Comptes Rendus

Biologies

Isabel Ballesteros, Pablo Castillejo, Adriana Paulina Haro, Cuthy Cristina Montes, Carla Heinrich and Eduardo Alexis Lobo

Genetic barcoding of Ecuadorian epilithic diatom species suitable as water quality bioindicators

News and events / *Actualités*

Genetic barcoding of Ecuadorian epilithic diatom species suitable as water quality bioindicators

Code-barres génétique des espèces de diatomées épilithiques équatoriennes susceptibles de servir de bioindicateurs de la qualité de l'eau

Isabel Ballesteros[⊕] *ᵃ*, **Pablo Castillejo**[⊕] *, ᵃ*, **Adriana Paulina Haro** *ᵃ*,
Cuthy Cristina Montes *ᵃ*, **Carla Heinrich** *ᵇ* and **Eduardo Alexis Lobo** *ᵇ*

ᵃ Facultad de Ingeniería y Ciencias Aplicadas, Grupo de Investigación en
Biodiversidad Medioambiente y Salud (BIOMAS), Universidad de Las Américas,
Quito, Ecuador

ᵇ Universidade de Santa Cruz do Sul (UNISC), Rio Grande do Sul, Brasil.

E-mails: maria.ballesteros@udla.edu.ec (I. Ballesteros), pablo.castillejo@udla.edu.ec
(P. Castillejo), adriana.haro@udla.edu.ec (A. P. Haro), cuthy.montes@udla.edu.ec
(C. C. Montes), carla.heinrich@hotmail.com (C. Heinrich), lobo@unisc.br (E. A. Lobo).

Abstract. Diatom identification is a key step in using these microorganisms as water quality bioindicators. Morphological diagnosis is a difficult task due to the enormous number of species and their microscopic size. This can be overcome using molecular tools to complement the diagnosis. The main goal of this work was to obtain the DNA barcode of Ecuadorian epilithic diatoms with a wide geographical distribution, a well-defined ecological range and characteristics that allow them to be reliable indicator species. Unialgal diatom cultures were obtained from environmental samples of Ecuadorian Andean streams. Morphological characterization of cultures was carried out under SEM microscopy. For molecular characterization, 18SV4 and rbcL barcodes were sequenced from each strain and blasted against a GenBank database. A phylogenetic tree for each barcode was constructed using the ML method including sequences of strains of the studied species from different geographical locations. The results showed the following five species to be suitable as bioindicators and these were isolated. *Sellaphora seminulum* (strain JA01b, c), *Nitzschia fonticola* (strain SP02a) and *N. palea* (strain CA01a) are tolerant to eutrophication; *Eolimna minima* (strain CH02a) is a mesotrophic water bioindicator, and *Achnanthidium minutissimum* (strain JA01a) is an oligotrophic water bioindicator. The comparison with the GenBank database of the barcoding regions supported the morphological identification. The barcoding sequences of the strains showed a high percentage of identity with the sequences reported in INSDC databases for the same species. The topology of the phylogenetic trees demonstrates that epilithic diatoms from Ecuador are closely related to those of same species isolated from other geographical regions. This study is a first attempt to establish a morphological and molecular taxonomic

* Corresponding author.

reference library for neotropical diatoms. This study demonstrates that it would be feasible to use the existing barcoding data for diatoms to develop molecular tools for the bioassessment of aquatic ecosystems in the Ecuadorian Andean region.

Résumé. L'identification des diatomées est une étape clé dans l'utilisation de ces microorganismes comme bioindicateurs de la qualité de l'eau. Le diagnostic morphologique est une tâche difficile en raison du nombre considérable d'espèces et de leur dimension microscopique. Il est possible de surmonter cette difficulté en utilisant des techniques moléculaires pour compléter le diagnostic. L'objectif principal de ce travail était d'obtenir le code-barre de l'ADN des diatomées épilithiques équatoriennes ayant une large distribution géographique, une niche écologique bien définie et des caractéristiques leur permettant d'être des espèces indicatrices fiables. Des cultures de diatomées unialgales ont été obtenues à partir d'échantillons environnementaux de cours d'eau des Andes équatoriennes. La caractérisation morphologique des cultures a été réalisée sous microscopie MEB. Pour la caractérisation moléculaire, les codes-barres 18SV4 et rbcL ont été séquencés à partir de chaque souche et comparés à la base de données GenBank. Pour chaque code-barres, un arbre phylogénétique a été construit à partir de la méthode ML comprenant des séquences de souches des espèces étudiées, provenant de différents lieux géographiques. Les résultats ayant montré que les cinq espèces suivantes étaient appropriées comme bioindicateurs, elles ont été isolées. *Sellaphora seminulum* (souche JA01b, c), *Nitzschia fonticola* (souche SP02a) et *N. palea* (souche CA01a) sont tolérantes à l'eutrophisation ; *Eolimna minima* (souche CH02a) est un bioindicateur d'eau mésotrophe, et *Achnanthidium minutissimum* (souche JA01a) est un bioindicateur d'eau oligotrophe. La comparaison avec la base de données GenBank des régions de code-barres a supporté leurs identifications morphologiques. Les séquences de code-barres des souches ont montré un pourcentage élevé d'identité génétique avec les séquences signalées dans les bases de données de l'INSDC pour la même espèce. La topologie des arbres phylogénétiques démontre que les diatomées épilithiques de l'Équateur sont étroitement liées à celles des mêmes espèces isolées d'autres régions géographiques. Cette étude est une première tentative d'établir une bibliothèque de référence morphologique et taxonomique moléculaire pour les diatomées néotropicales. Cette étude démontre qu'il serait possible d'utiliser les données de code-barres existantes pour les diatomées afin de développer des instruments moléculaires pour la bioévaluation des écosystèmes aquatiques dans la région andine équatorienne.

1. Introduction

Diatoms are an enormous group of microalgae containing more than 100,000 living species [1, 2]. A lot of work has been done to document diversity and relationships within the group. As a result, species-level taxonomy has become adequate to support the extensive and successful use of diatoms in biomonitoring based on their species-specific response to their environmental changes, especially organic pollution and eutrophication, with a broad spectrum of tolerance, from oligotrophic to eutrophic conditions [3–5].

Diatom taxonomy is based almost exclusively on the morphological features of the silica frustules [6] and, even after training and long experience, diatomists often find it difficult to agree about iden-

tifications. Moreover, morphological treatment is a time-consuming process and costly. In contrast, DNA-based identification has been proposed as an alternative method to complement or even replace traditional identification methods of species [7]. DNA barcoding is based on DNA sequences that are linked to morphologically identified specimens. It is considered a faster and universally applicable approach and also has the potential for refined analyses [8, 9]. Consequently, DNA barcoding deserves careful consideration as a means of improving the reliability of identifications and discovering species and also of increasing the quality and quantity of other taxonomic information. DNA-based methods are being used for assessing biodiversity in environmental samples, the so-called DNA metabarcoding [10],

which allows one to determine a species' presence and abundance from bulk samples of soil, water, or air [11]. Several studies have demonstrated that metabarcoding may be an applicable tool for ecological monitoring based on epilithic diatoms [12–16]. The preferred barcoding markers in these studies were 18S rRNA and rbcL (Ribulose-1,5-bisphosphate carboxylase/oxygenase).

The basis for a reliable identification with molecular methods is a comprehensive reference library where molecular and morphological data are tied together with a taxonomic name. However, the taxonomic consistency of the sequence information in these databases is very low for diatoms [17]. Rimet *et al.* [18] developed the R-syst::diatom barcoding library in open-access (http://www.rsyst.inra.fr/) that includes barcodes for microalgae strains of the culture collection maintained in INRA (Institut National de la Recherche Agronomique, France). Another limitation to DNA-based identification approaches is the natural intraspecific divergence of the barcoding marker in specimens from different geographic origins. To date, no information on barcoding sequences has been reported for diatom species from Ecuador. This is a constraint on the application of these methods in freshwater diversity studies and in biomonitoring.

In this context, this research aimed at obtaining the barcodes of Ecuadorian epilithic diatoms with a wide geographical distribution, a well-defined ecological range and characteristics that allow them to be reliable indicator species [19–21]. This study is a humble first step towards coupling the morphological and molecular data of benthic diatoms in Ecuador. We have therefore studied morphology as a diagnostic feature of each unialgal diatom strain via light and scanning electron microscopies and also determined the DNA barcode 18SV4 rRNA [9, 22] as well as rbcL [23, 24] in order to establish the foundation of a reference library. The original strains are kept alive in the Universidad de las Américas' epilithic diatoms collection. By publishing these data online, we are contributing to a morphological and molecular taxonomic reference library for neotropical diatoms, making them available for comparative studies. Thus, we expect this to be a key step in developing a trophic water quality index based on a molecular biology analysis of epilithic diatoms in Ecuador.
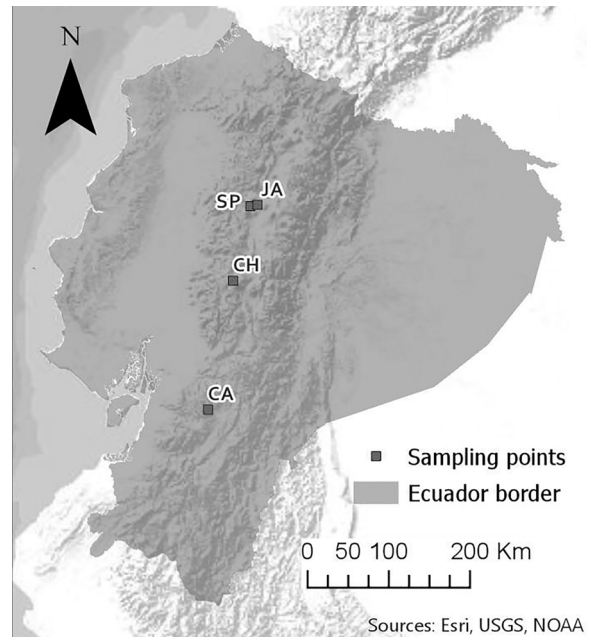

Figure 1. Location of sampling sites along Ecuadorian Andes. JA: Jambeli River at Pichincha province. SP: San Pedro River at Illiniza Ecological Reserve. CH: stream at Chimborazo Ecological Reserve. CA: stream at Cajas National Park.

2. Materials and methods

2.1. *Sample collection*

Samples were collected under the MAE-DNB-CM-2018-093 authorization conceded by the Environment Ministry of Ecuador. Epilithic diatom samples were collected from four freshwater streams within the Andean region of Ecuador (Figure 1; Table 1). Three of them characterized as páramo (highland wilderness) landscapes at the Iliniza Volcano (SP), Chimborazo volcano (CH) and Cajas national park (CA). The fourth one is located at the Jambelí stream, in the parish of Machachi, surrounded by agricultural livestock mosaics (Table 1). Location data were taken with a GPS navigator (WGS 84 UTM Zone 17 with 5 meters of precision). The sampling design aimed to capture indicator species from pristine streams (SP, CH and CA) and a eutrophic stream (JA) from the Ecuadorian Andean region (Figure 1).

The physical and chemical variables analyzed were water temperature, pH, turbidity, dissolved oxy-

Table 1. Location of the sampled streams in the central Ecuadorian Andean Region with geographical coordinates, altitude (masl) and the code assigned to samples collected in each stream

Code	Province	Description	Latitude	Longitude	Altitude
SP	Pichincha	Illiniza Ecological Reserve	−0.58661	−78.67343	3601 masl
JA	Pichincha	Quito	−0.57296	−78.59418	3162 masl
CH	Chimborazo	Chimborazo Ecological Reserve	−1.41337	−78.86458	3978 masl
CA	Azuay	Cajas National Park	−2.84271	−79.14182	3165 masl

gen (DO) and conductivity, measured in the field with a multiparameter analyzer, and nitrate and phosphate, measured in the laboratory. Diatom samples were scrubbed off the upper surface of three to five submerged stones, 10 to 20 cm in diameter, using a toothbrush [25]. Diatom and water samples were kept cold until they were transferred to the laboratory.

2.2. *Isolation and culturing*

The environmental samples were cultured on BBM agar plates [26] supplemented with sodium silicate (10 mM) at 15–18 °C, light intensity of 50 $\mu Em^{-2}\ s^{-1}$ and 12/12 hr light/dark regime. Individualized macroscopic colonies on agar plates were inoculated into liquid media and grown until they turned brown. This process was repeated until unialgal cultures were obtained. The achievement of unialgal cultures was checked by light microscopy using an Olympus BX51 microscope. Diatom strains were maintained on agar plates and liquid medium.

2.3. *Morphology-based analysis*

One fraction of each unialgal culture was used for morphological analysis. Living cells as well as cleaned frustules were examined and photographed by light microscopy. To remove organic material, the cells were oxidized with 50% nitric acid at 80 °C for an hour and rinsed several times with H_2O [27]. Permanent slides were mounted with the high refraction index mounting medium Naphrax®. For scanning with the electron microscope, a few drops were dried on stubs and sputter-coated with gold using a Quorum Tech Q150RES. SEM images were taken using an ultra-high-resolution analytical field emission (FE) scanning electron microscope Tescan MIRA3, operated at 10 kV.

Species identification was performed using a microscope Olympus BX-40 according to the following taxonomic references [4, 28–30].

2.4. *Molecular-based analysis*

A second fraction of each unialgal culture was used to extract total genomic DNA. Cells were pelleted by centrifugation of 1 mL culture medium at 8000 g for 2 min and disrupted using ceramic beads by vortexing. The extraction method was then performed according to Edwards *et al.* [31]. The rbcL and 18SV4 rRNA regions were submitted to PCR amplification using GoTaq® Green Master Mix by Promega with 0.2 μM of each primer. The V4 region of the 18S rRNA was amplified with the primer pair DIV4for/DIV4rev3 [13, 32]. Primer used for the rbcL fragment were Diat_rbcL_708F [33] and reverse primer R3 [34]. The selected primer pairs amplify a fragment around 300 pb and have been used previously for DNA metabarcoding analysis [12, 14, 24]. PCR regime included an initial denaturation at 95 °C for 2 min, then 35 cycles of denaturation at 95 °C for 45 s, annealing at 50 °C for 45 s, elongation at 72 °C for 30 s, and a final elongation at 72 °C for 2 min. Sequence reactions were performed with an Applied Biosystems® 3130 Genetic Analyzer. Sequences were edited with MEGA 7 software [35] and blasted against the GenBank database at NCBI (National Center for Biotechnology Information). Sequences of strains of the studied species from different geographical locations were downloaded from GenBank (Supplementary Tables S1 and S2) and aligned with barcoding sequences from Ecuadorian strains using the Muscle Algorithm [36]. *Bolidomonas pacifica* was used as an outgroup taxon, due to its genetic proximity to diatoms [37]. The ends of the alignments were trimmed to minimize missing characters. Tree topologies and branch lengths were computed separately for the

Table 2. Sequence identity of the closest relative of each isolated strain found in GenBank using BLASTn

Strain	18SV4	% identity	N hits	Closest match	rbcL	% identity	N hits	Closest match
CA01a	*Nitzschia palea*	99–100%	37	KU341755.1	*Nitzschia palea*	98–100%	104	HF675122
CH02a	*Eolimna minima*	95–100%	7	KM084877.1	*Eolimna minima*	95–97%	3	KM084939.1
JA01a	*Achnanthidium minutissimum*	93–100%	24	KY863464	*Achnanthidium minutissimum*	97–100%	15	KY863482.1
JA01b	*Sellaphora seminulum*	93–100%	6	KR150677.1	*Sellaphora seminulum*	95–100%	4	KM084937.1
JA01c	*Sellaphora seminulum*	93–100%	5	KR150677.1	*Sellaphora seminulum*	95–100%	4	KM084937.1
SP02a	*Nitzschia fonticola*	90–98%	3	AJ867022.1	*Nitzschia fonticola*	99–100%	4	HF675068.1

two markers with the maximum-likelihood method (ML) using Tamura-Nei distance [38] with gamma distributed rates among sites followed by a statistical test of the tree topologies with 500 bootstrap replications.

3. Results

Five suitable bioindicator species were sampled from the Ecuadorian Andean Region (Table 1): *Sellaphora seminulum* (Grunow) D. G. Mann (strain JA01b, c), *Nitzschia fonticola* (Grunow) Grunow (strain SP02a) and *N. palea* (Kützing) W. Smith (strain CA01a) are tolerant to eutrophication; *Eolimna minima* (Grunow) Lange-Bertalot (strain CH02a) is a mesotrophic water bioindicator and *Achnanthidium minutissimum* (Kützing) Czarnecki (strain JA01a) is an oligotrophic water bioindicator. The comparison with the GenBank database of the barcoding regions obtained supported the morphological identification. Blasting results against NCBI are shown in Table 2. *N. palea* had the most hits (37 sequences for 18SV4; 104 for rbcL) followed by *A. minutissimum* (24 sequences for 18SV4; 15 for rbcL). GenBank accession numbers for 18SV4 rRNA and rbcL sequences from this study are MN589666-MN589670 and MN603956-MN603960 respectively.

JA01b strain presents linear valves with rounded apices. The axial area is narrow, linear. The central area is wide, rectangular. The raphe is filiform with the terminal fissures curved to the same side of the valve with distinct central pores. Striae are radiate, multiseriate, and more distant and curved near the central area, with 2 or 3 shorter striae. The areolas are rounded. Valve dimensions: Length: 9–12 μm; Width: 3–3.5 m, 20–22 striae in 10 μm (Figure 2i–l). This result is consistent with *S. seminulum* description [39]. Morphological description of JA01c strain also matches *S. seminulum* although its length ranges from 6 to 10 μm and width reaches up to 4.5 μm (Figure 2m–p). Thus, JA01c strain presents elliptical valve shape while JA01b has linear valves. Despite the morphological differences, both strains have 100% identity for 18SV4 and rbcL sequences. The number of 18S and rbcL sequences for *Sellaphora seminulum* in the GenBank database is scarce. There are strains from the Geneva basin [32] Seoul and Reunion Island [40] with a homology of 98% with 18SV4 and rbcL. Evans *et al.* [41] and Kermarrec *et al.* [24] described two strains of *S. seminulum* from United Kingdom and Mayotte Island at the Komora Archipelago, with a homology below 95%. There is a third case in the GenBank for a strain from Berlin that presents 100% of identity with rbcl, but less than 95% with 18SV4 [17].
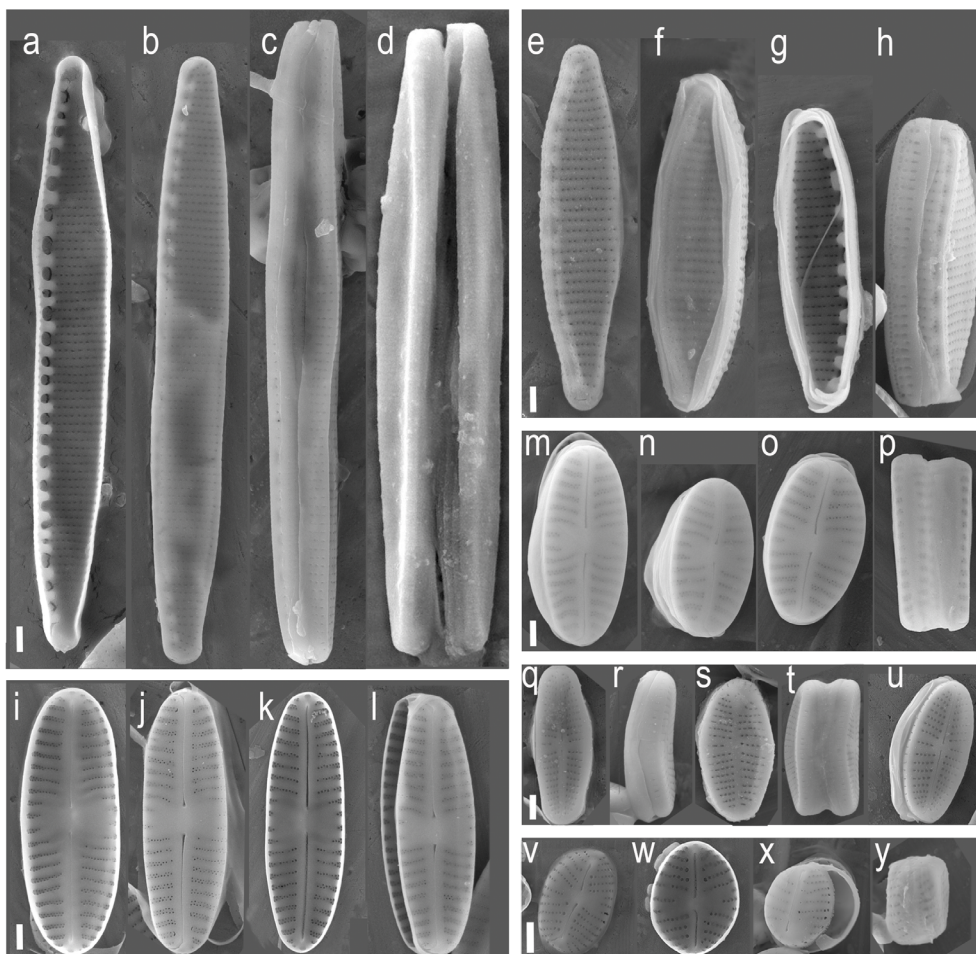
Figure 2. Isolated diatoms under SEM: a–d: strain CA01a (*N. palea* (Kützing) W. Smith); e–h: strain SP02 (*Nitzschia fonticola* (Grunow) Grunow); a, i–l: strain JA01b (*Sellaphora seminulum* (Grunow) D. G. Mann); m–p: JA01c (*Sellaphora seminulum* (Grunow) D. G. Mann); q–t, u: JA01a (*Sellaphora seminulum* (Grunow) D. G. Mann); v–y: CH02a (*Eolimna minima* (Grunow) Lange-Bertalot). Scale bar = 1 µm.

The morphological traits of strain SP02 match the species *Nitzschia fonticola* [42]: linear-lanceolate valves, acutely rounded apices, eccentric, marginal raphe, uniseriate striae on the valvar face and biseriate in the canal of the raphe, practically equidistant fibulae, except in the central area that presents two fibulae more distant from the others. Valve dimensions: Lenght: 15–18 µm; width: 2.5–3 µm; 10–14 fibulae in 10 µm (Figure 2e–h). The sequences obtained from this strain have higher identity with rbcL sequences (99–100%) than with 18SV4 sequences (90–98%) uploaded to the GenBank. There is no strain described as *N. fonticola* with both rbcL an

18SV4 barcodes reported. The sequences matching rbcL belong to strains collected from the United Kingdom, Spain, and Reunion Island [17, 40] while 18SV4 matches only one sequence (from France, data not published) with over 98% similarity. There are two other sequences for 18SV4 with less than 90% similarity from microorganisms isolated from karstic-stream biofilms in Germany (data unpublished).

CA01a strain, described as *Nitzschia palea* [43], presents linear-lanceolate valves, rostral apices, eccentric raphe, marginal, very thin uniseriate striae, equidistant fibulae including the central pair. Valve

dimensions: Lenght: 25–27 μm; width: 2.8–3 μm; 16 fibulae in 10 μm (Figure 2a–d). There are plenty of sequences for *N. palea* [32, 37, 44] in the GenBank from all around the world, most of them with over 99% identity with CA01a strain for both rbcL and 18SV4 barcodes.

CH02a strain shows elliptical valves with rounded apices. The axial area is narrow, linear. The central area is little expanded, slightly rectangular, limited by the irregular size of the median striae. The raphe is filiform, straight, with proximal ends slightly flexed. The central pores are small. Striae are radiate, uniseriate, with irregular sizes near the central area, composed of rounded areolas (Figure 2v–y). Valve dimensions: Length: 6–10 μm; Width: 3–4 μm; 25–26 stretch marks in 10 μm. This is consistent with the morphological description of *E. minima* [45]. The closest match with both barcodings belongs to the *Eolimna* genera, but it is not identified at a species level [17]. There are other strains reported with a 99% identity match for 18SV4 described as *Eolimna, Sellaphora* and *Navicula minima*, which are homotypic synonyms. The rbcL sequence obtained showed 96% similarity with a strain identified as *E. minima* from Germany [17].

JA01a morphological traits are consistent with *Achnantidium minutissimum* [46]. It shows linear-lanceolate valves, rounded and rostrate ends, narrow axial area, linear, widening towards the center, straight raphe, filiform, curved terminal fissures, radiate striae, more widely spaced in the central area of the valve. Valve dimensions: Length: 6–7 m; Width: 2–3 m; 30–32 striae in 10 m. (Figure 2q–u). Blast results for both barcodes show from 98 to 100% identity with *A. minutissimum* from Europe, Asia and North America [9, 32, 37, 40]. There are a few strains from Japan and Île de Reunion that have less than 96% identity for 18SV4 (Data unpublished).

Phylogenetic trees were constructed based on 18SV4 and rbcL regions separately due to the lack of strains with both barcode sequences deposited in INSDC databases. The phylogenetic analyses include sequence representatives of each species from different geographical locations reported previously. In the phylogeny, both barcode sequences of the same species clustered into one clade (Figures 3 and 4). Species level clades are well supported statistically. Short branches in species clades mean low divergence between the Ecuadorian strains and their homologous strains. In the phylogenetic tree based on the 18SV4 region (Figure 3), *N. fonticola* sequences are in the same clade as *E. minima* and *S. seminulum* but separated from *N. palea*. However bifurcations representing the relationship between the species are not well supported by bootstrap values. Nevertheless, in the phylogeny of rbcL (Figure 4), sequences belonging to *Nitzschia* genera clustered together. *E. minima* and *S. seminulum* are grouped in the same clade in both phylogenetic analyses.

4. Discussion

To the best of our knowledge, this study is the first attempt to establish a barcode reference library for epilithic diatoms from Andean streams of Ecuador. Five strains considered as water quality bioindicators have been cultured and sequenced for rbcL and 18SV4 barcodes. *N. palea, N. fonticola* and *S. seminulum* are tolerant to organic pollution and eutrophication. *N. palea* regularly occurs downstream of sewage treatment plants [47]. *E. minima* is abundant in mesotrophic water. *A. seminulum* is a small monoraphid diatoms of oligo/mesosaprobic rivers [47]. These species are very important from an ecological point of view, since they are often among the most abundant benthic species in freshwater systems. Furthermore, *N. palea, A. minutissimum* and *E. minima* have been used to assess sensitivity to herbicides [37, 48] and metals [49].

The occurrence and/or abundance of epilithic diatoms to define ecological status through morphological characteristics has some limitations [7]. The use of DNA sequences could potentially alleviate these limitations [50]. The barcoding sequences of the strains presented here have a high percentage of identity with the sequences reported in INSDC databases for the same species. Hence, DNA barcoding, based on 18SV4 and rbcL, emerges as a useful tool for diatom species identification in Andean streams. Nonetheless, we found that there are reported sequences for the same species with a divergence of at least 5%. This could be due to an inaccurate identification, since there are poor circumscriptions and a lack of reliable information about the epithets '*minima*', and '*seminulum*' even though they are apparently well–established and often referred to in ecological and taxonomic literature [51–53].
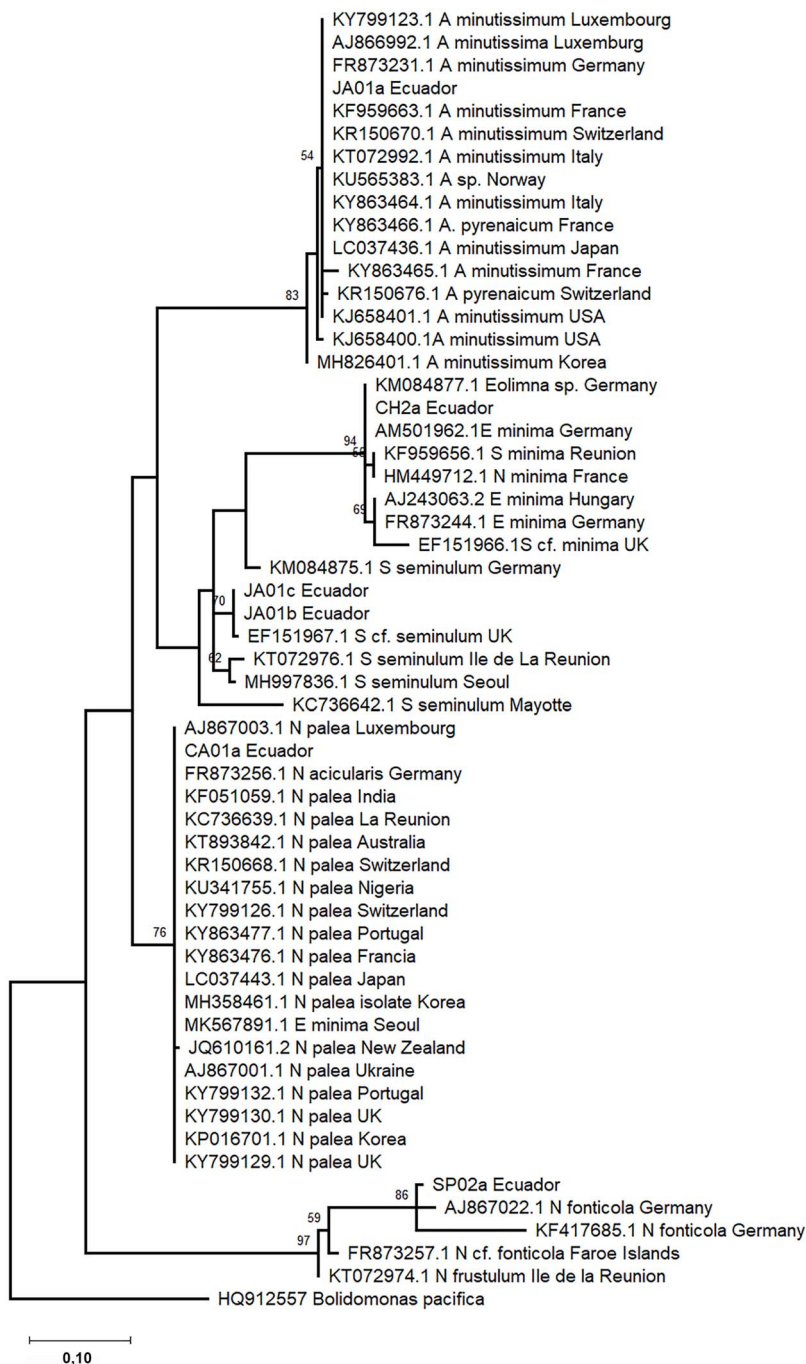
KY799123.1 A minutissimum Luxembourg
AJ866992.1 A minutissima Luxemburg
FR873231.1 A minutissimum Germany
JA01a Ecuador
KF959663.1 A minutissimum France
KR150670.1 A minutissimum Switzerland
54 KT072992.1 A minutissimum Italy
KU565383.1 A sp. Norway
KY863464.1 A minutissimum Italy
KY863466.1 A. pyrenaicum France
LC037436.1 A minutissimum Japan
KY863465.1 A minutissimum France
83 KR150676.1 A pyrenaicum Switzerland
KJ658401.1 A minutissimum USA
KJ658400.1A minutissimum USA
MH826401.1 A minutissimum Korea
KM084877.1 Eolimna sp. Germany
CH2a Ecuador
AM501962.1E minima Germany
94 KF959656.1 S minima Reunion
66 HM449712.1 N minima France
AJ243063.2 E minima Hungary
69 FR873244.1 E minima Germany
EF151966.1S cf. minima UK
KM084875.1 S seminulum Germany
70 JA01c Ecuador
JA01b Ecuador
EF151967.1 S cf. seminulum UK
62 KT072976.1 S seminulum Ile de La Reunion
MH997836.1 S seminulum Seoul
KC736642.1 S seminulum Mayotte
AJ867003.1 N palea Luxembourg
CA01a Ecuador
FR873256.1 N acicularis Germany
KF051059.1 N palea India
KC736639.1 N palea La Reunion
KT893842.1 N palea Australia
KR150668.1 N palea Switzerland
KU341755.1 N palea Nigeria
KY799126.1 N palea Switzerland
76 KY863477.1 N palea Portugal
KY863476.1 N palea Francia
LC037443.1 N palea Japan
MH358461.1 N palea isolate Korea
MK567891.1 E minima Seoul
JQ610161.2 N palea New Zealand
AJ867001.1 N palea Ukraine
KY799132.1 N palea Portugal
KY799130.1 N palea UK
KP016701.1 N palea Korea
KY799129.1 N palea UK
86 SP02a Ecuador
AJ867022.1 N fonticola Germany
59 KF417685.1 N fonticola Germany
97 FR873257.1 N cf. fonticola Faroe Islands
KT072974.1 N frustulum Ile de la Reunion
HQ912557 Bolidomonas pacifica

0,10

Figure 3. Phylogram constructed using 18SV4 rRNA sequences. The phylogram was constructed using the maximum-likelihood method with Tamura-Nei protocol. Numerical values at the nodes of the branches indicate bootstrap values above 50%. Accession number and geographical origin is indicated for each sequence (Supplementary Table S1).

For instance, *E. minima* can barely be identified using light microscopy (LM) and the lack of illustra- tions from a scanning electron microscope (SEM) is a major impediment [7]. Moreover, there are iden-
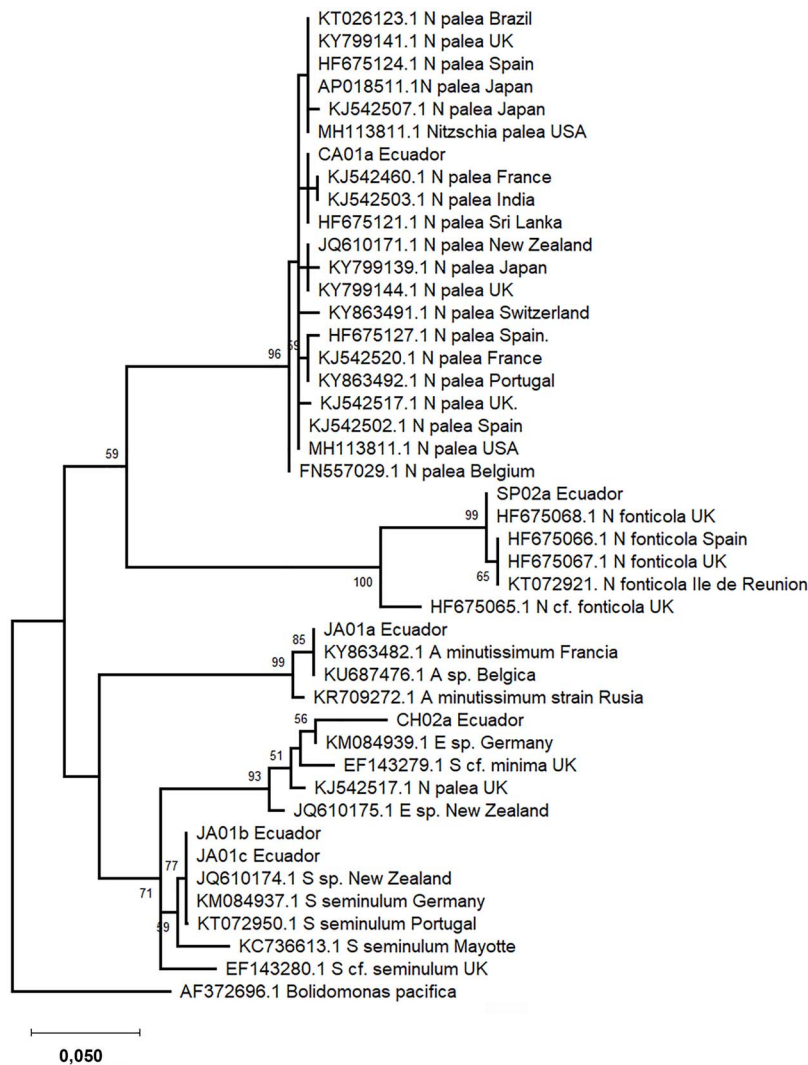
Figure 4. Phylogram constructed using rbcL sequences. The phylogram was constructed using the maximum-likelihood method with Tamura-Nei protocol. Numerical values at the nodes of the branches indicate bootstrap values above 50%. Accession number and geographical origin is indicated for each sequence (Supplementary Table S2).

tification problems derived from isolation and culturing, which are needed to obtain barcodes, given that some species experience cell size reduction in culture [54]. That could be the case for *S. seminulum* strains, JA01b and c, which are different in size but identical for rbcL and 18SV4 sequences. This supports the necessity of barcode reference libraries that couple morphological and molecular data for comparative analyses [12, 17].

The usefulness of 18SV4 and rbcL barcodes to assess phylogenetic relationships in diatoms has been proven extensively [6, 17, 18, 55, 56]. The topology of the phylogenetic trees demonstrates that epilithic diatoms from Ecuador show close relatedness to those of same species isolated from other geographical regions. This could mean low phylogeographic variability in diatoms. However, the *E. minima* clade presents the highest divergences, but

this could be due to identification problems related with this species. The genera clades are congruent with previous analysis where *Sellaphora* falls into one group with *Eolimna* [17] and these are separate from *Nitzschia* and *Achnantidium* [18].

This research highlights the complementary aspects of classical taxonomy and DNA barcoding. All reference sequences presented here are linked to morphological detailed images in order to initiate a complete barcode reference library for diatoms in the neotropical region. In addition, this study demonstrates that it would be feasible to use the existing barcoding data for diatoms to develop molecular tools for bioassessment of aquatic ecosystems in the Ecuadorian Andean region.

Acknowledgements

Supplementary data

Supporting information for this article is available on the journal's website under https://doi.org/10.5802/crbiol.2 or from the author.

References

[1] D. G. Mann, J. Droop, "Biodiversity, biogeography and conservation of diatoms", *Hydrobiologia* **336** (1996), no. 1–3, p. 19-32.

[2] D. Mann, P. Vanormelingen, "An inordinate fondness? The number, distributions, and origins of diatom species", *J. Eukaryot. Microbiol.* **60** (2013), no. 4, p. 414-420.

[3] M. Kelly, L. King, B. Ní Chatháin, "The conceptual basis of ecological-statusassessments using diatoms", *Biol. Environ. Proc. R. Irish Acad.* **109** (2009), no. 3, p. 175-189.

[4] E. A. Lobo, C. G. Heinrich, M. Schuch, C. E. Wetzel, L. Ector, "Diatoms as bioindicators in rivers", in *River Algae*, Springer International Publishing, Cham, 2016, p. 245-271.

[5] E. A. Lobo, M. Schuch, C. G. Heinrich, A. B. da Costa, A. Düpont, C. E. Wetzel, L. Ector, "Development of the Trophic Water Quality Index (TWQI) for subtropical temperate Brazilian lotic systems", *Environ. Monit. Assess.* **187** (2015), no. 6, p. 354.

[6] B. Beszteri, E. Acs, J. Makk, G. Kovács, K. Márialigeti, K. T. Kiss, "Phylogeny of six naviculoid diatoms based on 18S rDNA sequences", *Int. J. Syst. Evol. Microbiol.* **51** (2001), no. Pt 4, p. 1581-1586.

[7] D. Hering, A. Borja, J. I. Jones, D. Pont, P. Boets, A. Bouchez, K. Bruce, S. Drakare, B. Hänfling, M. Kahlert, F. Leese, K. Meissner, P. Mergen, Y. Reyjol, P. Segurado, A. Vogler, M. Kelly, "Implementation options for DNA-based identification into ecological status assessment under the European Water Framework Directive", *Water Res.* **138** (2018), p. 192-205.

[8] P. D. N. Hebert, S. Ratnasingham, J. R. Dewaard, "Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species", *Proc. Biol. Sci.* **270** (2003), no. Suppl 1, p. S96-S99.

[9] J. Zimmermann, R. Jahn, B. Gemeinholzer, "Barcoding diatoms: evaluation of the V4 subregion on the 18S rRNA gene, including new primers and protocols", *Org. Divers. Evol.* **11** (2011), no. 3, p. 173-192.

[10] F. Pompanon, E. Coissac, P. Taberlet, "Metabarcoding a new way to analyze biodiversity", *Biofutur* **319** (2011), p. 30-32.

[11] K. M. Ruppert, R. J. Kline, M. S. Rahman, "Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: a systematic review in methods, monitoring, and applications of global eDNA", *Glob. Ecol. Conserv.* **17** (2019), e00547.

[12] D. Mora, N. Abarca, S. Proft, J. H. Grau, N. Enke, J. Carmona, O. Skibbe, R. Jahn, J. Zimmermann, "Morphology and metabarcoding: a test with stream diatoms from Mexico highlights the complementarity of identification methods", *Freshw. Sci.* **38** (2019), no. 3, p. 448-464.

[13] L. Apothéloz-Perret-Gentil, A. Cordonier, F. Straub, P. Esling, J. Pawlowski, "Taxonomy-free molecular diatom index for high-throughput eDNA biomonitoring", *Mol. Ecol. Resour.* **17** (2017), no. 6, p. 1231-1242.

[14] V. Vasselon, F. Rimet, K. Tapolczai, A. Bouchez, "Assessing ecological status with diatoms DNA metabarcoding: scaling-up on a WFD monitoring network (Mayotte island, France)", *Ecol. Indic.* **82** (2017), p. 1-12.

[15] J. Zimmermann, G. Glückner, R. Jahn, N. Enke, B. Gemeinholzer, "Metabarcoding vs. morphological identification to assess diatom diversity in environmental studies", *Mol. Ecol. Resour.* **15** (2015), no. 3, p. 526-542.

[16] L. Kermarrec, A. Franc, F. Rimet, P. Chaumeil, J. M. Frigerio, J. F. Humbert, A. Bouchez, "A next-generation sequencing approach to river biomonitoring using benthic diatoms", *Freshw. Sci.* **33** (2014), no. 1, p. 349-363.

[17] J. Zimmermann, N. Abarca, N. Enke, O. Skibbe, W. H. Kusber, R. Jahn, "Taxonomic reference libraries for environmental barcoding: a best practice example from diatom research", *PLoS ONE* **9** (2014), no. 9, e108793.

[18] F. Rimet, P. Chaumeil, F. Keck, L. Kermarrec, V. Vasselon, M. Kahlert, A. Franc, A. Bouchez, "R-Syst::diatom: an open-access and curated barcode database for diatoms and freshwater monitoring", *Database* (2016), p. 1-21, baw016.

[19] V. Geissler, "Experimentelle Untersuchungen zur Vari-abilität der Schalenmerkmale bei einigen zentrischenSü sswasser-Diatomeen", *Nova. Hedwig. Beih.* **73** (1982), p. 211-246.

[20] M. Schuch, M. Oliveira, E. A. Lobo, "Spatial response of

epilithic diatom communities to downstream nutrient increases", *Water Environ. Res.* **87** (2015), no. 6, p. 547-558.

[21] P. Castillejo, S. Chamorro, L. Paz, I. Carrillo, J. Salazar, J. C. Navarro, C. Heinrich, E. A. Lobo, "Response of epilithic diatom communities to environmental gradients along an Ecuadorian Andean River", *C. R. Biol.* **341** (2018), no. 4, p. 256-263.

[22] J. Pawlowski, F. Lejzerowicz, L. Apotheloz-Perret-Gentil, J. Visco, P. Esling, "Protist metabarcoding and environmental biomonitoring: time for change", *Eur. J. Protistol.* **55** (2016), p. 12-25.

[23] M. L. MacGillivary, I. Kaczmarska, "Survey of the efficacy of a short fragment of the rbcL gene as a supplemental DNA barcode for diatoms", *J. Eukaryot. Microbiol.* **58** (2011), no. 6, p. 529-536.

[24] L. Kermarrec, A. Franc, F. Rimet, P. Chaumeil, J. F. Humbert, A. Bouchez, "Next-generation sequencing to inventory taxonomic diversity in eukaryotic communities: a test for freshwater diatoms", *Mol. Ecol. Resour.* **13** (2013), no. 4, p. 607-619.

[25] H. Kobayasi, S. Mayama, "Most pollution-tolerant diatoms of severely polluted rivers in the vicinity of Tokyo", *J. Phycol.* **30** (1982), p. 188-196.

[26] R. A. Andersen, J. A. Berges, P. J. Harrison, "Recipes for freshwater and seawater media", in *Algal Culturing Techniques* (R. A. Andersen, ed.), Elsevier Academic Press, London, 2005, p. 429-538.

[27] M. G. Potapova, D. F. Charles, "Benthic diatoms in USA rivers: distributions along spatial and environmental gradients", *J. Biogeogr.* **29** (2002), no. 2, p. 167-187.

[28] D. Metzelin, H. Lange-Bertalot, "Tropical Diatoms of South America II. Special remarks on biogeographic disjunction", *Iconogr. Diatomol.* **18** (2007), p. 1-877.

[29] D. Metzelin, F. Las García-Rodríguez, *Las diatomeas Uruguayas*, División de Relaciones y Actividades Culturales (DI.R.A.C), Universidad de la República, Montevideo, Uruguay, 2003, 1-207 pages.

[30] U. Rumrich, H. Lange-Bertalot, M. Rumrich, "Diatomeen der Anden. Von Venezuela bis Patagonien (Feuerland)", *Iconogr. Diatomol.* **9** (2000), p. 1-649.

[31] K. Edwards, C. Johnstone, C. Thompson, "A simple and rapid method for the preparation of plant genomic DNA for PCR analysis", *Nucleic Acids Res.* **19** (1991), no. 6, p. 1349.

[32] J. A. Visco, L. Apothéloz-Perret-Gentil, A. Cordonier, P. Esling, L. Pillet, J. Pawlowski, "Environmental monitoring: inferring the diatom index from next-generation sequencing data", *Environ. Sci. Technol.* **49** (2015), no. 13, p. 7597-7605.

[33] K. R. Stoof-Leichsenring, L. S. Epp, M. H. Trauth, R. Tiedemann, "Hidden diversity in diatoms of Kenyan Lake Naivasha: a genetic approach detects temporal variation", *Mol. Ecol.* **21** (2012), no. 8, p. 1918-1930.

[34] K. Bruder, L. K. Medlin, "Molecular assessment of phylogenetic relationships in selected species/genera in the naviculoid diatoms (Bacillariophyta). I. The genus Placoneis", *Nova. Hedwigia.* **85** (2007), no. 3–4, p. 331-352.

[35] S. Kumar, G. Stecher, K. Tamura, "MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets", *Mol. Biol. Evol.* **33** (2016), no. 7, p. 1870-1874.

[36] R. C. Edgar, "MUSCLE: multiple sequence alignment with high accuracy and high throughput", *Nucleic Acids Res.* **32** (2004), no. 5, p. 1792-1797.

[37] S. M. Esteves, F. Keck, S. F. P. Almeida, E. Figueira, A. Bouchez, F. Rimet, "Can we predict diatoms herbicide sensitivities with phylogeny? Influence of intraspecific and interspecific variability", *Ecotoxicology* **26** (2017), no. 8, p. 1065-1077.

[38] K. Tamura, M. Nei, "Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees", *Mol. Biol. Evol.* **10** (1993), p. 512-526.

[39] S. Blanco, C. Cejudo-Figueiras, I. Álvarez Blanco, E. Bécares, L. Hoffmann, L. Ector, Atlas de las Diatomeas de la cuenca del Duero, Diatom atlas of the Duero Basin. 2010.

[40] F. Keck, F. Rimet, A. Franc, A. Bouchez, "Phylogenetic signal in Diatom ecology: perspectives for aquatic ecosystems biomonitoring", *Ecol. Appl.* **26** (2016), no. 3, p. 861-872.

[41] K. M. Evans, A. H. Wortley, G. E. Simpson, V. A. Chepurnov, D. G. Mann, "A molecular systematic approach to explore diversity within the *Sellaphora pupula* species complex (Bacillariophyta)", *J. Phycol.* **44** (2008), no. 1, p. 215-231.

[42] R. Trobajo, D. G. Mann, V. A. Chepurnov, E. Clavero, "Taxonomy, life cycle, and auxosporulation of *Nitzschia Fonticola* (Bacillariophyta)", *J. Phycol.* **42** (2006), p. 1353-1372.

[43] E. Mora, D. Jiménez, J. Cantoral, "Epilithic diatoms in the Upper Laja River Basin, Guanajuato, Mexico", *Rev. Mex. Biodivers.* **86** (2015), p. 1024-1040.

[44] F. Rimet., R. Trobajo, D. G. Mann, L. Kermarrec, A. Franc, I. Domaizon, A. Bouchez, "When is sampling complete? The effects of geographical range and marker choice on perceived diversity in Nitzschia palea (Bacillariophyta)", *Protist* **165** (2014), no. 3, p. 245-259.

[45] G. Moser, H. Lange-Bertalot, D. Metzeltin, "Insel der Endemiten Geobotanisches Phänomen Neukaledonien (Island of endemics New Caledonia - a geobotanical phenomenon)", *Bibl. Diatomol.* **38** (1998), p. 464.

[46] M. Potapova, P. B. Hamilton, "Morphological and ecological variation within the *Achnanthidium minutissimum* (Bacillariophyceae) species complex", *J. Phycol.* **43** (2007), no. 3, p. 561-575.

[47] H. Van Dam, A. Mertens, J. Sinkeldam, "A coded checklist and ecological indicator values of freshwater diatoms from The Netherlands", *Neth. J. Aquat. Ecol.* **28** (1994), no. 1, p. 117-133.

[48] S. Moisset, S. Kim Tiam, A. Feurtet-Mazel, S. Morin, F. Delmas, N. Mazzella, P. Gonzalez, "Genetic and physiological responses of three freshwater diatoms to realistic diuron exposures", *Environ. Sci. Pollut. Res. Int.* **22** (2015), no. 6, p. 4046-4055.

[49] S. Kim Tiam, A. Feurtet-Mazel, F. Delmas, N. Mazzella, S. Morin, G. Daffe, P. Gonzalez, "Development of q-PCR approaches to assess water quality: effects of cadmium on gene expression of the diatom *Eolimna minima*", *Water Res.* **46** (2012), no. 4, p. 934-942.

[50] J. Pawlowski, M. Kelly-Quinn, F. Altermatt, L. Apothéloz-Perret-Gentil, P. Beja, A. Boggero, A. Borja, A. Bouchez, T. Cordier, I. Domaizon, M. J. Feio, A. F. Filipe, R. Fornaroli, W. Graf, J. Herder, B. van der Hoorn, J. I. Jones, M. Sagova-Mareckova, C. Moritz, J. Barquín, J. J. Piggott, M. Pinna, F. Rimet, B. Rinkevich, C. Sousa-Santos, V. Specchia, R. Trobajo, V. Vasselon, S. Vitecek, J. Zimmerman, A. Weigand,

F. Leese, M. Kahlert, "Science of the Total Environment The future of biotic indices in the ecogenomic era : integrating (e) DNA metabarcoding in biological assessment of aquatic ecosystems", *Sci. Total Environ.* **638** (2018), p. 1295-1310.

[51] E. A. Lobo, H. Kobayasi, "Shannon's diversity index applied to some freshwater diatom assemblages in the sakawa river system", *Jpn. J. Phycol.* **38** (1990), p. 229-243.

[52] P. Siver, P. B. Hamilton, "Diatoms of North America: the Freshwater Flora of Waterbodies on the Atlantic Coastal Plain", *Iconogr. Diatomol.* **22** (2011), p. 1-916.

[53] J. Żelezna Wieczorek, "Diatom flora in springs of Łódź Hills (Central Poland)", in *Biodiversity, Taxonomy, and Temporal Changes of Epipsammic Diatom Assemblages in Springs Affected by Human Impact* (A. Witkowski, ed.), Diatom Monographs, Gantner Verlag, Ruggell, Liechtenstein, 2011, p. 419.

[54] J. S. Ki, S. Y. Cho, T. Katano, S. W. Jung, J. Lee, B. S. Park, S. H. Kang, M. S. Han, "Comprehensive comparisons of three pennate diatoms, *Diatoma tenuae, Fragilaria vaucheriae*, and *Navicula pelliculosa*, isolated from summer Arctic reservoirs (Svalbard 79° N), by fine-scale morphology and nuclear 18S ribosomal DNA", *Polar Biol.* **32** (2009), no. 2, p. 147-159.

[55] C. Santhosh Kumar, V. A. Prabu, C. P. Kumar, "DNA Barcode Genes (rbcL, 18s rRNA and ITS Phylogeny) in *Skeletonema costatum* Grevelli (Cleve, 1873)", *Int. J. Curr. Microbiol. Appl. Sci.* **4** (2015), no. 9, p. 195-203.

[56] L. Guo, Z. Sui, S. Zhang, Y. Ren, Y. Liu, "Comparison of potential diatom 'barcode' genes (The 18S rRNA gene and ITS, COI, rbcL) and their effectiveness in discriminating and determining species taxonomy in the Bacillariophyta", *Int. J. Syst. Evol. Microbiol.* **65** (2015), no. 4, p. 1369-1380.