Bull. Iraq nat. Hist. Mus.
(2020) 16 (1): 39- 61. https://doi.org/10.26842/binhm.7.2020.16.1.0039

FIVE DIATOM SPECIES IDENTIFIED BY USING POTENTIAL APPLICATION OF NEXT GENERATION DNA SEQUENCING

Warqaa Yehia Al-Meshhdany* Fikrat M. Hassan**

*Institute of Genetic Engineering and Biotechnology for Postgraduate Studies, University of Baghdad, Baghdad, Iraq

**Department of Biology, College of Science for Women University of Baghdad, Baghdad, Iraq

**corresponding author: fikrat@csw.uobaghdad.edu.iq

Received Date: 28 January 2020; Accepted Date: 06 April 2020; Published Date: 24 June 2020

ABSTRACT

Molecular barcoding was widely recognized as a powerful tool for the identification of organisms during the past decade**;** the aim of this study is to use the molecular approach to identify the diatoms by using the environmental DNA. The diatom specimens were taken from Tigris River. The environmental DNA(e DNA) extraction and analysis of sequences using the Next Generation Sequencing (NGS) method showed the highest percentage of epipelic diatom genera including *Achnanthidium minutissimum* (Kützing) Czarnecki, 1994 (21.1%), *Cocconeis placentula* Ehrenberg, 1838 (%21.3) and *Nitzschia palea* (Kützing) W. Smith, 1856 (16.3%).

Five species of diatoms: *Achnanthidium minutissimum*; *Fistulifera saprophila* (Lange-Bertalot & Bonik) Lange-Bertalot, 1997; *Gomphonema pumilum* (Grunow) E. Reichardt & Lange-Bertalot, 1991; *Navicula veneta* Kützing, 1844 and *Thalassiosira pseudonana* Hasle Heimdal, 1970 were registered in NCBI under the accession numbers as follows: MN749640.1, MN749641.1, MN749642.1, MN749643.1 and MN749646.1 for the first time; while the two algae *Fistulifera saprophila* and *Thalassiosira pseudonana* are regarded as a new record to algal flora in Iraq.

The environmental DNA study will be a catalyst for new studies of biodiversity and environmental studies in Iraq and the region.

Keywords: Algae, Diatoms, Edema, Freshwater, NGS, Tigris River.

INTRODUCTION

Significant environmental problems are caused by rapid population growth in the world; lack of environmental knowledge in society and changes in the industry, particularly during

the last century. Freshwater habitats are without doubt one of the biosphere elements most impacted by this pollution. Monitoring water quality is therefore essential to the health of the water ecosystem's sustainability and protection (Çiçek *et al*., 2013; Campbell *et al*., 2017). Monitoring of water quality by physical and chemical methods is inadequate; in the recent years, particularly in the scientific community, the biological monitoring methods and biological indicator organisms were widely used for effective research (Chang, 2008; Tokatlı and Dayıoğlu, 2011; Adebayo *et al*., 2013; Berthold *et al*., 2018).

Diatoms are considered to be a large part of the benthos (often 90–95 percent), and are present all the time in all surface waters. They are also one of the most important groups of aquatic producers and react quickly to the environmental variables change. Diatoms, which are recognized as an important component of bioindicator species, have, therefore, been used as water pollution indicators for environmental assessments in many countries (Gürbüz and Kivrak, 2002; Passy *et al*., 2004; Godhe and Härnström, 2010; Aydın and Büyükışık, 2014; Tan *et al*., 2017).

Recently, scientists and researchers can use a basic reality to obtain information and produce more informed choices; this material persists, giving insight into the creature's past and present that left it behind. The eDNA samples were taken from different environments and for that it called environmental DNA (Thomsen and Willerslev, 2015).

During the past decade, molecular barcoding has been widely recognized as a powerful tool for identifying species. The assumption is that there is sufficient information in a short DNA sequence (DNA barcode) to identify the organisms. The major advantage in design studies of the use of DNA barcodes is that standardization and process implementation are simpler than the conventional morphology-based approach (Gao, 2019).

Metabarcoding, which refers to the employment of universal primers for the amplification of DNA from various organisms collected in one sample, is the approach that is most commonly applied in Next Generation Sequencing (NGS) (Taberlet *et al*., 2012). NGS approaches are being increasingly employed to characterize water-living organisms from the eDNA specimens (Yu *et al*., 2015).

Current advances in NGS approaches have made it possible to employ molecular barcoding in readily and efficiently investigating the diversity in the environment. The NGS -based environmental monitoring has been shown to be of less time and cost-consuming as compared to the conventional morphology-based methods (Baird and Hajibabaei, 2012). It is important to use the molecular concept as a solution to revise the mis-identification of diatoms and it could be also useful for biodiversity studies (Vasselon *et al*., 2017). The diatoms identification are often collected as a mixed species in taken sample and this is the main challenges for this purpose (Zimmerma *et al*., 2015).

Al-Rawi *et al.* (2018) reported that the traditional classification is not accurate to identify the algal species in Iraq and confirmed the use of molecular concept to identify algal taxa; Abed *et al*. (2018) used the molecular concept to identify the algal *Coelastrella* Chodat, 1922.

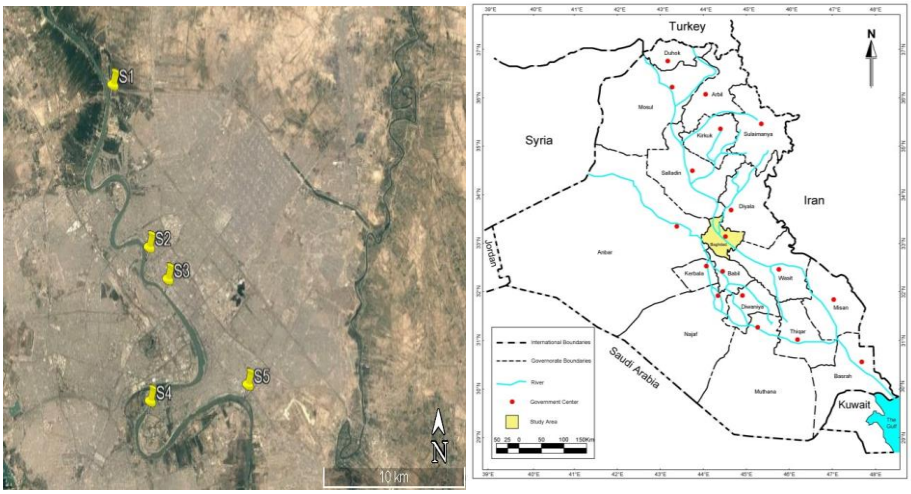
This study is aimed to assess the suitability of amplicon sequencing in Next Generation Sequencing (NGS) approach using Illumina platform for identifying epipelic diatoms in the sediment of the Tigris River for the evaluation and development of molecular biological methods in water quality.

MATERIAL AND METHODS

Specimens collection

Algal specimens were collected from five sites along the Tigris River from November 2018 to July 2019 (Map 1, Tab. 1); specimens of epipelic were collected randomly by scraping the clay from the surface layer with a depth 0.1-0.5 cm in area (50 m^2) and (3-5 mm) using a spatula, samples were placed in polyethylene bottles and sample water was added, the bottle was closed and shaken well and placed in a dark place until returning to the laboratory.

Epipelic diatoms were trapped by lens tissues as described by Eaton and Moss (Salman *et al*., 2017). Epipelic cell was identified after cleaning the silica skeleton by placing the glass slide on a heating plate (75-80°C), then placing a droplet of the sample on the slide and letto dry completely; followed by a concentrated nitric acid was added on the dry spot and the acid was left to evaporate completely. Then slides were mounted by Canada balsam and the cover was flipped on the dry spot and the lid of the slide was pressed gently to distribute the material in a homogeneous manner to avoid the emergence of bubbles near the edges of the sliding lid (Salman *et al*. 2017); diatoms were identified according to Round *et al* (1990).



Map (1): Map of study areas (Source: https://earth.google.com).

Five diatom species identified

Table (1): Geographical positions (GPS) of the five study sites.

No.	Symbol	Area	Coordinate	
			North	East
1	S1	Al-Muthanna Bridge	33°25'41.85″	44°20'49.63″
2	S2	Al-Sarafiya Bridge	33°2112.99″	44°22'28.77″
3	S3	Al-Shuhadaa Bridge	33°2019.99″	44°23'19.91″
4	S4	Al-Jadriya	33°16'58.35″	44°22'31.87″
5	S5	Al-Zafraniya	33°17'25.44″	44°26'58.23″

Culturing of diatoms

Diatoms were cultured by using the purchase F2 medium following Guillard (1975)method. Each epipelic diatom cell suspension was inoculated with 20 to 250 ml off/2+0.025 SiO_3 medium gradually, the culture was incubated under cycles of illumination with 12 h/12 h light-dark and constant temperatures 20°C (Al-Hussieny *et al*., 2014).One ml of culture was transferred to 1.5 ml tubes in the exponential growth phase (14-20 days of incubation), and the sample was centrifuged at 4000 xg for ten minutes. In the final step, the supernatant was discarded and the resulted pellets were stored at -20 °C, this step is to freeze the pellet in order to block the action of the enzymes like RNAase and protease. The pellets were kept for further use as recommended by Visco *et al*. (2015).

Molecular identification of Diatoms

In order to identify unknown diatoms at a molecular base, four genes were selected (XXXX) (Tab. 2). Primers were designed and manufactured in Macrogen company laboratories (Seoul, South Korea).

Table (2): Primers design used in this study.

Genes	Primer Sequence FWD 3`-5`	Primer Sequence Re 3`-5`	Reference
ITS3-ITS4	GCATCGATGAAGAACGCAGC	TCCTCCGCTTATTGATATGC	Moniz and Kaczmarska (2010)
18S V4F-V4R	CCAGCAGCCGCGGTAATTCC	ACTTTCGTTCTTGATTAA	Luddington *et al*. (2012)
18S V9F-V9R	CCCTGCCHTTTGTACACAC	CCTTCYGCAGGTTCACCTAC	Luddington *et al*. (2012)
D2/D3 of LSU rRNA	ACAAGTACCGTGAGGGAAAGTTG	TCGGAAGGAACCAGCTACTA	Hamsher *et al*. (2011)

Genomic DNA manipulation:

For DNA purification, the genomic DNA of 20 isolated samples of unknown diatoms were extracted according to the protocol of QIAamp DNA Mini Kit, QIAGEN, and the isolated DNA was subjected to PCR (Gene Amp, PCR system 9700; Applied Biosystem) according to manufacturer's instructions.

Multiplex polymerase chain reaction (PCR)

The total volume of PCR amplification reaction was performed 25µ land included10ng/µl DNA, (1X) Taq PCR PreMix (Intron, Korea), and 1µM of each primer, and then distilled water was added into the tubes. Conditions of the thermal cycling containing denaturation at 95 °C for 5 min, were followed by 30 cycles of 95 °C for 30s, 60 °C for 30s and 72 °C for 30s, with a final incubation at 72 °C for 7 min using a thermal Cycler (Gene Amp, PCR system 9700; Applied Biosystem). The PCR products were separated by 2% agarose gel electrophoresis and visualized by exposure to ultraviolet light (302 nm) after staining with red stain (Intron Korea). For PCR products, 10µl was directly loaded into the well. Electrical power was turned on at 100v/m Amp for 75 minutes and DNA was migrating from the Cathode to plus Anode poles. Ethidium bromide-stained bands in gel were visualized using gel imaging system.

For standard genes sequencing, PCR amplification of *18S rRNA* products of all isolated diatoms was sent to macrogen company laboratories for sequencing using the Illumina platform by Next Generation Sequencing (NGS) workflow, which includes 4 basic steps.

Calculating Phred Quality Scores (Q scores)

Q scores are a measure of the quality of the identification of the nucleobase generated by automated DNA sequencing, that is logarithmically related to the base call error probabilities (P)(Ewing and Green,1998).

$Q = -10 \log 10^{P}$

RESULTS AND DISCUSSION

A total of 186 epipelic diatoms taxa were identified according to the traditional concept (by compound microscopy model GX- 140105) which belong to a 59 genera according to (Round *et al*., 1990). The most abundant taxa are illustrated in Table (3).

Five diatom species identified

Table (3): The most abundant diatomic taxa (identified by compound microscope) during the study period.

Classes	**Taxa**
Bacillariophycaeae	*Achnanthidium minutissimum* (Kutzing) Czarnecki
	Cocconeis placentula Ehrenberg
	C. placentula var. *euglypta* (Ehrenberg) Grunow
	Gomphonema gracile Ehrenberg
	Nitzschia frustulum var. *minuta* Pantocsek
	Rhopalodia musculus (Kützing) O.Müller
Fragilariophyceae	*Fallacia enigmatica* (H. Germain) Lange-Bertalot & Werum
	Fragilaria intermedia (Grunow) Grunow
	F. pygmaea (Kützing) A. J. Stickle & D.G.Mann
Coscinodiscophyceae	*Aulacoseira granulata* (Ehrenberg) Simonen
	Melosira varians C.Agardh
	Pantocsekiella ocellata (Pantocsek) K.T.Kiss & E.Ács

Characterization of Diatoms by *18S rRNA* and eDNA

To unequivocally determine the diatoms in sediment samples, diatoms were isolated by two means through *18S rRNA* and eDNA. The gene of interest was screened for *18SrRNA* using different primer pairs (Table 4).The resulted PCR products of *18S rRNA* were obtained from unknown samples and analyzed on 2% agarose gel and subsequently sequenced by NGS. The PCR products were (778bp) for *A. minutissimum*, (877 bp) for *F. saprophila*, (1110 bp) *G. pumilum*, (679bp) *N. veneta*, and (484 bp) for *T. pseudonana* (Pl. 1).

Table (4): Data Statistics for diatoms.

Genes	Total read bases (bp)	Total reads	GC (%)	AT (%)	Q20 (%)	Q30 (%)
18S_V9FV9R	94,387,580	313,580	47.957	52.04	77.987	72.456
D2D3_LSU	98,594,356	327,556	50.856	49.14	89.683	78.397
ITS3_ITS4	102,630,164	340,964	43.632	56.37	95.028	88.079

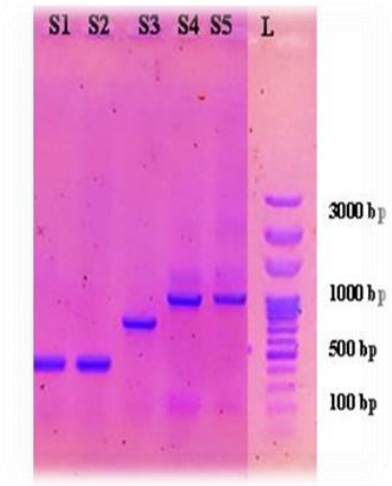


Plate (1):PCR products were electrophoresed on a 2% agarose gel (2 h., 5V/cm, 1X TBE) and visualized under U.V. light after staining Lane: L (M: 100bp ladder, S: sample. Lane S1, S2, S3, S4 and S5 represent the PCR products of isolated diatoms.

In Illumina MiSeq by NGS, the sequencing generated total number of bases sequenced, and total number of reads sequences, quinine- cytosine (GC %) content and adenine - thymine (AT%). As is explained table (4). While the high quality of the phred score for each gene sequences with an average Q20% and Q30% was illustrated in Diagram (1).

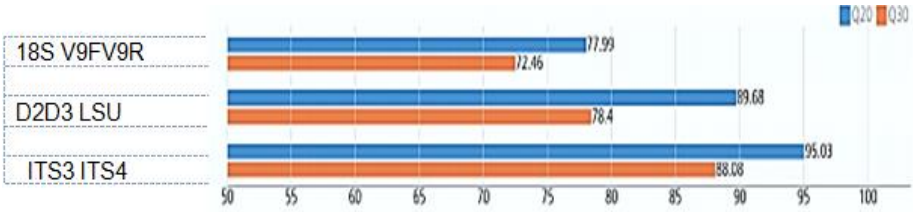


Diagram (1): Quality values line about sequences of the three genes with Q20/Q30 scores of sequences data.

The following diatom species were obtained with the relative abundance by the laboratories of Macrogen Corporation laboratories in Korea using Illumina platform by NGS for each DNA samples. These samples were identified by a three encoding genomic sequence described in previous table 2. *A. minutissimum* and *C. placentula* were diagnosed with the highest relative abundance with a slight difference (21.1and 21.3%), respectively, followed by *N. palea* with a percentage (16.3%), while the least abundance diatom was recorded for *N. cf. frustulum* with abundance of (0.7 %) (Tab. 5).

Five diatom species identified

Table (5): Relative abundance of diatom species by NGS

Taxa	Proportion (%)	Notes
Achnanthidium minutissimum	21.1	For the first time identified by molecular analysis in Iraq
Amphora montana	1.6	
Cocconeis placentula	21.3	
Cyclotella meneghiniana	2.3	
Fistulifera saprophila	1.9	New record in Iraq
Fragilaria pinnata	2.9	
Gomphonema parvulum	9.8	
Gomphonema pumilum	9.6	For the first time identified by molecular analysis in Iraq
Nitzschia amphibia	1.4	
Nitzschia cf. frustulum	0.7	
Nitzschia palea	16.3	
Navicula veneta	3.4	For the first time identified by molecular analysis in Iraq
Thalassiosira pseudonana	4.34	Unclear. *Thalassiosira pseudonana* is considered widespread. It is known from freshwater habitats (Kiss, 1984). New record in Iraq Confirm by Prof. Dr.Bahram K. Maulood (personal communication, March 14, 2020)
Ulnaria ulna	3.36	

The NGS sequencing were aligned online using Basic Local Alignment Search Tool (BLAST) at the National Center for Biotechnology Information (NCBI). The *18S rRNA* sequence of all diagnostic diatom samples showed 99% homology with other global diatoms registered in the NCBI under the accession numberin NCBIMN602030.1, MH997844.1, AM501970.1, KU900218.1, KC736629.1, respectively. The sequence analysis, types of polymorphism, location of nucleotide of *18S rRNA* gene for isolated diatoms were shown in Table (6) and demonstrated in (Diags. 2, 6).

Table (6): Types of polymorphism of *18S rRNA* gene from isolated diatoms.

No. of sample	Type of substitution	Location	Nucleotide	Sequence ID	Score	Identities	Taxa
1	Transition	762	**G>A**	ID:MN602030.1	1372	99%	***Achnanthidium minutissimum***
	Transition	977	**A>G**				
	Transvertion	1211	**C>G**				
	Transition	1231	**G>A**				
	Transition	1250	**A>G**				
	Transvertion	1354	**T>G**				
	Transition	1365	**A>G**				
2	Transition	554	**A>G**	ID:H997844.1	1564	99%	***Fistulifera saprophila***
	Transvertion	556	**T>A**				
	Transition	671	**A>G**				
	Transvertion	886	**C>A**				
3	Transvertion	556	**T>G**	ID:M501970.1	1198	99%	***Navicula veneta***
	Transvertion	711	**T>G**				
	Transvertion	735	**T>G**				
	Transition	765	**A>G**				
	Transition	792	**T>C**				
	Transvertion	849	**G>C**				
4	Transition	785	**A>G**	ID:KU900218.1	869	99%	***Thalassiosira pseudonana***
5	Transvertion	711	**G>C**	ID:KC736629.1	1994	99%	***Gomphonema pumilum***
	Transvertion	923	**G>C**				

Five diatom species identified

Achnanthidium minutissimum strain (*18S rRNA*) gene, partial sequence. Sequence ID:MN602030.1 Length: 1651Number of Matches: 1
Range 1: 644 to 1421Genbank Graphics Next Match Previous Match

Score	Expect	Identities	Gaps	Strand
1372 bits(1521)	**0.0**	**771/778(99%)**	**0/778(0%)**	**Plus/Plus**

Query61GTTCAAAGCAGGCTTATGCCGTTGAATGTCTTAGCATGGAATAATAAGAT
AGGACCTTAG120
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct704
GTTCAAAGCAGGCTTATGCCGTTGAATGTCTTAGCATGGAATAATAAGATAGGAC
CTTGG763
Query301
CCATCGTAGTCTTAACCATAAACTATGCCGACAGGGGATTGGTGGGGTTTCGTTA
CGTCT360
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct944
CCATCGTAGTCTTAACCATAAACTATGCCGACAAGGGATTGGTGGGGTTTCGTTA
CGTCT1003
Query541
TCTTTCTTGATTCTATGGGTGGTGGTGGATGGCCGTTCTTAGTTGGTAGAGTGATT
TGTC600
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct1184TCTTTCTTGATTCTATGGGTGGTGGTGCATGGCCGTTCTTAGTTGGTGGA
GTGATTTGTC1243

Query601
TGGTTAGTTCCGTTAACGAACGAGACCGCTGCCTGCTAAATAGTCCAGTGAGTGA
ATTTC660
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct1244TGGTTAATTCCGTTAACGAACGAGACCGCTGCCTGCTAAATAGTCCAGT
GAGTGAATTTC1303
Query661
ACTGACGAGGACTTCTTAGAGGGACGTGCGTTCTATTAGACGCAGGAAGAGAGC
GGCAAT720
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct1304ACTGACGAGGACTTCTTAGAGGGACGTGCGTTCTATTAGACGCAGGAAG
ATAGCGGCAAT1363
Query721
AGCAGGTCTGTGATGCCCTTAGATGTTCTGGGCCGCACGCGCGCTACACTGATGC
ATT778
|||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct1364AACAGGTCTGTGATGCCCTTAGATGTTCTGGGCCGCACGCGCGCTACAC
TGATGCATT1421

Diagram (2): Sequences analysis of *18S rRNA* gene for *Achnanthidium minutissimum.*

Fistulifera saprophila isolate HYU-D033 small subunit ribosomal RNA gene, partial sequence, Sequence ID: MH997844.1 Length: 1654Number of Matches: 1
Range 1: 305 to 1181Genbank Graphics Next MatchPrevious Match

Score	**Expect**	**Identities**	**Gaps**	**Strand**
1564 bits(1734)	**0.0**	**873/877(99%)**	**0/877(0%)**	**Plus/Plus**

Query241
CGTAGTTGGGTATGTGGTGTGCGTTGCGGCGTCCATTTGTTTGGTTCTGCCGTGAC
CGCG300
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct545
CGTAGTTGGATTTGTGGTGTGCGTTGCGGCGTCCATTTGTTTGGTTCTGCCGTGAC
CGCG604
Query361
CTGTGAGAAAATTAGAGTGTTCAAAGCAGGCTTATGCCGTTGAATATATTAGCAT
GGAAT420
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct665
CTGTGAAAAAATTAGAGTGTTCAAAGCAGGCTTATGCCGTTGAATATATTAGCAT
GGAAT724
Query541
GAACTACTGCGAAAGCATTTACCAAGGATGTTTTCATTAATAAAGAACGAAAGTT
AGGGG600
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct845
GAACTACTGCGAAAGCATTTACCAAGGATGTTTTCATTAATCAAGAACGAAAGTT
AGGGG904

Diagram (3): Sequences analysis of *18S rRNA* gene for *Fistulifera saprophila*

Naviculaveneta18S rRNA gene, strain AT-108Gel01 Sequence ID: AM501970.1
Length: 1745Number of Matches: 1
Range 1: 492 to 1170GenBank Graphics Next Match Previous Match

Score	Expect	Identities	Gaps	Strand
1198 bits(1328)	0.0	673/679(99%)	0/679(0%)	Plus/Plus

Query61CAGCGCCAATAGCGTATATTAAAGTTGTTGCAGTTAAAAAGCTCGTAGTT
GGATTTGTGG120

||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct552
CAGCTCCAATAGCGTATATTAAAGTTGTTGCAGTTAAAAAGCTCGTAGTTGGATT
TGTGG611

Query181
AACCTGTGTGGCATTAGGTTGTCGTGCAGGGGATGCCCAGCGTTTACTGTGAAAA
AATTA240

||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct672
AACCTGTGTGGCATTAGGTTGTCGTGCAGGGGATGCCCATCGTTTACTGTGAAAA
AATTA731

Query241
GAGGGTTCAAAGCAGGCTTATGCCGTTGAATATGTTAGCATGGAATAATGAGATA
GGACT300

|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct732
GAGTGTTCAAAGCAGGCTTATGCCGTTGAATATATTAGCATGGAATAATGAGATA
GGACT791

Query301
CTTTCGCTATTTTGTTGGTTTGCGCGAGAAGGTAATGATTAATAGGGACAGTTGG
GGCTA360

||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct792
TTTTCGCTATTTTGTTGGTTTGCGCGAGAAGGTAATGATTAATAGGGACAGTTGG
GGGTA851

Diagram (4): Sequences analysis of *18S rRNA* gene for *Navicula veneta*.

Thalassiosira pseudonana strain CCAP 1085/12 *18S ribosomal RNA* gene, partial sequence,
Sequence ID: KU900218.1 Length: 1755Number of Matches: 1
Range 1: 615 to 1098GenBankGraphics Next Match Previous Match

Score	Expect	Identities	Gaps	Strand
869 bits(963)	**0.0**	**483/484(99%)**	**0/484(0%)**	**Plus/Plus**

Query121
GGGATACCCATCGTTTACTGTGAAAAAATTAGAGTGTTTAAAGCAGGCTTGTGCC
GTTGA180

||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct735
GGGATACCCATCGTTTACTGTGAAAAAATTAGAGTGTTTAAAGCAGGCTTATGCC
GTTGA794

Diagram (5): Sequences analysis of *18S rRNA* gene for *Thalassiosira pseudonana*.

Gomphonema pumilum clone TCC536 18S ribosomal RNA gene, partial sequence
Sequence ID: KC736629.1 Length: 1683Number of Matches: 1
Range 1: 255 to 1364GenBank Graphics Next Match Previous Match

Score	Expect	Identities	Gaps	Strand
1994 bits(2210)	0.0	1108/1110(99%)	0/1110(0%)	Plus/Plus

Query421
ACGTTTACTGTGAAAAAATCAGCGCGTTCAAAGCAACCTTATGCTGTGAATGTAT
TAGCA480
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct675
ACGTTTACTGTGAAAAAATCAGCGCGTTCAAAGCAAGCTTATGCTGTGAATGTAT
TAGCA734
Query661
TAGGGGATCCAAGATGATTAGATACCATCGTAGTCTTAACCATAAACTATGCCGA
CAAGG720
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct915
TAGGGGATCGAAGATGATTAGATACCATCGTAGTCTTAACCATAAACTATGCCGA
CAAGG974

Diagram (6): Sequences analysis of *18S rRNA* gene for *Gomphonema pumilum*.

NGS data analysis

The results were analyzed using genius software. Sequencing of genes was performed by the Seoul National Instrumentation Center for Environmental Management (SNU NICEM) online at: http:/www.mbio.ncsu.edu/bioedit/bioedit.html, using a DNA sequencer 3730XL by Applied Biosystem. A homology search was conducted using Basic Local Alignment Search Tool (BLAST) program which is available at the National Center Biotechnology Information (NCBI) online at http://www.ncbi.nlm.nih.gov and software BioEditPro. Version: 7.0.0 program. An expected value is defined to give an estimation of the number of times expected to get the same similarity coincidental and the lower the value of expecting. This indicates that the degree of similarity was high between sequences which give greater confidence; a value close to zero means that these sequences are identical and the Bit Score, which is a statistical measure of the sequence similarity and the higher value indicates a high degree of similarity. Isolated diatom samples were confirmed by sequence-based phylogenetic tree (aligned sequences were conducted using MEGA 6 program) structuring analysis using 18S ribosomal RNA (*18SrRNA*) gene sequencing in Diagrams (7-11).

Five diatom species identified

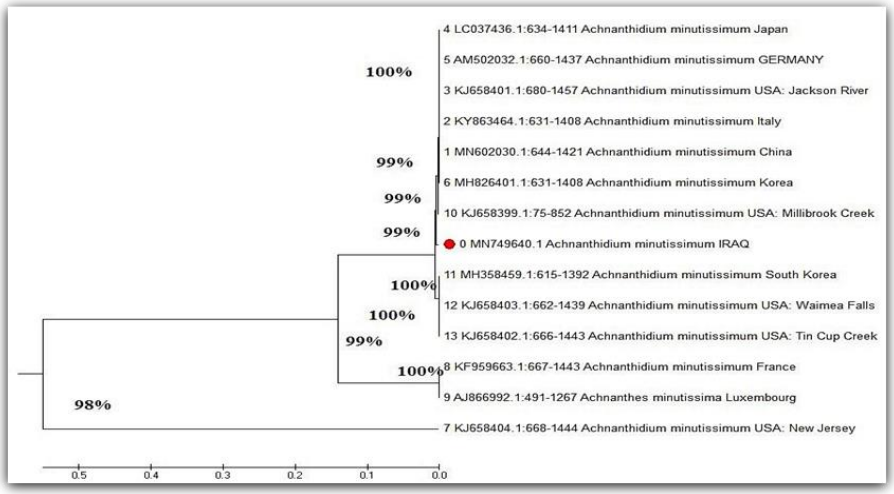


Diagram (7): Phylogenetic tree of *Achnanthidium minutissimum* based on *18S rRNA* gene sequences conferred by GeneBank data base, were analyzed and aligned through BLAST from NCBI using the Neighbor-Joining Analyses of 778bp of corresponding position of *18S rRNA* gene sequence. MEGA 6 program was used for phylogenetic tree.

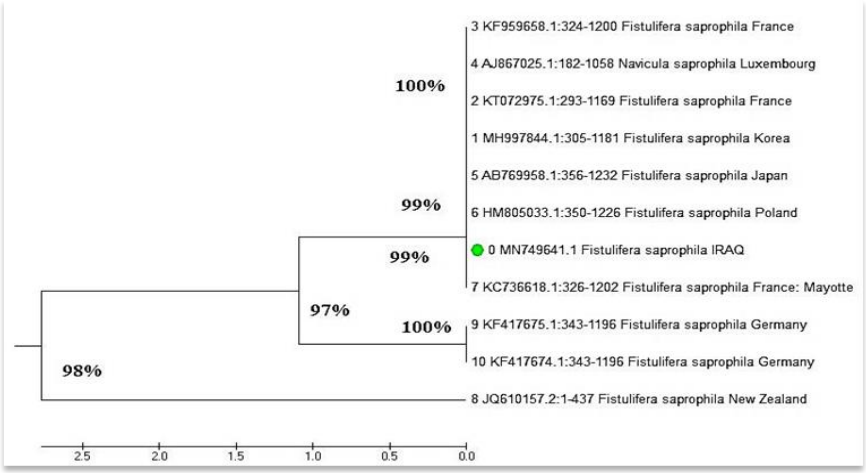


Diagram (8): Phylogenetic tree of *Fistulifera saprophila* based on *18S rRNA* gene sequences conferred by GeneBank data base, were analyzed and aligned through BLAST from NCBI using the Neighbor-Joining Analyses of 877 bp of corresponding position of *18S rRNA* gene sequence. MEGA 6 program was used for phylogenetic tree.

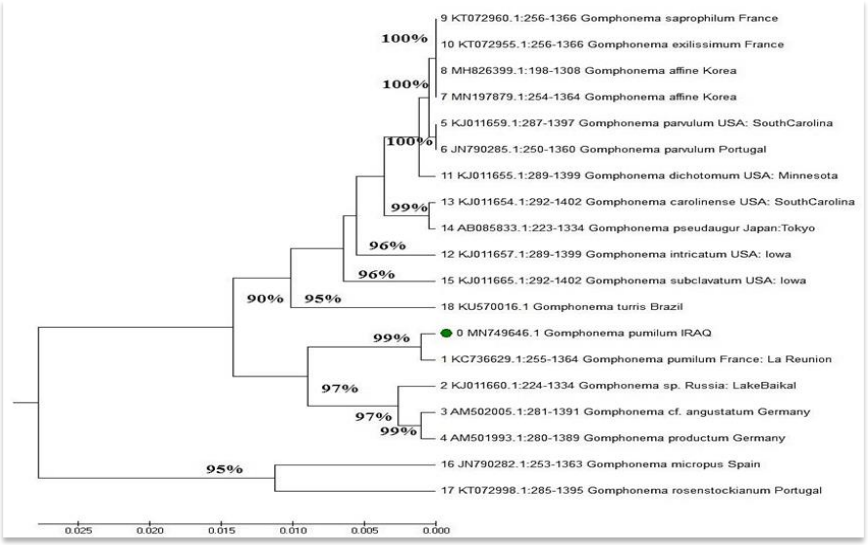


Diagram (9): Phylogenetic tree of *Gomphonema pumilum* based on *18S rRNA* gene sequences conferred by GeneBank data base, were analyzed and aligned through BLAST from NCBI using the Neighbor-Joining Analyses of 1110 bp of corresponding position of *18S rRNA* gene sequence. MEGA 6 program was used for phylogenetic tree.

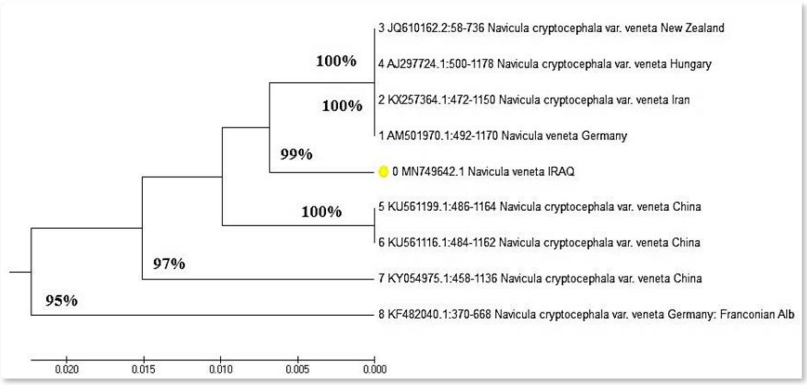


Diagram (10): Phylogenetic tree of *Naviculaveneta* based on *18S rRNA* gene sequences conferred by GeneBank data base, were analyzed and aligned through BLAST from NCBI using the Neighbor-Joining Analyses of 679 bp of corresponding position of *18S rRNA* gene sequence. MEGA 6 program was used for phylogenetic tree.

Five diatom species identified

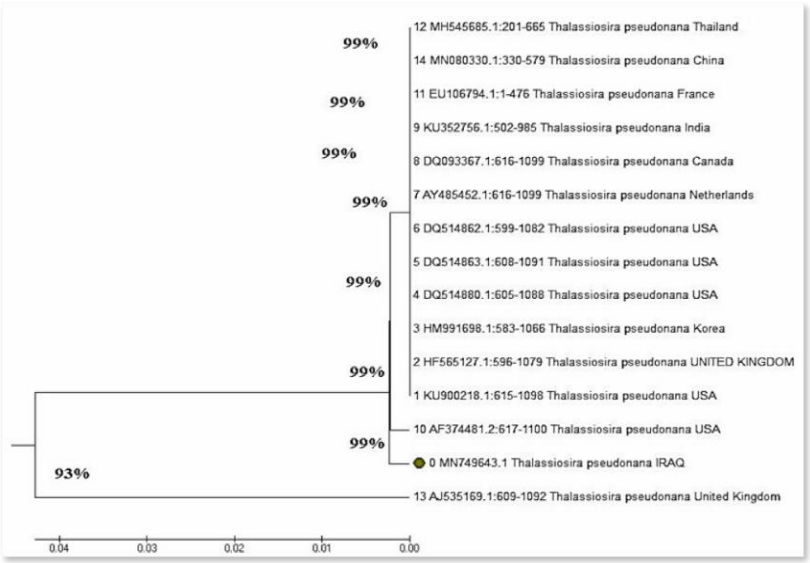


Diagram (11): Phylogenetic tree of *Thalassiosira pseudonana* based on *18S rRNA* gene sequences conferred by GeneBank data base, aligned together with yeast, were analyzed and aligned through BLAST from NCBI using the Neighbor-Joining Analyses of 484 bp of corresponding position of *18S rRNA* gene sequence. MEGA 6 program was used for phylogenetic tree.

The molecular analysis revealed five diatom species which were identified for the first time by molecular analysis, while two species were recorded as new species of Iraqi algal flora and were registered in NCBI under the accession number as follows:

(1) *Achnanthidium minutissimum* (accession numberMN749640.1).
(2) *Fistulifera saprophila* (accession number MN749641.1) new record.
(3) *Gomphonema pumilum* (accession number MN749642.1).
(4) *Navicula veneta* (accession number MN749643.1).
(5) *Thalassiosira pseudonana* (accession number MN749646.1) new record.

By comparing Phylogenetic tree of *A. minutissimum* with neighboring countries, it was 99% closer to China. When compared, Phylogenetic tree of *F. saprophila* turned out to be more similar 99% to the ID number diagnosed in Korea. The Phylogenetic tree for the species *G. pumilum* was more closely related to the registration number that was diagnosed in France as 99%.The affinity ratio was 99% phylogenetic tree of *N. veneta* with registration number ID: AM501970.1 which registered in Germany. Phylogenetic tree of *T. pseudonana* based on *18S rRNA* gene sequences conferred by GeneBank data base, were analyzed and aligned through

BLAST from NCBI using the Neighbor-Joining Analyses and was also 99% recorded in the USA.

The morphological and molecular (phylogenetic) determination of diatomic organisms is another potential conflict source. Firstly, there is a range of genetically distinctive forms that reflect almost all morphospecies. Secondly, some species have their own auto-ecological values subdivided into subspecies or morphological varieties. In the first case, a significant benefit for biomonitoring may be the cryptic diversity, especially when cryptic species relate to certain specific ecological conditions. The second case is more troubling because the sub-specific taxa are generally not genetically characterized (Visco *et al*., 2015).

The quantitative analysis of NGS data gives the greatest challenge in efforts to alleviate biases in the calculation of diatom indices. In fact, numerous NGS environmental studies display contradiction between the number of sequences assigned to a given species and the number of specimens of the same species in microscopic preparations (Gibson *et al*., 2014) or even microbially diverse communities (Amend *et al*., 2010).This unbalance correlation between the multiple reading and individuals could be interpreted either by technical biases introduced during DNA extraction, PCR amplification or sequencing or by biological factors such as the variations of rRNA gene copies (Weber and Pawlowski, 2013; Pawlowski *et al*., 2014), which may depend on number of nuclei in genome size, or variety in size of cell (Prokopowich *et al*., 2003; Heyse *et al*., 2010).

The results given in this experience study will need validation by more NGS-based surveys of diatom diversity. Indeed, substantial efforts must be done by diatom taxonomists and biologists to complete the DNA barcoding reference database and to determine the rate of genetic and morphological differences in diatom species.

A total of 186 taxa were identified of epipelic algae by microscopy (Tab. 3), whereas only a few of identified epipelic (5.4%) were observed by using molecular analysis in this study (Tab. 5). While *Amphora montana*, *Fistulifera saprophila*, *Nitzschia cf. frustulum* and *Thalassiosira pseudonana* were detected by molecular analysis and not identified by microscopy. Another study also observed only 19% of identified diatoms by using molecular analysis while they identified 63 taxa by microscopy (Vasselon *et al*., 2017). Vasselon *et al*. (2017) mentioned that about 68% of diatom species identified by microscopy were incomplete in the reference database; moreover, it is important to use suitable DNA extraction methods. This finding will encourage the researcher to use the molecular analysis for identifying algae in the environment. These diatoms were found in freshwater habitats and reorganized in different regions worldwide (Reichardit, 1997; Wojtal, 2003, Novais *et al*., 2015).

CONCLUSION

The use of molecular concept of classification is important to re-check the list of the algal flora in Iraq to confirm or to amend them. The application of eDNA revealed five diatom species were a new record species of Iraqi algal flora and it will be a catalyst for new studies

of biodiversity and environmental studies in Iraq and the region. The molecular application will resolve the misclassification and the persistent problems of misidentification of algae. Moreover, the NGS will decrease the period of the specimen process with using automation of the protocols of molecular works and led to the increase in the number of sampling, in addition to reduce the cost of this tech.

ACKNOWLEDGMENTS

Authors thank the department of Biology, College of Science and Iraqi Hereditary company for the facilities support and they extend their thanks to Asst. Prof. Dr. Halah Al- Haideri and Asst. Prof. Dr. Safa Al-Deen A. Al-Qaysi for their comments on the manuscript.

LITERATURE CITED

Abed, I. J., Abdulhasan, G. A. and Najem, A. M. 2018. Genotype versus phenotype to determine the definitive identification of the genera *Chlorella* Berjrinck, 1890 (Chlorophyceae) and Coelastrella Chodat, 1922 (Scenedesmaceae). *Bulletin of Iraq Natural History Museum*, 15 (1):101-111.

Adebayo, O. O., Sangodoyin, A.Y., Ogedengbe, K., and Taiwo, O. 2013. Mapping of river water quality using inverse distance weighted interpolation in Ogun-Osun river basin, Nigeria. *Acta Geographica Debrecina Landscape and Environment*, 7 (2): 48-62.

Al-Hussieny, A. A., Hussein, H. T. and Hmood, A. H. 2014. Propagation of algae farms using several methods by different farming media. *Journal of the College of Basic Education*, 20 (84/Scientific): 121-142.

Al-Rawi, A., Alwash, B. M. J., Al-Essa, N. E. and Hassan, F. M. 2018. A new record of *Coelastrella terrestris* (Reisigle) Hegewald & N. Hangata, 2002 (Sphaeropleales, Scenedesmaceae) in Iraq. *Bulletin of Iraq Natural History Museum*, 15 (2): 153-161.

Amend, A. S., Seifert, K. A., Bruns, T. D. 2010. Quantifying microbial communities with 454 pyrosequencing: does read abundance count? *Molecular ecology*, 19(24): 5555-6565.

Aydın, G. Ş. and Büyükışık, B. 2014. Effects on the species-specific variables nutrient pulses: *Thalassiosira allenii* (Takano). *Journal of Tekirdag Agricultural Faculty*, 11(3): 82-90.

Baird, D. J. and Hajibabaei, M. 2012. Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. *Molecular Ecology*, 21(8): 2039-2044.

Berthold, M., Karsten, U., von Weber, M., Bachor, A. and Schumann, R. 2018. Phytoplankton can bypass nutrient reductions in eutrophic coastal water bodies. *Ambio*, 47(1): 146-158.

Campbell, B. M., Beare, D.J., Bennett, E. M., Hall-Spencer, J. M., Ingram, J. S., Jaramillo, F., Ortiz, R., Ramankutty, N., Sayer, J. A. and Shindell, D. 2017. Agriculture production as a major driver of the Earth system exceeding planetary boundaries. *Ecology and Society*, 22(4): 1-8.

Chang, H. 2008. Spatial analysis of water quality trends in the Han River basin, South Korea. *Water Research*, 42 (13): 3285-3304.

Çiçek, A., Bakiş, R., Uğurluoğlu, A., Köse, E. and Tokatli, C. 2013. The effects of large borate deposits on ground water quality. *Polish Journal of Environmental Studies*, 22(4): 1031- 1037.

Ewing, B. and Green, P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research*, 8(3):186-194.

Gao, Z., Wang, X., Wei, X., Liu, Y. and Han, J. 2019. DNA Mini-Barcoding: A Derived Barcoding Method for Herbal Molecular Identification. Frontiers in plant science, 10: 987.

Gibson, J., Shokralla, S., Porter, T. M., King, I., van Konynenburg, S., Janzen, D. H., Hallwachs, W. and Hajibabaei, M.2014. Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasystematics. *Proceedings of the National Academy of Sciences*, 111(22): 8007-8012.

Godhe, A. and Härnström, K. 2010. Linking the planktonic and benthic habitat: genetic structure of the marine diatom Skeletonema marinoi. *Molecular Ecology*, 19(20): 4478-4490.

Guillard, R. R. 1975. Culture of phytoplankton for feeding marine invertebrates. *In*: Smith, W. L., Chanley, M. H. (eds), Culture of marine invertebrate animals (pp. 29-60). Springer, Boston, MA.

Gurbuz, H. and Kivrak, E. 2002. Use of epilithic diatoms to evaluate water quality in the Karasu River of Turkey. *Journal of Environmental Biology*, 23(3): 239-246.

Hamsher, S. E., Evans, K .M., Mann, D. G., Poulíčková, A. and Saunders, G. W. 2011. Barcoding diatoms: exploring alternatives to COI-5P. *Protist*, 162 (3): 405-422.

Heyse, G., Jönsson, F., Chang, W.J. and Lipps, H.J. 2010. RNA-dependent control of gene amplification. *Proceedings of the National Academy of Sciences*, 107(51): 22134-22139.

Kiss, K. T. 1984. Occurrence of *Thalassiosira pseudonana* Hasle et Heimdal (Bacillariophyceae) in some rivers of Hungary. *Act Botiunica Hungarica*, 30 (3-4): 277-287.

Luddington, I.A., Kaczmarska, I. and Lovejoy, C. 2012. Distance and character-based evaluation of the V4 region of the 18S rRNA gene for the identification of diatoms (Bacillariophyceae). *PLoS ONE*, 7 (9): e45664.

Moniz, M. B. and Kaczmarska, I. 2010. Barcoding of diatoms: nuclear encoded ITS revisited. *Protist*, 161(1): 7-34.

Novais, M. H., Juettner, I., Van de Vijver, B., Morais, M. M., Hoffmann, L. and Ector, L. 2015. Morphological variability within the *Achnanthidium minutissimum* species complex (Bacillariophyta): comparison between the type material of *Achnanthes minutissima* and related taxa, and new freshwater Achnanthidium species from Portugal. *Phytotaxa*, 224 (2): 101-139.

Passy, S. I., Bode, R. W., Carlson, D. M. and Novak, M. A. 2004. Comparative environmental assessment in the studies of benthic diatom, macroinvertebrate, and fish communities. International Review of Hydrobiology: *A Journal Covering all Aspects of Limnology and Marine Biology*, 89 (2): 121-138.

Pawlowski, J., Lejzerowicz, F. and Esling, P. 2014. Next-generation environmental diversity surveys of foraminifera: preparing the future. *The Biological Bulletin*, 227(2): 93-106.

Prokopowich, C. D., Gregory, T. R. and Crease, T. J. 2003. The correlation between rDNA copy number and genome size in eukaryotes. *Genome*, 46(1): 48-50.

Reichardit, E. 1997. Taxonomic revision of species complex of *Gomophonema pumilum* (Bacillariophyceae). *Nova Hedwigia*, 65:99-130.

Round, F. E., Crawford, R. M. and Mann, D. G. 1990. The Diatoms. Morphology and biology of the genera. Cambridge University Press, Cambridge, UK, ix+747 pp.

Salman, J. M., Hassan, F. M. and Baiee, M. A. 2017. Practical methods in environmental and pollution laboratory. Issued by Environmental Research Center, University of Babylon. ISBN2463.

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. and Willerslev, E. 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular ecology*, 21(8): 2045-2050.

Tan, X., Zhang, Q., Burford, M. A., Sheldon, F. and Bunn, S. E. 2017. Benthic diatom based indices for water quality assessment in two subtropical streams. *Frontiers in microbiology*, 8: 601.

Thomsen, P.F. and Willerslev, E. 2015. Environmental DNA–An emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation*, 183: 4-18.

Tokatlı, C. and Dayıoğlu, H. 2011. Use of epilithic diatoms to evaluate water quality of Murat Stream (Sakarya River Basin, Kütahya): different saprobity levels and pH status. *Journal of Applied Biological Sciences*, 5(2): 55-60.

Vasselon, V., Domaizon, I., Rimet, F., Kahlert, M. and Bouchez, A. 2017. Application of high-throughput sequencing (HTS) metabarcoding to diatom biomonitoring: Do DNA extraction methods matter?. *Freshwater Science*, 36(1): 162-177.

Visco, J. A., Apothéloz-Perret-Gentil, L., Cordonier, A., Esling, P., Pillet, L. and Pawlowski, J. 2015. Environmental monitoring: inferring the diatom index from next-generation sequencing data. *Environmental Science and Technology*, 49 (13): 7597-7605.

Weber, A. A. and Pawlowski, J. 2013. Can abundance of protists be inferred from sequence data: a case study of Foraminifera. *PloS ONE*, 8 (2): e56739.

Wojtal, A. 2003. Diatoms of the genus *Gomphonema* Ehr. [Bacillariophyceae] from a karstic stream in the Krakowsko-Czestochowska Upland. *Acta Societatis Bot Anicorum Poloniae*, 72 (3): 213- 220.

Yu, L., Zhang, W., Liu, L. and Yang, J. 2015. Determining microeukaryotic plankton community around Xiamen Island, southeast China, using Illumina MiSeq and PCR-DGGE techniques. *PLoS ONE*, 10(5): e0127721

Zimmerman, J., Glöckner, G., Jahn, R., Enke, N. and Gemeinholzer, B. 2015.. Metabarcoding vs. morphological identification to assess diatom diversity in environmental studies. *Molecular Ecology Resources*, 15(3):526-42.

Bull. Iraq nat. Hist. Mus.
(2020) 16 (1): 39-61.

تشخيص خمسة انواع دايتومية باستعمال التطبيقات الممكنة لتسلسل الحمض النووي البيئي من الجيل التالي

ورقاء يحيى المشهداني* و فكرت مجيد حسن**
*معهدالهندسية الوراثية و التقنيات الاحيائية للدراسات العليا، جامعةبغداد، بغداد، العراق
**قسم علوم الحياة، كليةالعلوم للبنات، جامعةبغداد، بغداد، العراق

تأريخ الاستلام: 2020/01/16، تأريخ القبول: 2020/04/27، تأريخ النشر: 2020/06/24

الخلاصة

عرفت تشفير الباركود بشكل واسع كأداة قوية لتحديد الكائنات الحية خلال العقد الماضي. لذلك هدفت الدراسة الحالية لاستخدام المفهوم الجزيئي لتحديد الدايتومات باستخدام الحمض النووي البيئي.

اخذت العينات الدايتومات من نهر دجلة، اذ بينت نتائج استخلاص وتحليل تسلسل الحمض النووي البيئي من خلال استخدام تسلسل الجيل التالي (NG) بان اعلى نسبة سجلت لكل من الدايتومات التالية :
(21.1%) *Achnanthidium minutissimum* (Kützing) Czarnecki, 1994 و *Cocconeis placentula* Ehrenberg, 1838(21.3%) و *Nitzschia palea* (Kützing) W. Smith, 1856 بنسبة 16.3% .

سُجلتْ خمسة اجناس للدايتومات لاول مرةفي المركز الوطني لمعلومات لتكنولوجيا الحيوية (NCBI) تحت ارقام الانضمام MN749640.1 و MN749641.1 و MN749642.1 و MN749643.1 و MN749646.1على التوالي و هي: *Achnanthidium minutissim* و *Fistulifera saprophila* (Lange-Bertalot & Bonik) Lange-Bertalot, 1997 و *Gomphonema pumilum* (Grunow) E. Reichardt & Lange-Bertalot, 1991 و *Navicula*

Thalassiosira pseudonana Hasle و *veneta* Kütz. 1844 Heimdal, 1970؛ كما يعد النوعان *Fistulifera saprophila* و *Thalassiosira pseudonana* *تسجيلا* جديدا للفلورا الطحلبية في العراق.

تعتبر دراسة الحمض النووي البيئي عاملا مساعدا في اجراء دراسات جديدة حول التنوع البيولوجي والدراسات البيئة في العراق والمنطقة.