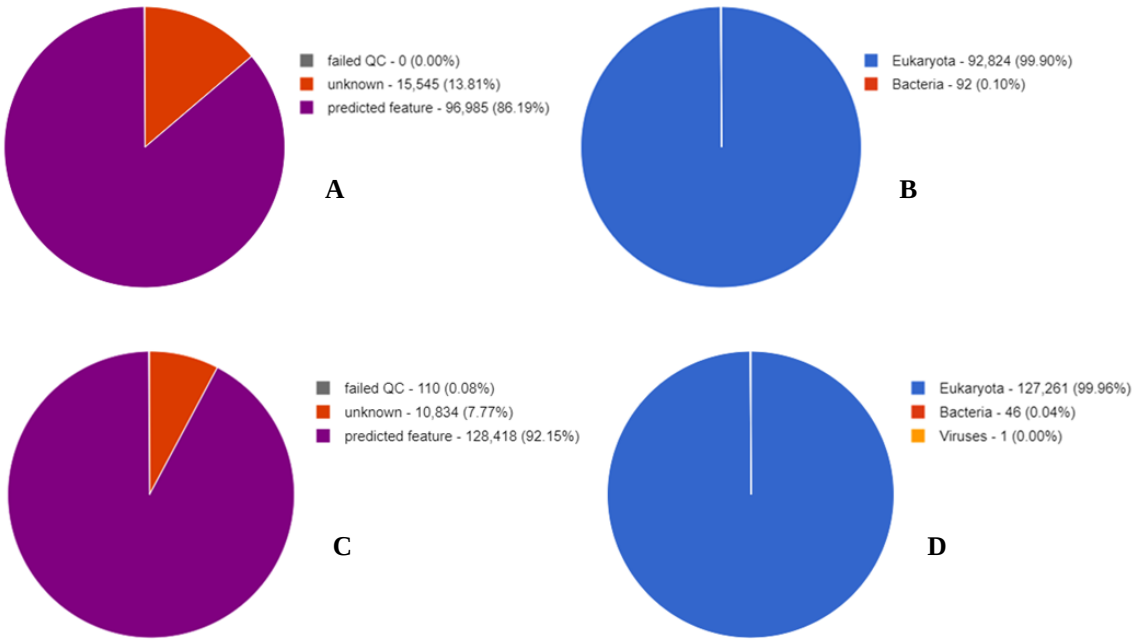

Figure 3. Sequence breakdown in Philippines. Site 1: A. Alabang-Cupang River and B. Distribution of taxa by domain. Sequence breakdown in Site 5: C. Biñan River and D. Distribution of taxa by domain; QC – quality control. *Failed QC means sequences were of low quality; unknown means sequence data not recognized in public databases; predicted feature means sequences have inferred taxonomy

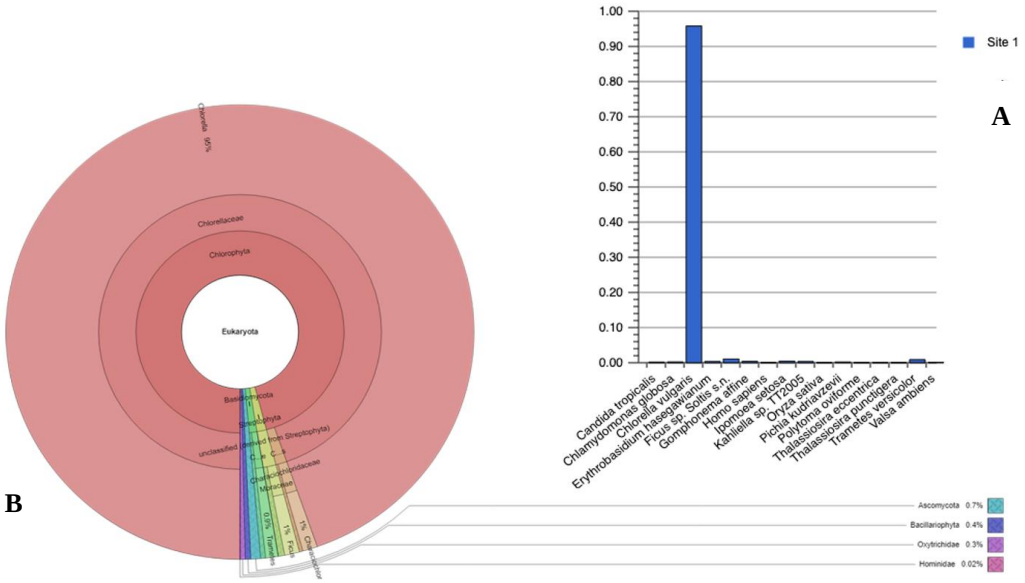

Figure 4. A. Site 1 (Alabang-Cupang River, Philippines) water sample was classified using the MG-RAST software and displayed using the bar chart of MG-RAST. B. The eukaryotic metagenome of Site 1 (Alabang-Cupang River) was imported from MG-RAST and displayed using Krona. Taxonomic nodes are shown as nested sectors arranged from the center on the top level of the hierarchy and progressing outward. The chart is zoomed to place the domain Eukarya at the root. Parameters were set to best hit with an E value of 0.00001, 99% identity, and a length of 301 bp

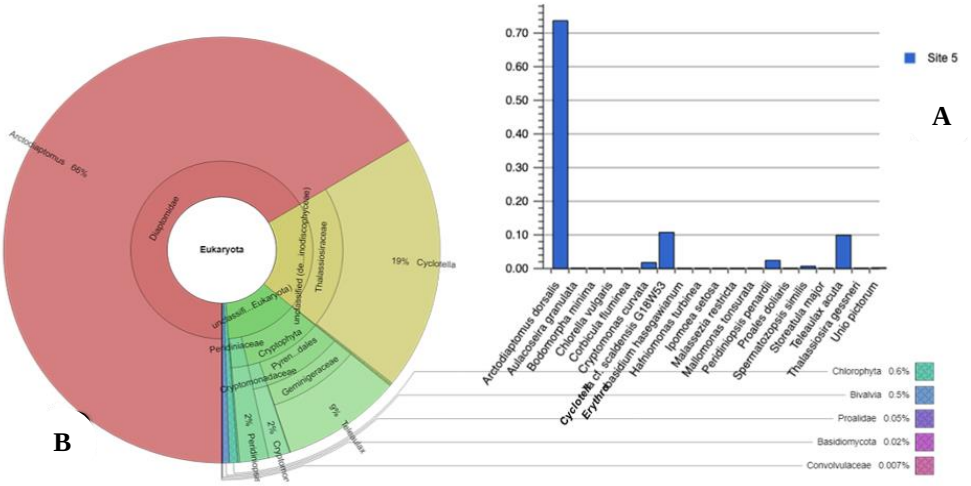

Figure 5. A. The eukaryotic metagenome of Site 5 (Biñan River, Philippines) was imported from MG-RAST and displayed using Krona. B. Taxonomic nodes are shown as nested sectors arranged from the center at the top level of the hierarchy and progressing outward. The chart is zoomed to place the domain Eukarya at the root (B). Parameters were set to best hit with an E value of 0.00001, 99% identity, and a length of 301 bp

Chlorella M.Beijerinck, reads dominated the microbial population in Site 1, which is 95% of the total population in the water sample (Figure 4). While in Site 5, relevant species found were *Cyclotella scaldensis* Muylaert & Sabbe, *Teleaulax acuta* (Butcher) Hill, *Cryptomonas curvata* Ehrenb., and *Peridiniopsis penardii* (Lemmerm.) Bourr. (Figure 5). Yaakob et al. (2021) report that various *Chlorella* species respond significantly to elevated levels of nitrogen and phosphorous as an increase in biomass. The study of Khalil et al. (2021) found Chlorophyta genera in various lakes and dams in Pakistan, which have been deemed probable organic pollution based on the Palmer algal genus index. Cryptophytes persist optimally in eutrophic waters, such as Glinki Lake in Poland (Poniewozik and Lenard 2022). However, they exhibit survival mechanisms in oligotrophic conditions, such as higher cell-specific grazing rates (Saad et al. 2016). The same is observed for diatoms and dinoflagellates, with dominance shifting depending on water nutrient profiles and other factors such as temperature (Spilling et al. 2018;

Xiao et al. 2018). Diatoms have also been observed in mountain lakes of Colombia, where different species were found in different nitrate-phosphorous ratios (Muñoz-López and Rivera-Rondón 2022). Dinoflagellate profiles also differ in several freshwater bodies of Colombia as a result of combined factors of altitude, acidity, temperature, and nutrients, with waters in subtropical regions exhibiting high species richness and biomass associated with low temperatures and high nutrient profiles (Bustamante-Gil et al. 2022).

It is estimated that 60% of the estimated 8.4 million people staying around Laguna Lake discharge their wastes indirectly to the lake through its 21 tributaries, where 40% are from agricultural, 30% are from domestic, and 30% are from industrial sources. This is reflected in the 2009-2017 LLDA data where the sampling sites Tunasan River, San Isidro (San Pedro) River, and Biñan River do not conform to the required Class C water classification and freshwater usage (Department of Transportation 2018).

Because of the increase in the use of water resources and natural processes, surface water degradation has become a worldwide issue. Testing for contaminants in aquatic systems involves biological assays, including a wide range of microalgae species, specifically through cytotoxicity assays. *Chlorella* spp., which belongs to Chlorophyta, is the most abundant genus in some aquatic ecosystems (He et al. 2022; Zhang et al. 2023); *Chlorella* spp. has been identified as a water quality bioindicator due to its abundance in nutrient-enriched aquatic systems (El-Kassas and Gharib 2016; Gökçe 2016; Yan et al. 2020). The use of *Chlorella vulgaris* Beij. has also been documented in toxicity, mutagenicity, and carcinogenicity assays for water reservoirs (Czaplicka-Kotas and Lowdowska 2014; Yan et al. 2020).

The *C. scaldensis* is one of the larger estuarine diatoms under the *Cyclotella meneghiniana* Kützing, 1844 species complex. It differs from the other species, having three satellite pores (instead of two) on the marginal fultoportulae (Muylaert and Sabbe 1996). *Cyclotella* spp. (F.T.Kützing) A.de Brébisson, 1838 is well-represented as one of the pollution-tolerant taxa and, hence, is considered as one of water pollution indicators (Saros et al. 2015; El-Kassas and Gharib 2016) according to the Palmer pollution index (Palmer 1969). Moreover, *Cyclotella* spp. indicates human-induced disturbance of eutrophication (Yang et al. 2020). However, no studies have focused on *Cyclotella*-induced water quality effects, as previous studies have collectively assessed correlations as units of phytoplankton communities (Onyema 2013; El-Kassas and Gharib 2016).

The *T. acuta* and *C. curvata* are cryptomonads commonly found in many aquatic systems, including freshwater, marine, and brackish environments (Kugrens and Clay 2003). Cryptomonads, in general, are also commonly found in deep drinking water reservoir ecosystems (Yan et al. 2020). *Teleaulax* spp. Hill, 1991 has been identified as contributing to red tide events in many countries (Yoo et al. 2017). The *C. curvata* has been associated with high levels of dissolved total phosphorus and dissolved total nitrogen (Xia et al. 2014; Dondajewska et al. 2019).

The dinoflagellates contribute to harmful algal blooms (Carty and Parrow 2015). *Peridiniopsis* spp. is documented to be a causative agent of dinoflagellate blooms in several countries (Satta et al. 2020; You et al. 2015). The *P. penardii* has been recognized as the primary cause of freshwater blooms in Italy (Hansen and Flaim 2007).

The development of molecular biological techniques has benefited society in many ways, including monitoring water quality in natural habitats like lakes, rivers, and seas, mainly where people depend on such surroundings for their livelihood. Moreover, the discovery of detection methods that target nucleic acids has widened our understanding of the microbial world beyond the minority of bacterial taxa cultivable by traditional or older microbiological methods. One of these newly discovered methods is the Next-Generation Sequencing (NGS), which revolutionized DNA sequencing by allowing massively parallel sequencing with millions of reactions running in the same experiment (Vierheilig et al. 2015). The rate and efficiency with which vast volumes of nucleic acid sequence data are acquired have significantly increased because of the intensive usage of high throughput NGS technology. This technology has increased the feasibility and routine implementation of several previously difficult applications. Library-independent methods mainly involved identifying a target gene of a bacterial species or a specific DNA sequence found in human or animal species. These methods could better characterize bacterial communities from environmental waters (Keshri et al. 2015; Unno et al. 2018; Nimnoi and Pongsilp 2020; Garner et al. 2021; Wani et al. 2021).

The LLDA reported the occurrence of an algal bloom in Laguna Lake, which depletes the dissolved oxygen in the area, which may cause fish kill. This phenomenon is considered inevitable owing to climate change in the Philippines, the natural characteristics of the lake, and societal-environmental concerns like illegal dumping of waste, illegal fish pens, and illegal settlers that occupied areas near the shoreland (LLDA 2019). The data generated from this study have shown that organisms associated with harmful algal blooms are present in selected sites of Laguna Lake. However, 18S rRNA metagenomic sequencing has given considerable data a more targeted approach for identifying eukaryotic microorganisms such as *Cryptosporidium* spp. and *Giardia lamblia* (Lambl) Kofoid & Christiansen must be used for future waterborne pathogen assessments.

NGS offers many advantages in comparison to other molecular biological methods. One of the apparent advantages of using NGS is that it has a very high sample throughput (König et al. 2015), which makes it convenient for obtaining information from targeted nucleic acids. The NGS, coupled with other approaches focusing on its use in water monitoring, has been unraveling the functions of microbial communities in many freshwater (Fondi and Liò 2015; Franzosa et al. 2015) and marine environments (Bourlat et al. 2013). Compared to Sanger Sequencing, NGS has a higher sensitivity to detect low-frequency variants (Jamuar et al. 2014), which means that it can detect small populations of microbes even in a large

sample. This is beneficial in assessing potential biological risk factors in the form of specific microbial populations and their genes for virulence, antibiotic resistance, or toxin biosynthesis, which can be crucial in microbial water quality management (Tan et al. 2015).

Although this technique can offer previously unattainable insights into the composition and operation of aquatic microbial communities, NGS is not without its difficulties or drawbacks. For example, NGS sequencing costs are decreasing, and the sequencing speed is increasing. However, the microbiological database is still insufficient, so it is difficult to assemble correctly and bin without sequencing data for reference (Zhang et al. 2021).

On a more positive note, NGS-based studies produce more data with the same amount of input DNA (Dudhagara et al. 2015; Stamps et al. 2018), which gives a more detailed result. Studies by Cai et al. (2014), Greay et al. (2019), and Numberger et al. (2019) using the small subunit ribosomal ribonucleic acid (SSU rRNA) genes, both revealing that, in general, wastewater treatment is capable of removing most microbial populations associated with human-feces. This proves that NGS might be able to assess and monitor the overall human population's health status in the future (Tan et al. 2015).

Yaakob et al. (2021) report that various *Chlorella* species respond significantly to elevated levels of nitrogen and phosphorous as an increase in biomass. The study of Khalil et al. (2021) found Chlorophyta genera in various lakes and dams in Pakistan, which have been deemed probable organic pollution based on the Palmer algal genus index. Cryptophytes persist optimally in eutrophic waters, such as Glinki Lake in Poland (Poniewozik and Lenard 2022). However, they exhibit survival mechanisms in oligotrophic conditions, such as higher cell-specific grazing rates (Saad et al. 2016). The same is observed for diatoms and dinoflagellates, with dominance shifting depending on water nutrient profiles and other factors such as temperature (Spilling et al. 2018; Xiao et al. 2018). Diatoms have also been observed in mountain lakes of Colombia, where different species were found in different nitrate-phosphorous ratios (Muñoz-López and Rivera-Rondón 2022). Dinoflagellate profiles also differ in several freshwater bodies of Colombia as a result of combined factors of altitude, acidity, temperature, and nutrients, with waters in subtropical regions exhibiting high species richness and biomass associated with low temperatures and high nutrient profiles (Bustamante-Gil et al. 2022).

In conclusion, the current study presented data that establishes that Next-Generation Sequencing can determine the diversity of microalgae in an environment, giving a more detailed result with the same amount of input DNA. The microbial composition obtained from the sample sites within Laguna Lake may be used as baseline data for further studies and assessment of the lake's water quality. Additionally, the data gathered will aid policymakers in formulating potential strategies to improve the lake's water quality, considering its multifunctional nature and significance as a water resource for nearby towns.

ACKNOWLEDGEMENTS

This work was financially supported by research grants from the Natural Sciences Research Institute, University of the Philippines, Philippines (Project Code: NSR-19-1-01 and Project Code: NSRI-23-1-05). We appreciate Manila HealthTek, Inc., Philippines for allowing us to use their laboratories for the experiments. We also thank Christian Jordan O. Dela Rosa, Kristine Marie Flores-Destura, and Arriane Joy P. Mejia for their technical assistance.

REFERENCES

Ana KMS, Espino MP. 2020. Occurrence and distribution of hormones and bisphenol A in Laguna Lake, Philippines. Chemosphere 256: 127122. DOI: 10.1016/j.chemosphere.2020.127122.

Aslam S, Tahir A, Aslam MF, Alam MW, Shedayi AA, Sadia S. 2017. Recent advances in molecular techniques for the identification of phytopathogenic fungi – A mini review. J Plant Interact 12 (1): 493-504. DOI: 10.1080/17429145.2017.1397205.

Banos S, Lentendu G, Kopf A, Wubet T, Glöckner FO, Reich M. 2018. A comprehensive fungi-specific 18S rRNA gene sequence primer toolkit suited for diverse research issues and sequencing platforms. BMC Microbiol 18: 190. DOI: 10.1186/s12866-018-1331-4.

Bourlat SJ, Borja A, Gilbert J, Taylor MI, Davies N, Weisenber SB, Griffith JF, Lettieri T, Field D, Benzie J, Glöckner FO, Rodríguez-ezpeleta N, Faith DP, Bean TP, Obst M. 2013. Genomics in marine monitoring: New opportunities or assessing marine health status. Mar Pollut Bull 74 (1): 19-31. DOI: 10.1016/j.marpolbul.2013.05.042.

Bustamante-Gil C, Amat E, Boltovskoy A, Ramírez-Restrepo JJ. 2022. The first floristic study of freshwater dinoflagellates (Dinophyceae) in Colombia. J Limnol 81 (2022). DOI: 10.4081/jlimnol.2021.2023.

Caballero I, Navarro G. 2021. Monitoring cyanoHABs and water quality in Laguna Lake (Philippines) with Sentinel-2 satellites during the 2020 Pacific typhoon season. Sci Total Environ 788: 147700. DOI: 10.1016/j.scitotenv.2021.147700.

Cai L, Ju F, Zhang T. 2014. Tracking human sewage microbiome in a municipal wastewater treatment plant. Appl Microbiol Biotechnol 98: 3317-3326. DOI: 10.1007/s00253-013-5402-z.

Carty S, Parrow MW. 2015. Dinoflagellates. In: Wehr JD, Sheath RG (eds). Freshwater Algae of North America (2nd ed.). Academic Press, New York. DOI: 10.1016/b978-0-12-385876-4.00017-7.

Cox MP, Peterson DA, Biggs PJ. 2010. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. BMC Bioinform 11 (1): 1-6. DOI: 10.1186/1471-2105-11-485.

Cui Q, Fang, T, Huang Y, Dong P, Wang H. 2016. Evaluation of bacterial pathogen diversity, abundances and health risks in urban recreational water by amplicon Next-Generation Sequencing and quantitative PCR. J Environ Sci 57: 137-149. DOI: 10.1016/j.jes.2016.11.008.

Czaplicka-Kotas A, Lodowska J. 2014. Biomonitoring of surface water by synchronous culture of *Chlorella vulgaris* algae. Environ Protect Eng 40 (4): 29-40. DOI: 10.5277/epe140403.

Department of Environment and Natural Resources - Environmental Management Bureau. 2014. National Water Quality Status Report 2006-2013. http://water.emb.gov.ph/wpcontent/uploads/2016/06/NWQSR2006-2013.pdf

Department of Transportation Philippines. 2018. Environmental Impact Assessment: Proposed Multitranche Financing Facility Republic of the Philippines: Malolos-Clark Railway Project (Solis to Blumentritt Section of North South railway project. https://www.adb.org/sites/default/files/project-documents/52083/52083-001-eia-en_4.pdf

Dondajewska R, Gołdyn R, Messyasz B, Kowalczewska-Madura K, Cerbin S. 2019. A shallow lake in an agricultural landscape–water quality, nutrient loads, future management. Limnol Rev 19 (1): 25-35. DOI: 10.2478/limre-2019-0003.

Dudhagara P, Ghelani A, Bhavsar S, Bhatt S. 2015. Metagenomic data of fungal internal transcribed spacer and 18S rRNA gene sequences from Lonar Lake sediment, India. Data Br 4: 266-268. DOI: 10.1016/j.dib.2015.06.001.

El-Kassas HY, Gharib SM. 2016. Phytoplankton abundance and structure as indicator of water quality in the drainage system of the Burullus Lagoon, Southern Mediterranean coast, Egypt. Environ Monit Assess 188: 1-14. DOI: 10.1007/s10661-016-5525-7.

Fondi M, Liò P. 2015. Multi-omics and metabolic modelling pipelines: Challenges and tools for systems microbiology. Microbiol Res 171: 52-64. DOI: 10.1016/j.micres.2015.01.003.

Franzosa EA, Hsu T, Sirota-Madi A, Shafquat A, Abu-Ali G, Morgan XC, Huttenhower C. 2015. Sequencing and beyond: Integrating molecular 'omics' for microbial community profiling. Nat Rev Microbiol 13 (6): 360-372. DOI: 10.1038/nrmicro3451.

Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: Accelerated for clustering the Next-Generation Sequencing data. Bioinformatics 28 (23): 3150-3152. DOI: 10.1093/bioinformatics/bts565.

Garner E, Davis BC, Milligan E, Blair MF, Keenum I, Maile-Moskowitz A, Pan J, Gnegy M, Liguori K, Gupta S, Prussin AJ, Marr LC, Heath LS, Vikesland PJ, Zhang L, Pruden A. 2021. Next-Generation Sequencing approaches to evaluate water and wastewater quality. Water Res 194: 116907. DOI: 10.1016/j.watres.2021.116907.

Goss MRM, Rivera WL, Torreta NK. 2020. Methods in Microalgal Studies. Philippine Science Letters, Laguna.

Greay TL, Gofton AW, Zahedi A, Paparini A, Linge KL, Joll CA, Ryan UM. 2019. Evaluation of 16S Next-Generation Sequencing of hypervariable region 4 in wastewater samples: An unsuitable approach for bacterial enteric pathogen identification. Sci Total Environ 670: 1111-1124. DOI: 10.1016/j.scitotenv.2019.03.278.

Gökçe D. 2016. Algae as an indicator of water quality. Algae - Organisms for Imminent Biotechnology: 81-101. InTech, London. DOI: 10.5772/62916.

Hansen G, Flaim G. 2007. Dinoflagellates of the Trentino Province, Italy. J Limnol 66 (2): 107-141. DOI: 10.4081/jlimnol.2007.107.

He L, Wang Q, Wang Z, Wang F, Sun S, Liu, X. 2022. Eukaryotic diversity based on high-throughput 18S rRNA sequencing and its relationship with environmental factors in a Salt Lake in Tibet, China. Water 14 (17): 2724. DOI: 10.3390/w14172724.

Hernandez EC, Henderson A, Oliver DP. 2012. Effects of changing land use in the Pagsanjan-Lumban catchment on suspended sediment loads to Laguna de Bay, Philippines. Agric Water Manag 106: 8-16. DOI: 10.1016/j.agwat.2011.08.012.

Huo S, Shang C, Wang Z, Zhou W, Cui F, Zhu F, Dong R. 2017. Outdoor growth characterization of an unknown microalga screened from contaminated *Chlorella* culture. Biomed Res Intl 2017: 5681617. DOI: 10.1155/2017/5681617.

Jamuar SS, Lam ATN, Kircher M, D'Gama AM, Wang J, Barry BJ, Walsh CA. 2014. Somatic mutations in cerebral cortical malformations. N Engl J Med 371 (8): 733-743. DOI: 10.1056/NEJMoa1314432.

Jo SW, Do JM, Kang NS, Park JM, Lee JH, Kim HS, Yoon HS. 2020. Isolation, identification, and biochemical characteristics of a cold-tolerant *Chlorella vulgaris* KNUA007 isolated from King George Island, Antarctica. J Mar Sci Eng 8 (11): 935. DOI: 10.3390/jmse8110935.

Keshri J, Mankazana BB, Momba MN. 2015. Profile of bacterial communities in South African mine-water samples using Illumina Next-Generation Sequencing platform. Appl Microbiol Biotechnol 99: 3233-3242. DOI: 10.1007/s00253-014-6213-6.

Khalil S, Mahnashi MH, Hussain M, Zafar N, Waqar-Un-Nisa, Khan FS, Afzal U, Shah GM, Niazi UM, Awais M, Irfan M. 2021. Exploration and determination of algal role as bioindicator to evaluate water quality - probing fresh water algae. Saudi J Biol Sci 28 (10): 5728-5737. DOI: 10.1016/j.sjbs.2021.06.004.

Khaw YS, Khong NMH, Shaharuddin NA, Yusoff FM. 2020. A simple 18S rDNA approach for the identification of cultured eukaryotic microalgae with an emphasis on primers. J Microbiol Methods 172: 105890. DOI: 10.1016/j.mimet.2020.105890.

Kugrens P, Clay BL. 2003. Cryptomonads. In: Wehr JD, Sheath RG, Kociolek JP (eds). Freshwater Algae of North America: Ecology and Classification. Academic Press, Massachusetts. DOI: 10.1016/b978-012741550-5/50022-2.

König K, Peifer M, Fassunke J, Ihle MA, Künstlinger H, Heydt C, Heukamp LC. 2015. Implementation of amplicon parallel sequencing leads to improvement of diagnosis and therapy of lung cancer patients. J Thorac Oncol 10 (7): 1049-1057.

Laguna Lake Development Authority (LLDA). 2019. Annual Report, 2019. LLDA, Laguna.

Laguna Lake Development Authority (LLDA). 2015. Laguna de Bay Basin Master Plan: 2016 and Beyond (Towards Climate-Resiliency and Sustainable Development). http://gwhs-stg03.i.gov.ph/~s3lldagov/wp-content/dox/ldbMP2016.pdf

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10 (3): 1-10. DOI: 10.1186/gb-2009-10-3-r25.

Meyer F, Bagchi S, Chaterji S, Gerlach W, Grama A, Harrison T, Wilke A. 2019. MG-RAST version 4 - Lessons learned from a decade of low-budget ultra-high-throughput metagenome analysis. Brief Bioinform 20 (4): 1151-1159. DOI: 10.1093/bib/bbx105.

Moreno Y, Moreno-Mesonero L, Amorós I, Pérez R, Morillo JA, Alonso JL. 2018. Multiple identification of most important waterborne protozoa in surface water used for irrigation purposes by 18S rRNA amplicon-based metagenomics. Intl J Hyg Environ Health 221 (1): 102-111.DOI: 10.1016/j.ijheh.2017.10.008.

Muñoz-López CL, Rivera-Rondón CA. 2022. Diatom response to environmental gradients in the high mountain lakes of the Colombia's Eastern Range. Aquat Sci 84: 15. DOI: 10.1007/s00027-021-00838-z.

Muylaert K, Sabbe K. 1996. *Cyclotella scaldensis* spec. nov. (Bacillariophyceae), a new estuarine diatom. Nova Hedwig 63 (3): 335-346.

Nimnoi P, Pongsilp N. 2020. Marine bacterial communities in the upper gulf of Thailand assessed by Illumina Next-Generation Sequencing platform. BMC Microbiol 20: 1-11. DOI: 10.1186/s12866-020-1701-6.

Numberger D, Ganzert L, Zoccarato L, Mühldorfer K, Sauer S, Grossart HP, Greenwood AD. 2019. Characterization of bacterial communities in wastewater with enhanced taxonomic resolution by full-length 16S rRNA sequencing. Sci Rep 9 (1): 9673. DOI: 10.1038/s41598-019-46015-z.

Ondov BD, Bergman NH, Phillippy AM. 2011. Interactive metagenomic visualization in a web browser. BMC Bioinform 12 (1): 1-10. DOI: 10.1186/1471-2105-12-385.

Onyema IC. 2013. Phytoplankton bio-indicators of water quality situations in the Iyagbe Lagoon, South-western Nigeria. Res J Pharm Biol Chem Sci 4 (3): 639-652.

Palmer CM. 1969. A composite rating of algae tolerating organic pollution. J Phycol 5: 78-82. DOI: 10.1111/j.1529-8817.1969.tb02581.x.

Poniewozik M, Lenard T. 2022. Phytoplankton composition and ecological status of lakes with Cyanobacteria dominance. Intl J Environ Res Public Health 19 (7): 3832. DOI: 10.3390/ijerph19073832.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Res 41 (D1): D590-D596. DOI: 10.1093/nar/gks1219.

Rho M, Tang H, Ye Y. 2010. FragGeneScan: Predicting genes in short and error-prone reads. Nucleic Acids Res 38 (20): e191-e191. DOI: 10.1093/nar/gkq747.

Saad JF, Unrein F, Tribelli PM, López N, Izaguirre I. 2016. Influence of lake trophic conditions on the dominant mixotrophic algal assemblages. J Plankton Res 38 (4): 818-829. DOI: 10.1093/plankt/fbw029.

Saguin K. 2014. Biographies of fish for the city: Urban metabolism of Laguna Lake aquaculture. Geoforum 54: 28-38. DOI: 10.1016/j.geoforum.2014.03.008.

Saros JE, Anderson N. 2015. The ecology of the planktonic diatom *Cyclotella* and its implications for global environmental change studies. Biol Rev 90 (2): 522-541. DOI: 10.1111/brv.12120.

Stamps BW, Nunn HS, Petryshyn VA, Oremland RS, Miller LG, Rosen MR, Spear JR. 2018. Metabolic capability and phylogenetic diversity of Mono Lake during a bloom of the eukaryotic phototroph *Picocystis* sp. strain ML. Appl Environ Microbiol 84 (21): e01171-18. DOI: 10.1128/aem.01171-18.

Satta CT, Reñé A, Padedda BM, Pulina S, Lai GG, Soru O, Lugliè A. 2020. First detection of the bloom forming *Unruhdinium penardii* (Dinophyceae) in a Mediterranean reservoir: Insights on its ecology, morphology and genetics. Adv Oceanogr Limnol 11 (2). DOI: 10.4081/aiol.2020.9500.

Shrestha RG, Tanaka Y, Malla B, Bhandari D, Tandukar S. 2017. Next-Generation Sequencing identification of pathogenic bacterial genes and their relationship with fecal indicator bacteria in different water sources in the Kathmandu Valley, Nepal. Sci Total Environ 601-602: 278-284.DOI: 10.1016/j.scitotenv.2017.05.105.

Spilling K, Olli K, Lehtoranta J, Kremp A, Tedesco L, Tamelander T, Klais R, Peltonen H, Tamminen T. 2018. Shifting diatom -

Dinoflagellate dominance during spring bloom in the Baltic Sea and its potential effects on biogeochemical cycling. Front Mar Sci 5: 327. DOI: 10.3389/fmars.2018.00327.

Symonová R, Majtánová Z, Sember A, Staaks GB, Bohlen J, Freyhof J, Ráb P. 2013. Genome differentiation in a species pair of coregonine fishes: An extremely rapid speciation driven by stress-activated retrotransposons mediating extensive ribosomal DNA multiplications. BMC Evol Biol 13: 1-11. DOI: 10.1186/1471-2148-13-42.

Tan B, Ng C, Nshimyimana JP, Loh LL, Gin KYH, Thompson, JR. 2015. Next-Generation Sequencing (NGS) for assessment of microbial water quality: Current progress, challenges, and future opportunities. Front Microbiol 6: 1027. DOI: 10.3389/fmicb.2015.01027.

Tayaban KMM, Pintor KL, Vital PG. 2018. Detection of potential harmful algal bloom-causing microalgae from freshwater prawn farms in Central Luzon, Philippines, for bloom monitoring and prediction. Environ Dev Sustain 20 (3): 1311-1328. DOI: 10.1007/s10668-017-9942-8.

Unno T, Staley C, Brown CM, Han D, Sadowsky MJ, Hur HG. 2018. Fecal pollution: New trends and challenges in microbial source tracking using next-generation sequencing. Environ Microbiol 20 (9): 3132-3140.

Vierheilig J, Savio D, Ley RE, Mach RL, Farnleitner AH, Reischer GH. 2015. Potential applications of next generation DNA sequencing of 16S rRNA gene amplicons in microbial water quality monitoring. Water Sci Technol 72 (11): 1962-1972. DOI: 10.2166/wst.2015.407.

Villaruel MJS, Camacho MVDC. 2022. Phytoplankton community dynamics in Tadlac Lake, Los Baños, Laguna, Philippines. J Fish Environ 46 (3): 55-71.

Wani GA, Khan MA, Dar MA, Shah MA, Reshi ZA. 2021. Next generation high throughput sequencing to assess microbial communities: An application based on water quality. Bull Environ Contam Toxicol 106: 727-733. DOI: 10.1007/s00128-021-03195-7.

White TJ, Bruns T, Lee SJ, Taylor J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. PCR Protocols: A Guide to Methods and Applications: 315-322. Academic Press, New York. DOI: 10.1016/B978-0-12-372180-8.50042-1.

Xia S, Cheng Y, Zhu H, Liu G, Hu Z. 2014. Distribution and population dynamics of cryptomonads in a Chinese lake with three basins varying in their trophic state. Phycol Res 63 (2): 93-101. DOI: 10.1111/pre.12079.

Xiao W, Liu X, Irwin A, Laws E, Wang L, Chen B, Zeng Y, Huang B. 2018. Warming and eutrophication combine to restructure diatoms and dinoflagellates. Water Res 128: 206-216. DOI: 10.1016/j.watres.2017.10.051.

Yaakob MA, Mohamed RMSR, Al-Gheethi A, Aswathnarayana RG, Ambati RR. 2021. Influence of nitrogen and phosphorus on microalgal growth, biomass, lipid, and fatty acid production: An overview. Cells 10 (2): 393. DOI: 10.3390/cells10020393.

Yan M, Chen S, Huang T, Li B, Li N, Liu K, Huang X. 2020. Community compositions of phytoplankton and eukaryotes during the mixing periods of a drinking water reservoir: Dynamics and interactions. Intl J Environ Res Public Health 17 (4): 1128. DOI: 10.3390/ijerph17041128.

Yang J, Ji Y, Yan R, Liu X, Zhang J, Wu N, Wang K. 2020. Applicability of benthic diatom indices combined with water quality valuation for dish lake from Nanjishan Nature Reserve, Lake Poyang. Water 12 (10): 2732. DOI: 10.3390/w12102732.

You X, Luo Z, Su Y, Gu L, Gu H. 2015. *Peridiniopsis jiulongensis*, a new freshwater dinoflagellate with a diatom endosymbiont from China. Nova Hedwig 101: 313-326. DOI: 10.1127/nova_hedwigia/2015/0272.

Yoo YD, Seong KA, Jeong HJ, Yih W, Rho JR, Nam SW, Kim, HS. 2017. Mixotrophy in the marine red-tide cryptophyte *Teleaulax amphioxeia* and ingestion and grazing impact of cryptophytes on natural populations of bacteria in Korean coastal waters. Harmful Algae 68: 105-117. DOI: 10.1016/j.hal.2017.07.012.

Zhang H, Yang Y, Liu X, Huang T, Ma B, Li N, Yang W, Li H, Zhao K. 2023. Novel insights in seasonal dynamics and co-existence patterns of phytoplankton and micro-eukaryotes in drinking water reservoir, Northwest China: DNA data and ecological model. Sci Total Environ 857: 159160. DOI: 10.1016/j.scitotenv.2022.159160.

Zhang L, Chen FX, Zeng Z, Xu M, Sun F, Yang L, Bi X, Lin Y, Gao YJ, Hao HX, Yi W, Li M, Xie Y. 2021. Advances in metagenomics and its application in environmental microorganisms. Front Microbiol 12 (December): 1-15. DOI: 10.3389/fmicb.2021.766364.