University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Faculty Publications from the Harold W. Manter Laboratory of Parasitology

Parasitology, Harold W. Manter Laboratory of

11-2018

Museum Metabarcoding: A Novel Method Revealing Gut Helminth Communities of Small Mammals across Space and Time

Stephen E. Greiman
Georgia Southern University, sgreiman@georgiasouthern.edu

Joseph A. Cook
University of New Mexico, cookjose@unm.edu

Vasyl V. Tkach
University of North Dakota, vasyl.tkach@und.nodak.edu

Eric P. Hoberg
University of New Mexico, geocolonizer@gmail.com

Damian M. Menning
United States Geological Survey, Alaska Science Center, dmenning@usgs.gov

See next page for additional authors

Follow this and additional works at: http://digitalcommons.unl.edu/parasitologyfacpubs

🔵 Part of the Biodiversity Commons, Bioinformatics Commons, Museum Studies Commons, Parasitology Commons, Terrestrial and Aquatic Ecology Commons, and the Zoology Commons

Authors

Stephen E. Greiman, Joseph A. Cook, Vasyl V. Tkach, Eric P. Hoberg, Damian M. Menning, Andrew G. Hope, Sarah A. Sonsthagen, and Sandra L. Talbot

Museum metabarcoding: A novel method revealing gut helminth communities of small mammals across space and time

Stephen E. Greiman [a,b,*], Joseph A. Cook [b], Vasyl V. Tkach [c], Eric P. Hoberg [b], Damian M. Menning [d], Andrew G. Hope [e], Sarah A. Sonsthagen [d], Sandra L. Talbot [d]

[a] Department of Biology, Georgia Southern University, 4324 Old Register Rd., Statesboro, GA 30460, USA
[b] Department of Biology and Museum of Southwestern Biology, University of New Mexico, CERIA, 302 Yale Blvd NE, Albuquerque, NM 87106, USA
[c] Department of Biology, University of North Dakota, 10 Cornell St., Grand Forks 58202, ND, USA
[d] US Geological Survey, Alaska Science Center, 4210 University Dr., Anchorage, AK 99508, USA
[e] Division of Biology, Kansas State University, 116 Ackert Hall, Manhattan, KS 66506, USA

ABSTRACT

Natural history collections spanning multiple decades provide fundamental historical baselines to measure and understand changing biodiversity. New technologies such as next generation DNA sequencing have considerably increased the potential of museum specimens to address significant questions regarding the impact of environmental changes on host and parasite/pathogen dynamics. We developed a new technique to identify intestinal helminth parasites and applied it to shrews (Eulipotyphla: Soricidae) because they are ubiquitous, occupy diverse habitats, and host a diverse and abundant parasite fauna. Notably, we included museum specimens preserved in various ways to explore the efficacy of using metabarcoding analyses that may enable identification of helminth symbiont communities from historical archives. We successfully sequenced the parasite communities (using 12S mtDNA, 16S mtDNA, 28S rDNA) of 23 whole gastrointestinal tracts. All gastrointestinal tracts were obtained from the Museum of Southwestern Biology, USA, and from recent field collections, varying both in time since fixation (ranging from 4 months to 16 years) and preservation method (70% or 95% ethanol stored at room temperature, or flash frozen in liquid nitrogen and stored at −80 °C). Our proof of concept demonstrates the feasibility of applying next generation DNA sequencing techniques to authoritatively identify the parasite/pathogen communities within whole gastrointestinal tracts from museum specimens of varying age and fixation, and the value of future preservation of host-associated whole gastrointestinal tracts in public research archives. This powerful approach facilitates future comparative examinations of the distributions and interactions among multiple associated groups of organisms through time and space.

© 2018 Australian Society for Parasitology. Published by Elsevier Ltd. All rights reserved.

1. Introduction

Museum collections were built to discover and document biodiversity and now serve as primary informatics resources for understanding the history and future of the biosphere (e.g., Hoberg et al., 2013; McLean et al., 2016). Specimen archives have also proved central to recent exploration of the effects of changing environments on organismal population decline (Shaffer et al., 1998; Suarez and Tsutsui, 2004; Rowe et al., 2011). Voucher specimens held in museum repositories, including ethanol (EtOH)-preserved organisms, are essential biodiversity infrastructure that can lead to a better understanding of the ecology, evolution and distribution of free living vertebrate and invertebrate animals as well as their diverse associated parasites, pathogens and other symbionts.

New technologies in molecular ecology, especially high throughput DNA sequencing, have rejuvenated the museum collection enterprise and prompted novel uses for museum samples. Specimens too degraded for classical PCR and Sanger sequencing can now be integrated through the use of next generation sequencing (NGS) technologies (Bi et al., 2013; Guschanski et al., 2013; Besnard et al., 2016). These new methods have been applied to studies of the evolutionary history and phylogeography of a variety of vertebrate (e.g., McCormack et al., 2012; Besnard et al., 2016; Melville et al., 2017) and invertebrate taxa (e.g., Haponski and Stepien, 2016; Yuan et al., 2016; Allen et al., 2017) and to explore the microbiomes of a wide range of animals (e.g., Phillips et al.,

2012; Bourne et al., 2013; Carrillo-Araujo et al., 2015; Jiménez and Sommer, 2017). Our study extends this approach to characterise communities of helminth parasites in preserved vertebrate specimens.

Helminths represent a wide array of parasitic organisms belonging to three phyla (Platyhelminthes, Nematoda, Acanthocephala) and most have complex life cycles, often using both invertebrates (commonly mollusks and arthropods) and vertebrates as hosts. Due to these obligate, diverse host relationships, parasites may act as indicators of ecosystem quality, with healthier and more heterogeneous ecosystems having higher parasite species richness (Hudson et al., 2006). Given current threats to biodiversity, particularly through anthropogenic environmental perturbation, understanding temporal changes in the diversity and distribution of both free-living organisms and their parasites will be critical for predicting future ecosystem responses. This is especially true for the changing dynamics of emerging pathogens and infectious diseases (Hoberg et al., 2013; Brooks et al., 2014). A number of parasitic worms are recognised pathogens of humans, domestic animals, and wildlife, although the level of pathogenicity of the majority of helminths, especially those parasitizing wild animals, is unknown. Rapid and comprehensive assessments of pathogens, including helminths, remain difficult (Hoberg et al., 2015). Identification of helminth species requires time-intensive processing by expert taxonomists, which becomes problematic given rigorous sample sizes, from multiple hosts, and with broad geographic coverage. This is particularly important for the identification of voucher specimens, as misidentification almost certainly leads to perpetuation of errors.

Our model system involves the helminth faunas of shrews, but could easily be scaled taxonomically, spatially and temporally. Many vertebrate species harbour diverse assemblages of helminths, and insectivorous small mammals in the genus *Sorex* (long-tailed shrews) are a prime example. Shrews in North America, north of Mexico, occupy diverse habitats and host a speciose and abundant parasite fauna, which comprises 97 currently known helminth species, including nine trematodes, 39 cestodes, 50 nematodes, and four acanthocephalans (likely representing only a fraction of true species diversity). This diversity reflects a diet consisting almost exclusively of invertebrates that act as intermediate hosts (Kinsella and Tkach, 2009). It is extremely rare to find a shrew (*Sorex* spp.) uninfected with helminths, especially cestodes. Indeed, most shrews harbour hundreds of individual cestodes ranging in size from less than 1 mm to 90 mm in length, with the majority in the range of 1–10 mm (Kinsella and Tkach, 2009). At any given time, a single shrew may be infected with representatives of several genera of cestodes and nematodes, including multiple congeneric species.

The 10 largest mammal archives in North America (National Museum of Natural History, Museum of Southwestern Biology (MSB), American Museum of Natural History, Berkeley Museum of Vertebrate Zoology, Field Museum, Kansas University Natural History Museum and Biodiversity Research Center, University of Michigan Museum of Zoology, University of Alaska Museum, Royal Ontario Museum, Carnegie Museum of Natural History) collectively house over 2.3 million small mammal specimens. Although most are stored as dried skins and skeletons without gastrointestinal (GI) tracts, a large number of whole specimens are stored in 70–95% EtOH or frozen at −20 °C. Relatively fewer specimens — over 500,000 — have parts or organs, including the GI tracts, frozen at −80 °C. Whether GI tracts are frozen or EtOH-preserved, the helminths in these preserved GI tracts are seldom of sufficient quality for morphological identification due to contraction or degradation. To document the helminth diversity in wild rats, Tanaka et al. (2014) developed a metabarcoding approach that targets the 18S rDNA of helminths in DNA extractions from fecal samples. In this

study we developed a similar approach, but one that increases detectability (Stat et al., 2017) by leveraging multiple genes to examine the helminth diversity within whole GI tracts of shrews as a convenient model, with the methodology being applicable to a broad spectrum of vertebrate hosts. We test the feasibility of using NGS technologies to identify parasite biodiversity and community structure within museum-archived whole GI tracts, specifically targeting collections where the quality of helminth samples is too poor for morphological identification. Herein, we outline our new approach to identifying the helminth community in a series of whole GI tracts from shrew specimens that had been preserved over varying lengths of time (4 months to 16 years).

2. Materials and methods

2.1. Sample collection

Twenty-three whole GI tracts from shrews in the genus *Sorex* were sampled from museum specimens from the Museum of Southwestern Biology at the University of New Mexico in Albuquerque, USA (Table 1). Museum samples varied, with GI tracts being preserved: (i) still within the shrew host and fixed in 95% EtOH (abdomen punctured and placed directly into EtOH) and moved to 70% EtOH for long-term storage at room temperature (*n* = 6); (ii) flash frozen in liquid nitrogen (LN) and stored at −80 °C (*n* = 17); or (iii) fixed and preserved in 95% EtOH and stored at room temperature (*n* = 2; GI tracts were split at time of collection, with the small intestine of each GI tract fixed in LN and the large intestine fixed in 95% EtOH) (Table 1). GI tracts were carefully removed from each animal or vial using UV sterilised and bleached (10%) micro-forceps and dissection scissors, and placed into a sterile glass Petri dish. GI tracts were straightened within the Petri dish, cut in half, and each half opened lengthwise using sterile fine dissection scissors under a dissecting microscope.

2.2. DNA extraction from whole GI tracts

DNA was extracted from each half of the opened GI tracts using the ZR Fecal DNA MiniPrep™ kit (Zymo Research, Irvine, CA, USA) following the manufacturer's instructions with minor modifications. Each half of the GI tract from an individual shrew was extracted separately to avoid overloading spin columns. DNA eluted from each GI tract half (≈150 μL each) was combined into a single sterile microcentrifuge tube. To determine the best method for whole gut extraction, opened GI tracts were either placed directly into the ZR BashingBead™ lysis tube or the intestinal content was scraped using sterile forceps into the lysis tube. Samples within the ZR BashingBead™ lysis tube were lysed by bead beating for ≈25 min utilizing the Disruptor Genie™ (Scientific Industries, Inc., Bohemia, NY, USA) or for ≈15 min using the TissueLyser II (Qiagen, Hilden, Germany). Eluted DNA (2 μL from each sample) was quantified using the Qubit™ dsDNA Broad Range Assay Kit (ThermoFisher Scientific, Waltham, MA, USA) on a Qubit™ fluorometer. Following quantification, the remaining eluted DNA from each sample (if possible) was standardised to 50 ng/μL in new sterile 1.5 mL microcentrifuge tubes.

2.3. Reference library generation

Helminths obtained from fresh shrew GI tracts were fixed for standard morphological examination (e.g. heat relaxed and fixed in 70–80% EtOH) and used to help design NGS primers and generate DNA reference libraries for each genetic locus used in this study. Processing varied depending on the particular helminth taxonomic group (cestodes, nematodes, etc.). For armed cestodes (presence of

Table 1
Shrew gastrointestinal tract samples obtained from the Museum of Southwestern Biology Division of Mammals, New Mexico USA.

Sample ID	Preservation type	Collection Year	Locality	Species	DNA (ng/μl)
MSB:Mamm:259288 (AF51992)	Whole organism in 70% EtOH	2001	Canada, Yukon	*Sorex cinereus*	93.1
MSB:Mamm:156005 (NK153868)	Whole organism in 70% EtOH	2004	Canada, Northwest territories	*Sorex cinereus*	1.0
MSB:Mamm:195886 (NK151517)	Whole organism in 70% EtOH	2007	New Mexico	*Sorex monticola*	1.6
MSB:Mamm:250051 (NK213643)	Whole organism in 70% EtOH	2011	Alaska	*Sorex monticola*	37.8
MSB:Mamm:266233 (NK216541)	Whole organism in 70% EtOH	2013	Alaska	*Sorex monticolus*	72.1
MSB:Mamm:266518 (NK216528)	Whole organism in 70% EtOH	2013	Alaska	*Sorex monticola*	9.5
MSB:Mamm:291145 (NK261915)	GI flash frozen in LN2, stored at −80 °C	2015	New Mexico	*Sorex cinereus*	303
MSB:Mamm:291146 (NK261916)	GI flash frozen in LN2, stored at −80 °C	2015	New Mexico	*Sorex monticola*	220
MSB:Mamm:291147 (NK261917)	GI flash frozen in LN2, stored at −80 °C	2015	New Mexico	*Sorex monticola*	309
MSB:Mamm:291148 (NK261920)	½ GI flash frozen in LN2, stored at −80 °C	2015	New Mexico	*Sorex monticola*	207
MSB:Mamm:291148 (NK261920)	½ GI in 95% EtOH, stored at −20 °C	2015	New Mexico	*Sorex monticola*	76.8
MSB:Mamm:291149 (NK261921)	½ GI flash frozen in LN2, stored at −80 °C	2015	New Mexico	*Sorex monticola*	225
MSB:Mamm:291149 (NK261921)	½ GI in 95% EtOH, stored at −20 °C	2015	New Mexico	*Sorex monticola*	58.3
MSB:Mamm:291150 (NK261932)	GI flash frozen in LN2, stored at −80 °C	2015	New Mexico	*Sorex cinereus*	195
MSB:Mamm:291151 (NK261933)	GI flash frozen in LN2, stored at −80 °C	2015	New Mexico	*Sorex cinereus*	151
MSB:Mamm:291152 (NK261934)	GI flash frozen in LN2, stored at −80 °C	2015	New Mexico	*Sorex monticola*	78.7
MSB:Mamm:291153 (NK261935)	GI flash frozen in LN2, stored at −80 °C	2015	New Mexico	*Sorex monticola*	144
MSB:Mamm:291154 (NK261936)	GI flash frozen in LN2, stored at −80 °C	2015	New Mexico	*Sorex cinereus*	110
MSB:Mamm:291155 (NK261937)	GI flash frozen in LN2, stored at −80 °C	2015	New Mexico	*Sorex monticola*	213
MSB:Mamm:291156 (NK261938)	GI flash frozen in LN2, stored at −80 °C	2015	New Mexico	*Sorex monticola*	164
MSB:Mamm:291157 (NK261939)	GI flash frozen in LN2, stored at −80 °C	2015	New Mexico	*Sorex monticola*	159
MSB:Mamm:291158 (NK261940)	GI flash frozen in LN2, stored at −80 °C	2015	New Mexico	*Sorex cinereus*	238
MSB:Mamm:291159 (NK261946)	GI flash frozen in LN2, stored at −80 °C	2015	New Mexico	*Sorex monticola*	207
MSB:Mamm:291160 (NK261947)	GI flash frozen in LN2, stored at −80 °C	2015	New Mexico	*Sorex monticola*	165
MSB:Mamm:291161 (NK261948)	GI flash frozen in LN2, stored at −80 °C	2015	New Mexico	*Sorex cinereus*	129

rostellar hooks), the scolex was removed from individual worms with armed rostellums (those that bear hooks for attachment) and mounted in Berlese's clearing medium as a voucher. DNA was extracted from the remaining part of the cestode using the Zymo Research Genomic DNA-tissue MicroPrep kit following the manufacturer's protocol. Unarmed cestodes (lacking rostellar hooks) were processed using the posterior end of the strobila for DNA extraction (ZR MicroPrep kit). The remaining part of these cestodes was either placed in 80% EtOH as a voucher or stained and permanently mounted in Damar gum. For nematodes, the anterior and posterior ends were preserved in 80% EtOH as a voucher of mouthparts and genitalia, respectively, and the middle section of the worm was used for DNA extraction (ZR MicroPrep kit).

PCR and Sanger sequencing was accomplished using standardised primers and annealing temperatures (Table 2), different from metabarcoding primers. For mitochondrial genes (16S and 12S) GoTaq Colorless Master Mix (Promega, Madison, WI, USA) was used in PCRs; for 28S rDNA, Quick load OneTaq mastermix (New England Biolabs, Ipswich, MA, USA) was used.

2.4. Primer design for NGS

Primers were designed using Python (van Rossum, 1995) and Biopython (Cock et al., 2009) scripts that are part of the U.S. Geological Survey Alaska Science Center Bioinformatics pipeline (Menning, D.M., Talbot, S.L., 2018. Python scripts for bioinformatics, 2017. U.S. Geological Survey data release, https://doi.org/10.5066/F74F1NZ4). Briefly, all available nematode and cestode 28S rDNA and12S and16S mtDNA sequences were downloaded from NCBI GenBank into respective FASTA files. Each FASTA file was aligned using MEGA6 (Tamura et al., 2013). Aligned FASTA files were used to locate potential primer sites (conserved regions greater than 17 bp). Once potential primers were developed, the original unaligned FASTA files were screened to verify that only individual species would be identified and that no potential sequences had more than one unique generic/specific epithet descriptor (primer sequences in Table 3). Locus-specific primers were appended with Illumina Nextera (Illumina Inc., San Diego, CA, USA) P5 and P7 adapters with dual indices.

Table 2
PCR and Sanger sequencing primers for helminth references libraries used in this study.

Genetic Locus (annealing temp)	Primer	Sequence (5′–3′)
28S (56 °C)	PCR	
	Cestl2 (Tkach et al., 2013)	AAGCATATCAATAAGCGG
	1500RC (Tkach et al., 2013)	GACGATCGATTTGCACGTC
	Sequencing	(PCR primers and the following)
	c250f (Tkach et al., 2013)	GTCGGGTTGTTTGAGATTGC
16S (48 °C)	PCR/Sequencing	
	16SF (Littlewood et al., 2008)	TGCCTTTTGCATCATGCT
	16SR (Littlewood et al., 2008)	AATAGATAAGAACCGACCTGG
12S (45 °C)	PCR/Sequencing	
	12SF (Casiraghi et al., 2004)	GTTCCAGAATAATCGGCTA
	12SR (Casiraghi et al., 2004)	ATTGACGGATGRTTTGTACC

Table 3
Primer sequences used for next generation amplicon sequencing in this study (without adapters and index sequences).

Primer name	Gene	Primer sequence (5′–3′)
12S_F_Long	12S mtDNA	ATAAATAAGTAAAATTTGGC
12S_R_Long		GTACCACCTCTAAATAATCTTC
28S_F_Ces	28S rDNA	GAGTAAACAGTACGTGAAGC
28S_R_Ces		CCACCGGTCGTGGTGTTC
16S_F_Ces	16S mtDNA	CAATTAATTATGCTACCTT
16S_R_Ces		CGTCTGTTTATYAAAAACATTTC

2.5. DNA library preparation and sequencing

Libraries were prepared for sequencing on an Illumina platform. Precautions against contamination were taken. DNA extraction and initial library preparation steps were completed in a laboratory designated for pre-PCR protocols conducted on low copy and highly degraded DNA. Filtered barrier tips where always used, and plastics and reagents (when appropriate) were all UV treated prior to use. DNA extracts were amplified by PCR in triplicate for each locus (16S mtDNA, 12S mtDNA, 28S rDNA). Positive and negative controls were included for all loci. Negative controls consisted of ultra-pure sterile water and positive controls for cestodes consisted of one *Monocercus*, five individual *Mathevolepis*, two individual *Lineolepis* sp., one *Staphylocystis*, three *Ditestolepis* and six *Soricinia*. Positive controls for nematodes consisted of a mix of 15 individuals of two species of *Longistriata*. PCR amplifications were carried out in a 25 μL volume; 4 μL of template DNA, 2.5 μL of each primer (1.0 mM), 0.5 μL of dNTPs (0.2 mM), 2.5 μL of 1X PCR Gold buffer (ThermoFisher), 2 μL of 2.0 mM MgCl$_2$, 0.5 μL of 2% BSA, and 1 unit (0.2 μL) of AmpliTaq Gold Polymerase (ThermoFisher). Thermocycler conditions were: initial denaturation at 95 °C for 10 min, 30–35 cycles of denaturation at 94 °C for 30 s, annealing at 55 °C for 30 s, extension at 72 °C for 60 s, followed by a final extension at 72 °C for 30 min and a final hold at 4 °C. Following PCR amplification, libraries were normalised. PCR triplicates for each locus were pooled by sample, mixed by vortexing and visualised on an agarose gel. Following verification of amplification, 25 μL of each pooled PCR product were subaliquoted and enzymatically purified using 1 μL of ExoSAP-IT (ThermoFisher) per aliquoted sample and the purified PCR product was quantified using a Broad Range Quant-iT dsDNA assay kit (ThermoFisher). After quantification, PCR products were diluted to equal concentrations (no less than 30 ng/μL) using ultra-pure water; if a sample was less than 30 ng/μL then equal volumes (5 μL) were used. Diluted samples were pooled by locus into a 1.5 mL microcentrifuge tube (5 μL each), and excess primers were removed from a 25 μL subaliquot from each pooled locus using a Qiagen MiniElute Gel Purification Kit (Qiagen, Valencia, CA, USA) following the manufacturer's protocol. Gel purified libraries were quantified using a High Sensitivity Quant-it dsDNA assay kit (ThermoFisher),

diluted to 4 nM, and pooled in equimolar amounts across loci. Libraries were sequenced on an Illumina MiSeq (paired-end reads; 2× 250 bp cycles with two 8 bp index reads; 15% PhiX). Machine-processed sequencing output has been deposited under National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under BioProject PRJNA397927, together with the NCBI Biosample Object Accession numbers SAMN07495800-SAMN07495824 for each shrew GI tract. Associated study metadata is available through the U.S. Geological data release, https://doi.org/10.5066/P9YH4C20.

2.6. Bioinformatics analysis

Paired-end reads were processed and analysed using the mothur software package (v 1.37.3) by following the MiSeq SOP analysis pipeline (Kozich et al., 2013). Each genetic locus was processed and analysed separately. Paired reads were assembled into contigs. Contigs with ambiguous bases and those outside a target range of lengths (for each locus) were removed (Table 4). Duplicate contigs were merged. Contigs were aligned to their respective reference databases based on sequences (16S and 12S mtDNA, and 28S rDNA) generated from morphologically identified helminths by authors S.E. Greiman and V.V. Tkach (personal reference databases). Contigs were filtered to remove overhangs at each end and poorly aligned reads. Duplicate contigs were again merged. Contigs were further clustered allowing two nucleotide differences between contigs to reduce noise. Chimeric sequences were identified and removed using the VSEARCH algorithm through mothur. Assembled reads were classified using the Bayesian classifier utilizing the appropriate reference and taxonomy file for each locus. Following classification, contigs were clustered into operational taxonomic units (OTUs) using the cluster.split command, which uses taxonomic information to split the reads into bins and then clusters within each bin.

Errors are inherently associated with amplification, sequencing and mapping of reads to OTUs, making it difficult to distinguish between amplification artifacts and low abundance OTUs. We therefore applied a filter to remove false positive detections by subtracting the number of reads detected in the negative controls for a given OTU from the number of reads assigned to that OTU

Table 4
Sequence length selection criteria used in the screen.seqs command of the bioinformatics programme Mothur for next generation amplicon sequence processing for each genetic locus (16S mitochondrial, 12S mitochondrial, and 28S ribosomal RNA) targeted in this study.

Gene	Minimum length	Maximum length
16S (helminth)	106	118
28S (helminth)	280	290
12S (helminth)	118	130

within each sample. Additionally, OTUs with fewer than 15 reads within a sample were excluded from the analysis; 15 reads represent the application of the following minimum threshold ranges (0.007–6.8% (16S mtDNA)), (0.07–20.8% (28S rDNA)), (0.006–16.5% (12S mtDNA)).

OTU abundances for each locus and sample were visualised on a stacked bar plot using the data visualisation package, ggPLOTs, in the statistical programming language R. Shannon diversity (alpha diversity statistic) indices were calculated separately for each focal shrew species (*Sorex cinereus* and *Sorex monticola*), fixation method (70% EtOH, 95% EtOH, LN), and extraction source (whole gut, intestinal scrape), and for each locus. A Kruskal–Wallis rank sum

test was used to look for significant differences amongst the Shannon diversity values and amongst taxon abundances for each genetic locus and metadata category. *P* values less than 0.05 were considered significant.

3. Results

3.1. Extraction quantities

Variation in DNA quantities was observed for the different sample fixations, with freshly collected GIs (collected in 2015
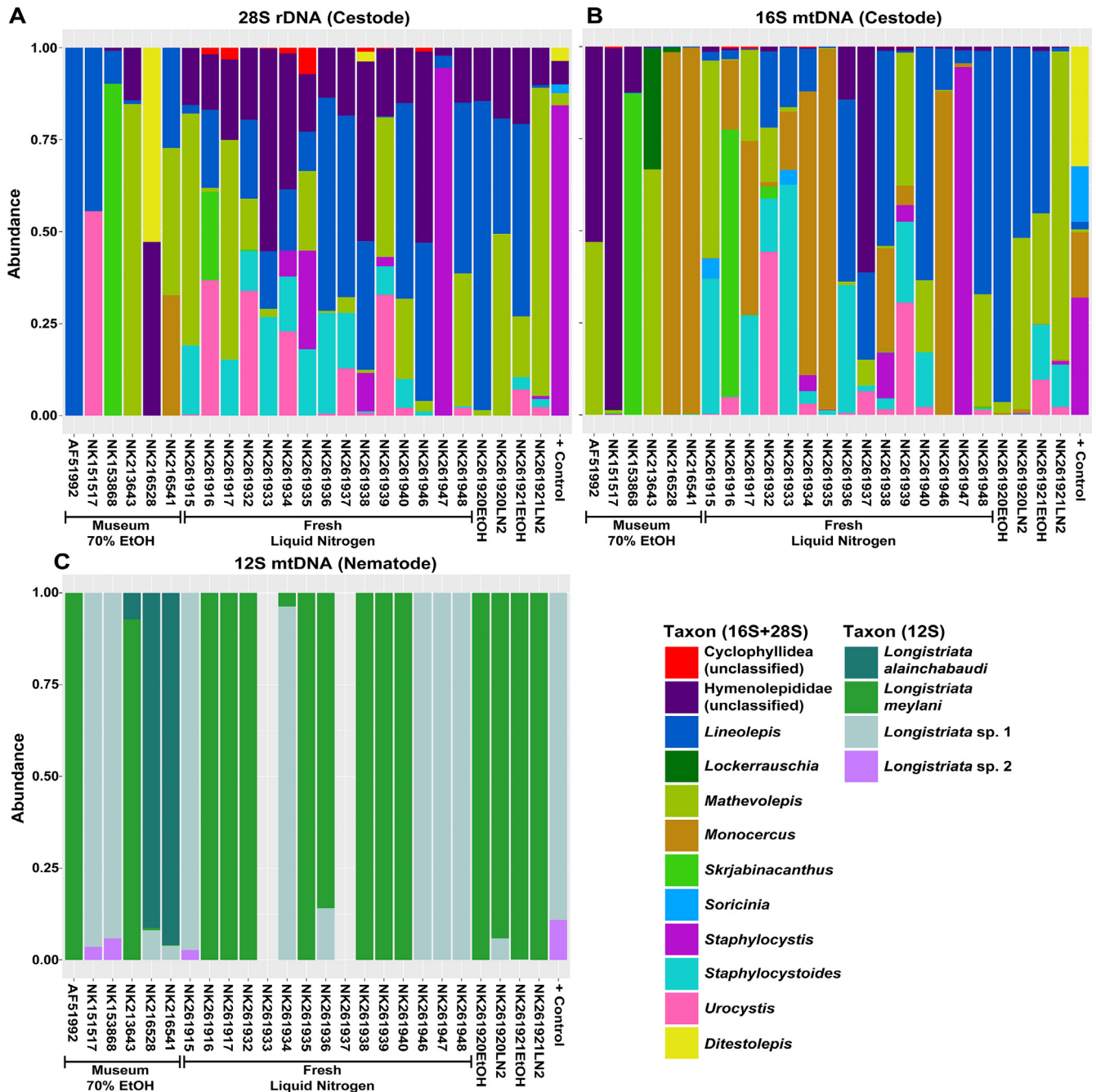
Fig. 1. Stacked bar plots showing the relative abundance of each operational taxonomic unit for each shrew gastrointestinal tract across the three genetic loci, 28S ribosomal RNA, 16S mitochondrial DNA, and 12S mitochondrial DNA. (A) Cestode genera abundances within 23 shrew gastrointestinal tracts based on 28S rDNA Illumina paired-end reads. (B) Cestode genera abundances within 23 shrew gastrointestinal tracts based on 16S mtDNA Illumina paired-end reads. (C) Nematode taxa abundances within 23 shrew gastrointestinal tracts based on 12S mtDNA Illumina paired-end reads.

and stored in LN) having higher concentrations of DNA after extraction (Table 1).

3.2. Illumina sequencing

Over 8.7 million paired reads were obtained for all genetic loci across the 23 whole GI tract samples and the positive and negative controls. Of the >8.7 million sequences, those passing filtering/quality criteria (Supplementary Tables S1–S9) included over 270,000 for 28S rDNA, over 3.5 million for 16S mtDNA and over 3.2 million for 12S mtDNA.

3.3. OTU identification

3.3.1. Positive controls

All six OTUs used for the positive controls were identified for 16S mtDNA, while only five OTUs were identified for 28S rDNA, with only *Monocercus* not found for 28S. Both OTUs used for the 12S *Longistriata* were identified for 12S mtDNA. *Ditestolepis* and *Soricinia* used for the positive controls were collected from Mongolia and represent the first available sequences for these taxa, and thus were not included in the reference database, thus *Ditestolepis* was likely identified as Cyclophyllidea unknown and *Soricinia* was likely identified as Hymenolepididae unknown.

3.3.2. 28S rDNA (cestodes)

Twenty-three unique OTUs were identified for the 28S rDNA gene; however, after applying filtering thresholds, 18 unique OTUs remained. OTUs were further grouped at the genus level (Fig. 1A, Supplementary Tables S1–S3). There were differences in the identified OTUs for both samples (MSB shrew accession numbers:

NK261920 and NK261921) where the GI tracts were split between 95% EtOH (stored at room temperature) and LN. In addition, *Urocystis* and *Staphylocystoides* were not detected from the EtOH preserved (large intestine) half of the GI tract for either sample, whereas both parasite genera were detected from the frozen half.

3.3.3. 16S mtDNA (cestodes)

Twenty-one unique OTUs were identified for the 16S mtDNA gene; however, after applying filtering thresholds, 18 unique OTUs remained. OTUs were further grouped at the genus level (Fig. 1B, Supplementary Tables S4–S6). There were differences in the identified OTUs for one (MSB NK261920) of the two samples where the GI tracts were split between 95% EtOH (stored at room temp) and LN, with the EtOH fixed GI missing *Urocystis* compared with the LN fixed GI. There was no difference in the identified OTUs for the other split GI tract (MSB NK261921).

3.3.4. 12S mtDNA (nematodes)

Five unique OTUs were identified for the 16S mtDNA gene, however, after applying thresholds, four unique OTUs remained (Fig. 1C, Supplementary Tables S7–S9).

3.4. Alpha diversity

3.4.1. 28S rDNA (cestodes)

A significant difference in alpha diversity was detected amongst the different sample fixation types (Kruskal–Wallis, $P = 0.016$; Fig. 2A), but not between shrew species (Kruskal–Wallis, $P = 0.683$), or between extraction methods (Kruskal–Wallis, $P = 0.082$) based on 28S rDNA data. Slight variations in OTUs present amongst the different fixation types were observed, with
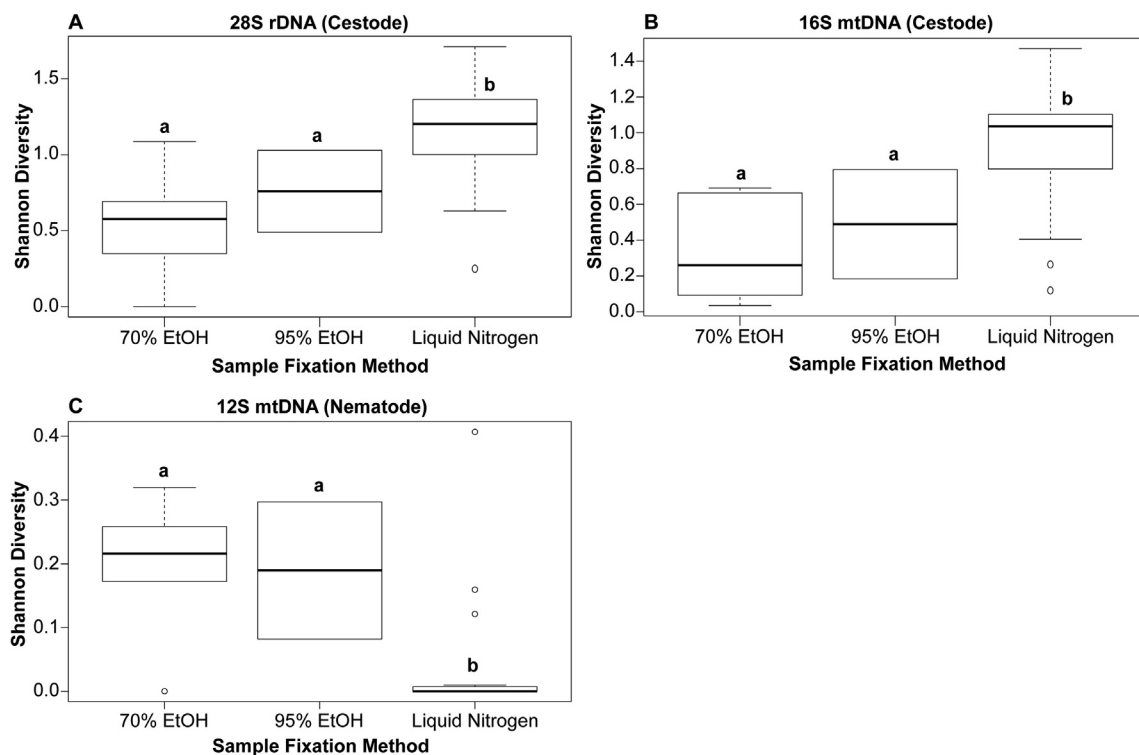
Fig. 2. Box plots comparing the Shannon Diversity index (alpha diversity) values for (A) cestode genera infecting 23 shrew gastrointestinal tracts based on 28S rDNA Illumina paired-end reads between the different sample fixation methods; (B) cestode genera infecting 23 shrew gastrointestinal tracts based on 16S mtDNA Illumina paired-end reads between the different sample fixation methods; (C) nematode taxa infecting 23 shrew gastrointestinal tracts based on 12S mtDNA Illumina paired-end reads between the different sample fixation methods. Significant differences ($P < 0.05$) between either fixation method or shrew species computed with a Kruskal–Wallis rank sum test are indicated with letters above the box plots. Outliers are indicated by empty circles above the maximum whisker or below the minimum whisker, the upper quartile is indicated by the top portion of the box, above the median value line, and the lower quartile is indicated by the lower portion of the box, below the median value line.

Staphylocystoides and *Urocystis* being absent from EtOH preserved samples. Based on a Kruskal–Wallis test among individual cestode taxa for sample fixation type, a higher abundance of tapeworms in the genus *Staphylocystoides* was detected in GI tracts preserved in LN ($P = 0.015$) compared to 70% or 95% EtOH preserved samples, although not significant ($P = 0.08$) (Supplementary Fig. S1).

3.4.2. 16S mtDNA (cestodes)

A significant difference was detected in alpha diversity among the different sample fixation types (Kruskal–Wallis, $P = 0.009$; Fig. 2B) but not between shrew species (Kruskal–Wallis, $P = 0.180$) or between extraction methods (Kruskal–Wallis, $P = 0.446$) based on 16S mtDNA data. Slight variations in OTUs amongst the different fixation types were observed, with *Urocystis* and *Soricinia* (although only two samples were infected with *Soricinia*) being absent from EtOH preserved samples and *Lockerrauschia* being absent from LN preserved samples. Based on a Kruskal–Wallis test amongst individual cestode taxa for sample fixation type, there was a higher abundance of tapeworms of the genus *Staphylocystoides* ($P = 0.04$) in LN samples compared with their abundances in either 70% or 95% EtOH preserved samples (Supplementary Fig. S2), and a higher abundance of tapeworms in the genus *Lineolepis* ($P = 0.007$) in 95% EtOH and LN preserved samples compared with their abundances in 70% EtOH preserved samples (Supplementary Fig. S3).

3.4.3. 12S mtDNA (nematodes)

There was a significant difference in alpha diversity amongst the different sample fixation types (Kruskal–Wallis, $P = 0.024$) (Fig. 2C), but not between shrew species (Kruskal–Wallis, $P = 0.73$), or between extraction method (Kruskal–Wallis, $P = 0.97$) for samples based on 12S mtDNA. Based on a Kruskal–Wallis test amongst individual nematode taxa for sample fixation type, there was a higher abundance of the nematode *Longistriata alainchabaudi* ($P = 0.029$) in 70% EtOH preserved samples compared with their abundances in either 95% EtOH or LN preserved samples (Supplementary Fig. S4).

4. Discussion

Emerging pathogen and other biological research relies on correct species determination, which requires taxonomic expertise; however, despite the importance of taxonomy, this discipline is in rapid decline (Drew, 2011; Brooks et al., 2014; Lees and Pimm, 2015). Fourteen years ago, Hebert et al. (2003) suggested that the best option for supporting a capability for authoritative and sustainable taxonomic identification was a system relying on DNA barcode sequences. Since then, large amounts of DNA barcode data have been generated and made available in public databases. However, the success of a DNA barcode-based approach relies initially on a broad taxonomic panel of authoritatively determined organisms that subsequently serve as the source for appropriate diagnostic DNA markers among a diverse array of often closely related taxa (Joly et al., 2014). The importance of authoritative knowledge of taxonomic experts is highlighted by the fact that no more than 10% of global pathogens have been documented (Brooks and Hoberg, 2013), and among these documented taxa, very little is known regarding their evolution and ecology (Hoberg et al., 2015). Approximately 75% of named parasitic flatworms (cestodes, trematodes) are known only from their original descriptions, and therefore little can be extrapolated regarding their host diversity, geographic range, or disease dynamics (Poulin and Morand, 2004). The availability of trustworthy helminth barcode sequences is limited, although the application of definitive sequence data for identification of parasites in geographically

extensive and site intensive sampling has been clearly demonstrated (e.g., Kutz et al., 2007; Brooks et al., 2014).

NGS methodologies have not been widely applied in studies of helminth communities, despite their obvious considerable potential in addressing a variety of questions. Tanaka et al. (2014) utilised primers amplifying the variable V9 region of the eukaryotic 18S rDNA gene, which has several drawbacks. First, it amplifies across a wide range of eukaryotic organisms, including mammals, thereby requiring mammal blocking primers. Because we used whole shrew GI tracts to keep helminth communities intact, the amount of host DNA was much greater than that of helminth DNA. Second, we attempted to design new 18S primers for helminth taxa only; however, we could not find a sufficiently conserved region for identification of closely related species. Third, in the Tanaka et al. (2014) study, the rats were infected with a fairly low diversity of helminth parasites, including only a single cestode species, and therefore they did not have to differentiate among multiple cestode species within a single host individual, a clear difference from our shrew system.

We have identified and generated our own DNA barcode libraries for 16S mtDNA, 28S rDNA, and 12S mtDNA, for a majority of helminth species currently known from shrews of the genus *Sorex* in North America. We combined multiple DNA markers, including both nuclear and mitochondrial rDNA, to more accurately identify helminth communities. For example, the 28S rDNA locus, although the most abundant in GenBank for cestodes, is highly conserved, so finding a barcode region variable enough to identify closely related species is challenging. 28S rDNA sequences can be identical in congeneric shrew cestode species that are otherwise easily distinguishable by morphology and mtDNA (Tkach et al., 2013). Conversely, mitochondrial barcoding genes alone may be problematic due to high intraspecific variability.

Taxonomic assignment for helminth sequences is based on specimen identified reference databases developed by S.E. Greiman and V.V. Tkach. This constitutes a reliable basis for accurate subsequent identification using NGS data. However, despite high richness of helminth parasites already known from shrews in North America, a substantial number of species have yet to be formally described. Often, well-fixed samples suitable for proper morphological identification are lacking, resulting in some sequences categorised into either Cyclophyllidea unclassified or Hymenolepididae unclassified. Increased sampling of fresh shrew samples for proper fixation and identification of helminths will help fill in the gaps in the reference databases, improving taxonomic designation of helminths in future metabarcoding studies. Our study emphasises the need for the integration of classical parasitological and molecular techniques with more modern techniques for identifying helminths from a large sample of hosts (Hoberg et al., 2015). At the same time, host identification should be similarly confirmed through integration of molecular and morphological approaches (Dunnum et al., 2017).

Additionally, the 28S rDNA and 16S mtDNA reference databases are not 100% matching, resulting in slight differences in OTU designation. This variation arose due to past sequencing efforts that focused solely on 28S rDNA. Our data demonstrated significant differences in estimates of taxon read abundance based on sample fixation for 16S mtDNA, but not 28S rDNA. This was evident in observations of higher read abundance of *Lineolepis* tapeworms from specimens preserved in 95% EtOH and LN compared with their abundances in samples fixed in 70% EtOH. We hypothesize that this result reflects differences in the reference databases, with several species of *Lineolepis* still missing 16S mtDNA sequences, but not 28S rDNA. Future efforts should aim to increase taxon availability in both reference databases. Existing metabarcoding sequence datasets can then be retroactively

enhanced and re-analysed as more reference sequences from identified taxa become available.

NGS sequencing allowed for accurate designation of helminths to genera, and in many cases to species. However, due to gaps in our sequence reference databases, and a significant number of undescribed species of cestodes from shrews in North America, we limited the reported taxonomic designations to the level of genus. With the exception that *Monocercus* (see below for explanation) was not identified using our 28S primers, positive controls were accurately identified using our sequencing method with all six cestode taxa identified using the 16S primers, five of six being correctly identified for 28S, and both nematode taxa being correctly identified for 12S. Twenty-two unique helminth OTUs were identified, including four nematodes and 18 cestodes. The 18 cestode OTUs were further reduced to nine genera and two unclassified (Cyclophyllidea and Hymenolepididae) (Fig. 1). Based on 28S rDNA, shrews were infected with an average of 3.4 genera of tapeworms (not including unclassified Hymenolepididae or Cyclophyllidea), with one shrew infected with seven different (identified) genera of tapeworms. Based on 16S, shrews were infected with 4.8 genera of tapeworms (not including unclassified Hymenolepididae or Cyclophyllidea), with one shrew infected with as many as eight different (identified) genera of tapeworms.

In regard to the slight variation seen in OTU identification for the two samples where the GI tract was split between LN (small intestine) and 95% EtOH (stored at room temperature) (large intestine) (i.e. *Staphylocystoides* and *Urocystis* missing from the EtOH preserved GI tract for 28S and *Urocystis* missing for one of the samples from the EtOH preserved GI for 16S), we hypothesize that this variation is not caused by the different fixative, but instead due to the portion of the GI tract used for 95% EtOH (i.e., large intestine). Both *Staphylocystoides* and *Urocystis* are minute cestodes that are found in the upper part of the small intestine, so most of the associated cestode tissue would be located in the portion of the small intestine fixed in LN, and therefore be preferentially amplified in that part of the GI tract compared with the large intestine.

Although we examined the differences in helminth alpha diversity among shrew species, this study was designed primarily to test the feasibility of utilizing NGS approaches to identify the helminth community within whole GI tracts of varying age and fixation type. Therefore, differences in alpha diversity should be taken with some caution. The sample sizes and geographic coverage used here for methods development constitute only a small subset of fluid preserved or frozen samples available in museum collections. *Sorex cinereus* and *S. monticola* are among the most widespread and abundant shrew species in North America, and are largely sympatric throughout their ranges, making them an ideal system for examining variation in host, helminth and microbiome communities over time (e.g., long term, seasonal) and space. The approach can be used, however, to study helminth communities and their temporal and spatial dynamics in a broad diversity of small mammals, other vertebrate taxa and a variety of ecosystems across terrestrial, freshwater and marine environments.

Slight variation in the perceived parasite fauna was observed between the 28S rDNA and 16S mtDNA sequencing (Fig. 1A and B). One contributing factor to this variation is the ability of the 16S primers to better amplify tapeworms in the genus *Monocercus* than the 28S primers. This was an expected result, as our 28S rDNA primers were designed almost exclusively to amplify tapeworms in the family Hymenolepididae, whereas *Monocercus* belongs to the family Dilepididae. All genera of cestodes infecting *Sorex* in North America, except for *Monocercus*, belong to the family Hymenolepididae (Kinsella and Tkach, 2009). Again, unequal availability among reference databases for 16S mtDNA and 28S rDNA may also contribute to differential detection, as evidenced by a substantial number of unclassified Hymenolepididae sequences in both our

28S rDNA and 16S mtDNA sequence data, with variation between the two. Future work will include development of primers to target an approximately 480 bp fragment of cytochrome c oxidase subunit I (COI) for improved identification of helminth taxa (i.e., cestodes, digeneans, nematodes) to species in combination with other loci. In addition, we anticipate development of primers to target the invertebrate diet of the shrews, coupled with surveys of the invertebrate community at a given collection site. This additional biodiversity will provide insight into complex parasite life cycles through identification of intermediate hosts, as well as seasonal changes in shrew diet.

Significant differences (albeit small) in alpha diversity were observed for helminth taxa for both 28S and 16S between locality (northern and southern North America) and fixation type. The 28S EtOH preserved samples were found to not be infected with *Staphylocystoides* or *Staphylocystis* (only one EtOH specimen was infected with *Staphylocystis* based on 16S). Based on this, it may be possible that the 28S primers amplify *Staphylocystoides* DNA less efficiently from more degraded samples. At the same time, the 16S primers more easily pick up this degraded DNA due to the shorter fragment size of the target region and greater number of remaining copies of the mitochondrial DNA. In addition, the 16S primers did not amplify *Urocystis* (only one shrew was found to be infected based on 28S) or *Soricinia* (only two shrews were found to be infected overall with *Soricinia*) DNA in EtOH preserved samples, but did identify *Lockerrauschia* DNA in only EtOH preserved samples. Given that slight variation in OTUs is not related solely to one type of fixative (i.e. *Lockerrauschia* only found in EtOH preserved samples), it is likely that variation is a result of locality and not preservation.

In this study we attempted to pool all target loci for sequencing on a single MiSeq run. Although we obtained usable data for each locus, we found that depending on the size (in base pairs) of a given target region, we would obtain different numbers of sequence reads. Smaller targets such as 16S mtDNA ($\approx$110 bp) and 12S mtDNA ($\approx$120 bp) produced greater numbers of sequences than larger targets, 28S rDNA ($\approx$280 bp) and 16S rDNA ($\approx$300 bp), with 3.5 million and 3.2 million sequences compared with 270,000 and 545,000 sequences, respectively. Therefore, we recommend sequencing each locus in a separate run, sequencing loci of similar sizes together (e.g. 28S rDNA with 16S rDNA and 12S mtDNA with 16S mtDNA), or including higher representation of longer fragments (relative to the shorter fragments) in the pooled library. Costs of sequencing will remain similar with these approaches, as we were not limited by the sequencing output of the MiSeq, and therefore, sequencing costs are largely defined by the total number of unique barcodes available. Through reduction of the number of loci per sequencing run, the number of samples can be increased to 384 samples for a single locus run.

Although our main goal was to develop and test a method to identify helminth parasites from museum archived samples, we believe a discussion of general sampling techniques to maximize long-term specimen use is warranted. Historically, field collection has focused on individual taxonomic groups, with little emphasis on collecting samples for other syntopic species or for disciplines other than systematics. With technological advancements in museum storage, DNA sequencing techniques, and field collection protocols, samples should be collected in ways that maximize their use by diverse specialists in the future (mammalogists, parasitologists, bacteriologists, stable isotope researchers, entomologists, etc.). Here we will briefly focus on sampling protocols from shrews as an example of how the utility of field collected samples can be maximised for future parasitological and microbial research.

Helminth parasites of shrews quickly degrade upon death of the host, making them difficult to study. Degradation can be limited, however, by using live traps (pitfalls) that are regularly checked

every few hours. Upon death, a quick method for dissection and preservation of the parasites and microbes for targeted sequencing is needed. One such method is to remove the intestinal tract with clean forceps, place it directly into a 2 mL sterile cryovial, and preserve it in LN. These frozen samples can be effectively used for both helminth and microbial metabarcoding identification. Further, a new 'gold standard' for biodiversity discovery involving metagenomics would also integrate a subset of parasite specimens from each host and locality properly fixed for morphological examination as vouchers to be distributed or held in recognised museum repositories. Thus, the process of biodiversity discovery would serve to build increasingly fine scale pictures of helminth communities and faunal structure over space and time.

Our study outlines a novel multi-locus metabarcoding approach for efficiently identifying the helminth symbiont communities within museum archived and freshly collected whole GI tracts of vertebrates. The protocol successfully amplified helminth DNA from variously fixed samples (4 months–16 years old) in museum archives. Metabarcoding approaches, as outlined here, show the potential to increase the use and impact of the vast number of samples held in natural history museums. As we shift from response to anticipation of emerging disease, such capacity is critical in public health research, surveillance, and in building accurate pictures of distribution for parasites across the biosphere. In particular, it allows for development of temporal baselines necessary for reliable comparisons across past and present biotic communities. Such comparisons can provide increased resolution in ecological modelling and other capacities to anticipate and identify accelerating changes in distributions and interactions among assemblages of both hosts and pathogens (e.g., Hope et al., 2013, 2016; Brooks et al., 2014; Cook et al., 2016).

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijpara.2018.08.001.

References

Allen, J.M., Boyd, B., Nguyen, N.-P., Vachaspati, P., Warnow, T., Huang, D.I., Grady, P. G.S., Bell, K.C., Cronk, Q.C.B., Mugisha, L., Pittendrigh, B.R., Leonardi, M.S., Reed, D.L., Johnson, K.P., 2017. Phylogenomics from whole genome sequences using aTRAM. Syst. Biol. 66, 786–798.

Besnard, G., Bertrand, J.A.M., Delahaie, B., Bourgeois, Y.X.C., Lhuillier, E., Thébaud, C., 2016. Valuing museum specimens: high-throughput DNA sequencing on historical collections of New Guinea crowned pigeons (Goura). Biol. J. Linn. Soc. Lond. 117, 71–82.

Bi, K., Linderoth, T., Vanderpool, D., Good, J.M., Nielsen, R., Moritz, C., 2013. Unlocking the vault: next-generation museum population genomics. Mol. Ecol. 22, 6018–6032.

Bourne, D.G., Dennis, P.G., Uthicke, S., Soo, R.M., Tyson, G.W., Webster, N., 2013. Coral reef invertebrate microbiomes correlate with the presence of photosymbionts. ISME J. 7, 1452–1458.

Brooks, D.R., Hoberg, E.P., 2013. The emerging infectious disease crisis and pathogen pollution: a question of ecology and evolution. In: Rohde, K. (Ed.), The Balance of Nature and Human Impact. Cambridge University Press, Cambridge UK, pp. 215–229.

Brooks, D.R., Hoberg, E.P., Boeger, W.A., Gardner, S.L., Galbreath, K.E., Herczeg, D., Mejía-Madrid, H.H., Rácz, S.E., Dursahinhan, A.T., 2014. Finding them before they find us: informatics, parasites, and environments in accelerating climate change. Comp. Parasitol. 81, 155–164.

Carrillo-Araujo, M., Taş, N., Alcántara-Hernández, R.J., Gaona, O., Schondube, J.E., Medellín, R.A., Jansson, J.K., Falcón, L.I., 2015. Phyllostomid bat microbiome composition is associated with host phylogeny and feeding strategies. Front. Microbiol. 6, 447.

Casiraghi, M., Bain, O., Guerrero, R., Martin, C., Pocacqua, V., Gardner, S.L., Franceschi, A., Bandi, C., 2004. Mapping the presence of Wolbachia pipientis on the phylogeny of filarial nematodes: evidence for symbiont loss during evolution. Int. J. Parasitol. 34, 191–203.

Cock, P.J.A., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., de Hoon, M.J.L., 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 22 (11), 1422–1423. https://doi.org/10.1093/bioinformatics/btp163.

Cook, J.A., Galbreath, K.E., Bell, K.C., Campbell, M.L., Carrière, S., Colella, J.P., Dawson, N.G., Dunnum, J.L., Eckerlin, R.P., Fedorov, V., Greiman, S.E., Haas, G.M.S., Haukisalmi, V., Henttonen, H., Hope, A.G., Jackson, D., Jung, T.S., Koehler, A.V., Kinsella, J.M., Krejsa, D., Kutz, S.J., Liphardt, S., MacDonald, S.O., Malaney, J.L., Makarikov, A., Martin, J., McLean, B.S., Mulders, R., Nyamsuren, B., Talbot, S.L., Tkach, V.V., Tsvetkova, A., Toman, H.M., Waltari, E.C., Whitman, J.S., Hoberg, E.P., 2016. The Beringian Coevolution Project: holistic collections of mammals and associated parasites reveal novel perspectives on evolutionary and environmental change in the North. Arct. Sci. 3, 1–33.

Drew, L.W., 2011. Are we losing the science of taxonomy? As need grows, numbers and training are failing to keep up. Bioscience 61, 942–946.

Dunnum, J.L., Yanagihara, R., Johnson, K.M., Armien, B., Batsaikhan, N., Morgan, L., Cook, J.A., 2017. Biospecimen repositories and integrated databases as critical infrastructure for pathogen discovery and pathobiology research. PLoS Negl. Trop. Dis. 11, e0005133.

Guschanski, K., Krause, J., Sawyer, S., Valente, L.M., Bailey, S., Finstermeier, K., Sabin, R., Gilissen, E., Sonet, G., Nagy, Z.T., Lenglet, G., Mayer, F., Savolainen, V., 2013. Next-generation museomics disentangles one of the largest primate radiations. Syst. Biol. 62, 539–554.

Haponski, A.E., Stepien, C.A., 2016. Two decades of genetic consistency in a reproductive population in the face of exploitation: patterns of adult and larval walleye (Sander vitreus) from Lake Erie's Maumee River. Conserv. Genet. 17, 1345–1362.

Hebert, P.D.N., Ratnasingham, S., de Waard, J.R., 2003. Barcoding animal life: cytochrome oxidase subunit 1 divergences among closely related species. Proc. Biol. Sci. 270, S96–S99.

Hoberg, E.P., Agosta, S.J., Boeger, W.A., Brooks, D.R., 2015. An integrated parasitology: revealing the elephant through tradition and invention. Trends Parasitol. 31, 128–133.

Hoberg, E.P., Kutz, S.J., Cook, J.A., Galaktionov, K., Haukisalmi, V., Henttonen, H., Laaksonen, S., Markarikov, A., Marcogliese, D.J., 2013. Parasites in terrestrial, freshwater and marine systems. In: Meltofte, H. (Ed.), Arctic Biodiversity Assessment- Status and Trends in Arctic Biodiversity. Conservation of Arctic Flora and Fauna, Arctic Council, Akureyi, Iceland, pp. 476–505.

Hope, A.G., Waltari, E., Payer, D.C., Cook, J.A., Talbot, S.L., 2013. Future distribution of tundra refugia in northern Alaska. Nat. Clim. Change 3, 931.

Hope, A.G., Greiman, S.E., Tkach, V.V., Hoberg, E.P., Cook, J.A., 2016. Shrews and their parasites: small species indicate big changes. Arctic Report Card, NOAA. https://www.arctic.noaa.gov/Report-Card/Report-Card-2016/ArtMID/5022/ArticleID/268/Shrews-and-Their-Parasites-Small-Species-Indicate-Big-Changes.

Hudson, P.J., Dobson, A.P., Lafferty, K.D., 2006. Is a healthy ecosystem one that is rich in parasites? Trends Ecol. Evol. 21, 381–385.

Jiménez, R.R., Sommer, S., 2017. The amphibian microbiome: natural range of variation, pathogenic dysbiosis, and role in conservation. Biodivers. Conserv. 26, 763–786.

Joly, S., Davies, T.J., Archambault, A., Bruneau, A., Derry, A., Kembel, S.W., Peres-Neto, P., Vamosi, J., Wheeler, T.A., 2014. Ecology in the age of DNA barcoding: the resource, the promise and the challenges ahead. Mol. Ecol. Resour. 14, 221–232.

Kinsella, J.M., Tkach, V.V., 2009. Checklist of helminth parasites of Soricomorpha (= Insectivora) of North America north of Mexico. Zootaxa 1969, 36–58.

Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., Schloss, P.D., 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.

Kutz, S.J., Asmundsson, I., Hoberg, E.P., Appleyard, G.D., Jenkins, E.J., Beckmen, K., Branigan, M., Butler, L., Chilton, N.B., Cooley, D., Elkin, B., Huby-Chilton, F., Johnson, D., Kuchboev, A., Nagy, J., Oakley, M., Polley, L., Popko, R., Scheer, A., Simard, M., Veitch, A., 2007. Serendipitous discovery of a novel protostrongylid (Nematoda: Metastrongyloidea) in caribou, muskoxen, and moose from high latitudes of North America based on DNA sequence comparisons. Can. J. Zool. 85, 1143–1156.

Lees, A.C., Pimm, S.L., 2015. Species, extinct before we know them? Curr. Biol. 25, R177–R180.

Littlewood, D., Waeschenbach, A., Nikolov, P., 2008. In search of mitochondrial markers for resolving the phylogeny of cyclophyllidean tapeworms (Platyhelminthes, Cestoda) — a test study with Davaineidae. Acta Parasitol., 133

McCormack, J.E., Maley, J.M., Hird, S.M., Derryberry, E.P., Graves, G.R., Brumfield, R. T., 2012. Next-generation sequencing reveals phylogeographic structure and a species tree for recent bird divergences. Mol. Phylogenet. Evol. 62, 397–406.

McLean, B.S., Bell, K.C., Dunnum, J.L., Abrahamson, B., Colella, J.P., Deardorff, E.R., Weber, J.A., Jones, A.K., Salazar-Miralles, F., Cook, J.A., 2016. Natural history collections-based research: progress, promise, and best practices. J. Mammal. 97, 287–297.

Melville, J., Haines, M.L., Boysen, K., Hodkinson, L., Kilian, A., Smith Date, K.L., Potvin, D.A., Parris, K.M., 2017. Identifying hybridization and admixture using SNPs: application of the DArTseq platform in phylogeographic research on vertebrates. R. Soc. Open Sci. 4, 161061.

Phillips, C.D., Phelan, G., Dowd, S.E., McDonough, M.M., Ferguson, A.W., Delton Hanson, J., Siles, L., OrdÓÑEz-Garza, N., San Francisco, M., Baker, R.J., 2012. Microbiome analysis among bats describes influences of host phylogeny, life history, physiology and geography. Mol. Ecol. 21, 2617–2627.

Poulin, R., Morand, S., 2004. Parasite Biodiversity. Smithsonian Institution Press, Washington, District of Columbia USA.

Rowe, K.C., Singhal, S., Macmanes, M.D., Ayroles, J.F., Morelli, T.L., Rubidge, E.M., Bi, K.E., Moritz, C.C., 2011. Museum genomics: low-cost and high-accuracy genetic data from historical specimens. Mol. Ecol. Resour. 11, 1082–1092.

Shaffer, H.B., Fisher, R.N., Davidson, C., 1998. The role of natural history collections in documenting species declines. Trends Ecol. Evol. 13, 27–30.

Stat, M., Huggett, M.J., Bernasconi, R., DiBattista, J.D., Berry, T.E., Newman, S.J., Harvey, E.S., Bunce, M., 2017. Ecosystem biomonitoring with eDNA: metabarcoding across the tree of life in a tropical marine environment. Sci. Rep. 7, 12240.

Suarez, A.V., Tsutsui, N.D., 2004. The value of museum collections for research and society. Bioscience 54, 66–74.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729.

Tanaka, R., Hino, A., Tsai, I.J., Palomares-Rius, J.E., Yoshida, A., Ogura, Y., Hayashi, T., Maruyama, H., Kikuchi, T., 2014. assessment of helminth biodiversity in wild rats using 18S rDNA based metagenomics. PLoS ONE 9, e110769.

Tkach, V.V., Makarikov, A.A., Kinsella, J.M., 2013. Morphological and molecular differentiation of *Staphylocystis clydesengeri* n. sp. (Cestoda, Hymenolepididae) from the vagrant shrew, *Sorex vagrans* (Soricimorpha, Soricidae) in North America. Zootaxa 3691, 389–400.

Van Rossum, G., 1995. Python Tutorial, Technical Report CS-R9526. Centrum voor Wiskunde en Informatica (CWI), Amsterdam.

Yuan, H., Jiang, J., Jiménez, F.A., Hoberg, E.P., Cook, J.A., Galbreath, K.E., Li, C., 2016. Target gene enrichment in the cyclophyllidean cestodes, the most diverse group of tapeworms. Mol. Ecol. Resour. 16, 1095–1106.