Validation and development of freshwater invertebrate metabarcoding COI primers for Environmental Impact Assessment

Vasco Elbrecht*, Florian Leese¹²

¹Aquatic Ecosystem Research, Faculty of Biology, University of Duisburg-Essen, Essen, Germany
²Centre for Water and Environmental Research (ZWU) Essen, University of Duisburg-Essen, Essen, Germany

* Correspondence:
Vasco Elbrecht
Vasco.Elbrecht@uni-due.de

Keywords: DNA Barcoding, Primer development, primer evaluation, primer bias, ecosystem assessment, in silico PCR

Abstract

To assess and manage human impacts on freshwater ecosystems is a central challenge in the present era of biodiversity loss. Macroinvertebrates are an ideal group for Environmental Impact Assessment (EIA). However, generating accurate macroinvertebrate inventories is difficult and error-prone based on larval morphology. Here, DNA metabarcoding provides new opportunities. Its potential to accurately identify bulk invertebrates on species level has been demonstrated in several case studies. However, DNA based identification is often limited by primer bias, potentially leading to taxa in the sample remaining undetected. Thus, the success of DNA metabarcoding as an emerging technique for EIA critically relies on carefully evaluated primers.

We used the R package PrimerMiner to obtain and process cytochrome c oxidase I (COI) sequence data for the 15 globally most relevant freshwater invertebrate groups in EIAs. Using these sequence alignments, we developed four primers combinations optimized for freshwater macrozoobenthos. All primers were evaluated by sequencing ten mock community samples each consisting of 52 freshwater invertebrate taxa. Additionally popular metabarcoding primers from the literature and the here developed primers were tested in silico against the 15 relevant invertebrate groups.

The developed primers varied in amplification efficiency and the amount of detected taxa, yet all retrieved more taxa than standard ‘Folmer’ barcoding primers. Two new primer combinations showed more consistent amplification than a previously tested ribosomal marker (16S) and detected all 42 insect taxa present in the mock community samples. In silico evaluation revealed critical design flaws in some commonly used primers from the literature.

We demonstrate a reliable strategy to develop optimized primers using the tool PrimerMiner. The developed primers detected almost all taxa present in the mock samples, and we argue that high base degeneracy is necessary to decrease primer bias as confirmed by experimental results and in silico primer evaluation. We further demonstrate that not all primers currently used in metabarcoding studies are likely not suitable for amplification of insect and freshwater taxa. Thus careful primer evaluation and more region / ecosystem specific primers might be needed, before DNA metabarcoding can be used for routine EIA of freshwater ecosystems.
1 Introduction

Freshwater resources worldwide are threatened by anthropogenic activities, and with exponential increase of human population the pressure on these sensitive ecosystems will further increase (Dudgeon et al., 2005; Vörösmarty et al., 2010). World-wide, ambitious water monitoring, management and restoration projects have been launched in the last decades for Environmental Impact Assessment (EIA) and to protect and restore freshwater ecosystems (EU Water Framework Directive, US Clean Water Act). Macroinvertebrates are often a key component to monitor stream health, as many are sensitive to stressors. It is however critical to determine these indicator taxa precisely, as they can show different stressor responses even on genus level (Macher et al., 2015). However, accurate determination of freshwater invertebrates is often difficult on larval specimens, often leading to low taxonomic resolution and misidentification (Haase et al., 2010; Sweeney et al., 2011). This of course translates directly into a decreased accuracy of generated taxa lists, which might lead to false EIA and misguided management in the worst case (Stein et al., 2014). Additionally, identification accuracy is affected by taxonomist experience, limiting the comparability of assessments (Haase et al., 2010). With the decline of taxonomic expertise and much of the world’s diversity not even being properly described, morphology based monitoring cannot keep pace with current challenges of sustainable water management.

A promising alternative to morphological identification is DNA based determination of macrozoobenthic invertebrates, which has been demonstrated in multiple case studies (Carew et al., 2013; Elbrecht and Leese, 2015; Hajibabaei et al., 2011; Stein et al., 2013; Sweeney et al., 2011). A short fragment of a standardised genetic marker of adult insect taxa (which can often be determined reliably to species level) can be amplified, sequenced and stored in a reference database. The same reference database can then be utilised for the identification of larval specimens. The cytochrome c oxidase I (COI) gene is typically used for this DNA barcoding technique and extensive reference sequences is already available in online databases (Ratnasingham and Hebert, 2007; 2013). However, identifying single specimens using DNA barcoding is still quite expensive because each individual has to be processed and sequenced individually (Cameron et al., 2006; Stein et al., 2014). Recent advances in high throughput sequencing (HTS) made it possible to extract DNA and sequence the barcoding region from complete bulk samples often containing hundreds to thousands of specimens. This technique, coined DNA metabarcoding, has been already widely used to generate comprehensive taxa lists for many ecosystems and environments (Taberlet et al., 2012). However, the usability of DNA metabarcoding remains limited because of severe primer bias which prevents the detection of all taxa present in a sample and hinders quantification of biomass and abundances (Elbrecht and Leese, 2015; Piñol et al., 2014).

A universal barcoding primer pair, which amplifies a marker sequence of suitable length for HTS, is thus the most critical component to assess macroinvertebrate bulk samples with DNA metabarcoding. The COI barcoding gene region shows high codon degeneracy throughout its sequence, making the design of "truly" universal primers difficult (Deagle et al., 2014; Sharma and Kobayashi, 2014). Several universal COI barcoding primers have been developed of which many are now used or could be suitable for metabarcoding studies (Figure 1, e.g. (Brandon-Mong et al., 2015; Folmer et al., 1994; Geller et al., 2013; Gibson et al., 2014; Hebert et al., 2004; Leray et al., 2013; Meusnier et al., 2008; Shokralla et al., 2011; 2015; Van Houde et al., 2010; Zeale et al., 2010). However, often these primers were developed for a specific taxonomic group, purpose or ecosystem, as for example the (Zeale et al., 2010) primers originally developed for gut content analysis on bats, but now more widely used. Thus, despite being "universal" and often including several degenerate bases, metabarcoding primers typically recover only 80-90% or even less of the taxa present in a sample.
(Brandon-Mong et al., 2015; Elbrecht and Leese, 2015; Leray et al., 2013). Furthermore, many primers have never been thoroughly evaluated concerning primer bias and the proportion of undetected taxa, making development and testing of universal primers a pressing issue.

Additionally, details on primer design and/or used sequence data are often not described extensively (e.g. (Hajibabaei et al., 2011; Shokralla et al., 2015). Primers are often developed with available reference barcode sequences for the taxonomic target groups are downloaded from NCBI or BOLD and aligned (Gibson et al., 2014; Leray et al., 2013; Zeale et al., 2010) or alternatively only mitochondrial genomes or a small subset of barcoding sequences are used (Brandon-Mong et al., 2015; Deagle et al., 2014; Geller et al., 2013). These two approaches can be biased, as sequences for certain taxa will be overrepresented in big datasets (e.g. from population genetic studies), while only mitochondrial genome datasets might be insufficient.

To explore these primer development problems, we used the recently developed R package PrimerMiner to explore the suitability of existing primers for freshwater invertebrate metabarcoding using in silico evaluations. We further developed our own optimised primer sets, which we used to amplify ten mock communities which have been used for primer evaluation in previous studies (Elbrecht & Leese 2015; Elbrecht et al. 2016) each containing 52 freshwater invertebrate taxa.

2 Material and Methods

2.1 Primer development and in silico evaluation

The PrimerMiner package v0.7 was used to download COI and cluster sequences for the 15 most relevant freshwater invertebrate groups for biodiversity assessment (Accessed September 2016, table S2, (Elbrecht and Leese, 2016)). Sequences were aligned with MAFFT v7.017 (Katoh et al., 2002) as implemented in Geneious 8.1.7 (Kearse et al., 2012). 26 bp in the HCO and 25 bp in the LCO region were selectively trimmed using the "selectivetrim" function and the alignment for each group was visualized with PrimerMiner to manually identify suitable primer binding sites. Two forward (BF1, BF2) and two reverse primers (BR1, BR2) were developed with high base degeneracy. Fusion primers were generated by adding Illumina adapters and inline barcodes as described in (Elbrecht and Leese, 2015) to increase sequence diversity while sequencing and allow for a one step PCR protocol.

Primers were also used to evaluate all primers shown in Figure 1 against alignments of the 15 freshwater invertebrate groups, using the default "Position_v1.csv" and "Type_v1.csv" table for mismatch scoring (tables are included in the PrimerMiner example data). Primers matching with a penalty score of above 120 where considered not working.

2.2 Testing of DNA metabarcoding primers on mock communities

Amplification success of the BF / BR primers was evaluated using ten mock communities, each containing a set of 52 freshwater invertebrates used in previous studies (Elbrecht and Leese, 2015; Elbrecht et al., 2016). The identical DNA aliquot and one step PCR protocol as in (Elbrecht and Leese, 2015) was used for all four primer combinations, but the number of PCR cycles was increased from 30 to 35. As in the previous studies, each sample was uniquely tagged from both sides, but for half of the samples only 25 ng instead of 50 ng DNA was used in PCR (see Figure S1). For each primer combination all ten samples were run in the same PCR setup, using one PCR replicate per sample. Ready-to-load products were magnet-bead purified (left sided, 0.8x SPRIselect, Beckman Coulter, Bread, CA, USA) and quantified using the Qubit HS Kit (Thermofisher Scientific, Carlsbad,
CA, USA). For each primer combination, equimolar amounts of amplicons were pooled into one library (amplicon concentrations have to be adjusted as amplicon length does vary, see Figure S1). The library was sequenced on one lane of a HiSeq 2500 (rapid run, 2x250 bp) with 5% PhiX spike-in, carried out by the DNA Sequencing Center of Brigham Young University, USA.

Bioinformatic processing of HTS data was kept as similar as possible to previous studies (Elbrecht and Leese, 2015; Elbrecht et al., 2016). In short, reads were demultiplexed (script S1) and paired end reads merged using Usearch v8.1.1831 -fastq_mergepairs with -fastq_merge_maxee 1.0 (Edgar and Flyvbjerg, 2015). Where necessary, reads were converted into reverse complement. For each primer combination all ten replicates were pooled and sequences which were present only one single time in the dataset (singletons) removed prior to clustering with Usearch (cluster_otus, 97% identity, strand plus, includes chimera removal) (Edgar, 2013). Dereplicated reads for each of the 40 samples (including singletons) were compared against the respective OTU dataset, using usearch_global with a minimum match of 97% and strand plus. Like in previous studies, low abundance OTUs without at least one sample above 0.003% sequences assigned, were considered unreliable and excluded from the dataset. Taxonomy of the remaining OTUs was identified and manually verified using the BOLD and NCBI database. To ensure that the same taxonomy was assigned across primer combinations and the reference COI study (Elbrecht and Leese, 2015), the most abundant sequence for each OTU in each sample was extracted using an R script (Script S2) and the haplotype of all individual specimens assembled, if amplified by more than one primer combination.

3 Results

3.1 Developed primers using PrimerMiner

We designed four primer pairs (Table 1) using the alignments of 15 freshwater taxa relevant for bioassessment (Figure S2). The two BF and two BR primers show high base degeneracy to amplify as many insect taxa as possible. Amplified regions range from 217 bp for internal barcodes and up to 421 bp for combinations using a degenerated version of the HCO2198 primer (Figure 1). While samples in this study were tagged uniquely from both sides using fusion primers (Figure S3), the inline barcodes allow for tagging of up to 72 samples for each primer combination (see Figure S4 for recommended primer combinations).

Table 1: Newly developed universal primers targeting freshwater invertebrates relevant for assessment.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Direction</th>
<th>Primer sequence (from 5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>BF1</td>
<td>Forward</td>
<td>ACWGGWTGRACWGTNTAYCC</td>
</tr>
<tr>
<td>BF2</td>
<td>Forward</td>
<td>GCHCHGHAYATRGCHTYTC</td>
</tr>
<tr>
<td>BR1</td>
<td>Reverse</td>
<td>ARYATDGTRATDGCCHCDGC</td>
</tr>
<tr>
<td>BR2</td>
<td>Reverse</td>
<td>TCDGGRTGNCCRAARAAYCA</td>
</tr>
</tbody>
</table>

All four BF / BR primer combinations were tested on ten invertebrate mock community samples on an Illumina HiSeq sequencer. PCR efficiency varied across primer combinations, with PCRs involving the BF2 primer showing good amplification whereas those with the BF1 primer always showing decreased yields (Figure S5). Amplification efficiency with fusion primers was always lower than the positive control (standard COI Folmer primers, without Illumina tail, data not shown). Sequencing was successful for all samples, with very similar amounts of sequences obtained for all
replicates (on average 1.55 million reads per sample, SD = 0.2, Figure S1A). Cluster density on the lane was low (402 k/mm²) yielding only 48.74% of the expected sequencing output, yet with good sequence quality (Q30 ≥ 92.17%, raw data deposited on SRA: SRX1619153). The amplified read lengths had an influence on the amount of sequences retained in bioinformatic processing. Longer amplicons showed less overlap when PE merged and were thus excluded more often due to expected errors > 1 (Figure S1B). Additionally, for primer combinations that used the P5_BF1_2 primer more sequences were discarded than with other primer combinations, as ~1/5 of the reads had poor Phred scores (See Figure S1B). There were also issues with the BF1 and BF2 primers which showed insertions or deletions on the 3' end affecting total sequence length by 1-2 bp across all replicates (Figure S6). Some primer combinations also amplified up to 1.35% shorter or longer fragments than expected (Figure S7).

### 3.2 Amount of taxa recovered

All insect taxa present in the mock samples were detected with each primer combination, with exception of the BF1 + BR1 combination that failed to amplify the Scirtidae (Coleoptera) specimens (Table 2, raw OTU data table S3, haplotype sequences data Script S2). All primers failed for some of the other metazoan taxa, with the BF1 + BR2 combination showing the least amount of undetected taxa. In comparison to the traditional Folmer primers (Folmer et al. 1994), all BF / BR freshwater primers showed a more consistent and equal read abundance across the mock samples (Figure 2). As in Elbrecht et al. (2016), the standard deviation from the expected abundance and precision for the primer pairs was estimated, which summarizes the variance in amplification for each morphotaxon. The primer combination BF1 + BR1 showed the highest inconsistencies in read abundance, while the BF2 + BR1 and BF2 + BR2 combination showed even higher precision than a previously tested 16S marker (Elbrecht et al. 2016). The proportion of detected non-insect metazoan taxa varied between primer combinations, with the combination BF1+BR2 detecting all but one taxon.

<table>
<thead>
<tr>
<th>Taxonomic group</th>
<th>Number of specimens</th>
<th>LCO1490+HCO2198 7 (88%)</th>
<th>16S ins 8 (100%)</th>
<th>BF2+BR2 8 (100%)</th>
<th>BF2+BR1 8 (100%)</th>
<th>BF1+BR2 8 (100%)</th>
<th>BF1+BR1 8 (100%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plecoptera</td>
<td>4</td>
<td>4 (100%)</td>
<td>4 (100%)</td>
<td>4 (100%)</td>
<td>4 (100%)</td>
<td>4 (100%)</td>
<td>4 (100%)</td>
</tr>
<tr>
<td>Trichoptera</td>
<td>15</td>
<td>13 (86%)</td>
<td>15 (100%)</td>
<td>15 (100%)</td>
<td>15 (100%)</td>
<td>15 (100%)</td>
<td>15 (100%)</td>
</tr>
<tr>
<td>Diptera</td>
<td>8</td>
<td>7 (88%)</td>
<td>7 (88%)</td>
<td>8 (100%)</td>
<td>8 (100%)</td>
<td>8 (100%)</td>
<td>8 (100%)</td>
</tr>
<tr>
<td>Other insects</td>
<td>7</td>
<td>7 (100%)</td>
<td>7 (100%)</td>
<td>7 (100%)</td>
<td>7 (100%)</td>
<td>7 (100%)</td>
<td>6 (86%)</td>
</tr>
<tr>
<td>Other metazoa</td>
<td>10</td>
<td>5 (50%)</td>
<td>2 (20%)</td>
<td>7 (70%)</td>
<td>6 (60%)</td>
<td>9 (90%)</td>
<td>6 (60%)</td>
</tr>
<tr>
<td>Σ All insects</td>
<td>42</td>
<td>38 (91%)</td>
<td>41 (98%)</td>
<td>42 (100%)</td>
<td>42 (100%)</td>
<td>42 (100%)</td>
<td>41 (98%)</td>
</tr>
<tr>
<td>SD*</td>
<td>1.01</td>
<td>0.62</td>
<td>0.54</td>
<td>0.65</td>
<td>0.71</td>
<td>0.84</td>
<td></td>
</tr>
<tr>
<td>Precision**</td>
<td>0.72</td>
<td>0.37</td>
<td>0.28</td>
<td>0.35</td>
<td>0.49</td>
<td>0.58</td>
<td></td>
</tr>
<tr>
<td>Σ All taxa</td>
<td>52</td>
<td>43 (83%)</td>
<td>43 (83%)</td>
<td>49 (94%)</td>
<td>48 (92%)</td>
<td>51 (98%)</td>
<td>47 (90%)</td>
</tr>
</tbody>
</table>
* Mean standard deviation (SD) of log10 sequence abundance from each insect taxon that was detected (specimens with < 0.003% read abundance discarded)

** Precision defined as the SD of the mean log10 distance to the expected abundance, calculated for each morphotaxon (all taxa).

### 3.3 In silico evaluation of primers

Performances of the 11 forward and 12 reverse primers were evaluated against OTUs of all insect orders (Figure 3). Reference data for binding sites of the standard Folmer primers HCO and LCO were very limited and Megaloptera and Turbellaria had below 100 OTUs. Primer efficiencies were very similar across orders but varied slightly between primers. However, Bivalvia, Turbellaria and Hirudinea showed higher penalty scores than other groups, while the high penalty scores for Amphipoda are likely due to the low sequence coverage and one mismatching sequence in the binding region (Figure 3). In silico and PCR (mock community samples) amplification success of BF/BR primer combinations was similar, but not always consistent. For example, while the BR1 primer shows a mean in silico amplification of only 77% (Figure 3), the BF2+BR1 primer combination performed well with actual samples (Figure 2). Generally speaking, primers incorporating wobble bases (jgLCO1490, BF1, BF2, BR1, BR2, jgHCO2198, H2123d) or inosin (Ill_B_F, ArF5, Il_C_R, ArR5) at the 3' end performed better than primers with no or just few wobble bases (linear regression mean penalty scores against log10 primer degeneracy: \( p = 0.004, \text{adj. } R^2 = 0.296 \)).

It should be noted that some primers from the literature are not only poorly matching because they lack wobble bases, but are affected by additional problems (see Figure S2, "critical mismatches"). For instance, near the 3’ ends, the EPT-long-univR has a completely unnecessary second inosine at a conserved position, while the Uni-MinibarF1 had a "T" at a position where more than half of the reference OTUs had an "A". Furthermore, the L499 primer targets a highly variable region. The mlCOIintR primer incorporates S (= C or G) leading to many mismatches (Figure S2), while the forward version of the same primer uses W (= A or T) wobble bases which match better. The reverse primers listed in the supplementary information of (Gibson et al., 2014) are not written in reverse complement, and will not work if ordered as provided (we evaluated the ArR5 primer in the reverse complement in silico). Finally, certain primers show mismatches to particular groups, e.g. the ZBJ-ArtFlc primer does not match well to sequences of Bivalvia and the BR1 primer shows an unambiguous mismatch to Turbellaria and Hirudinea at the fifth position (Figure S2).

### 4 Discussion

#### 4.1 Amplification success of mock communities

Environmental Impact Assessment (EIA) requires standardized and reliable data on biodiversity. Metabarcoding holds the potential to assess biodiversity of freshwater ecosystems faster and more reliable, if in particular suitable primers can be designed. We used PrimerMiner to obtain freshwater invertebrates specific sequence information, based on OTU sequence alignments generated of mitochondrial and COI barcodes from NCBI and BOLD. On this well balanced dataset we developed and tested four primer sets targeting freshwater invertebrates. We deliberately decided to not factor in nucleotide variability present in only few groups (mostly non-insect Metazoa) to limit the degeneracy of the primers to a reasonable level.
All four BF / BR primer combinations amplified the ten mock communities successfully, especially for insect taxa. By factoring in the different amplicon lengths in library pooling we obtained similar amounts of reads for each sample. All degenerated COI primers showed superior detection rates (up to 100% of insects and 98% of all morphotaxa) and more consistent read abundances compared to the standard Folmer barcoding primers that lacked any base degeneracy (Elbrecht and Leese, 2015; Folmer et al., 1994). The primer BF2 in combination with BR1/BR2 even showed better detection rates and higher precision than a previously used primer targeting a more conserved region of the mitochondrial 16S rRNA gene, which was tested on the same communities (Elbrecht et al., 2016). In silico analysis of the BF / BR primers against 15 freshwater groups on NCBI and BOLD confirmed their good detection rates (especially the BF2+BR2 combination), but also other primer sets from the literature are suitable for amplification of insect taxa, based on our in silico testing (e.g. the primers by (Geller et al., 2013; Gibson et al., 2014; Shokralla et al., 2015)). (Deagle et al., 2014) argued strongly against the use of degenerated primers to be used in DNA metabarcoding and instead proposed the use of ribosomal markers with more conserved binding regions. However, using a highly standardized approach with 10 independent taxa-rich mock communities, we here clearly show that the application of highly degenerated COI primers is not only feasible but even superior to ribosomal metabarcoding of animals with respect to primer performance and available reference databases.

While our developed primers show very reliable amplification results, there are also problems associated with the primers and the metabarcoding protocol. First, the use of fusion primers potentially decreases the chance of tag switching and reduces needed laboratory work, it also reduces PCR efficiency substantially (Schnell et al., 2015). Primer combinations involving BF2 primers were less affected by this issue, it was more pronounced with the BF1 primer (especially in combination with BR1). Furthermore, concerns have been raised by biases associated with use of tagged primers (O’Donnell et al., 2016). While we could not observe any obvious effects in our current dataset (most taxa are detected to equal proportions regardless of primer tag), there was a drop in sequence quality when using the P5_BF1_2 primer. Whether this was a systematic effect associated with the tag of the P5_BF1_2 primer or a problem in primer synthesis / quality could not be determined from this dataset. Independently of the source of this possible bias no effects on the number of detected taxa was observed. Further, 17% of reads from the BF2+BR2 primer combinations were discarded due to high expected error values, as the overlap was limited with 250 PE sequencing of a 421 bp region on the HiSeq system. Further, with highly degenerated primers the specificity of the primers decreases (Deagle et al., 2014) potentially amplifying non target regions. This effect was often minimal, with few sequences deviating from the expected length (below <0.5 % for most primers sets). These numbers were potentially inflated by PCR / sequencing errors and pseudogenes. More problematically, the BF1 and BF2 primers were affected by indel effects making up to 40% of the sequences 1-2 bp shorter or longer at the primer binding site. The reasons for this effects, which were also observed to a lesser degree in datasets from previous studies (Elbrecht and Leese, 2015; Elbrecht et al., 2016), are unclear. Possibly, the high degeneracy of the forward primers in combination with low diversity nucleotides at the primer’s 3’ end (e.g. C[cta]TT[tc]CC in BF2) makes this effect particularly pronounced. Thus, we recommend designing primers with two unique nucleotides on the 3’ end. The effect of this minimal shifting shortens the read length by 1-2 bp which has no effect on detection of taxa (OTUs will still match the same reference taxon, regardless of 1-2 bp being clipped from the sequence). However, when calculating OTU based biodiversity indices, the small shift might lead to a bias in these metrics due to inflated OTU numbers. This might be solved by aligning OTU sequences, and trimming them to the same length, but we still advise to take OTU based diversity measures with caution when using the BF / BR primer set. Finally, we have to acknowledge...
that the BF / BR primer sets showed poor performance on non-insect Metazoa like Bivalvia, Turbellaria, Amphipoda and Hirudinea, which are genetically distant to insects, making the development of an universal primer difficult.

While the primer sets developed and thoroughly evaluated in this study are a step in the right direction, they by no means perfect. We can recommend using the BF2+BR2 or BF2+BR1 primer set for targeting freshwater taxa with DNA metabarcoding, we also have to stress that for routine monitoring better primers are desirable.

4.2 Primer success is determined by base degeneracy and reference data

In silico analysis of 23 potentially suitable primers for COI DNA metabarcoding showed that high primer degeneracy leads to the best amplification of freshwater and insect taxa. This was also confirmed experimentally with the tested macroinvertebrate mock communities, which showed high primer bias with standard Folmer primers (Elbrecht and Leese, 2015) but a very consistent amplification with higher detection rates with the primers developed in this study. Possibly, other primers ((Gibson et al., 2014), (Shokralla et al., 2015) may also lead to equally good amplification. However, a lack of degeneracy can lead to substantial bias in many of the other evaluated primers. These biases might not strongly affect PCR for DNA barcoding on single organisms, but they may substantially skew detection rates of complex multispecies bulk samples, in the worst case leading to taxa remaining undetected (Elbrecht and Leese, 2015; Piñol et al., 2014). For example, the mlCOIint primers which have a maximum degeneracy of two nucleotides at each position (Leray et al., 2013), were previously tested with two mock communities and up to 35% of taxa remained undetected (Leray and Knowlton, 2015). Probably even more problematic are primers that lack base degeneracy. Despite primer bias associated with the high variation of the COI gene have been well documented (Clarke et al., 2014; Deagle et al., 2014; Elbrecht and Leese, 2015; Piñol et al., 2014; Sharma and Kobayashi, 2014), primers without base degeneracy like ZBJ-Art by (Zeale et al., 2010) are widely used e.g. for gut content analysis (123 citations as of June 2016).

Additionally we could reveal that several popular primers from the literature contain critical design flaws, possible introduced by accident (e.g. EPT-long-univR, mlCOIintR, Uni-MinibarF1). One has to keep in mind that a typo, or just one false base at the 3" end can make or break a primer. Additionally primers are often developed for on a small set of taxa, and thus might not work well for the ecosystem, geographic region or groups you want to target. For example (Clarke et al., 2014) evaluated the L499+H2123d as a metarcoding primer, but It was originally only developed to target Tephritid fruit flies, and probably never intended to be used beyond this Diptera Family (Van Houdt et al., 2010). Thus careful in silico evaluation and mock community testing of newly developed primers or primers from the literature against the specific groups of interest is crucial for metabarcoding projects. We highly recommend evaluation primers not only in silico but also using mock communities of known composition, to validate the primers work well for the targeted groups and purpose. Unfortunately, resources are limited, and metabarcoding primers gent not always tested and validated before being used in larger scale ecological or monitoring studies.

4.3 Recommended approaches for freshwater bio-assessment using macroinvertebrates

The success of DNA metabarcoding for bio-assessment and EIA of freshwater ecosystems depends on well-designed primers that reliably amplify the target communities. The more conserved primer
binding regions, the greater the amplification efficiency (Deagle et al., 2014). Therefore, 18S and 16S ribosomal markers have been proposed as suitable alternative markers to the COI gene, despite lacking comprehensive reference databases for animal taxa (Clarke et al., 2014; Deagle et al., 2014; Elbrecht et al., 2016). However, the *in silico* evaluations and documented good performance of the BF2+BR1 and BF2+BR2 primer sets of the COI gene shown in this study suggest clearly that ribosomal markers are not necessary for reliable DNA metabarcoding on animal species. The COI marker can lead to equally good results, but already has large reference databases available for animals. Thus we strongly encourage to focus efforts developing optimized ecosystem or community-specific COI primers.

When using DNA metabarcoding approaches for bio-assessment, protocols from the literature should be critically evaluated as success may be flawed by unsuitable primer design. Additionally, we recommend to include replicates to reduce the chance of tag switching and exclude false OTUs from the dataset (Lange et al., 2015). While we have previously encouraged the use of fusion primers due to their ease of use (single step PCR, (Elbrecht and Leese, 2015)), we have to acknowledge that they decrease PCR efficiency (Schnell et al., 2015). Thus, two step PCRs might be better suited for environmental samples because they often contain PCR inhibitors, even though two step PCR can be more prone to tag switching (Esling et al., 2015; Schnell et al., 2015).

Besides metabarcoding, metagenomic approaches using enrichment for mitochondrial genomes may become suitable for bio-assessment as well, with potentially less bias as the PCR amplification step can be omitted (Liu et al., 2015). However, as briefly discussed in (Elbrecht et al., 2016), metagenomics methods have to be further validated and mitochondrial reference genome libraries need to be completed (Dowle et al., 2015).

Thus, the probably most suitable approach for bio-assessment to date is to use the DNA metabarcoding approach with the COI marker for DNA based monitoring of stream ecosystems. However, primers for DNA metabarcoding of macroinvertebrates are ideally further optimized and primers from the literature should be tested more extensively on mock communities.

### 4.4 Conclusions

Reliable and quick bio-assessment is of critical importance for Environmental Impact Assessment of aquatic ecosystems. DNA metabarcoding has the potential to meet this challenge, if suitable primers can be obtained. We were able to show with *in silico* evaluations as well as experimental data with COI metabarcoding almost the complete aquatic invertebrate community of freshwater ecosystem can be reliably detected. We provide novel degenerated primer sets with high detection rates and greatly reduced primer bias. In view of the still incomplete data bases we encourage further *in silico* and *in vivo* evaluation of existing primers and further development of suitable metabarcoding primers to unlock the full potential of metabarcoding for Environmental Impact Assessment and biomonitoring. However, already now our data suggests that for freshwater ecosystems the technique of DNA metabarcoding is mature enough to be used for stream monitoring on a large scale.

### 5 Figures
**Figure 1**: Selection of potential COI primer sets for DNA metabarcoding of insects, targeting the Folmer region. Primer pairs shown are typically used / suggested combinations from the literature. Table S1 gives an overview of the exact primer sequences and references.
Figure 2: Comparison of the COI Folmer primer performance and the four tested primer combinations newly developed. All primer combinations were tested with the same ten bulk samples each containing 52 morphologically distinct macroinvertebrate taxa. The 52 taxa are shown on the x-axis with the relative number of reads obtained for each morphotaxon by black dots on the logarithmic y-axis (mean read abundance indicated by red circles), for each respective primer combination. Sequence abundance was normalized across the ten replicates and the amount of tissue used in each DNA extraction. Only OTUs with a minimum read abundance of 0.003% in at least one of the ten samples were included in analyses. Number of samples for which a morphotaxon was not detected is indicated by orange and red numbers in each plot. A thick vertical line in light red indicates if a morphotaxon was not detected.
Figure 3: Overview of in silico evaluation of primer performance using PrimerMiner v0.7 with OTU data from 15 freshwater assessment relevant invertebrate groups. Primer performance is shown for each group in pie charts (red = failure, green = working, grey = missing data / gaps). Every primer sequence match with a mismatch penalty score of above 120 is considered a failure, and the amplification success displayed in each circle (excluding missing data). The box plot is based on the mean penalty scores for each group, with the mean penalty score and degeneracy given for each primer. For metabarcoding potentially suitable primers have a yellow background. For detailed evaluation parameters see scripts S2. The L499 primer for the Trubellaria group could not be evaluated due a 3 bp deletion in the reference sequences, but the primer are likely to not amplify well.
Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

VE and FL conceived the ideas and designed methodology, VE did carry out the laboratory work and analyzed the data, VE led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

Funding

FL and VE are supported by a grant of the Kurt Eberhard Bode Foundation to FL.

Acknowledgments

We thank Bianca Peinert for her help with primer development, and Edward Wilcox and ScienceExchange for HiSeq sequencing. We would like to thank Edith Vamos, Jan Macher and Vera Zizka for their helpful suggestions that improved this manuscript.

References


