Conservation Genet Resour (2016) 8:323–327
DOI 10.1007/s12686-016-0550-y

METHODS AND RESOURCES ARTICLE

Freshwater vertebrate metabarcoding on Illumina platforms using double-indexed primers of the mitochondrial 16S rRNA gene

Miguel Vences[1] · Mariana L. Lyra[2] · R. G. Bina Perl[1] · Molly C. Bletz[1] · David Stanković[3] · Carla Martins Lopes[2] · Michael Jarek[4] · Sabin Bhuju[4] · Robert Geffers[4] · Célio F. B. Haddad[2] · Sebastian Steinfartz[1]

Received: 10 February 2016 / Accepted: 21 May 2016 / Published online: 28 May 2016

Abstract Metabarcoding is a promising tool for biodiversity inventories and other applications in conservation genetics. We developed a new pair of primers for efficient and affordable high-throughput analysis of a 250 base pair stretch of DNA from the mitochondrial 16S rRNA gene of vertebrates, especially amphibians and fishes. By adapting a double-indexed protocol for Illumina platforms, our approach allows pooling of hundreds of samples in a single sequencing run. We obtained high detection rates of 82–93 % for fish in two German streams, 70 % for mock mixes of DNA from amphibians and fishes, and could distinguish multiple gene copies in amphibians, probably caused by nuclear-mitochondrial transposed DNA or heteroplasmy.

Keywords eDNA · Double-indexed primers · Amphibians · *Discoglossus* · NUMT

Electronic supplementary material The online version of this article (doi:10.1007/s12686-016-0550-y) contains supplementary material, which is available to authorized users.

✉ Miguel Vences
m.vences@tu-bs.de

[1] Zoological Institute, Braunschweig University of Technology, Mendelssohnstr. 4, 38106 Brunswick, Germany

[2] Departamento de Zoologia, Instituto de Biociências, Universidade Estadual Paulista "Júlio de Mesquita Filho", Rio Claro, SP 13506-900, Brazil

[3] Laboratory of Genetics, Department of Life Science, University of Trieste, Via Licio Giorgieri 5, 34127 Trieste, Italy

[4] Department of Genome Analytics, Helmholtz Centre for Infection Research, Brunswick, Germany

Introduction

Metabarcoding, i.e., the simultaneous molecular identification of multiple species from pooled samples of DNA or tissue, is a powerful tool for high-throughput biodiversity inventories (Barnes and Turner 2016), especially if coupled with direct isolation of environmental DNA (eDNA; Taberlet et al. 2012a; Thomsen and Willerslev 2015).

Usually, eDNA is a complex mixture of DNA from nuclei, mitochondria and chloroplasts of many different organisms, and is more or less ubiquitous in water and different substrates (Taberlet et al. 2012a). Microbiologists have been analyzing eDNA from water for more than 15 years, and this approach has since been extended to algae, zooplankton, macroinvertebrates and vertebrates (Ficetola et al. 2008; Foote et al. 2012; Goldberg et al. 2011; Lindeque et al. 2013, Thomsen et al. 2012a, b; Thomsen and Willerslev 2015).

In aquatic environments, rapid diffusion of DNA from its source means that the presence of a specific animal can theoretically be detected anywhere within the water body and not just at its point of origin (Pilliod et al. 2014). This approach is particularly useful for those species that are difficult to detect using conventional methods or are very rare (Rees et al. 2014).

Studies based on eDNA can either be focused on detecting DNA from a single taxon with specific primers (eDNA barcoding), through classical polymerase chain reaction (PCR; Ficetola et al. 2008) or quantitative PCR (qPCR; Foote et al. 2012). Alternatively, completing PCR with universal primers followed by high-throughput sequencing techniques has the potential to simultaneously determine entire species communities by eDNA metabarcoding, even if the precise composition of these

assemblages is a priori unknown (Taberlet et al. 2012b; Valentini et al. 2016).

Here, we propose a protocol for metabarcoding freshwater vertebrate communities based on newly developed primers for a ~250 base pair (bp) segment of the mitochondrial 16S rRNA gene, one of the markers with highest taxon coverage among published sequences. We also report on an example that highlights the potential of these primers to quantify heteroplasmy and nuclear pseudocopies of mitochondrial DNA in individual organisms.

Methods

We designed new primers for 16S sections conserved in vertebrates and flanking a hypervariable loop region of the 16S rRNA gene (approximately 250 bp): Vert-16S-eDNA-F1 (AGACGAGAAGACCCYdTGGAGCTT), Vert-16S-eDNA-R1 (GATCCAACATCGAGGTCGTAA). These were modified for usage in a dual-index framework for Illumina platforms. Each construct contained the Illumina adapter, a unique 8-bp index sequence, a 10-bp pad sequence, a 2-bp link sequence, and the primer sequence. Pad and primer sequences were adapted so that the sequencing primers had melting temperatures near 65 °C and the link sequence was anti-complementary to the known sequences (Table S1). Read primers consisted of pad + link + forward primer, and pad + link + reverse primer; the index read primer was the reverse complement of pad + link + reverse primer. We also adopted the same approach for the V9 region of the nuclear 18S rRNA gene (primers as in Amaral-Zettler et al. 2009; Table S2).

A diverse series of 218 PCR products was obtained with Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific) in 20 µl reactions. Three different types of DNA samples were analyzed with our approach including eDNA samples from water filters, in vitro prepared mock mixes of DNA templates, and single-species genomic DNA, as follows: (1) eDNA from eight water filters from three ponds at Itapé (Rio Claro municipality, São Paulo State, Brazil), obtained by filtering 20–100 ml water with sterile Millipore Millex syringe filters (Durapore—pore size 0.22 µm, diameter 33 mm; ref SLGV033RS), aided by a standard laboratory vacuum pump. Water was collected in the field, brought to the lab, and filtered 2–3 h after collection. Samples included two replicates from one ponds, and three replicates from two ponds. The filter holders were broken and the filter membranes were removed for DNA extraction. (2) eDNA from six samples obtained by filtering 20 l water in the field, in two of the three Itapé ponds, using a debit of 1.6 l/min, with two Envirochek HV sampling capsules (Pall Life Sciences) and a peristaltic pump (Solinst, model 410). For each pond, one filtering capsule was used. DNA was extracted twice from each capsule: once by directly adding 1 ml of the buffer contained in the capsule, and once by cutting a small piece of the filter itself. (3) eDNA from seven samples obtained by filtering 1–2 l water from two streams and two ponds in the area of Braunschweig (Lower Saxony, Germany), and from three aquaria at the corresponding author's institution. A vacuum pump (Pall Life Sciences, Model DOA-P730-BN) was used to pull collected water through sterile Nalgene filters (CN membrane) of 0.45 µm pore size and 47 mm diameter. Filters were opened and the filter membranes were used for DNA extraction. DNA extraction from all eDNA samples was performed using MoBio power water extraction kits, following the manufacturer's instructions. (4) 36 artificially constructed mock mixes of DNA templates from different numbers of amphibian or fish species. These were mixed from DNA that had been extracted using standard salt extraction protocols from roughly similar amounts of muscle tissue, with equal volumes (1 µl) per template included in a mix. (5) Single-species genomic DNA extracted from muscle tissues of 161 individuals of painted frogs (genus *Discoglossus*) previously found to be affected by phenomena of nuclear-mitochondrial transposed DNA (NUMTs), or by heteroplasmy, i.e., the presence of different variants of mitochondrial genes in the same individual (Vences et al. 2014).

PCRs with unique combinations of indexed forward and reverse primers for each sample were performed with Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific). Amplicon DNA concentration was roughly quantified based on the intensity of their signal on agarose gels and 2–6 µl added to a pooled library. The whole library was gel-purified by cutting out the band of the correct amplicon size and cleaning it up with a Qiagen MinElute kit (the entire library spinned down in two columns and each eluted in 12 µl per column). Quality control of the normalized amplicon library pool was done using a Bioanalyzer (Agilent) and the concentration was determined using a Qubit 2.0 fluorometer. Sequencing was performed on the Illumina MiSeq platform using MiSeq Reagent Kit v2 for 250 cycles in both directions. The sequencing cartridge was prepared by piercing the foil over wells 12, 13, 14 and 17. A volume of 3 µl of Read 1 primer (100 µM) was pipetted and properly mixed into reservoir 12. Similarly, Index primer and Read 2 primer were pipetted and mixed into well 13 and 14 respectively. For sequencing 600 µl of 10 pM of the denatured library with 10 % phiX was pipetted into well 17.

Sequences were processed with MacQIIME v1.9.1 (Caporaso et al. 2010), filtering the forward reads as follows: ambiguous calls, <75 % consecutive base call with a quality score >5, >3 consecutive low-quality base calls, length <100 or >300 base pairs. Sequences were clustered

Table 1 Detection of fish by eDNA, compared to results of fishing surveys by the Niedersächsisches Landesamt für Verbraucherschutz und Lebensmittelsicherheit in 2013, in two streams (Mittelriede and Schunter) in Braunschweig, Germany

Species	BLAST identity (%)	Schunter amplicon reads	Schunter fishing survey (locality Querum—2013)	Mittelriede amplicon reads	Mittelriede fishing survey (locality Vossenkamp—2014)
Rutilus rutilus	100	4046	Yes	89	Yes
Perca fluviatilis	100	826	Yes	11	Yes
Blicca bjoerkna	100	819	Yes	7	–
Leuciscus leuciscus	99.2	607	Yes	19	Yes
Esox lucius	100	294	Yes	52	Yes
Gobio gobio	100	260	Yes	34	Yes
Gasterosteus aculeatus	100	258	Yes	46	Yes
Barbatula barbatula	93.2	125	Yes	–	Yes
Tinca tinca	100	122	Yes	–	Yes
Scardinius erythrophthalmus	100	115	Yes	–	Yes
Pungitius pungitius	98.52	28	Yes	–	–
Chondrostoma toxostoma	98.2	14	–	–	–
Cottus gobio	100	–	–	19	Yes
Cyprinus carpio	100	–	–	9	–
Anguilla anguilla	99.1	–	Yes	7	Yes
Salmo trutta	–	–	–	–	Yes
Squalius cephalus	–	–	Yes	–	–
Pseudorasbora parva	–	–	Yes	–	Yes

into operational taxonomic units at 90 % sequence identity. For *Discoglossus* samples, paired-end reads from each sample were joined into consensus sequences using Fastq-join with default settings (Aronesty 2011). Representative sequences were locally compared with BLAST+ against custom reference databases. Original data were deposited in the NCBI short read database (Bioproject PRJNA320967).

Results and discussion

The number of sequence reads per amplicon ranged from 895–101,146, with averages of 10,948–21,891. Overall, the primers successfully amplified a wide array of freshwater vertebrates.

Amplicons obtained from eight water filters from three ponds in Itapé (Brazil) revealed, with high read numbers, six common amphibians from the site (*Rhinella ornata, Dendropsophus minutus, Hypsiboas albopunctatus, H. faber, H. lundii,* and *Elachistocleis cesarii*), plus three other amphibians, 5–6 species of fish, and 11 other vertebrates. The different types of filters used (Envirochek capsule vs. Durapore syringe filters) all had comparable success rates, although our sampling design with only two Envirocheck filters was not suitable for a statistical comparison of the performance of filtering methods.

From the seven water filters sampled in Germany, two refer to streams for which comparative fish survey data were available (Table 1). The method identified 82–93 % of the fish species known from these streams, plus three additional species. Filtered aquarium water yielded all species contained in those, including a turtle (*Pelusios*) and a newt (*Calotriton*). Filters from ponds near Braunschweig yielded, besides numerous fish species, also other vertebrates associated to the ponds or their shores, such as birds (*Anas*, *Anser*, *Fulica*) and mammals (*Microtus, Sus*).

Mock mixes of non-normalized DNA of amphibians and fishes yielded a detection rate of 85–100 % in mixes of 40 species or less, and an overall rate of 70 %, with only single false positives (Tables S3-S4).

We also tested our approach with 161 samples of genomic DNA of frogs of the genus *Discoglossus* known to be affected by NUMTs or heteroplasmy (Vences et al.

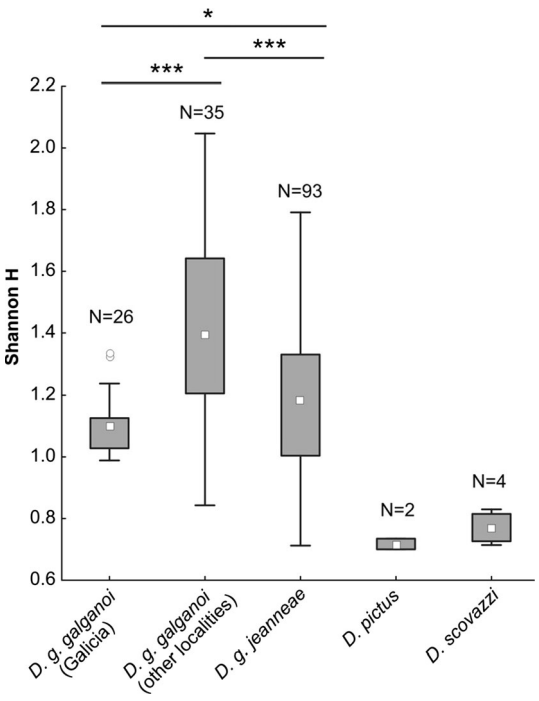

Fig. 1 16S allele diversity (calculated as Shannon H index) in tissue samples of *Discoglossus* individuals. Statistical comparisons by Mann–Whitney *U* tests: * $P < 0.05$, *** $P < 0.001$

2014). All *Discoglossus* individuals had one predominant sequence that corresponded to the taxon inferred previously by Sanger sequencing. Proportional to the total number of reads per samples, this predominant sequence made up on average 79 % (43–96 %) in *D. galganoi galganoi*, 84 % (52–100 %) in *D. g. jeanneae*, 98–99 % in *D. pictus*, and 94–99 % in *D. scovazzi*. In *D. g. galganoi* on average 1 % (0–8 %) of the reads corresponded to the "wrong" subspecies *D. g. jeanneae* whereas in *D. g. jeanneae* 10 % of reads (0–37 %) blasted to the "wrong" subspecies *D. g. galganoi*.

Shannon diversity indices (Fig. 1) indicated sequence diversity was lowest in *D. pictus* and *D. scovazzi*. Within *D. galganoi*, the populations of *D. g. galganoi* from Galicia differed only slightly from *D. g. jeanneae*, whereas *D. g. galganoi* from other localities had distinctly higher diversity values.

While a detailed discussion of these results is beyond the scope of the present paper, these results confirm an influence of the hybridization between *D. g. galganoi* and *D. g. jeanneae* on the presence of NUMTs or, less likely, heteroplasmy. The fact that *D. g. galganoi* alleles were detected rather commonly in *D. g. jeanneae* but not vice versa suggests an asymmetric introgression, with mtDNA of *D. g. galganoi* invading the gene pool of *D. g. jeanneae*. This distinct pattern indicates that the sequence variation is indeed a biological phenomenon (i.e., different sequence variants present in the same individual) and not explainable by errors of PCR, sequencing, or sequence assembly.

The main goal of this study was the development of a cost-effective set of eDNA primers and to confirm their utility under a variety of conditions. Our sampling design and sample sizes do not allow to draw thorough comparative conclusions, e.g., about the performance of different filters and filtering protocols. Yet, the proposed primers and protocols proved remarkably robust and versatile, given that all sets of samples yielded meaningful results despite the diversity of sampling methods and sample characteristics.

Our results show that eDNA from water filters can yield reasonable amounts of reads for a comparatively long stretch of the mitochondrial DNA (ca. 250 bp). Despite some contamination issues (*Discoglossus* sequences were present among eDNA amplicons), we did not detect cross contamination among eDNA samples and only a small number of false positives in the mock mix analyses. It is obvious however, that results must be interpreted with thought and prudence (Gilbert et al. 2005), especially when applying the method to monitoring rare species.

For metabarcoding and eDNA analysis, the *Discoglossus* example provides a cautionary tale. First, in cases of hybrid zones, the detection of additional haplotypes of one taxon might not reliably indicate its real presence, but in some cases can be caused not only by introgression but also by NUMTs in individuals of another taxon. Second, describing species diversity and detecting undescribed candidate species with eDNA requires extreme caution because NUMTs can be frequent in some species (e.g., Podnar et al. 2007; Baldo et al. 2011; Miraldo et al. 2012). Although unambiguous reports of heteroplasmy in non-model organisms are rare, heteroplasmy is today seen as common in animals and plants (Kmiec et al. 2006) and is a well-studied phenomenon in humans (Stewart and Chinnery 2015). In metabarcoding studies, these DNA variants can be mistaken for sequences of unknown or additional species, and biodiversity thus overestimated (Song et al. 2008), as has been previously noted for other eDNA data sets (e.g., Vestheim and Jarman 2008).

The method presented here allows for a simple and high-throughput sequencing of eDNA, combining hundreds of samples in one library. This will be particularly useful if detection of only the more common community members is sufficient. If the goal is instead to detect rare species, it will be worth comparing the efficiency of strategies using few samples at great sequencing depth, versus many replicates from the same water bodies at low sequencing depth as is possible with our approach. Double barcoded primer systems following Kozich et al. (2013) could also be developed for shorter DNA stretches and serve to sequence more degraded DNA (Boyer et al. 2012), e.g.,

from preserved museum specimens (Shokralla et al. 2011). Adapting the approach to primers for "minibarcodes" of the mitochondrial cox1 gene should also be possible, as tools for optimal primer design for such applications have been developed (e.g., Boyer et al. 2012).

Acknowledgments We are grateful to Bárbara Zaidan, Michelle Abstein, Vanessa Kuschnereit (field work), Reinhard Huwe, Meike Kondermann (technical assistance), Eva Mosch, Uwe Kirchberger (data on fish surveys), Maria José O. Campos and Luiz Marcelo de Carvalho (access to field site). We acknowledge financial support by CAPES to MV, CFBH and ML (88881.062205/2014-01), and São Paulo Research Foundation (Grant #2013/50741-7) and CNPq to CFBH.

References

Amaral-Zettler LA, McCliment EA, Ducklow HW, Huse SM (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

Aronesty E (2011) ea-utils: Command-line tools for processing biological sequencing data. https://code.google.com/p/ea-utils/

Baldo L, De Queiroz A, Hedin M, Hayashi CY, Gatesy J (2011) Nuclear-mitochondrial sequences as witnesses of past interbreeding and population diversity in the jumping bristletail *Mesomachilis*. Mol Biol Evol 28:195–210

Barnes MA, Turner CR (2016) The ecology of environmental DNA and implications for conservation genetics. Conserv Genet 17:1–17

Boyer S, Brown SD, Collins RA, Cruickshank RH, Lefort MC, Malumbres-Olarte J, Wratten SD (2012) Sliding window analyses for optimal selection of mini-barcodes, and application to 454-pyrosequencing for specimen identification from degraded DNA. PLoS ONE 7:e38215

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R (2010) QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335–336

Ficetola GF, Miaud C, Pompanon F, Taberlet P (2008) Species detection using environmental DNA from water samples. Biol Lett 4:423–425

Foote AD, Thomsen PF, Sveegaard S, Wahlberg M, Kielgast J, Kyhn LA, Salling AB, Galatius A, Orlando L, Gilbert MTP (2012) Investigating the potential use of environmental DNA (eDNA) for genetic monitoring of marine mammals. PLoS ONE 7:e41781

Gilbert MTP, Bandelt HJ, Hofreiter M, Barnes I (2005) Assessing ancient DNA studies. Trends Ecol Evol 20:541–544

Goldberg CS, Pilliod DS, Arkle RS, Waits LP (2011) Molecular detection of vertebrates in stream water: a demonstration using Rocky Mountain tailed frogs and Idaho giant salamanders. PLoS ONE 6:e22746

Kmiec B, Woloszynska M, Janska H (2006) Heteroplasmy as a common state of mitochondrial genetic information in plants and animals. Curr Genet 50:149–159

Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl Environ Microbiol 79:5112–5120

Lindeque PK, Parry HE, Harmer RA, Somerfield PJ, Atkinson A (2013) Next generation sequencing reveals the hidden diversity of zooplankton assemblages. PLoS ONE 6:e23398

Miraldo A, Hewitt GM, Dear PH, Paulo OS, Emerson BC (2012) Numts help to reconstruct the demographic history of the ocellated lizard (*Lacerta lepida*) in a secondary contact zone. Mol Ecol 21:1005–1018

Pilliod DS, Goldberg CS, Arkle RS, Waits LP (2014) Factors influencing detection of eDNA from a stream-dwelling amphibian. Mol Ecol Res 14:109–116

Podnar M, Haring E, Pinsker W, Mayer W (2007) Unusual origin of a nuclear pseudogene in the Italian Wall Lizard: intergenomic and interspecific transfer of a large section of the mitochondrial genome in the genus *Podarcis* (Lacertidae). J Mol Evol 64:308–320

Rees HC, Maddison BC, Middleditch DJ, Patmore HRM, Gough KC (2014) The detection of aquatic animal species using environmental DNA—a review of eDNA as a survey tool in ecology. J Appl Ecol 51:1450–1459

Shokralla S, Zhou X, Janzen DH, Hallwachs W, Landry JF, Jacobus LM, Hajibabaei M (2011) Pyrosequencing for mini-barcoding of fresh and old museum specimens. PLoS ONE 6:e21252

Song H, Buhay JE, Whiting MF, Crandall KA (2008) Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. Proc Natl Acad Sci USA 105:13486–13491

Stewart JB, Chinnery PF (2015) The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. Nat Rev Genet 16:530–542

Taberlet P, Coissac E, Hajibabae M, Rieseberg LH (2012a) Environmental DNA. Mol Ecol 21:1789–1793

Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012b) Towards next-generation biodiversity assessment using DNA metabarcoding. Mol Ecol 21:2045–2050

Thomsen PF, Willerslev E (2015) Environmental DNA—an emerging tool in conservation for monitoring past and present biodiversity. Biol Conserv 183:4–18

Thomsen PF, Kielgast J, Iversen LL, Møller PR, Rasmussen M, Willerslev E (2012a) Detection of a diverse marine fish fauna using environmental DNA from seawater samples. PLoS ONE 7:e41732

Thomsen PF, Kielgast J, Iversen LL, Wiuf C, Rasmussen M, Gilbert MTP, Orlando L, Willerslev E (2012b) Monitoring endangered freshwater biodiversity using environmental DNA. Mol Ecol 21:2565–2573

Valentini A, Taberlet P, Miaud C, Civade R, Herder J, Thomsen PF, Bellemain E, Besnard A, Coissac E, Boyer F, Gaboriaud C, Jean P, Poulet N, Roset N, Copp GH, Geniez P, Pont D, Argillier C, Baudoin JM, Peroux T, Crivelli AJ, Olivier A, Acqueberge M, Le Brun M, Møller PR, Willerslev E, Dejean T (2016) Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. Mol Ecol 25:929–942

Vences M, de Pous P, Nicolas V, Díaz-Rodríguez J, Donaire D, Hugemann K, Hauswaldt JS, Amat F, Barnestein JAM, Bogaerts S, Bouazza A, Carranza S, Galán P, González de la Vega JP, Joger U, Lansari A, El Mouden EH, Ohler A, Sanuy D, Slimani T, Tejedo M (2014) New insights on phylogeography and distribution of painted frogs (*Discoglossus*) in northern Africa and the Iberian Peninsula. Amphibia-Reptilia 35:305–320

Vestheim H, Jarman SN (2008) Blocking primers to enhance PCR amplification of rare sequences in mixed samples—a case study on prey DNA in Antarctic krill stomachs. Front Zool 5:12