



Aquatic bioassessment and monitoring using DNA-based methods

Martina Weiss, Arne Beermann, Cristina Hartmann-Fatu, Till-Hendrik Macher, Florian Leese

Disclaimer

University of Duisburg-Essen, Aquatic Ecosystem Research,
Universitätsstr. 5, 45141 Essen

Disclaimer

This book includes conceptual and methodological guidance with respect to the use of DNA and environmental DNA (eDNA) based methods in aquatic bioassessment and ecological monitoring. It was developed as part of the ERASMUS+ project ECOBIAS for student and lifelong learning classes. Because genetic methods develop rapidly, protocol information will need to be updated from time to time. Therefore, prior to using protocols verify that you have the latest versions (field, lab, bioinformatics). The website <https://dnaquahub.eu> provides links to guidance documents and protocols.

The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Erasmus + Project No ECOBIAS_609967-EPP-1-2019-1-RS-EPPKA2-CBHE-JP
Development of master curricula in ecological monitoring and aquatic bioassessment for Western Balkans HEIs

Co-funded by the
Erasmus+ Programme
of the European Union



Preface

Biological assessments are essential to monitor the ecological state of ecosystems and identify biodiversity loss. Typically, this assessment requires taxonomic assessments of specific indicator organism groups (e.g., insects, specific vertebrates, microalgae etc.). By taxonomic experts, the diversity and abundance of these taxa are registered and compared to the situation in undisturbed, near-natural habitats. When the aim is to monitor biodiversity change at sites as holistic as possible, this process of taxonomic identification takes rather long because for many taxa such as small invertebrates or unicellular eukaryotes, few qualified experts exist - and even then, it takes a long time to identify hundreds to thousands of specimens from a subsample. While the taxonomic name of a species and the associated ecological knowledge is the essential basis for bioassessment and monitoring, the traditional identification is often too slow to provide fast and comprehensive pictures of biodiversity across the globe. Several new methods are in place that can complement traditional bioassessment. One very promising one is identification based on organismal DNA, so-called "DNA barcoding". DNA-based identification is common practice in forensics. For decades, criminals have been successfully convicted and sent to prison based on DNA traces left at the crime scene. For decades, fathers who deny paternity have been successfully convicted of their legal payment obligations using DNA fingerprints. And also for decades, cases of food fraud can be successfully detected using DNA screens and manufacturers can be convicted. Now the time has come to utilize the methods, in particular DNA barcoding, for assessment and monitoring of ecosystems and biodiversity. A game changer here was the invention of high-throughput sequencers and the possibility to extract DNA from whole environmental samples in bulk or from the environment (eDNA). Using short DNA fragment to amplify barcode genes for the target biological quality elements - or in general: the target of the environmental monitoring - it is possible nowadays to perform DNA metabarcoding for entire communities or ecosystems. Based on the obtained sequence data it is possible to assess the composition of the distinct community at a site. Data can be stored and reanalyzed easily by basically any scientist around the globe, thus offering easy access to biodiversity data for

Preface

biodiversity monitoring. With this textbook we will introduce the basic idea of DNA-based biomonitoring for different bioindicator taxa in streams and lakes. We provide guidance for field and lab work as well as analysis steps to facilitate knowledge transfer and capacity building as part of the ERASMUS+ project ECOBIAS.

1. Introduction

Water is the essential basis of all life on Earth. Therefore, a key global challenge in the 21st century is to maintain the supply of clean water and other aquatic ecosystem services to humans, without affecting the underlying biodiversity and ecosystem processes that underpin their sustainability. Accordingly, extensive national and international regulations have been adopted to protect water resources.

1.1 Classical ecological status assessment in Europe

In the European Union the Water Framework Directive (WFD, Directive 2000/60/EC) is an important regulation that aims at achieving and protecting good ecological status of all groundwater and surface water bodies. In other countries, similar regulations are established, e.g., the SwissWater Protection Ordinance (WPO, Swiss Federal Council 1998) or the CleanWater Act (CWA, 33 U.S.C. §1251 et seq. 1972) of the US Environmental Protection Agency. All of these regulations aim at protecting aquatic ecosystems from damage and restoring degraded systems to at least “good status”, defined as conditions only slightly altered by anthropogenic activities. The deviations from the ideal status (“high status”) are defined using reference conditions (where possible) or modelling. In the EU, the status of aquatic ecosystems has been monitored by characterizing biological communities, and physicochemical and hydromorphological conditions since the early 2000s.

To achieve the ambitious aim of a good ecological status of all surface water bodies by 2027, many restoration measures have been implemented to address the requirements of the above-mentioned legislation. Quality assessments are being performed by calculating a large number of different biotic metrics and indices. All of these are based on morphological identification of various groups of aquatic indicator organisms at different taxonomic identification levels (see Birk et al. (2012)).

Introduction

The WFD requires ecological status assessment of surface waters to be based on 'Biological Quality Elements' (BQEs), which, depending on the water body type, include "phytoplankton", "diatoms", "aquatic flora", "macroalgae and angiosperms", "benthic invertebrate fauna", and "fish fauna". The resulting lists of taxa and their abundances are used to compute biotic metrics/indices and to define the ecological quality ration (EQR) of a given water body (Fig. 1). The EQR in this context is a ratio between the value of the observed biological parameter (nominator) for a given surface water body and the expected value under reference conditions (denominator). The ratio shall be expressed as a numerical value between 0 and 1, with high ecological status represented by values close to one and bad ecological status by values close to zero. The EQR is typically based on several calibrated biotic metrics/indices. These define certain measures of the structure, function, or some other characteristics of biological assemblages and show a predictable response to anthropogenic disturbances. Over 300 methods have been developed to assess the ecological status of aquatic ecosystems, including rivers, lakes, transitional and marine coastal waters across countries implementing the WFD. Most of them have been formally intercalibrated to allow for comparisons among countries.

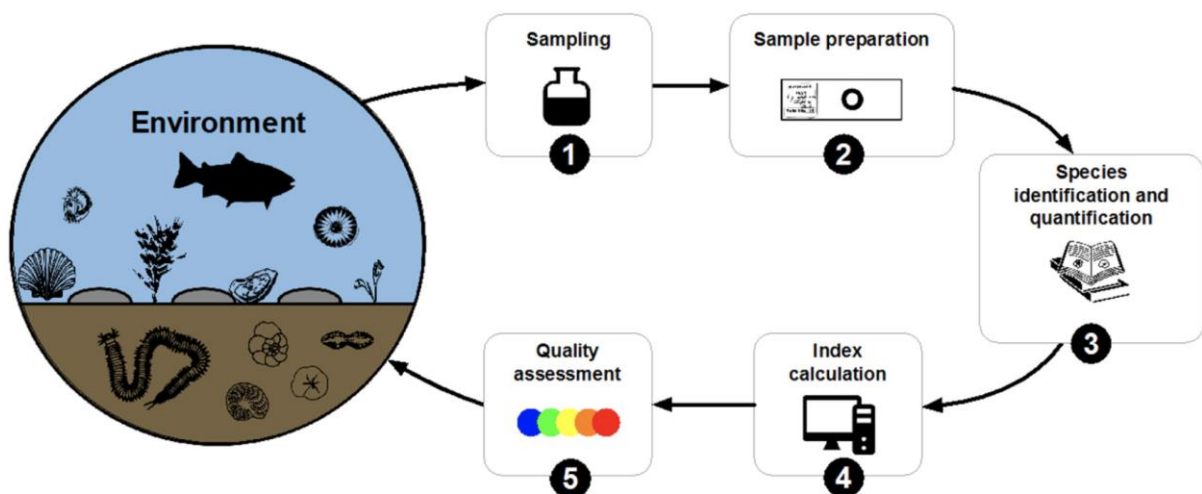


Figure 1: Ecological status assessment ("quality assessment") using traditional indicator taxa. The EQR is calculated in step 4 based on BQE data. Source: Pawlowski et al. (2018).

1.2 DNA-based bioassessment and monitoring

The traditional assessment methods are based on the direct observation of the organisms used to calculate biotic metrics/indices and ultimately the EQR. The approach of sampling and identification of organisms is typically a time and resource-intensive process (Haase et al., 2004). Rather than using identification keys and microscopes it is also possible to use the DNA of an organism to identify it. This process of identification uses DNA isolated from organismal cells (e.g., isolated from mucus, muscle tissue or other cells), amplifies and sequences a characteristic gene fragment and bases the taxonomic assignment by comparing this sequence against a reference database that contains hundreds of thousands of reference data. This process is called "DNA barcoding" (P. D. Hebert et al., 2003). In its simplest manner, it is nothing but the process of reading a "species fingerprint" similar to forensic genotyping. The process of using DNA isolated from individual specimens and amplifying the so-called barcode gene for sequencing and sequence comparison has proven to significantly improve stream bioassessment and monitoring. Stein et al. (2014) showed that 10 out of 16 stream assessment metrics had significantly greater taxonomic resolution based on the DNA barcoding data and thus statistical power to identify effects between impacted and unimpacted streams improve (Figure 2).

While this study, as well as several other studies, showed the power of DNA barcoding to improve assessment, it also became obvious that the main challenge of processing speed and costs are not decreased. Individual specimen-based barcoding takes even longer than morphological assessment because DNA has to be extracted per individual. And given that sequencing had to be performed for each specimen, the estimated costs were between a factor of 1.7-3.4 higher compared to traditional assessment (Stein, Martinez, et al., 2014). Thus, while the information content undoubtedly was and is excellent, single-specimens based bioassessment does not play a straight-forward way with respect to high-

Introduction

throughput ecosystem and biodiversity monitoring, despite novel technologies facilitating the process and leading to a drop of costs (de Kerdrel et al., 2020; P. D. N. Hebert et al., 2018).

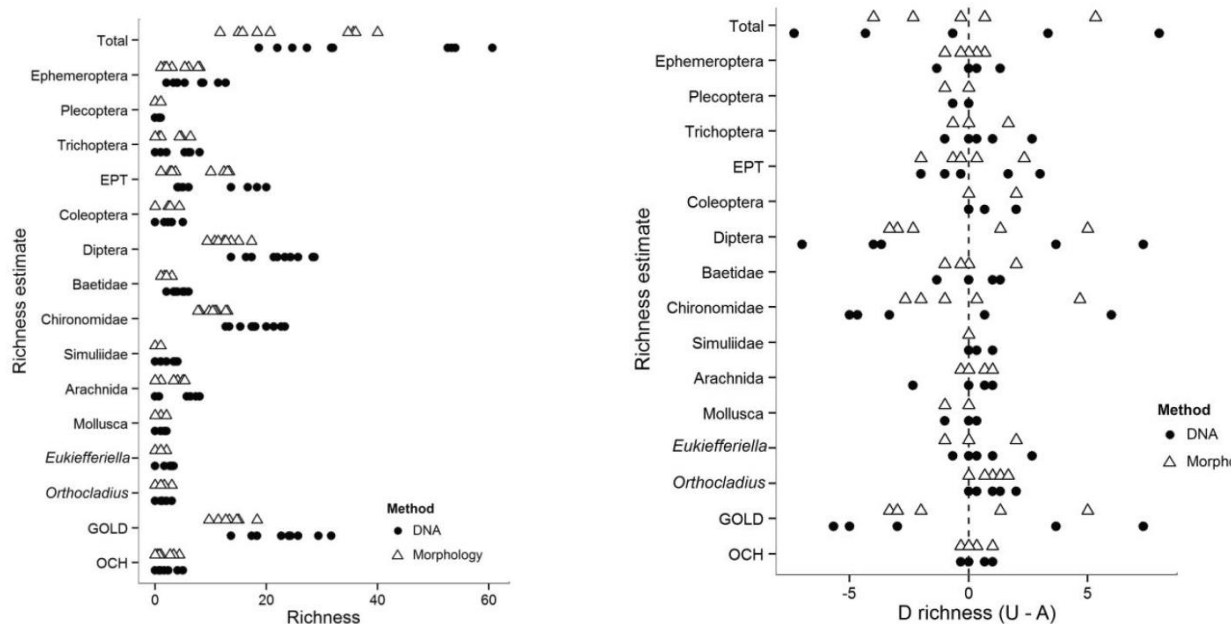


Figure 2: DNA barcoding of individual stream macroinvertebrates (black dots) significantly increased richness estimates compared to traditional taxonomic assessment (triangles) in a comparative stream assessment. Left: Total as well as individual richness estimates each for the morphological and DNA-based assessment. Right: Difference (D) in the richness estimates for unarmoured vs. armoured (=unimpacted vs. impacted) streams. With DNA-barcoding, the differences are much stronger than for traditional morphology-based assessment. Source: Stein et al. (2014).

Recently, the new field of “Biomonitoring 2.0” emerged from advances in DNA barcoding (Baird et al., 2012). Rather than using individual specimens for the analysis, it was proposed to assess ecological status by simultaneous analysis of DNA isolated from many specimens in bulk (“Bulk metabarcoding”, (Baird & Hajibabaei, 2012; Yu et al., 2012) or even by just analyzing traces of DNA of organisms, so-called environmental DNA (eDNA), from the water (“eDNA metabarcoding”(Taberlet et al., 2012, 2018). Environmental DNA (eDNA) refers to DNA present in the environment, and that can be extracted from environmental samples (such as water, or soil, sensu Taberlet et al. 2012). Thus, eDNA contains DNA that exists free

Introduction

in the environment, DNA contained in cells shed from organisms, as well as whole organisms (see Figure 3 for a comparison of the three different DNA-based identification methods).

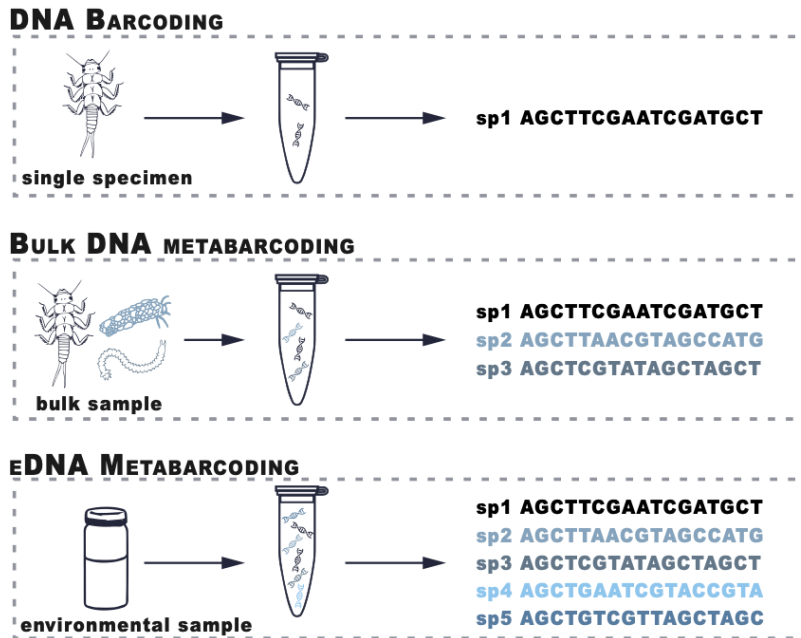


Figure 3: Differences of the three DNA-based identification methods. DNA barcoding refers to the identification of a single species by using cellular material from this specimen for DNA extraction and sequencing. Bulk DNA metabarcoding refers to the simultaneous analysis of many specimens through total sample homogenisation and extraction of DNA from the homogeneous in bulk. eDNA metabarcoding (bottom) refers to metabarcoding based on DNA molecules extracted from the environment (e.g., water or sediment) rather than from collected and sorted individuals. After barcoding, the “sp1, sp2 etc.” is replaced by the species name. Source: Pawlowski et al. (2020)

In this textbook we focus on the use of DNA for community studies, which can be done either by analysis of DNA extracted from bulk samples (bulk DNA metabarcoding) or by analyzing the total eDNA extracted from water or biofilm samples (eDNA metabarcoding; Pawlowski et al., 2014; Deiner et al., 2016; Valentini et al., 2016). The development of DNA metabarcoding has been boosted by advances in high-throughput sequencing (HTS) technologies that overcome most of the limitations of classical cloning/Sanger sequencing approaches, and generate millions of sequences in a relatively rapid and inexpensive way

Introduction

(Shokralla et al., 2012). One major advance was the development of multiplexing protocols, which allowed many samples to be processed at the same time (Herbold et al., 2015). The number of HTS-based metabarcoding studies is growing exponentially, leading to spectacular advances in our knowledge of the global patterns of diversity in aquatic ecosystems, of both prokaryotic (e.g. Besemer et al., 2012; Yilmaz et al., 2016; Thompson et al., 2017) and eukaryotic organisms (e.g. Thomsen et al., 2012; de Vargas et al., 2015; Massana et al., 2015; Leray and Knowlton, 2015; Hänfling et al., 2016; Deiner et al., 2016; Debroas et al., 2017). There is now a growing body of literature summarizing the potential of environmental DNA metabarcoding for biological monitoring (e.g. Bohmann et al., 2014; Cristescu, 2014; Valentini et al., 2016; Keck et al., 2017; Leese et al., 2018; Deiner et al., 2017; Darling et al., 2017) and highlighting its importance for environmental management (Kelly et al., 2014; Jackson et al., 2016; Hering et al., 2018). All of these papers present DNA metabarcoding as a faster, cheaper and easier-to-use alternative to conventional biomonitoring.

The method starts with sampling either communities as traditionally done or via sampling eDNA from water or sediment. Next, the DNA is extracted, and, for the biological quality elements under study, amplified in a PCR. Depending on the taxon under study, different barcode marker genes are used. For macroinvertebrates this is typically a fragment of the mitochondrial cytochrome oxidase I gene (Elbrecht and Leese 2017). For fish eDNA metabarcoding, typically a fragment of the mitochondrial 12S gene is used (Miya et al. 2015). For diatoms the chloroplast ribulose biphosphate gene (*rbcl*) is used or a fragment of the small ribosomal subunit gene (18S). For small meiofaunal taxa such as nematodes or rotifers, often the small or large ribosomal subunit genes (18S or 28S) are targeted. For bacteria, the 16S ribosomal gene is the commonly accepted amplicon marker of choice (see Table 1 in Pawlowski et al. 2018 for an overview).

The barcode marker fragments are analyzed using a high-throughput sequencer (HTS) before then being processed bioinformatically (quality control and filtering, clustering according to similarity, taxonomic assignment and data analysis for final index calculation)(Fig. 4).

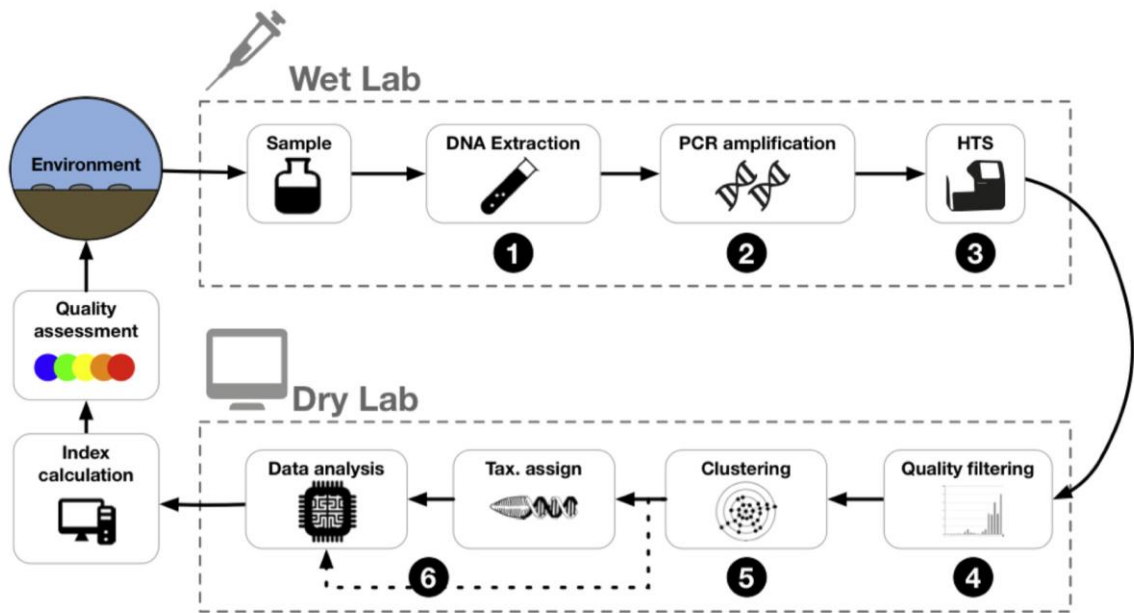


Figure 4: Key steps during sample processing for DNA and eDNA metabarcoding. Source: Pawlowski et al. (2018).

1.3 Targeted vs. untargeted DNA analysis for bioassessment and monitoring

The above-mentioned (1.2) workflow describes a typical DNA and eDNA metabarcoding analysis to recover whole communities ('untargeted approach'). Here, short DNA oligos (primers) are used that bind to regions of ideally all species that could appear in the target community. Thus, these primers are often called universal primers. Due to the fact that mutations also occur at the sites where primers bind, different primers are used for the amplification. For example, if at one position of the binding region some specimens have an adenine base while some of the target specimens have a guanine base (A vs. G), a 50:50 primer mix is used with half of the primer containing the complementary T and half the complementary C (T/C). Because both are pyrimidine bases (IUPAC code Y), the primer contains a "Y" at this position. Several primers have to be used in cases of complex

Introduction

communities, leading to the use of so-called “degenerated” universal primers that contain many such ambiguity IUPAC codes.

However, there are also monitoring approaches that specifically aim to detect only specific taxa of concern (TOC). In these cases, specific regions are targeted where only the TOC contain a mutation for which TOC-specific primers can be designed. Then, the absence or presence of the TOC can be assessed using only a diagnostic PCR, i.e. a PCR where the presence of a product is a proof of presence of the respective taxon. For this purpose, specifically validated primers need to be developed that do not amplify DNA of any other target organisms (false positives). Specific guidance for the development of TOC exists (www.edna-valuation-com).

1.4 Aim of this text book

The aim of this textbook is to highlight the approaches for aquatic bioassessment and monitoring or biological quality elements (BQEs) used as part of the European Water Framework Directive. We aim to introduce the key methods and underlying concepts as well as provide hands-on guidance with respect to field, laboratory, and data analysis steps. Finally, also considerations regarding potential limitations (false positives, false negatives) will be provided.

2. Methods

For all BQEs, eDNA analysis starts with sampling. Here, depending on the targeted organisms and water bodies, slightly different sampling strategies exist and different aspects, e.g. regarding replicates, or sampling spots have to be considered. While for fish and macroinvertebrates, water samples can be used, diatoms need to be scratched with the biofilm from stones. When considering water bodies, main differences exist between lentic (standing water bodies) and lotic (running waters) ecosystems and also for large and small water bodies. Especially for larger lakes, seasonal stratification needs to be taken into account, when designing sample campaigns and also when comparing for example estimated diversity over time or between lakes. Further, more samples need to be taken when assessing lake biodiversity with eDNA in comparison to running waters, because DNA will be distributed more patchy due to the lack of constant mixing of the water. Samples should be taken in different depths and sites in the lake to cover multiple habitats. However, also in streams and rivers, it is important to think about targeted organism groups when planning sampling, as also here DNA is not equally distributed (e.g. Macher & Leese 2017). Species detection can differ for samples taken from the surface or the bottom in the middle of the stream or at the shore, depending on where target species have their preferred habitats. Therefore, when conducting eDNA sampling for bioassessment, it is preferable to include multiple habitats, but also a single sample can be used to get an overview over species living in the river system. However, in general it is of great importance to precisely document how samples were taken and to do it similarly when aiming to compare different streams.

In the following, the sampling and filtration of eDNA for fish and macroinvertebrate bioassessment and of biofilm eDNA for diatoms is described in detail as well as the sample-specific homogenization. After this step, all samples can be processed in a similar way. Here, the processing in the lab starts with DNA extraction, followed by PCR with primers for different target organism groups. Here, depending on the aim, qPCR can be conducted for single-species detection, or metabarcoding PCRs if the aim is to assess community

composition. For both aims, different approaches exist, but we will focus here on commonly used and cost-efficient approaches. For metabarcoding, PCRs are followed by high-throughput sequencing, and sequence data analysis to eventually obtain taxa lists.

2.1 eDNA sampling

2.1.1 eDNA from water samples

To obtain eDNA from water samples, the most commonly used method is filtration. With this technique, DNA from e.g. cells or organelles can be bound to a filter membrane, from which DNA can then be extracted. Filtration can be conducted in different ways, using e.g. peristaltic or vacuum pumps (see, e.g., Pawlowski et al. 2020 for details). Another method, which is easy to apply in the field also at remote sites and does not require much material, is hand filtration using disposable syringes. We will describe this method here in detail, using material from a commercial company in Austria (Sinsoma) that was used as part of a large German biomonitoring project (GeDNA). However, the analysis workflow is easily transferable to other hand filtration methods using syringes and encapsulated filters and the principle is the same for all filtration methods. Generally, filtration can be carried out in the field or in the lab, but risk of cross-contamination is lower in the field and it is easily applicable with the hand filtering method. Afterwards samples need to be transported in cool conditions to prevent DNA degradation. Here, filters are also to be preferred as they take up less space than bottles.

Another important point to consider before starting the sampling is how much water needs to be sampled. In general, the more water is filtered, the more DNA can be obtained. However, filtering of large volumes is often not feasible in the field, especially when using hand pumps and also filters will clog after a time (also depending on suspended particles in the water). In general, for bioassessment in streams often 1 – 2 L are sufficient, while for large rivers and lakes more water needs to be filtered. If the aim is to detect specific rare species, also more water is needed. Here, also the sampling depth needs to be taken into account for example when targeting benthic or pelagic fish.

Methods

Before we can start with sampling, we need to think of possible sources of contamination and how to avoid these. In general, we aim to detect low amounts of DNA, making it extremely important to work as sterile and clean as possible using either single-use, sterile packed material, or cleaning equipment that needs to be reused before each sampling site. For cleaning, chlorine bleach (5 – 10%) or other decontamination solutions can be used. However, it is best to conduct the cleaning in the lab, as large volumes of bleach should not be used in the field. After cleaning with bleach, equipment needs to be rinsed with ddH₂O to remove remaining bleach, so that it is not transported into the field. All these measures are important to reduce the risk of cross-contamination between samples and with this to avoid false-positives, meaning here the detection of a species, which was not present at the site. To further avoid cross-contamination as well as possible human contamination, it is important to wear sterile gloves at all sampling steps and to discard them after each sample. Another important measure to be able to detect and to remove possible false positives, is to integrate negative/blank controls and replicates in each step of the process. To obtain negative controls from the field (also called blanks), deionized water is sometimes taken to the field sites and filtered there; alternatively field blank are obtained by exposing the filter at the site of filtration. If possible, this should be done at each sampling site, or at least at the first and the last site sampled at one day and it is very important that these samples are processed in the same way as the field samples in the whole workflow.

The sampling process is summarized in [Fig X. \(Guidelines eDNA sampling\)](#). When using the commercial kit (Sinsoma, Austria) it is important to store the small 1 ml syringe at 4°C until sampling. For filter storage and DNA extraction (lysis), it contains TNES buffer and proteinase K and the latter needs to be stored under cool conditions. An alternative solution is to fill up an appropriate 1 ml syringe with 500 µl TNES and add Proteinase K only when starting with DNA extraction in the lab. In this case, syringes could be stored at room temperature.

When aiming at bioassessment of freshwater ecosystems, often different kinds of samples are taken at one sampling site, for example macrozoobenthos kick samples, biofilm samples and water samples for eDNA analysis. Here, in order to reduce the contamination risk, it is important to start with taking the water sample before stirring the sediment in the

Methods

process of taking other samples. When visiting more than one sampling site in a stream, or connected stream system, start with sampling the most downstream site to prevent distribution of DNA in the flow direction from more upstream to downstream sites. Before starting to sample, put on sterile gloves (you need one pair for each sampling site) and take a sterile plastic bottle. When sampling is conducted in running waters, walk against the flow direction to reach the sampling site without stirring the sediment. In general, if only one sample should be taken and general biodiversity is targeted, the best position for sampling is in the middle of the river. For deep rivers, the water sample should be taken as far away from the river shore as possible without safety concerns. To get a more complete picture of the biodiversity of a stream or if specific taxa are targeted, it is beneficial to sample at different habitats, e.g., at the shore near to the ground on both sides of the river, or at the specific habitat preferred by the target taxon. However, for general bioassessment, also one sample can be sufficient and at least 1 l of water should be sampled. When reaching the sampling site, it is important to first rinse the sample bottle with stream water, before taking the sample. To do so, dip the bottle gently into the water with the opening facing upstream, until it is approximately half-filled and empty the bottle downstream of the sampling site. Then take the real sample by dipping the plastic bottle again gently into the water with the opening facing upstream until it is completely filled. Now, you can return to the shore and start with the filtration (Fig Y, step 3). For this you will first need the big 60 ml syringe and the filter capsule, which needs to be unpacked first. As a first step, draw 50 ml water from the sample bottle up into the big syringe. To filter the DNA out of the water, screw the syringe onto the filter capsule and slowly push the water through the filter. Twist off the filter and put it onto the syringe or the filter package. Never just put it on the ground or onto other unsterile surfaces to prevent contamination. Repeat drawing and filtering water until 1 L of water is filtered, or the filter is clogged. Clogging means that it is not possible anymore to push water through the filter. An increase of resistance during the filtration is normal. Depending on the project and availability of filters per site, either proceed with a second filter until 1 L is filtered, or note the filtered volume and proceed with the next step. Unscrew the filter, push any remaining water out of the syringe, and draw in air instead. Screw the filter again on the syringe and push air through the filter to remove the remaining water. There should remain as little

Methods

water as possible in the filter as it would otherwise dilute the buffer added in the next step and enhance degradation.

For preservation and preparation of filters for DNA extraction (Fig X, step 4), lysis buffer (TNES) needs to be injected into the filter capsule. To do so, remove the plug from the small 1 ml syringe containing either only TNES or a mix of TNES and proteinase K, and insert it into the filter capsule. Turn the filter to ensure that the unused opening is positioned in horizontal orientation and push the buffer carefully and slowly through the filter. Use the cap ('plug') to close the unused opening of the filter capsule by gentle turning. Unscrew the syringe and use the second cap to close the other opening. When the filter is closed on both sides, sway it for some seconds to ensure that the whole membrane is wetted with the buffer. Now it is important to label the filter capsule unambiguously with a waterproof pen. Put the filter into a zip bag and store it under dark and cool conditions at 4°C until processed in the lab.

Guidelines: eDNA sampling

1) Preparation

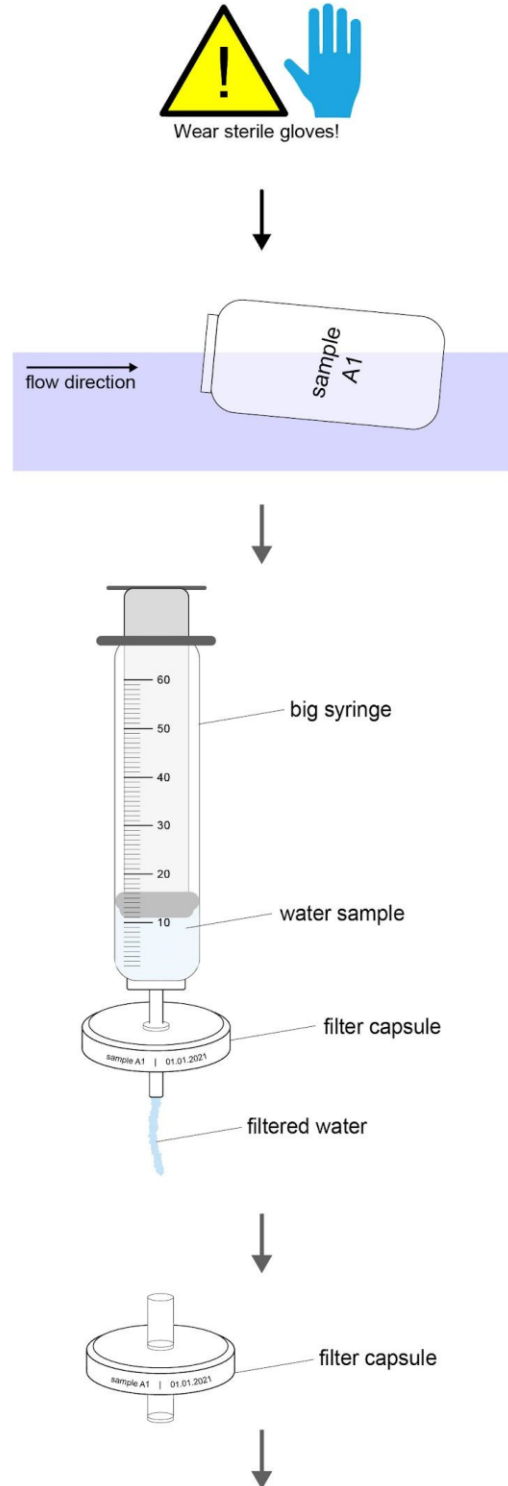
- a) Open sampling kit and store zip bag with small syringe at 4°C in a fridge until sampling.

2) Sampling (using Sinsoma technology)

- a) Put on sterile gloves.
- b) Take sterile plastic bottle from sampling kit.
- c) The best position for the sampling site is in the middle of the river. For deep rivers, sampling should be conducted as far away from the river bank as possible without safety concerns.
- d) To get to the sampling site, walk against flow direction to prevent stirring the sediment.
- e) Use the DNA-free plastic bottle to take the water sample. Before taking the sample, the plastic bottle must be rinsed: Dip the bottle gently into the water against the flow direction until it is approximately half-filled. Empty bottle downstream of sampling site.
- f) Take the water sample by dipping the plastic bottle gently into the water against the flow direction until it is completely filled.

3) Filtration

- a) Draw 50 mL water from the sample bottle up into the big syringe
- b) To filter the DNA out of the water, take the filter out of the packing, screw the syringe onto the filter capsule, and slowly push the water through the filter.
- c) Twist off the filter and put it onto the syringe package. Attention! Do not put it onto the ground.
- d) Repeat drawing of 50 mL water from the bottle into big syringe, screw the syringe again onto the filter capsule, and slowly push the water through the filter.
- e) Repeat this until 1 L of water is filtered, or the filter is clogged, i.e. it is not possible to push water through the filter anymore; an increasing resistance during filtration is normal.
- f) Depending on the project requirements, either proceed with a second filter until 1 L is filtered, or note filtered volume and proceed with next step.
- g) Unscrew filter, push any remaining water out of the syringe, and draw in air instead.
- h) Screw filter again on syringe and push air through the filter, this will remove remaining water. Important: There should remain as little water as possible in the filter.



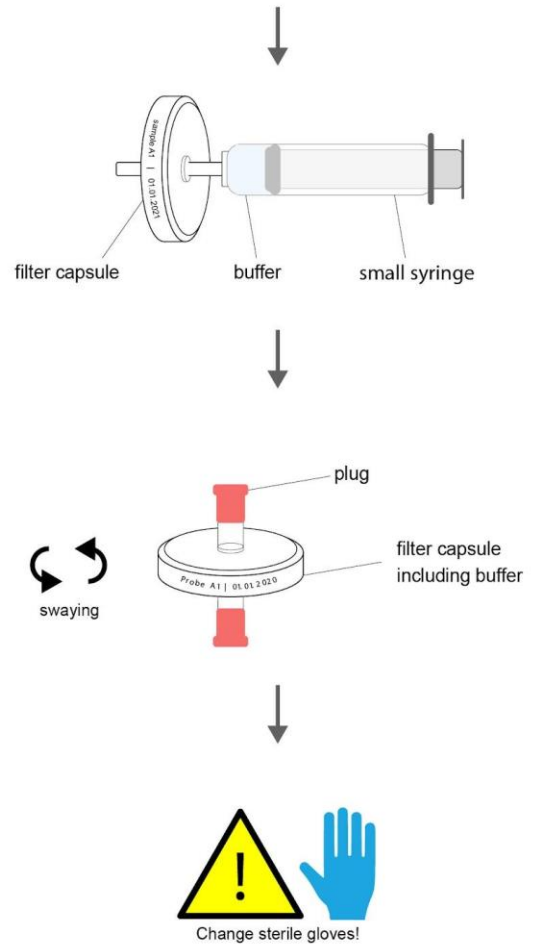
Guidelines: eDNA sampling

4) Preservation

- a) Remove plug from small syringe, insert it into the filter capsule and turn it to ensure that the unused opening is positioned in horizontal orientation. Push the liquid included in the syringe carefully and slowly through the filter.

The buffer is not fit for human consumption, should not get in contact with the eyes and should only be removed from the fridge shortly before usage.

- b) Use red plug to close unused opening of filter capsule by gentle turning. Unscrew small syringe and use the second plug to close the other opening of the filter capsule. When filter capsule is closed on both sides, sway it for some seconds to ensure that the whole membrane is wetted by the buffer.
- c) Label the filter capsule unambiguously with a waterproof pen. Put the filter into the zip bag and close it and store it at 4°C.
- d) Change sterile gloves after every sample!



Materials

- sterile gloves
- package with sterile big syringe (50 - 60 mL)
- plastic bottle (500 mL)
- package with sterile filter
- small syringe (2 mL) including 500 µL buffer*
- two sterile packaged plugs
- sterile zip bag

* The buffer included in the small syringe is not fit for human consumption and should not get into contact with the eyes. Please store at 4°C and keep out of the reach of children.

Protocol modified after Sinsoma, Lanes 6, 6176 Völs, Austria

2.1.2 eDNA from biofilms (targeting diatoms)

The sampling strategy for assessing biodiversity of diatoms in streams differs from the sampling strategy of filtering water described before for fish and macroinvertebrates. Diatom DNA is not sampled from the water, but from the biofilm associated with larger stones. If larger stones are not present, it is also possible to sample from dead wood, sediment or macrophytes, but solid substrate is to be preferred. The sampling steps are summarized in Fig. Y (Guidelines diatoms). The number of subsamples or replicates that need to be taken for bioassessment differs between countries. For example, 10 subsamples are taken at the sampling sites. In general, 3 to 5 larger stones are sampled per site and biofilm is scratched from the surface using single-use toothbrushes or a sterile knife. To avoid contamination, wear sterile gloves during the sampling process and discard them after each sampling site. The sampling should preferably be conducted using a single-use toothbrush, which has to be changed for every sample to prevent cross-contamination. As an alternative, a pocket knife can be used, but it needs to be thoroughly cleaned between samples with bleach. As described for water samples, approach the sampling site from downstream against the flow direction to prevent stirring of the sediment. Take the toothbrush or knife with you and also a sterile sampling container. Sampling directly into the sampling container is important to prevent contamination by interim storage in e.g. sampling trays. If this is not possible, sampling trays or used interim storage containers need to be cleaned with bleach after sampling. Start sampling by collecting a stone at the sampling site, wet the toothbrush in river water and scrape the surface of the stones over the sampling container. Repeat this for 3 to 5 stones per subsample, so that in the end the sample container is filled with approximately 20 ml water-biofilm solution. If metabarcoding as well as morphological analyses are planned, the sample should be directly divided into two bottles. For morphological determination, add 10 mL sample to 10 mL 96% technical ethanol, resulting in a final concentration of approximately 50% ethanol. For DNA metabarcoding, add 10 mL of sample to 40 mL 96% technical ethanol (target concentration of ethanol approx. 75%). Mix ethanol and sample by inverting the bottle. If sediment samples had to be taken by using a syringe, take the sample container to the lab and let it sediment for one day. Discard the upper water phase and take two samples (10 mL each) for metabarcoding and morphological analysis and mix it with

Methods

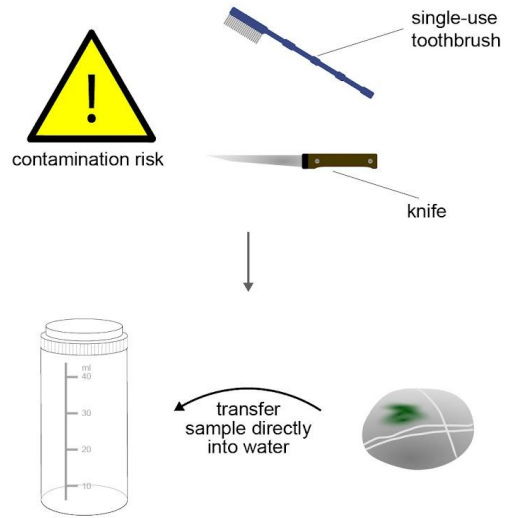
ethanol as described before. The samples need to be stored at dark and cool conditions (4°C or -20°C). Always make sure that samples are correctly and unambiguously labeled from the outside on the lid and at the side with a waterproof pen.

Guidelines: Diatoms Sampling for DNA metabarcoding



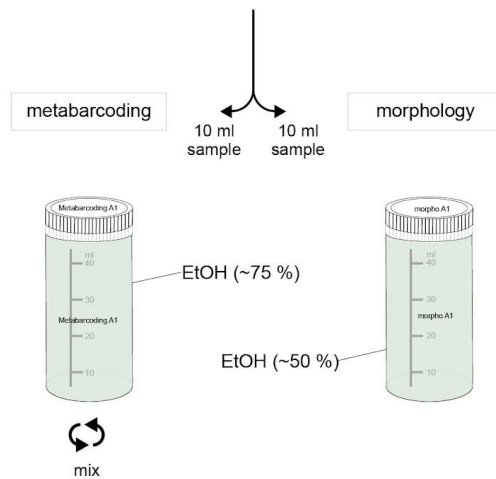
1) Sampling

- Use the multi-habitat sampling protocol usually used in your country (e.g. 10 sub samples).
- Approach sampling site from downstream against flow direction to prevent stirring of the sediment
- Sampling is conducted using a single-use toothbrush, which has to be changed for every sample to prevent contamination. Cleaning of toothbrushes with e.g. bleach is not feasible. As an alternative, a pocketknife can be used, if it is thoroughly cleaned between samples with chlorine bleach.
- The diatom film from the substrate should be directly filled into the sampling container into water to prevent contamination by interim storage e.g. in a sampling tray (otherwise it needs to be cleaned with chlorine bleach after each sample).
- Solid substrate should be preferably sampled. Alternatively, also dead wood, sediment, or macrophytes can be sampled.



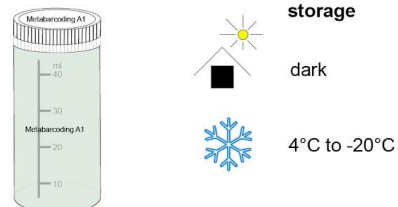
2) Sample splitting & preservation

- If metabarcoding and morphological analysis are conducted, the sample should be divided in two bottles:
 - For morphological determination, add 10 ml sample to 10 ml 96% technical ethanol resulting in a final concentration of approx. 50% ethanol, or use alternative protocol.
 - For DNA metabarcoding, add 10 ml sample to 40 ml 96% technical ethanol resulting in a final concentration of approx. 75% ethanol. Mix ethanol and sample by inverting the bottle.
- If sediment samples were taken with a syringe, take the sample container to the lab and let it sediment for one day. On the next day, discard the upper water phase and take two samples (10 ml, each) for metabarcoding and morphological analysis and proceed as described in 2a.
- Store samples in the dark and ideally cool (at 4°C or -20°C). Check if samples are correctly labeled. The sample container should be labeled from the outside on the lid and at the side using a waterproof pen.



Materials

- equipment for classical diatom sampling:
 - sterile single-use toothbrushes, and/or knife
 - use additional sampling containers only after thorough cleaning with chlorine bleach, otherwise risk of contamination!
- 96% technical ethanol
- sterile, leak-proof 50 ml sample container
- waterproof pen



2.2 eDNA extraction

The first step, when aiming at analyzing the community composition of a (eDNA) sample with metabarcoding is to extract the DNA from the filter, the biofilm, or other environmental or tissue samples. We want to isolate the DNA from the environmental sample and separate it from proteins, membranes and other cellular material contained in cells. There are various methods that can be used for DNA extraction, which vary extremely in price, time, effort but also accuracy. In general, the workflow is similar for all methods:

1. **Lysis:** DNA is isolated from nucleus and organelles by breaking down the cell membranes. The lysis can be conducted either by using a lysis buffer and other chemical components, or by mechanical destruction of the walls and cell membranes using for example beads and a bead-beating device. These beads are small balls with a size of 0.03 mm up to 25 mm and can be made of glass (silica), ceramic (zirconium) or steel. Often, also a combination of both methods is applied when dealing for example with tissue samples. For water and biofilm eDNA it is sufficient to go for a chemical lysis using for example the lysis buffer TNES and proteinase K to degrade the double lipoprotein layers of cells and digest enzymes.
2. **Removal of organic and inorganic components** except DNA. Here often a concentrated salt solution is used to make these components (e.g. proteins and lipids) clump together, which can then be separated from the DNA by centrifugation.
3. **Purification of DNA** from detergents, proteins, salts and other reagents, using e.g. precipitation, silica membranes or magnetic beads. For salt precipitation, ice-cold ethanol is added in the presence of Na^+ ions (from sodium chloride, NaCl) interrupting the hydrogen bonds between the phosphate molecules of the DNA and water. In addition, ethanol induces a structural change of DNA molecules that causes them to aggregate and precipitate and can be pelleted by centrifugation. After precipitation, washing steps are conducted using e.g. 70% ethanol to remove salt and other water-soluble ingredients without solubilizing the DNA. Another method for purification is the use of spin-columns containing a silica membrane. Here the principle is based on the binding of DNA to the silica membrane dependent on the pH and the salt

Methods

concentration (chaotropic salts) of the buffer. After binding of the DNA to the membrane, also ethanol washing steps are conducted as described before. Similarly, DNA can be bound to magnetic beads instead of the membrane and purified by washing out the other ingredients using 80% ethanol.

4. **Resuspension:** after removal of remaining ethanol, the DNA pellet (precipitation method) or DNA bound to the silica membrane or magnetic beads can be resuspended in ultrapure water, TE-minimum or other slightly alkaline elution buffers.

The DNA extraction can be performed using commercially manufactured kits (e.g. Qiagen, ThermoFisher, Macherey-Nagel) or custom-made protocols, which are usually much less expensive, but might not result in the same high quantities of DNA. Here, we will describe in detail two different extraction protocols, one using magnetic beads (NucleoMag Tissue kit for DNA purification from cells and tissue, Macherey-Nagel, Düren), and the second using a salt precipitation protocol.

One general remark before starting with the lab work, is that extra care has to be taken when processing the samples to prevent contamination, as we aim at analyzing very low amounts of DNA. Here, different



opinions exist which measures are necessary and are also dependent on the target of the analysis (detection of single rare species/ community analyses). However, in general, it is important that all material, equipment and working places are decontaminated using either UV light or bleach solution before starting to process samples. Further single-use protective clothing, face masks and sterile gloves as can be seen in the pictures (Fig X) are important to wear to prevent human contamination. To be able to detect contamination it

is further very important to include negative controls in all laboratory steps and to use extraction and PCR replicates.

2.2.1 eDNA extraction using NucleoMag B-Beads

An alternative to using commercially manufactured kit, is to use custom-made beads and also magnet racks to reduce costs. Detailed instructions for preparation of ingredients and material can be found at <https://bomb.bio> and are similar to the ones described here. Steps which are needed for both extracting DNA from filters and biofilm are written in black, steps only important for filters in

green, and only for biofilm samples in blue.

Fig x: Working in an eDNA lab (Photo: Till-Hendrik Macher, Robin Schütz)

Chemicals, materials and devices:

- 10 ml syringe (1 per sample)
- Filter tips of different sizes and pipettes
- 1.5 ml Eppendorf tubes, 2 ml twist-top tubes
- *Optional: Zirconia beads and bead beating device*
- TNES buffer (50 mM Tris Base, 400 mM NaCl, 20 mM EDTA, 0.5% SDS pH 7.5; 1000 µl per sample)
- Proteinase K (10 µl per sample)
- NucleoMag Tissue kit for DNA purification from cells and tissue:
 - NucleoMag B-Beads (8 µl per sample)
 - Binding Buffer MB2 (320 or 640 µl per sample)
 - Washing Buffer MB3 (200 µl per sample)
 - Washing Buffer MB4 (200 µl per sample)
 - Washing Buffer MB5 (300 µl per sample)
 - Elution Buffer MB6 (40-50 µl per sample)

Methods

- Centrifuge
- Heating block
- Magnet separator rack
- *Optional: bead-beating device*
- *Optional: vacuum centrifuge*

Lysis filter and sample transfer

1. Preheat the heating block to 55 °C.
2. Place the filter in a Falcon tube rack with the inlet facing upwards. Remove the inlet cap ('plug') **and add 10 µl of Proteinase K** to each filter if filters had been stored in TNES. When using the SIMSONA kit, this step is not necessary as Proteinase K is already included in the buffer. Close the inlet cap and shake the filter 10 times to mix the sample.
3. **Incubate samples for 3h at 55 °C.**
4. Label one Eppendorf 1.5 ml tube for each sample.
5. Remove both caps of the filter and place it sideways on a Falcon tube rack. The inlet and outlet have to be oriented horizontally.
6. Take a new **10 ml sterile syringe** and place it on the respective filter.
7. Gently **draw the mix of sample, TNES and Proteinase K out of the filter**: Face the inlet downwards and the syringe upwards. Draw the syringe and wait until the sample stops running into the syringe.
8. Gently remove the syringe without spilling droplets from the inlet. Note: The sample volume can vary between 400 µl to 1200 µl depending on how much water was left before the buffer injection.
9. **Transfer the sample** to the respective Eppendorf 1.5 ml tube by gently clearing the syringe. Prevent the occasionally produced foam from exiting the tube.

Lysis biofilm sample:

1. Preheat the heating block to 55 °C.

Methods

2. If available, start the vacuum centrifuge and set temperature to 60°C.
3. OPTIONAL: Prepare 2 ml twist-top tube for each sample. Label both the tube and the lid. Add a small spatula tip of zirconia beads to each tube if a bead-beating device is available. Otherwise skip this step and continue with the chemical lysis.
4. Thoroughly mix the diatom sample before transferring it to the tube.
5. Pipette 2 ml of mixed sample to twist-top tube. Make sure that the sample remains mixed and no separate phases are visible.
6. Centrifuge samples at full speed for 5 min.
7. Remove and discard supernatant by pipetting. Make sure not to come in contact with the pellet and remove as much ethanol as possible.
8. Open tubes and place them in the pre-heated vacuum centrifuge or a heating block (55°C). Leave one position empty between tubes and make sure not to touch the insides of the lids and tubes.
9. Start vacuum centrifuge and let the remaining ethanol evaporate for 6 min (5-15 min heating block) or until the sample is dried (green-greyish pellet).
10. Add 1000 µl TNES and 10 µl Proteinase K to the dried pellet. Mix by vortexing and spin down shortly.
11. OPTIONAL: Grind the samples for 1 min in the bead beater.
12. Incubate samples for 25 min at 55 °C, 1400 rpm.

Clear lysate and bind DNA to NucleoMag B-Beads

1. Set the thermocycler to 20 °C.
2. Centrifuge the samples for **5 min at full speed (5,600-6,000 rcf)**.
3. Prepare premix for all samples consisting of **8 µl NucleoMag B-Beads** and **320 µl binding buffer MB2** per sample and subsequently distribute 328 µl of the Bead/Buffer-Mix into new, labelled Eppendorf 1.5 mL tubes. Instead of using tubes, it is also possible to use deep-well plates (96 wells).
4. Transfer **200 µl of the cleared lysate** into a labelled tube containing Bead/Buffer-Mix and pipette up and down 6 times. If extraction replicates are performed, transfer another 200 µl into a second tube. Make sure to discard pipet tips in all cases where it could have been in contact with a sample or other ingredients.

Methods

5. Shake the tubes for **5 min at 1000 rpm** at 20 °C (room temperature, benchtop shaker).
6. Separate the magnetic beads against the side of the tubes by placing them in the magnetic separator rack. Wait **2 min** until all the magnetic beads are attracted to the magnets.
7. **Remove and discard clear supernatant** by pipetting. Do not discard the beads!
8. Remove the tubes from the magnetic separator rack.
9. Recommended for higher yield: Add 200 µl of sample and 320 µl Buffer MB2 to increase the extraction input volume and repeat steps 2-5 once.

Wash with washing buffer MB3

1. Remove the tubes from the magnetic separator rack.
2. Add **200 µl Buffer MB3** to each tube and mix by pipetting up and down 6 times.
3. Shake the tubes for **5 min at 1000 rpm at 20 °C**.
4. Separate the magnetic beads against the side of the tubes by placing them in the magnetic separator rack. Wait **2 min** until all the magnetic beads are attracted to the magnets.
5. **Remove and discard clear supernatant** by pipetting. Do not discard the beads!

Wash with washing buffer MB4

1. Remove the tubes from the magnetic separator rack.
2. Add **200 µl Buffer MB4** to each tube and mix by pipetting up and down 6 times.
3. Shake the tubes for **5 min at 1000 rpm at 20 °C**.
4. Separate the magnetic beads against the side of the tubes by placing them in the magnetic separator rack. Wait **2 min** until all the magnetic beads are attracted to the magnets.
5. Remove and discard the supernatant by pipetting. Do not discard beads!

Methods

6. *Do not remove the tubes from the magnetic separator rack!*
7. Set the thermocycler to 56 °C.

Wash with washing buffer MB5

1. *Leave the tubes in the magnetic separator rack!*
2. Gently **add 300 µl Buffer MB5** to each tube without disturbing the pellet and incubate for 45 – 60 s. *Do not resuspend the magnetic beads in the Buffer MB5!* Note: It is possible to incubate 4-5 tubes at the same time. Sequentially add Buffer MB5 and immediately after the last sample (4 or 5) start aspirating Buffer MB5 again. When conducted in plates, more samples can be simultaneously processed and multichannel pipettes used.
3. Aspirate and discard the supernatant by pipetting.

Elution with elution buffer MB6

1. Remove the tubes from the magnetic separator rack.
2. Add 40 - 50 µl Buffer MB6 to each tube and mix by pipetting up and down 10 times.
3. Incubate the tubes for 10 min at 1000 RPM at 56 °C.
4. Label a new Eppendorf 1.5 ml tube for each sample.
5. Separate the magnetic beads against the side of the tubes by placing them in the magnetic separator rack. Wait 2 min until all the magnetic beads are attached to the magnets.
6. Transfer the supernatant containing the purified DNA to the new tube.
7. Store the samples at -20 °C until further processing.

2.2.2 eDNA extraction using salt precipitation

Lysis and sample transfer need to be conducted in the same way as described above. For easier usability of the full protocol, it is repeated here.

Chemicals, materials and devices:

- 10 ml syringe (1 per sample)
- Filter tips of different sizes and pipettes
- 1.5 ml Eppendorf tubes, 2 ml twist-top tubes
- Optional: Zirconia beads and bead beating device
- TNES buffer (50 mM Tris Base, 400 mM NaCl, 20 mM EDTA, 0.5% SDS pH 7.5; 1000 μ l per sample)
- Proteinase K (10 μ l per sample)
- NaCl 5M
- -20°C ice-cold Ethanol p.a. (100%)
- Ethanol p.a. 70%
- TE-minimum buffer (or H₂O PCR grade)
- Centrifuge and cooling centrifuge
- Heating block
- Optional: bead-beating device
- Optional: Vacuum centrifuge

Lysis filter and sample transfer

1. Preheat the heating block to 55 °C.
2. Place the filter in a Falcon tube rack with the inlet facing upwards. Remove the inlet cap **and add 10 μ l of Proteinase K** to each filter, if filters had been stored in TNES. When using the SIMSONA kit, this step is not necessary, as Proteinase K is already included in the buffer. Close the inlet cap and spin the filter 10 times to mix the sample.
3. **Incubate samples for 3h at 55 °C.**
4. Label one Eppendorf 1.5 ml tube for each sample.
5. Remove both caps of the filter and place it sideways on a Falcon tube rack. The inlet and outlet have to be oriented horizontally.

Methods

6. Take a new **10 ml sterile syringe** and place it on the respective filter.
7. Gently **pull the sample out of the filter**: Face the inlet downwards and the syringe upwards. Pull the syringe and wait until the sample stops running into the syringe.
8. Gently remove the syringe without spilling droplets from the inlet. Note: The sample volume can vary between 400 μl to 1200 μl depending on how much water was left before the buffer injection.
9. **Transfer the sample** to the respective Eppendorf 1.5 ml tube by gently clearing the syringe. Prevent the occasionally produced foam from exiting the tube.

Lysis biofilm sample:

1. Preheat the heating block to 55 °C.
2. If available, start vacuum centrifuge and set temperature to 60°C.
3. *Prepare 2 ml twist-top tube for each sample. Label both the tube and the lid. Add small spatula tip of zirconia beads to each tube if bead-beating device is available. Otherwise skip this step and just go for the chemical lysis.*
4. Thoroughly mix the diatom sample before transferring it to the tube.
5. Pipette 2 ml of mixed sample to twist-top tube. Make sure that the sample remains mixed and no separate phases are visible.
6. Centrifuge samples at full speed for 5 min.
7. Remove and discard supernatant by pipetting. Make sure not to come in contact with the pellet and remove as much ethanol as possible.
8. Open tubes and place them in the pre-heated vacuum centrifuge or a heating block (55°C). Leave one position empty between tubes and make sure not to touch the insides of the lids and tubes.
9. Start vacuum centrifuge and let the remaining ethanol evaporate for 6 min (5-15 min heating block) or until the sample is dried (green-greyish pellet).
10. Add 1000 μl TNES and 10 μl Proteinase K to the dried pellet. Mix by vortexing and spin down shortly.
11. *Grind the samples for 1 min in the bead beater.*
12. Incubate samples for 25 min at 55 °C, 1400 rpm.

Salt precipitation

Switch on the cooling centrifuge at 4°C prior to next steps

1. Add **170 µl** of 5 M **NaCl** and mix for 15 s to precipitate proteins.
2. Centrifuge for **5 min at 18,400 rcf, RT** (room temperature).
3. Pipette clear supernatant into new previously labelled tube, discard the protein pellet.
4. Centrifuge again for **5 min at 18,400 rcf, RT**.
5. Pipette clear supernatant into new tube, discard the protein pellet.
6. Add **800 µl** of **-20°C 100% ethanol**, invert tubes multiple times, and centrifuge for **15 min at 18,400 rcf** in a **cooling centrifuge** to pellet the DNA.

Washing with EtOH

1. Discard the clear supernatant (do not discard the DNA pellet), add **180 µl of 70% ethanol**, and centrifuge again for **15 min at 18,400 rcf** in a **cooling centrifuge**.
2. Remove ethanol and air dry the DNA pellet (heating block at 55°C).

Elution

1. Dissolve the DNA in 50 µl TE-minimum buffer or PCR grade H₂O.
2. Store the samples at -20 °C until further processing.

2.2.3 Evaluation of DNA extraction success

To assess the quantity and quality of the obtained DNA after extraction, different methods can be applied. The verification on an agarose gel is a quick and comparatively cheap method to get an overview about extraction success and DNA quality. Here, DNA fragments of different size are separated through electrophoresis and compared to DNA standards of known size and concentration (ladder). If the expected DNA concentration is very low as for example for eDNA from water samples, quantification of DNA can instead be conducted for example with a Qubit Fluorometer. The downside of this measurement

Methods

is that it is more expensive and only shows the quantity (better than in gels, where estimation of exact concentration is not possible) but not the quality of the DNA, so if it mostly consists of high molecular weight or degraded DNA.

Gel electrophoresis

Agarose can be dissolved in either TAE (TRIS-Acetate-EDTA-Buffer) or TBE (TRIS-Borate-EDTA-Buffer) with a concentration of 1 %. Here, also the intercalating dye, needed for visualisation of the DNA, can be directly added (e.g. HD-green or Gel Red).

1. Heat the gel bottle (including agarose powder, buffer and intercalating dye) in a microwave until no particles are visible in the liquid anymore and pour the liquid agarose into the gel chambers.
2. Add combs directly afterwards (will create wells in which the DNA can be pipetted). Let the gel stand until it is fully solid again.
3. Add 1X TAE or TBE buffer (same as the buffer in which the agarose was dissolved) to the gel chamber and remove combs.
4. Mix 3 μl of DNA with 2 μl loading dye by pipetting up and down.
5. Load 5 μl DNA/dye mix into the wells on the gel. Make sure not to pierce the gel wells, but also that the pipette is deep enough into the pocket so that DNAmix will not diffuse into the buffer.
6. Add 1.5 μl of 1 kb ladder (includes loading dye) into a separate well on the gel
7. Apply current (120 V) and let the DNA migrate for 30 mins to the positive electrode (anode).

The gel can be visualized using a UV table (see Fig X for an example).

Methods

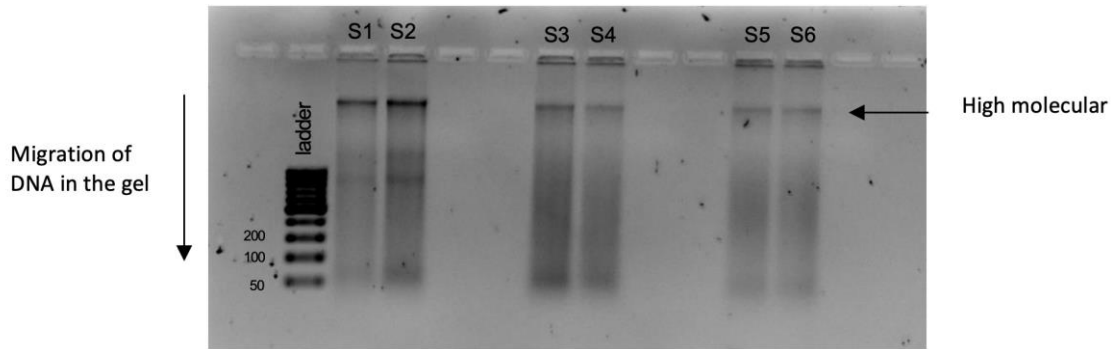


Fig. X: Agarose gel showing DNA molecules of six samples after extraction. Strong bands are visible for extremely long fragments. For other size categories a light smear is visible indicating also smaller fragments with low concentrations.

Qubit measurement

1. Prepare a premix of Qubit buffer (High sensitivity Kit) and dye according to the following calculation:

$$\text{BR dye} = \text{Number of samples} + 2 \text{ standards} + 1 \text{ extra} = X (\text{---}) \mu\text{l}$$

$$\text{Buffer} = X (\text{---}) \mu\text{l} * 199$$

2. Mix premix thoroughly.
3. Prepare samples for measurement in provided 0.5 ml Qubit tubes:

DNA samples: **199 μl** dye/buffer mix + **1 μl** DNA (each sample separately)

Standards Qubit : **190 μl** dye/buffer mix + **10 μl** DNA (standard 1 and 2)

4. Vortex all samples and let them incubate for two minutes.
5. First measure standards following the instructions on the device (choose correct kit).
6. Read samples; press "calculate stock", set Qubit to 1 μl DNA and ng/ μl .
7. Note the concentration for each sample.

2.3 Two-step Polymerase Chain Reaction (PCR)

The PCR is used to amplify a short, exactly defined DNA fragment (called amplicon) to obtain millions of copies of this fragment. The fragment can be a whole gene, part of a gene or a non-coding DNA sequence. The amplification is critically important before sequencing.

For the process of PCR different ingredients are needed:

1. Original DNA template.
2. Two primers (forward and reverse) are needed for amplification of the target sequence on both DNA strands from a distinct starting to a distinct end point.
3. DNA-Polymerase to replicate the specified section.
4. Deoxyribonucleotide triphosphates (dNTPs) which are the bricks for the strands synthesised by the polymerase.
5. A buffer solution which is specific for the used polymerase and ensures optimal conditions for the enzymatic activity.
6. Magnesium (Mg^{2+}) which is essential for the function of the polymerase and for the stabilization of primer bonds to the targeted region, sometimes already included in the PCR buffer.

The whole process of the PCR is conducted in a thermocycler, which is heated or cooled to the optimal temperatures for the process (see protocol below), including three main steps:

1. **Denaturation:** heating up of the double strand DNA up to a temperature of 94-96 °C to break up the hydrogen bonds between the strands and produce single strands.
2. **Annealing:** temperature is lowered for approx. 30 seconds to a specific value optimal for primer attachment on the single strands (primer specific). The exact temperature is set through the length and sequence of the primers (approx. 5-10 °C below melting temperature).

Methods

3. **Elongation:** the polymerase fills up the strands with free nucleotides from the 3' end of attached primers to the 5' end copying the original strand.

To prepare the samples for high-throughput sequencing, more steps are needed than e.g. for single specimen barcoding (Sanger-Sequencing), where the PCR is only used to amplify the target region and samples are sequenced separately. For high-throughput sequencing (here specifically Illumina sequencing), additional steps necessary to enable binding of the fragments on the flow cell and to tag samples with indices, so that samples can be pooled for sequencing and bioinformatically identified later after sequencing. Therefore, sequencing primer binding sites and flow cell adapters as well as sample-specific index sequences need to be added to the target fragment. Several methods have been developed for this purpose, e.g. 1-step PCR, 2-step PCR or ligation-based tagging approaches (see [Zizka et al. 2019](#) for details). Here, we will describe the commonly used 2-step approach in more detail. Also, here different strategies for primer design exist, but we will focus here only on one option.

In the first step, universal primers designed for the target group are used to amplify the specific target fragment (see [Table X](#) for commonly used primer pairs for different taxa). These primers further comprise the sequence of the Illumina sequencing primer (universal tail) and a shift of 0 – 3 not specified bases (N-shift, Fig. [Y](#)). The universal tail is needed as a template for the 2nd step PCR in which individual short sequences (tags) are added specific for each sample. The N-shift is needed, so that during sequencing not all amplicons start with the same base, because this could reduce the sequencing quality. This 1st-step PCR is usually run for 15 – 25 cycles. Primers for the 2nd step (Fig. [Y](#)) are complementary at the universal tail and further contain the sample-specific Illumina-Index sequence (6 or 8 bp long) and the flow cell binding sequence. These three parts together are called Illumina adapters and are needed for Illumina sequencing. The sample-specific index sequences allows for pooling of different samples on one sequencing lane and the subsequent bioinformatic assignment of the retrieved sequences to each of the original samples based on the index combination. Here, it is recommended to use the same index in both forward

Methods

and reverse primer to identify potential tag-switching. This 2nd-step is run for 25 – 45 cycles depending on the starting concentration of the samples.

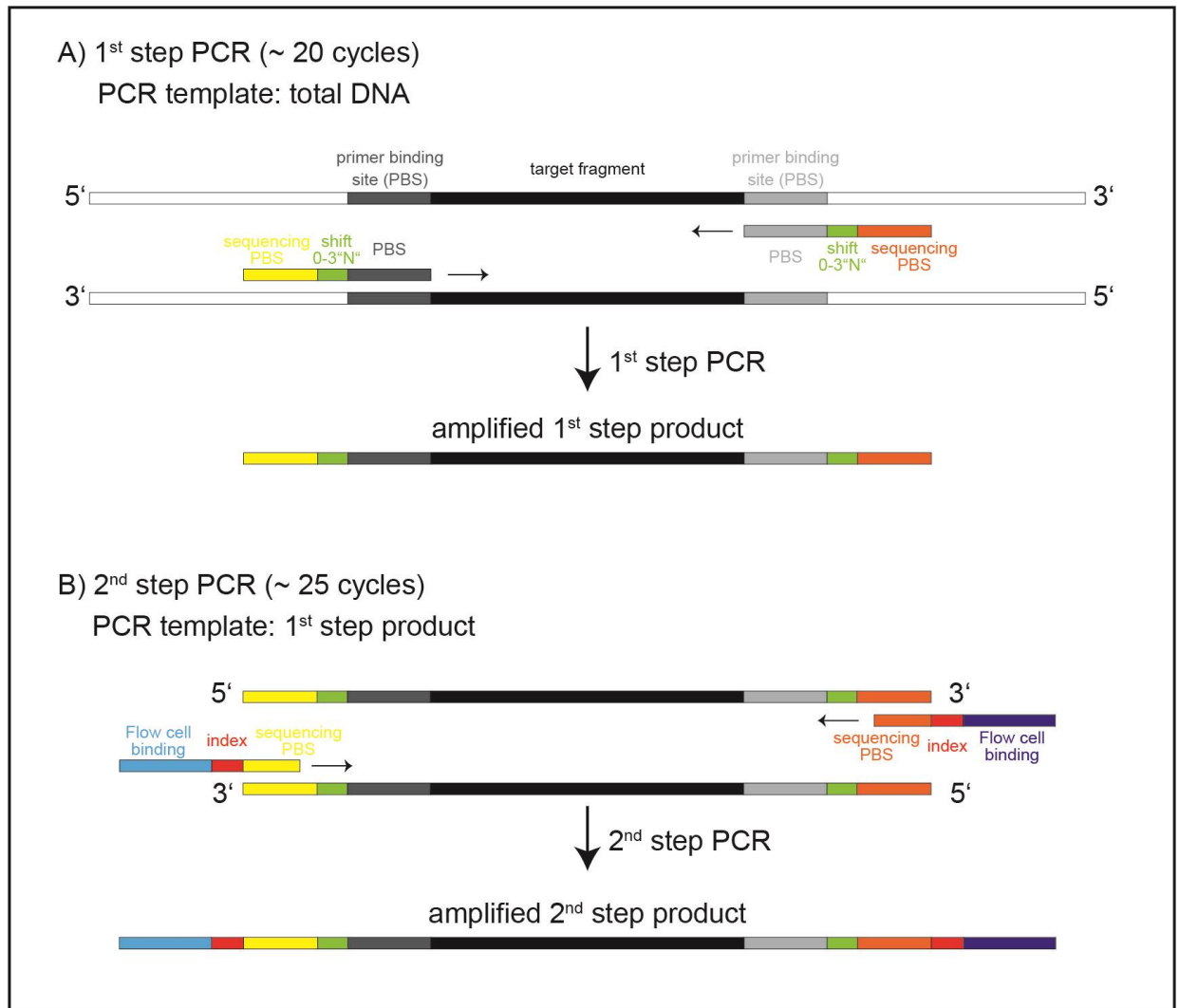


Fig. X: Scheme of the 2-step approach. 1st step primers are the same for all samples, 2nd step primers are sample-specific differing in the index enabling later multiplexing. Flow cell binding, index and sequencing primer binding sites are called adapters and need to be included to enable Illumina sequencing.

Depending on the DNA polymerase used, the PCR needs to be set up on ice or cooling racks. We will here describe the PCR reaction using the Qiagen Multiplex PCR Plus Kit

Methods

(Qiagen, Hilden, Germany), but it is similar for all Taq polymerases (specific protocols are always delivered with the respective polymerase). For the 1st-step PCR, we will first prepare a premix, containing Multiplex PCR Master Mix (consisting of Buffer, Taq polymerase and dNTPs), forward and reverse primer, CoralLoad Dye (for later visualization on an agarose gel) and PCR water. The volume of the reaction can be adjusted for the specific needs. Often 25 μ l reaction volumes are used, but 10 μ l can be sufficient. We here describe a 25 μ l reaction, as it will enable an easier normalization of samples for pooling later. Stock and target concentrations and needed volumes per sample are given in Table X. After preparing the premix, it is distributed in all wells and DNA is added to the reaction, e.g. 9 μ l premix and 1 μ l DNA. After vortexing and spinning down, the reaction can be placed in the thermocycler, running the PCR program, where the annealing temperature is dependent on the primer used.

Table X: Ingredients for the 1st step PCR.

Ingredients	Stock concentration	Target concentration	Volume per sample (in μ l), total V = 10 μ l
ddH ₂ O			8.5
2x Multiplex PCR Master Mix	2x	1x	12.5
1 st step forward primer	10 μ M	0.1	0.25
1 st step reverse primer	10 μ M	0.1	0.25
CoralLoad Dye	10x	1x	2.5
DNA	variable		1

Exemplary PCR program:

- 1) Initial denaturation 95°C for 5 min
- 2) 20 cycles of each
 - a. Denaturation 95°C for 30 s
 - b. Annealing 50°C – 65°C for 90 s (depending on the primer)
 - c. Extension 72°C for 1 min
- 3) Final extension: 68°C for 10 min

The second step will be conducted similarly, but here primers cannot be added to the premix but must be added separately to each sample as they contain the individual index sequence and the 1st-step PCR product is used as input for the PCR.

The success of the 2nd PCR can be checked on an agarose gel similar as described for DNA extraction. Here, 1 – 3 µl PCR product can be loaded directly on the gel as the loading dye is already included in the PCR reaction. Do not forget to also pipette ladder in one well in order to compare the size of the PCR product to a predefined standard.

2.4 Normalization of PCR products and pooling

In order to purify the amplicon, and to normalize PCR product concentrations between samples prior to pooling, the SequalPrep™ Normalization Plate Kit can be used which works with a limited binding capacity solid phase. Each well of the plate can bind ~ 25 ng of PCR amplicon. Following the normalization, the eluted PCR products can be pooled for sequencing. The SequalPrep™ Normalization Plate kit works by using a switchable surface charge (ChargeSwitch® Technology) depending on the pH of the surrounding buffer to facilitate nucleic acid purification. Under low pH conditions, the positive surface charge of the ChargeSwitch® coating binds the negatively charged nucleic acid backbone. Proteins and other contaminants (such as short oligonucleotide primers) are not bound and

Methods

therefore simply washed away. The lab workflow consists of adding the PCR products to the normalization plate wells, mixing them with binding buffer, incubation (binding step), clean-up with washing buffer (washing step), and elution of PCR products (elution step):

Binding Step:

1. Transfer **25 µl PCR product** into wells of Normalization Plate.
2. Add an equivalent volume of **Binding Buffer** to the PCR products (**25 µl**) and **mix** completely by pipetting up and down.
3. **Incubate** plate for **1 hour** at **room temperature** to allow binding of DNA to the plate surface.

Washing Step:

1. Aspirate liquid from wells. Make sure not to scrape the well sides during aspiration.
2. Add **50 µl Wash Buffer** to the wells. **Mix** by pipetting up and down twice.
3. Completely aspirate the buffer from the wells and discard it. To ensure complete removal, you may need to invert and tap the plate on paper towels. A small amount of residual Wash Buffer (1-3 µl) is typical and does not affect the subsequent elution or downstream applications.

Elution Step:

1. Add **20 µl Elution Buffer** to each well of the plate (do not use water for elution!).
2. **Mix** by pipetting up and down 5 times and ensure that the buffer contacts the entire plate coating (up to 20 µl level).
3. **Incubate** at room temperature for **5 min**, transfer and pool the purified DNA to a new Eppendorf tube.

Instead of using a Normalization plate, it is also possible to measure the concentration e.g. with a Qubit Fluorometer, calculate the volume needed for an equal amount of PCR product for all samples and pool accordingly. However, here it is important that either no primer dimer is present, or that the amount is similar between samples, as only the total

Methods

concentration can be measured with the Qubit. Another option would be to conduct a left-sided size selection (works similar to the described bead extraction and will exclude small fragments) prior to measuring or measuring the samples on e.g. a Fragment Analyzer, that not only determines DNA concentrations but also DNA concentrations per fragment size.

In any case, after pooling the volume of the library pool needs to be reduced either using a vacuum centrifuge or a spin-column purification kit. Before the library can be sequenced, a last quality check should be conducted e.g. on a Fragment Analyzer to see if no primer dimer is present, because they can otherwise lower the sequencing depth of the targeted fragment. An exemplary picture of a Fragment Analyzer profile is shown in Fig X.

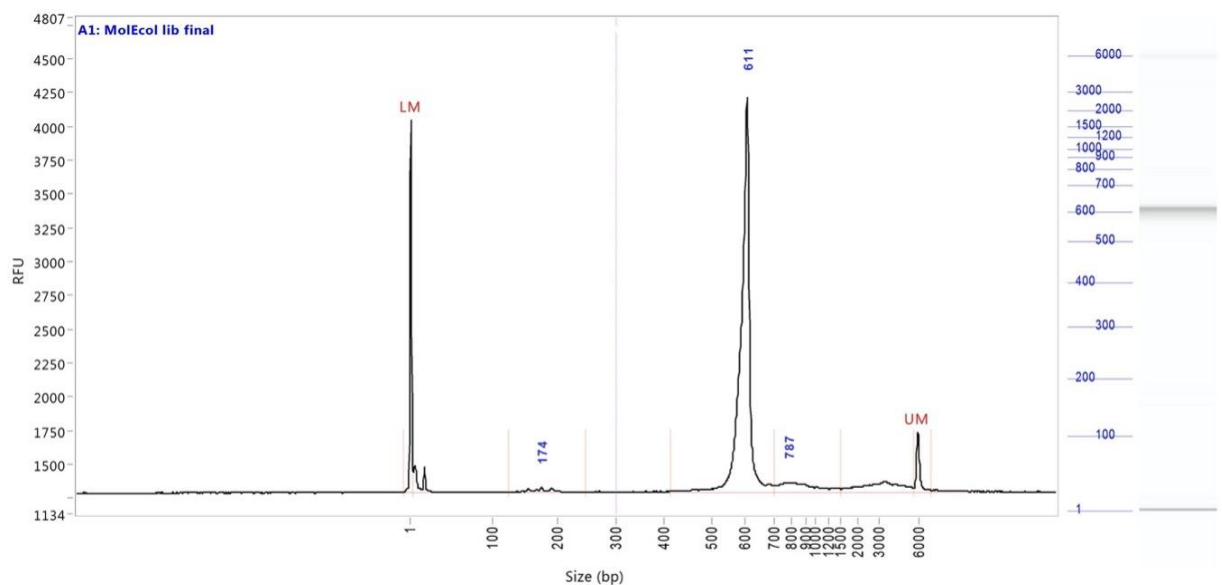


Fig X: Fragment Analyzer profile of a DNA library. Primer dimer (length around 174 bp) are minimal and library concentration was with 5.7 ng/μl of the target fragment (611 bp peak) high enough for sequencing.

If quality is sufficient, the final library can now be sequenced on a Illumina Sequencing device, e.g. MiSeq or HiSeq depending on the required sequence length which depends on the length of the target fragment. The length should be chosen in a way that there will be a sufficient overlap (XY bp) between forward and reverse read obtained by paired-end sequencing, which will be described in more detail in the following paragraph. In the

presented case here (Fig. X), the fragment size is on the upper range of feasible fragment length and would need to be sequenced on a MiSeq with 2x300 bp length (paired end).

2.5 Illumina sequencing

After we have prepared our libraries by amplifying our target fragment and adding Illumina adapters consisting of P5 and P7 flow cell binding sites, index sequences and sequencing primer binding region, sequencing can be started. Here, the first step is cluster generation, followed by sequencing and raw data analysis. See also e.g. this Illumina video - <https://www.youtube.com/watch?v=fCd6B5HRaZ8> - to get an overview over the different steps, or this video - <https://www.youtube.com/watch?v=olJaA6h2bFM> - to learn more about details. There are different sequencing platforms, e.g. MiSeq, HiSeq, NextSeq and NovaSeq which vary in details at different steps, but the general principle is similar. Here we will focus specifically on the MiSeq technology, because longer reads (up to 300 bp) can be generated and are thus often used for metabarcoding. Sequencing takes place on a flow cell which is a glass slide with different lanes and each lane is coated with a lawn of oligos which are complementary to the flow cell bind (P5 and P7) added to our sequences during library preparation. The first step of sequencing is cluster generation, where a cluster is a group of clonal DNA strands which are positioned in close proximity to each other. Each cluster contains up to 1000 identical copies of the same DNA strand in a 1-2 micron spot. This cluster generation is needed to enhance the signal emitted in the sequencing process for each strand to be detectable in the images taken. For details, especially on the different index systems see the Indexed Sequencing Overview Guide provided by Illumina. In the following the different steps during the sequencing process are described and visualized in Fig. X.

Cluster generation

1. To obtain single strands, libraries are denatured as a first step.

Methods

2. Single strands are hybridized to the oligos on the lane and 3'-extension is realized by a DNA polymerase. Here, flow cell oligos act as primers for the synthesis of the new strand.
3. In the next step, double stranded molecules are denatured and the original template is washed away. The newly synthesized strand now binds covalently to the flow cell surface.
4. To generate clonal clusters, the next step is the bridge amplification. Here, the single stranded DNA molecule bends over and forms a bridge by hybridizing to an adjacent complementary oligo on the flow cell. This hybridized oligo on the flowcell then acts as a primer for the polymerase which now extends the strand to form a double stranded bridge.
5. This bridge is now denatured, leading to two copies of covalently bound single stranded templates, a forward and a reverse strand bound either at the P7 end (forward strand), or at the P5 end (reverse strand).
6. The process of bridge amplification is now repeated with both strands forming a new bridge, hybridizing to the complementary oligo acting as primer for synthesis of a complementary strand and denaturation, resulting in four copies, half of them forward and half of them reverse strands. This process is repeated multiple times until a cluster with approx. 1000 identical sequences is generated and ends with the denaturation of the double stranded bridges.

Sequencing by synthesis (SBS) – 1st read (forward read)

1. The strands in the cluster are now bound in both, forward and reverse direction to the P7 and the P5 adapters. Sequencing of DNA fragments (1st and 2nd read for paired-end sequencing) will always start from the 3'-end, so the top of the sequence, which is not bound to the flow cell oligos. So if both, the forward and the reverse strand would be sequenced simultaneously, two different signals would be emitted at the same time and no clear signal could be generated. Therefore, only strands with the same orientation can be present at the flow cell, when sequencing starts. In the first step, we want to sequence the forward strand, i.e. the ones bound to the P7 adapters, so that the P5 end is free for priming the

Methods

sequencing. Hence, as a first step, reverse strands are cleaved and washed away, leaving only clusters in forward direction bound on the flow cell.

2. To prevent further bridge amplification and priming from the hybridized P5 adapter, 3'-ends are blocked for forward strands and flow cell oligos.
3. Sequencing can now start with binding the first sequencing primer to the strands. For all Illumina platforms, sequencing is performed with the SBS technology. Depending on the platform, there are different chemistries, differing in the number of dyes used, but we will here focus on the four channel, or four dye chemistry used on MiSeq and HiSeq platforms. Here, each of the four fluorescently labeled DNA bases (fl-NTPs) emits an intensity of a unique wave length, collected in 4 images per cycle.
4. After binding the sequencing primer, SBS can start: In each cycle, all four fl-NTPs and a polymerase are added simultaneously. The matching base is incorporated into the complementary strand. All fl-NTPs include a terminator that prevents a second NTP from binding in the same cycle. Unused NTPs are washed away.
5. A light source is used to excite the newly added base and a series of four images is taken for each cluster in each wavelength (4 images per cycle).
6. After images have been taken the base is deblocked by cleaving the terminator and the fluorescent dye from the incorporated fl-NTP.
7. With this the next base can be incorporated after adding again all 4 fl-NTPs. The steps described before are repeated until the number of cycles for the desired read length (e.g. 300 cycles for a 300 bp read) are done. As a last step, the sequencing products are washed away.

Sequencing by synthesis (SBS) – 1st index read (forward read)

1. When using single or dual indexing, in the next step the i7 index read is generated. For this step, the i7 index primer is annealed to the same template strand.
2. The index read is then generated using the same SBS technology as described before for up to 20 cycles. However, normally indexes of 6 or 8 bp are used, so only this respective number of cycles is needed.

Methods

3. After the required cycles are finished, the read product is washed away and the 3'-ends of the template strand and lawn oligos are unblocked

Sequencing by synthesis (SBS) – 2nd index read (forward strand workflow)

1. Before generating the Index 2 (i5) read, the template strand anneals to the P5 oligo attached to the flow cell. The P5 oligo now acts as a primer for the index 2 read, which is started by seven chemistry-only cycles (no imaging occurs), followed by up to 20 cycles of sequencing. Because the sequencing occurs before the bridge amplification, the 2nd index read is still read in forward direction.
2. The index read product is washed away and the forward template strand is used to generate the complementary reverse strand similar as described for bridge amplification during cluster generation. The bridge is denatured and the forward strands are cleaved off and washed away, enabling sequencing of the reverse strand.
3. For other platforms, e.g. HiSeq X, a reverse complement workflow is used. Here, the bridge amplification is conducted before the i5 index read is generated. Here, instead of using the P5 oligo attached to the flow cell as primer, the index 2 (i5) sequencing primer is annealed and up to 10 cycles of sequencing are performed to generate the i5 index read. This workflow does not require the seven additional chemistry-only cycles. The read is generated in the reverse direction. After completing the required cycles, the read product is removed.

Sequencing by synthesis (SBS) – 2nd read (reverse read)

1. The 3'-end of reverse template and lawn oligos are blocked to prevent unwanted priming and the Read 2 primer is annealed to the reverse strand.
2. Now the sequencing can start as described for the first read, ending after the desired number of cycles.

Processing of raw reads

1. The data analysis starts with identifying the location of all clusters on the images.
2. Then the intensity of the signal is extracted and translated into base calls and Quality scores (Q scores). The Q score is based on the Phred scale and indicates the chance that the base call is wrong. A Phred score of 30 means that the probability of an incorrect base call is 1 in 1000, so a base call accuracy of 99.9% and a score of 40 stands for the probability that 1 in 10000 bases is incorrect, resulting in a base call accuracy of 99.99%. These scores can be used later in the analysis for filtering.
3. As a last step, FASTQC files are generated, which will be used in further data analysis steps.

Methods

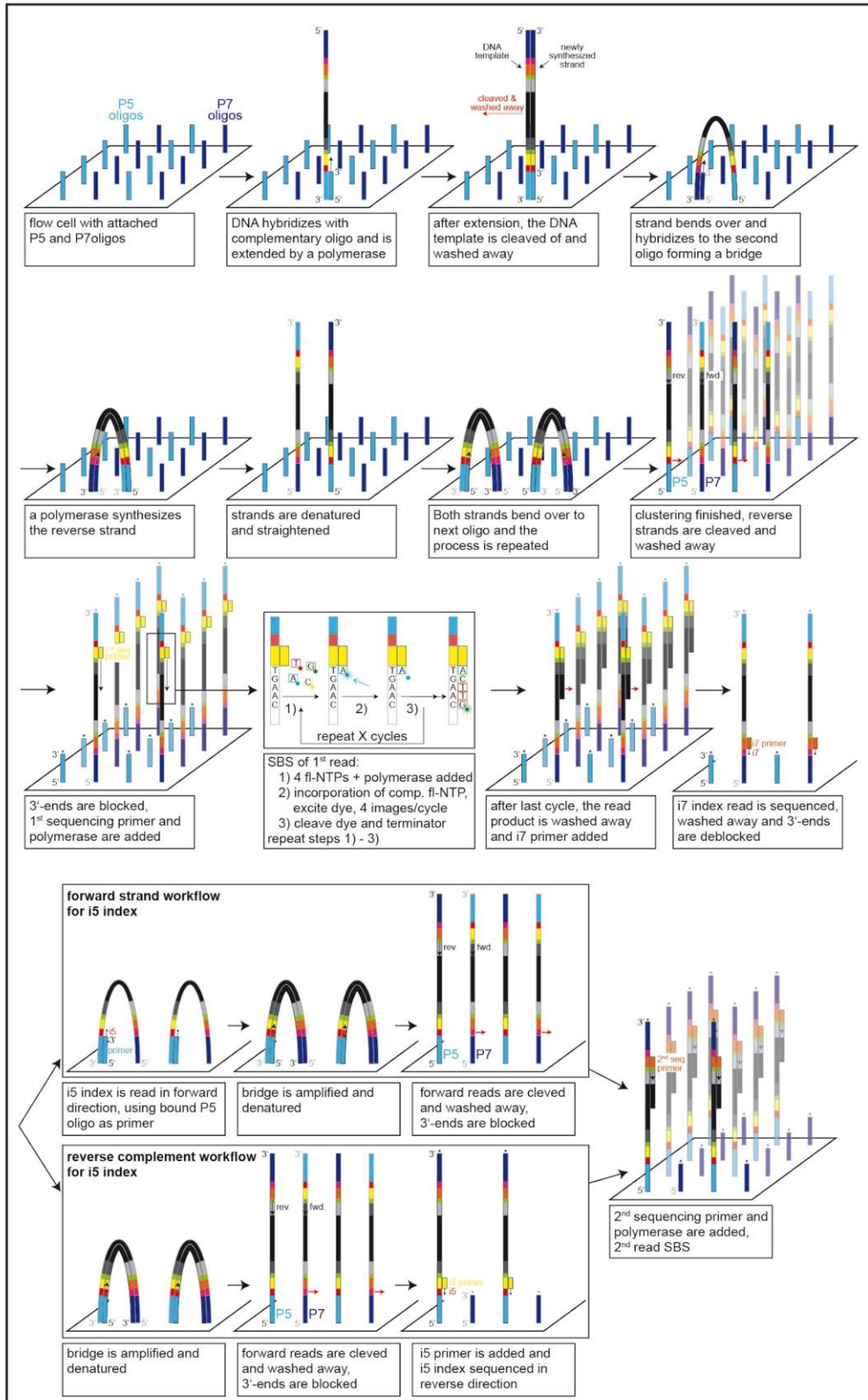



Fig y: Schematic illustration of the Illumina sequencing workflow

2.6 Data analysis of sequence data

In the tagging approach described here, we used Illumina index reads and will therefore receive the data already demultiplexed according to the respective index, so two fastq files per sample, one for the forward (indicated with “_1.fastq”) and one for the reverse read (“_2.fastq”). There is also a possibility to use further inline tags, which are then included in the received sequences. Here, an additional demultiplexing step using the inline tags is needed. However, the first step when receiving data is to check the integrity of the downloaded files and the general quality of the sequencing data. To check if data are corrupted, you can compare the md5 sums of the original and the downloaded file, which should be identical. To check the sequencing quality of the raw reads, the program FastQC  can be used. When running this program, you will receive different general information such as number of reads, average read length, per base sequence quality, per sequence GC content, per base N content (percent of bases at each position with no base call), sequence duplication level, overrepresented sequences, adapter and kmer content. To help interpreting the results, e.g. this tutorial can be useful: <https://rtsf.natsci.msu.edu/genomics/tech-notes/fastqc-tutorial-and-faq/>, where it is explained how the different sequencing summary statistics should look like.

Per base sequence quality

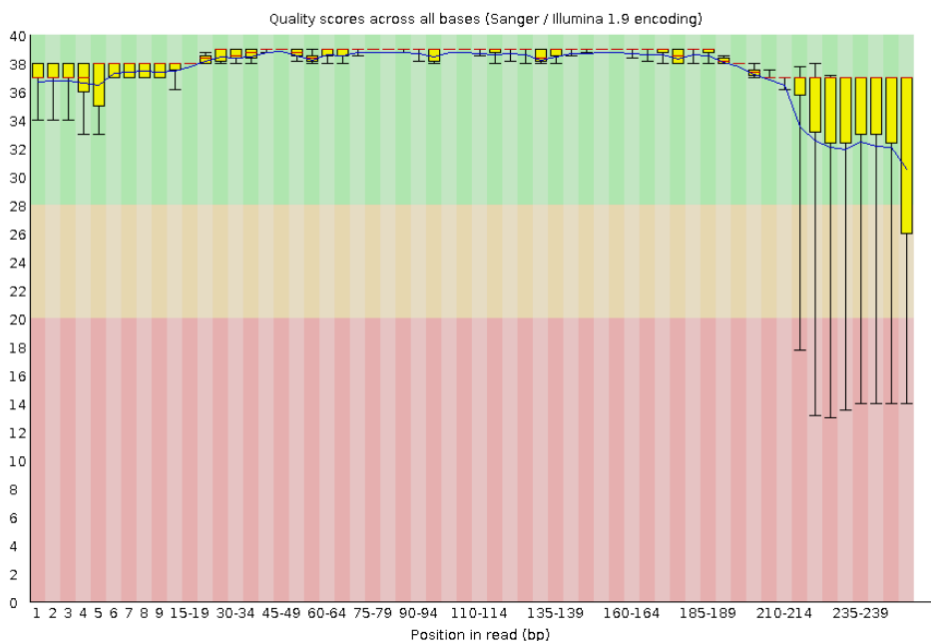


Figure Y: Analysis of your sequence data using the software FastQC allows checking if sequence quality was sufficiently good. In this example, sequence quality dropped towards the end of the sequence read but average quality was high (y-axis). The y-axis Phred score lists the error probability per base pair and is 10% for a score of 10, 1% for a score of 20, 0.1% for score of 30.

If a sequencing run produced sequences with in general sufficient quality, you can proceed with data analysis. Here, many pipelines for metabarcoding analysis exist, but for all the general workflow is similar. First, different quality filtering and preprocessing steps are required: paired-end sequences need to be merged into a continuous sequence, primer regions of the sequences need to be removed as well as sequences with a low quality or ambiguous bases and sequences which do not have a sufficient length. Another important step is to remove possible chimeras, i.e., sequences that originate from more than one parental sequence, which can e.g. originate from PCR amplification, when the extension of an amplicon is aborted and then serves in the next PCR cycle as a primer.

After quality filtering, the next step is the clustering of the remaining sequences according to their similarity. Here, different approaches exist, but generally sequences are either clustered into operational taxonomic units (OTUs) or unclustered amplicon/exact sequence variants (ASVs/ESVs) are used. For OTU clustering, often a threshold of 3% differences between sequences is applied based on the general barcoding concept (e.g., using USEARCH (Edgar, 2010) or VSEARCH (Rognes et al., 2016). However, this threshold is not valid for all species or taxonomic groups. Here, another approach is e.g. SWARM (Mahé et al., 2014), where the OTU clustering can be conducted using a dynamic (=adaptive) threshold.

Taxonomic assignment and reference databases

After clustering sequences into OTUs or assigning them to ASVs, taxonomy needs to be assigned to the respective entities by comparing sequencing data against (reference)

Methods

databases containing units of sequence and taxonomic information. There are several databases containing information on both different biota and different genetic markers such as as NCBI Genbank (<https://www.ncbi.nlm.nih.gov/genbank/>) and the the Barcode of Life Datasystems (BOLD) database (<http://www.boldsystems.org/>) or more specialised databases such as Diat.Barcode (<https://www6.inrae.fr/carrtel-collection/Barcoding-database>) for diatoms. In the process of taxonomic assignment, OTUs or ASVs are compared against the reference database resulting in similarity values of queried sequences to sequences stored in the database. Based on the highest similarity values of sequences in the database, a taxon can be assigned to the queried sequence. As it is often not feasible to manually assign every single OTU to a taxon, a taxonomic assignment step is carried out automatically in many metabarcoding pipelines or separate programs. For instance, the BOLDigger program (<https://github.com/DominikBuchner/BOLDigger>) automatically queries metabarcoding datasets of hundreds or thousands of OTU sequences against the BOLD database and downloads the “Top 20 Hits” for every OTU sequence including information on the taxonomy (phylum, class, order, family, genus, species and subspecies) as well as the respective similarity values. As a subsequent filtering step, the pipeline assigns different taxonomic levels (species, genus, family, order) to OTUs based on the similarity of reference sequences (98%, 95%, 90%, 85%) and takes the most common name in case of conflicting results (JAMP pipeline criteria, <https://github.com/VascoElbrecht/JAMP>). Additionally, BOLDigger offers the option to flag suspicious hits, e.g. when reference sequences with high similarity values but two or more different species names are present, or in case the most likely assignment result is represented by only one reference sequence. While this approach of taxonomic assignment is straightforward and fast, it can be inaccurate in certain cases. Keep in mind that bioinformatic processing, including automatic taxonomic assignment, does not replace carefully verifying and checking assignments and interpreting the data. Note that you will almost always find non-target taxa in your metabarcoding data. In case you study aim is to analyse one specific taxon or biological quality element (e.g., macrozoobenthos), all found non-target taxa (in this example phytobenthos or fish) should be discarded for downstream analyses to not incorrectly inflate estimations on richness or other metrics.

Methods

The taxa list resulting from taxonomic assignment can be used for further data analysis, for example, to compare species composition between sites or monitoring the changes in community due to stressor impacts. It can also be directly used for bioassessment by transforming it into the respective format need for EQR calculation, yet several aspects need to be considered (see below and Hering et al., 2018).

Basic statistics

As a first insight into your metabarcoding data, directly after the taxonomic assignment, overview statistics can be explored. A common strategy in metabarcoding projects is to achieve about equal numbers of sequence reads per sample by equimolar pooling of all samples into one DNA library for sequencing. As there are many confounding factors (e.g., inaccuracy in pipetting or DNA quantification) it is useful to compare the achieved sequencing depth per sample and to check for potential (unexpected) outliers (Fig. X).

Figure X: Overview of sequence number per sample obtained after sequencing. Not all samples are sampled with the same depths. With rarefaction / subsampling approaches it is possible to test if for all samples a sufficient number of sequences reads was obtained.

In case PCR replicates were used to account for PCR stochasticity, the number of reads per OTU/ASV of replicates can be checked for consistency. As a conservative approach, only OTUs/ASVs with reads in both replicates of a sample should be used for downstream analyses. A subsequent and straightforward step is to check your data for OTU/ASV occupancy (number of samples an OTU/ASV was found in) and OTU/ASV richness (how many OTUs/ASVs per sample). As metabarcoding data usually contains more OTUs/ASVs than species (similarity threshold of many OTUs/ASVs will be below your chosen species similarity threshold), all analyses steps can be carried out with only OTUs/ASVs on species

level. In the same sense, all OTUs/ASVs bearing the same taxonomic name can theoretically be merged into one entry. While this is more conservative in some scenarios (e.g., a high similarity clustering threshold may lead to oversplitting of species into several entries and hence inflate richness estimations) it may limit the derived information (OTUs/ASVs without species names may very well reflect species) or incorrectly reflect the information (OTUs/ASVs with the same species name assigned but high intersequence distances may reflect cryptic species).

TaxonTableTools

The options for downstream analyses are plentiful and can be suited to individual requirements of a project. And while it is possible to have a first exploration of your data with simple table calculation tools, it is recommended to use readily available tools for more complex analyses and visualization steps. One such very useful tool, to quickly analyze and visualize the metabarcoding data, is the GUI-based program TaxonTableTools (TTT) (<https://github.com/TillMacher/TaxonTableTools>). TTT requires a taxonomy table and a read table which are combined for all subsequent analyses. Replicates can be checked for consistency and merged and taxa can be filtered before conducting basic statistics. As such, the number of OTUs/ASVs on different taxonomic levels can be inferred, the number of OTUs/ASVs per higher taxon counted and the number of OTUs per sample as well as the overlap of OTUs/ASVs between samples be analyzed. All mentioned steps can be visualised. TTT offers the possibility to perform rarefaction curves, analyze read proportions and site occupancies and generate Krona charts for visualising taxonomic compositions. It further offers several ecological analyses such as calculating alpha or beta diversity estimates and implementation of bioassessment calculations ("Perloides procedure", other procedures to be included in the future). See appendix for a full tutorial of TTT.

Bioassessment

After having obtained metabarcoding data the essential question is how these data can inform bioassessment. Two different approaches exist: i) fit DNA data and the taxonomic information to existing bioassessment schemes, such as those applied in the European Water Framework Directive, or ii) rebuild novel indices that do not rely on existing reference data (Pawlowski et al. 2018).

As The biggest drawback of (e)DNA metabarcoding is the lack of accurate abundance data as the resulting read abundance is biased by differences in primer binding efficiencies between different taxa and biomass (in case of homogenised bulk sample metabarcoding) or DNA shedding rates (in case of eDNA metabarcoding). One analysis strategy is to convert all results into binary incidence data (0 = absent, 1 = present). This will result in robust data for richness estimations, but poses a challenge for analyses that require abundance data, which is often the case in bioassessment procedures (e.g., WFD assessments).

Abb. Buchner et al. 2019; Beentjes et al. 2018.

However, even in case of used binary data rather than real abundance data, assessment results have been shown to be consistent (e.g., Beentjes 1,2; Buchner, Aylagas uswl.)

Furthermore, despite being biased, read abundance data holds some information on real abundance or otherwise having 100.000 or 1 read would be accepted as the same information.

Outlook in the line of Hering et al. 2018

Inklusive Machine Learning

Outlook

Übersicht Pawlowski Training Machine Learning - Bioassessment.

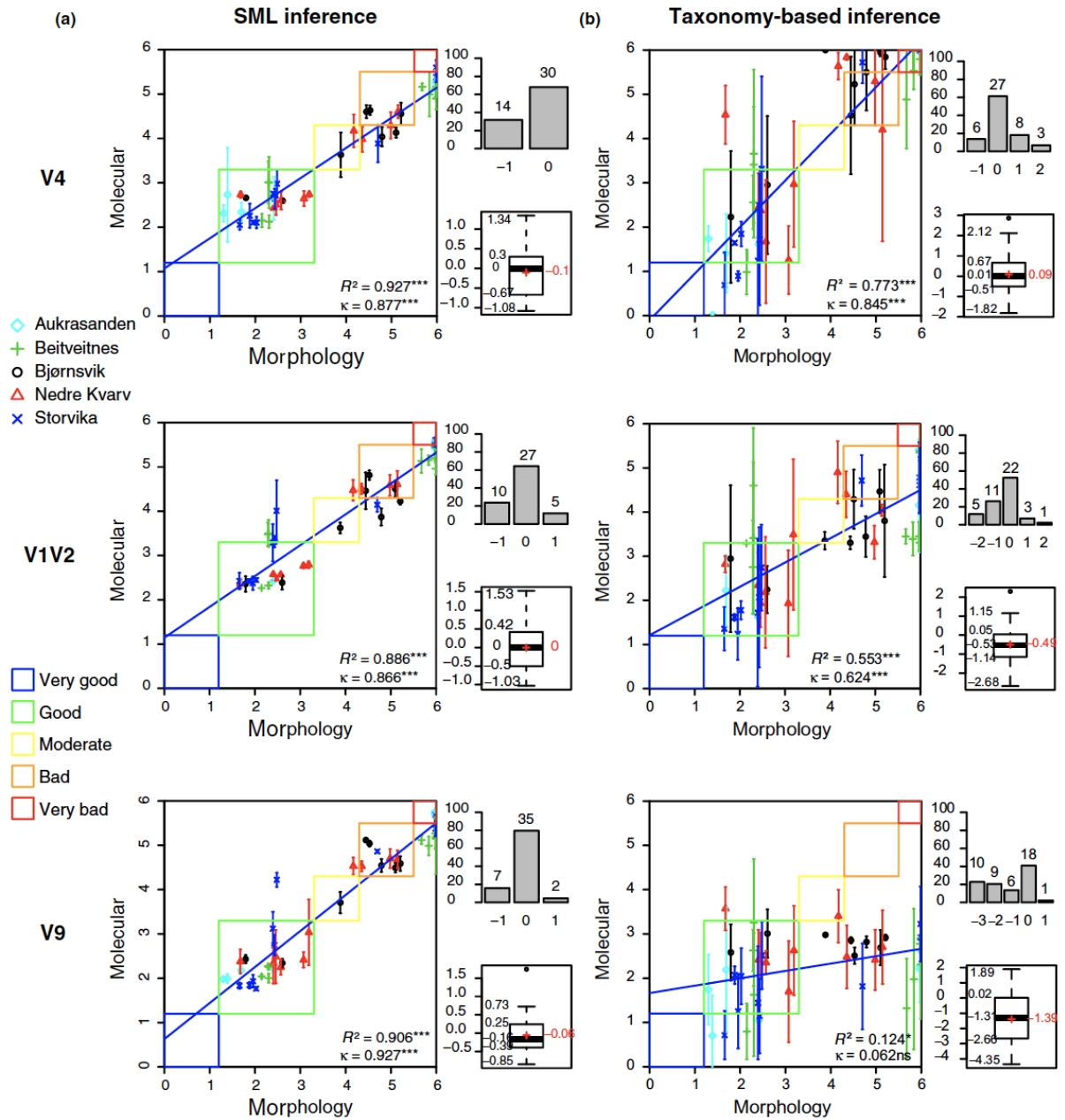


FIGURE 1 AMBI index predictions from three eukaryotic markers using (a) SML approach and (b) taxonomy-based approach. The reference AMBI values, as inferred by morphotaxonomic inventories, are on the horizontal axis, and the predicted ones, as inferred from the molecular data, are on the vertical axis. Coloured rectangles indicate the discrete quality classes, from blue “very good” to red “very bad.” The R^2 value of linear relationships between reference and predicted AMBI values, as well as the kappa value measuring the agreement between reference and predicted AMBI-derived ecological quality status, is indicated on each plot. Their significance is indicated on each plot as follow: ns $p > 0.05$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. Barplots on the top right side of each plot indicate the amount (number over bars) and percentage (y-axis) of correct classifications (0 on x-axis) and misclassifications. Boxplots on the bottom right side of each plot indicate the median and quartile values (black numbers) and the mean value (red number) of the difference between the predicted and the reference AMBI values [Colour figure can be viewed at wileyonlinelibrary.com]

Bibliography

- Baird, D. J., & Hajibabaei, M. (2012). Biomonitoring 2.0: A new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. *Molecular Ecology*, *21*(8), 2039–2044. <https://doi.org/10.1111/j.1365-294X.2012.05519.x>
- Birk, S., Bonne, W., Borja, A., Brucet, S., Courrat, A., Poikane, S., Solimini, A., van de Bund, W. V., Zampoukas, N., & Hering, D. (2012). Three hundred ways to assess Europe's surface waters: An almost complete overview of biological methods to implement the Water Framework Directive. *Ecological Indicators*, *18*, 31–41.
- de Kerdrel, G. A., Andersen, J. C., Kennedy, S. R., Gillespie, R., & Krehenwinkel, H. (2020). Rapid and cost-effective generation of single specimen multilocus barcoding data from whole arthropod communities by multiple levels of multiplexing. *Scientific Reports*, *10*(1), 1–12. <https://doi.org/10.1038/s41598-019-54927-z>
- Haase, P., Lohse, S., Pauls, S., Schindehütte, K., Sundermann, A., Rolaufts, P., & Hering, D. (2004). Assessing streams in Germany with benthic invertebrates: Development of a practical standardised protocol for macroinvertebrate sampling and sorting. *Limnologica-Ecology and Management of Inland Waters*, *34*(4), 349–365.
- Hebert, P. D., Cywinska, A., & Ball, S. L. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B: Biological Sciences*, *270*(1512), 313–321.
- Hebert, P. D. N., Braukmann, T. W. A., Prosser, S. W. J., Ratnasingham, S., deWaard, J. R., Ivanova, N. V., Janzen, D. H., Hallwachs, W., Naik, S., Sones, J. E., & Zakharov, E. V.

Methods

- (2018). A Sequel to Sanger: Amplicon sequencing that scales. *BMC Genomics*, 19(1), 219. <https://doi.org/10.1186/s12864-018-4611-3>
- Herbold, C., Pelikan, C., Kuzyk, O., Hausmann, B., Angel, R., Berry, D., & Loy, A. (2015). A flexible and economical barcoding approach for highly multiplexed amplicon sequencing of diverse target genes. *Frontiers in Microbiology*, 6, 731. <https://doi.org/10.3389/fmicb.2015.00731>
- Mahé, F., Rognes, T., Quince, C., Vargas, C. de, & Dunthorn, M. (2014). Swarm: Robust and fast clustering method for amplicon-based studies. *PeerJ*, 2, e593. <https://doi.org/10.7717/peerj.593>
- Pawlowski, J., Kelly-Quinn, M., Altermatt, F., Apotheloz-Perret-Gentil, L., Beja, P., Boggero, A., Borja, A., Bouchez, A., Cordier, T., Domaizon, I., Feio, M. J., Filipe, A. F., Fornaroli, R., Graf, W., Herder, J., van der Hoorn, B., Iwan Jones, J., Sagova-Mareckova, M., Moritz, C., ... Kahlert, M. (2018). The future of biotic indices in the ecogenomic 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>
- Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: A versatile open source tool for metagenomics. *PeerJ*, 4, e2584. <https://doi.org/10.7717/peerj.2584>
- Stein, E. D., Martinez, M. C., Stiles, S., Miller, P. E., & Zakharov, E. V. (2014). Is DNA Barcoding Actually Cheaper and Faster than Traditional Morphological Methods: Results from a Survey of Freshwater Bioassessment Efforts in the United States? *PLOS ONE*, 9(4), e95525. <https://doi.org/10.1371/journal.pone.0095525>

Methods

Stein, E. D., White, B. P., Mazor, R. D., Jackson, J. K., Battle, J. M., Miller, P. E., Pilgrim, E. M., & Sweeney, B. W. (2014). Does DNA barcoding improve performance of traditional stream bioassessment metrics? *Freshwater Science*, 33(1), 302–311.

Taberlet, P., Bonin, A., Zinger, L., & Coissac, E. (2018). *Environmental DNA. From biodiversity research to application*. Oxford University Press.

Taberlet, P., Coissac, E., Hajibabaei, M., & Rieseberg, L. H. (2012). Environmental DNA. *Molecular Ecology*, 21(8), 1789–1793.

Yu, D. W., Ji, Y. Q., Emerson, B. C., Wang, X. Y., Ye, C. X., Yang, C. Y., & Ding, Z. L. (2012). Biodiversity soup: Metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution*, 3(4), 613–623.

Appendix:

List of primers, software, chemicals that is useful

Co-funded by the
Erasmus+ Programme
of the European Union



The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Erasmus + Project No ECOBIAS_609967-EPP-1-2019-1-RS-EPPKA2-CBHE-JP

Development of master curricula in ecological monitoring and aquatic bioassessment for Western Balkans HEIs