

Development of DNA barcode for freshwater microalgae species revealing it's evolutionary insights on diversity in Noyyal river of Western ghats, India

Pandian Sureshkumar

Karunya Institute of technology and Science: Karunya Institute of Technology and Sciences

Jibu Thomas (✉ jibuthomas.t@gmail.com)

Karunya Institute of Technology and Sciences <https://orcid.org/0000-0001-8132-6229>

Sukumar Bhakta

Botanical Survey of India

Research Article

Keywords: DNA barcode, Microalgae, molecular identification, GENBANK, BOLD database, Noyyal river

Posted Date: February 24th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1330489/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

The identification of microalgae is extremely difficult solely based on morphological features. DNA based classification is particularly performed for unveiling cryptic diversity at various taxonomic levels and identifying species where structural characteristics are few or difficult to ascertain. The efficiency of DNA barcoding in recognizing microalgae involves generation of multiple gene markers and phylogenetic approaches that are however attaining significance. In this study, the feasibility of using DNA markers to identify a diverse and also poorly known group of green microalgae isolated from freshwater of Siruvani, Coimbatore was evaluated. DNA extracted from 23 freshwater microalgae cultures (Karunya Algae Culture Collection) was amplified using 18S, 28S, ITS and 18S region specific for Chlorophyceae with amplification success rate of 86.95%, 82.60%, 43.47% and 69.59% respectively. Ambiguity among the 22 out of 23 microalgae isolates was unveiled using the results generated from the phylogenetic analysis and morphological data, thus confirming the taxonomical identity. Generated sequences of KACC microalgae were deposited in GENBANK and BOLD database for further use as reference barcodes for microalgae identification. In conclusion our findings highlight barcode database as powerful tool in microalgae identification and further it uncover new perspectives in species diversity on a global scale.

Introduction

Algal systematics is highly based on the morphospecies concept i.e., a species is described and identified on the basis of morphological criteria. These methods are tedious, time-consuming and direct species identification is often difficult and hence it is necessary to find an alternative method of technologies that can assist traditional taxonomical identification (Songdong 2008). Though algae have been studied over the decades, the taxonomic and phylogenetic relationship was poorly understood. The criterion used in identification and classification of algae are morphology, ultrastructure of cell, cell division- mitosis, cytokinesis, flagellar apparatus, patterns of life cycles and molecular markers like ITS, *rbcL*, rRNA genes etc., Knowing the morphological and biochemical characteristics alone in most cases is derisory to elucidate the evolutionary relationships of microalgae. Morphology of algae may vary within each chemical race and even within the same strain in relation to age and culture conditions. Morphological heterogeneity of the alga also makes the identification often difficult by microscopy and somewhat inaccessible to non-specialist (Metzger and Largeau 2005).

Many reports have suggested that there are lots of taxonomic problems in the traditional classification system based on the morphology (Krienitz et al. 2001;2004; Luo et al. 2006). However, molecular techniques can be used to differentiate between the species of algae and allow reliable identification of the species (Fawley and Fawley 2002). Literature evidences showed that morphological taxonomy and molecular phylogenetics studies using ITS, *rbcL*, 18S rRNA, *COI* have been reported for Chlorophyta and green algae (Huss et al. 1999; Lo'pez-Bautista et al. 2006; Hall et al. 2010), Scenedesmus (Kessler et al. 1997; Zou et al. 2016) Coelastrella [Ancona-canché et al. 2017) ulvophyceae (Zhu et al. 2015) and brown algae (McDevit and Saunders 2010). Phylogenetic studies on these algae had demonstrated the prevalence of homoplasy and polyphyly of various taxa (Hall et al. 2010). The small subunit ribosomal RNA has

proved to be an invaluable tool in molecular evolution (Woese 1987; Sogin 1989; Doolittle and Brown 1994). Analysis of small subunit rRNA sequence data has facilitated the classification of phylogenetic relationships among microalgae (Gunderson et al. 1987; Wilcox et al. 1992). However, molecular-based phylogenetic analysis of microalgal species is very much limited in India as most of the diversity analysis rely on the morphological characterization.

Sequencing of algal DNA and the comparison of its genetic material have become the powerful analytical tool in modern taxonomy. This method gives an indication on the blueprints of the organisms, and to correlate them to see for the changes occur if any. Similar approach was gauged by the flooding of research papers in many scientific journals (Hall et al. 2010; Hadi et al. 2016; Zou et al. 2016). This also has given a widespread acceptance for sequence based studies on phylogenetic speculation. The impressive results obtained using molecular methods have also created a caution for the interpretation of their phylogenetic significance. The uncertain amount of genomic data of algae in NCBI database limits the development of algae biotechnology makes it challenging and taxonomically complex.

DNA barcoding methods have been applied for the identification of number of micro and macro algal groups (Lo ´pez-Bautista et al. 2006; Kessler et al. 1997; Hadi et al. 2016). The lack of resources among the algal culture collections to create DNA barcodes and to utilize DNA sequencing technology as a method to confirm the identity of strains remains to be the limitation at the time of accession. Morphologically identical specimens could also be genetically distinct. The ideal DNA barcode would be sufficiently vary to distinguish between closely related species. The marker should be found in all taxa, and it amplifies using universal primers. Finding an ideal molecular barcode for all algae is complicated because of the fact that algae are an ancient and are of diverse lineage. The Barcode of Life Data System (BOLD) provides an integrated bioinformatics workbench aiding the analytical pathway from specimen collection to tightly validated barcode library. By assembling molecular, morphological and distributional data, it bridges a traditional bioinformatics chasm. It is a repository for the specimen and sequence records that serve as a basic data for all forms of barcode lives (Ratnasingham and Hebert 2007). Currently 4,449 *rbcL* sequences from chlorophyte species and few rRNA sequences for green algae have been deposited at the BOLD database (Hadi et al. 2016). Although DNA barcoding is not an immediate remedy for the taxonomic problems, it reduces the need for referring it to a physical type specimens to a greater extent, provided the barcodes are linked to type specimens and made effective in nomenclature. But a classified data remains mostly unavailable and unaccessible. A barcode, sequence-essentially a molecular type is unambiguous and can be communicated easily, contrasting to the morphology of a given specimen.

Our earlier study reported the population dynamics and diversity indices of 142 fresh water microalgae species of river Noyyal and their response to the seasonal variation with physico – chemical characters of water (Kumar and Thomas 2019). The study also provided the checklist of microalgae in abundance, the distribution and community structure in the five collection sites from catchment region of river Noyyal. Another study from our lab highlighted the morphological (ultra-structural features) and biochemical characterisation (FT-IR and Nile-red) of these isolated microalgae and their biomass (Sureshkumar and Thomas 2020). So the present study focuses on the identification of 23 fresh water green microalgae which are coded as Krunya Algae Culture Collection (KACC 1-23) isolated from catchment region of river

Noyyal using a combination of DNA sequence analysis and light microscopy examination. Hence the study investigates the following goals (1) to characterize the eukaryotic algal isolates of freshwater using morphological traits, to incorporate molecular phylogenetics for further identification and to place the isolates into a larger known green algal tree of life, (2) to enrich our knowledge in the diversity of eukaryotic algae from unexplored Siruvani Waters, from catchment region of river Noyyal, Western Ghats India and to confirm and authenticate the microalgae identification through DNA Barcode database.

Materials And Methods

Microalgae isolation and morphological characterization

Microalgae cultures isolated from water samples collected from river Noyyal of Western ghats at five different locations as mentioned in our earlier studies (Kumar and Thomas 2019, Sureshkumar and Thomas 2020). They were maintained as pure cultures and periodically observed under a light microscope at 40X and 100X magnifications using Accuscope Light Microscope (Accuscope, USA). The morphological characteristics were observed and documented using a camera equipped with a microscope (Kumar and Thomas 2019).

PCA discrimination

FT-IR spectra was collected from microalgae to provide insight for discrimination of the microalgae population based on their biomolecular profile. In the present study, Principal Component Analysis (PCA) has been done to examine the changes across the whole population of spectra and cells which contained the majority of the spectral information. It was performed over full spectra range (4000–400 cm^{-1}) collected using FT-IR spectrometer equipped with an air-cooled Deuterated Tri Glycine Sulfate (Madison, USA) according to the protocol (SureshKumar et al. 2017). Resulting PCA scores and loading plots were compared to determine the optimum spectral range for the microalgae population. All the analyses were carried out using SPSS IBM software (Driver et al. 2015).

DNA Extraction, PCR amplification and sequencing

DNA from microalgae cells was extracted using a modified CTAB method (Doyle and Doyle, 1987). PCR amplification reactions were performed on a gradient thermocycler (Applied Biosystems, USA) using following primers 18S rDNA (Fwd 5'-ACCTGGTTGATCCTGCCAGT-3'; Rev 5'-TCAGCCTTGCGACCATAC-3'), 18S for Chlorophyceae (Fwd 5'- TGGCCTATCTTGTTGGTCTGT-3'; Rev 5'- GAATCAACCTGACAAGGC-3'), ITS-1 (5'- TCCGTAGGTGAACCTGCGG-3'). ITS 4 (5'- TCCTCCGCTTATTGATATGC-3'), 28S rDNA (Fwd 5' GGACAGAAAGACCCTATGAA-3'; Rev 5'- TCAGCCTGTTATCCCTAGAG-3'). PCR was performed in 25- μL reaction volume as described in the protocol (Piligaev et al. 2015). Sequencing reactions were performed using ABI 373096 capillary, Big Dye Terminator Cycle Sequencing Ready Reaction Kit and the sequencing fragments were analyzed on Applied Biosystems 3730xl DNA analyser.

Sequence alignment and phylogenetic analysis

The raw sequences were edited and trimmed using Bioedit v.7.2.5 and contig sequence formed was analysed using nucleotide BLAST tool. Phylogenetic analysis for KACC isolates was performed on a highly similar sequences obtained from NCBI BLAST. The sequences showing the highest homology were subjected to multiple sequence alignment using ClustalW and MEGA7.0. Finally, the aligned nucleotide sequences were exposed to phylogenetic tree construction using neighbour joining (NJ) algorithm with 1000 bootstrap values through MEGA v7.0 (Tamura et al. 2007). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test was shown next to the branches (Felsenstein 1985). The tree has drawn to scale, with branch lengths same as that of evolutionary distances used to infer the phylogenetic tree.

Identification of the taxa

Taxonomic description of each taxon was studied in depth from standard literature to identify the specimens. In addition to the published papers, standard book, algal floras, monographs, journals, websites, and algal repositories were also referred for the identification of the collections. All taxa identified in the study were confidently identified to the species level, and species name checked and confirmed with AlgaeBase (www.algaebase.org) (Guiry and Guiry 2018) as well as the primary literature for distributional records.

BOLD database

Data submission to BOLD was done under the Project ID, KACC. A total of 30 sequences were submitted. Only two data elements were considered per project to inject each specimen record with a sample ID number and a taxonomic assignment (although the latter can simply be to a phylum-level). In the BOLD database preliminary submission of specimen details for all the species prelude the sequence submission process. However, to record the specimen the following seven data elements need to be uploaded. 1. Species name, 2. Voucher data 3. Collection record (collector, collection date and location with GPS coordinates) 4. Identifier of the specimen, 5. Nucleotide sequence, 6. PCR primers used 7. Trace files. On submission of DNA sequence data to BOLD, the unique four color coded candidate DNA barcode was obtained.

Results

Total of 142 microalgal cultures isolated from Catchment region of river Noyyal, the Western Ghats were downsized to 23 different cultures based on their unique culture morphology and distinct characteristics (Table 1). These cultures were coded as Karunya Algae Culture Collection (KACC) 1-23. Microscopic observations of these microalgal isolates revealed its colonial existence and purity. Based on its different morphological features like cell shape, size and color revealed that the cultures belongs to the family of Asteromonadaceae, Brachysiraceae, Bacillariaceae, Chlorococcaceae, Chlorellaceae, Oscillatoriaceae, Prasiolales incertae sedis, Scenedesmaceae, Selenastraceae, and Sphaeropleaceae. Prior to identification, a study was attempted to find the diversity of KACC microalgae cultures based on its biochemical profile. For that purpose FT-IR was used in the present study which provides insight for discrimination of the

microalgae population based on their biomolecular contents. PCA examined the changes across the whole population of spectra and results showed that there was a clear clustering pattern between KACC microalgae isolates in FT-IR spectra and that the clustering was based on more biological relevant information. In PCA analyses, three PCs were considered, and PC1 (triglycerides region) accounted for 95% of the variance in clustering. Scoring plots showed significant discrimination between KACC microalgae due to PC1 with PC2 and PC3, contributing to some further discrimination (Fig. 1). The loading plot for PC1 showed the difference between the clusters due to absorbance band of triglycerides and differing C–O stretching band absorbance of esters. Our findings also proved that KACC strains exhibited cluster pattern due to their close proximity of the clusters of biomolecule derived from FT-IR spectra.

Table 1
Morphological description of KACC microalgae cultures

Culture code	Morphological identification	Distinct Morphology	Cell Size (µm)
KACC 1	<i>Chlorella vulgaris</i>	Very minute cell size	4.39
KACC 2	<i>Actinochloris sphaerica</i>	Green colored oval shape	8.26
KACC 3	<i>Bracteacoccus anomalus</i>	Circular minute size	5.35
KACC 4	<i>Myrmecia astigmatica</i>	Cells with numerous conspicuous vacuoles. Cells uninucleate, apparently without centrioles	6.28
KACC 5	<i>Myrmecia incisa</i>	Always found in cluster/colonies	4.62
KACC 6	<i>Scenedesmus acutus</i>	Colonial non motile	8.94
KACC 7	<i>Desmodesmus dispar</i>	Rotational ellipsoid structure	9.24
KACC 8	<i>Monoraphidium contorium</i>	Irregular comma shaped	11.69
KACC 9	<i>Secendesmus acunae</i>	Rotational ellipsoid structure	4.62
KACC 10	<i>Chlorella sp</i>	Subspherical to subcylindrical unicells	1.4
KACC 11	<i>Chlorococcum minutum</i>	Spherical or slightly oblong cells of varied size	8.95
KACC 12	<i>Navicula exilis</i>	Unicellular symmetric bivalve, straight raphe	
KACC 13	<i>Follicularia texensis</i>	Spherical or slightly oblong cells of varied size	13.31
KACC 14	<i>Nitzshia paleacea</i>	The valves are narrowly linear-lanceolate and taper gradually to very narrow, rounded ends	27.76
KACC 15	<i>Kirchneriella incurvata</i>	Individual sickle shaped cells form a colony in a mucus envelope	6.52
KACC 16	<i>Ankistrodesmus fusiformis</i>	Fibre shaped straight cells	26.26
KACC 17	<i>Geiterinema exile</i>	Long ladder shaped	60.20

Culture code	Morphological identification	Distinct Morphology	Cell Size (µm)
KACC 18	<i>Tetraedron caudatum</i>	Unicell with an irregular outline, and four or five principal "points" on the corners	5.91
KACC 19	<i>Planophila Laetevirens</i>	Spherical or slightly oblong cells of varied size	8.19
KACC 20	<i>Ankistrodesmus fusiformis</i>	Elliptical single cells	12.31
KACC 21	<i>Chlorella neustonica</i>	Green spherical to ellipsoidal unicells	14.06
KACC 22	Un identified	Green spherical cells	14.93
KACC 23	<i>Elliptochloris subsphaerica</i>	Green spherical cells in colonies	5.55

Description of culture morphology

Based on the morphological characteristics, 23 different microalgae cultures were identified. The morphological identification with the description of each culture was given in the Table 1. The microscopic images of each culture was given in the supplementary (Fig S1-S4).

Kalenjinia: Cells are either solitary or in 2–4 celled colonies, with a gelatinous envelope, planktonic. Cells mostly ovoid, drop-shaped, or spherical (4–14 × 3.5–9 µm). Single cup-shaped chloroplast with starch-covered pyrenoid.

Scenedesmus: Spine-like appendages seen among the mostly elliptical single cells fuse together in rows, ranging from two to 32.

Chlorella: Green spherical to ellipsoidal unicells with one chloroplast, often growing in groups. Asexual reproduction by forming 2-8 autospores.

Chlorococcum: Nonflagellated green unicells that become multi-nucleate prior to forming zoospores (biflagellated). Spherical and/or broad ellipsoidal, chloroplasts - cup shaped to several in number, or plate-like, reticulated or star-shaped, with or without pyrenoids.

Monoraphidium: Solitary or sometimes attached by extremities to form chains, cells 2-6 µm wide and 5-19 µm long, short to broadly spindle-shaped, mostly asymmetrical, slightly pointed or rounded at apices, autospores 4 or 8 per cell.

Nitzschia: The outline of the cells vary widely which are usually elongated and in valve view. The valves can be straight to sigmoid and can lie in either valve or girdle view. The center can be slightly swollen, rounded edges at each pole. The cells can be unicellular or form chain-like colonies, sometimes forming within mucilage tubes.

Kirchneriella: Individual sickle-shaped cells form a colony of 2-64 cells in a mucus envelope, shapes of colonies vary. The cells are uninucleated and contain a single chloroplast. Cells cylindrical, lunate, sickle-shaped, twisted fusiform or spirally twisted, $3-40 \times 1-7 \mu\text{m}$, with smooth cell walls. Cells uninucleate, chloroplast single and parietal, pyrenoids 1-4 per cell.

Nanochloropsis: Unicells lacking flagella, spherical to ovoid, 2 – 4 μm diameter.

Brachysira: Unicellular symmetric bivalve 4-9 μm wide and 15-45 μm long, rhombic-lanceolate, straight raphe, striae punctuate with 9-12 punctae per 10 μm .

Ankistrodesmus: Unicellular or in clusters, of loosely bound unicells. Fiber-shaped cells are either straight, bent, or spiral. Needle-like, sometimes curving.

Tetraedron: Unicell with an irregular outline, and four or five principal "points" on the corners

Coelastrella: Cells are an ovoid to ellipsoidal cell shape and rough cell wall. Each sporangium was composed of 2-6 young autospores with visible pyrenoid. The cell arrangement in sporulation showed cells in two superposed lines. The cell measurements were $7.01 \pm 0.28 \mu\text{m}$ in length and $4.56 \pm 0.31 \mu\text{m}$ in width with some granulations of 159 nm dispersed on the cell wall. The pyrenoid was defined by two or three divisions of starch granules.

Molecular identification

Total genomic DNA from the freshwater microalgae (KACC 1 -23) was extracted and the size of the genomic DNA was found as $>10\text{kb}$. PCR was amplified using 18S rRNA, 28S rRNA and ITS gene and the product was found to be $\sim 1100 \text{ bp}$, 590 bp , 700 bp respectively. PCR amplification success rate for 18S, 28S, ITS and 18S for Chlorophyceae were found as 86.95%, 82.60%, 43.47% and 69.59% respectively. The phylogenetic tree revealed the evolutionary relationships of the KACC isolates with closely associated microalgae strains deposited in GENBANK and some new reports were also generated. The ribosomal sequences were deposited to GenBank (NCBI) to record our isolated microalgal strains.

Phylogenetic analysis was performed on a linked data set obtained from an NCBI BLAST search using nucleotide sequence of KACC isolates. For the cultures which are difficult to identify the organism using single marker gene (18S rRNA), the other markers like 28S rRNA and ITS genes were used. In our study, 28S rRNA was used for 6 organisms and ITS for one organism. For identification of KACC 2, 18S, 28S, ITS genes were used in the study. Fig. 2A and 2B showed the phylogenetic identification of KACC 2 microalgae using 18S rRNA gene and ITS gene respectively along with the BARCODE data obtained from the BOLD. Phylogenetic evidence using 28S rRNA was depicted in Fig. 3 for the culture KACC 2, 5, 9, 10, 14, 15, 16, and 20. The KACC identified cultures with closest similarity and close homology in phylogenetic tree was depicted in the Table 2. The taxonomical classification of identified freshwater microalgae KACC isolates were given in detail in Supplementary table S1. The Phylogenetic tree with DNA barcode of 18S rRNA sequences for all the KACC isolates was provided as supplementary Fig. S5 – S21.

Table 2

GenBank accession numbers for KACC microalgae cultures with closest relative sequence and their similarity

Code	Marker (base pairs)	Closest relative	Similarity %	Phylogeny close homology	Present study Culture identified as	Accession Numbers
KACC 1	18S (911bp)	<i>Kalenjinia gelatinosa</i> (GQ477061)	99	<i>Kalenjinia</i> sp. (KU361145)	<i>Kalenjinia gelatinosa</i>	MF693860
KACC 2	18S (893 bp)	<i>Asterarcys quadricellularae</i> (KT991532) <i>Scenedesmus</i> sp. (KT279490), <i>Pectinodesmus</i> sp. (AB917099), <i>Scenedesmus regular</i> (FR865732)	99	<i>Asterarcys quadricellularae</i> (KT991532) <i>Scenedesmus</i> sp. (KT279490), <i>Pectinodesmus</i> sp. (AB917099), <i>Scenedesmus regular</i> (FR865732)	<i>Scenedesmus rubescens</i>	MG009232
	C- 18S (1236 bp)	<i>Tetradesmus obliquus</i> (MH137236), <i>Scenedesmus</i> sp. (MH277469) and <i>Coelastrella saipanensis</i> (MF407353)	100	<i>Tetradesmus obliquus</i> (MH137236), <i>Scenedesmus</i> sp. (MH277469) and <i>Coelastrella saipanensis</i> (MF407353)	<i>Scenedesmus rubescens</i>	MG009234
	ITS (541 bp)	<i>Scenedesmus</i> sp (KT932960), <i>Scenedesmus</i> sp (JX519261), <i>Scenedesmus</i> sp (JQ782746) and <i>Chlorophyta</i> sp. (JX046434)	100	<i>Scenedesmus</i> sp (KT932960), <i>Scenedesmus</i> sp (JX519261), <i>Scenedesmus</i> sp (JQ782746) and <i>Chlorophyta</i> sp. (JX046434)	<i>Scenedesmus rubescens</i>	MF967210
KACC 3	18S (992 bp)	<i>Chlorella minutissima</i> (KJ849830), <i>Neodesmus pupukensis</i> (MG257914), <i>Neodesmus pupukensis</i> (MF523241)	94	<i>S. pupukensis</i>	<i>Neodesmus pupukensis</i>	MF693861

Code	Marker (base pairs)	Closest relative	Similarity %	Phylogeny close homology	Present study Culture identified as	Accession Numbers
KACC 4	18S (917 bp)	<i>Coelastrella</i> sp. (MG257917), <i>Coelastrella</i> sp (JQ731679), <i>Tetradesmus</i> sp. (JQ796860) and <i>Scenedesmus</i> sp. (KC342184)	100, 98, 98, 98	<i>Coelastrella</i> sp.	<i>Coelastrella</i> sp.	MF693862
KACC 5	18S (574 bp)	<i>Myrmecia picta</i> (AY703498), <i>Myrmecia</i> <i>pilosula</i> (AY703498), <i>Myrmecia</i> <i>fulvipes</i> (AH012892)	86	<i>Myrmecia picta</i>	<i>Myrmecia</i> <i>incisa</i>	MF693863
	28S (411 bp)	uncultured organism clone (KC284764)	96	uncultured organism clone	<i>Myrmecia</i> <i>incisa</i>	MF716572
KACC 6	18S (1310 bp)	<i>Scenedesmus</i> sp. (KY816917), <i>Scenedesmus</i> sp. (KY315600), <i>Scenedesmus</i> sp. (JX286515)	99 98 92	<i>Scenedesmus</i> sp.	<i>Scenedesmus</i> <i>acutus</i>	MF693864
KACC 7	18S (402 bp)	<i>Scenedesmus</i> sp. (KY315600), <i>Tetranebris</i> <i>brasiliensis</i> (MF523234), <i>Tetradesmus</i> <i>brasiliensis</i> (MG971386)	99	<i>Tetranebris</i> <i>brasiliensis</i>	<i>Desmodesmus</i> <i>dispar</i>	MF693865
KACC 8	18S (465 bp)	<i>Monoraphidium</i> sp. (KT598261), <i>Monoraphidium</i> <i>griffithi</i> (KP726255), <i>Monoraphidium</i> sp. (KT279485)	95	<i>Monoraphidium</i> <i>contortum</i>	<i>Monoraphidium</i> <i>contortum</i>	MF693866
KACC 9	18S (589 bp)	<i>Scenedesmus</i> sp	98	<i>Scenedesmus</i> sp.	<i>Scenedesmus</i> <i>acunae</i>	MF693867

Code	Marker (base pairs)	Closest relative	Similarity %	Phylogeny close homology	Present study Culture identified as	Accession Numbers
KACC 10	18S (882 bp)	<i>Nanochloris</i> sp. (AB080306),	96	Nanochloris sp	<i>Nannochloris</i> sp	MF693868
		<i>Nanochloris normandiane</i> (KM020167),	95			
		<i>Nanochloris</i> sp. (AY195968)	94			
	28S (411 bp)	<i>uncultured organism clone</i> (JX877610)	99	uncultured organism clone	<i>Nannochloris</i> sp	MF716574
KACC 11	18S (979 bp)	<i>Chlorococcum lobatum</i> (AB936289),	94	Chlorococcum sp.	<i>Chlorococcum</i> sp.	MF693869
		<i>Chlorococcum</i> sp. (KY816916),				
		<i>Chlorococcum</i> sp. (KP726224)				
KACC 12	18S (403 bp)	<i>Brachysira</i> sp. (KU951588),	94	Brachysira sp.	<i>Brachysira exilis</i>	MF693870
		<i>Nupela</i> sp. (KU951548),				
		<i>Nupela</i> sp. (KU951583)				
KACC 13	18S (724 bp)	<i>Desmodesmus intermedius</i> (KP10179),	92	Chlorococcum sp.	<i>Chlorococcum</i> sp.	MF693871
		<i>Scenedesmus naegeli</i> (KC218499),	92			
		<i>Chlorococcum</i> sp. (KC218477)	91			
KACC 14	18S (562 bp)	<i>Nitzschia</i> sp. (MH063483),	86	Nitzschia dissipata	<i>Nitzschia paleacea</i>	MF693872
		<i>Nitzschia dissipata</i> (AJ867018),				
		<i>Nitzschia spathulata</i> (MH063491)				
	28S (330 bp)	<i>Chloromonas perforata</i> (KT625416)	97	uncultured organism clone	<i>Nitzschia paleacea</i>	MF716575
		uncultured organism clone (KY919359)				

Code	Marker (base pairs)	Closest relative	Similarity %	Phylogeny close homology	Present study Culture identified as	Accession Numbers
KACC 15	18S (421 bp)	<i>Kirchneriella lunaris</i> (KP726238), <i>Kirchneriella lunaris</i> (KP726236), <i>Kirchneriella obesa</i> (HM483513) <i>Kirchneriella dianae</i> (HM483512)	98	<i>Kirchneriella obesa</i>	<i>Kirchneriella incurvata</i>	MF693873
	28S (326 bp)	<i>Kirchneriella aperta</i> (KT199250)	85	<i>Kirchneriella aperta</i>	<i>Kirchneriella incurvata</i>	MF716576
KACC 16	28S (334)	<i>Ankistrodesmus stipitatus</i>	82	<i>Ankistrodesmus stipitatus</i>	<i>Ankistrodesmus sp.</i>	MF716577
KACC 18	18S (425 bp)	<i>Tetraedron caudatum</i> (AY780669), <i>Tetraedron bitridens</i> (AY663043), <i>Tetraedron pentaedricum</i> (AY780671) <i>Tetraedron minimum</i> (AY780670)	99	<i>Tetraedron bitridens</i>	<i>Tetraedron caudatum</i>	MF693874
KACC 19	18S (1014 bp)	<i>Graesiella emersonii</i> (MG022718), <i>Graesiella emersonii</i> (FR865687), <i>Graesiella vacuolata</i> (FR865685) <i>Chlorella emersonii</i> (FR865661)	99	<i>Graesiella emersonii</i> (MG022718), <i>Graesiella emersonii</i> (FR865687), <i>Graesiella vacuolata</i> (FR865685) and <i>Chlorella emersonii</i> (FR865661)	<i>Graesiella sp.</i>	MF693875

Code	Marker (base pairs)	Closest relative	Similarity %	Phylogeny close homology	Present study Culture identified as	Accession Numbers
KACC 20	18S (855 bp)	<i>Scenedesmus regularis</i> (AB037095), <i>Ankistrodesmus</i> sp (AY846374), <i>Chlorella emersoni</i> (KF673368) <i>Ankistrodesmus</i> sp (AY846372)	84	Ankistrodesmus sp. Scenedesmus sp.	<i>Scenedesmus</i> sp.	MF693876
	28S (347 bp)	<i>Chloromonas perforata</i> (KT625416) uncultured organism clone (KY919388)	92	uncultured organism clone	<i>Scenedesmus</i> sp.	MF716578
KACC 21	18S (1018 bp)	<i>Chlorella emersonii</i> (FR865657), <i>Chlorella</i> sp (KM985412), <i>Chlorella</i> sp. (KM985403)	99	<i>Chlorella emersonii</i> (FR865657), <i>Chlorella</i> sp (KM985412), <i>Chlorella</i> sp. (KM985403)	<i>Chlorella neustonica</i>	MF693877
KACC 22	18S (427 bp)	<i>Asterarcys quadricellulare</i> (MH159122), <i>Chlorella</i> sp. (KM985412), <i>Chlorella</i> sp. (KM985403)	100	<i>Chlorella</i> sp.	<i>Chlorella</i> sp.	MF693878
KACC 23	18S (850 bp)	<i>Elliptochloris</i> sp. (HQ317309), <i>Elliptochloriis</i> sp (FJ217365), <i>Elliptochloris</i> sp. (FJ217360)	96	Elliptochloriis sp	<i>Elliptochloris</i> sp.	MF693879

Generation of DNA BARCODE database

Database for 23 fresh water microalgae isolated from Siruvani waters of Western Ghats, India have been prepared. Datasheet comprising the strain details, taxonomical information, and geographical details were prepared for the respective microalgae stains. Partial sequences of 18S rRNA, 28S rRNA, and ITS regions along with NCBI GenBank accession numbers of KACC microalgae were submitted in the BOLD database

and BOLD accession numbers were obtained. Sequence data and Barcode data depicted as color-coded nucleotide barcodes, generated using BOLD is presented (Fig. 2 & 3). Further, a detailed account of the magnitude of data currently available on microalgae at the national and global level as made. Accordingly, the DNA sequence data generated for the barcode regions of microalgae had been submitted to the NCBI and BOLD databases. Nucleotide and barcode sequence data for each species and the statistics data available with respect to contributions made in microalgae research community worldwide are represented (Fig. S22A-E).

Discussion

Our earlier study reported the diversity of microalgae species in catchment region of river Noyyal and their response to the seasonal variation in the physico – chemical characteristics of water (Kumar and Thomas, 2019). We also reported that microalgae were characterized by high H' index, lower Simpson dominance value and Margalef index value with indefinite patterns of annual variations. In the current study, the KACC cultures were identified using morphological and molecular traits, the identified cultures were authenticated and validated using four colored DNA barcode. In general, the morphological features were able to differentiate the microalgae between the genus. Microscopic observation of KACC microalga isolate revealed the difference in the morphological patterns and the isolates were identified based on the taxonomical keynote (Table 1). Since there was an ambiguity in the identification of those isolates to the genus and species level, the isolates were grouped into 10 different families based on their unique morphological features. But ascertaining of morphological heterogeneity of the microalgae through microscopic examination alone was difficult. Thus, all the isolated limnic microalgae were identified by molecular tools using rRNA as capable molecular markers for species-level characterization. Molecular phylogenetic analysis was carried out using 18S ribosomal RNA sequences, 28S rRNA and ITS sequences which are most commonly studied for DNA barcoding in microalgae. Phylogenetic trees were derived with the help of NJ algorithms (Saitou and Nei, 1987). rRNA gene present in high copy numbers, with increased sensitivity of detection and thus they serve as a good candidates gene for identification of microalgae. Among the rRNA genes, SSU rRNA (18S) gene has a higher evolutionary rate and should be a better tool for the discrimination of closely related species using short diagnostic sequences (Abou-Shanab et al. 2011). They are highly conserved and easier for amplification by using universal primer; thus it became an important marker for higher ranks such as classes and orders (Andersen et al. 1999). In the present study, for the cultures with ambiguity in identification of species using 18S rRNA, other regions like 28S rRNA (LSU) and ITS were used for exact identification. The cultures KACC 2, 5, 10, 14, 15 and 20 showed vagueness in identification with 18S rRNA sequences. So LSU 28S rRNA was used for identification of species (Fig. 3). Similarly ITS region was also used for KACC 2, for which it was difficult to identify based on SSU and LSU ribosomal sequences. ITS sequence mainly composed of ITS1, 5.8S, and ITS2, has a high variation, so it was considered as an effective region to analyse the phylogenetic relationships of taxa below the level of the genera. The identified microalgae sequences with their code, specific accession numbers and the proximity to nearest identifiable match in the GenBank nucleotide database were shown in Table 3.

Our phylogenetic analysis showed that the newly isolated strain for KACC 22 has a deeply rooted arm, distant from the main *Chlorella* species. It was connected with some of the other strains of uncertain taxonomy. Therefore, more work on the molecular and physiological level would be required to firmly establish the taxonomic evolution of the microalgae strain. So, the culture KACC 22 was named as *Chlorella* sp. KACC 22. The same was observed for *Coelastrella* sp. KACC 4, *Nannochloris* sp. KACC 10, *Chlorococcum* sp. KACC 11, *Ankistrodesmus* sp. KACC 16, *Graesiella* sp. KACC 19, *Scenedesmus* sp. KACC 20 and *Elliptochloris* sp. KACC 23. The detailed taxonomical nomenclature of microalgae cultures were tabulated in Table S1. Although much biological research depends upon species diagnoses, taxonomic expertise is collapsing. We are convinced that the sole prospect for a sustainable identification capability lies in the construction of systems that employ DNA sequences as taxon 'barcodes'. We establish that the 18S, 28S rRNA sequences can serve as the core of a global bio identification system for microalgae.

The present work is the pioneer attempt to use DNA Barcoding techniques to confirm the identity of freshwater algae from India. Twenty-three species of microalgae from freshwater of Siruvani were subjected to sequence analysis in order to confirm the genetic identity of these taxa. The samples obtained from freshwater of Siruvani showed slight morphological differences like the increased length of the filaments, cell length and size from the already reported species. The morphological observations further corroborated molecular phylogenetic identification of freshwater Indian isolates.

Submission to BOLD helped to compare the volume of data collected with that of already available in the database. Our submissions of barcode sequences of microalgae to the database revealed that few of our submissions were reported to be new and adds to the contributions of various other institutions across the world for the first time to the BOLD (Fig S22). This gave the work more significance as it made known the need and importance of the study and proved to be a task worth pursuing. The obtained sequences were further used to narrow down on the most variant and thus suitable candidate barcode by using various sequence analysis software viz. MEGA 7.0, ClustalW, BioEdit, etc were identified. This data is also now available for further study on microalgae found elsewhere in the world since geographical variation in the barcode regions has been found to be highly negligible (Lahaye et al. 2007). The DNA sequence data can also be used by regulatory agencies for species identification of these freshwater microalgae species.

Conclusion

In order to reveal the species diversity, evolutionary history and geographic distributions of microalgae species of Siruvani freshwater, all the cultures were first identified based on morphological and molecular characteristics. In our study, phylogenetic and systematic analysis of microalgae was determined by a molecular approach using the sequencing of 18S, 28S rRNA and ITS regions. We have established that species-level identifications can be achieved, using sequence analysis by providing a phylogeny for these microalgae. The barcodes were developed for the identified freshwater microalgae determined by a molecular approach using the sequence data. Hopefully, this study can serve as a pioneer step towards species-level molecular research in freshwater microalgae complex.

Declarations

Acknowledgements

The authors gratefully acknowledge the financial support funded by Science and Engineering Research Board – Department of Science and Technology, (SERB – DST), Government of India (S.O: No/SB/FT/LS-389/2012) and Karunya Institute of Technology and Sciences for providing the research facilities.

Author contributions

PS and JT conceived the research and designed the experiments. PS performed the experiments, analyzed and interpreted the data. BK interpreted the morphological identification of algae. PS wrote the original draft. JT supervised and edited the manuscript. All authors edited and approved the final manuscript.

Compliance with Ethical Standards

Disclosure of potential conflicts of interest

The authors declare no conflict of interest.

Research involving human participants and/or animals

No conflicts, or human or animal rights are applicable to this study

Informed consent

No informed consent applicable in this study.

Declaration of the authors' agreement

All authors agreed to the authorship and submission of the manuscript to Environmental Science and Pollution Research for peer review.

Availability of Data and materials

The data availability will be provided on request.

References

1. Abou-Shanab RAI, Hwang JH, Cho Y, Min B, Jeon BH (2011) Characterization of microalgal species isolated from freshwater bodies as a potential source for biodiesel production. *Appl Energy* 88:3300–3306
2. Ancona-canché K, López-adrián S, Espinosa-aguilar M, Garduño-solórzano G, Toledano-thompson T, Narváezzapata J, Valdez-ojeda R (2017) Molecular phylogeny and morphologic data of strains of the genus *Coelastrella* (Chlorophyta, Scenedesmaceae) from a tropical region in North America (Yucatan Peninsula). *Bot Sci* 95(3):527–537

3. Andersen RA, Peer YVD, Potter D, Sexton JP, Kawachi M, Lajeunesse T (1999) Phylogenetic Analysis of the SSU rRNA from Members of the Chrysophyceae. *Protist* 150:71–84
4. Doolittle WF, Brown JR (1994) Tempo, mode, the progenote, and the universal root. *Proc Natl Acad Sci USA* 91:6271–6728
5. Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* 19:11–15
6. Driver T, Bajhaiya AK, Allwood JW, Goodacre R, Pittman JK, Dean AP (2015) Metabolic responses of eukaryotic microalgae to environmental stress limit the ability of FT-IR spectroscopy for species identification. *Algal Res* 11:148–155
7. Fawley MW, Fawley KP, Hoffman MJ (2002) Diversity of coccoid algae in Minnesota and North Dakota lakes. *J Phycol* 38:8
8. Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*;783–791
9. Guiry MD, Guiry GM (2018) *AlgaeBase*. World-wide electronic publication, National University of Ireland, Galway. ; <http://www.algaebase.org> Last accessed on 16 June 2020
10. Gunderson JH, Elwood H, Ingold A, Kindle K, Sogin ML (1987) Phylogenetic relationship between Chlorophytes, Chrysophytes and Oomycetes. *Proc Nat Acad Sci USA* 4(84):5823–5827
11. Hadi SIA, Santana H, Brunale PPM, Gomes TG, Oliveira MD, Matthiensen A, Oliveira MEC, Silva FCP, Brasil BSAF (2016) DNA Barcoding Green Microalgae Isolated from Neotropical Inland Waters. *PLOS One*;1–18
12. Hall JD, Lewis LA, Karol KG (2010) An assessment of proposed DNA barcodes in freshwater green algae. *Cryptogamie Algologie* 31(4):529–555
13. Huss VAR, Frank C, Hartmann EC, Hirmer M, Kloboucek A, Seidel BM, Wenzeler P, Kessler E (1999) Biochemical taxonomy and molecular phylogeny of the genus *Chlorella* Sensu Lato (Chlorophyta). *J Phycol* 35:587–598
14. Kessler E, Schäfer M, Hümmer C, Kloboucek A, Huss VAR (1997) Physiological, biochemical, and molecular characters for the taxonomy of the subgenera of *scenedesmus* (Chlorococcales, Chlorophyta). *Bot Acta* 110(3):244–250
15. Krienitz L, Hegewald EH, Hepperle D, Huss VAR, Rohrs T, Wolf M (2004) Phylogenetic relationship of *Chlorella* and *Parachlorella* gov.nih. (Chlorophyta, Trebouxiophyceae). *Phycologia* 43:529–542
16. Krienitz L, Ustinova I, Friedl T, Huss VRR (2001) Traditional generic concepts versus 18s rRNA gene phylogeny in the green algal family Selenastraceae (Chlorophyceae, Chlorophyta). *J Phycol* 37:852–865
17. Kumar PS, Thomas J (2019) Seasonal distribution and population dynamics of limnic microalgae and their association with physico-chemical parameters of river Noyyal through multivariate statistical analysis. *Sci Rep* 9(1):15021
18. Lahaye R, Bank M, Bogarin D, Warner J, Pupulin F, Gigot G, Maurin O, Duthoit S, Barraclough TG, Savolainen V (2007) DNA barcoding the floras of biodiversity hotspots. *Proc Natl Acad Sci USA*

105(8):2923–2928

19. Lo´pez-Bautista JM, Rindi F, Guiry MD (2006) Molecular systematics of the subaerial green algal order Trentepohliales: an assessment based on morphological and molecular data. *Int J Syst Evol Microbiol*;1709–1715
20. Luo W, Pflugmacher S, Proschold T, Walz N, Krienitz L (2006) Genotype versus phenotype variability in *Chlorella* and *Miractinium* (Chlorophyta, Treboxiophyceae). *Protist* 157:315–333
21. McDevit DC, Saunders GW (2010) A DNA barcode examination of the Laminariaceae (Phaeophyceae) in Canada reveals novel biogeographical and evolutionary insights. *Phycologia* 49(3):235–248
22. Metzger P, Largeau C (2005) *Botryococcus braunii*: A rich source for hydrocarbon and related ether lipids. *Appl Microb Biotechnol* 66(5):486–496
23. Piligaev AV, Sorokina KN, Bryanskaya AV, Peltek SE, Kolchanov NA, Parmon VN (2015) Isolation of prospective microalgal strains with high saturated fatty acid content for biofuel production. *Algal Res* 12:368–376
24. Ratnasingham S, Hebert PDN (2007) BOLD: the barcode of life data system. *Mol Ecol Notes* 7:355–364
25. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425
26. Sogin ML (1989) Evolution of eukaryotic microorganisms and their small subunit ribosomal RNAs. *Amer Zool* 29:487–499
27. Songdong S (2008) Genetic diversity analysis with ISSR PCR on green algae *Chlorella vulgaris* and *Chlorella pyrenoidosa*. *Chin J Oceanol and Limnol* 26(4):380–384
28. SureshKumar P, Thomas J, SivaSubramanian J (2017) Rapid method for simultaneous discrimination of microalgae and determination on biochemical composition based on vibrational spectroscopy. *Int J Algae* 19(4):385–396
29. Sureshkumar P, Thomas J (2020) Exploring the distinctiveness of biomass and biomolecules from limnic microalgae of unexplored waters of Noyyal River, Western Ghats, for exploitation. *Environ Sci Pollut Res* 27:23309–23322
30. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596–1599
31. Wilcox LW, Lewis LA, Fuerst PA, Floyd GL (1992) Assessing the relationships of autosporic and zoosporic Chlorococcalean green algae with 18s rRNA sequence data. *J Phycol* 28:381–386
32. Woese CR (1987) Bacterial evolution. *Microbiol Rev* 51(2):221–271
33. Zhu H, Zhao Z, Xia S, Hu Z, Liu G (2015) Morphological examination and phylogenetic analyses of *phycopeltis* spp (trentepohliales, ulvophyceae) from tropical china. *PLoS ONE* 10(2):1–19
34. Zou S, Fei C, Wang C, Gao Z, Bao Y, He M, Wang C (2016) How DNA barcoding can be more effective in microalgae identification: a case of cryptic diversity revelation in *Scenedesmus* (Chlorophyceae). *Sci rep* 6:36822

Figures

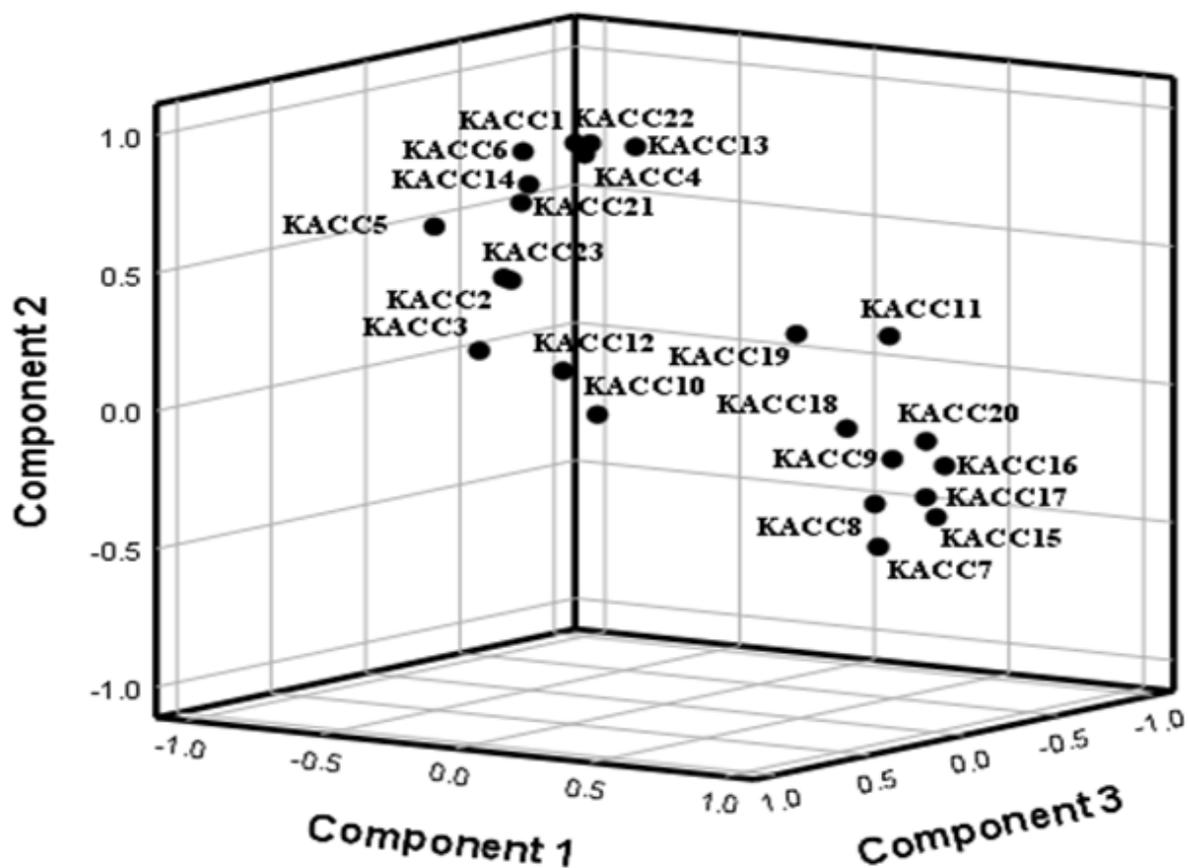


Figure 1

PCA discrimination of KACC microalgae strains based on biochemical profile using FT-IR spectra

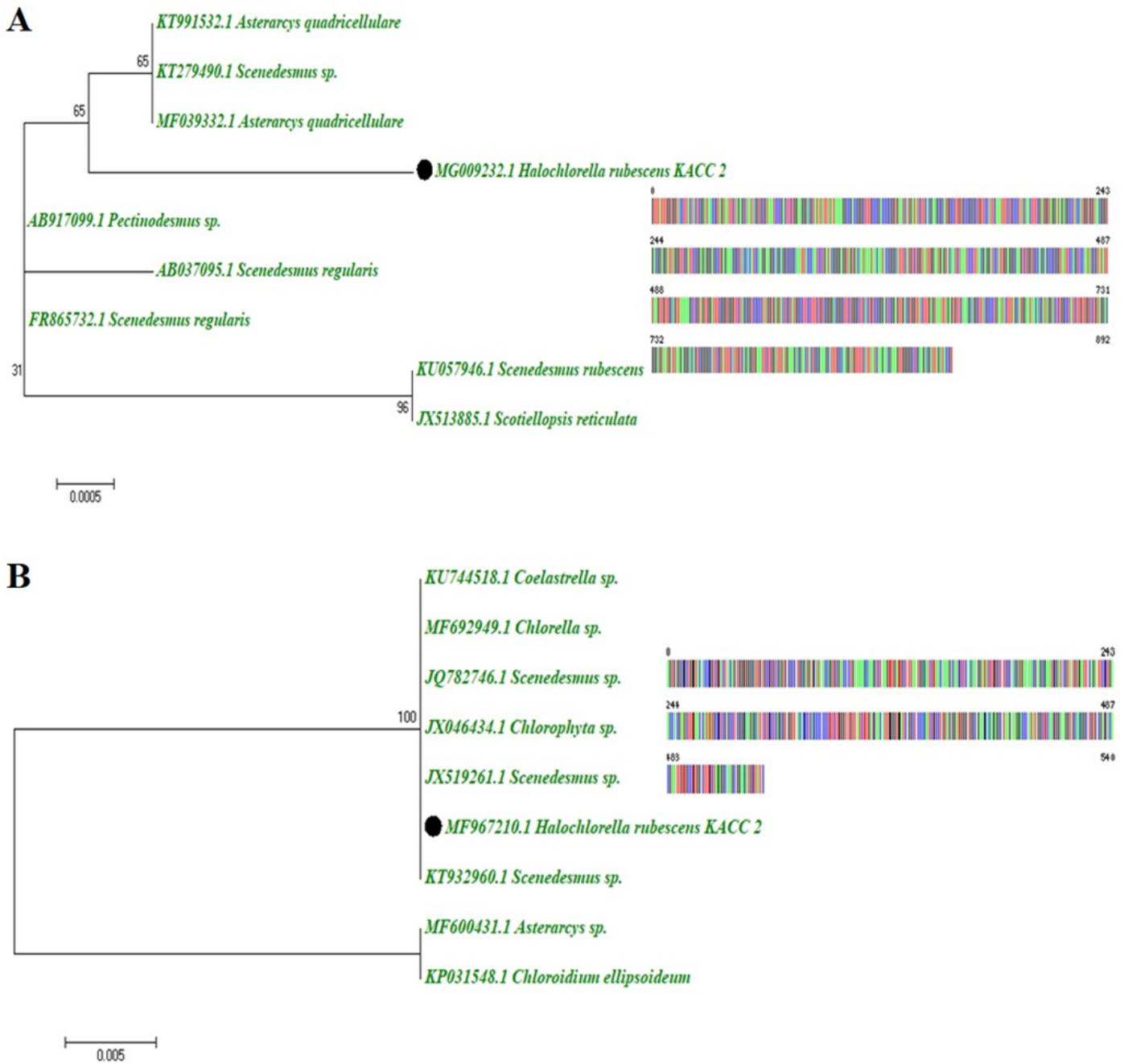


Figure 2

Evolutionary relationship of KACC microalgae isolates of KACC 2 with DNA barcode of A. 18S rRNA and B. ITS sequences.

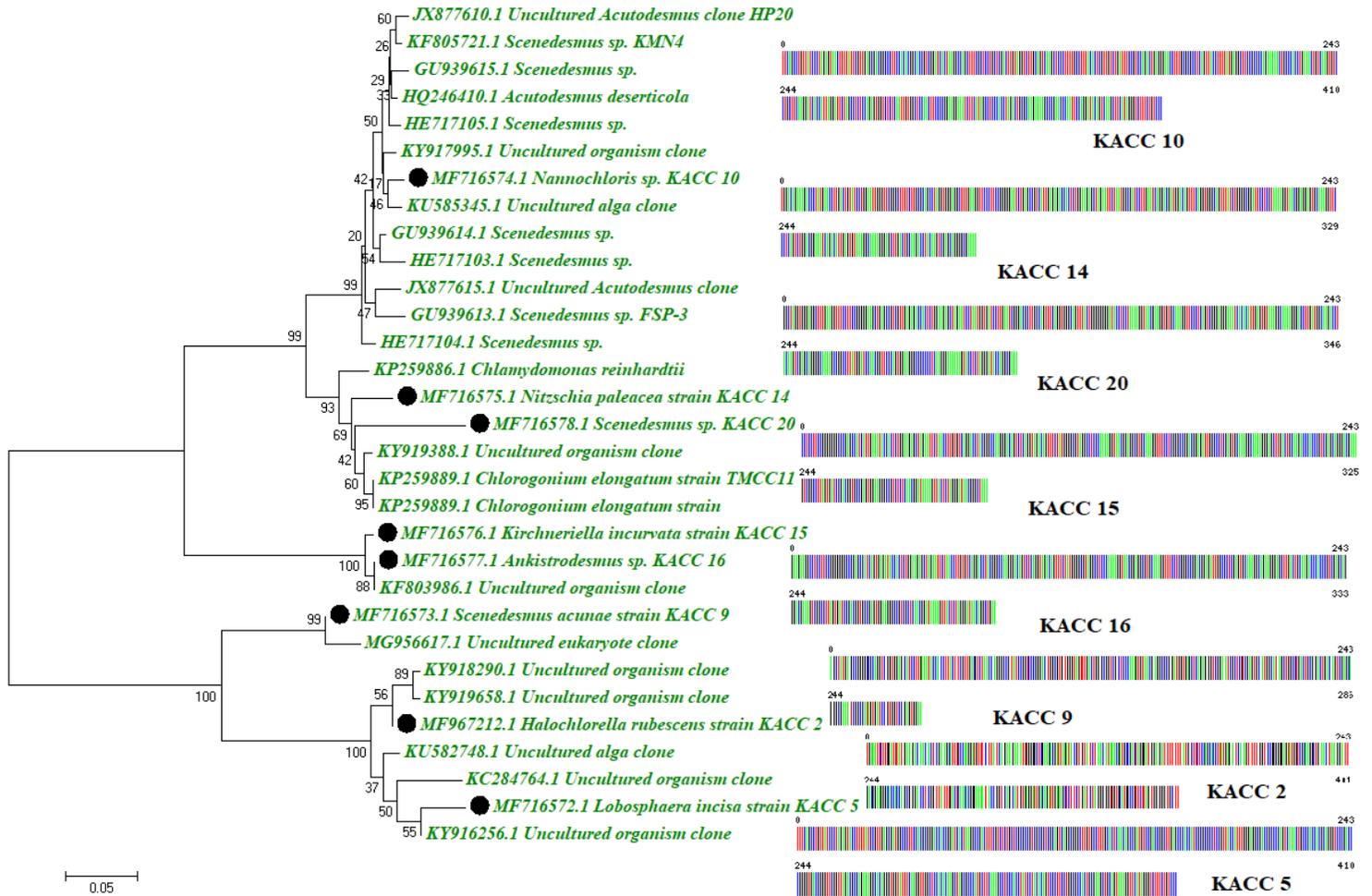


Figure 3

Evolutionary relationship of KACC microalgae isolates with DNA barcode of 28S rRNA sequences

Supplementary Files

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

- [Supplementarydata.docx](#)