Evaluation of V2 18S rDNA barcode marker and assessment of sample collection and DNA extraction methods for metabarcoding of autotrophic euglenids

Natalia Gumieńska, Maja Łukomska-Kowalczyk,*, Katarzyna Chaber, Bożena Zakryś and Rafał Milanowski
Institute of Evolutionary Biology, Faculty of Biology, Biological and Chemical Research Centre, University of Warsaw, ul. Zwirki i Wigury 101, PL-02-089, Warsaw, Poland.

Summary

Even though the interest in metabarcoding in environmental studies is growing, euglenids are still underrepresented in both sea and freshwater bodies researches. The reason for this situation could be the unsuitability of universal eukaryotic DNA barcodes and primers as well as the lack of a verified protocol, suitable to assess euglenid diversity. In this study, using specific primers for the V2 hypervariable region of 18S rDNA for metabarcoding resulted in obtaining a high fraction (85%) of euglenid reads and species-level identification of almost 90% of them. Fifty species were detected by the metabarcoding method, including almost all species observed using a light microscope. We investigated three biomass harvesting methods (filtering, centrifugation and scraping the side of a collection vessel) and determined that centrifugation and filtration outperformed scrapes, but the choice between them is not crucial for the reliability of the analysis. In addition, eight DNA extraction methods were evaluated. We compared five commercially available DNA isolation kits, two CTAB-based protocols and a chelating resin. For this purpose, the efficiency of extraction, quality of obtained DNA, preparation time and generated costs were taken into consideration. After examination of the aforementioned criteria, we chose the GeneMATRIX Soil DNA Purification Kit as the most suitable for DNA isolation.

Introduction

With the advancement of next-generation sequencing (NGS) technologies, DNA metabarcoding has become a powerful molecular tool frequently used for biodiversity assessment (Stefanni et al., 2018; Gran-Stadniczeñko et al., 2019). It was proven to be a sensitive, noninvasive, time-efficient and inexpensive monitoring technique that averts many limitations of microscopic observation, such as the lack of distinct morphological diagnostic traits (Pawlowski et al., 2014; Jeunen et al., 2019).

However, metabarcoding is not a bias-resistant strategy. Errors can be introduced at each stage of the workflow: field sampling, laboratory procedure and data processing (McKee et al., 2015; Zizka et al., 2019). For example, the selection of biomass harvesting, and DNA extraction methods are known to affect the yield and quality of isolated DNA (Deiner et al., 2015; Jeunen et al., 2019). Another drawback of metabarcoding surveys is associated with potential PCR flaws and sequencing artefacts. It has been pointed out that tagged primers can bias the sequence counts in multiplexed samples. It comes mainly through the differential amplification efficiency among heterogeneous templates, mispriming and tag jumps (Binladen et al., 2007; Zizka et al., 2019). The selection of bioinformatic pipelines and high-quality reference databases are also crucial, as data processing can critically impact the conclusions of the study - since both false negatives and false positives may not be screened out at the computational stage (Lamb et al., 2018; Doi et al., 2019). Nevertheless, the experimentally validated methodology ensures reproducibility, thus allowing to face the abovementioned challenges (Dickie et al., 2018; Lamb et al., 2018).

Euglenids are one of the several groups of eukaryotic microorganisms known to be underrepresented in metabarcoding studies (Simon et al., 2015; Łukomska-Kowalczyk et al., 2016; Gran-Stadniczeñko et al., 2019). Phototrophic euglenids (Euglenophyceae) are free-living, unicellular flagellated algae, widespread in predominantly freshwater ecosystems. They usually inhabit lentic eutrophic environments and occur abundantly in small reservoirs, such as ponds (Zakryś et al., 2017). Most taxa are
probably ubiquitous, although no comprehensive research on ecology of photosynthetic euglenids has been carried out yet (Triemer and Zakryś, 2015; Bicudo and Menezes, 2016). Green euglenids display a limited number of diagnostic features, thus species recognition is often arguable (Karnkowska-Ishikawa et al., 2010, 2013; Kim and Shin, 2014; Łukomska-Kowalczyk et al., 2015). Currently, an increasing number of species is being represented in sequence databases, thus phylogenetic relationships of taxa become resolved and supported (Kim et al., 2015; Łukomska-Kowalczyk et al., 2020a; Łukomska-Kowalczyk et al., 2020b). Furthermore, DNA extraction from euglenids is hampered by their cell cover (pellicle), secreted mucus and loricae of Trachelomonas and Strombomonas genera (Guminska et al., 2018). Finally, long and highly variable nuclear SSU rRNA gene (nSSU rDNA) of euglenids is presumed to be responsible for hindering amplification with universal eukaryotic primers (Łukomska-Kowalczyk et al., 2016). Recently, a database of nSSU rDNA reference sequences of taxonomically validated species was established and variable regions V2–V3 and V4 of 18S rDNA were suggested as DNA barcodes of phototrophic euglenids (above 95% and 90% efficiency respectively) (Łukomska-Kowalczyk et al., 2016).

In addition to the challenges caused by the peculiarity of euglenids, there are also those arising from the properties of their habitat (Łukomska-Kowalczyk et al., 2016; Zakryś et al., 2017). In eutrophic ponds, the water is usually murky, as it contains high amounts of planktonic organisms (e.g. bacteria, algae), cellular debris, suspended silt and organic compounds. Therefore filtering, commonly used in researching marine environments, is impeded by clogged filters and spin columns, which may reduce DNA recovery (Klymus et al., 2017; Harper et al., 2019). Alternatively, samples can be preprocessed via centrifugation (Eichmiller et al., 2016; Klymus et al., 2017). However, the decantation of the supernatant may cause a considerable loss of suspended DNA (Eichmiller et al., 2016; Jeunen et al., 2019). What is more, pond water often contains humic acids and other substances reducing amplification efficiency or completely restraining PCR (McKee et al., 2015; Goldberg et al., 2016). It has been proven that PCR inhibition can provide false-negative results, thus it must be alleviated through optimization of the workflow (McKee et al., 2015; Harper et al., 2019).

Despite the growing interest in metabarcoding surveys of freshwater ponds, the lack of a verified protocol to assess the diversity of euglenids in natural environments prompted us to undertake the present study. Herein, we applied the V2 18S rDNA region for the surveillance of phototrophic euglenids from the chosen hypereutrophic reservoir. The research covered morphological and molecular identification of taxa. We investigated the performance of three biomass harvesting and eight DNA extraction methods in terms of DNA yield and quality, preparation time and generated costs. In addition, we assessed amplicon sequence variants (ASVs) and species richness per replicate. Finally, we investigated whether the proposed genetic marker can be used for the identification of autotrophic euglenids.

Results

Spectrophotometric measurements

The average DNA concentrations varied from 2.18 to 120.58 ng μl⁻¹, depending on the method of biomass harvesting and the isolation procedure (Fig. 1, Supplementary Table S1). Among the biomass collecting protocols, the highest total average amounts of DNA were obtained from material captured by centrifugation (47.81 ng μl⁻¹), and the lowest from scrapes (9.28 ng μl⁻¹) (Table 1). Among the extraction protocols of the commercial, silica membrane-based kits, the highest average concentration of DNA (20.93 ng μl⁻¹) was yielded with SOI(E), whereas the lowest average concentrations (6.07 ng μl⁻¹) were obtained with WAT(O) (Table 2). It is also the lowest average value among all investigated methods of DNA purification. According to the spectrophotometric measurements, solution-based DNA extraction methods using CTAB and CHEL can be considered much more efficient than the commercial kits. On average, 54.94 ng μl⁻¹ of DNA was obtained with three manual protocols. Of these, CTAB yielded the highest DNA concentration (with an average value of 64.38 ng μl⁻¹).

According to the 260/280 nm absorbance ratios, SOI(E) extracted the purest DNA, followed by two other commercial kits: ENV(E) and WAT(O) (Table 2, Supplementary Table S1). The Commercial kits produced purer DNA samples compared to the standard in-solution methods with regards to this criterion. In the case of DNE(Q), two CTAB-based protocols and CHEL, the average values of this ratio were lower than 1.7, which could indicate either contamination by proteins or residual reagents (i.e. chaotropic salts).

In terms of the 260/230 nm absorbance ratios, ENV(E) and SOI(E) were associated with the highest, hence the most desirable values (Table 2, Supplementary Table S1). In the case of the six methods (DNE(Q), SOI(O), WAT(O) and three non-spin column protocols), the average values of this ratio were below 1.0, which
may be either the result of reagent contamination or carbohydrate carryover.

**Fluorometric measurements**

The highest average concentrations of DNA were obtained in the centrifuged samples (12.50 ng μl⁻¹), and the lowest in the samples collected by scraping (3.79 ng μl⁻¹) (Table 1; Fig. 1, Supplementary Table S1). According to the Qubit 3.0 (Thermo Scientific), out of the commercial, silica membrane-based kits, SOI(E) extracted the most concentrated DNA (17.77 ng μl⁻¹ on average), whereas DNE(Q) was, on average, the least efficient (2.51 ng μl⁻¹). Among the manual methods, CTAB+D (14.58 ng μl⁻¹ on average) outperformed the other two protocols (Table 2; Fig. 1).

The results of spectrophotometric and fluorometric measurements differed from each other, in some samples even by an order of magnitude. This is because fluorescence is more sensitive than UV–vis. Therefore, we considered the difference in the concentrations recorded with both methods as sample quality criterion (for review see Gumińska et al., 2018). In this term, ENV(E) produced the most consistent average values, followed by SOI(E) and WAT(O). The greatest discrepancy was observed for samples isolated without using the commercial kits.

**Gel electrophoresis**

Agarose gel electrophoresis revealed that SOI(E) yielded higher quantities of more intact DNA compared to the other tested procedures, regardless of the method of biomass harvesting (Supplementary Fig. S2A). Lower amounts and slight degradation of DNA were observed in the case of CTAB-based protocols. Furthermore, samples extracted with CHEL resin did not allow to obtain the
the most suitable PCR template, whereas the amplification results for the kits dedicated to DNA isolated from scraped and centrifuged samples turned out to be the other spin-column methods. Of the three methods of biomass collection, DNA isolated from centrifuged and scraped samples was the cheapest among the eight extraction methods. The fastest of the assessed extraction methods was CHEL, followed by the two commercial kits: SOI(O) and WAT(O), centrifuged then isolated with DNE(Q), scraped and then extracted with SOI(O), WAT(O) or CTAB + D. The fastest of the assessed extraction methods was CHEL, followed by the two commercial kits: SOI(O) and WAT(O), centrifuged then isolated with DNE(Q), scraped and then extracted with SOI(O), WAT(O) or CTAB + D.

**Time, workload and cost analysis**

The fastest of the assessed extraction methods was CHEL, followed by the two commercial kits: SOI(E) and DNE(Q), while the most time-consuming was the CTAB + D protocol (Table 3). The extraction with CHEL was also the least expensive among the eight extraction methods, while DNE(Q) was the most expensive protocol. WAT(O) turned out to be the cheapest among commercially available kits, although not as easy to handle as the other spin-column methods.

**Amplification performance**

Of the three methods of biomass collection, DNA isolated from scraped and centrifuged samples turned out to be the most suitable PCR template, whereas the amplification of DNA isolates from the filtered samples was clearly hampered. For the latter, amplification of the undiluted template obtained with DNE(Q), CTAB and CHEL failed (Supplementary Fig. S2B). Of all DNA extraction protocols assessed, commercial methods based on silica-column adsorption of DNA dedicated for environmental samples provided the best amplifiable templates. The amplification results for the kits dedicated to

### Table 1. Comparison of three biomass capture methods using two spectrophotometer measurements (260/280 and 260/230 nm) and fluorometric measurement.

| Capture method | Implen average concentration (ng \(\mu\)l\(^{-1}\)) | Implen average concentration SD (ng \(\mu\)l\(^{-1}\)) | Average A260/280 | Average A260/230 | Qubit average concentration (ng \(\mu\)l\(^{-1}\)) | Qubit concentration SD (ng \(\mu\)l\(^{-1}\)) | Implen - Qubit difference |
|----------------|---------------------------------|---------------------------------|-----------------|-----------------|---------------------------------|---------------------------------|-----------------------------|
| Centrifuged    | 47.81                           | 44.96                           | 1.76            | 0.22            | 0.8                             | 0.45                            | 12.5                        |
| Filtered       | 26.6                            | 27                              | 1.61            | 0.43            | 0.42                            | 0.33                            | 5.82                        |
| Scrapped       | 9.28                            | 4.79                            | 1.73            | 0.18            | 0.78                            | 0.51                            | 3.79                        |

**Table 2. Comparison of eight DNA extraction methods using two spectrophotometer measurements (260/280 and 260/230 nm) and fluorometric measurement.**

| Extraction method | Implen average concentration (ng \(\mu\)l\(^{-1}\)) | Implen average concentration SD (ng \(\mu\)l\(^{-1}\)) | Average A260/280 | Average A260/230 | Qubit average concentration (ng \(\mu\)l\(^{-1}\)) | Qubit concentration SD (ng \(\mu\)l\(^{-1}\)) | Implen - Qubit difference |
|-------------------|---------------------------------|---------------------------------|-----------------|-----------------|---------------------------------|---------------------------------|-----------------------------|
| DNE(O)            | 14.82                           | 15.60                           | 1.64            | 0.23            | 0.44                            | 0.19                            | 2.51                        |
| ENV(E)            | 10.18                           | 1.84                            | 1.89            | 0.04            | 1.07                            | 0.32                            | 8.62                        |
| SOI(E)            | 20.93                           | 7.56                            | 1.77            | 0.16            | 1.07                            | 0.25                            | 17.77                       |
| SOI(O)            | 6.36                            | 3.96                            | 1.94            | 0.18            | 0.33                            | 0.12                            | 2.91                        |
| WAT(O)            | 6.07                            | 4.47                            | 1.93            | 0.31            | 0.16                            | 0.14                            | 2.71                        |
| CTAB + D          | 54.33                           | 60.29                           | 1.45            | 0.24            | 0.82                            | 0.49                            | 14.58                       |
| CTAB              | 64.38                           | 52.91                           | 1.50            | 0.47            | 0.97                            | 0.77                            | 6.70                        |
| CHEL              | 46.11                           | 28.79                           | 1.47            | 0.06            | 0.51                            | 0.08                            | 3.17                        |

Symbols: DNE(Q) - DNeasy Blood & Tissue (Qiagen); ENV(E) - GeneMATRIX Environmental DNA & RNA Purification Kit (Eurx); SOI(E) - GeneMATRIX Soil DNA Purification Kit (Eurx); SOI(O) - E.Z.N.A. Soil DNA Kit (Omega Bio-tek); WAT(O) - E.Z.N.A. Water DNA Kit (Omega Bio-tek); CTAB + D - CTAB with DMSO; CTAB - CTAB without DMSO; CHEL - Chelex.

**Table 3. Evaluation of cost per sample and duration of eight DNA extraction methods.**

| Extraction method | Extraction duration | Cost per sample (€) |
|-------------------|---------------------|---------------------|
| DNE(Q)            | 47 min              | 4.10                |
| ENV(E)            | 2 h 1 min           | 2.32                |
| SOI(E)            | 41 min              | 2.08                |
| SOI(O)            | 54 min              | 3.01                |
| WAT(O)            | 1 h 28 min          | 2.05                |
| CTAB + D          | 3 h 11 min          | 0.26                |
| CTAB              | 2 h 30 min          | 0.23                |
| CHEL              | 37 min              | 0.18                |

Cost (in €) includes only kit components/reagents, not the additional expendables nor the laboratory equipment required to perform the procedure. Symbols: DNE(Q) - DNeasy Blood & Tissue (Qiagen); ENV(E) - GeneMATRIX Environmental DNA & RNA Purification Kit (Eurx); SOI(E) - GeneMATRIX Soil DNA Purification Kit (Eurx); SOI(O) - E.Z.N.A. Soil DNA Kit (Omega Bio-tek); WAT(O) - E.Z.N.A. Water DNA Kit (Omega Bio-tek); CTAB + D - CTAB with DMSO; CTAB - CTAB without DMSO; CHEL - Chelex.
environmental samples were the most consistent and reproducible among all analysed methods, regardless of the method of biomass harvesting (Supplementary Fig. S2B).

**NGS results and quota of euglenids**

The HTS of 18 samples yielded a total number of 938,816 paired-end reads. After the bioinformatic filtering, merging and chimera deletion 710,615 reads (75.7% of the total number) were retained. Filtered reads were assigned to a total of 1018 ASVs. 831 ASVs and 603,938 filtered reads were assigned to Euglenida, their total percentage equaled 85.0% and it varied between samples and their types. In filtered samples, Euglenida constituted 90.0% on average (between 87.6% and 92.2%), in centrifuged 85.9% (73.5%–95.4%) and in scraped samples 80.5% (62.6%–94.6%) (Fig. 2).

In the scraped samples, reads of Alveolates (Dinophyceae, representing mainly Gymnodiniphycidae) were the distinctly dominant non-euglenid reads, as they constituted 18.9% of the total number of reads in that type of samples. Percentage of other non-euglenid reads (including Holozoa and two groups of Stramenopiles: Chrysophyceae and Eustigmatophyceae) was below 0.2% each. In filtered and centrifuged samples, besides euglenids, there were mainly reads representing Stramenopiles, including mostly Chrysophyceae (4.9% and 8.3%) and Eustigmatophyceae (2.7% and 1.8% respectively). In both types of samples there are also reads of Dinophyceae (0.5% and 2.2%) and Holozoa, Ciliophora, Centrohelida and Chlorophyta (below 1% each) (Fig. 2; Supplementary Table S3).

All 603,938 reads belonging to Euglenida represented autotrophic euglenids (Euglenophyceae) and 99.8% of them were identified at least at the genus level. Among them 88.1% were assigned to species. It equalled 88.3%
in centrifuged samples, 84.9% in filtered samples and 91.9% of reads from scraped samples (Fig. 2).

**Taxonomic composition of autotrophic euglenids**

Obtained autotrophic euglenid reads were classified into 10 genera and 50 species. Among them there were species with various numbers of reads: 11 abundant species (with more than 1% of reads), 17 moderately abundant species (between 0.1% and 1%) and 22 rare species (with fraction below 0.1%), including 14 species with less than 50 reads (Supplementary Table S4).

Reads of *Euglena archaeoplastidiata*, *Euglena splendens* and *Euglenaria clavata* were the most abundant ones with a fraction of reads in the whole analysis equaling 29.8%, 21.0% and 11.2% respectively. In total they constituted 62.0% of all autotrophic euglenid reads. The following most abundant reads fractions were *Colacium* sp. and *Trachelomonas lefevrei*, each with over 5% of reads.

The reads belonging to genus *Euglena* constituted over half of all obtained euglenid reads (53%) and they were assigned to 11 species, including the two most abundant species in the analysis (*E. archaeoplastidiata* and *E. splendens*). The following highly numerous groups are formed by reads classified to genus *Trachelomonas* (13%), with five species, and many reads unassigned to species (17.2% of all *Trachelomonas* reads), and *Euglenaria* (11%) including Eu. clavata (the third most abundant species). We also detected members of the following genera: *Phacus* (8% of all reads, 13 species), *Colacium* (6% of all reads, over 98.7% reads unassigned to species), *Lepocinclis* (5% of all reads, seven species), *Discoplastis* (2% of all reads), *Monomorphina* (1%), *Strombomonas* (below 1% of reads, one species) and four reads assigned to *Cryptoglena*.

**Comparison of results of three types of samples**

Species composition in the three types of samples was similar but far from identical (Fig. 3; Supplementary Table S4). In filtered samples, reads were assigned to 45 species, in centrifuged and scraped samples to 41 and 39 species respectively. Most of the observed species were present in all types of samples. However, six of them were observed exclusively in filtered samples, two only in centrifuged and three only in scraped samples (Supplementary Table S4). There were also reads of three species observed in filtered and centrifuged samples but not scraped samples. The fraction of reads of the species that were absent in some types of samples did not exceed 0.02%. The observed differences in species composition between types of samples are statistically significant in most analyses, based on the Jaccard index distance matrix (Table S5). Pairwise comparisons of species composition in centrifuged versus scraped samples and filtered versus scraped samples indicated significant differences in all analyses with all ASVs or only ASVs with a fraction of at least 0.1% (datasets: 1a - all ASVs with at least 0.1% of all reads; 1b - all ASVs with at least 10 reads; 2a - ASV assigned to euglenids constituting at least 0.1% of all reads; 2b - all ASVs assigned to euglenids with at least 10 reads) and differences between filtered and centrifuged samples - based on datasets 1b and 2a were also highly supported (analysis of similarities (ANOSIM); *p* < 0.004, permutational analysis of variance (PERMANOVA): *p* < 0.003). Differences between filtered and centrifuged samples based on datasets 1a (all euglenids ASVs) and 2b (all ASVs with fraction of at least 0.1%) had lower confidence (ANOSIM: *p* < 0.031, *p* < 0.021; PERMANOVA: *p* < 0.017, *p* < 0.030 respectively). Analysis based on Jaccard distance matrix calculated for dataset 3 (presence of euglenid species with at least 0.1% reads based on taxonomic assignment) gave different results and showed no statistically significant differences between centrifuged–scraped and filtered–scraped types of samples (ANOSIM: *R* = 0.17, *p* < 0.062, *R* = 0.17, *p* < 0.062; PERMANOVA: *p* < 0.063, *p* < 0.059 respectively) and no differences at all between centrifuged and filtered samples (Table S5).

Fractions of reads belonging to each species also vary between types of samples and the differences between them based on Bray–Curtis distance matrix are statistically significant in each analysis (ANOSIM: *p* < 0.006, PERMANOVA: *p* < 0.005). The pseudo-F value in PERMANOVA and R-statistic in ANOSIM in all analyses is the lowest for the comparison of filtered and centrifuged samples, indicating higher similarity between those two types of samples (Table S5). In filtered samples reads of *E. archaeoplastidiata* and *E. splendens* were dominant with very similar numbers of reads, 32.1% and 29.9% respectively. The third most abundant fraction was reads classified as *Phacus* sp. (7.4%). In centrifuged samples, *E. archaeoplastidiata* reads were clearly dominant (35.2%), followed by *E. splendens* (19.1%), *Colacium* sp. (10.2%) and *Eu. clavata* (7.3%) reads. The most numerous reads in scraped samples were *Eu. clavata* (22.0%), *E. archaeoplastidiata* (21.5%), *E. splendens* (16.7%) and *Trachelomonas lefevrei* (8.8%). Percentage of reads of any other ASVs did not exceed 4.5% in any sample.

**Comparison of microscopic observations and DNA metabarcoding results**

Microscopic observations resulted in the identification of 22 species of autotrophic euglenids representing seven
genera (Supplementary Table S4 - marked with bold type). Representatives of *E. archaeoplastidiata*, *E. splendens*, *Euglenaria clavata*, *Lepocinclis caudata* and *Phacus salinus* were the most abundant species (marked with ++++). The cells of *Lepocinclis gilboa*, *Lepocinclis oxyuris*, *Phacus anacoelus*, *Trachelomonas lefevrei* and *Trachelomonas volvocina* were less numerous (+++). In the next category (+++) there were *Monomorpha pyrum*, *Phacus hamatus*, *Phacus mangini*, *Trachelomonas volvocina* and *Trachelomonas echinata*. Cells of *Euglena ehrenbergii*, *Discoplastis spathirhyncha*, *Discoplastis excavata* and *Phacus paraorchicularis* occurred rarely (+) and cells of *Euglena deses*, *Euglena cantabrica* and *Lepocinclis acus* - occasionally (marked with s). Beside autotrophic euglenids, high numbers of cells of *Synura sp.* (+++) and dinoflagellates (++++) were observed in the sample.

Twenty of 22 observed euglenid species were detected also by the metabarcoding method in all samples except for four of the scraped samples in which reads of 18 or 19 such species were noticed (Fig. 4). Reads of all species marked with ++++, +++ and ++ were noticed in each sample. Two species were seen exclusively under the microscope: *E. ehrenbergii* (+; the species lacking a representative in the reference database) and *E. cantabrica* (s). The following two species (*P. paraorchicularis* (+) and *E. deses* (s)) were observed under the microscope and their reads were present in all but three scraped samples. The rest of the microscopically detected species were identified by the NGS method as well. Total fraction of reads belonging to species marked as the most abundant in microscopic observations (++++) vary between 55% and 70% among samples and they constitute 64% of all euglenid reads (Fig. 5). The reads of such species comprised various percentages of all reads: *E. archaeoplastidiata* 29.8%, *E. splendens* 21.0%, *Eu. clavata* 11.2%, *L. caudata* 1.7%, *P. salinus* 0.5% and it varied between sample types. Total percentage of reads of taxa that were not observed using the microscope did not exceed 5.5% in any sample.

![Figure 3](wileyonlinelibrary.com)
and were the least abundant in scraped samples (on average 1.5%) (Fig. 3, Supplementary Table S4).

There were 30 species reported exclusively by the NGS method. In filtered samples, 25 such species were noticed, in centrifuged samples 21 and 19 in scrapes. There were also three whole genera that were not observed using the microscope: *Colacium* (5.7% of reads, mainly *Colacium* sp.), *Strombomonas* (about 0.4% reads belonging to *Strombomonas acuminata*) and *Cryptoglena* (three reads of *Cryptoglena pigra*). Only seven unseen species constituted above 0.1% of the total number of reads each: *Euglena laciniata* (0.7%), *Euglena pseudostellata* (0.6%), *Euglena agilis* (0.5%), *Strombomonas acuminata* (0.4%), *Monomorphina piriformis* (0.2%), *Phacus orbicularis* (0.2%) and *Euglena pseudoviridis* (0.2%). All of them were present in all samples but one scraped sample lacking *Euglena pseudoviridis*. The remaining eight species (with fraction of reads between 0.01% and 0.1% of the total number of reads) were three species of *Colacium*, *Discoplasia constricta*, *Euglena geniculata*, *Phacus caudatus*, *Phacus granum* and *Trachelomonas magdaleniana* that were noticed in at least one sample of each type.

**Discussion**

Towards the optimized protocol of capture and extraction of environmental DNA for detection and quantification of green euglenids

A growing body of research highlights the importance of protocol optimization and the urge for standard methodological approach to ensure comparability between metabarcoding surveys (Dickie et al., 2018; Harper et al., 2019; Jeunen et al., 2019). The effect of the capture and DNA extraction methods on the detected biodiversity has been addressed by other researchers using various types of samples (e.g. faecal, clinical, soil and
Our results confirm that protocol choice significantly affects the DNA yield from freshwater samples. Maximizing DNA recovery and quality is crucial in research focused on poorly studied organisms, such as euglenids, therefore we evaluated three capture methods and eight DNA extraction protocols.

In the first experimental module, we aimed to select the most suitable method for DNA isolation. For this purpose, we tested eight methods (Supplementary Fig. S6). Among the commercial kits, SOI(E) outperformed the other methods in terms of effectiveness (DNA concentration, integrity and suitability for downstream applications), workload and time consumption (Tables 3 and 4). This kit produced the highest DNA concentrations measured both spectroscopically and fluorometrically of all tested spin-column methods, although according to spectrophotometric data it was less efficient than both CTAB-based protocols and CHEL. Also, the values of the A260/280 and A260/230 ratios provided by SOI(E) were the closest to the desired outcome (1.8). However, none of the methods allowed to obtain a sample with ideal parameters. This probably is a result of the specificity of the analysed material. Environmental samples collected from hypereutrophic reservoirs contain a lot of organic matter (e.g. humic acids, metabolites), which affects the spectrophotometric measurement. The inconsistency between the results of spectrophotometry and fluorimetry in all samples can be explained by the fact that the fluorimeter allowed the specific measurement of DNA, unaffected by coprecipitating compounds.

The results of the spectrophotometric and fluorometric analyzes were inconsistent with the results of the DNA electrophoresis (Tables 1 and 2; Fig. 1, Supplementary Fig. S2). In the case of the spectrophotometer, the differences can be explained by overestimation of the DNA concentration in manually obtained samples (both CTAB-based protocols and CHEL) due to the presence of...
contaminations. In turn, in the case of the fluorimeter, the most likely explanation is the fragmentation and degradation of the isolated DNA (the fluorophore binds to fragmented DNA less effectively).

The undoubted advantage of spin-column methods is the lack of tediousness caused by many rounds of incubation and precipitation during processing (Table 3). According to our observations, PCR templates extracted by manual methods (both CTAB-based protocols and CHEL) may amplify less well than those obtained otherwise. This can be explained by an inhibitory effect of humic acids derived from the dead biomass and/or substances carried by algae (e.g. polysaccharides and phenols) which may be co-extracted and thereafter obstruct the PCR (Schrader et al., 2012). CTAB isolation protocols have been developed for laboratory cultures that are free from decomposition products, wastewater, agricultural fertilizers, and so on, typically found in environmental samples (Gumińska et al., 2018). Therefore, the isolates produced with these methods require cleanup prior to further processing. Extraction with CHEL, while rapid, also provides samples that should be diluted and/or purified before downstream applications. Interestingly, same as Eichmiller et al. (2016), we also noticed a decrease in amplification efficiency of samples processed with DNE(Q) (Supplementary Fig. S2B).

Considering all criteria of purity, quality and usefulness of the assessed methods in high throughput analyze, we selected SOI(E) as the most suitable DNA extraction method. We also conducted a pilot metabarcoding study of green euglenids (Supplementary Figs S7 and S8).

**DNA metabarcoding performance**

Our attempt at DNA metabarcoding of autotrophic euglenids was successful regardless of the method of biomass harvesting. The majority of obtained reads was assigned to Euglenida, almost all of them were identified at least at genus level and almost 90% were assigned to species. Our study is the first metabarcoding survey that resulted in finding so many euglenid taxa. Hitherto, members of that group were overlooked in many environmental surveys both in seas (Salonen et al., 2018; Gran-Stadniczeñko et al., 2019) and various freshwater environments (Banerji et al., 2018; Boenigk et al., 2018) or their reads constituted very low fractions (Simon et al., 2016; Sawaya et al., 2019; Preston et al., 2020; Wolf and Vis, 2020). The reason for that could be the primer bias that was reported as one of the main problems concerning DNA metabarcoding (Pawlowski et al., 2018; Sawaya et al., 2019). 18S rDNA sequences of euglenids are known to be poorly amplified by universal eukaryotic primers targeting, e.g. V4 and V9 region (Banerji et al., 2018; Gran-Stadniczeñko et al., 2019) and in silico analysis confirmed poor matching of many metabarcoding primers for Excavata (Geisen et al., 2019). Using group-specific primers, designed for autotrophic euglenids, targeting the V2 18S rDNA fragment in this study allowed us to avoid the above-mentioned problems with PCR primer bias and barcode sequence length.

Almost 90% of obtained euglenid reads were assigned to species. The factors impeding species assignment of the rest of the reads, belonging mainly to three genera: *Colacium*, *Trachelomonas* and *Phacus*, could have been various. One of them is probably low genetic variability of 18S rDNA sequences in some parts of the phylogenetic tree of autotrophic euglenids, for example in the *Phacus oscillans* clade (Karnkowska-Ishikawa et al., 2010; Kim and Shin, 2014) or the *Phacus anomalous* clade (Łukomska-Kowalczyk et al., 2020b), where most of the reads assigned as *Phacus* sp. are placed. Furthermore, in the whole genus *Colacium*, ranges of its interspecific and intraspecific variability significantly overlap (Łukomska-Kowalczyk et al., 2016).

On the other hand, an incomplete database and errors in the taxonomic assignment of some reference sequences are considered to be the main issues that can hamper metabarcoding studies (Pawlowski et al., 2018; Sawaya et al., 2019; Fan et al., 2020). Euglenids are one of the taxa with a limited number of reference sequences, thus their molecular identification is strongly affected by this problem. It could explain the high number of unclassified reads of *Trachelomonas*, the very big and poorly studied genus with 375 taxonomically accepted species (Algaebase) and sequences of only 23 species in the database. Reads labelled as *Lepocinclis* sp., *Euglena* sp. and Euglenea could also represent new phylogenetic lines (species or strains of unknown sequences of 18S rDNA). Despite enriching the database with a thoroughly curated euglenid set of sequences, it still lacks sequences of many species and requires increasing of their number to improve the DNA barcoding performance of that group in the future.

**NGS-based identification versus microscopic analyses**

In all analysed samples autotrophic euglenid species richness based on the DNA metabarcoding method was similar and almost all (18–20) of 22 species observed under the microscope were noticed. One of the HTS unnoticed but observed species was *E. ehrenbergii* - the species without 18S rDNA sequences in the reference database despite many attempts to obtain it (unpublished data). The other three species unnoticed by the NGS method occurred in very low numbers. This is consistent with other studies that also overlooked less abundant species without 18S rDNA sequences in the reference database despite many attempts to obtain it (unpublished data). The other three species unnoticed by the NGS method occurred in very low numbers. This is consistent with other studies that also overlooked less abundant species without 18S rDNA sequences in the reference database despite many attempts to obtain it (unpublished data).
species in parts of replicates in metabarcoding surveys. This is interpreted as the result of the heterogeneity within the DNA extracts (Beentjes et al., 2019).

On the other hand, we detected 30 species that were not observed using the microscope, which comprises more than half of all (52) detected species. However, the fraction of their reads did not exceed 6% and most of them (23) were rare species (<0.1%). In most studies comparing lists of species detected based on microscopic observation and metabarcoding, not all species are detected by both methods. In some surveys, most of the observed species are HTS detected (Bálint et al., 2020) but more often the opposite is true (Steffanni et al., 2018). The rate of detection of observed species can also vary between analysed samples (Fujii et al., 2019) or depend on the taxonomic group (De Vargas et al., 2015; Gran-Stadniczek et al., 2019).

The reasons for the observed discrepancies could be the higher sensitivity of HTS detecting methods and overlooking of some species in microscope analyses. They could have been omitted due to their similarity to cells of other species present in the sample, especially those of very high abundances. In our study, this situation concerns some unobserved species: *E. laciniata* that is morphologically similar to very common *E. splendens* (Karnkowska-Ishikawa et al., 2013); *E. agilis* - to *E. archaeoplastidiata*, *P. caudatus* - to *P. mangini*, *P. orbicularis* - to *P. paraorbicularis* and *M. piriformis*, morphologically indistinguishable from *M. pyrum*. Another factor that can result in inconsistencies between HTS and microscope detection of species can be the temporal loss of their diagnostic features. It can take place in the case of many euglenids: *E. cantabrica* and other species with star-shaped chloroplasts that can lose some of their chloroplasts (Kosmala et al., 2009) or *S. acuminata*, which can lose its characteristic lorica. Members of genus *Colacium* could be overlooked in microscopic observation because of their occurrence in attachment to mesoplankton animals that were omitted during the making of microscopic slides.

Another possible reason for not observing HTS detected species could be their very low abundance. It is possible that species that were never seen in the pond (like, *Euglena gracilis*, *Lepocinclis fominii*, *Phacus granum*, *Phacus mariae*, *Phacus tortus*, *Trachelomonas magdalenianna*) always occur in very low densities as it was reported in another study, where above 93% of planktonic eukaryotic operational taxonomic units (OTUs) occurred at low abundance during the whole survey (<0.05% reads per sample) (Simon et al., 2015). On the other hand, some species that were not observed in our study were present in the water body many times before and after. Some of them always occurred in low densities (Discoplastis constricta, Discoplasts gasterosteus, Monomorphina aenigmatica), others were sometimes very numerous (*Lepocinclis fusiformis*, Phacus pleuronectes*). Such rapid fluctuations in abundances and shifting of proportions above 100 times are documented for some OTUs of planktonic eukaryotic microorganisms in consecutive months and weeks (Simon et al., 2015; Beentjes et al., 2019). In such a situation, we can detect so-called ‘ghost’ MOTUs, corresponding to the taxa represented by extracellular DNA only (Pawlowski et al., 2018) or the species occurring in abundances too low to be detected by microscopic observation due to the possible higher resolution of HTS identification. It is also possible that the presence of species especially with the smallest number of reads is an artefact resulting from sequencing errors.

### Comparison of material gathering methods

According to numerous reports, filtration is currently the predominant capture technique in aquatic metabarcoding analyses since it surpasses other methods, such as centrifugation and precipitation (Deiner et al., 2015; Jeunen et al., 2019). Contrary to this, our findings demonstrate that centrifugation outperforms both filtering and scraping in terms of the amount and quality of obtained material (Table 1; Fig. 1, Supplementary Fig. S2). However, in terms of the taxa detected, the results from centrifuged samples were slightly inferior to the filtered ones. Scraping turned out to be the least reliable in terms of biodiversity assessment of euglenids. Furthermore, scraped samples contained the least sequencing reads assigned to the least numerous species observed under the microscope.

In our study all replicates of the same biomass harvesting method have very similar species composition, although some species are not equally represented. The most pronounced differences between types of samples are based on fractions of reads of various species (Bray–Curtis distances). No straightforward relation between the number of reads and observed abundance of cells was detected, thus it is difficult to assess which method gave us the most reliable results regarding that issue. Such a situation is often mentioned and relative abundance of reads does not necessarily directly correspond to the number of specimens in many metabarcoding studies (Carew et al., 2013; Stoeck et al., 2014). There are statistically significant differences between the taxonomic composition of samples differing in the way of the material gathering in almost all analyses, despite the analysis based on detected species with above 0.1% of reads and presence–absence data (Jaccard index, dataset 3). The most significant
differences were observed between samples with scraped material in comparison to the two other categories.

Some discrepancies between types of samples, especially the relatively far distance between scrapes and other types, can be the result of the size and weight of cells and phototactic activity. The presence of stigma suggests that all euglenids exhibit positive phototaxis. Nevertheless, various species can react differently to the light intensity - therefore not all of them tend to be found in the most illuminated part of the vessel. The size of the stigma may also matter; species of small cell sizes have a stigma relatively bigger than those of large cells. Although detailed studies on this issue have not been conducted so far, our long-term microscopic observations confirm the differences in phototaxy in particular species of euglenids (unpublished data). Differences between centrifuged and filtered samples can be a result of detection of species represented by cells debris or extracellular DNA only (Pawlowski et al., 2018; Doi et al., 2019; Sawaya et al., 2019). Free DNA can be efficiently bound to the filter but not easily centrifuged and thus, using filtration can result in detection not only of species present in the water body at the moment but also those present earlier. Therefore, if the goal is to study the actual species composition, centrifugation could be a convenient way to harvest the biomass.

Conclusions

We propose a highly efficient protocol for HTS based identification of autotrophic euglenids in environmental samples using the V2 18S rDNA fragment, euglenid specific primers and the database enriched with curated sequences. The GeneMATRIX Soil DNA Purification Kit (Eurx) was chosen as the most suitable method of DNA isolation due to the best quality and quantity of obtained DNA as well as the relatively low costs and workload. From the studied biomass harvesting methods, both centrifugation and filtering outperformed scrapes and the choice between them is not crucial for the reliability of the analysis. Therefore any of them can be used depending on the aim of the study, analysed environment and organisms.

Materials and methods

Study site and water sampling

Prior to sampling, the equipment was washed in 10% Trilux (Analab) solution, rinsed in distilled water and UV irradiated. Water samples were taken from the shallow, hypereutrophic mid-field pond located in Masovian District, Poland (52°08'21.00"N, 20°32'03.2"E) in June 2018. Two methods of water collection were used. First, approximately 20 L of water from the sunlit surface layer was concentrated with a plankton net with a mesh size of 10 μm to the final volume of 2 L and placed in two sterile vessels. These samples were used to obtain biomass by centrifugation and sedimentation respectively. Second, 1 L of water from the coastal surface layer was collected in a sterile vessel, then used to capture biomass by filtering. The sample intended for filtration was not concentrated to avoid filter clogging due to the turbidity of the water.

Biomass harvesting

Biomass was obtained from the samples using three methods: (i) 100 ml of unconcentrated water taken directly from the pond was vacuum-filtered using 0.2 μm pore size Nucleopore (Whatman) track-etched polycarbonate membrane filters, (ii) 100 ml of water taken with a plankton net was centrifuged for 25 min at 8000g, the supernatant was discarded and the pellet was collected (iii) 1 L of water sample taken with a plankton net was left in a transparent glass container for 24 h on the bench for sedimentation, then the cell pellet accumulated on the sunlit wall of the vessel (cells with positive phototaxis) was scraped off with a sterile spatula, transferred to sterile 1.5 ml tubes and stored at −80°C until use. Before freezing, 100 μl of the scraped cell pellet was taken for microscopic examination. Negative controls with distilled water were also performed.

Microscopic observations

Microscopic studies were performed using Nikon Eclipse E-600 light microscope with differential interference contrast, equipped with the NIS Elements Br 3.1 (Nikon, Japan) measurement program. From the 100 μl of the scraped cell pellet, a few microscopic preparations were made. The species of euglenids were identified and their quantity was assigned to five categories: (++++) the most abundant species, ~10 cells in one field of view under microscopic magnification ×10, (+++) ~5 cells in the field of view, (+) 1–2 individuals in one field of view, (+) single individual in a few fields of view and (s) single cell in one drop across the microscopic slide.

Assessment of DNA purification methods

In the first module of the project, various DNA extraction methods were evaluated for their suitability for DNA isolation from pond samples. Pellets obtained via centrifugation and scraping were suspended in nuclease-free water (GE Healthcare) to the final volume of 800 μl, then divided into aliquots of 100 μl. Filtration membranes with...
immobilized biomass were divided into eight equal parts and placed in separate sterile 1.5 ml tubes. Eight DNA isolation protocols were assessed. All experimental steps (including negative controls) were performed in a sterile laminar flow cabinet. To avoid cross-contamination each extraction was carried out separately.

DNA isolation protocols were assessed based on (i) sample purity (A260/230 and A260/280 absorbance ratios), (ii) DNA concentration, (iii) amplification efficiency (e.g. presence of amplification inhibitors, such as humic acids), (iv) cost of reagents, (v) preparation complexity and time required to process a single sample.

**Extraction with silica-membrane spin-column kits**

Five commercially available DNA isolation kits were compared: DNeasy Blood & Tissue Kit (Qiagen; DNE(Q)), GeneMATRIX Environmental DNA & RNA Purification Kit (Eurx; ENV(E)), GeneMATRIX Soil DNA Purification Kit (Eurx; SOI(E)), E.Z.N.A. Soil DNA Kit (Omega Bio-tek; SOI(O)) and E.Z.N.A. Water DNA Kit (Omega Bio-tek; WAT(O)). The selected methods are commonly used in metabarcoding surveys (Kaden and Krolla-Sidenstein, 2015; Braun and Szewzyk, 2016; Wang et al., 2018; Gołębiewski and Tretyn, 2019; Jeunen et al., 2019). All isolation steps were performed according to the manufacturer’s instructions. Finally, the DNA was eluted or suspended in 100 μl of nuclease-free water (GE Healthcare).

**Extraction with CTAB-based protocols**

The DNA isolation with CTAB (AppliChem) was performed in two variants (with/without dimethyl sulfoxide; DMSO; AppliChem; CTAB+D and CTAB respectively) as described elsewhere (Healey et al., 2014; Gumińska et al., 2018). The volumes of reagents were downscaled according to the amount of the harvested biomass.

**Extraction with chelating resin**

In previous reports, Chelex (CHEL) was used for DNA extraction from various organisms, including algae (Hamilton et al., 2015; Lienhard and Schäffer, 2019). Herein, a 10% (wt./vol.) stock solution of Chelex 100 (Bio-Rad) in nuclease-free water (GE Healthcare) was prepared. To distribute the beads evenly, the solution was vortexed for 3 min and pipetted several times before use. For each sample 100 μl of suspended resin was added, then thoroughly vortexed for 5 min. Samples were incubated at 95°C for 30 min with interval mix (800 rpm: 30 s, every 2 min), vortexed for 30 s and centrifuged for 1 min at 12 000g. The supernatant containing DNA was transferred to a separate, sterile tube.

**Assessment of DNA integrity, purity and yield**

The quality of the DNA isolates was evaluated spectrophotometrically with NanoPhotometer NP80 (Implen) and fluorometrically with Qubit 3.0 (Thermo Scientific). DNA integrity was assessed by standard agarose gel electrophoresis with 5 μl of each sample: 1.5% agarose gel in 1× TAE buffer, staining with 0.5% Midori Green (Nippon).

**Final DNA purification**

In the second module of the project, the method that outperformed the others, SOI(E), was used to extract the DNA captured with filtering, centrifugation and scraping. The biomass obtained with each capture method was divided into eight equal parts as previously described. DNA isolates were quality checked and amplified (Supplementary Fig. S7).

**Amplification**

A 1 μl of the DNA isolates or nuclease-free water (negative controls) were used as templates in PCR reactions. Amplification was performed with primers universal for autotrophic euglenids. The primers encompassing the V2 18S rDNA region were newly designed for this study. The PCR efficiency was tested with primers 18S_V2i_F1 and 18S_V2i_R1, while library amplification was performed with the remaining pairs (Supplementary Table S9). A 25 μl reaction mixture contained 0.2 mM dNTPs, 10 pmol of forward and reverse primers, appropriate 1× reaction buffer and 0.5 U of Phusion (Thermo Scientific) polymerase. The amplification protocol consisted of 2 min at 98°C, followed by 35 cycles comprising 20 s at 98°C, 20 s at 57°C and 30 s at 72°C. The final extension step was performed for 5 min at 72°C. Each sample was amplified in triplicates. The amplification efficiency was assessed based on the 1.5% agarose gel electrophoresis (Supplementary Fig. S7).

**Library preparation and sequencing**

Six best-amplified samples for each of three capture methods were selected for the next step (Supplementary Fig. S7). Corresponding PCR triplicates were pooled and purified from the reaction mixture using GeneMATRIX PCR/DNA Clean-Up Purification Kit (Eurx) following the manufacturer’s instruction. The concentration of amplified DNA fragments was assessed spectrophotometrically. All 18 amplicons were pooled at equal concentration to receive a multiplexed sequencing library. The
amplicons were sequenced by the synthesis in an external facility (Genomed, Warsaw, Poland) with MiSeq (Illumina) platform on a single run with the 1/10 capacity of the flow cell using the v2 chemistry (2 × 250 bp).

**Bioinformatic data processing**

Raw Fastq files were demultiplexed and preprocessed (trimmed for the presence of Illumina adapter sequences) by the sequencing company (Genomed). The reads were then quality checked with the FASTQC program (Andrews, 2010). Subsequently, reads were analysed using the QIIME 2.0 (version 2019.1) software pipeline (Bolyen et al., 2019). Using DADA2 (Callahan et al., 2016) within QIIME 2.0, demultiplexed reads were trimmed by removing 50 nt from the forward and reverse reads (containing primer sequences), denoised, quality filtered (truncated at the first instance of a quality score less than 10), pair-ends were joined, chimeras were removed with consensus method. Afterwards, the obtained sequences were dereplicated to produce ASVs.

Taxonomy was assigned to each ASV using the database constructed based on SILVA_132_rep_set_all_99 database (Quast et al., 2013). The sequences of autotrophic euglenids (taxa Phacaceae, Euglenaceae and Eutriptaleae) from the database were replaced with 466 18S rDNA sequences with a verified species assignment, representing 154 species belonging to all 14 genera of autotrophic euglenids (34 sequences unpublished). Prior to assigning taxonomy, the database sequences were trimmed using the primers used in this study. A machine learning naive Bayes classifier was trained and used to assign taxonomy (confidence = 0.7, classifier: QIIME2). The list of species obtained from taxonomic assignments of ASVs was compared with the list of species observed under the microscope.

Reference V2 18S rDNA sequences of autotrophic euglenids has been deposited with COEN/RepOD Repository for Open Science (Interdisciplinary Centre for Mathematical and Computational Modelling, Warsaw, Poland) as a dataset resource with DOI: https://doi.org/10.18150/9MKYCN. Raw reads are deposited in SRA database (NCBI) with accession number: PRJNA686234.

The differences of taxonomic composition between types of samples (filtered, centrifuged and scraped) were statistically tested by permutational analysis of similarities (ANOSIM) and permutational analysis of variance (PERMANOVA), based on the Jaccard similarity index and Bray–Curtis dissimilarity index using QIIME2 platform and Past 4 program (Hammer et al., 2001). Both analyses were conducted based on five various subsets of data: (1a) all ASVs with at least 0.1% of all reads across the whole dataset; (1b) all ASVs with at least 10 reads; (2a) ASVs assigned to euglenids constituting at least 0.1% of all reads; (2b) all ASVs assigned to euglenids with at least 10 reads; (3) species of euglenids based on taxonomic assignment with at least 0.1% reads.

**Acknowledgements**

This work was supported by the OPUS 2016/23/B/ NZ28/00919 grant from the National Science Centre, Poland.

We are grateful for the input of two anonymous reviewers, who has been a great help in improving the manuscript. The Open access for this publication was partially funded by the University of Warsaw.

**References**

Andrews, S. (2010) FastQC: a quality control tool for high throughput sequence data. URL http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Balint, M., Nowak, C., Marót, O., Pauls, S.U., Wittwer, C., Luis, J., et al. (2020) Accuracy, limitations and cost efficiency of eDNA - based community survey in tropical frogs. *Mol Ecol Resour* 18: 1415–1426. https://doi.org/10.1111/1755-0998.12934.

Banerji, A., Bagley, M., Elk, M., Pilgrim, E., Martinson, J., and Santo Domingo, J. (2018) Spatial and temporal dynamics of a freshwater eukaryotic plankton community revealed via 18S rRNA gene metabarcoding. *Hydrobiologia* 818: 71–86. https://doi.org/10.1007/s10750-018-3593-0.

Beentjes, K.K., Speksnijder, A.G.C.L., and Schilthuizen, M. (2019) The effects of spatial and temporal replicate sampling on eDNA metabarcoding. *PeerJ Life Environ* 7: 1–18. https://doi.org/10.7717/peerj.7335.

Bicudo, C.E.D.M., and Menezes, M. (2016) Phylogeny and classification of euglenophyceae: a brief review. *Front Ecol Evol* 4: 17. https://doi.org/10.3389/fevo.2016.00017.

Binladen, J., Gilbert, M.T.P., Bolllback, J.P., Panitz, F., Bendixen, C., Nielsen, R., and Willerslev, E. (2007) The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS One* 2: e197. https://doi.org/10.1371/journal.pone.000197.

Boenigk, J., Wodniok, S., Bock, C., Beisser, D., Hempel, C., Grossmann, L., et al. (2018) Geographic distance and mountain ranges structure freshwater protist communities on a European scale. *Metabarcoding Metagenomics* 2: e21519. https://doi.org/10.3879/mbmg.2.21519.

Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., et al. (2019) Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 37: 852–857. https://doi.org/10.1038/s41587-019-0209-9.

Braun, B., and Szewzyk, U. (2016) Complete genome sequence of “Candidatus Viadribacter manganicus” isolated from a German floodplain area. *Genome Announc* 4: e00897-16. https://doi.org/10.1128/genomeA.00897-16.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016) DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods* 13: 581–583. https://doi.org/10.1038/nmeth.3869.
Carew, M.E., Pettigrove, V.J., Metzeling, L., and Hoffmann, A.A. (2013) Environmental monitoring using next generation sequencing: rapid identification of macroinvertebrate biodicator species. *Front Zool* **10**: 45. https://doi.org/10.1186/1742-9994-10-45.

De Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., et al. (2015) Eukaryotic plankton diversity in the sunlit ocean. *Science* **348**: 1–12. https://doi.org/10.1126/science.1261605.

Deiner, K., Walser, J.-C., Mächler, E., and Altermatt, F. (2015) Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. *Biol Conserv* **183**: 53–63. https://doi.org/10.1016/j.biocon.2014.11.018.

Dickie, I.A., Boyer, S., Buckley, H.L., Duncan, R.P., Dickie, I.A., Boyer, S., Buckley, H.L., Duncan, R.P., et al. (2018) Towards robust and repeatable sampling methods in eDNA-based studies. *Mol Ecol Resour* **18**: 940–952. https://doi.org/10.1111/1755-0998.12907.

Dif, H., Fukaya, K., Oka, S., Sato, K., Kondo, M., and Miyah, M. (2019) Evaluation of detection probabilities at the water-filtering and initial PCR steps in environmental DNA metabarcoding using a multispecies site occupancy model. *Sci Rep* **9**: 3581. https://doi.org/10.1038/s41598-019-40233-1.

Eichmiller, J.J., Miller, L.M., and Sorensen, P.W. (2016) Optimizing techniques to capture and extract environmental DNA for detection and quantification of fish. *Mol Ecol Resour* **16**: 56–68. https://doi.org/10.1111/1755-0998.12421.

Fan, J., Wang, S., Li, H., Yan, Z., Zhang, Y., Zheng, X., and Wang, P. (2020) Modeling the ecological status response of rivers to multiple stressors using machine learning: a comparison of environmental DNA metabarcoding and morphological data. *Water Res* **183**: 116004. https://doi.org/10.1016/j.watres.2020.116004.

Fujii, K., Doi, H., Matsuoka, S., Nagano, M., Sato, H., and Yamanaka, H. (2019) Environmental DNA metabarcoding for fish community analysis in backwater lakes: a comparison of capture methods. *PLos One* **14**: 1–17. https://doi.org/10.1371/journal.pone.0210357.

Geisen, S., Vaulot, D., Mahé, F., Lara, E., and de Vargas, C. (2019) A user guide to environmental protistology: primers, metabarcoding, sequencing, and analyses. *BioRxiv* **850610**: 1–34. https://doi.org/10.1101/850610.

Goldberg, C.S., Turner, C.R., Deiner, K., Klymus, K.E., Thomsen, P.F., Murphy, M.A., et al. (2016) Critical considerations for the application of environmental DNA methods to detect aquatic species. *Methods Ecol Evol* **7**: 1299–1307. https://doi.org/10.1111/2041-210X.12595.

Gołębiewski, M., and Tretyn, A. (2019) Generating amplicon reads for microbial community assessment with next-generation sequencing. *J Appl Microbiol* **128**: 330–354. https://doi.org/10.1111/jam.14380.

Gran-Ślądniczerko, S., Egge, E., Hostyeva, V., Logares, R., Elkem, W., and Edwardsen, B. (2019) Proist diversity and seasonal dynamics in Skagerrak plankton communities as revealed by metabarcoding and microscopy. *J Eukaryotic Microbiol* **66**: 494–513. https://doi.org/10.1111/jeu.12700.

Gumińska, N., Plechá, M., Walkiewicz, H., Haloń, P., Zakryś, B., and Milanowski, R. (2018) Culture purification and DNA extraction procedures suitable for next-generation sequencing of euglenids. *J Appl Phycol* **30**: 3541–3549. https://doi.org/10.1007/s10811-018-1496-0.

Hamilton, P.B., Lefebvre, K.E., and Bull, R.D. (2015) Single cell PCR amplification of diatoms using fresh and preserved samples. *Front Microbiol* **6**: 1–16. https://doi.org/10.3389/fmicb.2015.01084.

Hammer, O., Harper, D.A.T., and Ryan, P.D. (2001) PAST: paleontological statistics software package for education and data analysis. *Palaeontol Electron* **4**: 9.

Harper, L.R., Buxton, A.S., Rees, H.C., Bruce, K., Brys, R., Halfmaerten, D., et al. (2019) Prospects and challenges of environmental DNA (eDNA) monitoring in freshwater ponds. *Hydrobiologia* **826**: 25–41. https://doi.org/10.1007/s10750-018-3750-5.

Healey, A., Furtado, A., Cooper, T., and Henry, R.J. (2014) Protocol: a simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. *Plant Methods* **10**: 1–8. https://doi.org/10.1186/1746-4811-10-21.

Jeuner, G., Knapp, M., Spencer, H.G., Taylor, H.R., Lamare, M.D., Stat, M., et al. (2019) Species-level biodiversity assessment using marine environmental DNA metabarcoding requires protocol optimization and standardization. *Ecol Evol* **9**: 1323–1335. https://doi.org/10.1002/ece3.4843.

Kaden, R., and Kroila-Sidensten, P. (2015) How to show the real microbial biodiversity? A comparison of seven DNA extraction methods for bacterial population analyses in matrices containing highly charged natural nanoparticles. *Microorganisms* **3**: 695–706. https://doi.org/10.3390/microorganisms3040695.

Kamikowska-Ishikawa, A., Milanowski, R., Kwiatkowski, J., and Zakryś, B. (2010) Taxonomy of the Phacus oscillans (Euglenaceae) and its close relatives-balancing morphological and molecular features. *J Phycol* **46**: 172–182. https://doi.org/10.1111/j.1529-8817.2009.00765.x.

Kamikowska-Ishikawa, A., Milanowski, R., Triemer, R.E., and Zakryś, B. (2013) A redescription of morphologically similar species from the genus Euglena: *E. laciniata*, *E. sanguinea*, *E. sociabilis*, and *E. splendens*. *J Phycol* **49**: 616–626. https://doi.org/10.1111/jpy.12072.

Kim, J.I., Linton, E.W., and Shin, W. (2015) Taxon-rich multi-gene phylogeny of the photosynthetic euglenoids (Euglenophyceae). *Front Ecol Evol* **3**: 98. https://doi.org/10.3389/fevo.2015.00098.

Kim, J.I., and Shin, W. (2014) Molecular phylogeny and cryptic diversity of the genus Phacus (Phacaceae, Euglenophyceae) and the descriptions of seven new species. *J Phycol* **50**: 948–959. https://doi.org/10.1111/jpy.12227.

Klymus, K.E., Richter, C.A., Thompson, N., and Hinck, J.E. (2017) Metabarcoding of environmental DNA samples to explore the use of uranium mine containment ponds as a water source for wildlife. *Diversity* **9**: 54. https://doi.org/10.3390/d9040054.

Kosmała, S., Kamikowska-Ishikawa, A., Milanowski, R., Kwiatkowski, J., and Zakryś, B. (2009) Phylogeny and systematics of Euglena (Euglenaceae) species with axial, stellate chloroplasts based on morphological and molecular data-new taxa, emended diagnoses, and
Methods for euglenids metabarcoding 3007

epytifications. J Phycol 45: 464–481. https://doi.org/10.1111/j.1529-8817.2009.00653.x.
Lamb, P.D., Hunter, E., Pinnegar, J.K., Creer, S., Davies, R. G., and Taylor, M.I. (2018) How quantitative is metabarcoding: a meta-analytical approach. Mol Ecol 28: 420–430. https://doi.org/10.1111/mec.14920.
Lienhard, A., and Schäffer, S. (2019) Extracting the invisible: obtaining high quality DNA is a challenging task in small arthropods. PeerJ Life Environ 7: e6753. https://doi.org/10.7717/peerj.6753.
Łukomska-Kowalczyk, M., Chaber, K., Fells, A., Milanowski, R., and Zakryś, B. (2020) Molecular and morphological delimitation of species in the group of Lepocinclis ovum-like taxa (Euglenida). J Phycol 56: 283–299. https://doi.org/10.1111/jpy.12949.
Łukomska-Kowalczyk, M., Fells, A., Chaber, K., Milanowski, R., and Zakryś, B. (2020b) Taxon-rich phylogeny and taxonomy of the genus Phacus (Euglenida) based on morphological and molecular data. J Phycol 56: 1135–1156. https://doi.org/10.1111/jpy.13028.
Łukomska-Kowalczyk, M., Karkowska, A., Krupska, M., Milanowski, R., and Zakryś, B. (2016) DNA barcoding in autotrophic euglenids: evaluation of COI and 18S rDNA. J Phycol 52: 951–960. https://doi.org/10.1111/jpy.12439.
Łukomska-Kowalczyk, M., Karkowska, A., Milanowski, R., Lach, Ł., and Zakryś, B. (2015) Delimiting species in the Phacus longicauda complex (Euglenida) through morphological and molecular analyses. J Phycol 51: 1147–1157. https://doi.org/10.1111/jpy.12352.
McKee, A.M., Spear, S.F., and Pierson, T.W. (2015) The effect of dilution and the use of a post-extraction nucleic acid purification column on the accuracy, precision, and inhibition of environmental DNA samples. Biol Conserv 183: 70–76. https://doi.org/10.1016/j.biocon.2014.11.031.
Pawlowski, J., Esling, P., Lejezerowicz, F., Cedhagen, T., and Wilding, T.A. (2014) Environmental monitoring through protist next-generation sequencing metabarcoding: assessing the impact of fish farming on benthic foraminifera communities. Mol Ecol Resour 14: 1129–1140. https://doi.org/10.1111/1755-0998.12261.
Pawlowski, J., Kelly-Quinn, M., Altermatt, F., Apothéloz-Perret-Gentil, L., Beja, P., Boggero, A., et al. (2018) The future of biotic indices in the ecogeneric era: integrating (e)DNA metabarcoding in biological assessment of aquatic ecosystems. Sci Total Environ 637–638: 1295–1310. https://doi.org/10.1016/j.scitotenv.2018.05.002.
Preston, C.M., Durkin, C.A., and Yamahara, K.M. (2020) DNA metabarcoding reveals organisms contributing to particulate matter flux to abyssal depths in the north East Pacific Ocean. Deep-Sea Res Part II: Trop Stud Oceanogr 173: 104708. https://doi.org/10.1016/j.dsr2.2019.104708.
Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41: 590–596. https://doi.org/10.1093/nar/gks1219.
Salonen, I.S., Chronopoulou, P.M., Leskinen, E., and Koho, K.A. (2018) Metabarocoding successfully tracks temporal changes in eukaryotic communities in coastal sediments. FEMS Microbiol Ecol 95: 1–11. https://doi.org/10.1093/femsec/fiy226.
Saway, N.A., Djuruhas, A., Closek, C.J., Hepner, M., Olesin, E., Visser, L., et al. (2019) Assessing eukaryotic biodiversity in the Florida keys National Marine Sanctuary through environmental DNA metabarcoding. Ecol Evol 9: 1029–1040. https://doi.org/10.1002/ece3.4742.
Schrader, C., Schielke, A., Ellerbroek, L., and Joune, R. (2012) PCR inhibitors - occurrence, properties and removal. J Appl Microbiol 113: 1014–1026. https://doi.org/10.1111/j.1365-2672.2012.05384.x.
Simon, M., Lopez-Garcia, P., Deschamps, P., Moreira, D., Restoux, G., Bertolino, P., and Jardillier, L. (2015) Marked seasonality and high spatial variability of protist communities in shallow freshwater systems. ISME J 9: 1941–1953.
Simon, M., Lopez-Garcia, P., Deschamps, P., Restoux, G., Bertolino, P., Moreira, D., and Jardillier, L. (2016) Resilience of freshwater communities of small microbial eukaryotes undergoing severe drought events. Front Microbiol 7: 812. https://doi.org/10.3389/fmicb.2016.00812.
Stefani, S., Stankovic, D., Borne, D., de Olazabal, A., Juretic, T., Pallavicini, A., and Tirelli, V. (2018) Multi-marker metabarcoding approach to study mesozooplankton at basin scale. Sci Rep 8: 1–13. https://doi.org/10.1038/s41598-018-30157-7.
Stoeck, T., Breiner, H.W., Filker, S., Ostermaier, V., Kammerlander, B., and Sonntag, B. (2014) A morphogenetic survey on ciliate plankton from a mountain lake pinpointed the necessity of lineage-specific barcode markers in microbial ecology. Environ Microbiol 16: 430–444. https://doi.org/10.1111/1462-2920.12194.
Sun, Z., Majaneva, M., Sokolova, E., Rauch, S., Meland, S., and Ekrem, T. (2019) DNA metabarcoding adds valuable information for management of biodiversity in roadside stormwater ponds. Ecol Evol 9: 9712–9722. https://doi.org/10.1002/ece3.5503.
Triemer, R.E., and Zakryś, B. (2015) Photosynthetic euglenoids. In Freshwater algae of North America: Ecology and Classification, pp. 459–483. Cambridge, Massachusetts: Academic Press. https://doi.org/10.1016/B978-0-12-385876-4.00010-4.
Wang, M., Xiong, W., Liu, P., Xie, X., Zeng, J., Sun, Y., and Zeng, Z. (2018) Metagenomic insights into the contribution of phages to antibiotic resistance in water samples related to swine feedlot wastewater treatment. Front Microbiol 9: 2474. https://doi.org/10.3389/fmicb.2018.02474.
Wolf, D.I., and Vis, M.L. (2020) Stream algal biofilm community diversity along an acid mine drainage recovery gradient using multimarker metabarcoding. J Phycol 56: 11–22. https://doi.org/10.1111/jpy.12935.
Zakryś, B., Milanowski, R., and Karkowska, A. (2017) Evolutionary origin of Euglena. Adv Exp Med Biol 979: 3–17. https://doi.org/10.1007/978-3-319-54910-1_1.
Zikza, V.M.A., Elbrecht, V., Macher, J., and Leece, F. (2019) Assessing the influence of sample tagging and library preparation on DNA metabarcoding. Mol Ecol Resour 19: 893–899. https://doi.org/10.1111/1755-0998.13018.

© 2021 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd., Environmental Microbiology, 23, 2992–3008.
Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher’s web-site:

Supplementary Table S1. Detailed comparison of three biomass capture methods and eight DNA extraction methods using two spectrophotometer measurements (260/280 and 260/230 nm) and fluorometric measurement.

Supplementary Figure S2. Agarose gel electrophoresis (1.5% v/w) images of (A) DNA isolated from the samples in the first experimental module, (B) PCR products obtained with primers designed for metabarcoding of green euglenids. In order to compare the performance of assessed DNA capture/extraction methods, the same volume of samples with various concentrations was applied (5 μl per lane; both panels).

Supplementary Table S3. Number of reads belonging to major groups found in individual samples, sums reads in the samples of three types of biomass harvesting (C - centrifugation, F - filtering, S - scraping).

Supplementary Table S4. Number and fraction of reads assigned to various euglenid taxa in samples of three types of biomass harvesting (C - centrifugation, F - filtering, S - scraping) and the total number and fraction of reads. Names of species observed under the microscope are indicated in bold type with the abundance categories in brackets.

Supplementary Table S5. Results of statistical comparisons of pairs of biomass capture methods based on Jaccard and Bray-Curtis distances.

Supplementary Figure S6. Schematic depiction of the experimental procedure applied in the first experimental module: assessment of eight DNA isolation methods: c - commercial kits; m - manual protocols. Chosen method is marked in red.

Supplementary Figure S7. Detailed overview of the experimental procedure applied in metabarcoding survey of green euglenids (second experimental module of this study).

Supplementary Figure S8. Schematic depiction of the experimental procedure applied in the second experimental module: assessment of biomass harvesting method – metabarcoding survey with chosen DNA extraction protocol.

Supplementary Table S9. Primers used in this study. Font code: underlined - Nextera transposase adapters; bold - tag; normal - sequence complementary to V2 18S rDNA region of euglenids.