

Review

Genetic Markers for Metabarcoding of Freshwater Microalgae: Review

Elena Kezlya , Natalia Tseplik  and Maxim Kulikovskiy

Laboratory of Molecular Systematics of Aquatic Plants, K.A. Timiryazev Institute of Plant Physiology RAS, IPP RAS, 127276 Moscow, Russia; ntseplik@gmail.com (N.T.); max-kulikovskiy@yandex.ru (M.K.)

* Correspondence: melosira@mail.ru

Simple Summary: The metabarcoding approach is widely used for studying the diversity and distribution of freshwater microalgae and for routine biomonitoring. Due to microalgae being a phylogenetically diverse group, the choice of a genetic marker directly affects the metabarcoding results. Specific markers are good for identifying only concrete groups, while universal markers may miss classes or lack the variability necessary for differentiating taxa at the species and sometimes genus levels. An analysis of publications on the subject showed that metabarcoding studies of eukaryotic freshwater microalgae used 12 markers (different nuclear regions 18S and ITS and plastid regions *rbcL*, 23S and 16S). Studies that compared outcomes from different markers show that the resulting lists of taxa do not match. The plastid marker *rbcL* is widely used for diatom metabarcoding, as it differentiates taxa at the species and intraspecies levels, and there is a specific set of primers designed for identifying Eustigmatophyceae. The V9 18S region is more variable than V4 18S and provides more diversity at higher taxonomic levels (supergroup and phylum). The ITS1 and ITS2 regions are used rarely and may be underestimated. These barcodes amplify well with the standard primers and are variable enough to identify sequences at the species level. Plastid markers (23S and 16S rDNA) focused on the plastid-containing eukaryotic algae and Cyanobacteria, conserved regions, identify taxa to the genus level and higher. Using specialized curated databases for data interpretation significantly improves the quality of the results.



Citation: Kezlya, E.; Tseplik, N.; Kulikovskiy, M. Genetic Markers for Metabarcoding of Freshwater Microalgae: Review. *Biology* **2023**, *12*, 1038. <https://doi.org/10.3390/biology12071038>

Academic Editors: Lijuan Ren, Xiaomin Xia and Jin Zeng

Received: 23 June 2023

Revised: 14 July 2023

Accepted: 18 July 2023

Published: 22 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: The metabarcoding methods for studying the diversity of freshwater microalgae and routine biomonitoring are actively used in modern research. A lot of experience has been accumulated already, and many methodological questions have been solved (such as the influence of the methods and time of sample conservation, DNA extraction and bioinformatical processing). The reproducibility of the method has been tested and confirmed. However, one of the main problems—choosing a genetic marker for the study—still lacks a clear answer. We analyzed 70 publications and found out that studies on eukaryotic freshwater microalgae use 12 markers (different nuclear regions 18S and ITS and plastids *rbcL*, 23S and 16S). Each marker has its peculiarities; they amplify differently and have various levels of efficiency (variability) in different groups of algae. The V4 and V9 18S and *rbcL* regions are used most often. We concentrated especially on the studies that compare the results of using different markers and microscopy. We summarize the data on the primers for each region and on how the choice of a marker affects the taxonomic composition of a community.

Keywords: barcode; metabarcoding; ecological assessment; microalgae; genetic markers

1. Introduction

Currently eDNA metabarcoding is a popular method for studying the diversity and functioning of various communities, from microbes to mammals. Interest in this method grows every year, and the number of studies increases. For example, a query in the SCOPUS database with the keyword “metabarcoding” returns 2215 results; a query with the keyword “eDNA” returns 26,034 results (date of search 12 February 2023).

Algae are a phylogenetically heterogeneous group of organisms that is very diverse in morphology and ecological preferences. In the eukaryotic tree of life, photosynthetic eukaryotes are spread across 12 separate phylogenetic lines at the level of phylum [1–3]. On a macrosystematic level, they belong to four to seven (according to different estimates) supergroups that also contain non-photosynthetic organisms in each clade [1–6]. This phylogenetical heterogeneity is connected with a gene locus “...which is variable enough to provide robust identification at the species level...” [7] (and references in it) “...and different markers are applied for species delimitation in different algal groups.” [8] (and references in it). For example, phylogenetic studies and species descriptions of diatoms and red algae do not use the rDNA ITS marker, whereas it is the main marker currently employed for DNA-based species of green microalgae, Dinoflagellates, Chrysophytes and Synurophytes [7,8] (and references in it). Well-documented nucleotide sequences are accumulated in databases, which are the basis of interpreting the metabarcoding data.

Thus, the choice of the barcode region and primer pairs, which can limit or bias the diversity of organisms observed, is a challenge with environmental metabarcoding studies [9,10]. The proportion of biodiversity covered by metabarcoding studies directly depends on the markers and primers used, so organisms that are not amplified by standard methods go undetected, even if they are common and play an important role in the ecosystem [11]. It is important for a “good barcode” to be taxonomically informative; it needs to be able to distinguish between species (i.e., the DNA region should mutate at the right rate), because most modern biomonitoring and biotic index programs require identification at the species level. At the same time, a barcode needs conserved primer binding areas, or degenerate primers, in order to be able to attach to the DNA of all the organisms in the sample [12]. The choice of primers also impacts the results of a biodiversity assessment of an ecosystem. Complete universality causes a loss of resolution and limits the depth of the biodiversity assessments of groups. Limiting the universality of the primers might, on the other hand, exclude important groups in the analysis and introduce biases, favoring some organisms or groups. Furthermore, the use of different universal primers makes direct comparisons between studies more challenging [13] (and references in it). All these conditions show that metabarcoding is not a simple and universal method of monitoring and biodiversity studies of algae, as it has its limitations, and further development and tuning are needed. The choice of marker also plays an important role in the interpretation of results.

There is a lot of experience already gained in using next-generation sequencing (NGS) approaches for studying algae. One of the high-priority research areas is the integration of metabarcoding into routine biomonitoring. Many methodological questions have been answered; bioinformatics pipelines have been assessed [14,15], sampling, DNA extraction methods and applications of global eDNA have been discussed [16–21], and recently, it has been shown that the preservation time and sample preservation methods have little effect on DNA metabarcoding results [22], the experience of integrating eDNA metabarcoding into routine freshwater biomonitoring has been summarized [23–25] and the terminology “eDNA” has been clarified [26]). In a recent study, Salmaso et al. [27] looked into the problem of a taxonomic gap in reference databases for aquatic cyanobacteria and eukaryotic microalgae and the effect it has on the interpretation of metabarcoding data. Extensive reviews of the methodology of DNA metabarcoding in marine bulk samples have been published, including benthic communities of sediment and hard substrate, plankton samples and dietary samples [12] and freshwater harmful algae *Microcystis aeruginosa* and *Prymnesium parvum* [28]. The state of DNA barcoding of macroalgae in the Mediterranean Sea has also been reviewed [29].

In metabarcoding studies dedicated to freshwater eukaryotic algae, various genetic markers (nuclear regions 18S V3, V4, V4–V5, V7, V7–V9, V9, V9-ITS1 and ITS2 and plastid regions *rbcL*, 16S and 23S) and various primer sets have been used (Table 1). The aim of this review is to summarize the available information and to critically assess which markers and primers are the most effective for metabarcoding freshwater algae, how they should be chosen, what the level of taxonomic coverage and resolution is and which databases are used for the taxonomic attribution of sequences.

Table 1. Different available markers and primer sets used in bulk metabarcoding studies of freshwater algae.

Number of Primer Set	Gene Region	Target Group	Primer Name	Primer Sequence 5' to 3' (Primer Author Reference)	Forward/Reverse	References	PCR Cycling
1	V3 18S	Eukaryotes		ATTAGGGTTTCGATTCCGGAGAGG CTGGAATTACCGCGGSTGCTG [31]	forward reverse	[30]	n.d.
	V4 18S	Diatoms	DIV4for: DIV4rev3	GCGGTAATTCCAGCTCCAATAG CTCTGACAATGGAATACGAATA [32]	forward reverse	[14,32–36]	94 °C—2 min (35 cycles: 94 °C—45 s, 50 °C—45 s, 72 °C—1 min), 72 °C—10 min
2	V4 18S	Protist, Diatoms	M13F–D512	TGT AAA ACG ACG GCC AGT ATT CCA GCT CCA ATA GCG	forward	[10,37–39]	94 °C—2 min, (5 cycles: 94 °C—45 s, 52/54 °C—45 s, 72 °C—1 min), (35 cycles: 94 °C—45 s, 50/52 °C—45 s, 72 °C—1 min), 72 °C—10 min.
			M13R–D978rev	CAG GAA ACA GCT ATG AC GAC TAC GAT GGT ATC TAATC [37]	reverse		
3	V4 18S	Eukaryotes	F574 1132r	GCGGTAATTCCAGCTCCAA [13] CCGTCAATTHCTTYAART [41] AATTCCAGCTCCAATAGCGTATAT	forward reverse	[40]	95 °C—5 min, (25 cycles: 98 °C—1 min, 98 °C—20 s, 51 °C—20 s, 72 °C—12 s), 72 °C—1 min.
4	V4 18S	Eukaryotes		TTTCAGCCTTGCGACCATAC [42]	forward reverse	[42]	98 °C—30 s, (30 cycles: 98 °C—10 s, 59 °C—30 s, 72 °C—30 s), 72 °C—10 min.
5	V4 18S	Eukaryotes	F574 R952	GCGGTAATTCCAGCTCCAA AAG ACG ATC AGA TAC C [13]	forward reverse	[13]	PCR <i>in silico</i> , Tm 55.3
6	V4 18S	Eukaryotes	TAReuk454FWD1	CCAGCA (G/C)C(C/T)GCGGTAATTCC [43]	forward	[10,27,44–52] *	94 °C—5 min, (15 cycles: 94 °C—30 s, 53 °C—45 s, 72 °C—1 min), (20 cycles: 94 °C—of 30 s, 48 °C—45 s, 72 °C—1 min), 72 °C—10 min.
			TAReukREV3	ACTTTCGTTCTTGAT(C/T)(A/G)A [43]	reverse		
			V4 forward	CCAGCAGCCGCGGTAATTCC [43] modified primers from [43]	forward	[53]	
			V4 reverse	ACTTTCGTTCTTGATTAA [43] modified primers from [43]	reverse	[53]	
7	V4 18S	Eukaryotes	TAReuk454FWD1	CCAGCA (G/C)C(C/T)GCGGTAATTCC [43]	forward	[54]	95 °C—5 min, (10 cycles: 94 °C—30 s, 57 °C—45 s, 72 °C—1 min), (15 cycles: 94 °C—30 s, 47 °C—45 s, 72 °C—1 min), 72 °C—10 min.
			V4r	ACTTTCGTTCTTGAT [54] modified primers from [43]	reverse		

Table 1. Cont.

Number of Primer Set	Gene Region	Target Group	Primer Name	Primer Sequence 5' to 3' (Primer Author Reference)	Forward/ Reverse	References	PCR Cycling		
1	V4–V5 18S	Eukaryotes	563f 1132r	GCCAGCAVCYGCGGTAAY CCGTCAATTHCTTYAART [41]	forward reverse	[41,55,56]	Two-step tailed PCR. Round 1: 95 °C for 3 min, (15 cycles: 95 °C—1 min, 55 °C—1 min, 72 °C—1 min), 72 °C—10 min (260 bp). Round 2: 98 °C—30 s, (10 cycles—98 °C—10 s, 55 °C—30 s, 72 °C—30 s), 72 °C—5 min. The annealing temperature of 52 °C The annealing temperature of 59.1 °C		
	V7 18S	Eukaryotic phytoplankton community	960F NSR1438	GGCTTAATTTGACTCAACRCG GGGCATCACAGACCTGTTAT [58]	forward reverse	[57]			
	V7–V8 18S	Eukaryotes	F-1183 R-1631	AAT TTG ACT CAA CAC GGG TAC AAA GGG CAG GGA CGT AAT [13]	forward reverse	[13]			
	V8–V9 18S	Eukaryotes	V8f 1422 1510R	ATAACAGGTCTGTGATGCCCT [54] GCCTTGCCAGCCCGCTCAG (eukaryotic) [59]	forward reverse	[30,54]			
	V9 18S	Eukaryotes	1391F EukBr	GTACACACCGCCCGTC [60] TGATCCTTCTGCAGGTTACCTAC [67]	forward reverse	[10,48,61–66]			
	2	V9 18S	Eukaryotes	1380F (eukaryotic)	CCCTGCCHTTTGTACACAC (eukaryotic)	forward		[53]	94 °C—3 min, 30 cycles: 94 °C—30 s, 57 °C—60 s, 72 °C—90 s), 72 °C—10 min 94 °C 10 min, (35 cycles: 94 °C—40 s, 58 °C—25 s, 72 °C—30 s), 72 °C—10 min.
				1389F	TTGTACACACCGCCC (universal)	forward			
				1510R	CCTTCYGCAGGTTACCTAC (eukaryotic) [59]	reverse			
	V9-ITS1	Protist	ITS2_Dino; 10% ITS2_broad; 90% ITS3	GTACACACCGCCCGTC GCTGCGCCCTTCATCGKTG GCTGCGTTCTTCATCGWTR GCATCGATGAAGAACGCAGC	forward reverse reverse forward	[68–71]		98 °C—3 min, (35 cycles: 98 °C—30 s, 52 °C—75 s, 72 °C—60 s), 72 °C—10 min.	
	ITS2	Chlorophyceae	ITS4	TCCTCCGCTTATTGATATGC [75]	reverse	[72–74]		n.d.	

Table 1. Cont.

Number of Primer Set	Gene Region	Target Group	Primer Name	Primer Sequence 5' to 3' (Primer Author Reference)	Forward/Reverse	References	PCR Cycling
1	<i>rbcL</i>	Diatoms	Diat_ <i>rbcL</i> _708F_1	AGGTGAAGTAAAAGGTTCTACTTAAA	forward	[14,16,22,33,34,36] ** [76–95]	95 °C—15 min, (30–40 cycles: 95 °C—45 s, 55 °C—45, 72 °C—45 s) (final extension).
			Diat_ <i>rbcL</i> _708F_2	AGGTGAAGTTAAAGGTTCTAYTTAAA	forward		
			Diat_ <i>rbcL</i> _708F_3	AGGTGAAACTAAAGGTTCTACTTAAA	forward		
			R3_1	CCTTCTAATTTACCWACWACTG	reverse		
			R3_2	CCTTCTAATTTACCWACAACAG [16]	reverse		
2	<i>rbcL</i>	Diatoms	<i>rbcL</i> 646F <i>rbcL</i> 998R	ATGCGTTGGAGAGARGTTTC GATCACCTTCTAATTTACCWACAACTG [17]		[17,46,86,96]	95 °C—15 min, (32–35 cycles: 95 °C—20 s, 55 °C—45 s, 72 °C—60 s), 72 °C—5 min.
3	<i>rbcL</i>	Eustigmatophyceae	EU <i>rbcL</i> 500FA Eustig <i>rbcL</i> -R900	GGNCGYGTWGTDTWYGAAGGT CACCWGCCATACGCATCC [97]	forward reverse	[97]	The annealing temperature of 53.5 °C
	23S	Protist	p23SrV_f1 p23SrV_r1	GGA CAG AAA GAC CCT ATG AA TCA GCC TGT TAT CCC TAG AG [100]	forward reverse	[10,98,99]	94 °C—2 min, (35 cycles: 94 °C—20 s, 55 °C—30 s, and 72 °C—30 s) 72 °C—10 min.
1	V3–V4 16S	Freshwater phytoplankton	341F 805R	CCTACGGGNGGCWGCAG GACTACHVGGGTATCTAATCC	forward reverse	[101]	95 °C—5 min, (25 cycles: 95 °C—40 s, 53 °C—40 s and 72 °C—1 min) 72 °C—7 min.
2	V4 16S	Diatom plastid	515F 806R	GTGYCAGCMGCCGCGGTAA [102] GGA CTA CHV GGG TWTCTA AT [104]	forward reverse	[103]	94 °C for 3 min, (30–35 cycles: 94 °C—30 s, 53 °C—40 s, 72 °C—1 min), 72 °C—5 min.

* in [48] forward named V4_1F, in [52] primers named 547F/V4R; ** in [36] used only Diat_ *rbcL*_708F_2 and R3_1 primers.

2. Materials and Methods

The search for literary sources was carried out in the SCOPUS database in February 2023 using the keywords “metabarcoding”, “algae”, “markers”, “barcode”, “freshwater”, “eDNA”, “diatom”, “protist”, “NGS”, “18S”, “ITS”, “23S”, “*rbcL*” and “16S” in combinations of two or three keywords. The results of each search were reviewed. The selection of appropriate publications was conducted according to the following criteria: (1) the research article or review was published in a peer-reviewed journal, (2) the research concerned freshwater algae or eukaryotic organisms in general, (3) the study examined the results of metabarcoding and (4) genetic markers were discussed. In total 70 studies published in the period from 2013 to 2023 were analyzed. Also, 5 studies of marine microalgal communities were included in the review in the section that discussed the comparison of metabarcoding results based on the V4 and V9 18S regions.

Simple histograms for a graphic representation of the obtained results were constructed using MS Excel. The list of publications used in analysis is provided in Table S1 (Supplementary Materials). The list of genetic markers and primer sets is provided in Table 1. The sets of primers were assigned conditional numbers for convenience (numbered in order for each region).

3. Results and Discussion

3.1. Gene Markers and Primer Sets for Freshwater Microalgae Metabarcoding

We found that 12 various genetic regions are used in the studies (Figure 1). The nuclear regions V3, V4, V7, V9 and V9-ITS1 are used for analyzing whole eukaryotic communities, as well as communities of microalgae, also focusing on individual groups of algae (dinoflagellates and diatoms [31] (Table S1)). The ITS2 region has been chosen for studying green algae s.l. (*Viridiplantae*) in a series of studies of the Antarctic region [72–74,105]. The plastid region *rbcL* is widely used for diatom metabarcoding, and also, primers for identifying Eustigmatophyceae have been designed and tested [97]. There are only three studies on cyanobacterial and eukaryotic algal diversity that were carried out using the universal plastid barcode 23S (Figure 1). The 16S rRNA gene has been used even less (in only two studies) as a universal marker for prokaryotes and eukaryotic algae.

Among nuclear markers, the V4 18S rRNA region is used for analyses most often (Figure 1). In the reviewed studies, we found seven options of primer sets, the most used of which was Set 6 (TAREuk454FWD1/TAREukREV3), developed by Stoeck et al. [43] (Table 1). This set is widely used in metabarcoding of both marine and freshwater eukaryotic plankton. Sets 1 (DIV4for/DIV4rev3) and 2 (M13F-D512/M13R-D978rev) are aimed at diatoms and used in seven and four studies, respectively. The remaining sets are all mentioned in only one publication each, apart from Set 7 (TAREuk454FWD1/V4r), which has been recently accepted as the standard for using environmental DNA in Finnish marine phytoplankton monitoring. The V9 18S region was chosen as a barcode in nine publications. The universal primer Set 1 (1391F/EukBr) was mostly used for the amplification of this region (in eight publications out of nine). In addition, this region and the primer set were used in a large-scale project called the Earth Microbiome Project (EMP; <http://www.earthmicrobiome.org> (accessed on 22 May 2023)). Set 2 (1380F(1389F)/1510R) was used in the research of a brackish lake [53].

Most of the studies on diatom metabarcoding used *rbcL* and primers Set 1 (Diat_ *rbcL* _708F_1, 2, 3/R3_1, 2, 312 bp), first suggested by Vasselon et al. [16]. The set designed by Kelly et al. [17] (*rbcL* 646F/*rbcL* 998R) for the adaptation of a DNA metabarcoding approach to ecological assessments within the Environment Agency’s routine monitoring program in the UK was used less often (4 publications out of 32).

We found only one publication each where the regions V3 18S, V4–V5 18S, V7, V7–V8 18S and V8–V9 18S were used as genetic markers. Studies on metabarcoding with the regions V9-ITS, ITS2, 23S and V4 16S all used one respective primer set (Table 1).

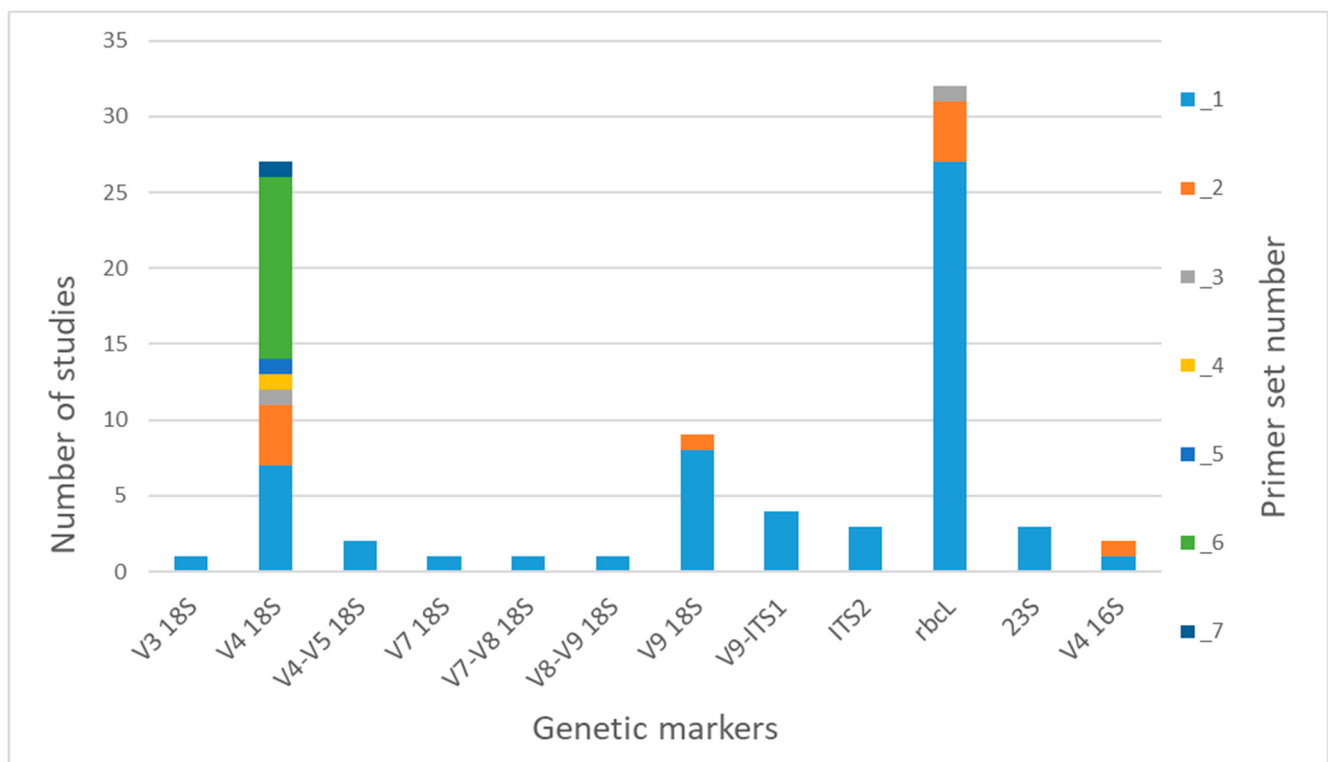


Figure 1. Gene markers and primer sets used for freshwater microalgae metabarcoding. The sets of primers have been assigned numbers that are listed in Table 1.

3.2. Reference Databases for Sequence Interpretation

During our literary analysis, we noticed that authors used different databases for taxonomic attributions of sequences (Figure 2). Studies on diatoms that use the *rbcL* region always use “Diat. barcode” (Rsyst:diatom database), a curated barcode library for diatoms [105] for sequence interpretation. Taxonomic attributions of sequences of various 18S rRNA (V3, V4, V7, V8, V9 and combinations) regions are usually carried out using GenBank, as well as quality-controlled databases of ribosomal RNA gene sequences such as “SILVA” [106]. The PR2 (Protist Ribosomal Reference) database—a catalog of unicellular eukaryote small subunit rRNA sequences with curated taxonomy—is used less often. In a series of studies on Antarctic green algae [72–74], the sequences were annotated using a recently established reference dataset PLANITS, which included the sequences of Viridiplantae ITS1, ITS2 and entire ITS sequences, including both Chlorophyta and Streptophyta [107]. To classify the 16S reads of freshwater diatom biofilm [103], PhytoREF, a reference database of the plastid 16S rRNA gene of photosynthetic eukaryotes, was used [108].

To sum up, metabarcoding studies most often use specialized reference datasets with curated taxonomy in order to interpret the sequences acquired during a study.

3.3. First Works on Testing Genetic Markers on Monoclonal Microalgal Cultures Provide Insight on the Effectiveness of Amplification and the Resolution of Species Differentiation

The first studies that tested the resolution of genetic markers for species differentiation were carried out using large collections of monoclonal algal cultures. It allowed to determine the effectiveness of primers in amplifying certain regions, directly compare the variability of sequences and morphological features (including cryptic species) and establish the regions that are most suitable for further research. These studies became the basis of choosing the markers for next-generation sequencing.

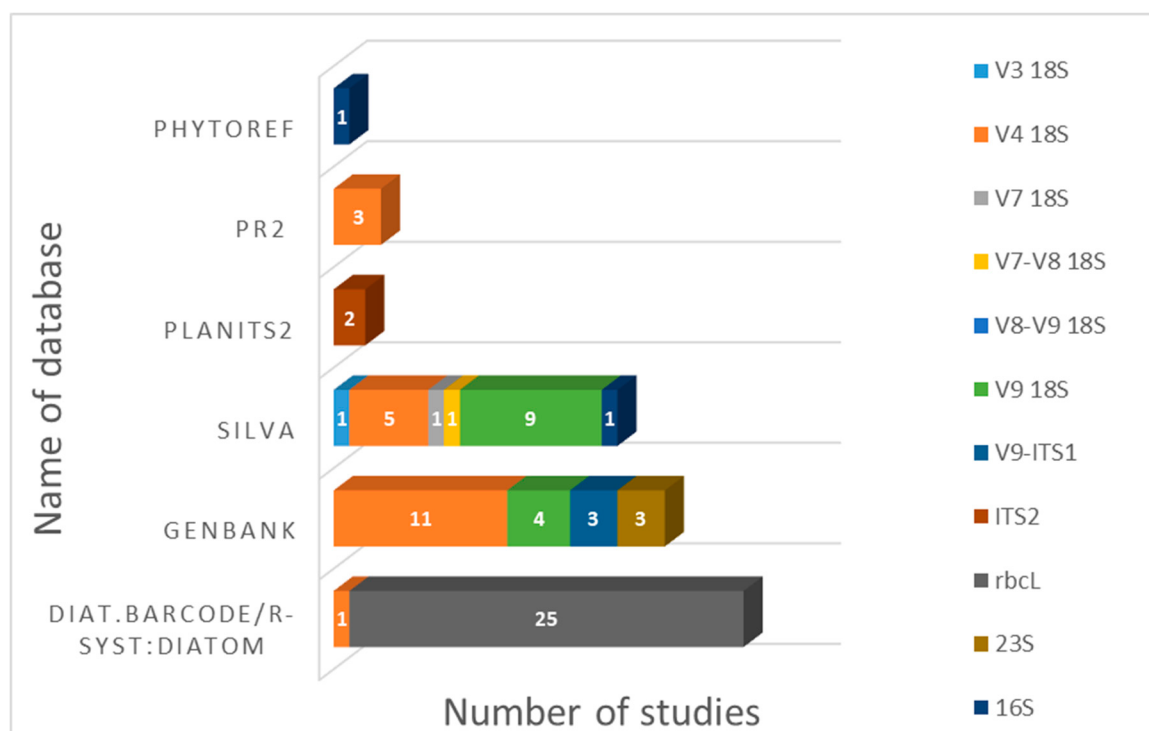


Figure 2. Usage of reference databases for annotating sequences in freshwater microalgae metabarcoding studies (Rsys:diatom version v7 renamed “Diat.barcode” in February 2018 [49]).

One of the first tests of diatom “barcode” genes (COI, *rbcL*, 18S and ITS rDNA) was done by Evans et al. in 2007 [109]. The study aimed to determine the effectiveness of markers in distinguishing cryptic species within the model “morphospecies” *Sellaphora pupula* agg. As a result of their analysis, the authors suggested the barcode region COI as a valuable phylogenetic marker. However, they also reported some difficulties with the amplification of this gene (a large primer set was used, sequences for *Seminavis* cf. *robusta* and for centric diatoms could not be obtained and only partial sequences were obtained for the araphid pennate diatom *Tabularia* sp.). According to the acquired data, the plastid gene *rbcL* is less variable than COI, but it supports all the phylogenetic lines of the latter. As for ITS, this barcode has a lot of variability in the length of the region, and there is also the problem of intraindividual variations. Behnke et al. [110] “recorded three types of ITS sequences that differed at 48 positions and two indels of 50 and 4 bp” within one *Sellaphora auldreekie* isolate.

Later, Moniz and Kaczmarek [111] tested as a barcode the small ribosomal subunit (SSU, 1600 bp), a 5′ end fragment of the cytochrome c oxidase subunit 1 (COI, 430 bp), and the second internal transcribed spacer region combined with the 5.8S gene (5.8S + ITS2, 300–400 bp) on 28 species from 22 genera of diatoms. COI showed the lowest rates of amplification (only 29% of good quality DNA amplified with COI, and of those, only 30% were sequenced successfully and found to be diatom DNA). For SSU, the authors noted the highest of all three success rates in amplification and easy alignment; however, a long fragment is required for species delimitation. 5.8S + ITS2 showed a higher rate of successful amplification and sequencing (79% and 84%, respectively), as it was the most variable of the three markers, but its secondary structure was needed to aid in alignment. As a result, the 5.8S + ITS2 fragment was proposed as the best candidate for a diatom DNA barcode. In their next work, M. Moniz and I. Kaczmarek [112] confirmed the successful use of 5.8S + ITS2 for differentiating diatoms on a large selection of sequences: 618 sequences representing 114 diatoms from classes Mediophyceae and Bacillariophyceae. In particular, a 99.5% success rate in separating species was shown and a 91% success rate in separating

species using a short barcode starting at the 5' end of 5.8S and ending in the conserved motif of helix III of ITS2 (300 to 400 bp).

A search for a universal marker for diatoms was carried out by Hamsher et al. [113]. The authors assessed the following markers: ~1400 bp of *rbcL*, 748 bp at the 3' end of *rbcL* (*rbcL*-3P), LSU D2/D3 and UPA. As a result, *rbcL*-3P was suggested as the primary marker for diatom barcoding, since it had the power to distinguish all species and could be sequenced more easily. LSU D2/D3 could distinguish all but the most closely related species (96%). UPA showed low resolution, distinguishing only 20% of the species. Relying on the authors' personal experiences (several copies were amplified, and the resulting sequences were different in length and unreadable), as well as the literary data, it was concluded that ITS is not a good barcode for diatoms.

The effectiveness of *rbcL* was discussed by M. MacGillivray and I. Kaczmarek [114]. A 540-bp fragment 417 bp downstream of the start codon of the *rbcL* gene was tested on a large selection of diatom taxa from classes Mediophyceae and Bacillariophyceae (381 sequences representing 66 genera and 245 species). This fragment was chosen after preliminary testing as the most variable. As a result, this fragment of *rbcL* correctly segregated 96% and 93% of the morphological congeners, respectively. The authors indicated a limitation in the resolution of biologically defined and closely related species (e.g., *Pseudo-nitzschia* and *Stephanodiscus*); using a $p = 0.02$ cut-off, only 80% of biological species were segregated. The authors noted that, with the total diversity of the diatoms (near 200,000 species), up to 40,000 species might be misidentified by their proposed *rbcL* barcode.

The effectiveness of three markers (SSU rDNA, *rbcL* and COI) for metabarcoding was tested on a mock community of diatom algae (30 strains belonging to 21 species) by Kermarrec et al. [115]. These markers are the primary ones used for the molecular identification of diatoms. The markers ITS and LSU were not considered in this study because of their high interclonal variability and the lack of available data for the establishment of reference libraries. In order to interpret the acquired sequences, reference libraries were created for each marker. Sequences from the authors' own collection and from GenBank were included in these libraries. Gene marker *rbcL* showed the best species composition assessment of the mock community, and SSU rDNA was next (it did not differentiate the complexes *Nitzschia palea* and *Gomphonema parvulum* at the intraspecific level). COI is variable and provides high resolution, but it was not recommended for routine metabarcoding due to difficulties in amplification and low representativity of the reference library.

A large work on assessing the utility of the gene markers COI, *rbcL*, ITS, *tufA*, UPA and 18S for freshwater green algae was done by Hall et al. [116]. They tested representatives of seven distantly related species groups from classes Chlorophyceae, Charophyceae and Zygnematophyceae (151 strains, 40 species total). As a result, the authors concluded that 18S, UPA and COI would be poor choices for a DNA barcode in green algae (18S and UPA proved insufficiently variable and COI difficult to amplify). ITS, *rbcL* and *tufA* were sufficiently variable to distinguish most species of Chlorophyceae, but additional primers were sometimes needed for amplification. For the charophytes, *rbcL* was noted as the most suitable primer but with a remark that it was impossible to differentiate species using this marker alone.

A detailed study of within-species and between-species genetic distances for ITS region (using 81 dinoflagellate species belonging to 14 genera) showed that "...the sequence of the dominant ITS region allele has the potential to serve as a unique species-specific "DNA barcode" that could be used for the rapid identification of dinoflagellates..." [117]. This idea has been supported by other research done on dinoflagellates [118–120].

Our search criteria did not reveal any similar research on other groups of algae; however, the review by Leliaert et al. [8] showed that the sets of main markers employed for DNA-based species delimitation in Chrysophytes, Cryptophytes and Raphidophytes included nuclear markers SSU rDNA and ITS, and for Xanthophytes, they also included ITS, whereas, for Euglenophytes, the barcode markers were plastid, and nuclear SSU rDNA, LSU rDNA and ITS were not used.

Summarizing the results of the first studies concerning the search of DNA barcodes for different groups of algae, we can conclude the following: the UPA region is insufficiently variable, COI is difficult to amplify, 18S can be amplified successfully but is insufficiently variable and LSU D2/D3 cannot distinguish the most closely related species in diatoms. For diatoms, the most effective genetic marker has proven to be *rbcL*; in green algae, this region is difficult to amplify (additional primers are needed). The ITS region successfully distinguishes species of Chlorophyceae and dinoflagellates, but in diatoms, alignment is difficult, and there are problems connected with a high level of intraspecific variability [109,110]. In charophytes, ITS is difficult in amplification. This fundamental research highlights the limitations of metabarcoding and explains the instances of common species being missed while using only one marker or taxonomic attribution being limited at the genus level.

3.4. 18S—Choosing a Variable Barcode Region for Eukaryotes In Silico

The eukaryotic gene 18S-rRNA is used for species delimitation in almost all groups of freshwater algae [8]. It contains nine hypervariable regions (V1 to V9), each of which has been considered as a short barcode for species identification (with the exception of V6, because this region is more conserved in eukaryotes) [121] (and references in it). The question of using hypervariable regions as barcode markers for eukaryotes in silico has been discussed in several publications.

Stoeck et al. [43] provided pairwise comparisons of 7503 publicly available sequences of dinoflagellates and showed that the V4 region is less variable compared to the V9 region (the number of homopolymers per sequence is 6.8 times higher in the V4 region compared to the V9 region). On the whole, V9 detected a wider range of higher taxonomic groups than V4.

Based on an alignment of eukaryotes containing 24,793 positions from the SILVA database, the characterization of the 18S rRNA gene and the design of universal eukaryote specific primers were provided by Hadziavdic et al. [13]. To describe the nucleotide variation in the alignment, the authors used Shannon entropy values. The results suggested that the V2, V4 and V9 regions were best suited for biodiversity assessments (they yielded the highest taxonomic resolutions at cut-off values ranging 95–100% for the sequence identity). The V1 region is rather short (ca 100 nt) and contains a highly conserved core segment, and the V3 and V5 regions lack highly variable segments and are not very long. V7 has a highly variable core of approximately 20–25 nt. The V8 region is over 150 nucleotides long with variable and conserved positions interspersed across the region, with a conserved segment towards the 3' end. The authors noted that there were no nucleotide segments of sufficient length for standard PCR along the whole gene that were entirely conserved within all eukaryotes while being absent in prokaryotes. Therefore, a single primer pair that will cover the full eukaryotic diversity and, at the same time, exclude prokaryotes cannot be designed. The authors mapped the available universal primers from the literature, as well as self-designed primers (total 100 non-degenerate eukaryote primers), and suggested two pairs of universal eukaryote-specific primers targeted to V4 (F574/R952) and V7–V8 (F-1183/R-1631) (Table 1). However, the authors noted that the coverage of eukaryotic taxa may be lower, as with the universal eukaryotic primers.

A comparative study of the validity of three regions of the 18S-rRNA gene (V1–3, V4–5 and V7–9) for the planktonic eukaryotic community was done by Tanabe et al. [121]. They showed that the V1–3 region (568 nt) has the highest variability and identification power, followed by the V7–9 region (484 nt), and the V4–5 region (415 nt) has the lowest variability. Based on in silico PCR analyses, the authors showed that the number of sequences from international nucleotide sequence databases (INSDs) such as DDBJ, EMBL and GenBank for the V4–5 region was 5–22 times higher than for V1–3 and 3–4 times higher than for V7–9. Nevertheless, the authors concluded that no significant difference was detected between the V1–3 and V7–9 regions, so the V1–3 region was suggested for the mass parallel sequencing-based monitoring of natural eukaryotic communities. Subsequently, the choice of genetic markers was limited by the use of the Illumina MiSeq platform (250–300 nt

single read length, resulting in ~450–500 nt-long combined reads with 50–150 bp overlap). Therefore, amplicons with length >500 nucleotides such as V1–3 (568 nt) were excluded [54].

Thus, based on in silico PCR analyses, it was concluded that the V1–3, V7–9 and V9 regions are more variable than V4 and V4–V5. V1–3 is too long for the Illumina MiSeq platform and cannot be used for metabarcoding. The number of sequences in international nucleotide sequence databases differs for the V4–5 and V7–9 regions.

3.5. 18S rRNA Gene Metabarcoding: V4 vs. V9

Several studies have been dedicated to comparing the efficiency of using the V4 and V9 regions for characterizing the diversity of eukaryotic communities.

Bradley et al. [54] examined the effect of PCR/sequencing bias of the V4 and V8–V9 regions on community structure and membership using seven microalgal mock communities consisting of 12 algal species across five major divisions of eukaryotic marine and freshwater microalgae. The authors found a critical shortcoming of the V4 primer set as used in the literature [43] and described the failed sequencing runs. The V4 region failed to reliably capture 2 of the 12 mock community members (the haptophytes *Prymnesium parvum* and *Isochrysis galbana*), whereas the V8–V9 hypervariable region more accurately represented the mean relative abundance and alpha and beta diversity. Bradley et al. [54] found that degeneracies on the 3' end of the current V4-specific primers impacted the read length and mean relative abundance. They modified the TAREukREV3 reverse primer and suggested the V4r primer without degeneracies on the 3' end for the subsequent sequencing (Table 1). Overall, the V4 and V8–V9 regions showed similar community representations, but their specific samples were markedly different. Therefore, the authors suggested that multiple primer sets might be advantageous for gaining a more complete understanding of community structures.

A comparative analysis of the V4 and V9 regions of 18S rDNA of the eukaryotic community of a pond [53] showed a remarkable discrepancy: the inventory of the major subdivision groups in the V9 region dataset did not correspond to that in the V4 region dataset. Eukaryotic OTUs for the V9 region were 20% more abundant than those for the V4 region at a 97% identity threshold. V9 also showed a larger diversity from the point of view of taxonomic coverage. The classes Karyorelictea, Prostomatea and Nasophorea in Ciliophora and the family Perkinsida ('Alveolata' group) were not detected using the V4 sequencing data, whereas they were detected using the V9 sequencing data. V4 missed Echinamoebida, Eumycetozoa and Euamoebida and green microalgae classes Chloropicophyceae, Pyramimonadophyceae and Mamiellophyceae. The authors noted "... the simultaneous application of two biomarkers may be suitable for understanding the molecular phylogenetic relationships".

In an investigation dedicated to a eukaryotic community in anaerobic wastewater treatment systems [48], the V4 and V9 regions also detected different taxonomic groups. The authors suggested that commonly used V4 and V9 primer pairs could produce a bias in eukaryotic community analyses. The number of sequences of the amplicon library for the V9 region was almost two times larger than the number of sequences of the V4 amplicon library (340,054 vs. 180,678). The V4 region-specific primer pair showed that the dominant group was fungi. However, the V9 region-specific primer pair showed a large portion of prokaryotic sequences (bacteria and archaea accounted for 52.2% and 35.6% of the total number of sequences, respectively.) Ultimately, the authors concluded that the V9 region-specific primer pair was not suitable for the analysis of eukaryotic communities in an upflow anaerobic sludge blanket reactor, because a large number of prokaryotes sequences was detected.

It is interesting to note that similar results were obtained in a comparison of these genetic markers in marine eukaryotic communities.

In the study of a eukaryotic community of marine anoxic waters, Stoeck et al. [43] showed similar results for these regions (V4 and V9) on the diversity profiles (higher rank taxon groups that were represented by a proportion $\geq 1\%$ of all unique tags in at least one of

the two sets of amplicons). However, the example of dinoflagellates showed that the V4 and V9 primer pairs detected very different taxonomic profiles at the genus and family levels. The authors connected these differences with the selectivity of primers that preferentially detect different dinoflagellate subgroups. On the other hand, sets of dinoflagellate taxa represented in GenBank by V4 and V9 SSU regions overlap only partially, which could artefactually lead to apparently different taxa being detected.

A comparison of the 18S rRNA V4 and V9 regions for coastal phytoplankton communities with a focus on Chlorophyta [122] showed that the V9 region provided 20% more OTUs built at 97% identity than V4. Interestingly, the expectations were the opposite: the authors assumed that V4 as the longer region would detect more OTUs. The authors noted that both markers work “...equally well to describe global communities at different taxonomic levels from the division to the genus and provided similar Chlorophyta distribution patterns”. The authors concluded that V9 was the better choice for Chlorophyta, as it was more discriminating than V4. In the same cases for prasinophytes clade VII, V9 OTUs allowed to discriminate all subclades defined to date, while, in V4, several clades collapsed together. However, there was also an opposite example: “The V9 region of some *Chlamydomonas* is very similar to that of prasinophytes clade VII A5”. The authors emphasized the importance of the existence of reference sequences in databases, the absence of which, for instance, prevented the assessment of Dolichomastigales (Chlorophyta and Mamiellophyceae) diversity using V9. Similar results were demonstrated on marine picoeukaryotes [123], amoebae [124] and zoonotic trichomonads [125].

Piredda et al. [126] reported similar patterns for the V4 and V9 markers. The authors compared data from metabarcoding and LM approaches using the example of marine planktonic protist assemblages. For Bacillariophyta, comparable taxonomic patterns were shown between the sequence and light microscopy data, whereas, for Dinophyta, there was an overrepresentation in the sequence dataset (authors explained it by the large genome size in this group and the relationships between genome size and rDNA copy numbers). The reassuring outcome of this study was the overall comparable results of taxonomic analyses obtained with V4 and V9 on the same samples. The diatom patterns across samples were rather similar between V4 and V9 at the levels of genera and species. Due to the failure in the identification of *Pseudo-nitzschia* in the V9 sequences, the authors associated this with the smaller reference dataset available for V9.

Overall, the taxonomic composition of the eukaryotic community in the V4 datasets differed from that in the V9 dataset. V9 provided more diversity on higher taxonomic levels (supergroup and phylum), whereas the V4 region missed some important eukaryotic groups (for example, the algae classes Chloropicophyceae, Pyramimonadophyceae and Mamiellophyceae). However, in the phylogenetic analyses of eukaryotes, the V4 region has a much better resolution than the V9 region [54]. It should also be taken into account that sets of taxa represented in databases by V4 and V9 SSU regions only partially overlap.

3.6. Internal Transcribed Spacer Ribosomal DNA (ITS) in Metabarcoding Researches

The ITS region is the accepted DNA barcode for fungi and a strong locus for delimiting or identifying species from different algal groups, such as Chlorophyta, Dinophyceae, Chrysophyceae, Xanthophyceae and Eustigmatophyceae [7,8]. Therefore, the usage of this region for metabarcoding has positive prospects, with a high probability of identifying nucleotide sequences at the species level. We found several studies that used ITS as a barcode region. As far as we are aware, there are no metabarcoding studies that compare ITS with other markers.

The V9-ITS1 region of the 18S was chosen for the large-scale research of freshwater protists from 217 freshwater lakes across Europe [68–70]. The studies were aimed at identifying the diversity dynamic of the protist communities relative to the geographic distance and mountain range structures [68], centers of endemism [70] and models of interactions between the protist community and bacteria [69]. In regard to algae, the diversity of the following groups was determined in these studies: Dinophyceae, Chrysophyceae, diatoms,

Cryptophyta and Viridiplantae (green algae). The same materials and the same methods were used for research on the phylogenetic and functional diversity of Chrysophyceae [71]. It was shown that Chrysophyceae are one of the most common groups in freshwater ecosystems (found in 213 out of 218 sample sites across Europe).

The ITS2 gene region is the best marker for DNA barcoding of Chlorophyta. This marker resolves major green algae lineages (some with high bootstrap support), has a high resolution for taxonomic assessment (enables the most species to be distinguished) and a high level of universality (i.e., in primers for PCR) [127] (and references in it). This region was successfully used in the first studies of the diversity of Viridiplantae (including green microalgae) in the Antarctic using the metabarcoding approach in soil and rock surfaces samples [72,128], sediments from lakes [74] and glacial ice [73]. The interpretation of sequences was carried out using the PLANITS2 database [107], and most of the taxa were identified to the species level.

3.7. Gene Markers for Diatoms

Diatoms are well-known ecological indicators of aquatic ecosystems and are widely used for routine monitoring. Indexes of the water quality in rivers and lakes have been developed on the basis of diatoms and are used in EU countries (the Water Framework Directive in Europe), the USA (the National Water Quality Assessment Program in the USA), Canada, Australia and New Zealand [105,129–131]. Therefore, adapting the metabarcoding method for use as a tool for ecological assessment is a relevant task of modern research.

The first works on metabarcoding of freshwater diatoms suggested the V4 18S region as a candidate for a barcode marker. Zimmermann et al. [37,132] demonstrated a high correlation of the results obtained by microscopy and by metabarcoding. The authors used effective specific primers M13F-D512 and M13R-D978rev (Table 1) that were tested on non-axenic unialgal cultures of 123 taxa of Bacillariophyta (including closely related species, the genus *Sellaphora* (incl. the *Sellaphora pupula* group)) and showed that the V4 18S rRNA fragment is variable enough for taxa identification. Still there is a balance between marker variability and primer universality. The latter is important for the reproducibility of laboratory protocols. Although 18S V4 does not allow sufficient resolution for cryptic species, the authors believe that this does not matter for ecological studies, because representatives of cryptic species groups usually have similar ecological preferences.

Visco et al. [32] showed a strong similarity between the DI-CH (the Swiss Diatom Index) values inferred from microscopic and V4 18S NGS analyses of diatom communities. However, the authors noted that the interspecies variability of this barcode might change between different genera, and its effectiveness would depend on the taxonomic composition of the diatom community. The V4 resolution did not allow to unambiguously assign *Navicula* species, but it was sufficient to distinguish most of the species of *Nitzschia* and *Gomphonema*.

The *rbcL* gene marker has a wider application for studying diatom communities, and thanks to the establishment of a quality reference database Diat.barcode/R-syst:diatom [105], it can already be considered the standard for diatom metabarcoding.

A region 263 bp long (or 312 bp, including primers) and a set of primers first suggested by Vasselon et al. [16,94] (Table 1, *rbcL* primers Set 1) is used in the overwhelming majority of studies. This marker choice is based on the works of Kermarrec et al. [115,133], who compared the nuclear gene 18S and the plastid gene *rbcL* and showed that the resolution of the *rbcL* gene provides detection at the species level, while 18S is efficient at the genus level.

The resolution of the *rbcL* 312 bp marker on the level of intraspecific and cryptic diversity was successfully demonstrated by Pérez-Burillo et al. [80]. Benthic diatom samples (n = 610) were studied with a special focus on several ecologically important diatom species that are also key for the Water Framework Directive monitoring of European rivers: *Fistulifera saprophila*, *Achnantheidium minutissimum*, *Nitzschia inconspicua* and *Nitzschia soratensis*. As a result, it was shown that intraspecific and cryptic diversity can be assessed and understood through the application of DNA metabarcoding. For example, the

genetic variants within *Achnanthyidium minutissimum* and *Fistulifera saprophila* were detected. There was no correlation between the phylogenetic lineages and ecological preferences, which emphasized the "...necessity to work at the lowest "taxonomic" level possible".

In a study of diatom endemism in high-altitude alpine lakes, Rimet et al. [84] showed the resolution of *rbcL* 312 bp at both the species and subspecies level. The analysis of the acquired data allowed the authors to draw important conclusions: high diversity was detected at the subspecies level, and the proportion of shared taxa equaled only 1.5% (in contrast, at the species level, the proportion of shared taxa equaled 15%); therefore, the level of endemism was very high, as the more sites were occupied by a species, the higher its intraspecific diversity. Finally, application of automated molecular species delimitation methods to *Achnanthyidium minutissimum* revealed a hidden diversity of five and seven putative species, which did not appear to be monophyletic on the tree and had no geographic structuring.

A longer *rbcL* region (331 bp) was suggested as a result of large-scale research (500 benthic samples from 250 sites in England) with the aim of adopting a metabarcoding approach for ecological status assessment using diatoms [17]. The choice of region was based on an analysis of 390 sequences from a database. Eleven conservative regions of the *rbcL* gene with >96% identity were identified. These regions were used for developing primers. Variable regions were also analyzed, and four of these showed good potential for species delimitation. Consequently, primers were developed for these latter regions, and tests were conducted in order to determine the most effective region. As a result, based on its taxonomic coverage, amplicon length, primer conservation and robust performance, amplicon K (331 bp) with the primer pair *rbcL*-646F/*rbcL*-998R (Table 1) was selected for use in all downstream Illumina analyses for benthic diatoms.

An evaluation of two overlapping *rbcL* markers of 263 (312 bp, including primers) and 331 bp (common region 263 bp) was done recently by Pérez-Burillo et al. [86]. A large dataset was used for the study (1703 benthic diatom samples), and the results were thoroughly analyzed, considering (i) the effect of marker choice on taxonomic assignment, (ii) in-depth analyses on species discrepancies, (iii) comparison of the nucleotide and amino-acid variability (Shannon entropy) and (iv) effects of the marker choice on ecological status assessment. It was shown that the 331 bp marker demonstrates a higher resolution of species and infraspecific variants (some ASVs were unambiguously classified at the species level based only on the 331 bp marker: *Surirella brebissonii*, *Halamphora montana* and *H. banzuensis*). The authors noted, however, that false negatives were possible (some ASVs were classified into the same species by both markers, but the identifications could be rejected for one or the other marker because the bootstrap support values were very low (≥ 85); some ASVs could not be identified to the species, because they were identical to the reference sequences for more than one taxon). However, the biotic index (IPS) scores derived from both markers were very highly correlated and the choice of the 263 bp or 331 bp the *rbcL* marker had no important effects on the ecological status assessments. But the higher resolution of the longer marker may be preferable in ecological or biogeographical studies.

3.8. Specific Primers Targeted to *rbcL* Region Detected a High Diversity of Eustigmatophyceae

A high diversity of Eustigmatophyceae was found in environmental DNA samples with the help of new specific primers targeted at the *rbcL* region [97] (Table 1). The authors compared their results to previous studies concerning Eustigmatophyceae and concluded that diversity of this group was underestimated. The designed primers allowed to detect 184 ASV haplotypes that were either Eustigmatophyceae (179) or possibly Eustigmatophyceae (15), while, in previous works, representatives of this group were reported only as rare or single finds. The sensitivity of eustigmatophyte-directed *rbcL* primers was compared higher to universal eukaryotic 18S primers. The authors suggested that the employed techniques can be used for future studies of the population structure, ecology, distribution and diversity of this class.

3.9. Comparison of *rbcL* and 18S Markers for Freshwater Diatoms Biomonitoring

Inconclusive results were obtained in a study using the *rbcL* and V4 18S rRNA markers [34] for benthic diatoms biomonitoring in freshwater habitats of Northern Europe. The classes of ecological condition differed significantly depending on the used method: only 48% of samples with the 18S marker and 37.5% of samples with the *rbcL* marker had the same ecological status as with the morphological analysis. The assessment of the ecological conditions gave different results using different markers. The authors connected this with the differences in the taxonomic scope of the corresponding reference databases and primer specificity. For example, *Tabellaria flocculosa* was always detected with the *rbcL* marker and never with 18S (even though they are represented in the reference database). Barcodes for green algae were present only in the 18S dataset and were completely absent from the *rbcL* dataset. According to the authors, the amplification of green algae in some samples while using the 18S marker led to a low percentage of detected diatoms in the sample. In general, however, the *rbcL* marker generated species lists were more similar to the ones generated by the morphological approach. In the end, the authors found it difficult to recommend one marker over the other.

Similar research was conducted by Apothélos-Perret-Gentil et al. [33]. They compared the same markers (fragment of the *rbcL* gene and the V4 region of the 18S rRNA gene) for the inference of the molecular diatom index. However, it was shown that, generally, a slightly better correlation with the morphological reference was observed with the *rbcL* marker due to the fact that it was more taxonomically resolutive, and the distinction of the diatom and other species was more accurate. As valuable advantages of the *rbcL* gene, the authors noted the primer specificity and the existence of the comprehensive curated Diat. barcode reference database [105]. The generated species lists based on *rbcL* were more exhaustive than the ones generated by the 18S marker. In the authors' opinion, *rbcL* so far represents the ideal candidate for the implementation of metabarcoding methods for routine river monitoring.

3.10. A 23S rDNA Plastid Marker for Simultaneous Detection of Eukaryotic Algae and Cyanobacteria

The universal plastid amplicon (UPA) is the variable Domain V of the 23S plastid rRNA gene ~330 bp in length. This region was proposed by Sherwood and Presting [99] as a marker for plastid-containing organisms, i.e., all lineages of eukaryotic algae and Cyanobacteria. In this research, a single pair of universal primers was designed, and it was indicated that these exact priming sequences are present only in cyanobacteria and plastids. However, comparisons with other markers showed the insufficient effectiveness of UPA. For example, Hamsher et al. [113] assessed four gene markers (COI, *rbcL*, LSU D1/D2 and UPA) for barcoding diatoms and concluded that the amplification of UPA was excellent, but this region was considerably more conserved among diatoms and distinguished only 20% of species. Hall et al. [116] reported UPA as the least variable locus in freshwater green algae. In charophytes (e.g., *Chara*, *Desmidium* and *Micrasterias*), there were difficulties with sequencing, and in the *Nitella* strains, the universal primers most often amplified a non-target region.

The low efficiency of this marker compared to 18S genes was confirmed by Cahoon et al. [10] based on a metabarcoding analysis of freshwater planktonic protists. The 18S barcode identified a much larger number of photoautotrophic genera OTUs (198) than 23S (75), from which 22 genera (9.5%) were uniquely identified by 23S and 145 (65.9%) by 18S. To our knowledge, this marker is used fairly rarely for metabarcoding studies of algae (Figure 1 and Table 1). Apart from the work of Cahoon et al. [10], 23S metabarcoding was conducted for a study of phytoplankton community structure and diversity of the aquaculture system for *Litopenaeus vannamei* [99], an examination of phytoplankton in the unique hypersaline system of Great Salt Lake's Gilbert Bay (Salt Lake City, UT, USA) [100] and a multimarker analysis of an algal biofilm community [134]. Bonfantine et al. [103] reported failing to detect diatoms in 23S reads from stream biofilm samples. The low

differentiating power of this marker should also be taken into account [117,120]; in the aforementioned studies, taxonomic attribution was done only to the genus level. 23S rRNA is also not included in the list of strong loci in use for delimiting or identifying species of algae [7,8].

3.11. The 16S rRNA Gene as a Marker for Simultaneous Detection of Prokaryotes and Eukaryotes

The 16S rRNA gene was first proposed as a metabarcoding marker by Eiler et al. in 2013 [101] on the basis of it being universally present in prokaryotes (including cyanobacteria), as well as in chloroplasts of eukaryotes. This enabled the simultaneous detection of prokaryotic and eukaryotic phytoplankton taxa. The authors analyzed the phytoplankton diversity from 49 lakes, including three seasonal surveys, and assessed the data using NGS and microscopy. The NGS approach detected 1.5–2 times more OTUs than there were taxa found by the microscopy approach. A more detailed comparison of taxonomic groups revealed that Heterokonta, Euglenophyta, Cryptophyta and Dinophyta were overrepresented in the microscopic biovolume dataset compared to the NGS data, whereas Cyanobacteria were proportionally overrepresented in the NGS dataset compared to microscopic biovolume data. The authors noted that Dinophyta, a major phylum in microscopic data, was poorly detected by NGS in some lakes. Discrepancies also included Euglenophyta and Heterokonta that were scarce in the NGS but were frequently detected by microscopy. The NGS approach detected a deep-branching taxonomically unclassified cluster that could not be linked to any group identified by microscopy.

Later, Huo et al. [57] opined that the chloroplast 16S rDNA gene might not be an appropriate choice for detecting eukaryotic phytoplankton diversity because of a bias toward bacteria. The common primers targeting this gene cover a wide spectrum of taxa, thus reducing the sequencing efforts aimed at phytoplankton diversity. The second problem is the endosymbiotic origin of chloroplasts in eukaryotic phytoplankton and endosymbionts retained in host cells permanently or temporarily. The authors noted that diatoms, cryptophytes and haptophytes have been reported to serve as endosymbiotic chloroplasts in diverse dinoflagellate species. “Therefore, the chloroplast 16S rDNA gene might not truly reflect host phytoplankton diversity.”

Recently, Bonfantine et al. [103] explored the potential of a standard V4 515F-806RB primer pair in recovering diatom plastid 16SrRNA sequences. PhytoREF was used to classify the 16S reads from 72 freshwater biofilm samples. Based on the Clustal nucleotide alignment, the authors confirmed the differences between eukaryotic chloroplast and prokaryotic sequences. “The Ochrophyta, and other eukaryote reads, showed high sequence conservation with no 3' mismatches in the last 5 bases of both forward and reverse 16S v4-515F and V4-806R primers. Two mismatches to the *E. coli* 16S rRNA (GT vs. TA) were observed across all aligned non-*E. coli* 16S RNA sequences 15 bases upstream of the V4-806 primer-binding site.” More than 90% of the diatom reads in each stream biofilm sample were identified. The authors found significant beta-diversity in diatom assemblages and discrimination among river segments. In an example of the three Australian environmental 16S rRNA datasets selected from NCBI-SRA, it was shown that most of the diatom OTUs (67 out of 71) were detected in other Australian ecosystems. As a result, the authors concluded that diatom plastid 16S rRNA genes are readily amplified with the standard primer sets. “Therefore, the volume of existing 16S rRNA amplicon datasets initially generated for microbial community profiling can also be used to detect, characterize, and map diatom distribution to inform phylogeny and ecological health assessments, and can be extended into a range of ecological and industrial applications.”

Overall, the 16S rRNA marker is rarely used for the simultaneous analysis of prokaryotic and eukaryotic communities (Figure 1). Although the cost of preparing a library and the volume of data increase, researchers prefer to separate one from the other and use different regions of 16S and 18S rRNA accordingly for studying prokaryotic and eukaryotic communities [30,40,50,51,135]. Nevertheless, as shown by Eiler et al. [101] and Bonfantine et al. [103], environmental 16S rRNA datasets can yield useful information on eukaryotes,

but the sensitivity and the level of taxonomic attribution in this case would be much lower than while using eukaryotic markers. It should also be considered that, according to Eiler et al. [101], Heterokonta, Dinophyta, Euglenophyta and Heterokonta are often poorly detected by NGS based on the 16S rRNA.

3.12. Comparing Approaches: Metabarcoding vs. Morphological Identification (Congruency between Methods)

Comparing the results acquired by using LM and NGS allows to reveal discrepancies between these methods and the causes of these discrepancies to determine the efficiency of the amplification of the chosen genetic markers and to identify problems in bioinformatical processing and taxonomic attribution. Every single work that compared the morphological and molecular approaches for studying the diversity of algal communities indicated a significant difference in the resulting taxonomic lists. The number of taxa detected by both approaches falls between 7.4 and 25.7% (Figure 3).

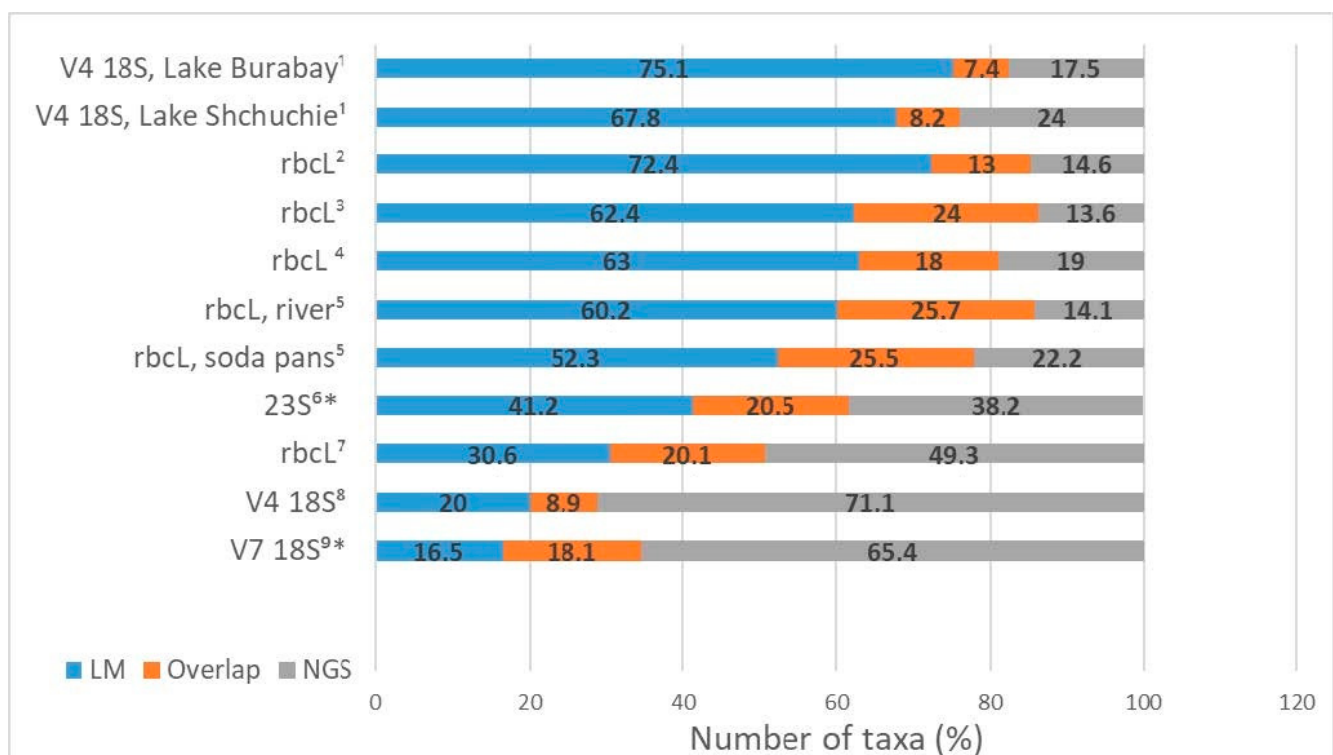


Figure 3. Shared taxa detected by the microscopy approach (LM), next-generation sequencing (NGS) and both (Overlap). * Publications with identification at the genus level. References: 1–[40], 2–[16], 3–[35], 4–[94], 5–[95], 6–[99], 7–[83], 8–[42], 9–[57].

Some studies have shown that diversity detected by NGS is much higher than that found in LM. In a research of diatom diversity, Zimmermann et al. [132] reported that about 2.5 times more taxa were found by the NGS approach (263 taxa vs. 102 taxa in LM). In an example of studying benthic diatoms, Baillet et al. [34] showed that the metabarcoding method using the 18S marker revealed 27% more taxa than the morphological method and 38% more taxa using the *rbcL* marker. As a result of a study on epiphytic diatoms, Borrego-Ramos et al. [83] showed that metabarcoding detected more taxa than LM (49.3% vs. 30.6%), and only 20.1% of the taxa were concordant when comparing both methodologies. In an investigation of phytoplankton, Huo et al. [57] reported that metabarcoding using the 18S V7 gene marker detected 3.5 times more OTUs than the number of morphospecies revealed by morphological identification. The same results were shown in the studies of phytoplankton conducted by Groendahl et al. [42] based on the 18S V4 gene marker: the

metabarcoding method detected 71.1% of the total taxa number, whereas LM identification found only 20% (Figure 3).

In contrast, other investigations have shown that the morphological approach allows to identify a greater diversity [16,35,40,90,95,99] (Figure 3). Using the example of diatoms, it has been shown that most of the species identified by LM are not represented in a database, including the special curated diatoms database Diat.barcode. Visco et al. [32] revealed that the GenBank database only covers 46% of the morphospecies found in microscopic analyses. Vasselon et al. [16] and Vasselon et al. [90] reported that 68% and 82% of morphological species are not represented in the database. Duelba et al. [91] revealed that 60% and 32% of taxa detected by LM in riverine and soda pan samples, respectively, are not recorded in the database. Bailet et al. [34] pointed out that only 15.4% of all Fennoscandian taxa are represented in the 18S database and 17.8% in the *rbcL* database. A comparative analysis of freshwater phytoplankton communities in two lakes conducted by Malashenkov et al. [40] showed that the NGS of 16S and 18S rRNA amplicons adequately identified phytoplanktonic taxa only at the genus level, while the species composition obtained by microscopic examination was significantly larger (67.8% and 75.1% vs. 24% and 17.5% by NGS, respectively) (Figure 3).

The challenges of taxonomic reference databases in metabarcoding analysis were recently summarized in a review by Keck et al. [136]. The authors discussed in detail the following problems: (i) mislabeling, (ii) sequencing errors, (iii) sequence conflict, (iv) taxonomic conflict, (v) low taxonomic resolution, (vi) missing taxa and (vii) missing intraspecific variants.

Here, we briefly summarize and supplement the list of reasons that explain discrepancies in the results obtained by using microscopic and metabarcoding methods on algae. These should all be taken into account while interpreting the results.

The challenges in metabarcoding analysis:

1. Gap in the reference database [16,27,32,34,38,46,62,82,90,99], etc.
2. The natural intraspecific and intragenomic variabilities of the barcoding marker (single taxon has multiple genotypes at the barcoding region, and members of that taxon might cluster into different Molecular Operational Taxonomic Units (MOTUs)) [35].
3. Cryptic diversity—a single morphological species can represent different genetic groups (e.g., diatoms *Sellaphora pupula*, *Pinnularia borealis*, *Hantzschia amphioxys* and *Nitzschia inconspicua* and species of *Stichococcus*, *Coccomyxa*, *Chlorokybus*, *Cryptomonas*, etc.) [78,137–142].
4. MOTU richness can be artificially inflated through technical errors at different steps of sample processing during amplification and sequencing [35].
5. The MOTU delimitation approach influences the richness estimation and interpretation [35] (and references in it) (assessment of the bioinformatics pipelines provided in [14,15]).
6. Complete absence of amplification on the whole due to a mismatch of the primer set used. For example, Salmaso et al. [27] did not find any species belonging to the Euglenales in the HTS results (with universal eukaryotic primers (TAREuk454FWD1 and TAREukREV3) for V4 18S), although they were present in LM. Hanžek et al. [66] reported that the taxa that contributed most to the biomass (*Actinotaenium/Mesotaenium* sp. and the species *Cosmarium tenue*, *Pantocsekiella comensis*, *Sphaerocystis schroeteri* and *Synedropsis roundii*) were not identified by eDNA metabarcoding (V9 18S region was amplified using the universal primer pair 1391F and EukB). Proeschöld and Darienko [140] noted that, although *Stichococcus*-like organisms are widely distributed in almost all habitats, they are not recorded in environmental studies based on HTS approaches, because the V4 or V9 regions of the SSU contain introns that obstruct amplification. Groendahl et al. [42] reported that *Monorhaphidium* sp., *Selenastrum* sp. and *Trachelomonas* sp. detected using the morphology-based approach were not identified by the metabarcoding approach, despite the fact that all three genera are included in the reference database.

7. Uncertainties and lack of sensitivity of reference databases for the selected DNA markers [27].

The challenges related to morphological identification:

1. Diatom extracellular skeletons are counted in LM even if they come from dead cells. The valves of dead cells can be transported from locations other than the target assemblage. Metabarcoding will not detect these dead cells [35,78].
2. The proportion of live diatoms found in environmental samples varies greatly, ranging from 2 to 98% [35].
3. Small-celled species and pico-sized cells are often overlooked or underestimated by the morphological approach. For example, the valves of *Fistulifera saprophila* tend to dissolve during sample processing, which can explain why this species is often missed during morphological identification [78,80,82,99].
4. LM misidentification (false LM positives) [27,78,99].
5. Differences in the detection limits of the two methods: morphological and molecular approaches do not give the same insight into communities of algae and, therefore, do not have the same detection capacity for species [27,92,99].
6. The different sample volumes settled for microscopy and metabarcoding [143].
7. Underlying units used for microscopy (individual cells) and those used for metabarcoding (ASV sequences) are quite different, making direct comparisons imperfect [24,143].
8. A short barcode gene fragment may have limited the taxonomic resolution [143]. For example, the resolution of the V4 18S region does not allow to unambiguously identify some species of *Navicula* [32]. For the V7–9 18S marker, a lack of inter-genus taxonomic resolution was found (the MOTUs matched multiple genera, e.g., *Alexandrium pseudogonyaulax* and *A. hiranoi*, *Chaetoceros neograticus* and *C. curvisetus* and *Thalassiosira eccentrica* and *T. antarctica*) [144]. In some *Chlamydomonas*, the V9 region is very similar to that of prasinophytes clade VII A5 [122].

Thus, each method is imperfect and has its limitations. For better understanding and interpretation of the metabarcoding results, studies that use both methods are still relevant. The filling of databases with identified nucleotide sequences with metadata will, in the future, greatly improve the quality of taxonomic attribution and metabarcoding data interpretation. Nevertheless, as pointed out by Bailet et al. [14], “...the longer-term goal should be to break free from the preconceptions we have brought with us from careers based around light microscopy and to recognize HTS data as distinct”.

4. Conclusions

Metabarcoding has already been accepted as an alternative (faster and more economical) method to the traditional microscopy method for the ecological assessment and monitoring of freshwater bodies of water, rivers and seas based on microalgae. Protocols, technical guidelines or standards for eDNA monitoring are developed and/or approved in many countries [145–149] (Table S2). The results of our review show that, currently, there is no one perfect marker for identifying microalgae across the whole diversity. Among the most popular genetic barcodes for freshwater metabarcoding, we can highlight the nuclear regions V4 and V9 18S rRNA (which allow to determine the composition of auto- and heterotrophic eukaryotes) and the region of the plastid gene *rbcL* (for diatoms). The regions ITS1 and ITS2 might be underestimated, but they show good potential for usage as a microalgae barcode; they can be easily amplified with the standard primers and are variable enough to identify sequences to the species level.

The choice of marker is determined by the focus of the study, for example, a certain group or groups of algae (a specific marker and/or primers can be used, like a marker for diatoms or Eustigmatophyceae); a screening of eukaryote diversity (universal markers V4 and V9 18S are suitable) or an assessment of interactions between different groups of organisms (a set of markers for various groups of prokaryotes or eukaryotes, invertebrates and vertebrates should be used). Some studies have used a multimarker approach. For example, Wolf and Vis [134] used four markers (V9 18S, *rbcL*, 23S and V4 16S rDNA) for identification of an

algal biofilm community. To investigate the dynamic evolution of multitrophic communities (bacterial and eukaryotic) under ecohydrological changes, Liang et al. [52] used three markers: V3–V4 16S rRNA (for bacteria), the V4 region of 18S rRNA and COI (for eukaryotes) and universal primers. Robinson et al. [89] used multimarker metabarcoding to study diatoms and macroinvertebrate indicators (*rbcL* and COI markers, respectively). In a large-scale study of benthic macroinvertebrates and diatoms in rivers, Seymour et al. [46] compared the 18S universal molecular marker and molecular markers that target traditional biomonitoring groups (*rbcL*, 12S and COI). For sequence identification, the most comprehensive databases for each marker were used: NCBI for COI and 12S, Silva for 18S and Diat.barcode for *rbcL*. Such investigations are tied to processing a massive amount of data for different groups of organisms (thousands of sequences are involved). Therefore, it is complicated to verify all taxonomic attributions obtained during automatic processing. Looking through the Supplemental Materials for the latest study [46], we noticed several misidentifications. For example, using the *rbcL* marker (eDNA method), several representatives of different classes were assigned to the class Bacillariophyceae: *Dictyosphaerium*, *Lobosphaera* (Trebouxiophyceae), *Kirchneriella*, *Oedogonium*, *Pandorina*, *Volvox*, *Pediastrum* (Chlorophyceae), *Tribonema* (Xanthophyceae), *Lagynion* (Chrysophyceae), etc. This shows that there are faults in the databases and confirms once more the necessity of verification for taxonomic lists obtained by NGS, since these mistakes can lead to incorrect descriptions of communities and inaccurate conclusions.

Below, we briefly summarize the main advantages (+) and disadvantages (–) of markers that are used for freshwater microalgae metabarcoding.

***rbcL*.**

“+” widely used for diatom metabarcoding, distinguishes between taxa at the species and intraspecies levels; a high-quality curated reference database for taxonomic attribution Diat.barcode [105].

“–” extremely heterogeneous in green algae; does not have a set of universal primers [127]; only diatoms are identified well.

Notes: the majority of studies use the region with the length of 263 bp (312 bp including primers) and a complicated set of primers (three forward primers and two reverse primers) suggested by Vasselon et al. [16,90] (Table 1, *rbcL* primers Set 1). However, recently, it has been shown that a longer region of 331 bp (common region 263 bp, proposed by Kelly et al. [17,18], primers set *rbcL* 646F–*rbcL* 998R) has a higher resolution for species and intraspecific variants [86].

V4 18S.

“+” widely used for metabarcoding of marine and freshwater eukaryotes; successfully amplified by a universal primer set. The V4 region (named pre-barcode) was designated as the starting point for the identification of protists in the International Barcode of Life Consortium Project (iBOL, <http://www.ibol.org/> (accessed on 22 May 2023) and the Protist Working Group (ProWG) [150]. Provides an understanding of molecular phylogenetic relationships.

“–” compared with V9 18S, misses haptophytes [54], many groups of heterotrophs and green algae from the classes Chloropicophyceae, Pyramimonadophyceae and Mamielophyceae [53]. The V4 region is less variable compared to the V9 region. Often does not differentiate species.

V9 18S.

“+” widely used for marine and freshwater eukaryotes metabarcoding; successfully amplified by a universal primer set; was chosen to amplify eukaryotes in the global project “The Earth Microbiome Project” (EMP; <http://www.earthmicrobiome.org> (accessed on 22 May 2023)); V9 is more variable than V4; provides more OTUs and diversity on higher level taxa (supergroup and phylum).

“–” a short region (96 bp–134 bp [36]; sometimes does not differentiate species.

Notes: the V4 and V9 regions detect different taxonomic profiles at the genus, family and macro-taxa levels. Both markers are recommended for a more complete understanding of community structures.

ITS (V9-ITS1 or ITS2 regions are used for metabarcoding).

“+” is a strong locus for some algae (Chlorophyta, Dinophyceae, Eustigmatophyceae and Xanthophyceae) [7]; sufficiently variable; allows to differentiate species; amplified by a universal primer set. A specialized curated reference dataset “PLANiTS” [107] exists for interpreting data, including ITS1, ITS2 and entire ITS sequences of Viridiplantae.

“−” for diatoms, this barcode has a great variability in the length of region and a problem in intraindividual variation [110,113].

23S (UPA).

“+” algal-specific markers focused on plastid-containing eukaryotic algae and Cyanobacteria; sufficiently amplified by a universal primer set.

“−” a conserved region; low resolution; identifies taxa only to the genus level or higher; is not a strong locus for algae.

16S (V3–V4 and V4 regions is used for metabarcoding).

“+” focused on the chloroplasts of eukaryotes and prokaryotes (Cyanobacteria); sufficiently amplified by a universal primer set; can be used to simultaneously detect prokaryotes and eukaryotic algae.

“−” biased towards bacteria in the community; might not accurately reflect the phytoplankton diversity due to the endosymbiotic origin of chloroplasts; is not a strong locus for algae.

Notes: is used very rarely for the simultaneous complex analysis of prokaryotes and eukaryotic algae.

In general, it should be taken into account that every marker will demonstrate a different image of the community that depends on successful amplification of the chosen region. The identification of the taxonomic composition and the level of taxonomic attribution depends on the region variability and the quality of the reference databases.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/biology12071038/s1>: Table S1: Reference list used in the analyses. Table S2: Barcode markers and primers used in the guidelines and standards for metabarcoding of various groups of algae and cyanobacteria.

Author Contributions: Conceptualization, E.K. and M.K.; methodology, E.K.; validation, M.K.; writing—original draft preparation, E.K.; writing—review and editing, M.K. and N.T.; visualization, E.K. and supervision, M.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was conducted with financial support by the Russian Science Foundation (project number 22-24-00965).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Adl, S.M.; Simpson, A.G.B.; Farmer, M.A.; Andersen, R.A.; Anderson, O.R.; Barta, J.R.; Bowser, S.S.; Brugerolle, G.; Fensome, R.A.; Fredericq, S.; et al. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J. Eukaryot. Microbiol.* **2005**, *52*, 399–451. [\[CrossRef\]](#)
2. Adl, S.M.; Simpson, A.G.B.; Lane, C.E.; Lukeš, J.; Bass, D.; Bowser, S.S.; Brown, M.W.; Burki, F.; Dunthorn, M.; Hampl, V.; et al. The revised classification of eukaryotes. *J. Eukaryot. Microbiol.* **2012**, *59*, 429–514. [\[CrossRef\]](#)
3. Kim, K.M.; Park, J.H.; Bhattacharya, D.; Yoon, H.S. Applications of next-generation sequencing to unravelling the evolutionary history of algae. *Int. J. Syst. Evol. Microbiol.* **2014**, *64*, 333–345. [\[CrossRef\]](#)

4. Burki, F. The convoluted evolution of eukaryotes with complex plastids. In *Secondary Endosymbioses*; Hirakawa, Y., Ed.; Elsevier: Tsukuba, Japan, 2017; Volume 84, pp. 1–30. [\[CrossRef\]](#)
5. Burki, F.; Roger, A.J.; Brown, M.W.; Simpson, A.G. The new tree of eukaryotes. *Trends Ecol. Evol.* **2020**, *35*, 43–55. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Gololobova, M.A.; Belyakova, G.A. Position of Algae on the Tree of Life. *Dokl. Biol. Sci.* **2022**, *507*, 312–326. [\[CrossRef\]](#)
7. Fawley, M.W.; Fawley, K.P. Identification of Eukaryotic Microalgal Strains. *J. Appl. Phycol.* **2020**, *32*, 2699–2709. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Leliaert, F.; Verbruggen, H.; Vanormelingen, P.; Steen, F.; López-Bautista, J.M.; Zuccarello, G.C.; De Clerck, O. DNA-based species delimitation in algae. *Eur. J. Phycol.* **2014**, *49*, 179–196. [\[CrossRef\]](#)
9. Leray, M.; Knowlton, N. Random sampling causes the low reproducibility of rare eukaryotic OTUs in Illumina COI metabarcoding. *PeerJ* **2017**, *3*, e3006. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Cahoon, A.B.; Huffman, A.G.; Krager, M.M.; Crowell, R.M. A meta-barcoding census of freshwater planktonic protists in Appalachia—Natural Tunnel State Park, Virginia, USA. *Metabarcoding Metagenom.* **2018**, *2*, e26939. [\[CrossRef\]](#)
11. Marcelino, V.; Verbruggen, H. Multi-marker metabarcoding of coral skeletons reveals a rich microbiome and diverse evolutionary origins of endolithic algae. *Sci. Rep.* **2016**, *6*, 31508. [\[CrossRef\]](#)
12. van der Loos, L.M.; Nijland, R. Biases in bulk: DNA metabarcoding of marine communities and the methodology involved. *Mol. Ecol.* **2021**, *30*, 3270–3288. [\[CrossRef\]](#)
13. Hadziavdic, K.; Lekang, K.; Lanzen, A.; Jonassen, I.; Thompson, E.M.; Troedsson, C. Characterization of the 18S rRNA Gene for Designing Universal Eukaryote Specific Primers. *PLoS ONE* **2014**, *9*, e87624. [\[CrossRef\]](#)
14. Bailet, B.; Apothéoz-Perret-Gentil, L.; Baričević, A.; Chonova, T.; Franc, A.; Frigerio, J.; Kelly, M.; Mora, D.; Pfannkuchen, M.; Proft, S.; et al. Diatom DNA metabarcoding for ecological assessment: Comparison among bioinformatics pipelines used in six European countries reveals the need for standardization. *Sci. Total Environ.* **2020**, *745*, 140948. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Czech, L.; Stamatakis, A.; Dunthorn, M.; Barbera, P. Metagenomic Analysis Using Phylogenetic Placement—A Review of the First Decade. *Front. Bioinform.* **2022**, *26*, 871393. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Vasselon, V.; Domaizon, I.; Rimet, F.; Kahlert, M.; Bouchez, A. Application of high-throughput sequencing (HTS) metabarcoding to diatom biomonitoring: Do DNA extraction methods matter? *Freshw. Sci.* **2017**, *36*, 162–177. [\[CrossRef\]](#)
17. Kelly, M.; Boonham, N.; Juggins, S.; Kille, P.; Mann, D.; Pass, D.; Sapp, M.; Sato, S.; Glover, R. A DNA based diatom metabarcoding approach for classification of rivers. In *Science Report SC140024/R*; Environment Agency: Bristol, UK, 2018; p. 157.
18. Kelly, M.; Boonham, N.; Juggins, S.; Mann, D.; Glover, R. Further development of a DNA based metabarcoding approach to assess diatom communities in rivers. Chief Scientist's Group report. In *Version: SC160014/R*; Environment Agency: Bristol, UK, 2020; p. 133.
19. Ruppert, K.M.; Kline, R.J.; Rahman, M.S. Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. *Glob. Ecol. Conserv.* **2019**, *17*, e00547. [\[CrossRef\]](#)
20. Bruce, K.; Blackman, R.; Bourlat, S.J.; Hellström, A.M.; Bakker, J.; Bista, I.; Bohmann, K.; Bouchez, A.; Brys, R.; Clark, K.; et al. *A Practical Guide to DNA-Based Methods for Biodiversity Assessment*; Pensoft Advanced Books: Sofia, Bulgaria, 2021. [\[CrossRef\]](#)
21. Pawlowski, J.; Bruce, K.; Panksep, K.; Aguirre, F.; Amalfitano, S.; Apothéoz-Perret-Gentil, L.; Baussant, T.; Bouchez, A.; Carugati, L.; Cermakova, K.; et al. Environmental DNA metabarcoding for benthic monitoring: A review of sediment sampling and DNA extraction methods. *Sci. Total Environ.* **2022**, *818*, 151783. [\[CrossRef\]](#)
22. Baricevic, A.; Chardon, C.; Kahlert, M.; Karjalainen, S.M.; Pfannkuchen, D.M.; Pfannkuchen, M.; Rimet, F.; Tankovic, M.S.; Trobajo, R.; Vasselon, V.; et al. Recommendations for the preservation of environmental samples in diatom metabarcoding studies. *Metabarcoding Metagenom.* **2022**, *6*, e85844. [\[CrossRef\]](#)
23. Kelly, M.G.; Juggins, S.; Mann, D.G.; Sato, S.; Glover, R.; Boonham, N.; Sapp, M.; Lewis, E.; Hany, U.; Kille, P.; et al. Development of a novel metric for evaluating diatom assemblages in rivers using DNA metabarcoding. *Ecol. Indic.* **2020**, *118*, 106725. [\[CrossRef\]](#)
24. Pawlowski, J.; Kelly-Quinn, M.; Altermatt, F.; Apothéoz-Perret-Gentil, L.; Beja, P.; Boggero, A.; Borja, A.; Bouchez, A.; Cordier, T.; Domaizon, I.; et al. The future of biotic indices in the ecogenomic era: Integrating (e)DNA metabarcoding in biological assessment of aquatic ecosystems. *Sci. Total Environ.* **2018**, *637–638*, 1295–1310. [\[CrossRef\]](#)
25. Sagova-Mareckova, M.; Boenigk, J.; Bouchez, A.; Cermakova, K.; Chonova, T.; Cordier, T.; Eisendle, U.; Elersek, T.; Fazi, S.; Fleituch, T.; et al. Expanding ecological assessment by integrating microorganisms into routine freshwater biomonitoring. *Water Res.* **2021**, *191*, 116767. [\[CrossRef\]](#)
26. Pawlowski, J.; Apothéoz-Perret-Gentil, L.; Altermatt, F. Environmental DNA: What's behind the term? Clarifying the terminology and recommendations for its future use in biomonitoring. *Mol. Ecol.* **2020**, *29*, 4258–4264. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Salmaso, N.; Vasselon, V.; Rimet, F.; Vautier, M.; Elersek, T.; Boscaini, A.; Donati, C.; Moretto, M.; Pindo, M.; Riccioni, G.; et al. DNA sequence and taxonomic gap analyses to quantify the coverage of aquatic cyanobacteria and eukaryotic microalgae in reference databases: Results of a survey in the Alpine region. *Sci. Total Environ.* **2022**, *834*, 155175. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Feist, S.M.; Lance, R.F. Genetic detection of freshwater harmful algal blooms: A review focused on the use of environmental DNA (eDNA) in *Microcystis aeruginosa* and *Prymnesium parvum*. *Harmful Algae* **2021**, *110*, 102124. [\[CrossRef\]](#)
29. Bartolo, A.G.; Zammit, G.P.; Akira, F.; Küpper, F.C. The current state of DNA barcoding of macroalgae in the Mediterranean Sea: Presently lacking but urgently required. *Bot. Mar.* **2020**, *63*, 253–272. [\[CrossRef\]](#)

30. Mikhailov, I.S.; Zakharova, Y.R.; Bukin, Y.S.; Galachyants, Y.; Petrova, D.; Sakirko, M.; Likhoshway, Y. Co-occurrence Networks Among Bacteria and Microbial Eukaryotes of Lake Baikal During a Spring Phytoplankton Bloom. *Microb. Ecol.* **2019**, *77*, 96–109. [CrossRef] [PubMed]
31. Nolte, V.; Pandey, R.V.; Jost, S.; Medinger, R.; Ottenwälder, B.; Boenigk, J.; Schlötterer, C. Contrasting seasonal niche separation between rare and abundant taxa conceals the extent of protist diversity. *Mol. Ecol.* **2010**, *19*, 2908–2915. [CrossRef] [PubMed]
32. Visco, J.A.; Apothéloz-Perret-Gentil, L.; Cordonier, A.; Esling, P.; Pillet, L.; Pawlowski, J. Environmental Monitoring: Inferring the Diatom Index from Next-Generation Sequencing Data. *Environ. Sci. Technol.* **2015**, *49*, 7597–7605. [CrossRef]
33. Apothéloz-Perret-Gentil, L.; Bouchez, A.; Cordier, T.; Cordonier, A.; Guéguen, J.; Rimet, F.; Vasselon, V.; Pawlowski, J. Monitoring the ecological status of rivers with diatom eDNA metabarcoding: A comparison of taxonomic markers and analytical approaches for the inference of a molecular diatom index. *Mol. Ecol.* **2021**, *30*, 2959–2968. [CrossRef]
34. Baillet, B.; Bouchez, A.; Franc, A.; Frigerio, J.-M.; Keck, F.; Karjalainen, S.-M.; Rimet, F.; Schneider, S.; Kahlert, M. Molecular versus morphological data for benthic diatoms biomonitoring in Northern Europe freshwater and consequences for ecological status. *Metabarcoding Metagenom.* **2019**, *3*, e34002. [CrossRef]
35. Mora, D.; Abarca, N.; Proft, S.; Grau, J.H.; Enke, N.; Carmona, J.; Skibbe, O.; Jahn, R.; Zimmermann, J. Morphology and metabarcoding: A test with stream diatoms from Mexico highlights the complementarity of identification methods. *Freshw. Sci.* **2019**, *38*, 448–464. [CrossRef]
36. Kutty, S.N.; Loh, R.K.; Bannister, W.; Taylor, D. Evaluation of a diatom eDNA-based technique for assessing water quality variations in tropical lakes and reservoirs. *Ecol. Indic.* **2022**, *141*, 109108. [CrossRef]
37. Zimmermann, J.; Jahn, R.; Gemeinholzer, B. Barcoding diatoms: Evaluation of the V4 subregion on the 18S rRNA gene, including new primers and protocols. *Org. Divers. Evol.* **2011**, *11*, 173–192. [CrossRef]
38. Zimmermann, J.; Abarca, N.; Enke, N.; Skibbe, O.; Kusber, W.-H.; Jahn, R. Taxonomic reference libraries for environmental barcoding: A best practice example from diatom research. *PLoS ONE* **2014**, *9*, e108793. [CrossRef] [PubMed]
39. Apothéloz-Perret-Gentil, L.; Cordonier, A.; Straub, F.; Iseli, J.; Esling, P.; Pawlowski, J. Taxonomy-free molecular diatom index for high-throughput eDNA biomonitoring. *Mol. Ecol. Resour.* **2017**, *17*, 1231–1242. [CrossRef]
40. Malashenkov, D.; Dashkova, V.; Zhakupova, K.; Vorobjev, I.; Barteneva, N. Comparative analysis of freshwater phytoplankton communities in two lakes of Burabay National Park using morphological and molecular approaches. *Sci. Rep.* **2021**, *11*, 16130. [CrossRef] [PubMed]
41. Hugerth, L.; Muller, E.; Hu, Y.; Lebrun, L.; Roume, H.; Wilmes, L.; Andersson, A. Systematic Design of 18S rRNA Gene Primers for Determining Eukaryotic Diversity in Microbial Consortia. *PLoS ONE* **2014**, *9*, e95567. [CrossRef] [PubMed]
42. Groendahl, S.; Kahlert, M.; Fink, P. The best of both worlds: A combined approach for analyzing microalgal diversity via metabarcoding and morphology-based methods. *PLoS ONE* **2017**, *12*, e0172808. [CrossRef]
43. Stoeck, T.; Bass, D.; Nebel, M.; Christen, R.; Jones, M.D.; Breiner, H.W.; Richards, T.A. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol. Ecol.* **2010**, *1*, 21–31. [CrossRef]
44. Filker, S.; Sommaruga, R.; Vila, I.; Stoeck, T. Microbial eukaryote plankton communities of high-mountain lakes from three continents exhibit strong biogeographic patterns. *Mol. Ecol.* **2016**, *25*, 2286–2301. [CrossRef]
45. Kammerlander, B.; Breiner, H.W.; Filker, S.; Sommaruga, R.; Sonntag, B.; Stoeck, T. High diversity of protistan plankton communities in remote high mountain lakes in the European Alps and the Himalayan mountains. *FEMS Microbiol. Ecol.* **2015**, *91*, fiv010. [CrossRef]
46. Seymour, M.; Edwards, F.; Cosby, B.; Kelly, M.; Bruyn, M.; Carvalho, G.; Creer, S. Executing multi-taxa eDNA ecological assessment via traditional metrics and interactive networks. *Sci. Total. Environ.* **2020**, *729*, 138801. [CrossRef] [PubMed]
47. Annenkova, N.V.; Giner, C.R.; Logares, R. Tracing the Origin of Planktonic Protists in an Ancient Lake. *Microorganisms* **2020**, *1*, 543. [CrossRef] [PubMed]
48. Hirakata, Y.; Hatamoto, M.; Oshiki, M.; Watari, T.; Kuroda, K.; Araki, N.; Yamaguchi, T. Temporal variation of eukaryotic community structures in UASB reactor treating domestic sewage as revealed by 18S rRNA gene sequencing. *Sci. Rep.* **2019**, *9*, 12783. [CrossRef] [PubMed]
49. Elerse, T. (Ed.) Technical Guidelines for eDNA Monitoring in Alpine Waters for Stakeholders and End-Users. 2021. Available online: <https://www.alpine-space.eu/project/eco-alpswater/> (accessed on 22 May 2023).
50. Brandani, J.; Peter, H.; Busi, S.B.; Kohler, T.J.; Fodelianakis, S.; Ezzat, L.; Michoud, G.; Bourquin, M.; Pramateftaki, P.; Roncoroni, M.; et al. Spatial patterns of benthic biofilm diversity among streams draining proglacial floodplains. *Front. Microbiol.* **2022**, *13*, 948165. [CrossRef]
51. Yang, N.; Wang, L.; Lin, L.; Li, Y.; Zhang, W.; Niu, L.; Zhang, H.; Wang, L. Pelagic-benthic coupling of the microbial food web modifies nutrient cycles along a cascade-dammed river. *Front. Environ. Sci. Eng.* **2022**, *16*, 50. [CrossRef]
52. Liang, D.; Xia, J.; Song, J.; Sun, H.; Xu, W. Using eDNA to Identify the Dynamic Evolution of Multi-Trophic Communities under the Eco-Hydrological Changes in River. *Front. Environ. Sci.* **2022**, *853*, 929541. [CrossRef]
53. Choi, J.; Park, J.S. Comparative analyses of the V4 and V9 regions of 18S rDNA for the extant eukaryotic community using the Illumina platform. *Sci. Rep.* **2020**, *10*, 6519. [CrossRef]
54. Bradley, I.M.; Pinto, A.J.; Guest, J.S. Design and evaluation of Illumina MiSeq-compatible, 18S rRNA gene-specific primers for improved characterization of mixed phototrophic communities. *Appl. Environ. Microbiol.* **2016**, *82*, 5878–5891. [CrossRef]

55. Carles, L.; Wulschleger, S.; Joss, A.; Eggen, R.I.; Schirmer, K.; Schuwirth, N.; Stamm, C.; Tlili, A. Impact of wastewater on the microbial diversity of periphyton and its tolerance to micropollutants in an engineered flow-through channel system. *Water Res.* **2021**, *203*, 117486. [CrossRef]
56. Carles, L.; Wulschleger, S.; Joss, A.; Eggen, R.; Schirmer, K.; Schuwirth, N.; Stamm, C.; Tlili, A. Wastewater microorganisms impact microbial diversity and important ecological functions of stream periphyton. *Water Res.* **2022**, *225*, 119119. [CrossRef] [PubMed]
57. Huo, S.; Li, X.; Xi, B.; Zhang, H.; Ma, C.; He, Z. Combining morphological and metabarcoding approaches reveals the freshwater eukaryotic phytoplankton community. *Environ. Sci. Eur.* **2020**, *32*, 37. [CrossRef]
58. Gast, R.J.; Dennett, M.R.; Caron, D.A. Characterization of Protistan assemblages in the Ross Sea, Antarctica, by denaturing gradient gel electrophoresis. *Appl. Environ. Microb.* **2004**, *70*, 2028–2037. [CrossRef]
59. Amaral-Zettler, L.A.; McCliment, E.A.; Ducklow, H.W.; Huse, S.M. A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. *PLoS ONE* **2009**, *4*, e6372. [CrossRef]
60. Lane, D.J. 16S/23S rRNA sequencing. In *Nucleic Acid Techniques in Bacterial Systematics*; Stackebrandt, E., Goodfellow, M., Eds.; John Wiley & Sons: New York, NY, USA, 1991; pp. 115–147.
61. The Earth Microbiome Project. Available online: <http://www.earthmicrobiome.org> (accessed on 22 May 2023).
62. Abad, D.; Albaina, A.; Aguirre, M.; Laza-Martínez, A.; Uriarte, I.; Iriarte, A.; Villate, F.; Estonba, A. Is metabarcoding suitable for estuarine plankton monitoring? A comparative study with microscopy. *Mar. Biol.* **2016**, *163*, 149. [CrossRef]
63. Yi, Z.; Berney, C.; Hartikainen, H.; Mahamdallie, S.; Gardner, M.; Boenigk, J.; Cavalier-Smith, T.; Bass, D. High-throughput sequencing of microbial eukaryotes in Lake Baikal reveals ecologically differentiated communities and novel evolutionary radiations. *FEMS Microbiol. Ecol.* **2017**, *93*, fix073. [CrossRef]
64. Ortiz-Álvarez, R.; Triadó-Margarit, X.; Camarero, L.; Casamayor, E.; Catalan, J. High planktonic diversity in mountain lakes contains similar contributions of autotrophic, heterotrophic and parasitic eukaryotic life forms. *Sci. Rep.* **2018**, *8*, 4457. [CrossRef]
65. Minerovic, A.D.; Potapova, M.G.; Sales, C.M.; Price, J.R.; Enache, M.D. 18S-V9 DNA metabarcoding detects the effect of water-quality impairment on stream biofilm eukaryotic assemblages. *Ecol. Indic.* **2020**, *113*, 106225. [CrossRef]
66. Hanžek, N.; Udovič, M.G.; Kajan, K.; Borics, G.; Várbiro, G.; Stoeck, T.; Žitunić, P.; Orlić, S.; Stanković, I. Assessing ecological status in karstic lakes through the integration of phytoplankton functional groups, morphological approach and environmental DNA metabarcoding. *Ecol. Indic.* **2021**, *131*, 108166. [CrossRef]
67. Medlin, L.; Elwood, H.J.; Stickel, S.; Sogin, M.L. The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. *Gene* **1988**, *71*, 491–499. [CrossRef]
68. Boenigk, J.; Wodniok, S.; Bock, C.; Beisser, D.; Hempel, C.; Grossmann, L.; Lange, A.; Jensen, M. Geographic distance and mountain ranges structure freshwater protist communities on a European scale. *Metabarcoding Metagenom.* **2018**, *2*, e21519. [CrossRef]
69. Bock, C.; Jensen, M.; Forster, D.; Marks, S.; Nuy, J.; Psenner, R.; Beisser, D.; Boenigk, J. Factors shaping community patterns of protists and bacteria on a European scale. *Environ. Microbiol.* **2020**, *22*, 2243–2260. [CrossRef] [PubMed]
70. Olefeld, J.L.; Bock, C.; Jensen, M.; Vogt, J.; Sieber, G.; Albach, D.; Boenigk, J. Centers of endemism of freshwater protists deviate from pattern of taxon richness on a continental scale. *Sci. Rep.* **2020**, *10*, 14431. [CrossRef] [PubMed]
71. Bock, C.; Olefeld, J.L.; Vogt, J.C.; Albach, D.; Boenigk, J. Phylogenetic and functional diversity of Chrysophyceae in inland waters. *Org. Divers. Evol.* **2022**, *2022*, 327–341. [CrossRef]
72. White, T.J.; Bruns, T.D.; Lee, S.B.; Taylor, J.W. Amplification and Direct Sequencing of Fungal Ribosomal RNA Genes for Phylogenetics. In *PCR Protocols: A Guide to Methods and Applications*; Innis, M.A., Gelfand, D.H., Sninsky, J.J., White, T.J., Eds.; Academic Press: New York, NY, USA, 1990; pp. 315–322. [CrossRef]
73. Câmara, P.E.; Menezes, G.C.; Pinto, O.H.; Silva, M.C.; Convey, P.; Rosa, L.H. Using metabarcoding to assess Viridiplantae sequence diversity present in Antarctic glacial ice. *Annu. Acad. Bras. Cienc.* **2022**, *94*, e20201736. [CrossRef]
74. Fonseca, B.M.; Câmara, P.E.A.S.; Ogaki, M.B.; Pinto, O.; Lirio, J.; Coria, S.; Vieira, R.; Carvalho-Silva, M.; Amorim, E.; Convey, P.; et al. Green algae (Viridiplantae) in sediments from three lakes on Vega Island, Antarctica, assessed using DNA metabarcoding. *Mol. Biol. Rep.* **2022**, *49*, 179–188. [CrossRef]
75. Rimet, F.; Vasselon, V.; A-Keszte, B.; Bouchez, A. Do we similarly assess diversity with microscopy and high-throughput sequencing? Case of microalgae in lakes. *Org. Divers. Evol.* **2018**, *18*, 51–62. [CrossRef]
76. Câmara, P.E.A.S.; Carvalho-Silva, M.; Pinto, O.H.B.; Amorim, E.; Henriques, D.; Holanda da Silva, T.; Pellizzari, F.; Convey, P.; Rosa, L. Diversity and Ecology of Chlorophyta (Viridiplantae) Assemblages in Protected and Non-protected Sites in Deception Island (Antarctica, South Shetland Islands) Assessed Using an NGS Approach. *Microb. Ecol.* **2021**, *81*, 323–334. [CrossRef]
77. Rimet, F.; Abarca, N.; Bouchez, A.; Kusber, W.-H.; Jahn, R.; Kahlert, M.; Keck, F.; Kelly, M.; Mann, D.; Piuze, A.; et al. The potential of High-Throughput Sequencing (HTS) of natural samples as a source of primary taxonomic information for reference libraries of diatom barcodes. *Fottea* **2018**, *18*, 37–54. [CrossRef]
78. Rivera, S.F.; Vasselon, V.; Ballorain, K.; Carpentier, A.; Wetzel, C.; Ector, L.; Bouchez, A.; Rimet, F. DNA metabarcoding and microscopic analyses of sea turtles biofilms: Complementary to understand turtle behavior. *PLoS ONE* **2018**, *13*, e0195770. [CrossRef]
79. Maitland, V.C.; Robinson, C.V.; Porter, T.M.; Hajibabaei, M. Freshwater diatom biomonitoring through benthic kick-net metabarcoding. *PLoS ONE* **2020**, *15*, e0242143. [CrossRef] [PubMed]

80. Pérez-Burillo, J.; Trobajo, R.; Leira, M.; Keck, F.; Rimet, F.; Sigró, J.; Mann, D.G. DNA metabarcoding reveals differences in distribution patterns and ecological preferences among genetic variants within some key freshwater diatom species. *Sci. Total Environ.* **2021**, *798*, 149029. [\[CrossRef\]](#)
81. Smucker, N.J.; Pilgrim, E.M.; Nietch, C.T.; Darling, J.A.; Johnson, B.R. DNA metabarcoding effectively quantifies diatom responses to nutrients in streams. *Ecol. Appl.* **2020**, *30*, e02205. [\[CrossRef\]](#)
82. Pissaridou, P.; Cantonati, M.; Bouchez, A.; Tziortzis, I.; Dörflinger, G.; Vasquez, M.I. How can integrated morphotaxonomy- and metabarcoding-based diatom assemblage analyses best contribute to the ecological assessment of streams? *Metabarcoding Metagenom.* **2021**, *5*, e68438. [\[CrossRef\]](#)
83. Borrego-Ramos, M.; Bécares, E.; García, P.; Nistal, A.; Blanco, S. Epiphytic Diatom-Based Biomonitoring in Mediterranean Ponds: Traditional Microscopy versus Metabarcoding Approaches. *Water* **2021**, *13*, 1351. [\[CrossRef\]](#)
84. Rimet, F.; Pinseel, E.; Bouchez, A.; Japoshvili, B.; Mumladze, L. Diatom endemism and taxonomic turnover: Assessment in high-altitude alpine lakes covering a large geographical range. *Sci. Total. Environ.* **2023**, *871*, 161970. [\[CrossRef\]](#)
85. Kahlert, M.; Karjalainen, S.M.; Keck, F.; Kelly, M.; Ramon, M.; Rimet, F.; Schneider, S.; Tapolczai, K.; Zimmermann, J. Co-occurrence, ecological profiles and geographical distribution based on unique molecular identifiers of the common freshwater diatoms *Fragilaria* and *Ulnaria*. *Ecol. Indic.* **2022**, *141*, 109114. [\[CrossRef\]](#)
86. Pérez-Burillo, J.; Mann, D.G.; Trobajo, R. Evaluation of two short overlapping *rbcL* markers for diatom metabarcoding of environmental samples: Effects on biomonitoring assessment and species resolution. *Chemosphere* **2022**, *307*, 135933. [\[CrossRef\]](#)
87. Yuan, L.L.; Mitchell, R.M.; Pollard, A.I.; Nietch, C.T.; Pilgrim, E.M.; Smucker, N.J. Understanding the effects of phosphorus on diatom richness in rivers and streams using taxon–environment relationships. *Freshw. Biol.* **2023**, *68*, 473–486. [\[CrossRef\]](#)
88. Bíró, T.; Duleba, M.; Földi, A.; Kiss, K.T.; Orgoványi, P.; Trábert, Z.; Vadkerti, E.; Wetzel, C.E.; Ács, É. Metabarcoding as an effective complement of microscopic studies in revealing the composition of the diatom community—A case study of an oxbow lake of Tisza River (Hungary) with the description of a new *Mayamaea* species. *Metabarcoding Metagenom.* **2022**, *6*, e87497. [\[CrossRef\]](#)
89. Robinson, C.; Porter, T.; Maitland, V.; Wright, M.; Hajibabaei, M. Multi-marker metabarcoding resolves subtle variations in freshwater condition: Bioindicators, ecological traits, and trophic interactions. *Ecol. Indic.* **2022**, *145*, 109603. [\[CrossRef\]](#)
90. Vasselon, V.; Rimet, F.; Tapolczai, K.; Bouchez, A. Assessing ecological status with diatoms DNA metabarcoding: Scaling-up on a WFD monitoring network (Mayotte island, France). *Ecol. Indic.* **2017**, *82*, 1–12. [\[CrossRef\]](#)
91. Duleba, M.; Földi, A.; Micsinai, A.; Várbiro, G.; Mohr, A.; Sipos, R.; Szabó, G.; Buczkó, K.; Trábert, Z.; Kiss, K.; et al. Applicability of diatom metabarcoding in the ecological status assessment of Hungarian lotic and soda pan habitats. *Ecol. Indic.* **2021**, *130*, 108105. [\[CrossRef\]](#)
92. Vasselon, V.; Rimet, F.; Domaizon, I.; Monnier, O.; Reyjol, Y.; Bouchez, A. Assessing pollution of aquatic environments with diatoms' DNA metabarcoding: Experience and developments from France Water Framework Directive networks. *Metabarcoding Metagenom.* **2019**, *3*, e39646. [\[CrossRef\]](#)
93. Mortágua, A.; Vasselon, V.; Oliveira, R.; Elias, C.; Chardon, C.; Bouchez, A.; Rimet, F.; Feio, M.; Almeida, S. Applicability of DNA metabarcoding approach in the bioassessment of Portuguese rivers using diatoms. *Ecol. Indic.* **2019**, *106*, 105470. [\[CrossRef\]](#)
94. Tapolczai, K.; Selmečzy, G.B.; Szabó, B.; B.-Béres, V.; Keck, F.; Bouchez, A.; Rimet, F.; Padišák, J. The potential of exact sequence variants (ESVs) to interpret and assess the impact of agricultural pressure on stream diatom assemblages revealed by DNA metabarcoding. *Ecol. Ind.* **2021**, *122*, 107322. [\[CrossRef\]](#)
95. Baker, L.A.; Beauger, A.; Kolovi, S.; Voldoire, O.; Allain, E.; Breton, V.; Chardon, P.; Miallier, D.; Bailly, C.; Montavon, G.; et al. Diatom DNA metabarcoding to assess the effect of natural radioactivity in mineral springs on ASV of benthic diatom communities. *Sci. Total Environ.* **2023**, *873*, 162270. [\[CrossRef\]](#)
96. Kang, W.; Anslan, S.; Börner, N.; Schwarz, A.; Schmidt, R.; Künzel, S.; Rioual, P.; Echeverría-Galindo, P.; Vences, M.; Wang, J.; et al. Diatom metabarcoding and microscopic analyses from sediment samples at Lake Nam Co, Tibet: The effect of sample-size and bioinformatics on the identified communities. *Ecol. Indic.* **2021**, *121*, 107070. [\[CrossRef\]](#)
97. Fawley, M.W.; Fawley, K.P.; Cahoon, A.B. Finding needles in a haystack—Extensive diversity in the eustigmatophyceae revealed by community metabarcode analysis targeting the *rbcL* gene using lineage-directed primers. *J. Phycol.* **2021**, *57*, 1636–1647. [\[CrossRef\]](#)
98. Qiao, L.; Chang, Z.; Li, J.; Chen, Z.; Yang, L.; Luo, Q. Phytoplankton community structure and diversity in the indoor industrial aquaculture system for *Litopenaeus vannamei* revealed by high-throughput sequencing and morphological identification. *Aquac. Res.* **2019**, *50*, 2563–2576. [\[CrossRef\]](#)
99. Brown, P.D.; Craine, J.M.; Richards, D.; Chapman, A.; Marden, B. DNA metabarcoding of the phytoplankton of Great Salt Lake's Gilbert Bay: Spatiotemporal assemblage changes and comparisons to microscopy. *J. Great Lakes Res.* **2022**, *48*, 110–124. [\[CrossRef\]](#)
100. Sherwood, A.R.; Presting, G.G. Universal primers amplify a 23S rDNA plastid marker in eukaryotic algae and cyanobacteria. *J. Phycol.* **2007**, *43*, 605–608. [\[CrossRef\]](#)
101. Eiler, A.; Drakare, S.; Bertilsson, S.; Pernthaler, J.; Peura, S.; Rofner, C.; Simek, K.; Yang, Y.; Znachor, P.; Lindström, E. Unveiling Distribution Patterns of Freshwater Phytoplankton by a Next Generation Sequencing Based Approach. *PLoS ONE* **2013**, *8*, e53516. [\[CrossRef\]](#)
102. Parada, A.E.; Needham, D.M.; Fuhrman, J.A. Primers for marine microbiome studies. *Environ. Microbiol.* **2016**, *18*, 1403–1414. [\[CrossRef\]](#) [\[PubMed\]](#)

103. Bonfantine, K.L.; Trevathan-Tackett, S.M.; Matthews, T.G.; Neckovic, A.; Gan, H.M. Dumpster diving for diatom plastid 16S rRNA genes. *PeerJ*. **2021**, *9*, e11576. [\[CrossRef\]](#) [\[PubMed\]](#)
104. Caporaso, J.G.; Lauber, C.L.; Walters, W.A.; Berg-Lyons, D.; Lozupone, C.; Turnbaugh, P.; Fierer, N.; Knight, R. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 4516–4522. [\[CrossRef\]](#)
105. Rimet, F.; Gusev, E.; Kahlert, M.; Kelly, M.; Kulikovskiy, M.; Maltsev, Y.; Mann, D.; Pfannkuchen, M.; Trobajo, R.; Vasselon, V.; et al. Diat. barcode, an open-access curated barcode library for diatoms. *Sci. Rep.* **2019**, *9*, 1–12. [\[CrossRef\]](#) [\[PubMed\]](#)
106. Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Glöckner, F.O. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **2013**, *41*, D590–D596. [\[CrossRef\]](#)
107. Banchi, E.; Ametrano, C.G.; Greco, S.; Stanković, D.; Muggia, L.; Pallavicini, A. PLANITS: A curated sequence reference dataset for plant ITS DNA metabarcoding. *Database* **2020**, *2020*, baz155. [\[CrossRef\]](#)
108. Decelle, J.; Romac, S.; Stern, R.F.; Bendif, E.M.; Zingone, A.; Audic, S.; Guiry, M.; Guillou, L.; Tessier, D.; Le Gall, F.; et al. PhytoREF: A reference database of the plastidial 16S rRNA gene of photosynthetic eukaryotes with curated taxonomy. *Mol. Ecol. Resour.* **2015**, *15*, 1435–1445. [\[CrossRef\]](#)
109. Evans, K.M.; Wortley, A.H.; Mann, D.G. An assessment of potential diatom “barcode” genes (*cox1*, *rbcL*, 18S and ITS rDNA) and their effectiveness in determining relationships in *Sellaphora* (Bacillariophyta). *Protist* **2007**, *158*, 349–364. [\[CrossRef\]](#) [\[PubMed\]](#)
110. Behnke, A.; Friedl, T.; Chepurnov, V.A.; Mann, D.G. Reproductive Compatibility and rDNA Sequence Analyses in the *Sellaphora* *Pupula* Species Complex (Bacillariophyta). *J. Phycol.* **2004**, *40*, 193–208. [\[CrossRef\]](#)
111. Moniz, M.B.J.; Kaczmarska, I. Barcoding diatoms: Is there a good marker? *Mol. Ecol. Resour.* **2009**, *9*, 65–74. [\[CrossRef\]](#)
112. Moniz, M.B.J.; Kaczmarska, I. Barcoding of diatoms: Nuclear encoded ITS revisited. *Protist* **2010**, *161*, 7–34. [\[CrossRef\]](#) [\[PubMed\]](#)
113. Hamsher, S.E.; Evans, K.M.; Mann, D.G.; Poulickova, A.; Saunders, G.W. Barcoding diatoms: Exploring alternatives to COI-5P. *Protist* **2011**, *162*, 405–422. [\[CrossRef\]](#) [\[PubMed\]](#)
114. MacGillivray, M.L.; Kaczmarska, I. Survey of the Efficacy of a Short Fragment of the *rbcL* Gene as a Supplemental DNA Barcode for Diatoms. *J. Eukaryot. Microbiol.* **2011**, *58*, 529–536. [\[CrossRef\]](#) [\[PubMed\]](#)
115. Kermarrec, L.; Franc, A.; Rimet, F.; Chaumeil, P.; Humbert, J.F.; Bouchez, A. Next-generation sequencing to inventory taxonomic diversity in eukaryotic communities: A test for freshwater diatoms. *Mol. Ecol. Resour.* **2013**, *13*, 607–619. [\[CrossRef\]](#) [\[PubMed\]](#)
116. Hall, J.D.; Fučíková, K.; Lo, C.; Lewis, L.A.; Karol, K.G. An assessment of proposed DNA barcodes in freshwater green algae. *Cryptogam. Algal.* **2010**, *31*, 529–555.
117. Litaker, R.W.; Vandersea, M.W.; Kibler, S.R.; Reece, K.S.; Stokes, N.A.; Lutzoni, F.M.; Yonish, F.M.; West, M.A.; Black, M.N.D. Recognizing dinoflagellate species using ITS rDNA sequences. *J. Phycol.* **2007**, *43*, 344–355. [\[CrossRef\]](#)
118. Stern, R.; Horak, A.; Andrew, R.; Coffroth, M.; Andersen, R.; Andersen, R.; Küpper, F.; Jameson, I.; Hoppenrath, M.; Véron, B.; et al. Environmental Barcoding Reveals Massive Dinoflagellate Diversity in Marine Environments. *PLoS ONE* **2010**, *5*, e13991. [\[CrossRef\]](#)
119. Stern, R.F.; Andersen, R.A.; Jameson, I.; Küpper, F.C.; Coffroth, M.A.; Vulot, D.; Le Gall, F.; Véron, B.; Brand, J.J.; Skelton, H.; et al. Evaluating the ribosomal internal transcribed spacer (ITS) as a candidate dinoflagellate barcode marker. *PLoS ONE* **2012**, *7*, e42780. [\[CrossRef\]](#) [\[PubMed\]](#)
120. La Jeunesse, T.C.; Thornhill, D.J. Improved Resolution of Reef-Coral Endosymbiont (*Symbiodinium*) Species Diversity, Ecology, and Evolution through *psbA* Non-Coding Region Genotyping. *PLoS ONE* **2011**, *6*, e29013. [\[CrossRef\]](#)
121. Tanabe, A.S.; Nagai, S.; Hida, K.; Yasuike, M.; Fujiwara, A.; Nakamura, Y.; Takano, Y.; Katakura, S. Comparative study of the validity of three regions of the 18S-rRNA gene for massively parallel sequencing-based monitoring of the planktonic eukaryote community. *Mol. Ecol. Resour.* **2016**, *16*, 402–414. [\[CrossRef\]](#)
122. Tragin, M.; Zingone, A.; Vulot, D. Comparison of coastal phytoplankton composition estimated from the V4 and V9 regions of the 18S rRNA gene with a focus on photosynthetic groups and especially Chlorophyta. *Environ. Microbiol.* **2018**, *20*, 506–520. [\[CrossRef\]](#)
123. Giner, C.R.; Forn, I.; Romac, S.; Logares, R.; de Vargas, C.; Massana, R. Environmental Sequencing Provides Reasonable Estimates of the Relative Abundance of Specific Picoeukaryotes. *Appl. Environ. Microbiol.* **2016**, *82*, 4757–4766. [\[CrossRef\]](#)
124. Maritz, J.M.; Rogers, K.H.; Rock, T.M.; Liu, N.; Joseph, S.; Land, K.; Carlton, J. An 18S rRNA Workflow for Characterizing Protists in Sewage, with a Focus on Zoonotic Trichomonads. *Microb. Ecol.* **2017**, *74*, 923–936. [\[CrossRef\]](#)
125. Zheng, X.; He, Z.; Wang, C.; Yan, Q.; Shu, L. Evaluation of different primers of the 18S rRNA gene to profile amoeba communities in environmental samples. *Water Biol. Secur.* **2022**, *1*, 100057. [\[CrossRef\]](#)
126. Piredda, R.; Tomasino, M.P.; D’Erchia, A.M.; Manzari, C.; Pesole, G.; Montresor, M.; Kooistra, W.; Sarno, D.; Zingone, A. Diversity and temporal patterns of planktonic protist assemblages at a Mediterranean Long Term Ecological Research site. *FEMS Microbiol. Ecol.* **2017**, *93*, fiw200. [\[CrossRef\]](#)
127. Buchheim, M.A.; Keller, A.; Koetschan, C.; Förster, F.; Merget, B.; Wolf, M. Internal Transcribed Spacer 2 (nu ITS2 rRNA) Sequence-Structure Phylogenetics: Towards an Automated Reconstruction of the Green Algal Tree of Life. *PLoS ONE* **2011**, *6*, e16931. [\[CrossRef\]](#) [\[PubMed\]](#)
128. Câmara, P.E.A.S.; de Menezes, G.C.A.; Oliveira, F.S.; Souza, C.D.; Amorim, E.T.; Schaefer, C.E.; Convey, P.; Pinto, O.H.; Carvalho-Silva, M.; Rosa, L.H. Diversity of Viridiplanta e DNA present on rock surfaces in the Ellsworth Mountains, continental Antarctica. *Polar. Biol.* **2022**, *45*, 637–646. [\[CrossRef\]](#)

129. Wang, Y.-K.; Stevenson, R.J.; Metzmeier, L. Development and evaluation of a diatom-based Index of Biotic Integrity for the Interior Plateau Ecoregion, USA. *J. North Am. Benthol. Soc.* **2005**, *24*, 990–1008. [\[CrossRef\]](#)
130. Chessman, B.C.; Bate, N.; Gell, P.A.; Newall, P. A diatom species index for bioassessment of Australian rivers. *Mar. Freshw. Res.* **2007**, *58*, 542–557. [\[CrossRef\]](#)
131. Lavoie, I.; Hamilton, P.B.; Wang, Y.-K.; Dillon, P.J.; Campeau, S. A comparison of stream bioassessment in Québec (Canada) using six European and North American diatom-based indices. *Nova Hedwig.* **2009**, *135*, 37–56.
132. Zimmermann, J.; Glöckner, G.; Jahn, R.; Enke, N.; Gemeinholzer, B. Metabarcoding vs. morphological identification to assess diatom diversity in environmental studies. *Mol. Ecol. Resour.* **2015**, *15*, 526–542. [\[CrossRef\]](#)
133. Kermarrec, L.; Franc, A.; Rimet, F.; Chaumeil, P.; Frigerio, J.M.; Humbert, J.F.; Bouchez, A. A next-generation sequencing approach to river biomonitoring using benthic diatoms. *Freshw. Sci.* **2014**, *33*, 349–363. [\[CrossRef\]](#)
134. Wolf, D.I.; Vis, M.L. Stream Algal Biofilm Community Diversity Along an Acid Mine Drainage Recovery Gradient Using Multimer Metabarcoding. *J. Phycol.* **2020**, *11–22*, 12935. [\[CrossRef\]](#)
135. Jackson, E.E.; Hawes, I.; Jungblut, A.D. 16S rRNA gene and 18S rRNA gene diversity in microbial mat communities in meltwater ponds on the McMurdo Ice Shelf, Antarctica. *Polar Biol.* **2021**, *44*, 823–836. [\[CrossRef\]](#)
136. Keck, F.; Couton, M.; Altermatt, F. Navigating the seven challenges of taxonomic reference databases in metabarcoding analyses. *Mol. Ecol. Resour.* **2022**, *23*, 742–755. [\[CrossRef\]](#)
137. Mann, D.G.; Evans, K.M. The species concept and cryptic diversity, Moestrup, Ø, Eds. In Proceedings of the 12th International Conference on Harmful Algae, International Society for the Study of Harmful Algae and Intergovernmental Oceanographic Commission of UNESCO, Copenhagen, Sweden, 4–8 August 2008; pp. 262–268.
138. Souffreau, C.; Vanormelingen, P.; Van de Vijver, B.; Isheva, T.; Verleyen, E.; Sabbe, K.; Vyverman, W. Molecular evidence for distinct Antarctic lineages in the cosmopolitan terrestrial diatoms *Pinnularia borealis* and *Hantzschia amphioxys*. *Protist* **2013**, *164*, 101–115. [\[CrossRef\]](#)
139. Darienko, T.; Gustavs, L.; Eggert, A.; Wolf, W.; Pröschold, T. Evaluating the Species Boundaries of Green Microalgae (*Coccomyxa*, *Trebouxiophyceae*, *Chlorophyta*) Using Integrative Taxonomy and DNA Barcoding with Further Implications for the Species Identification in Environmental Samples. *PLoS ONE* **2015**, *10*, e0127838. [\[CrossRef\]](#)
140. Proeschöld, T.; Darienko, T. The green puzzle *Stichococcus* (*Trebouxiophyceae*, *Chlorophyta*): New generic and species concept among this widely distributed genus. *Phytotaxa* **2020**, *441*, 113–142. [\[CrossRef\]](#)
141. Irisarri, I.; Darienko, T.; Pröschold, T.; Fürst-Jansen, J.M.R.; Jamy, M.; de Vries, J. Unexpected cryptic species among streptophyte algae most distant to land plants. *Proc. Biol. Sci.* **2021**, *288*, 1963. [\[CrossRef\]](#) [\[PubMed\]](#)
142. Hoef-Emden, K. Revision of the genus *Chroomonas* Hansgirg: The benefits of DNA-containing specimens. *Protist* **2018**, *169*, 662–681. [\[CrossRef\]](#) [\[PubMed\]](#)
143. MacKeigan, P.W.; Garner, R.E.; Monchamp, M.È.; Walsh, D.; E Onana, V.; Kraemer, S.; Pick, F.; Beisner, B.; Agbeti, M.; Barbosa da Costa, N.; et al. Comparing microscopy and DNA metabarcoding techniques for identifying cyanobacteria assemblages across hundreds of lakes. *Harmful Algae* **2022**, *113*, 102187. [\[CrossRef\]](#) [\[PubMed\]](#)
144. Dzhenbekova, N.; Moncheva, S.; Ivanova, P.; Slabakova, N.; Nagai, S. Biodiversity of phytoplankton cyst assemblages in surface sediments of the Black Sea based on metabarcoding. *Biotechnol. Biotechnol. Equip.* **2018**, *32*, 1507–1513. [\[CrossRef\]](#)
145. CEN/TR 17245:2018; Water Quality—Technical Report for the Routine Sampling of Benthic Diatoms from Rivers and Lakes Adapted for Metabarcoding Analyses. iTeh Standards: Newark, DE, USA, 2018.
146. CEN/TR 17244:2018; Water Quality—Technical Report for the Management of Diatom Barcodes. iTeh Standards: Newark, DE, USA, 2018; CEN/TC 230/WG 23—Aquatic Macrophytes and Algae: 2018.
147. Xie, Y.; Giesy, J.P. *UofS-ETL-EDNA-30 Metabarcoding of Cyanobacteria Assembly*; University of Saskatchewan: Saskatoon, SK, Canada, 2018; Version 1.
148. Yarimizu, K.; Fujiyoshi, S.; Kawai, M.; Norambuena-Subiabre, L.; Cascales, E.-K.; Rilling, J.-I.; Vilugrón, J.; Cameron, H.; Vergara, K.; Morón-López, J.; et al. Protocols for monitoring harmful algal blooms for sustainable aquaculture and coastal fisheries in Chile. *Int. J. Env. Res. Pub. Health* **2020**, *17*, 7642. [\[CrossRef\]](#)
149. Jerney, J.; Hällfors, H.; Oja, J.; Reunamo, A.; Suikkanen, S.; Lehtinen, S. Guidelines for using environmental DNA in Finnish marine phytoplankton monitoring—Improved biodiversity assessment through method complementation. *Rep. Finn. Environ. Inst.* **2022**, *40*, 69.
150. Pawlowski, J.; Audic, S.; Adl, S.; Bass, D.; Belbahri, L.; Berney, C.; Bowser, S.; Cepicka, I.; Decelle, J.; Dunthorn, M.; et al. CBOL protist working group: Barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. *PLoS Biol.* **2012**, *10*, e1001419. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.