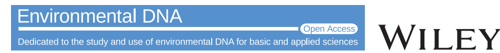
Received: 5 August 2020 | Revised: 26 April 2021 | Accepted: 10 May 2021
DOI: 10.1002/edn3.222

ORIGINAL ARTICLE



WILEY

High-throughput identification of non-marine Ostracoda from the Tibetan Plateau: Evaluating the success of various primers on sedimentary DNA samples

Paula Echeverría-Galindo[1] | Sten Anslan[2] | Peter Frenzel[3] | Sven Künzel[4] | Miguel Vences[2] | Liseth Pérez[1] | Nicole Börner[1] | Wengang Kang[1] | Anja Schwarz[1] | Junbo Wang[5] | Ping Peng[5] | Liping Zhu[5] | Antje Schwalb[1]

[1]Institute of Geosystems and Bioindication, Technische Universität Braunschweig, Braunschweig, Germany

[2]Zoological Institute, Technische Universität Braunschweig, Braunschweig, Germany

[3]Institute of Earth Sciences, Friedrich-Schiller-Universität, Jena, Germany

[4]Department of Evolutionary Genetics, Max Planck Institut für Evolutionsbiologie, Plön, Germany

[5]Key Laboratory of Tibetan Environment Changes and Land Surface Processes, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing, China

Correspondence
Sten Anslan, Zoological Institute, Technische Universität Braunschweig, Braunschweig, 38106, Germany.
Email: sten.anslan@gmail.com

Paula Echeverría-Galindo, Institute of Geosystems and Bioindication, Technische Universität Braunschweig, 38106, Braunschweig, Germany.
Emails: p.echeverria-galindo@tu-braunschweig.de; paulaegalindo@gmail.com

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: DFG grant 317513741 / GRK 2309

Abstract

Dwelling in a variety of aquatic habitats, one of the most abundant groups of microcrustaceans, ostracodes, are widely used indicator organisms in paleolimnological research. Typically, they are identified via traditional methods using morphological features but this may be excessively time-consuming and prone to inter-investigator variation. DNA barcoding and metabarcoding have become important tools for specimen identification, with a great impact in the field of taxonomy, (paleo-)ecology and evolution. Despite its potential, metabarcoding has been rarely used to analyze the community structure of ostracodes. Here, we evaluate the performance of a metabarcoding approach for ostracode identification in surface sediment samples collected from Lake Nam Co on the Tibetan Plateau. We tested six different primer pairs amplifying fragments of three different genes, and compared their success in inferring ostracode communities, coupled with morphological identification of ostracodes from the same sediment samples. In total, depending on the primer pair used, seven to nineteen ostracode amplicon sequence variants (ASVs) were identified. Via microscopy, eight morphospecies were identified. We found considerable differences between primer pairs in yielding ostracode sequences via metabarcoding. In general, the highest proportions of ostracode reads and ASVs were found with primers amplifying fragments of the 18S rRNA gene, whereas primers for COI gene had the highest *in silico* amplification success and highest sequencing depth per sample but only contained <1% of ostracode sequences. As a consequence, the metabarcoding results with 18S rRNA gene were more consistent with the morphological data compared to those obtained with COI or mitochondrial 16S rRNA primers. No significant effects of treatment with different sediment quantities for DNA extraction (10 g vs. 0.5 g) were found on ostracode ASVs community composition. These results indicate that DNA metabarcoding can serve as an efficient tool for ostracode-based environmental

Echeverría-Galindo and Anslan are shared first authorship.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2021 The Authors. *Environmental DNA* published by John Wiley & Sons Ltd

reconstructions but requires an informed decision on primers and target gene, as well as extending the barcoding database for improved accuracy.

KEYWORDS
metabarcoding, ostracodes, primer design, sample size, sediment sample, sedDNA, Tibetan Plateau

1 | INTRODUCTION

Ostracodes (Crustacea, Ostracoda) are small aquatic crustaceans (mostly 0.5–1.5 mm in size) inhabiting all types of waterbodies from freshwater, brackish to marine environments. Because of their abundance in the aquatic ecosystems and sensitivity to the changes in environmental conditions such as temperature or conductivity (Meisch, 2000), these organisms are suitable bioindicators (Echeverría Galindo et al., 2019; Pérez et al., 2013). Furthermore, their low-magnesium calcite valves preserve well in sediments, thus Ostracoda displays one of the oldest continuous fossil records among arthropods (Griffith & Holmes, 2000). Therefore, they also serve as paleoecological and paleoenvironmental indicators (Frenzel et al., 2010; Ruiz et al., 2013; Wrozyna et al., 2010; Zhu et al., 2010). Despite the intensive use of ostracodes as (paleo)bioindicators, in Tibetan Plateau environments and elsewhere, their identification through morphological traits often remains challenging. Moreover, the Tibetan Plateau is known as an area of high endemism, and the ostracode diversity of the area has not been fully assessed taxonomically (Mischke, 2012). Specific environmental conditions of the area may drive phenotypic variation in the local populations of widespread species. However, not all species can be identified based on morphological differences. In some cases, the accumulation of genetic and ecological differences is not correlated with the accumulation of morphological variations, this situation drove the appearance of cryptic species (Schön et al., 2016). Together, this may lead to misidentifications and to inter-investigator variation in species identification (Fürstenberg et al., 2015).

Given the complex and partly incomplete taxonomy of ostracodes, molecular specimen identification through DNA barcodes is an effective tool for complementary morphological assessments, allowing accurate species delimitation regardless of morphotypes and ontogenetic stages. DNA barcoding is reported to be efficient for delimiting crustacean species (Costa et al., 2007), and studies have demonstrated the method's usefulness in identifying ostracodes (Bode et al., 2010; Nigro et al., 2016; Schön et al., 2014). Further, with the development of high-throughput sequencing (HTS) methods, molecular species identification (via DNA barcodes) has considerably boosted the acquisition of biodiversity data in all biomes (Bahram et al., 2018; Pesant et al., 2015; Tedersoo et al., 2014; Thomsen & Willerslev, 2015). Consequently, the simultaneous molecular identification of multiple species directly from environmental samples (i.e., eDNA metabarcoding), has become an alternative way for fast biomonitoring purposes (Baird & Hajibabaei, 2012; Ruppert et al., 2019). Although ostracodes are great bioindicators, to the best of our knowledge, there are no published high-throughput DNA metabarcoding studies targeting specifically this taxonomic group from environmental samples.

The aims of this study were to (1) evaluate the success of various metabarcoding markers (commonly used for invertebrates and newly designed) with a specific focus on identifying ostracodes from lake sediment samples (sedDNA metabarcoding), (2) test the effect of sample mass (0.5 vs. 10 g) as a determinant of ostracode community composition, and (3) compare the metabarcoding results with the morphologically identified ostracodes data from the same sediment samples.

2 | METHODS

2.1 | Study area

Sediment samples were collected in Nam Co, a lake located in the monsoon-influenced transition zone between semi-humid and semi-arid areas of the Tibetan Plateau. The mean annual air temperature measured at the Nam Co Monitoring and Research Station for Multisphere Interactions (NAMORS) between 2006 and 2017 was −0.6°C, and the annual precipitation ranged from 291 to 568 mm (mean =405 mm) with most of the precipitation occurring during the monsoon season from May to September (Anslan et al., 2020). The precipitation rates are subject to spatial variations due to the >7,000 m high Nyainqêntanglha mountain range which represents the southern border of the Nam Co catchment. Nam Co (30° 40′N, 90° 30′E) (Figure 1) is a dimictic lake, endorheic, and located at high altitude (4730 m a.s.l.) in the south-eastern Tibetan Plateau (Wang et al., 2019). The lake has a surface area of 2,026 km^2 making it the third largest lake in the Tibetan Plateau region (Kai et al., 2020), with a maximum water depth of ~99 m (Wang et al., 2020). As expected for an endorheic lake, Nam Co is brackish and alkaline, with a salinity of 0.9 g/L, conductivity of 1850 µS/cm and pH 9.2 (Keil et al., 2010; Wang et al., 2020), with dominance of Na^+ and HCO_3^- in the lake water. The water temperature ranged from 12 to 2°C, dropping to close to 0°C during winter (November–May) when snow and ice cover the lake (Wang et al., 2020).

2.2 | Sample collection

Surface sediment samples from Nam Co were collected in July 2018 from different depths (≤35 m) using an Ekman grab, and from the littoral zone (0.5–1 m) using a hand net (125-µm mesh) (Table S1). Top 2–3 cm sediment from each Ekman grab were collected and transferred to sterile Whirl-Pak bags (ca. 200 g of wet weight per sample). Each surface

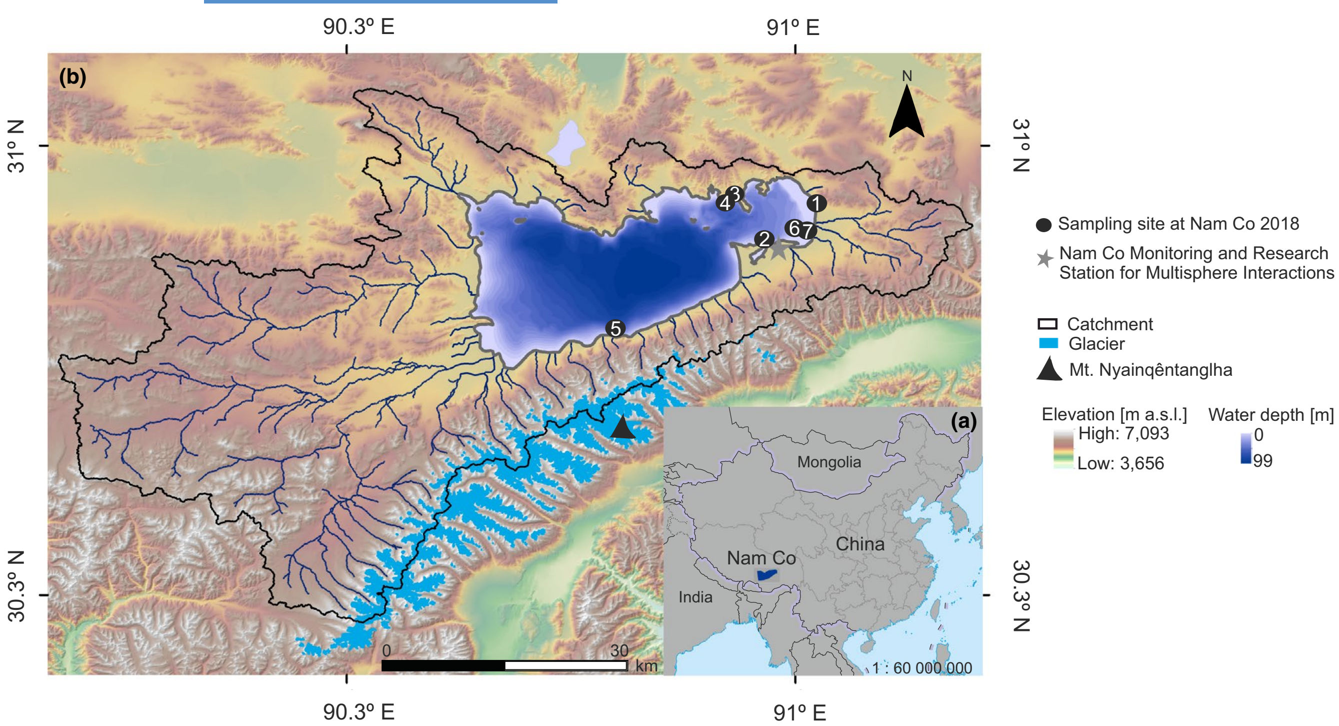


FIGURE 1 Location of the seven sampling sites (black dots), and the Nam Co Monitoring and Research Station for Multisphere Interactions (NAMORS) (gray star) in Nam Co region. (Map "a" modified from standard map service for China [http://bzdt.ch.mnr.gov.cn])

sediment sample was sub-sampled for (1) morphological identification of ostracodes and (2) metabarcoding. Sampling equipment was cleaned by thoroughly rinsing the equipment after each sampling and prior to taking a new sample in the water of the respective sampling location. In the field laboratory, surface sediment samples for the morphological identification were preserved with 95% ethanol and stored at 4°C until further processing. Samples for metabarcoding analyses were sieved through 2 mm sieves to remove larger particles (a 20 µm mesh was used as a collection dish), and ca. 50 g of sediment were divided in three 50 ml tubes filled with 95% ethanol (4:1, ethanol: sediment ratio). Sieve was cleaned with bleach (5%) and distilled water prior each sample processing. All samples were stored and transported in a freezer. Additionally, surface sediment samples were collected from the littoral zones (<0.5 m) of the lake and adjacent water bodies using a hand net (125-µm mesh) to pick individual living ostracode specimens. These sediments were transferred to sterile Whirl-Pak bags, transported to the field laboratory for manually sorting out ostracode specimens, which were then morphological identified. Identified specimens were subjected to molecular DNA analyses to add reference sequences for metabarcoding analyses and primer design (see Methods below).

2.3 | Morphological identification of ostracodes

From the hand net samples, isolated ostracode specimens (with soft parts) were extracted and identified in the field laboratory using a BMS 76095 stereomicroscope. *Leucocytherella sinensis* individuals were provided by Peter Frenzel collected from previous fieldtrips at Nam Co, Tibetan Plateau during September 2008–2012 (Börner et al., 2017). Specimens (from A-3 to Adult stages) were stored in 1.5 ml vials with 95% ethanol until DNA extraction (see Methods below). From the seven surface sediment samples, 1 g of wet sediment was sub-sampled and sieved using a 63 µm mesh size sieve. Distilled water was used to rinse the sieve between samples. All adult and juvenile ostracode carapaces and valves (broken and intact) were extracted using fine brushes under a Leica MZ 7.5 dissecting binocular microscope, and stored in micropaleontological slides. Valves were visualized using scanning electron microscopy (SEM), Philips ESEM XL30 at the Institute of Zoology, FSU Jena, Germany. Carapaces with well-preserved soft parts were stored in Eppendorf vials with 95% ethanol and then dissected and mounted following methods described by (Holmes & Chivas, 2002). Morphospecies were counted if the individuals had well-preserved (hyaline) valves. High number of juveniles and adults with soft parts and closed carapaces (articulated valves) were considered as autochthonous individuals (Wrozyna et al., 2009a). In order to exclude allochthonous (transported) material, opaque, heavily coated (encrusted) or abraded valves were excluded from counting (Frenzel et al., 2010). All ostracode identification followed Meisch (2000), Wrozyna et al., (2009b), and Fürstenberg et al., (2015).

2.4 | *Molecular analysis of net-collected ostracodes*

DNA from net-collected ostracode specimens was extracted using Platinum™ Direct PCR Universal Master Mix (Invitrogen) following

the manufacturer's instructions. PCRs were performed using primers LCO1490 and HCO2198 (Folmer et al., 1994) to amplify the 658 base pairs (bp) barcoding region of the mitochondrial gene for cytochrome oxidase subunit 1 (COI); 16SH and 16SL (Palumbi et al., 2002) to amplify ca. 490 bp fragment of the mitochondrial gene for large subunit rRNA (16S); and 18S-F1 and 18S-R9 (Yamaguchi & Endo, 2003) to amplify ca. 1500 bp fragment of the nuclear gene for small subunit rRNA (18S). For the failed 18S-F1 and 18S-R9 runs, newly developed primers 18SV4F and 18SV4R (Table 1; see methods below), amplifying ca. 328 bp nested fragment of 18S-F1 and 18S-R9, were used. Total of 20 µl PCR mix per sample consisted of 10 µl Platinum™ Direct PCR Universal Master Mix, 0.4 µl of forward primer (10 µM), 0.4 µl of reverse primer (10 µM), 7.2 µl of nuclease-free water and 2 µl of template DNA. The PCR program for COI included an initial activation at 94°C for 2 min, followed by 5 cycles at 94°C for 1 min, 45°C for 1 min, 72°C for 1 min; 35 cycles at 94°C for 1 min, 51°C for 1 min, 72°C for 1 min and final extension at 72°C for 5 min. The PCR program for 16S rRNA gene was as follows: 94°C for 2 min, 35 cycles for 30 s at 95°C, 50 s at 50°C, 1 min at 72°C, and final extension for 10 min at 72°C. PCRs for 18S with 18S-F1 and 18S-R9 primers consisted of 94°C for 2 min, 35 cycles at 95°C for 30 s, 50°C for 30 s, 72°C for 1 min 20 s, and final extension at 72°C for 10 min. PCR for 18S with18SV4osF and 18SV4osR primers consisted of 94°C for 2 min, 35 cycles of 94°C for 30 s, 54°C for 1 min, 72°C for 1 min, and 72°C for 10 min. Amplification success was checked via electrophoresis using 1% agarose gel and 5 µl of PCR product. PCR products were cleaned using Exo-SAP enzymes and sequenced in LGC Genomics GmbH (Berlin), using ABI 3730 XL. Sequences were quality-checked and filtered in CodonCode Aligner (CodonCode Corporation). Generated sequences are deposited in NCBI database, under accession numbers MT731602-MT731610 (18S rRNA gene), MT732947-MT732955 (16S rRNA gene), MT830920-MT830929 (COI gene).

2.5 | Molecular analysis of sediment samples and primer selection

For metabarcoding, sediment samples were centrifuged at 2600*g* for 10 min, followed by supernatant removal and mixing of the three subsamples. Two replicate DNA extractions were performed for each sample using DNeasy PowerMax Soil Kit and DNeasy PowerSoil Kit (Qiagen, Germany). For the PowerMax kit, up to 10 grams of wet sediments were used (in the following text named "treatment 10 g"), and for the PowerSoil kit, 0.5 grams of wet sediments (in the following text named "treatment 0.5 g"). The initial DNA extraction step was modified by adding Proteinase K (10 mg/ml) and 1 M DTT (dithiothreitol) together with the C1 solution from the extraction kits. For the PowerMax Kit (up to 10 g of sediments) 60 µl of Proteinase K and 100 µl of DTT were added, and for the PowerSoil Kit (0.5 g of sediments), 4 µl of Proteinase K and 25 µl of DTT was added; following overnight incubation at 56°C. Final DNA elution was performed twice by adding half of the recommended amount of the buffer onto spin column membrane and incubated at room temperature for 3 minutes. The rest of the steps were performed following manufacturer's instructions.

PCRs for metabarcoding were performed using six different primer pairs (Table 1). **I:** BF2 and BR1n primers (Elbrecht & Leese, 2017), a universal primer set for aquatic invertebrates, that amplify ca. 322 bp fragment of COI region (in the following text as COI). BR1n is slightly modified version of BR1 from Elbrecht and Leese (2017); replacing H and D with N in the 15th and 18th positions, respectively. Although a set of COI primers for zooplankton exists (Prosser et al., 2013), which included only one marine ostracode species in the primer development, based on *in silico* test (EcoPCR, see below), we found that these primers are not suitable for the majority of freshwater ostracodes (Table S2), whereas BF2 and BR1n primers demonstrated high *in silico* amplification success (Table S2). **II:**

TABLE 1 Primers used for metabarcoding

Gene	Primer name	Primer sequence (5′–3′)	Length	Reference
COI	BF2	GCH CCH GAY ATR GCH TTY CC	322	Elbrecht and Leese (2017)
	BR1n	ARY ATD GTR ATD GCN CCN GC		
16S (-long)	16S_osF	TGA CYG TRC DAA GGT AGC A	313	This study
	16S_osR	CAA CAT CGA GGT CRC AAA C		
16S (-short)	16S_osFs	AGK GAC RAG AAG ACC CT	132	This study
	16S_osRs	GCT GTT ATC CCT RRR GTA		
18S (-long)	18SV4osF	GCT CGT AGT TGG ATC TCA GT	328	This study
	18SV4osR	CGA ACC TCT GAC TTT CGT TC		
18S (-short)	18SV4osF	GCT CGT AGT TGG ATC TCA GT	118	This study
	18SV4osRs	TYG CCT GCT TTR AGC ACT C		
18S (-Euk)	Euk_1391f	GTA CAC ACC GCC CGT C	132	Amaral-Zettler et al. (2009); Stoeck et al. (2010)
	EukBr	TGA TCC TTC TGC AGG TTC ACC TAC		

Note: "Length" denotes the amplification length using corresponding primers.

16S_osF and 16S_osR are newly designed primers that amplify ca. 313 bp region from ostracode mitochondrial 16S (in the following text as 16S-long). **III:** 16S_osFs and 16S_osRs amplify ca. 132 bp region from mitochondrial 16S rRNA (amplification region is nested in between amplification region of primer set II; in the following text as 16S-short). 16S_osFs is modified from Ins16S_1 forward primer from (Clarke et al., 2014) and 16S_osRs is modified from 16Sins_R primer from (Elbrecht et al., 2016). A set of 16S primers previously designed for ostracodes (Jarman et al., 2006) were not used in our study as *in silico* amplification analyses demonstrated >5 mismatches for many freshwater ostracodes (Table S2). **IV:** 18SV4F and 18SV4R that amplify ca. 330 bp of V4 region of the small subunit of rRNA (18S; in the following text as 18S-long). 18SV4R is based on TAReukREV3 from Stoeck et al., (2010), modified according to available ostracode 18S sequences. 18SV4F is newly designed primer, located 40 bp toward 3′-end compared with TAReuk454FWD1 from Stoeck et al., (2010). **V:** 18SV4F and newly designed 18SV4Rs amplifies a nested ca. 118 bp fragment of previous primer pair (in the following text as 18S-short). **VI:** Euk_1391f and EukBr that amplify ca. 132 bp region from ostracode 18S V9 region (based on those of Amaral-Zettler et al., 2009; Stoeck et al., 2010) (in the following text as 18S-Euk).

The modifications or design of the primers were based on available ostracode sequences from the EMBL database (Kanz et al., 2005) and newly generated sequences from collected ostracode specimens. Primer design was mainly focused on covering taxa that are expected to inhabit Nam Co and related taxa, since genetic sequences of all species living in the lake are not available so far. Ostracode sequences were aligned using online service of MAFFT (Katoh et al., 2017) and inspected in MEGA (Kumar et al., 2016) to manually select PCR primers. *In silico* amplification success was evaluated using EcoPCR (Ficetola et al., 2010) by allowing maximum of two mismatches per primer.

The PCR mix per sample consisted of 5 µl of Hot Start FirePol Master Mix (Solis BioDyne, Estonia), 0.5 µl of uniquely tagged (8 bp tag +2–4 bp heterogeneity spacer) forward and reverse primers (10 µM), 17 µl of nuclease-free water and 2 µl of template sedDNA (total volume 25 µl). PCR conditions were as follows: for COI, 95°C for 15 min, 35 cycles of 94°C for 30 s, 50°C for 30 s, 65°C for 2 min 30 s, final extension at 65°C for 5 minutes. For 16S-long, 95°C for 15 min, 35 cycles of 95°C for 45 s, 53°C for 1 min and 72°C for 1 min, final extension at 72°C for 10 min. For 16S-short, 95°C for 15 min, 35 cycles of 95°C for 45 s, 54.5°C for 1 min and 72°C for 1 min, final extension at 72°C for 5 min. For 18S-long and 18S-short, 95°C for 15 min; 35 cycles of 95°C for 45 s, 54°C for 1 min and 72°C for 1 min; 72°C for 10 min. For 18S-Euk, 95°C for 15 min; 35 cycles of 94°C for 45 s, 57°C for 1 min and 72°C for 1 min 30 s, final extension at 72°C for 10 min. Three replicate PCRs were performed per sample, products were pooled, and amplification success was checked via gel electrophoresis by pipetting 5 µl PCR product on 1% agarose gel. All PCR products per sample were pooled as based on their relative quantity and purified using Favor-Prep™ Gel/PCR Purification Kit (Favorgen-Biotech Corp., Austria), following the manufacturer's instructions. To alleviate the sequencing bias toward generating higher amounts of shorter amplicons, the total pool for sequencing consisted of 2 to 1 ratio of larger amplicons *vs*. shorter amplicons (i.e., >300 bp vs. <300 bp). PCR amplicons were subjected to Illumina adapter ligation and sequencing by Illumina MiSeq (2x250) using MiSeq Reagent Kit v2 (10% Phix). Illumina sequencing data sets have been deposited in the Sequence Read Archive (SRA), BioProject ID: PRJNA647726. Steps of DNA extraction, PCR and sequencing included both negative and positive controls. All sample preparations were conducted under laminar flow clean bench, using 30 min UV sterilization prior each process.

2.6 | Illumina sequencing data processing

Raw Illumina paired-end reads were reoriented to 5′-3′ based on PCR primers, using PipeCraft software (v1.0; Anslan et al., 2017). Reads without matches to primer sequences (2 mismatches allowed), were discarded. Primers were clipped from the reoriented reads using cutadapt (v2.10; Martin, 2011). Resulting reads were subjected to DADA2 pipeline (v1.16; Callahan et al., 2016) to generate amplicon sequence variants (ASVs). Reads were quality filtered with the following quality filtering options: maxN=0, maxEE=1, truncQ=2. Default parameters were used for error models and merging paired-end reads. Putative chimeras were removed using consensus method. Taxonomic annotation of the ASVs was performed using the blastn algorithm (Camacho et al., 2009) against EMBL (v143) database and against locally generated sequences from ostracode specimens.

The "raw" ASV tables were further filtered to remove singleton ASVs (i.e., ASVs with only one sequence) and to include only ASVs from Ostracoda. For COI data set, an ASV was annotated as an Ostracoda, when the sequence similarity and coverage against the reference sequence in the database was higher than 80% and 98%, respectively. This threshold was based on the examination of the 10 best blast hits, where latter threshold revealed hits only to various Ostracoda. Lower thresholds exhibited comparable hits also to various Diptera and thus were considered unreliable Ostracoda annotations. For 16S rRNA genes data sets, an ASV was annotated as an Ostracoda, when the sequence similarity and coverage against the reference sequence in the database was higher than 75% and 98%, respectively. ASVs in all 18S data sets that got a first blast hit against Ostracoda had >90% sequence similarity and >95% coverage against reference sequences and were considered valid ostracode ASVs. Additionally, neighbor-joining phylogenetic trees (with 100 bootstraps) were generated for 18S and 16S data, using online service of MAFFT (Katoh et al., 2017), to visualize the placement of ostracode ASVs (Figures S1, S2). *Vargula hilgendorfii* (AF363301) and *Manawa staceyi* (AF363295) served as outgroups. Based on the placement of the ostracode ASVs on the trees, the taxonomically unidentified ASVs (due to a lack of reference sequences) were assigned, when possible, to a species that corresponded morphologically identified species from the same sediment samples (Figures S1, S2). The final ostracode ASV tables were compared with the "raw" ASV tables and

morphological identification table to identify the best performing primer pair for amplifying ostracodes from sediment samples.

2.7 | Statistics

Wilcoxon Matched Pairs Test was used to test the difference between 10 g and 0.5 g treatments per primer pair metabarcoding data in number of ostracode ASVs and sequences. To test the effect of 10 g versus 0.5 g treatment on the ostracode ASV composition, Bray–Curtis similarity of Hellinger-transformed data was analyzed using permutational analysis of variance (PERMANOVA) using "vegan" package (Oksanen et al., 2015) in R (R-Core-Team, 2019), with function "adonis" and 999 permutations. Additionally, Bray–Curtis distance matrices of Hellinger-transformed data per metabarcoding treatment (10 g vs. 0.5 g) were compared with Mantel tests (method="spear") to assess the correlations between sample similarities as implemented in the "vegan" package (Oksanen et al., 2015). Procrustes analyses were used to compare the correlations (in the ordination space) between morphological and metabarcoding data using "vegan" package in R (R Core Team, 2019). Presence–absence transformed Bray–Curtis distance matrices were used for Procrustes analyses with metaMDS (4999 permutation).

3 | RESULTS

3.1 | Ostracode morphological identification

In total, eight ostracode morphospecies were identified in the surface sediment samples: *Candona candida* (O.F. Müller, 1776), *Candona xizangensis* Huang, 1982, *Fabaeformiscandona gyirongensis* (Huang, 1982), *Ilyocypris bradyi* Sars, 1890, ?*Leucocythere dorsotuberosa* f. *typica sensu* Wrozyna et al. (2009), Huang, 1982, ?*Leucocythere dorsotuberosa* f. *postilirata* (Pang, 1985), *Leucocytherella sinensis* Huang, 1982, and *Tonnacypris gyirongensis* (Yang, 1982) (Figure 2; Table S3). *L. sinensis* and *T. gyirongensis* where identified in all seven sediment samples, from shallow to deep waters (≤35 m; Table S3). The predominant species is *L. sinensis* (average abundance 82% in terms of counted valves), followed by ?*L. dorsotuberosa* f. *typica*, and ?*L. dorsotuberosa* f. *postilirata* (>200 valves/wet gram), where ?*L. dorsotuberosa* f. *typica* tolerates water from 10 to 20 m depth, while forma *postilirata* is more adapted to deeper water (35 m). The species *C. candida*, *I. bradyi*, and *C. xizangensis* are comparatively rare in the samples, and had the lowest mean abundance (<2%; Table S3). All the ostracode valves and carapaces did not present any transport signal (opaque, heavily coated or abraded valves), which indicates that the identified morphospecies are living in the lake (autochthonous taxa).

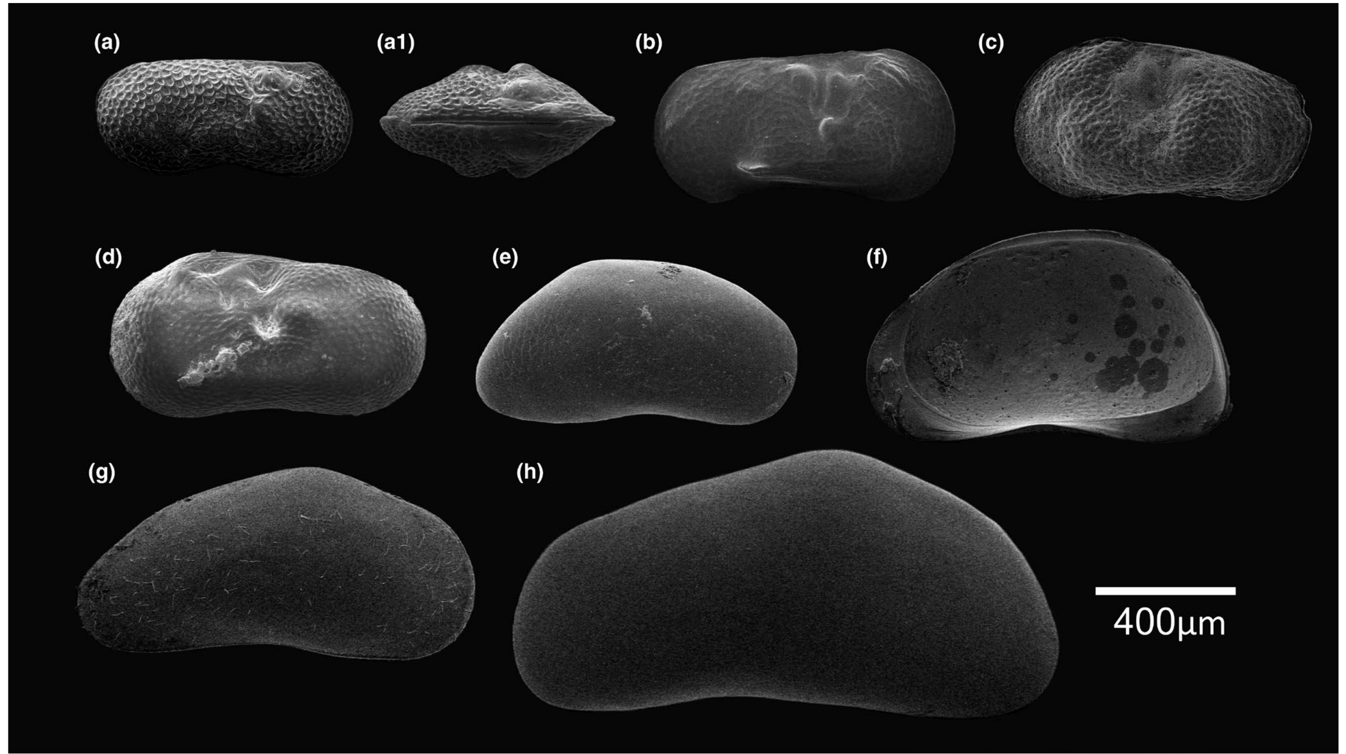


FIGURE 2 Scanning electron microscopy (SEM) pictures of eight species of ostracodes morphologically identified from Lake Nam Co sediment samples. LV: left valve, RV: right valve, Ev: external view, Iv: internal view. A. RV, Ev, male, A1. Carapace, dorsal view, female, *Leucocytherella sinensis*. B. RV, Ev, male? *Leucocythere dorsotuberosa* f. *postilirata*. C. LV, Ev, female? *Leucocythere dorsotuberosa* f. *typica*. D. Carapace from left side, *Ilyocypris bradyi*. E. RV, Ev, female, *Candona xizangensis*. F. RV, Iv, *Candona candida*. G. RV, Ev, *Tonnacypris gyirongensis*. H. LV, Ev, *Fabaeformiscandona gyirongensis*

3.2 | In silico amplification success

We estimated the potential of the primer pairs (*in silico* amplification success) using EcoPCR software by allowing maximum of 2 mismatches per primer. The sequences of the most diverse order of ostracodes, Podocopida (Karanovic, 2012) COI, 16S rRNA and 18S rRNA genes from the NCBI (Geer et al., 2009) were used as a reference database (data download: 28.04.2020). Total number of dereplicated (i.e., unique) ostracode sequences that contained primer binding sites was 1279 for COI gene, but considerably less for 16S and 18S rRNA genes (142, 156, 138, 139 and 50 for 16S-long, 16-short, 18S-long, 18S-short and 18S-Euk, respectively). All included reference sequences (and taxa) and *in silico* amplification success results are outlined in Table S2. Highest relative *in silico* amplification rate was observed for COI gene primers which matched with 99.3% (1270 sequences out of 1279) of the reference data. The success rates for 16S rRNA gene primers were 71.1% (101 sequences out of 142) and 73.7% (115 sequences out of 156) for 16S-long and 16S-short, respectively. The majority of sequences that did not share matching regions with 16S rRNA gene primers were annotated as *Romecytheridea ampla* and *R. bacata*, but also two species from genera *Cytherissa* were not amplified during the *in silico* process (Table S2). The matches for primers in the 18S rRNA gene regions were 95.7% (132 sequences out of 138), 94.2% (131 sequences out of 139) and 90.0% (45 reads sequences out of 50) for 18S-long, 18S-short and 18S-Euk, respectively.

3.3 | Molecular analysis of sediment samples

For this analysis, we pooled the results from the 10 and 0.5 g treatments to focus on the performance of different primer pairs. After quality filtering, average sequencing depths in the metabarcoding data sets were 11,765 sequences per sample for COI gene, 4603 for 16S-long, 3442 for 16S-short, 4768 for 18S-long, 3709 for 18S-short and 3180 for 18S-Euk. Universal COI gene primers amplified wider variety of Metazoa, but also high proportion of sequences, 36%, originated from various prokaryotes. In 16S rRNA gene data sets, majority of sequences from non-target taxa were assigned to other Metazoa (Mollusca, Crustacea, Hexapoda, Annelida), whereas sequences from prokaryotes were represented in less than 4%. The non-target taxa from 18S-long and 18S-short data sets were assigned mostly to Gastrotricha and Annelida; non-eukaryotic sequences were represented in <0.05%. Universal eukaryotic 18S rRNA gene primers (18S-Euk) amplified wide range of eukaryotes, but 17% of sequences originated from prokaryotes.

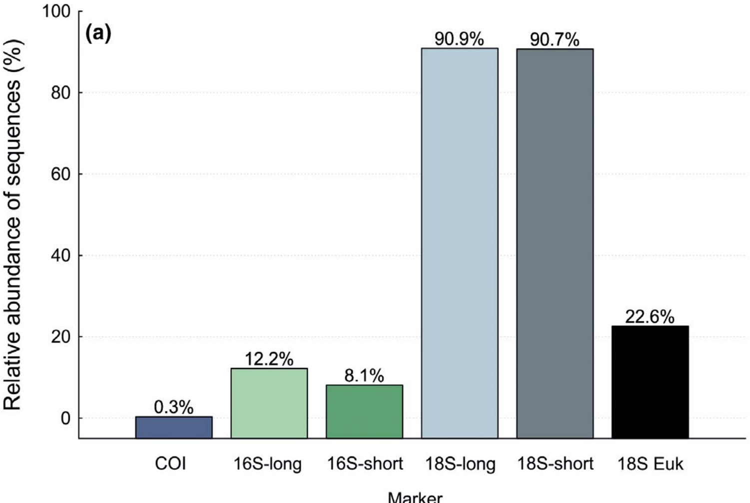
Among all sequences in the metabarcoding data sets, the highest proportion of ostracode sequences was found in 18S rRNA gene data sets (18S-long, 90.9%; 18S-short, 90.7%; and 18S-Euk, 22.6%) (Figure 3a). Data set of 16S-long consisted of 12.2%, but 16S-short data set harbored only 8.1% ostracode sequences. Although COI primers had the highest *in silico* amplification success and highest sequencing depth per sample, the data set contained only 0.3% of ostracode sequences (Figure 3a). The relative proportion of ostracode sequences per samples demonstrated a similar pattern, 18S-long and 18S-short data sets yielding the highest proportions (Figure 3b).

The overall metabarcoding data revealed seven ostracode ASVs for COI gene, 19 for 16S-long and 11 for 16-short; 15 for 18S-long data, 12 for 18S-short and 9 for 18S-Euk data (Table S3). Ostracodes were morphologically identified from all analyzed samples (Table S3). Accordingly, all samples in the18S rRNA gene data sets contained ostracode ASVs (Table S3). However, three samples from the COI gene data set, one samples from 16S rRNA gene data sets and three samples from 16S-short data sets yielded no ostracode ASVs (Table S3). Therefore, the ostracode ASV richness per sample was generally higher in the 18S rRNA gene data sets (Figure 4). The Procrustes analyses between morphological and metabarcoding data indicated the strongest correlations with 18S-short data (n = 7; Procrustes correlation = 0.618; Table S3), however this correlation was statistically non-significant (p = 0.135). Details on other Procrustes correlations are listed in Table S3.

Due to a lack of reference sequences for ostracodes, we were not able to firmly confirm the presence of all detected ostracode morphospecies in the metabarcoding data sets (Figure 6, Table S3). For example, *Candona xizangensis*, *Fabaeformiscandona gyirongensis* and the genera *Leucocythere* have no public genetic information available (i.e., four species in the morphological data). Therefore, to construct the species abundance heatmap per treatment (Figure 6), the presence of latter taxa in the metabarcoding data sets was conjectured from neighbor-joining phylogenetic trees (Figures S1, S2) as based on the placement of ostracode ASV on the trees (Table S3). *Leucocytherella sinensis* Huang 1982 (by far the most abundant species on central and southern Tibetan Plateau) was collected from previous fieldtrips by Peter Frenzel, and barcoded for the first time in this study (short 18S sequence was obtained). From our net-collected ostracode specimens, we were able to capture and barcode three species, that inhabit shallow waters: *Tonnacypris gyirongensis*, *Heterocypris* cf. *salina* (Brady, 1868), and *Ilyocypris bradyi* Brady & Norman, 1889. However, high-quality barcodes were not successfully generated for all markers used in this study (Table S4). Interestingly, 16S-long, 18S-short and 18S-long data contained *Heterocypris salina* (>99.7% identity to reference sequence),18S-short and 18S-long data set contained *Limnocythere inopinata* (100% identity to reference sequence) that were not identified via morphological analyses from the sediment samples (Figure 6; Table S3).

3.4 | Impact of sediment sample size for metabarcoding

On average, DNA extraction treatment with higher quantity of sediments, 10 g, resulted in slightly higher numbers of ostracode ASVs and sequences, except for COI gene (Figure 5a). However, the 10 g versus 0.5 g treatments only differed significantly for read numbers in the 18S-short data (p = 0.018; Figure 5b). PERMANOVA analyses indicated no significant effect of treatment with different sediment quantities for DNA extraction (10 g *vs*. 0.5 g) on ostracode ASVs community composition (p > 0.7 for all cases). Community similarity correlations between 10 g versus 0.5 g, however, varied among used primer sets (Figure S3). Sample size for COI gene data was only 3 (4 out of



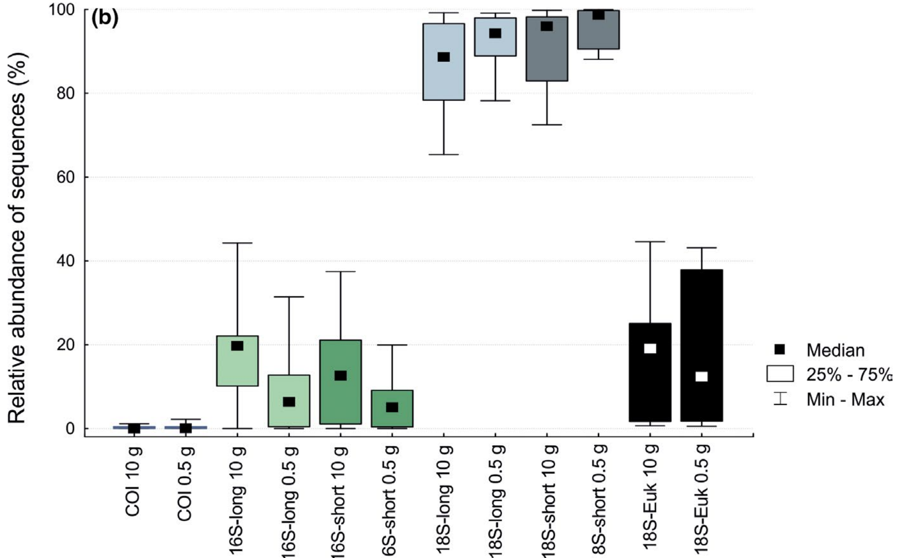


FIGURE 3 (a) Relative abundance of ostracode sequences (%) from each data set, 10 g and 0.5 g treatments are combined; (b) box plots for relative abundance of ostracode sequences (%) per treatment and metabarcoding marker

7 samples did not contain ostracode ASVs; Figure S4), thus Mantel test was not performed. The ostracode ASVs community correlations between 10 g and 0.5 g treatments were high and significant for all other cases (Mantel $R > 0.7$, $p < 0.011$ for all cases), except for 16S-long ($n = 6$; Mantel $R = 0.655$, $p = 0.1$; Figure S3). However, correlations between the number of ASVs in the corresponding samples from 10 g versus 0.5 g treatments, demonstrated significant positive correlations only for 18S-long and 18S-short data sets (Spearman $R = 0.925$, $p = 0.003$ and Spearman $R = 0.840$, $p = 0.017$, respectively).

4 | DISCUSSION

4.1 | Primer performance

Here we compared the performance of six primer sets in identifying ostracodes from lake sediment samples. The highest number of ostracode sequences (and ASVs) were retrieved with primers for the 18S rRNAV4 region designed herein (primer sets of 18S-long and 18S-short), whereas degenerate primers for COI and 16S rRNA gene amplified mostly other non-target taxa.

For routine DNA barcoding as well as metabarcoding, COI gene is the standard molecular identification marker for Metazoa (including crustaceans) and in general has a high species delimitation rate. COI gene also benefits from a large public reference sequence database, has been effectively used for delimiting ostracode species (Nigro et al., 2016), and holds sufficient sequence variance to discriminate between cryptic species (Bode et al., 2010; Karanovic, 2015). Accordingly, in the current study, we found that COI primers (BF2 and BR1n; Table 1) had high *in silico* amplification, however, the metabarcoding data set contained very few (<1%) ostracode sequences. This confirms the inapplicability of degenerate COI primers for environmental DNA applications when metazoan groups are targeted (Hajibabaei et al., 2019; Horton et al., 2017; Macher et al., 2018; Weigand & Macher, 2018). For example, the meiofauna study from sediment samples by Weigand and Macher (2018) reported that only about 1% of the operational taxonomic units (OTUs) in the COI metabarcoding data were assigned to Metazoa. Similarly, studying the

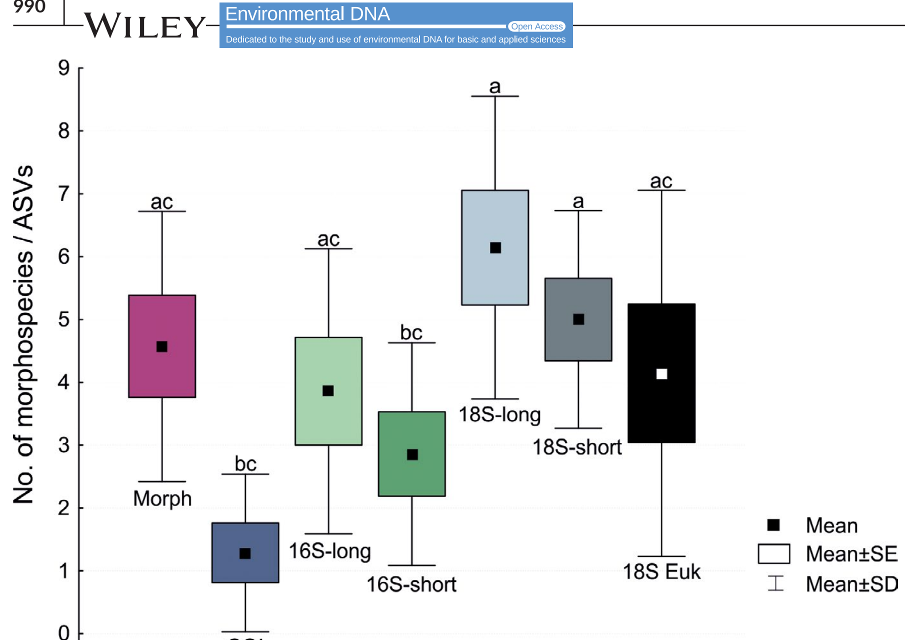


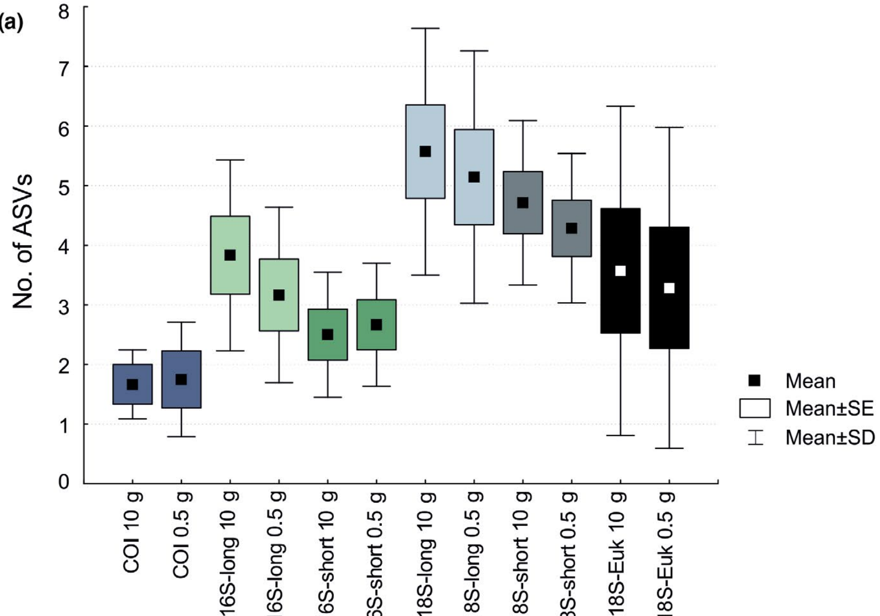
FIGURE 4 Box plots for ostracode morphospecies and ASVs per treatment. Different letters above the whiskers indicate significant differences according to Wilcoxon Matched Pairs Test. Metabarcoding data includes merged data from 10 g and 0.5 g treatments

invertebrate communities from soil and leaf litter samples, Horton et al., (2017) found that the vast majority of sequences in the data set obtained with degenerate COI primers originated from bacteria. Because of the difficulty to design taxon-specific primers for the COI gene, the developed primers for metabarcoding are not only universal across Metazoa (Elbrecht & Leese, 2017; Leray et al., 2013), but also capture genomic DNA fragments from microbes which may have several orders of magnitude higher biomass and DNA content in the substrate compared to fauna in the sediments or soil (Bar-On et al., 2018). This results in overrepresentation of microbial sequences, which hinders the signal from the target group. Therefore, prior to DNA extraction, an extra step of isolating the animals of interest from the substrate is performed (Aylagas et al., 2016; Brannock & Halanych, 2015; Macher et al., 2018). However, the process of isolating specimens is time-consuming, thus costly, which limits its application at larger scales.

Compared with primers for COI gene, the success in amplifying ostracode DNA was slightly higher with the primers for 16S rRNA gene, but still the majority of data consisted of sequences of non-target taxa. Although the primers used for the 18S rRNA V9 region (18S-Euk set) are also universal primers, but for eukaryotes, their success rate in amplifying ostracodes was higher compared to COI and 16S primers. About 15% of sequences in the 18S-Euk data set were not assigned to Eukaryota, yet more than 20% of sequences originated from ostracodes. This corresponds with the findings from Zhan et al., (2014), Horton et al., (2017) and Ahmed et al., (2019) who reported a higher success of 18S rRNA primers in identifying target metazoan groups directly from environmental DNA. However, the species delimitation power of 18S rRNA gene region is lower compared to 16S or COI genes (Tang et al., 2012). Nevertheless, the 18S rRNA gene primer sets used herein had the highest success in metabarcoding ostracodes from sediment samples. The vast majority of sequences in the data set produced with 18S-long and 18S-short primers were identified as ostracode sequences. This resulted in higher numbers of detected ostracode ASVs and more consistent results with morphological data compared to the COI or 16S genes metabarcoding data sets (Figure 6). Nevertheless, further efforts toward taxon-specific COI or 16S primer development for environmental DNA applications could favor the use of these more taxonomically informative regions (as in e.g., Leese et al., 2020, for freshwater macroinvertebrates), which would also benefit from the larger barcoding databases.

4.2 | Sediment sample size

Identifying ostracodes via DNA metabarcoding from 10 versus 0.5 g of sediments demonstrated a slightly higher number of sequences and ASVs from 10 g of sediments (interestingly, except for COI data). But in most cases the difference was not significant, especially for ASVs richness (Figure 5). Similarly, the study by Brannock and Halanych (2015) indicated that the overall number of meiofaunal OTUs did not vary significantly between various DNA extraction amounts from sediment samples (0.3, 5, 10 g). In contrast, Nascimento et al., (2018) reported that the data from sample volumes of 0.2 g contained lowest number of metazoan OTUs and insufficiently captured the metazoan communities in the sediment samples. In their study, the diversity estimates for 0.2, 4 and 6 g did not differ substantially, but varied compared with much higher amounts (up to 28.2 g, see Nascimento et al., 2018), but nevertheless, metazoan community composition from 0.2 g of sediments was significantly different from larger quantities of sediments (including from 6 g). Contrary to latter results, our analyses showed no significant difference in the ostracode communities (based on ASVs) between 10 and 0.5 g treatments for any metabarcoding data set. Therefore, while aiming to assess "complete" beta-diversity of the (e.g., metazoan) communities in the environment, larger quantities of sample results in a more complete picture, whereas "less" would be sufficient for certain target groups as demonstrated in the current study.



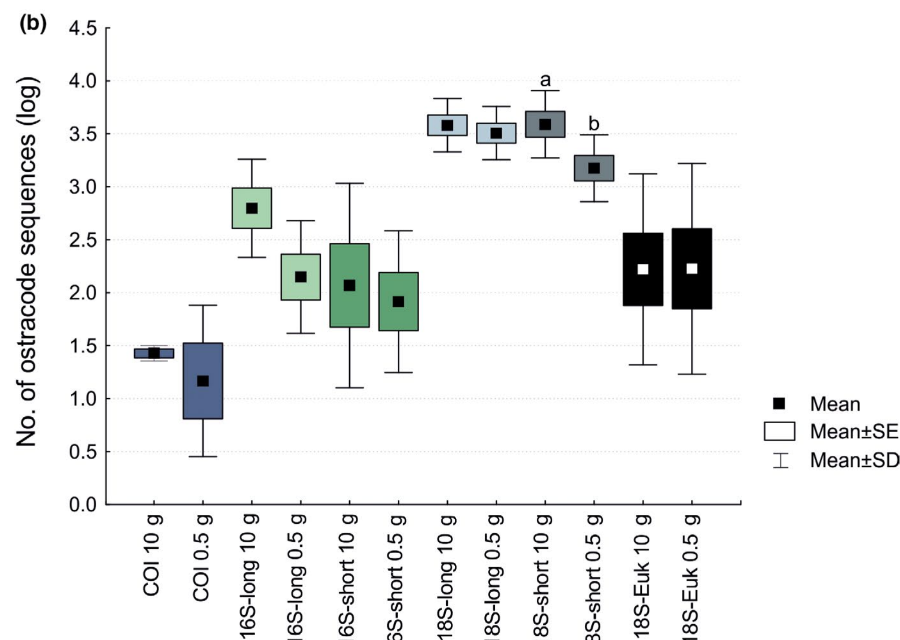


FIGURE 5 Box plots of the number of ostracode ASVs (a) and sequences (b) per metabarcoding data set. Different letter above whiskers denote Wilcoxon Matched Pairs Test statistical significance between groups of 10 g versus 0.5 g treatments per primer set (tests were performed only for 10 g versus 0.5 g for a corresponding primer set data, not among all data)

4.3 | Comparison of molecular and morphological analyses

Due to the issue of an incomplete barcoding database, that is, missing reference sequences for ostracodes, especially from Tibetan Plateau, we were not able to directly identify all morphologically detected species in the DNA metabarcoding data sets. The issue of incompleteness of the reference sequence databases has been reported in many other studies that have aimed to compare metabarcoding data with morphological identifications (Cahill et al., 2018; Elbrecht et al., 2017; Kang et al., 2021). Nevertheless, the most abundant species (*Leucocytherella sinensis* and *Tonnacypris gyirongensis*) were detected in majority of metabarcoding data sets (Table S3) due to a herein generated reference sequences. Although the 18S gene has limited species-level resolution in ostracodes (Macario-González et al., 2018), we found that the most successful primer pairs, amplifying 18S V4 region, detected more than 10 ostracode ASVs, which speculatively corresponded to 10 species (Figure 6; Figures S1, S2). However, it is important to notice that 18S V4 metabarcoding data sets harbored ASVs that had 100% sequence similarity against *Ilyocypris angulata* (Table S3), but we considered these ASVs as *Ilyocypris bradyi* (Figure 6). This, because of invariability of short 18S sequences among congeneric species (Karanovic et al., 2020) and the fact that only *I. bradyi* has been reported from our study lake (this study; Mischke, 2012; Echeverría-Galindo, unpublish). Therefore, in a certain case, such as in this

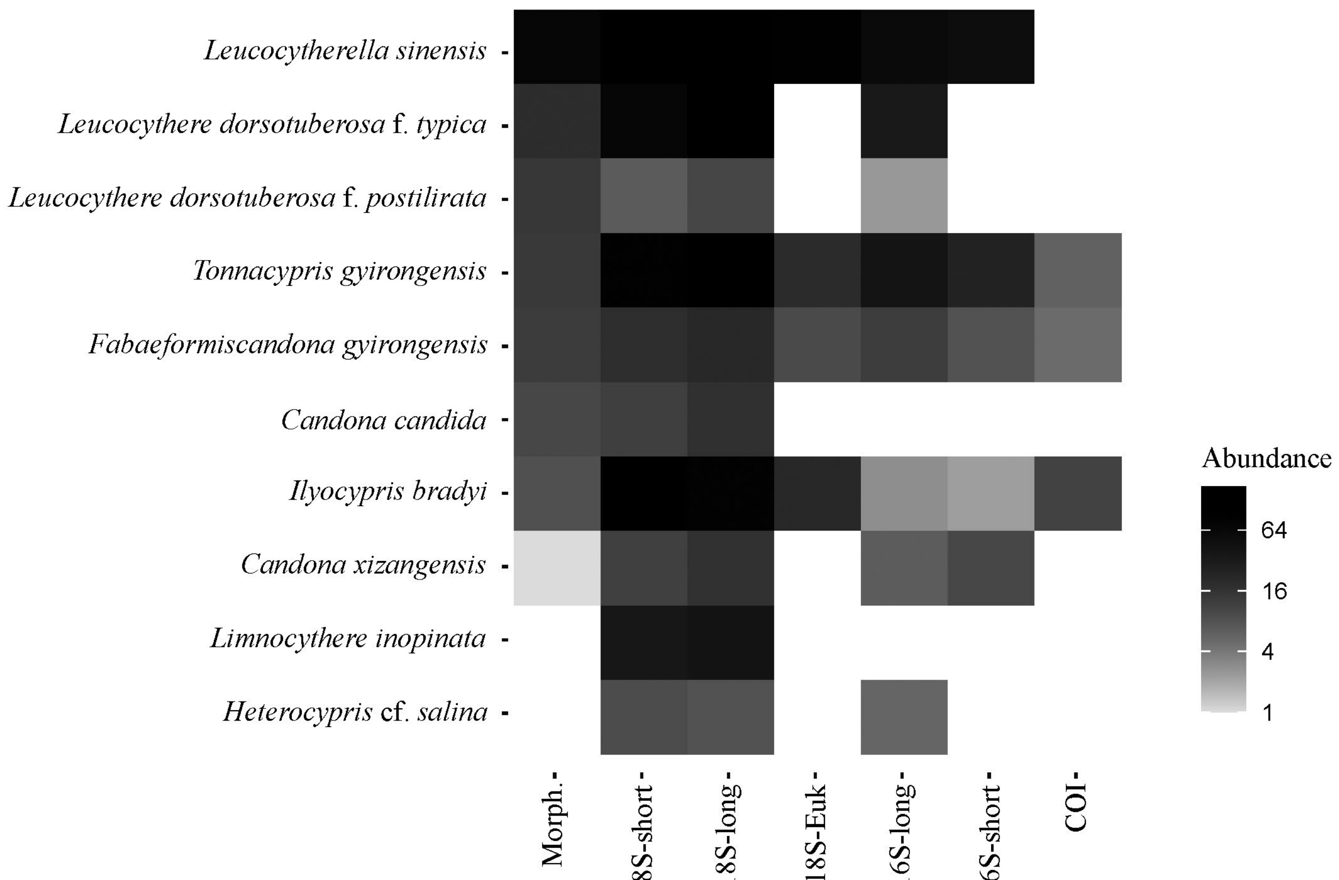


FIGURE 6 Heatmap of ostracode species abundances across treatments; "Morph" denotes morphological identification treatment, followed by metabarcoding treatments. Treatments with 18S-short and 18S-long markers are most similar to morphological identification treatment. The abundance values have been square root transformed for plotting. For morphologically identified species, the abundance values denote counted number of valves across 7 study samples. Abundance values in metabarcoding treatments represent number of sequences. White colored areas in the heatmap denote 0 abundance of a corresponding species in a treatment. Note that that due to a lack of reference sequences, not all molecularly detected ostracodes (metabarcoding) were assigned to species level. However, the phylogenetic placement (Figures S1, S2) and sample distribution patters (Table S3) allowed to speculatively connect ostracode molecular units (ASVs) with morphologically identifies species. See Table S3 for details

study where most species in the ecosystem are not congeneric, the 18S marker gene may provide sufficient resolution to detect ostracode species richness. Interestingly, our metabarcoding data contained also two species of ostracodes that were not detected via morphological analyses form the sediment samples (Figure 6; Table S3). A total of eight species has been reported from Lake Nam Co (Wrozyna et al., 2009a). However, our metabarcoding results suggested slightly higher species richness. It is not uncommon to detect higher richness of taxa through metabarcoding as demonstrated in many other studies, for example for benthic macroinvertebrates (Elbrecht et al., 2017), for zooplankton (Schroeder et al., 2020) and for phytoplankton (Zimmermann et al., 2015). This may be associated with the higher taxonomic resolution of identification through DNA barcodes (Gibson et al., 2015) and/or higher sensitivity of DNA methods to detect rare taxa (Zhan and MacIsaac, 2015). We highlight that some juvenile ostracode valves can be hardly identified morphologically, which allows one to speculate that these "extra" species in the metabarcoding data sets went unnoticed during morphological analyses. Ostracode morphospecies in our study were only counted if well-preserved specimens (hyaline valves) and with high number of juveniles and adults with soft parts or closed carapaces (articulated valves) were found. This allows more accurate morphological identifications, as well as ensures that the morphospecies inhabit the collected sediment rather than being transported from other environments or depths (Macario-González et al., 2018; Meisch, 2000). However, whether these missed species via morphological analyses were misidentified juveniles or represent extracellular environmental DNA that was captured during metabarcoding remains unclear.

4.4 | Perspectives

The endemism and abundance of ostracodes on the Tibetan Plateau make them a relevant study group as they are excellent bioindicators of past climate and environmental conditions

(Mischke, 2012). Our results indicate that DNA metabarcoding may be an efficient tool for ostracode-based paleoenvironmental rseconstructions. This may especially apply to environments like on the Tibetan Plateau, characterized by strong seasonality, with long, cold winters and short summers, where species develop different strategies that allow them to adapt to the harsh environments (e.g., changes in the valves or soft parts morphology). All morphological changes as well as the life cycles of ostracodes in the Tibetan Plateau lakes are not known in detail (Akita et al., 2016). However, molecular identification through metabarcoding could enable biosurveys regardless of ontogenetic stages of the specimens. Moreover, processing biological replicates, to account for spatial heterogeneity of fauna in the sediments and improve community structure estimates (Lanzen et al., 2017), is more time- and cost-efficient with metabarcoding approach due to its high-throughput workflow design. Furthermore, isolating intact specimens from sediment archives is not always feasible, due to the significant effects of diagenetic processes, where age-related post-burial degradation occur (Bennett et al., 2011; Karanovic et al., 2020; Mezquita et al., 2005). But unlike with the traditional morphological analyses, the recovery of species abundance data with metabarcoding is not that straightforward. Various processes in the metabarcoding workflow, such as DNA extraction, PCR, primer bias and target gene copy number in the cell, can introduce biases so that the number of sequences might not correlate with the organism abundance (Deagle et al., 2013; Fonseca, 2018; Nichols et al., 2018). This has noted to be an issue especially when comparing the inter-specific abundances, but less pronounced intra-specifically (Amend et al., 2010; Elbrecht & Leese, 2015). Nevertheless, the absolute abundances in the metabarcoding studies could be approximated via “spike-in” controls by adding known amounts (e.g., synthetic) DNA molecules to study samples (Harrison et al., 2020; Zemb et al., 2020).

This study also highlights the need for further commitment in generating appropriate barcoding database for accurate metabarcoding purposes. This also allows to address the issue of possible existence of cryptic species in the ecosystem, which is an important factor to take into account for paleoenvironmental inferences, because this can mislead ecological interpretation. However, not only species assemblages, but also valve ornamentations may be related to the environmental conditions as been suggested, for example for *Leucocytherella sinensis* (Fürstenberg et al., 2015). Based on the ornamentations on the valves, initially twelve morphospecies distributed across the Tibetan Plateau were described for *Leucocytherella* (Wrozyna et al., 2009), but are now considered synonyms for *L. sinensis* based on the revised morphological characteristics and ontogenetic stages of single valves (Fürstenberg et al., 2015). Yet, the DNA analyses still await to confirm or rebut the assignment validity of previously described morphospecies to one species. Thus, high-throughput molecular DNA identification of ostracodes cannot be applied “blindly” to infer paleoenvironmental conditions but needs further alliance with taxonomists and ecologists to redefine the limits between species and their relations with the environment.

ACKNOWLEDGEMENTS

We are thankful to the Nam Co Monitoring and Research station for Multisphere Interactions (NAMORS) team of the Institute of Tibetan Plateau Research, Chinese Academy of Sciences. This research is a contribution to the International Research Training Group "Geo-ecosystems in transition on the Tibetan Plateau (TransTiP)", funded by Deutsche Forschungsgemeinschaft (DFG grant 317513741/GRK 2309). We acknowledge support by the German Research Foundation and the Open Access Publication Funds of the Technische Universität Braunschweig.

CONFLICT OF INTEREST

The authors declare no competing interests.

AUTHORS CONTRIBUTIONS

WK, NB, AS and PP performed sampling. PEG and PF conducted ostracodes morphological identifications and scanning electron microscopy (SEM) pictures. SA performed molecular, bioinformatics and statistical analyses. SK performed high-throughput Illumina sequencing. MV, LP, PF and LZ revised and helped improve the manuscript. AS developed the program (DFG-GRK 2309) that funded this research. PEG and SA wrote the manuscript with contributions from all authors.

DATA AVAILABILITY STATEMENT

Illumina data sets have been deposited in the Sequence Read Archive (SRA) under BioProject PRJNA647726. Herein generated Sanger sequences are deposited in NCBI database, under accession numbers MT731602-MT731610 (18S rRNA gene), MT732947-MT732955 (16S rRNA genes), MT830920-MT830929 (COI gene).

ORCID

Paula Echeverría-Galindo https://orcid.org/0000-0003-0210-4607
Sten Anslan https://orcid.org/0000-0002-2299-454X
Peter Frenzel https://orcid.org/0000-0002-3821-4632
Miguel Vences https://orcid.org/0000-0003-0747-0817
Liseth Pérez https://orcid.org/0000-0002-5256-3070
Nicole Börner https://orcid.org/0000-0001-8833-6665
Wengang Kang https://orcid.org/0000-0003-2640-2019
Antje Schwalb https://orcid.org/0000-0002-4628-1958

REFERENCES

Ahmed, M., Back, M. A., Prior, T., Karssen, G., Lawson, R., Adams, I., & Sapp, M. (2019). Metabarcoding of soil nematodes: the importance of taxonomic coverage and availability of reference sequences in choosing suitable marker (s). *Metabarcoding and Metagenomics*, *3*, e36408. https://doi.org/10.3897/mbmg.3.36408

Akita, L. G., Frenzel, P., Wang, J., Börner, N., & Peng, P. (2016). Spatial distribution and ecology of the Recent Ostracoda from Tangra Yumco and adjacent waters on the southern Tibetan Plateau: a key to palaeoenvironmental reconstruction. *Limnologica*, *59*, 21–43.

Amaral-Zettler, L. A., McCliment, E. A., Ducklow, H. W., & Huse, S. M. (2009). A method for studying protistan diversity using massively

parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. *PLoS One*, *4*, e6372. https://doi.org/10.1371/journal.pone.0006372

Amend, A. S., Seifert, K. A., & Bruns, T. D. (2010). Quantifying microbial communities with 454 pyrosequencing: does read abundance count? *Molecular Ecology*, *19*, 5555–5565. https://doi.org/10.1111/j.1365-294X.2010.04898.x

Anslan, S., Azizi Rad, M., Buckel, J., Echeverria Galindo, P., Kai, J., Kang, W., Keys, L., Maurischat, P., Nieberding, F., Reinosch, E., Tang, H., Tran, T. V., Wang, Y., & Schwalb, A. (2020). Reviews and syntheses: How do abiotic and biotic processes respond to climatic variations in the Nam Co catchment (Tibetan Plateau)? *Biogeosciences*, *17*, 1261–1279.

Anslan, S., Bahram, M., Hiiesalu, I., & Tedersoo, L. (2017). PipeCraft: flexible open-source toolkit for bioinformatics analysis of custom high-throughput amplicon sequencing data. *Molecular Ecology Resources*, *17*, e234–e240.

Aylagas, E., Mendibil, I., Borja, Á., & Rodríguez-Ezpeleta, N. (2016). Marine sediment sample pre-processing for macroinvertebrates metabarcoding: mechanical enrichment and homogenization. *Frontiers in Marine Science*, *3*, 203. https://doi.org/10.3389/fmars.2016.00203

Bahram, M., Hildebrand, F., Forslund, S. K., Anderson, J. L., Soudzilovskaia, N. A., Bodegom, P. M., Bengtsson-Palme, J., Anslan, S., Coelho, L. P., Harend, H., Huerta-Cepas, J., Medema, M. H., Maltz, M. R., Mundra, S., Olsson, P. A., Pent, M., Põlme, S., Sunagawa, S., Ryberg, M., … Bork, P. (2018). Structure and function of the global topsoil microbiome. *Nature*, *560*, 233–237.

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*, 2039–2044. https://doi.org/10.1111/j.1365-294X.2012.05519.x.

Bar-On, Y. M., Phillips, R., & Milo, R. (2018). The biomass distribution on Earth. *Proceedings of the National Academy of Sciences*, *115*, 6506–6511.

Bennett, C. E., Williams, M., Leng, M. J., Siveter, D. J., Davies, S. J., Sloane, H. J., & Wilkinson, I. P. (2011). Diagenesis of fossil ostracods: Implications for stable isotope based palaeoenvironmental reconstruction. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *305*(1–4), 150–161. https://doi.org/10.1016/j.palaeo.2011.02.028

Bode, S. N. S., Adolfsson, S., Lamatsch, D. K., Martins, M. J. F., Schmit, O., Vandekerkhove, J., Mezquita, F., Namiotko, T., Rossetti, G., Schön, I., Butlin, R. K., & Martens, K. (2010). Exceptional cryptic diversity and multiple origins of parthenogenesis in a freshwater ostracod. *Molecular Phylogenetics and Evolution*, *54*, 542–552. https://doi.org/10.1016/j.ympev.2009.08.022

Börner, N., De Baere, B., Akita, L. G., Francois, R., Jochum, K. P., Frenzel, P., Zhu, L., & Schwalb, A. (2017). Stable isotopes and trace elements in modern ostracod shells: implications for reconstructing past environments on the Tibetan Plateau, China. *Journal of Paleolimnology*, *58*, 191–211. https://doi.org/10.1007/s10933-017-9971-1

Brannock, P. M., & Halanych, K. M. (2015). Meiofaunal community analysis by high-throughput sequencing: Comparison of extraction, quality filtering, and clustering methods. *Marine Genomics*, *23*, 67–75.

Cahill, A. E., Pearman, J. K., Borja, A., Carugati, L., Carvalho, S., Danovaro, R., Dashfield, S., David, R., Féral, J.-P., Olenin, S., Šiaulys, A., Somerfield, P. J., Trayanova, A., Uyarra, M. C., & Chenuil, A. (2018). A comparative analysis of metabarcoding and morphology-based identification of benthic communities across different regional seas. *Ecology and Evolution*, *8*, 8908–8920.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: high-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*, 581.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, *10*, 421.

Clarke, L. J., Soubrier, J., Weyrich, L. S., & Cooper, A. (2014). Environmental metabarcodes for insects: in silico PCR reveals potential for taxonomic bias. *Molecular Ecology Resources*, *14*, 1160–1170.

Costa, F. O., DeWaard, J. R., Boutillier, J., Ratnasingham, S., Dooh, R. T., Hajibabaei, M., & Hebert, P. D. (2007). Biological identifications through DNA barcodes: the case of the Crustacea. *Canadian Journal of Fisheries and Aquatic Sciences*, *64*, 272–295.

Deagle, B. E., Thomas, A. C., Shaffer, A. K., Trites, A. W., & Jarman, S. N. (2013). Quantifying sequence proportions in a DNA-based diet study using Ion Torrent amplicon sequencing: which counts count? *Molecular Ecology Resources*, *13*, 620–633.

Echeverría Galindo, P. G., Pérez, L., Correa-Metrio, A., Avendaño, C., Moguel, B., Brenner, M., Cohuo, S., Macario, L., & Schwalb, A. (2019). Tropical freshwater ostracodes as environmental indicators across an altitude gradient in Guatemala and Mexico. *Revista De Biologia Tropical*, *67*, 1037–1058.

Elbrecht, V., & Leese, F. (2015). Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass-sequence relationships with an innovative metabarcoding protocol. *PLoS One*, *10*, e0130324.

Elbrecht, V., & Leese, F. (2017). Validation and development of COI metabarcoding primers for freshwater macroinvertebrate bioassessment. *Frontiers in Environmental Science 5*.

Elbrecht, V., Taberlet, P., Dejean, T., Valentini, A., Usseglio-Polatera, P., Beisel, J. N., Coissac, E., Boyer, F., & Leese, F. (2016). Testing the potential of a ribosomal 16S marker for DNA metabarcoding of insects. *PeerJ*, *4*, e1966. https://doi.org/10.7717/peerj.1966.

Elbrecht, V., Vamos, E. E., Meissner, K., Aroviita, J., & Leese, F. (2017). Assessing strengths and weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine stream monitoring. *Methods in Ecology and Evolution*, *8*, 1265–1275.

Ficetola, G. F., Coissac, E., Zundel, S., Riaz, T., Shehzad, W., Bessiere, J., Taberlet, P., & Pompanon, F. (2010). An In silico approach for the evaluation of DNA barcodes. *BMC Genomics*, *11*, 434.

Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrdme c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, *3*, 294–299.

Fonseca, V. G. (2018). Pitfalls in relative abundance estimation using eDNA metabarcoding. *Molecular Ecology Resources*, *18*, 923–926.

Frenzel, P., Wrozyna, C., Xie, M. P., Zhu, L. P., & Schwalb, A. (2010). Palaeo-water depth estimation for a 600-year record from Nam Co (Tibet) using an ostracod-based transfer function. *Quaternary International*, *218*, 157–165.

Fürstenberg, S., Frenzel, P., Peng, P., Henkel, K., & Wrozyna, C. (2015). Phenotypical variation in Leucocytherella sinensis Huang, 1982 (Ostracoda): A new proxy for palaeosalinity in Tibetan lakes. *Hydrobiologia*, *751*, 55–72.

Geer, L. Y., Marchler-Bauer, A., Geer, R. C., Han, L., He, J., He, S., Liu, C., Shi, W., & Bryant, S. H. (2009). The NCBI biosystems database. *Nucleic Acids Research*, *38*, D492–D496.

Gibson, J. F., Shokralla, S., Curry, C., Baird, D. J., Monk, W. A., King, I., & Hajibabaei, M. (2015). Large-Scale Biomonitoring of Remote and Threatened Ecosystems via High-Throughput Sequencing. *PLoS ONE*, *10*, e0138432.

Griffith, H., & Holmes, J. (2000). *Non-marine ostracods and Quaternary palaeoenvironments* (p. 188). Quaternary Research Association.

Hajibabaei, M., Porter, T. M., Robinson, C. V., Baird, D. J., Shokralla, S., & Wright, M. T. (2019). Watered-down biodiversity? A comparison of metabarcoding results from DNA extracted from matched water and bulk tissue biomonitoring samples. *PLoS One*, *14*.

Harrison, J., Calder, W. J., Shuman, B. N., & Buerkle, C. A. 2020. The quest for absolute abundance: the use of internal standards for DNA-barcoding in microbial ecology.

Holmes, J. A., & Chivas, A. R. (2002). *The Ostracoda: Applications in Quaternary research*. : American Geophysical Union. Geophysical Monograph Series 131.

Horton, D. J., Kershner, M. W., & Blackwood, C. B. (2017). Suitability of PCR primers for characterizing invertebrate communities from soil and leaf litter targeting metazoan 18S ribosomal or cytochrome oxidase I (COI) genes. *European Journal of Soil Biology*, *80*, 43–48.

Jarman, S. N., Redd, K. S., & Gales, N. J. (2006). Group-specific primers for amplifying DNA sequences that identify Amphipoda, Cephalopoda, Echinodermata, Gastropoda, Isopoda, Ostracoda and Thoracica. *Molecular Ecology Notes*, *6*, 268–271.

Kai, J., Wang, J., Ju, J., Huang, L., Ma, Q., Daut, G., & Zhu, L. 2020. Spatio-temporal variations of hydrochemistry and modern sedimentation processes in the Nam Co basin, Tibetan Plateau: Implications for carbonate precipitation. *Journal of Great Lakes Research*, *46*(4), 961–975. https://doi.org/10.1016/j.jglr.2020.04.006

Kang, W., Anslan, S., Börner, N., Schwarz, A., Schmidt, R., Künzel, S., Rioual, P., Echeverría-Galindo, P., Vences, M., Wang, J., & Schwalb, A. (2021). Diatom metabarcoding and microscopic analyses from sediment samples at Lake Nam Co, Tibet: The effect of sample-size and bioinformatics on the identified communities. *Ecological Indicators*, *121*. https://doi.org/10.1016/j.ecolind.2020.107070

Kanz, C., Aldebert, P., Althorpe, N., Baker, W., Baldwin, A., Bates, K., Browne, P., van den Broek, A., Castro, M., & Cochrane, G. (2005). The EMBL nucleotide sequence database. *Nucleic Acids Research*, *33*, D29–D33.

Karanovic, I. (2012). *Recent freshwater ostracods of the world: Crustacea, Ostracoda, Podocopida*. Springer. https://doi.org/10.1007/978-3-642-21810-1.

Karanovic, I. (2015). Barcoding of ancient lake ostracods (Crustacea) reveals cryptic speciation with extremely low distances. *PLoS One*, *10*, e0121133.

Karanovic, I., Huyen, P. T. M., Yoo, H., Nakao, Y., & Tsukagoshi, A. (2020). Shell and appendages variability in two allopatric ostracod species seen through the light of molecular data. *Contributions to Zoology*, *89*, 247–269.

Katoh, K., Rozewicki, J., & Yamada, K. D. (2017). MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, *20*, 1160–1166.

Keil, A., Berking, J., Mügler, I., Schütt, B., Schwalb, A., & Steeb, P. (2010). Hydrological and geomorphological basin and catchment characteristics of Lake Nam Co, South-Central Tibet. *Quaternary International*, *218*, 118–130.

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, *33*, 1870–1874.

Lanzen, A., Lekang, K., Jonassen, I., Thompson, E. M., & Troedsson, C. (2017). DNA extraction replicates improve diversity and compositional dissimilarity in metabarcoding of eukaryotes in marine sediments. *PLoS One*, *12*, e0179443.

Leese, F., Sander, M., Buchner, D., Elbrecht, V., Haase, P., & Zizka, V. M. A. (2020). Improved freshwater macroinvertebrate detection from eDNA through minimized non-target amplification. *bioRxiv*, 2020.2004.2027.063545.

Leray, M., Yang, J. Y., Meyer, C. P., Mills, S. C., Agudelo, N., Ranwez, V., Boehm, J. T., & Machida, R. J. (2013). A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: Application for characterizing coral reef fish gut contents. *Frontiers in Zoology*, *10*(1), 34.

Macario-González, L., Cohuo, S., Elías-Gutiérrez, M., Vences, M., Pérez, L., & Schwalb, A. (2018). Integrative taxonomy of freshwater ostracodes (Crustacea: Ostracoda) of the Yucatán Peninsula, implications for paleoenvironmental reconstructions in the northern Neotropical region. *Zoologischer Anzeiger*, *275*, 20–36.

Macher, J. N., Vivancos, A., Piggott, J. J., Centeno, F. C., Matthaei, C. D., & Leese, F. (2018). Comparison of environmental DNA and bulk-sample metabarcoding using highly degenerate cytochrome c oxidase I primers. *Molecular Ecology Resources*, *18*, 1456–1468.

Meisch, C. (2000). Freshwater Ostracoda of western and central Europe. *Crustacea*.

Mezquita, F., Roca, J. R., Reed, J. M., & Wansard, G. (2005). Quantifying species–environment relationships in non-marine Ostracoda for ecological and palaeoecological studies: examples using Iberian data. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *225*(1–4), 93–117. https://doi.org/10.1016/j.palaeo.2004.02.052

Mischke, S. (2012). Quaternary ostracods from the Tibetan Plateau and their significance for environmental and climate-change studies. In *Developments in Quaternary Sciences* (pp. 263–279). : Elsevier.

Nascimento, F. J., Lallias, D., Bik, H. M., & Creer, S. (2018). Sample size effects on the assessment of eukaryotic diversity and community structure in aquatic sediments using high-throughput sequencing. *Scientific Reports*, *8*.

Nichols, R. V., Vollmers, C., Newsom, L. A., Wang, Y., Heintzman, P. D., Leighton, M., Green, R. E., & Shapiro, B. (2018). Minimizing polymerase biases in metabarcoding. *Molecular Ecology Resources*, *18*, 927–939.

Nigro, L. M., Angel, M. V., Blachowiak-Samolyk, K., Hopcroft, R. R., & Bucklin, A. (2016). Identification, discrimination, and discovery of species of marine planktonic ostracods using DNA barcodes. *PLoS One*, *11*, e0146327.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R., Simpson, G. L., Solymos, P., Stevens, M., & Wagner, H. (2015). vegan: Community Ecology Package. R package version 2.4-2. http://CRAN.R-project.org/package=vegan

Palumbi, S., Martin, A., Romano, S., McMillan, W., Stice, L., Grabowski, G., & MacMillan, W. (2002). *The simple fool's guide to PCR, version 2.0*. Dept. of Zoology and Kewalo Marine Laboratory, University of Hawaii.

Pérez, L., Lorenschat, J., Massaferro, J., Pailles, C., Sylvestre, F., Hollwedel, W., Brandorff, G. O., Brenner, M., Islebe, G., Lozano, M. D. S., Scharf, B., & Schwalb, A. (2013). Bioindicators of climate and trophic state in lowland and highland aquatic ecosystems of the Northern Neotropics. *Revista De Biologia Tropical*, *61*, 603–644.

Pesant, S., Not, F., Picheral, M., Kandels-Lewis, S., Le Bescot, N., Gorsky, G., Iudicone, D., Karsenti, E., Speich, S., Troublé, R., Dimier, C., & Searson, S. (2015). Open science resources for the discovery and analysis of Tara Oceans data. *Scientific Data*, *2*, 150023.

Prosser, S., Martínez-Arce, A., & Elías-Gutiérrez, M. (2013). A new set of primers for COI amplification from freshwater microcrustaceans. *Molecular Ecology Resources*, *13*, 1151–1155.

R-Core-Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ruiz, F., Abad, M., Bodergat, A. M., Carbonel, P., Rodriguez-Lazaro, J., Gonzalez-Regalado, M. L., Toscano, A., Garcia, E. X., & Prenda, J. (2013). Freshwater ostracods as environmental tracers. *International Journal of Environmental Science and Technology*, *10*, 1115–1128.

Ruppert, K. M., Kline, R. J., & Rahman, M. S. (2019). Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. *Global Ecology and Conservation*, *17*, e00547. https://doi.org/10.1016/j.gecco.2019.e00547

Schön, I., & Martens, K. (2016). Ostracod (Ostracoda, Crustacea) genomics — Promises and challenges. *Marine Genomics*, *29*, 19–25. http://dx.doi.org/10.1016/j.margen.2016.03.008

Schön, I., Poux, C., Verheyen, E., & Martens, K. (2014). High cryptic diversity and persistent lineage segregation in endemic Romecytheridea

(Crustacea, Ostracoda) from the ancient Lake Tanganyika (East Africa). *Hydrobiologia*, *739*, 119–131.
Schroeder, A., Stanković, D., Pallavicini, A., Gionechetti, F., Pansera, M., & Camatti, E. (2020). DNA metabarcoding and morphological analysis - Assessment of zooplankton biodiversity in transitional waters. *Marine Environmental Research*, *160*, 104946.
Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M. D. M., Breiner, H.-W., & Richards, T. A. (2010). Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Molecular Ecology*, *19*, 21–31.
Tang, C. Q., Leasi, F., Obertegger, U., Kieneke, A., Barraclough, T. G., & Fontaneto, D. (2012). The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in biodiversity surveys of the meiofauna. *Proceedings of the National Academy of Sciences of USA*, *109*, 16208–16212.
Tedersoo, L., Bahram, M., Polme, S., Koljalg, U., Yorou, N. S., Wijesundera, R., Ruiz, L. V., Vasco-Palacios, A. M., Thu, P. Q., Suija, A., Smith, M. E., Sharp, C., Saluveer, E., Saitta, A., Rosas, M., Riit, T., Ratkowsky, D., Pritsch, K., Poldmaa, K., … Abarenkov, K. (2014). Global diversity and geography of soil fungi. *Science*, *346*, 1078.
Thomsen, P. F., & Willerslev, E. (2015). Environmental DNA – An emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation*, *183*, 4–18.
Wang, J., Huang, L., Ju, J., Daut, G., Ma, Q., Zhu, L., Haberzettl, T., Baade, J., Mäusbacher, R., Hamilton, A., Graves, K., Olsthoorn, J., & Laval, B. E. (2020). Seasonal stratification of a deep, high-altitude, dimictic lake: Nam Co, Tibetan Plateau. *Journal of Hydrology*, *584*, 124668.
Wang, J., Huang, L., Ju, J., Daut, G., Wang, Y., Ma, Q., Zhu, L., Haberzettl, T., Baade, J., & Mäusbacher, R. (2019). Spatial and temporal variations in water temperature in a high-altitude deep dimictic mountain lake (Nam Co), central Tibetan Plateau. *Journal of Great Lakes Research*, *45*, 212–223.
Weigand, A. M., & Macher, J.-N. (2018). A DNA metabarcoding protocol for hyporheic freshwater meiofauna: Evaluating highly degenerate COI primers and replication strategy. *Metabarcoding and Metagenomics*, *2*, e26869.
Wrozyna, C., Frenzel, P., Manping, X., Zhu, L., & Schwalb, A. (2009b). A taxonomic and ecological overview of recent and Holocene ostracodes of the Nam Co region. *Southern Tibet*. *29*(4): 665-677.
Wrozyna, C., Frenzel, P., Steeb, P., Zhu, L., & Schwalb, A. (2009a). Recent lacustrine Ostracoda and a first transfer function for palaeo-water depth estimation in Nam Co, southern Tibetan Plateau. *Revista Española De Micropaleontología*, *41*, 1.
Wrozyna, C., Frenzel, P., Steeb, P., Zhu, L. P., van Geldern, R., Mackensen, A., & Schwalb, A. (2010). Stable isotope and ostracode species assemblage evidence for lake level changes of Nam Co, southern Tibet, during the past 600 years. *Quaternary International*, *212*, 2–13.
Yamaguchi, S., & Endo, K. (2003). Molecular phylogeny of Ostracoda (Crustacea) inferred from 18S ribosomal DNA sequences: implication for its origin and diversification. *Marine Biology*, *143*, 23–38.
Zemb, O., Achard, C. S., Hamelin, J., De Almeida, M.-L., Gabinaud, B., Cauquil, L., Verschuren, L. M. G., & Godon, J.-J. (2020). Absolute quantitation of microbes using 16S rRNA gene metabarcoding: A rapid normalization of relative abundances by quantitative PCR targeting a 16S rRNA gene spike-in standard. *MicrobiologyOpen*, *9*, e977.
Zhan, A., Bailey, S. A., Heath, D. D., & Macisaac, H. J. (2014). Performance comparison of genetic markers for high-throughput sequencing-based biodiversity assessment in complex communities. *Molecular Ecology Resources*, *14*, 1049–1059.
Zhan, A., & MacIsaac, H. J. (2015). Rare biosphere exploration using high-throughput sequencing: research progress and perspectives. *Conservation Genetics*, *16*, 513–522.
Zhu, L. P., Peng, P., Xie, M. P., Wang, J. B., Frenzel, P., Wrozyna, C., & Schwalb, A. (2010). Ostracod-based environmental reconstruction over the last 8,400 years of Nam Co Lake on the Tibetan plateau. *Hydrobiologia*, *648*, 157–174.
Zimmermann, J., Glöckner, G., Jahn, R., Enke, N., & Gemeinholzer, B. (2015). Metabarcoding vs. morphological identification to assess diatom diversity in environmental studies. *Molecular Ecology Resources*, *15*, 526–542.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Echeverría-Galindo, P., Anslan, S., Frenzel, P., Künzel, S., Vences, M., Pérez, L., Börner, N., Kang, W., Schwarz, A., Wang, J., Peng, P., Zhu, L., & Schwalb, A.. (2021). High-throughput identification of non-marine Ostracoda from the Tibetan Plateau: Evaluating the success of various primers on sedimentary DNA samples. *Environmental DNA*, 3, 982–996. https://doi.org/10.1002/edn3.222