Uncovering the Competences of Commonly Used Nutrient Media, Multi-primer Approach and Reference Databases in the Identification of Microalgae Diversity from Environmental Samples

Amal A. Badr

The American University in Cairo

Walid M. Fouad (✉ wfouad@aucegypt.edu)

The American University in Cairo

Research Article

Keywords: BG-11, BBM, Modified media (MM), half-strength Murashige, Skoog medium (MS)

Posted Date: September 1st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-855064/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Microalgae are highly diverse microorganisms and have a variety of benefits and use across different fields. On the other hand, their overgrowth can be extremely dangerous to our environment, thus, making it particularly important to continuously manage and track their abundance and diversity to oversee any potential of extinction or overgrowth. The vast diversity of microalgae imposes the challenge of their identification through the most common and economical identification method, morphological identification, and the more recent molecular-level identification tools. To enhance the identification of microalgae, we targeted enrichment of total microalgae diversity present in an environmental sample using four different enrichment media (BG-11, BBM, Modified media (MM), and half-strength Murashige and Skoog medium (MS)). Morphological identification of the enriched microalgae diversity was conducted every 4-days to monitor the population dynamics. After 14-day the DNA was extracted from the enriched population for molecular-level identification using 16S rRNA gene regions V1-V3 and V4-V5 and 18 rRNA gene V4 region. To further enhance microalgae identification through molecular-level identification, we evaluated three reference databases (SILVA, Greengenes, and Protist Ribosomal Reference (PR²)) to reveal their competence in microalgae diversity identification. A total of 38 microalgae were identified morphologically to the genus level, and the highest number of microalgae were identified through MM media (36), followed by BG-11 and BBM, 31 and 26, respectively. While sequencing the three-primer sets using the three databases, 87 microalgae were identified to the genus level. The highest diversity was identified using the MM media (71 genera) followed by BG-11 (69 genera), BBM (67 genera). Our multiple-media, primer, and reference database approach enabled us to identify a high microalgae diversity that would have been missed if a single approach was used over the other.

Introduction

Microalgae widespread arises mainly due to their aptitude to adapt and survive in different aquatic and terrestrial environments ¹. Their rich diversity provides us with various benefits that include but are not limited to the following: pigments (carotenoids), antioxidants, toxins, and sterols ^{2,3}; act as environmental bioindicators for aquatic ecosystems management. Since certain algae species in an ecosystem are directly correlated to the nutrient availability and concentration present in that habitat ^{4,5}, it is essential to constantly study the algal diversity present in various water resources.

One of the approaches to identifying microalgae from environmental samples is the traditional morphological identification method. This can be done directly by analysing the sample microscopically; however, the drawback of such an approach is the limiting detection threshold of microalgae present in a collected sample. An alternative approach is the enrichment of the environmental sample prior to morphological identification. The commonly used media for microalgae culture in the literature include Blue-Green media (BG-11) and Bold's Basal media (BBM), which have been widely used to identify and maintain species diversity as it can sustain a wide range of species ⁶.

Another means of microalgae identification is through the molecular level analysis of conserved regions of the DNA, such as 16S and 18S ribosomal RNA (rRNA) genes used for the profiling of prokaryotic and eukaryotic microalgae, respectively. The reference database to be used to align the sequenced results is of equal importance to the DNA barcodes sequenced. Hence, it is crucial to identify the most suitable database to be used for microalgae identification. The SILVA database is one of the predominantly used databases and is based on the small sub-unit (SSU) rRNAs that include the 16S and 18S rRNA in prokaryotes and eukaryotes, respectively. Therefore, they cover the three domains of life: Bacteria, Archaea and Eukarya. The SILVA database can assign taxonomic ranks to the genus level ⁷. Greengenes (GG) database is another database that constitutes bacterial/archaeal 16S rRNA gene. The GG database is constructed from public databases that are aligned and checked for chimera ^{8,9}). At the same time, the Protist Ribosomal Reference (PR²) database contains sequences of the SSU rRNA and rDNA of the kingdom Eukaryota. The PR² database is used to identify eukaryotes through nuclear-encoded sequences, the 16S rRNA, present in the mitochondria or plastid of eukaryotic organisms ¹⁰. Both GG and PR² provide taxonomic annotation down to the species level.

Several studies have relied on selective enrichment or enrichment using a single common nutrient media to improve microalgae identification. The influence of culture media on microalgae population growth and diversity in the literature remains elusive; however, still utilised ^{11–13}. There has been a lack in investigating the capabilities of these different enrichment media in the enrichment of certain microalgae. Additionally, the capabilities/limitation of the molecular identification approach regarding the use of different primer sets and reference databases have not been extensively studied for microalgae. In contrast, similar approaches have been extensively studied for bacteria ^{14–18}. We have previously investigated the efficiency of the commonly used BG-11 and BBM in microalgae enrichment for diversity identification and the effect of vitamin enrichment and reduced media nutrient on diversity enrichment. The results obtained using the reduced media provided us with an insight into the limitation of both media and the need to investigate other media compositions ¹⁹. Hence, the current study investigates the effect of a modified media (MM) that is considered our intermediate media and a high nitrogen-rich media and half-strength Murashige & Skoog Basal Salt Mixture (MS) on the enrichment of microalgae diversity from an environmental sample. Therefore, this study aims to improve the detection and quantification of microalgae through (1) increasing algal abundance above the detection threshold using four different enrichment media and; (2) the utilisation of multi-primer sets of small subunit (SSU) rRNAs (16S and 18S rRNA) and; (3) using different reference databases (SILVA, Greengenes (GG), and Protist Ribosomal Reference database (PR²) to evaluate their competence in the identification of microalgae.

Results

Overview of the freshwater microalgae taxa identified across different nutrient media

The morphological assessment of the samples using the four different nutrient media aided in identifying 38 microalgae to the genus level. The highest diversity was belonging to *Chlorella*, *Scenedesmus*, *Ankistrodesmus*, and *Selenastrum* species. Figure 1A shows that the highest diversity was identified when the sample was cultured on MM media (36 genera), followed by BG-11 (31 genera), BBM (26 genera), and MS media (20 genera). The half-strength MS media had the lowest identification, and all the genera identified through the MS media have been commonly identified through the other media.

Using the three primer sets to analyse the samples and sequencing through the Illumina HiSeq platform and using the three databases enabled us to identify 87 microalgae to the genus level. Similar to the results of the morphological analysis, the highest diversity was identified using the MM media (71 genera) as a sample enrichment media followed by BG-11 (69 genera), BBM (67 genera), and MS media (45 genera) (Fig. 1B). However, through the molecular-level identification and unlike the analysis results conducted through the morphological features of microalgae, only the half-strength MS media enriched and identified *Raphidonema*.

Combining the two-identification method, molecular and morphological, and the enrichment of the samples on the four-enrichment media enabled us to uncover a total of 104 microalgae to the genus level (Fig. 1C). The highest diversity was identified using the three primer sets sequencing sections of the hypervariable regions of 16S and 18S rRNA genes. When looking at the morphological and molecular level analysis results side by side, 17 genera have been uniquely identified only through the morphological level analysis and not through the sequencing of the hypervariable regions (Fig. 1C). While the molecular-based approach uniquely identified 67 microalgae (Fig. 1C).

Taxonomic composition and diversity comparison from different databases

The diversity of the microalgae cultured on the four media was analysed through two regions of the 16S rRNA gene and one region in the 18S rRNA gene. In total, the two 16S rRNA regions sequenced (V1-V3 and V4-V5) combined had a total of 852,517 sequences. In comparison, the 18S rRNA V4 region had a total of 423,623 sequences. After the filtering and processing, the two 16S rRNA regions combined had 271,511 unique sequences, and the 18S rRNA region had 344,887 sequences. The 16S rRNA OTUs assigned by aligned them to three reference databases: SILVA, Greengenes (GG), and the Protist Ribosomal Reference (PR²). The 18S rRNA sequences were aligned SILVA database.

A total of 5298 taxa were identified using the 16S V1-V3 and SILVA as a reference database, while using the GG and the PR² database, 5498 and 5942 taxa were classified, respectively. Moreover, 397 taxa were classified using the 18S rRNA through the SILVA database. Further filtering using RStudio packages was performed to eliminate unclassified phyla to focus on the classified microalgae phyla/genera.

Cyanobacteria Identification

Through the 16S rRNA genes, 19 Cyanobacteria genera have been revealed. The primer sets used for the 16S rRNA hypervariable regions V1-V3 and V4-V5 and three databases aided in identifying 18 and 17 genera, respectively (Fig. 2A & B). Through the analysis of the two SSU 16S rRNA gene regions sequenced, the SILVA database aided in identifying higher diversity of cyanobacteria over the other tested databases (Fig. 3A & D) and the two 16S rRNA gene regions identified the same 14 cyanobacteria genera. Yet, certain cyanobacteria were only identified through GG and PR² databases only. The 16S rRNA genes aligned to the GG database identified 11 genera through the V1-V3 region, and V4-V5 identified 10 genera (Fig. 3B & E), respectively. Using the GG database, two genera of cyanobacteria, *Aphanizomenon* and *Synechococcus*, were identified through the V1-V3 and V4-V5 regions of the 16S rRNA gene and were not identified using SILVA. At the same time, *Planktothrix* was picked up by the V1-V3 and not the V4-V5 primers in the GG databases. We did not expect much from the PR² database to identify cyanobacteria (Fig. 3C & F); however, two cyanobacteria genera were only identified through the PR² database. Using the V1-V3 region, *Gloeothoece* was identified in the BG-11 media, and the V4-V5 region revealed *Nostoc* in the nutrient media BG-11, BBM, and MM. Nonetheless, we can consider that using the primers to the V1-V3 and V4-V5 regions helped uncover more cyanobacteria through SILVA and GG databases.

Eukaryotic Microalgae Identification

Both the 16S and the 18S rRNA sequenced regions were used to identify a total of 68 eukaryotic microalgae (Fig. 2C). The SILVA database and the 18S rRNA V4 region revealed the highest diversity of 43 eukaryotic microalgae, with 27 genera belonging to the phylum Chlorophyta and 12 genera belonging to the Ochrophyta. The remaining identified genera belonging to the phylum Streptophyta (3 genera) and Cryptophyta (1 genus) (Fig. 4A-D). The PR² database and the two 16S rRNA genes sequenced together resulted in the classification of 37 eukaryotic microalgae genera. The 16S V1-V3 fragment divulged 28 genera, out of which 10 genera were uniquely identified by the PR² database and the V1-V3 fragment. In contrast, the V4-V5 fragment alignment identified 22 microalgae genera (Fig. 4E-H). Using only the SILVA database and the 18S rRNA V4, 25 eukaryotic microalgae would have been missed from the study as these genera were only revealed through the sequencing of the V1-V3 and the V4-V5 regions of the 16S rRNA gene and annotated using the PR².

Alpha Diversity Indices of Total Microbial Community based on OTUs

The alpha diversity was calculated for each location and based on total phyla identified through each region. The 16S V1-V3 region through the three databases had similar Good's coverage index ranging from 94–95% across the samples using GG database while PR² coverage index ranging from 91–93% (Fig. 5A-C). Nevertheless, the observed richness (Sobs) in terms of OTUs was the highest in the Nile2 sample, followed by the Nile3 and Nile1 samples across the three databases (Fig. 5A-C). However, the V4-V5 region of the 16S rDNA had a higher Good's coverage (< 96%) when calculated across the three databases. Furthermore, the Sobs is high when analysed through the V4-V5 region over the V1-V3 region.

Nile2 still had the highest Sobs, followed by Nile1 than Nile3 using either of the 16S regions (Fig. 5D-F). The Shannon index for Nile2 was the highest amongst the other Nile samples when analysing the sequences for the V4-V5 region (Fig. 5D-F), while the higher Shannon index was observed in Nile3 when examining the V1-V3 region (Fig. 5A-C), followed by Nile2 and Nile1 samples.

The high InvSimpson index reflects the increase in diversity. The InvSimpson values increase is observed across Nile3 samples across all the sequenced regions, followed by Nile2 and Nile1 samples. However, the 16S V4-V5 and the 18S V4 of Nile1 and Nile2 have a similar median of InvSimpson (Fig. 5D – F & G). The final alpha diversity index evaluated regarding the OTUs abundance is the Berger Parker index. The decrease in this index indicates an increase in diversity. The Nile3 sample has the lowest Berger Parker index indicated the presence of high diversity followed by Nile2 and Nile1 across all the databases and regions (Fig. 5). This is in line with the other two calculated indices, Shannon and InvSimpson, for the three locations. The 18S V4 region for eukaryotic microorganisms analysed through the SILVA database had a Good's coverage > 99%, while the highest Sobs was observed in the Nile1 sample followed by Nile3 and Nile2. However, Shannon and InvSimpson index across all the samples were similar Nile1 (1.81, 3.75), Nile2 (1.73, 3.33), and (1.74, 3.69) (Shannon, InvSimpson, respectively) (Fig. 5G). The Berger Parker index was also similar for all the three samples 0.45, 0.49, and 0.48 for Nile1, Nile2, and Nile3 samples.

The relationship between identified phyla, media, and location was examined through the Bray-Curtis similarity index based on the microbial composition identified through SILVA 18S V4 results as it identified the highest overall diversity. An unexpected clustering pattern was observed, and samples were clustered with samples enriched on the same media rather than being clustered by their corresponding location (Fig. 6).

Discussion

This study aims to fully evaluate the culture-based morphological identification method and the multi-primer/multi-database molecular approach in identifying microalgae diversity from an enriched environmental sample, in our case, the River Nile. The challenge presented was to identify the capability of the currently used media in the enrichment of the diversity to identify culturable microalgae. Additionally, pursuing the possibility of improving the media for better enrichment of microalgae population. This could be through finding a media that would provide an evenness to the present diversity that would enable the enrichment of rare species in the community. Growth analysis studied confirmed that none of the media hindered the growth of the microalgae (data not shown).

Through morphological identification of microalgae in the four tested media (BG-11, BBM, MM and MS) used to enrich the River Nile environmental samples, a total of 37 microalgae have been identified across the three samples combined. MM media alone identified 35 microalgae, while the commonly used media BG-11 and BBM identified 30 and 27 microalgae, respectively. Through a closer look at the four tested nutrient media used, each media had all the necessary macronutrients and micronutrients (such as N, P, K, Fe, Cu, Mn, Zn, Co and Mo) that would enhance microalgae and not hinder its growth. However, each

nutrient media had different macro and micronutrient composition/concentrations. Differences in media components may have caused the favoring growth of certain microalgae over the other. One of the main differences in media composition is the source of nitrogen. BG-11, BBM, and MM media contain sodium nitrate as the main source of nitrogen. While MS media, containing the highest nitrogen content, has nitrogen in two forms, ammonium nitrate, and potassium nitrate. Microalgae can use nitrogen in different forms such as nitrate, nitrite, ammonium, and urea, an organic nitrogen source. The different nitrogen sources are first reduced to ammonium which is assimilated into amino acids through different pathways in microalgae²⁰. While ammonium is a preferred nitrogen source due to its assimilation being a low energy process, sodium and potassium nitrates are still preferred in media over ammonium. This is attributed to ammonium ions (NH_4^+) being converted to ammonia gas (NH_3) under aeration at alkaline conditions and thus considered a loss of nitrogen source²¹. Hence, this may be considered a reason why MS media had the lowest identified diversity even though it was rich with nitrogen content.

Citric acid and EDTA are other critical media components. The citric acid present in BG-11 and MM media is responsible for solubilising salts components in the media and preventing iron precipitation, making it readily available to the microalgae, thus resulting in an enhanced growth rate as iron is considered an essential limiting micronutrient²¹. Therefore, it can be a leading cause for the higher diversity present in MM and BG-11 media. BG-11, BBM, and MM media also have EDTA, disodium salt acting as a chelating agent. In contrast, MS medium only has its iron chelated in the form of Ferric sodium EDTA, which has been reported to be a better source of iron^{22,23}. Yet, MS medium did not uniquely enrich any microalgae over the other tested media when morphologically analysed. This indicated that the ratio of iron is relatively low compared to the other media and its increase could have resulted in a higher diversity enrichment.

Nevertheless, we have decided to continue molecular analysis, including MS medium, because morphological analysis could be limited by the volume analysed and the number of microscopic fields visualised from a single sample and the possibility of human error. The morphological analysis was conducted every seven days, and the highest diversity was observed on day 14. Hence, DNA extraction for the molecular level analysis was conducted on the samples on the 14th day. Even though nutrient content was not measured during the experiment, we assume that after day 14, nutrient depletion and accumulation of waste products from the enriched microalgae biomass can drastically increase pH. pH can also increase due to the microalgae uptake of inorganic carbon, and photosynthesis increases O_2 production, which increases the pH of the culture^{24,25}. Exceeding the 14 days may also cause the reduction of the intermediate and rare microalgae (low in abundance) and further boost the dominant and most tolerant microalgae such as *Ankistrodesmus*, *Chlorella*, and *Scenedesmus*^{26,27}. Furthermore, this is common as green algae are usually known to dominate and push down the growth of cyanobacteria²⁸.

The combination of the hypervariable regions sequenced using universal primers specific to the 16S V1-V3, V4-V5 rRNA, and the 18S V4 rRNA region provided us with complete coverage of the algal diversity in

the enriched population. Each primer set aligned to the three reference databases contributed substantially to profiling the different eukaryotic and prokaryotic microalgae present in the River Nile samples studied. Eighty-seven microalgae were characterised from the River Nile samples collected. Additionally, each individual reference database, along with each primer set, had a limited capacity to profile a certain microalgae diversity in the samples analysed. Combining the three primer sets results resulted in a superior coverage that would have been missed if only a single set of primers and one reference database had been used.

Yet, several microalgae (17 genera) have been identified only using morphological characteristics and were missing from the list of species identified using molecular methods. This can be explained by the possibility that some microalgae have been incorrectly morphologically classified due to the close resemblance to other microalgae in the same phylum. For example, *Tetraspora* (Chlorophyta) may be confused with *Chlamydomonas* or *Gloeococcus*. Other genera such as *Leptolyngbya* and *Spirulina* (Cyanobacteria) can also be incorrectly classified with other filamentous Cyanobacteria²⁹. Furthermore, the presence of high nutrients and asexual reproduction can hinder morphological identification of unicellular microalgae due to their phenotypic plasticity under nutrient and environmental alteration^{30,31}.

The presence of synonyms of genera between databases can cause a double count between morphological and molecular-level identification or result in “hit and miss” between one database and another³². Genera such as *Selenastrum* identified through morphological-level identification has a synonym *Monophidium* profiled through SILVA 18S V4 region. Moreover, there is a complete sequence entry of the 18S rRNA gene of *Selenastrum* present in the SILVA database along with a sequence for *Monophidium*³³. This limitation can be ascribed to SILVA database being curated database, and there are several entries with different synonyms of the same genera³⁴. However, some of the microalgae identified only on a morphological level have characteristic features that can be easily distinguished microscopically and are not mistaken with other genera such as *Actinastrum*, *Coelastrum*, *Dictyosphaerium*, *Largerheimia*, *Tetraedron*, *Pediastrum*, and *Selenastrum*^{35,36}. Nevertheless, they were not identified on a molecular level through any of the regions analysed or through the three reference databases used. The complete 18S rRNA gene of the genera *Actinastrum*, *Coelastrum*, *Dictyosphaerium*, *Tetraedron*, *Pediastrum*, and *Selenastrum* are present in the SILVA database, while *Largerheimia* 18S rRNA gene was not available on SILVA but is present in a different database ENA (European Nucleotide Archive) that has not been used in this study for the analysis of the 18S rRNA sequences³³. These results suggest that the genera listed above could be new variants to the genera and, thus, the need for their isolation and characterisation. This also points out that databases that are not specific to microalgae are not constantly updated with microalgal sequences compared to the examined continuously and updated microbial taxonomies such as in studies on the gut and oral microbiomes or general microbial communities in environmental samples^{37–41}.

When it comes to eukaryotic microalgae, they were best identified through the SILVA database, and 18S V4 rRNA gene ~ 63.24% ($n = 43$ (total = 68)) of the identified genera was through the SILVA and 18S V4

region. Even though 25 eukaryotic genera were identified using the 16S V1-V3 region and PR² database. SILVA identified a greater diversity; this can be accredited to the fact that SILVA is a bigger database ⁷. But the microalgae specifically identified using PR² reference database ¹⁰ and the 16S rRNA gene can still be considered high and of added value to the profiling of microalgae diversity from an environmental sample.

Through the alpha diversity indices calculated for the OTUs analysed through the different databases, the Good's coverage index was higher in the V4-V5 region in general. Moreover, through the V4-V5 region annotated by the three databases, the Nile2 sample had a high Shannon index indicating the presence of high diversity in terms of abundance, and the Sobs index confirms that. While, the InvSimpson index, a dominance index, indicates that Nile2 has a lower InvSimpson index than Nile3. Nile3 sample has a higher diversity with more evenness between the genera. Furthermore, this is also confirmed by the decrease of the Berger Parker index of the Nile3 sample across all the databases and regions sequenced. The diversity indices indicate that the Nile3 is the most abundant and with an even distribution of genera and the absence of the dominance of a genus. This is followed by the Nile2 and Nile1 samples. Regarding the 18S V4 region annotated by SILVA, the diversity indices indicate that Nile1 has the richest diversity with a high evenness between the different genera identified. This is followed by Nile3 and then Nile2.

The biplot constructed using the SILVA 18S V4 region dataset confirms that media usage does enrich specific diversity and enriched diversity is biased to the media. This further confirms that using one medium can lead to the enrichment of certain microorganisms over the other. Hence, the usage of one enrichment media will lead to the loss of identifying certain microalgae and provide misleading data concerning the diversity of studied environmental samples. Using BG-11, BBM and MM were the most suitable for culturing and enrichment of the highest diversity from an environmental sample, compared to MS media that did not identify any unique species. This enrichment culture-based method can also be considered an alternative to the more expensive approach of increasing the number of sequencing reads to investigate further the low abundant species that are not enough to appear due to the presence of a limiting detection threshold. The culture-based method followed by the molecular level through several different hypervariable regions eliminates these biases that can result from using one media or one sequencing region. Moreover, it is critical to recognise that using different enrichment media and sequencing regions complement each other and provide completeness to the diversity profiled.

Thereby, we may conclude that the combined use of both the culture-based method and sequence analysis of multiple rRNA regions/genes (targeting both nuclear and plastid genomes ⁴²) facilitates a more reliable and comprehensive approach to identifying total microalgae biodiversity. This approach can be considered a more efficient methodology as it targets microalgae identification from an environmental sample over the commonly used metagenomic sequencing that utilises the environmental sample directly, which are usually high in microbial biomass and low in microalgae biomass.

Methods

Study areas and sample enrichment

To investigate the competence of the different databases and the multi-primers studied, three water samples were collected from the Nile River in the Cairo governate (**Supplementary Table S1**) and enriched with different nutrient media. A populated city surrounds the site from which the samples have been collected. The Nile River provides the population with water supply, fishery, and transportation. Hence, predicting the presence of functional biodiversity in the site. Water samples were collected 20 cm below the water surface from each location and immediately kept in the fridge until processed in the lab. Portions of each sample were used for physicochemical analysis, and the remaining water samples were used to inoculate the nutrient media, as follows. Four different nutrient media were used for algal enrichment: Blue-Green medium (BG-11), Bold's Basal medium (BBM), half-strength Murashige and Skoog medium (MS), and a Modified medium (MM) that fits in the middle of the BG-11 and BBM in-terms of the nitrogen and phosphorus composition (**Supplementary Table S2**). For all tested factors, cultures were carried out in triplicate and were maintained for three weeks in a JSR-Growth Chamber 3-Side Illumination (model JSPC-960C2) at a temperature $22 \pm 1^\circ\text{C}$. The light/dark regime was adjusted to 12 h:12 h, at a light intensity of $28 \mu\text{mol photons m}^{-2} \text{s}^{-1}$. Microalgae population growth was monitored by performing a cell count every four days using a Neubauer Improved Hemocytometer counting chamber for each replica under Zeiss AxioStar plus light microscope at 400x magnification based on ^{43,44}.

Physicochemical analysis

The physicochemical analysis was performed to assess the composition of the microalgae's natural habitat identified in each location's water samples (**Supplementary Table S3**). The physical and chemical characteristics measured include salinity, pH and Electrical conductivity (EC), were measured using a multimeter-probe (330i, WTW, Germany). The presence of cation and anions, such as sodium, chloride, sulfate, potassium, and calcium, were measured according to standard methods ⁴⁵. Total ammonia, nitrate and phosphorus were evaluated based on the standard protocol ⁴⁶.

Morphological identification of microalgae

The identification of microalgae was based on the morphological characteristics observed under a bright-field microscope following the standard procedures ^{35,36}. The strains were examined under a light microscope (Leica) using the software LAS EZ (Leica DM500). Genes-level identification of the microalgae was performed.

DNA extraction and sequencing

Total genomic DNA was extracted after 14 days of media enrichments of sample from a volume of 200 ml ($\sim 4.7 \times 10^6 \text{ cells ml}^{-1}$) of each culture and centrifuged at 4000 rpm for 30 minutes. The wet pellets collected were then weighed, and aliquots of 150–250 mg were transferred into 1.5ml Eppendorf tubes for total DNA extraction. The Eppendorf was then dropped in liquid nitrogen and grind to powder form to

extract DNA. Genomic DNA was extracted from the frozen wet pellet using the PowerMax Soil DNA Isolation kit (MoBio, Carlsbad, CA) following the manufacturer's instructions. The concentration and purity of the DNA extracted were monitored using an LVis Plate SPECTROstar® Nano (BMG LABTECH, UK) and 1% agarose gel.

Three primer sets hyper-variable regions of 16S rRNA V1-V3 43,44, V4-V5 43 and 18S rRNA V4 ^{49,50} were used for sequencing (**Supplementary Table S4**). Sequencing was performed by BGI (BGI-Tech, BGI-Shenzhen, Shenzhen, China) using Illumina HiSeq 2500 platform to generate 250/300bp paired-end (PE) reads.

Raw sequence processing and data analysis

Trimmed fastq files were filtered using MOTHUR v1.42.6 following the pipeline from the MiSeq standard operating procedure (SOP) available on their website (www.mothur.org/wiki/MiSeq_SOP) ⁵¹. Briefly, the fastq files were converted to fasta, and the files were screened using the MOTHUR command screen.seq. Reads shorter or more than 490, 420 and 690bp were discarded from the 16S rDNA, V1-V3, V4-V5 and 18S rDNA sequences, respectively, along with homopolymers longer than 8bp. The remaining sequences were de-replicated (unique.seqs) to merge duplicates and reduce the number of sequences to analyse. The remaining sequences were aligned to a database. The sequences of the 16S rDNA were aligned to the Greengenes database (Gg_13_8_99) and Protist Ribosomal Reference database (PR²) ¹⁰ 18S rDNA was aligned to the SILVA database (Silva.nr_v132). The lowest level of the taxonomic hierarchy present in SILVA is the genus, while in Greengenes lowest taxonomic level that can be identified is the species level ^{7,38}. Chimeras in the sequences were also filtered and removed (chimera.uchime and remove.seqs). Sequences were split into groups base on their taxonomy at the order level, and OTUs (operational taxonomic units) were assigned using the dist.seqs command. Alpha and beta diversities were calculated based on normalised OTU abundance information obtained using the sample with the fewest sequences as a standard. The diversity indices calculated include Good's coverage, species observed (S_{obs}), Shannon's diversity index, Simpson's diversity index (InvSimpson), and they were calculated using the MOTHUR SOP ⁵².

The molecular data were further analysed and visualised with R 3.6.1 packages executed via RStudio ("Open source and enterprise-ready professional software for data science-RStudio"; "R: The R Project for Statistical Computing").

Data Availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Author contributions statement

Writing- Original draft: A.B.; Review & Editing: A.B. and W.F.; Oversaw the project: W.F.

Competing interests

The authors declare no competing interests.

References

1. Barsanti, L. et al. The World of Algae. in *Algal Toxins: Nature, Occurrence, Effect and Detection* (eds. Evangelista, V., Barsanti, L., Frassanito, A. M., Passarelli, V. & Gualtieri, P.) 1–15 (Springer Netherlands, 2008).
2. Winwood, R. J. Recent developments in the commercial production of DHA and EPA rich oils from micro-algae. *OCL - Oilseeds and fats, crops and lipids* **20**, (2013).
3. Mobin, S. & Alam, F. Some Promising Microalgal Species for Commercial Applications: A review., **110**, 510–517 (2017).
4. Alp, M. T., Fakioğlu, Y. E., Özbay, Ğ. & Koçer, M. A. T. A study on water quality and trophic state of Akgöl Lagoon (Mersin, Turkey). *Aquatic Ecosystem Health and Management*, **19**, 58–63 (2016).
5. Gökçe, D. Algae as an Indicator of Water Quality. *Algae - Organisms for Imminent Biotechnology*, <https://doi.org/10.5772/62916> (2016).
6. Khaw, Y. S., Khong, N. M. H., Shaharuddin, N. A. & Yusoff, F. M. A simple 18S rDNA approach for the identification of cultured eukaryotic microalgae with an emphasis on primers. *Journal of Microbiological Methods*, **172**, 105890 (2020).
7. Balvočiute, M., Huson, D. H. & SILVA, R. D. P. Greengenes, NCBI and OTT - how do these taxonomies compare? *BMC Genomics* **18**, (2017).
8. DeSantis, T. Z. et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology*, **72**, 5069–5072 (2006).
9. McDonald, D. et al. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *The ISME Journal*, **6**, 610–618 (2012).
10. Guillou, L. et al. The Protist Ribosomal Reference database (PR2): A catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Res*, **41**, D597 (2013).
11. Kulichová, J., Škaloud, P. & Neustupa, J. Molecular diversity of green corticolous microalgae from two sub-Mediterranean European localities. *European Journal of Phycology*, **49**, 345–355 (2014).
12. Sánchez Roque, Y. et al. Evaluation of the population dynamics of microalgae isolated from the state of Chiapas, Mexico with respect to the nutritional quality of water. *Biodiversity Data Journal*, **6**, e28496 (2018).

13. Wu, Y. *et al.* Identification and Characterization of an Electrogenic Microalgae Strain. *PLoS ONE*, **8**, e73442 (2013).
14. Peto, L. *et al.* Selective culture enrichment and sequencing of feces to enhance detection of antimicrobial resistance genes in third-generation cephalosporin resistant Enterobacteriaceae. *PLOS ONE*, **14**, e0222831 (2019).
15. Biteen, J. S. *et al.* Tools for the Microbiome: Nano and Beyond. *ACS Nano*, **10**, 37 (2016).
16. Jin, S. *et al.* An assessment of the taxonomic reliability of DNA barcode sequences in publicly available databases. *Algae*, **35**, 293–301 (2020).
17. Groendahl, S., Kahlert, M. & Fink, P. The best of both worlds: A combined approach for analyzing microalgal diversity via metabarcoding and morphology-based methods. *PLOS ONE*, **12**, e0172808 (2017).
18. Nistal-García, A. *et al.* DNA metabarcoding and morphological methods show complementary patterns in the metacommunity organization of lentic epiphytic diatoms. *Science of The Total Environment*, **786**, 147410 (2021).
19. Badr, A. A. & Fouad, W. M. Identification of culturable microalgae diversity in the River Nile in Egypt using enrichment media. *African Journal of Biological Sciences*, **3**, 50 (2021).
20. Flynn, K. & Butler, I. Nitrogen sources for the growth of marine microalgae: role of dissolved free amino acids. *Mar. Ecol. Prog. Ser.*, **34**, 281–304 (1986).
21. Gour, R. S., Bairagi, M., Garlapati, V. K. & Kant, A. Enhanced microalgal lipid production with media engineering of potassium nitrate as a nitrogen source. <https://doi.org/10.1080/21655979.2017.1316440> **9**, 98–107 (2017).
22. Rana, M. S. & Prajapati, S. K. Resolving the dilemma of iron bioavailability to microalgae for commercial sustenance. *Algal Res*, **59**, 102458 (2021).
23. Botebol, H. *et al.* Different iron sources to study the physiology and biochemistry of iron metabolism in marine micro-algae. *BioMetals*, **27**, 75–88 (2014).
24. Juneja, A., Ceballos, R. & Murthy, G. Effects of Environmental Factors and Nutrient Availability on the Biochemical Composition of Algae for Biofuels Production: A Review. *Energies*, **6**, 4607–4638 (2013).
25. Yang, Z. *et al.* Effects of air bubble size on algal growth rate and lipid accumulation using fine-pore diffuser photobioreactors. *Algal Res*, **32**, 293–299 (2018).
26. Berge, T., Daugbjerg, N. & Hansen, P. J. Isolation and cultivation of microalgae select for low growth rate and tolerance to high {pH}. *Harmful Algae*, **20**, 101–110 (2012).
27. Vadlamani, A., Viamajala, S., Pendyala, B. & Varanasi, S. Cultivation of Microalgae at Extreme Alkaline pH Conditions: A Novel Approach for Biofuel Production. *ACS Sustainable Chemistry and Engineering*, **5**, 7284–7294 (2017).
28. Nolan, M. P. & Cardinale, B. J. Species diversity of resident green algae slows the establishment and proliferation of the cyanobacterium {Microcystis} aeruginosa. *Limnological*, **74**, 23–27 (2019).

29. Shubert, E. & Gärtner, G. Nonmotile Coccoid and Colonial Green Algae. in *Freshwater Algae of North America* 315–373 (Elsevier, 2015). doi:10.1016/B978-0-12-385876-4.00007-4.
30. Sassenhagen, I., Wilken, S., Godhe, A. & Rengefors, K. Phenotypic plasticity and differentiation in an invasive freshwater microalga. *Harmful Algae*, **41**, 38–45 (2015).
31. Dzhenbekova, N. *et al.* Comparative analysis of morphological and molecular approaches integrated into the study of the dinoflagellate biodiversity within the recently deposited Black Sea sediments – benefits and drawbacks. *Biodiversity Data Journal*, **8**, e55172 (2020).
32. Hugenholtz, P., Chuvochina, M., Oren, A., Parks, D. H. & Soo, R. M. Prokaryotic taxonomy and nomenclature in the age of big sequence data. *The ISME Journal* 2021 15:7 15, 1879–1892(2021).
33. RNAcentral: The non-coding RNA sequence database. <https://rnacentral.org/> (2021).
34. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research* **41**, (2013).
35. Bellinger, E. G. & Sigeo, D. C. *Freshwater Algae: Identification and Use as Bioindicators. Freshwater Algae: Identification and Use as Bioindicators* (Wiley-Blackwell, 2010). doi:10.1002/9780470689554.
36. Evans, R. I. & Prescott, G. W. How to Know the Fresh-Water Algae. Bulletin of the Torrey Botanical Club vol. 83(W.C. Brown Company, 1956).
37. Parks, D. H. *et al.* A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature Biotechnology* 2018 36:10 36, 996–1004(2018).
38. Sierra, M. A. *et al.* The Influences of Bioinformatics Tools and Reference Databases in Analyzing the Human Oral Microbial Community. *Genes*, **11**, 1–12 (2020).
39. Hernández-Pérez, A., Zamora-Briseño, J. A., Söderhäll, K. & Söderhäll, I. Gut microbiome alterations in the crustacean *Pacifastacus leniusculus* exposed to environmental concentrations of antibiotics and effects on susceptibility to bacteria challenges. *Developmental & Comparative Immunology*, **104181**, <https://doi.org/10.1016/J.DCI.2021.104181> (2021).
40. Wang, H. *et al.* Gut microbiome-host interactions in driving environmental pollutant trichloroethene-mediated autoimmunity. *Toxicology and Applied Pharmacology*, **424**, 115597 (2021).
41. Ramakodi, M. P. Effect of Amplicon Sequencing Depth in Environmental Microbiome Research. *Current Microbiology* 2021 78:3 78, 1026–1033(2021).
42. Eiler, A. *et al.* Unveiling Distribution Patterns of Freshwater Phytoplankton by a Next Generation Sequencing Based Approach. *PLoS ONE* **8**, (2013).
43. Karlson, B., Godhe, A., Cusack, C. & Bresnan, E. Microscopic and molecular methods for quantitative phytoplankton analysis. *Intergovernmental Oceanographic Commission Manuals and Guides*, **55**, 5–12 (2010).
44. Raeesossadati, M. J., Ahmadzadeh, H., McHenry, M. P. & Moheimani, N. R. CO₂ bioremediation by microalgae in photobioreactors: Impacts of biomass and CO₂ concentrations, light, and temperature. *Algal Res*, **6**, 78–85 (2014).

45. ASTM Standard D4327. *Test Method for Anions in Water by Suppressed Ion Chromatography*. (2017) doi:10.1520/D4327-17.
46. Rice, E. W. & Franson, M. A. H. *21st Edition 2005* (American Public Health Assoc, 2005).
47. Turner, S., Pryer, K. M., Miao, V. P. W. & Palmer, J. D. Investigating Deep Phylogenetic Relationships among Cyanobacteria and Plastids by Small Subunit rRNA Sequence Analysis. *The Journal of Eukaryotic Microbiology*, **46**, 327–338 (1999).
48. Watts, G. S. *et al.* 16S rRNA gene sequencing on a benchtop sequencer: accuracy for identification of clinically important bacteria. *Journal of Applied Microbiology*, **123**, 1584–1596 (2017).
49. Heidelberg, K. B. *et al.* Characterization of eukaryotic microbial diversity in hypersaline Lake Tyrrell, Australia. *Frontiers in Microbiology* **4**, (2013).
50. Wolf, C., Kiliyas, E. S. & Metfies, K. Evaluating the potential of 18S rDNA clone libraries to complement pyrosequencing data of marine protists with near full-length sequence information. *Mar. Biol. Res*, **10**, 771–780 (2014).
51. Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. & Schloss, P. D. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the miseq illumina sequencing platform. *Applied and Environmental Microbiology*, **79**, 5112–5120 (2013).
52. Mundra, S. *et al.* Soil depth matters: shift in composition and inter-kingdom co-occurrence patterns of microorganisms in forest soils. *FEMS Microbiology Ecology* **97**, (2021).

Figures

Figure 1

Freshwater microalgae taxa identified across different nutrient media. Venn diagram illustrating the number of shared and unique freshwater microalgae taxa across different culture media. The number of microalgae taxa identified on media BG-11, BBM, MM, and MS using two identification methods (A) Morphological identification using a light microscope (B) Molecular-level identification using different DNA barcodes. (C) The total number of microalgae taxa commonly shared or uniquely identified using morphological and molecular identification methods.

Figure 2

Venn diagrams illustrating the unique and shared microalgae identified through the different databases (SILVA, GG, and PR2) and the different DNA barcodes sequenced. Cyanobacteria genera were identified through (A) 16S V1-V3 region and (B) 16S V4-V5 rRNA region. (C) Eukaryotic genera identified through 18S rRNA gene annotated using SILVA database and 16S rRNA gene annotated using PR2 database.

Figure 3

Relative abundance of cyanobacteria identified through 16S rDNA V1-V3 (top-half) and 16S rDNA V4-V5 region (bottom-half) annotated through (A & D) SILVA, (B & C) GG and (C & F) PR2 databases.

Figure 4

Relative abundance of eukaryotic microalgae identified through the different databases (SILVA, GG, and PR2) and the different DNA barcodes sequenced. (A-D) are eukaryotic microalgae identified through 18S rRNA gene annotated using SILVA database. Relative abundance of eukaryotic microalgae genera identified using 16S V1-V3 through GreenGenes and PR2 (E & F), respectively, 16S V4-V5 rRNA gene (G & H) GreenGenes and PR2, respectively.

Figure 5

Alpha Diversity Indices of Total Microbial Community based on OTUs. The total calculated alpha diversity are as follows: Good's Coverage; Sobs (total number of OTUs observed); Shannon; InvSimpson and BregerParker. The first half belonging to the V1-V3 region (A-C) and the second-half represents the V4-V5 region (D-F), and last plot belongs to 18S V4. (A, D) SILVA, (B, E) GG, and (C, F) PR2 database. (G) Alpha diversity indices calculated through 18S rRNA gene and SILVA.

Figure 6

Non-metric multidimensional scaling analysis (nMDS) computed on Bray-Curtis similarity index obtained for microbial community-identified using SILVA 18S V4, water samples, and media as factors.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTables.docx](#)