RESOURCE ARTICLE

WILEY | MOLECULAR ECOLOGY RESOURCES

An improved method for utilizing high-throughput amplicon sequencing to determine the diets of insectivorous animals

Michelle A. Jusino[1,2]* | Mark T. Banik[1]* | Jonathan M. Palmer[1] | Amy K. Wray[3] | Lei Xiao[4] | Emma Pelton[5,6] | Jesse R. Barber[7] | Akito Y. Kawahara[4] | Claudio Gratton[5] | M. Zachariah Peery[3] | Daniel L. Lindner[1]

[1]United States Department of Agriculture, Forest Service, Northern Research Station, Center for Forest Mycology Research, Madison, Wisconsin

[2]Department of Plant Pathology, University of Florida, Gainesville, Florida

[3]Department of Forest and Wildlife Ecology, University of Wisconsin-Madison, Madison, Wisconsin

[4]McGuire Center for Lepidoptera and Biodiversity, Florida Museum of Natural History, University of Florida, Gainesville, Florida

[5]Department of Entomology, University of Wisconsin-Madison, Madison, Wisconsin

[6]The Xerces Society for Invertebrate Conservation, Portland, Oregon

[7]Department of Biological Sciences, Graduate Program in Ecology, Evolution and Behavior, Boise State University, Boise, Idaho

Correspondence
Michelle A. Jusino, Department of Plant Pathology, University of Florida, Gainesville, FL.
Emails: michellejusino@gmail.com; mjusino@ufl.edu
and
Daniel L. Lindner, United States Department of Agriculture, Forest Service, Northern Research Station, Center for Forest Mycology Research, One Gifford Pinchot Drive, Madison, WI.
Emails: dlindner@fs.fed.us; dlindner@wisc.edu

Abstract

DNA analysis of predator faeces using high-throughput amplicon sequencing (HTS) enhances our understanding of predator–prey interactions. However, conclusions drawn from this technique are constrained by biases that occur in multiple steps of the HTS workflow. To better characterize insectivorous animal diets, we used DNA from a diverse set of arthropods to assess PCR biases of commonly used and novel primer pairs for the mitochondrial gene, cytochrome oxidase C subunit 1 (COI). We compared diversity recovered from HTS of bat guano samples using a commonly used primer pair "ZBJ" to results using the novel primer pair "ANML." To parameterize our bioinformatics pipeline, we created an arthropod mock community consisting of single-copy (cloned) COI sequences. To examine biases associated with both PCR and HTS, mock community members were combined in equimolar amounts both pre- and post-PCR. We validated our system using guano from bats fed known diets and using composite samples of morphologically identified insects collected in pitfall traps. In PCR tests, the ANML primer pair amplified 58 of 59 arthropod taxa (98%), whereas ZBJ amplified 24–40 of 59 taxa (41%–68%). Furthermore, in an HTS comparison of field-collected samples, the ANML primers detected nearly fourfold more arthropod taxa than the ZBJ primers. The additional arthropods detected include medically and economically relevant insect groups such as mosquitoes. Results revealed biases at both the PCR and sequencing levels, demonstrating the pitfalls associated with using HTS read numbers as proxies for abundance. The use of an arthropod mock community allowed for improved bioinformatics pipeline parameterization.

KEYWORDS
AMPtk, arthropod mock community, bat guano, dietary analysis, insectivore, next-generation sequencing

---

1 | INTRODUCTION

High-throughput amplicon sequencing (HTS) has become the pre-ferred method for rapid molecular identification of members of mixed ecological communities. HTS is now also increasingly used to identify the arthropod dietary components of a wide taxonomic range of animals including mammals (Bussche et al., 2016; Clare et al., 2014; Clare, Symondson, & Fenton, 2014; Mallott, Malhi, & Garber, 2015; Rydell et al., 2016; Vesterinen et al., 2016), birds (Cri-sol-Martínez, Moreno-Moyano, Wormington, Brown, & Stanley, 2016; Jedlicka, Vo, & Almeida, 2016; Trevelline, Latta, Marshall, Nut-tle, & Porter, 2016), reptiles (Kartzinel & Pringle, 2015), fish (Harms-Tuohy, Schizas, & Appeldoorn, 2016) and arthropods (Krehenwinkel, Kennedy, Pekár, & Gillespie, 2016). Identification of the DNA of dietary components is accomplished by "metabarcoding," which involves extracting DNA from faecal samples, amplifying one or more barcoding loci, preparing DNA libraries and finally sequencing, bioinformatics and data analysis. Each of these steps involves deci-sions and assumptions that significantly affect results. For example, biases are unavoidable when amplifying environmental DNA with PCR-based methods (Brooks et al., 2015) and careful consideration should be exercised when selecting a primer pair for HTS. Thus, while DNA metabarcoding is a powerful tool for studying trophic interactions, conclusions should take into account the shortcomings and parameters of the techniques (e.g., Brooks et al., 2015; D'Amore et al., 2016; Lindahl et al., 2013; Nguyen, Smith, Peay, & Kennedy, 2015; Pompanon et al., 2012).

The mitochondrial cytochrome oxidase C subunit 1 locus (COI) is the most frequently used barcoding locus for identifying a wide range of taxonomic groups, including arthropods. Because COI has the most extensive reference library for arthropods (BOLD systems, Ratnasingham and Hebert, 2007), it is the most commonly used locus for dietary studies of insectivorous animals (Clarke, Soubrier, Weyrich, & Cooper, 2014). The entire COI barcoding region is about 658 base pairs (bp) and currently too long to be used efficiently with most HTS platforms. Therefore, it is necessary to sequence shorter regions of the COI locus, which has proven challenging due to a lack of conserved priming sites within the COI region (Deagle, Jarman, Coissac, Pompanon, & Taberlet, 2014). Therefore, novel primer pairs should be tested against as many expected target DNA sequences as possible.

Zeale, Butlin, Barker, Lees, and Jones (2011) developed the ZBJ-ArtF1c/ZBJ-ArtR2c (hereafter ZBJ) primer pair for detecting arthro-pod prey DNA in bat guano by amplifying a 157-bp fragment of the COI region. In the initial study (Zeale et al., 2011), which employed cloning and sequencing rather than HTS, the ZBJ primers amplified 37 taxa from 13 arthropod orders but did not amplify bat COI DNA. The ZBJ primers were designed to target a short fragment to amplify the presumably degraded DNA present in guano, and coincidentally, the length of the amplicon generated is compatible with many HTS platforms. Subsequently, numerous researchers have employed the ZBJ primers in HTS studies that analyse diets of insectivorous ani-mals, including bats (Bussche et al., 2016; Clare et al., 2014, 2014;

Rydell et al., 2016; Vesterinen et al., 2016) and birds (Crisol-Martínez et al., 2016; Jedlicka et al., 2016; Trevelline et al., 2016). Although the ZBJ primers have been widely utilized, there are indications that they have a narrow taxonomic range (Brandon-Mong et al., 2015; Clarke et al., 2014; Mallott et al., 2015).

Along with primer choice, many other assumptions and parame-ters commonly employed in HTS environmental DNA analyses have a large impact on the operational taxonomic units (OTUs) that are recovered and appropriate positive controls are necessary to deter-mine those impacts. Bioinformatics clustering algorithms can influ-ence apparent diversity within a sample, or an entire library of samples, and trimming and filtering parameters can impact the result-ing community composition (Deagle, Thomas, Shaffer, Trites, & Jar-man, 2013). A validation or control is needed to accurately parameterize bioinformatics pipelines; therefore, the use of mock communities as positive controls in HTS is increasingly becoming common, especially among researchers who work with fungal and bacterial communities (Bokulich & Mills, 2013; Bokulich et al., 2013; Nguyen et al., 2015; Palmer, Jusino, Banik, & Lindner, 2018). Mock communities can be used to examine biases, starting at the sampling step and ending at the bioinformatics and community analysis steps.

Here, we provide a comprehensive assessment of an improved molecular and analytical pipeline for molecular-based arthropod diet determination. We used a reference arthropod community to iden-tify specific amplification biases associated with three commonly used primer pairs, including ZBJ, and two novel primer pairs, LCOI-1490/COI-CFMRa (hereafter ANML) and LCOI-1490/COI-CFMRb (hereafter CFMRb), for the COI region (Table 1). To further test pri-mers, we compared HTS results from the ZBJ primers to our novel ANML primer pair using field-collected bat guano samples. We designed an arthropod mock community based on single-copy (cloned) mitochondrial COI sequences, which can serve as a standard in HTS sequencing and to help parameterize a bioinformatics pipe-line. Finally, we validated the accuracy of our system of novel pri-mers, the mock community control and our bioinformatics pipeline using guano from bats fed known insect diets and composite sam-ples of morphologically identified arthropods from pitfall traps.

2 | METHODS AND MATERIALS

2.1 | Testing of primer pairs against known insect samples

DNA was extracted from 59 arthropod taxa belonging to 12 orders following the protocol in Lindner and Banik (2009) with modifica-tions for insects (Supporting Information Appendix S1). Briefly, DNA was extracted from excised leg muscles of larger insects, or for smal-ler insects, the thorax was punctured and the entire insect was used for extraction. Leg muscles and small insects with punctured tho-raxes were placed in 100 μL of filtered cell lysis solution (CLS; Lind-ner & Banik, 2009) and frozen at −20°C, and the extraction proceeded. Following DNA extraction, the effectiveness of the fol-lowing five primer pairs in amplifying the 59 purified DNAs was

TABLE 1 Sequences and references, primer pair names and thermocycler conditions for the primers tested against known arthropod samples

Pair	Primers	Primer sequences	References	Thermocycler conditions
CO1 L/H	LCO1490 HCO2198	5′-GGTCAACAAATCATAAAGATATTGG–3′ 5′-TAAACTTCAGGGTGACCAAAAAATCA–3′	Folmer et al. (1994) Folmer et al. (1994)	94°C for 60 s; 5 cycles of: 94°C for 60 s, 45°C for 90 s, 72°C for 90 s; 35 cycles of: 94°C for 60 s, 50°C for 90 s, 72°C for 60 s; 72°C for 7 min (Hebert et al., 2003)
LEP	LEPF1 mLEPR	5′-ATTCAACCAATCATAAAGATATTGG–3′ 5′-CTTGTTCCAGCTCCATTTT–3′	Hebert et al. (2004) Smith et al. (2006)	94°C for 2 min; 5 cycles of: 94°C for 40 s, 45°C for 40 s, 72°C for 60 s; 35 cycles of: 94°C for 40 s, 51°C for 40 s, 72°C for 60 s; 72°C for 5 min (Smith et al., 2006)
ZBJ	ZBJ-ArtF1c ZBJ-ArtR2c	5′-AGATATTGGAACWTTATATTTTATTTTTGG–3′ 5′-WACTAATCAATTWCCAAATCCTCC-3′	Zeale et al. (2011)	94°C for 3 min; 16 cycles of: 94°C for 30 s, 61C for 30 s (decrease by ½ C per cycle), 72C for 30 s; 24 cycles of: 94C for 30 s, 53°C for 30 s, 72C for 30 s; 72C for 10 min (Zeale et al., 2011)
ANML	LCO1490 CO1-CFMRa	5′-GGTCAACAAATCATAAAGATATTGG-3′ 5′-GGWACTAATCAATTTCCAAATCC-3′	Folmer et al. (1994) This study	94°C for 60 s; 5 cycles of: 94°C for 60 s, 45°C for 90 s, 72°C for 90 s; 35 cycles of: 94°C for 60 s, 50°C for 90 s, 72°C for 60 s; 72°C for 7 min (Hebert et al., 2003)
CFMRb	LCO1490 CO1-CFMRb	5′-GGTCAACAAATCATAAAGATATTGG-3′ 5′-GGNACTAATCAATTHCCAAATCC-3′	Folmer et al. (1994) This study	94°C for 60 s; 5 cycles of: 94°C for 60 s, 45°C for 90 s, 72°C for 90 s; 35 cycles of: 94°C for 60 s, 50°C for 90 s, 72°C for 60 s; 72°C for 7 min (Hebert et al., 2003)

evaluated: LCOI-1490/HCOI-2198 (Folmer, Black, Hoeh, Lutz, & Vrijenhoek, 1994; hereafter COI L/H), ZBJ-ArtF1c/ZBJ-ArtR2c (Zeale et al., 2011; ZBJ), LCOI-1490/COI-CFMRa (ANML), LCOI-CFMRb (CFMRb) and LepF1/mLepR (Hebert, Penton, Burns, Janzen, & Hallwachs, 2004; Smith, Woodley, Janzen, Hallwachs, & Hebert, 2006; LEP). The COI-CFMRa and COI-CFMRb primers designed for this study were derived from the ZBJ-ArtR2c primer and had sequences of 5'-GGWACTAATCAATTTCCAAATCC-3' and 5'-GGNACTAATCAATTHCCAAATCC-3', respectively. Specifically, we compared Sanger sequences obtained from the COI L/H products (methodology described below) from the arthropod taxa included in the primer test and realized that an improved reverse HTS primer could be made by modifying the 5' and 3' ends of ZBJ-ArtR2c. We also made additional changes to the middle base pairs according to the in silico analysis of our arthropod test set. The COI-CFMRa and COI-CFMRb priming sites are located in the COI gene approximately 180 bp from the LCOI-1490 priming site, with which they are paired (Supporting Information Figure S1). We used the LCOI-1490 primer rather than modifying the forward ZBJ-ArtF1c primer because a) ZBJ-ArtF1c has some suboptimal primer characteristics (e.g., excessive length, multiple homopolymer runs and low GC content), b) it did not perform well in our analysis, and c) it is located near the LCOI-1490 site. A list of all primers used is presented in Table 1.

Amplification of the extracted DNA using all primer pairs, except ZBJ, used the following reagent volumes and concentrations per 15 µl reaction: 7.88 µl DNA-free molecular grade water, 3 µl Green GoTaq 5x buffer (Promega)—final concentration 1x, 0.12 µl of 20 mg/ml BSA (New England BioLabs), 0.3 µl of 10 mM dNTPs (Promega)—final concentration 200 µM of each dNTP, 0.3 µl of each 10 µM primer—final concentration of 0.2 µM of each primer, 0.1 µl

of 5u/µl GoTaq polymerase (Promega)—final concentration of 0.033u/µl of the reaction and 3 µl of extracted arthropod template DNA. The ZBJ primer pair was used with two different reagent protocols. One, termed the modified protocol, was the same as above except 1.0 µl of each 10 µM primer was added, and the second protocol was that described by original authors (Zeale et al., 2011). Thermocycler conditions for all primer pairs are found in Table 1. Briefly, the thermocycler parameters for the COI L/H, ANML and CFMRb primer pairs were those described by Hebert et al. (2003) with one modification: the final extension at 72°C was increased from 5 to 7 min. The LEP amplification parameters were those of Smith et al. (2006), while the ZBJ primer pair amplification parameters were those described by Zeale et al. (2011). Following amplification, 3 µl of product was run in a 2% agarose gel for 20 min at 110 V, stained with ethidium bromide and visualized using UV light. The presence or absence of bands was recorded for each primer pair and DNA combination. To provide reference sequences for each species, the COI L/H PCR products were Sanger-sequenced with ABI Prism BigDye (Applied Biosystems) sequencing following the method of Lindner and Banik (2009). The resulting sequences were subjected to an NCBI BLAST (National Center for Biotechnology Information; Basic Local Alignment Search Tool) search to confirm the identities of the insect species.

2.2 | HTS of field-collected guano samples using two different primer pairs

The arthropod DNA present in three field-collected bat guano samples was analysed using the ANML and ZBJ primer pairs. DNA was extracted from three *Myotis lucifugus* guano samples from three

different locations in southern Wisconsin. Each sample consisted of 100 mg of guano, approximately 10 pellets. DNA was extracted from each sample using a QIAamp DNA Stool Mini Kit (Qiagen) following the procedure in Appendix S2 of the Supporting Information. The DNA from each sample was then amplified (in two separate reactions) using both the ANML and ZBJ primers modified for metabarcoding by adding an Ion Torrent Xpress trP1 adapter sequence on the reverse primer and a unique barcode sequence and Ion Torrent Xpress A adapter sequence on each forward primer (see Supporting Information Table S1, for barcoded primer sequences). Amplification conditions for the ANML pair followed the protocol used for the primer pair test, and conditions for the ZBJ pair followed the modified protocol for ZBJ described in the primer pair test. Following amplification, each of the uniquely barcoded PCR products was purified via size selecting E-Gel CloneWell Gels (Invitrogen) at approximately 157 bp for ZBJ and approximately 180 bp for ANML. The size-selected products were then quantified on an Invitrogen Qubit 2.0 Fluorometer and brought to a concentration of 2 nM using DNA-free, molecular grade water. We then combined the products in equal amounts to produce the sequencing library. The library was diluted to 13 pM prior to templating onto ion sphere particles (ISPs) with the Ion OneTouch 2 system (Life Technologies) and a PGM Hi-Q OT2 templating kit (Thermo Fisher #A27739), according to the manufacturer's recommendations. The templated ISPs were then purified, and the templated DNA was sequenced using the Ion Torrent Personal Genome Machine (PGM; Thermo Fisher) with the Ion PGM Hi-Q Sequencing Kit (Thermo Fisher #A25592), according to the manufacturer's protocol for 400-bp sequencing. The samples from both primer pairs were sequenced on the same PGM run, using an Ion 314 chip (Thermo Fisher).

2.3 | Bioinformatics

HTS data were processed using the "DADA2" method via the AMPtk pipeline (https://amptk.readthedocs.io; Palmer et al., 2018). Briefly, the AMPtk pipeline processes (de-multiplexes) HTS amplicon sequencing reads by (a) identifying a valid barcode index in each read (with no mismatches allowed), (b) identifying forward and reverse primer sequences, (c) trimming barcode and primer sequences, (d) renaming the read based on barcode index and (e) trimming/padding the reads to a set length. The DADA2 algorithm (Callahan et al., 2016) is an alternative to widely used sequence-clustering algorithms (e.g., UPARSE, UCLUST, nearest neighbour and SWARM) and functions to "denoise" HTS sequencing reads. DADA2 has been shown to be very accurate and is sensitive to single base-pair differences between sequences (Callahan et al., 2016). AMPtk implements a modified DADA2 algorithm that produces the standard "amplicon sequence variants (ASVs)" output of DADA2 and then clusters the ASVs into biologically relevant OTUs using the UCLUST (Edgar, 2010) algorithm employed in VSEARCH (Rognes, Flouri, Nichols, Quince, & Mahé, 2016). This clustering step also removes any single-read OTUs. The resulting AMPtk OTU tables can be filtered based on spike-in mock communities (described below).

Taxonomy for COI is assigned in AMPtk using a combination of global sequence alignment, UTAX (https://www.drive5.com/manual/utax_algo.html) and SINTAX (Edgar, 2016) using a COI reference database. The current COI database distributed with AMPtk was derived from collating sequences from representative barcode index numbers (BIN) from chordates and arthropods in the Barcode of Life v4 database (BOLD; Ratnasingham & Hebert, 2007) and is available at https://amptk.readthedocs.io/en/latest/taxonomy.html.

2.4 | Development and testing of an arthropod mock community

To produce a mock community to serve as a control for HTS data analysis, 43 of the arthropod taxa used in the primer pair test were chosen as candidates (Table 2). DNA from each arthropod was amplified using COI L/H primers as described previously. To remove intragenomic variation (Song, Buhay, Whiting, & Crandall, 2008), the resulting amplicons were cloned into *E. coli* using the Promega pGEM-T vector system following the manufacturer's instructions with the modifications used by Lindner and Banik (2009). Three clones of each arthropod taxon were subsequently Sanger-sequenced to verify the presence of the COI insert sequence. Two of the cloned arthropods produced cloned sequence variants, and these variants (three in total) were also included in the mock community, bringing our mock community total to 46 plasmids. Plasmids were purified using standard alkaline lysis, and the resultant DNA was then quantified on an Invitrogen Qubit 2.0 Fluorometer and brought to a concentration of 1,500 pM using DNA-free, molecular grade water. Plasmids were then diluted to a 1:20 concentration using DNA-free molecular grade water and individually amplified using the ion ANML primers with the same barcode. The individual PCR products were then visualized on a 2% agarose gel, cleaned and size-selected at ≥150 bp using Zymo Research Select-A-Size DNA Clean & Concentrator spin columns, quantified and equilibrated to 2 nM as described previously and subsequently combined in equal amounts. This amplicon mixture is referred to as our "post-PCR combined mock community," which serves as a control to validate sequencing efficiency of each mock member. To measure initial PCR bias and to parameterize our bioinformatics pipeline, we also created "a pre-PCR combined mock community" by combining our 1,500 pM plasmids in equal amounts. The pre-PCR combined mock community was then diluted to a 1:8,000 concentration prior to amplification with ANML-barcoded primers. The resulting barcoded PCR product was then visualized, size-selected, quantified and brought to 2 nM as described before. The resulting barcoded PCR products were then prepared and sequenced on an Ion Torrent PGM as described above, but using an Ion 318 chip (Thermo Fisher), and data were bioinformatically processed as described above.

2.5 | Testing of known mixed samples with mock community and our pipeline

To test prey DNA recovery from bat guano, two bats, one *Eptesicus fuscus* and one *Lasiurus cinereus*, were fed known diets of *Galleria*

TABLE 2 Results from testing the five primer pairs listed in Table 1 on known arthropod samples, listed by order. Representative shading indicates the proportion of representatives of each order amplified. See Table S2 of Supporting Information for results by arthropod sample. Amplification was attempted on a variety of DNA concentrations for each template DNA sample before assigning a value of zero (no amplification)

Order	ANML	CFMRb	CO1 L/H	ZBJ Zeale et al. 2011 protocol	ZBJ modified protocol	LEP	
Blattodea	2/2 (100%)	2/2 (100%)	2/2 (100%)	1/2 (50%)	1/2 (50%)	2/2 (100%)	90-100%
Coleoptera	9/10 (90%)	9/10 (90%)	9/10 (90%)	4/10 (40%)	4/10 (40%)	8/10 (80%)	75-89%
Dermaptera	1/1 (100%)	1/1 (100%)	0/1 (0%)	0/1 (0%)	0/1 (0%)	1/1 (100%)	50-74%
Diptera	8/8 (100%)	8/8 (100%)	7/8 (88%)	4/8 (50%)	5/8 (63%)	8/8 (100%)	25-49%
Ephemeroptera	4/4 (100%)	4/4 (100%)	3/4 (75%)	1/4 (25%)	1/4 (25%)	3/4 (75%)	1-24%
Hemiptera	6/6 (100%)	6/6 (100%)	5/6 (83%)	1/6 (17%)	2/6 (34%)	3/6 (50%)	0%
Hymenoptera	6/6 (100%)	6/6 (100%)	6/6 (100%)	2/6 (34%)	2/6 (34%)	2/6 (34%)	
Lepidoptera	11/11 (100%)	11/11 (100%)	11/11 (100%)	9/11 (82%)	9/11 (82%)	11/11 (100%)	
Neuroptera	2/2 (100%)	2/2 (100%)	2/2 (100%)	0/2 (0%)	1/2 (50%)	2/2 (100%)	
Opiliones	1/1 (100%)	1/1 (100%)	1/1 (100%)	0/1 (0%)	0/1 (0%)	1/1 (100%)	
Orthoptera	3/3 (100%)	3/3 (100%)	3/3 (100%)	1/3 (33%)	1/3 (33%)	3/3 (100%)	
Trichoptera	5/5 (100%)	5/5 (100%)	3/5 (60%)	1/5 (20%)	1/5 (20%)	4/5 (80%)	
Total	58	58	52	24	27	48	
% Total	98.31	98.31	88.14	40.68	45.76	81.36	

mellonella, *Tenebrio molitor* and *Antheraea polyphemus* alone and in combination (Table 4). The biomass of the diet was the same regardless of the combination, and all bats were weighed each day. The bats were fed each known diet for one day, dietary components were added sequentially, and guano pellets were collected during the following 24 hr (approved by Boise State University Institutional Animal Care and Use Committee 006-AC14–018). We analysed three known-diet combinations from the *E. fuscus* individual and two known-diet combinations from the *L. cinereus* individual. DNA was extracted from guano samples using Qiagen QIAamp mini stool kits, following the modified protocol described in Zeale et al. (2011). DNA from the known-diet samples was amplified with barcoded

ANML primers, and the resulting PCR products were then visualized, size-selected at ≥150 bp using Zymo Research Select-A-Size DNA clean and concentrator spin columns, quantified, brought to 2 nM, and sequenced and processed as described before. These samples and the samples described below were sequenced on the same chip/run as the arthropod mock community samples.

To test the effectiveness of the method on complex insect communities, five samples from pitfall traps from the Snake River Birds of Prey Conservation Area in Kuna, Idaho, were analysed. Each pitfall trap consisted of a glass jar containing propylene glycol. Traps were left outside for 2–3 days, at which point the contents of the traps were rinsed with 100% ethanol and subsequently transferred

to glass vials containing 100% ethanol for storage at room temperature. All trap samples were sent to the Florida Museum of Natural History in Gainesville, Florida, for visual identification to arthropod family and long-term storage at −20°C. Initial identities of the arthropods present in the insect trap samples were obtained using traditional morphological keys, and most were identified by eye to the family level, with the following exceptions: all springtails were identified to order (Collembola), centipedes were identified to class (Chilopoda), and mites were identified to subclass (Acari).

The samples were sent to the US Forest Service, Northern Research Station, Center for Forest Mycology Research in Madison, Wisconsin, where they were processed for molecular analysis. Arthropods from the trap samples were rinsed in DNA-free molecular grade water and prepared for DNA extraction in two ways: (a) the excised leg muscles of larger arthropods and smaller arthropods with open thoraxes were combined and submersed in CLS and vortexed (dissected sample) or (b) the intact arthropods were added to 15 ml CLS and macerated with a sterile pestle and vortexed (macerated sample). DNA extraction followed details described in Appendix S1 of the Supporting Information; metabarcoding PCR and HTS then proceeded as previously described. Data were bioinformatically processed as described before. We used permutational multivariate analysis of variance (PERMANOVA) of a modified Raup–Crick distance matrix (Chase, Kraft, Smith, Vellend, & Inouye, 2011) using the adonis function in the vegan package of R (Oksanen et al., 2011, R Core Team, 2017) to test for an effect of: extraction method (dissected vs. macerated), identification method (key vs. HTS), and finally trap number on recovered community composition. To further test the efficacy of each extraction method, we calculated the difference in taxonomic richness recovered from dissected and macerated samples. We then used a *t*-test to compare these differences between dissected and macerated samples. To test for an effect of identification method on recovered taxonomic richness, we used a paired *t*-test to compare the richness recovered from each trap by both methods.

3 | RESULTS

3.1 | Testing of primer pairs against known insect samples

Fifty-eight of the 59 taxa (98%) amplified with the ANML (LCOI490/COI-CFMRa) and CFMRb (LCOI490/COI-CFMRb) primer pairs, with both pairs failing to amplify the same carabid beetle (Table 2). Fifty-two of 59 taxa (89%) amplified with the COI L/H primer pair, and 48 of 59 (81%) amplified with the LEP primer pair; the LEP pair amplified 100% of the Lepidopterans and Dipterans tested. The ZBJ primer pair amplified 24 of the 59 (41%) taxa tested with the Zeale et al.'s (2011) protocol and 27 of the 59 (46%) taxa using our modified protocol (Table 2; Supporting Information Table S2).

In an attempt to increase the performance of ZBJ, we also tested the pair using (a) the protocol used for ANML (described above) and (b) the standard protocol described by Clarke et al. (2014). The performance increased to 40 of 59 (68%) and 37 of 59 (63%),

respectively, still lower than the 98% amplified by ANML (Supporting Information Table S2).

3.2 | HTS of field-collected guano samples using two different primer pairs

Both the ZBJ and the ANML primers produced an amplification product from the three *Myotis lucifugus* guano samples. For both primer sets combined, a total of 64 OTUs (Table 3) were detected, of which 59 could be identified to the family level, representing 10 orders comprised of 28 families. The ANML primers detected 56 OTUs, and the ZBJ primers detected 15 OTUs. Seven of the 64 total OTUs were detected with both sets of primers, 49 were detected only with the ANML primers, while eight were only detected with the ZBJ primers. Representatives from all ten orders and 26 families were recovered using the ANML primer pair, while the ZBJ pair recovered representatives from three orders and eight families. The most often detected family was the dipteran midge family Chironomidae, with 27 OTUs, 24 of which were detected by the ANML primers and six by ZBJ. The second most often detected family was the mosquito family Culicidae, with five OTUs detected by ANML but only one by ZBJ. All but one of the remainder of the families were represented by only one OTU each (Table 3). An average of 25% of reads produced by ANML from the *Myotis lucifugus* samples were from an OTU representing bat DNA; however, this loss of reads to bat DNA did not offset the far greater taxonomic coverage of ANML primers and the resulting higher diversity of detected arthropod taxa.

3.3 | Development and testing of an arthropod mock community

Despite equal input DNA concentrations, the pre-PCR combined mock community generated highly uneven read counts among the individual plasmid components after sequencing. In contrast, the post-PCR combined mock community generated relatively even read counts. The individual plasmid components of our post-PCR combined mock community generated read counts that ranged from 3,740 to four; the mean was 2,119 and standard deviation ±799, with 89% (41 of 46) yielding greater than 1,500 reads (Figure 1). All mock members in the post-PCR combined community were recovered, although three generated final read counts below 100 (range 4 to 12). In contrast, individual members of our pre-PCR combined mock community generated read counts that ranged from 10,577 to 0 with a mean of 2,174 and standard deviation of ±2,238, with 54% (25 of 46) yielding more than 1,500 reads. Two of our mock members did not generate any sequences in the pre-PCR combined community and an additional four generated final read counts below 100 (range 2 to 39; Figure 1).

3.4 | Testing of known mixed samples with mock community and our pipeline

The results of the known-diet HTS samples are summarized in Table 4. We detected DNA from all of the expected dietary

ANML	ZBJ	Class	Order	Family	Species
1	0	Arachnida	Araneae	Theridiidae	*Theridion frondeum*
2	0	Arachnida	Trombidiformes	Limnesiidae	*Limnesia* sp.
1	0	Arachnida	Trombidiformes		Trombidiformes sp.
1	0	Arachnida			Arachnida sp.
1	0	Insecta	Coleoptera	Coccinellidae	*Harmonia* sp.
2	0	Insecta	Coleoptera	Elateridae	*Melanotus similis*
1	0	Insecta	Coleoptera	Hydrophilidae	*Helocombus bifidus*
1	0	Insecta	Coleoptera	Scarabaeidae	Scarabaeidae sp.
1	0	Insecta	Coleoptera	Tenebrionidae	*Tenebrio* sp.
1	0	Insecta	Coleoptera		Coleoptera sp.
1	0	Insecta	Diptera	Bibionidae	*Bibio* sp.
1	0	Insecta	Diptera	Ceratopogonidae	*Bezzia* sp.
0	2	Insecta	Diptera	Chaoboridae	*Chaoborus punctipennis*
1	1	Insecta	Diptera	Chironomidae	*Ablabesmyia americana*
1	0	Insecta	Diptera	Chironomidae	*Ablabesmyia annulata*
1	0	Insecta	Diptera	Chironomidae	*Ablabesmyia* sp. 1
1	0	Insecta	Diptera	Chironomidae	*Ablabesmyia* sp. 2
3	0	Insecta	Diptera	Chironomidae	*Chironomus plumosus*
1	1	Insecta	Diptera	Chironomidae	*Chironomus* sp. 1
0	2	Insecta	Diptera	Chironomidae	*Chironomus* sp. 2
1	0	Insecta	Diptera	Chironomidae	*Coelotanypus* sp.
1	0	Insecta	Diptera	Chironomidae	*Conchapelopia* sp.
1	0	Insecta	Diptera	Chironomidae	*Cryptochironomus* sp. 1
2	0	Insecta	Diptera	Chironomidae	*Cryptochironomus* sp. 2
2	0	Insecta	Diptera	Chironomidae	*Dicrotendipes tritomus*
2	1	Insecta	Diptera	Chironomidae	*Parachironomus* sp. 1
1	0	Insecta	Diptera	Chironomidae	*Parachironomus* sp. 2
1	0	Insecta	Diptera	Chironomidae	*Parachironomus* sp. 3
2	0	Insecta	Diptera	Chironomidae	*Polypedilum* sp. 1
1	0	Insecta	Diptera	Chironomidae	*Polypedilum* sp. 2
2	2	Insecta	Diptera	Chironomidae	*Procladius* sp. 1
1	0	Insecta	Diptera	Chironomidae	*Procladius* sp. 2
1	0	Insecta	Diptera	Chironomidae	*Procladius* sp. 3
0	1	Insecta	Diptera	Chironomidae	*Procladius* sp. 4
1	0	Insecta	Diptera	Chironomidae	*Xenochironomus* sp.
3	0	Insecta	Diptera	Chironomidae	Chironomidae sp. 1
1	0	Insecta	Diptera	Chironomidae	Chironomidae sp. 2
2	0	Insecta	Diptera	Chironomidae	Chironomidae sp. 3
2	0	Insecta	Diptera	Chironomidae	Chironomidae sp. 4
0	1	Insecta	Diptera	Chironomidae	Chironomidae sp. 5
1	0	Insecta	Diptera	Culicidae	*Aedes abserratus*
1	0	Insecta	Diptera	Culicidae	*Aedes excrucians*
1	0	Insecta	Diptera	Culicidae	*Aedes provocans*
1	1	Insecta	Diptera	Culicidae	*Aedes vexans*
1	0	Insecta	Diptera	Culicidae	*Culiseta melanura*
1	0	Insecta	Diptera	Hybotidae	*Platypalpus* sp.
0	1	Insecta	Diptera	Limoniidae	*Shannonomyia lenta*
1	0	Insecta	Diptera	Psychodidae	*Psychoda alternata*
1	1	Insecta	Diptera	Tachinidae	Tachinidae sp.
1	0	Insecta	Diptera	Tipulidae	*Nephrotoma ferruginea*
0	1	Insecta	Diptera	Tipulidae	*Tipula kennicotti*
1	1	Insecta	Ephemeroptera	Caenidae	*Caenis youngi*
1	0	Insecta	Ephemeroptera	Palingeniidae	*Pentagenia vittigera*
1	0	Insecta	Ephemeroptera	Siphlonuridae	*Siphlonurus typicus*
1	0	Insecta	Hemiptera	Corixidae	*Trichocorixa borealis*
1	0	Insecta	Hemiptera	Miridae	*Lygus lineolaris*
0	1	Insecta	Lepidoptera	Blastobasidae	*Blastobasis glandulella*
1	0	Insecta	Lepidoptera	Depressariidae	*Antaeotricha leucillana*
1	0	Insecta	Lepidoptera	Tortricidae	*Argyrotaenia pinatubana*
1	0	Insecta	Lepidoptera		Lepidoptera sp.
1	0	Insecta	Megaloptera	Corydalidae	*Chauliodes rastricornis*
1	0	Insecta	Trichoptera	Hydroptilidae	*Oxyethira serrata*
0	2	Insecta			Insecta sp.
3	0	Mammalia	Chiroptera	Vespertilionidae	*Myotis lucifugus*

TABLE 3 Operational taxonomic units (OTUs) recovered using high-throughput amplicon sequencing (HTS) and either the ANML primers or the ZBJ primers on three field-collected guano samples. Numbers (0–3) and representative shading indicate the number of guano samples each OTU was detected in for each primer pair
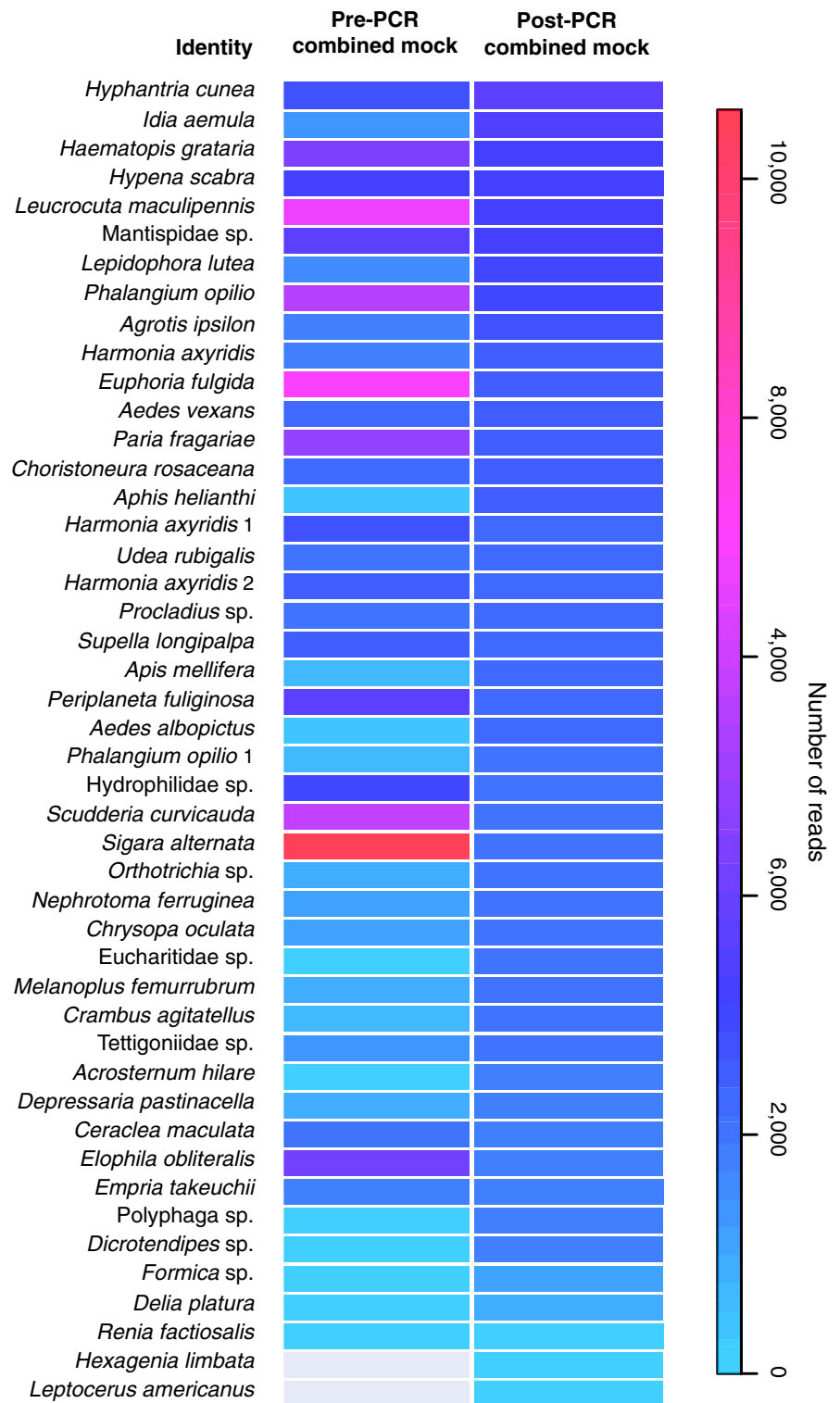
FIGURE 1 Heat map of the high-throughput amplicon sequencing read numbers of the arthropod mock community, equilibrated and combined both pre- and post-PCR (with ANML primers). The post-PCR combined mock community was far more even and representative of the equal amounts of DNA added for each mock member than the pre-PCR combined mock community [Colour figure can be viewed at wileyonlinelibrary.com]

components in all five of the known-diet samples tested. As each new dietary component was added, it was detected in subsequent faecal samples. Additionally, we detected DNA from two possible accidental dietary components (*Empria takeuchii* and *Agrotis ipsilon*) in big brown bat (*Eptesicus fuscus*) diet samples that included both *Galleria* and *Tenebrio* as dietary components. We also detected DNA from a parasitoid wasp (family: Ichneumonidae) in three of four (75%) diet samples that included *Galleria* larvae. Finally, we detected

big brown bat (*E. fuscus*) DNA in two of the three samples from big brown bats and hoary bat (*Lasiurus cinereus*) DNA in both (2 of 2) of the samples from hoary bats (Table 4). Although we did detect Chi-ropteran DNA in four of the five samples, this only accounted for 2.3% (7,804 of 343,035) of our total reads from *E. fuscus* and 0.3% (736 of 236,231) of our total reads from *L. cinereus*. All other reads attained were from arthropods (Table 4). These data were processed bioinformatically with DADA2, with and without 97% clustering

TABLE 4 Number of OTUs identified from bats fed known insect diets, broken down by expected dietary components, possible accidental dietary components and bat DNA. Blanks are zeros. DADA2 is data from DADA2, without clustering. DADA2 97% cluster is data from DADA2 with 97% clustering applied to the OTU table. Shaded cells are dietary components that were expected (i.e., known to be fed to the bat). EPFU1, EPFU2 and EPFU3 are from big brown bats (*Eptesicus fuscus*), and LACI1 and LACI2 are from hoary bats (*Lasiurus cinereus*)

| | Expected dietary components | | | | | | Possible accidental dietary components | | | | | | Bat DNA | | | |
| | *Galleria mellonella* | | *Tenebrio molitor* | | *Antheraea polyphemus* | | *Empria takeuchii* | | *Agrotis ipsilon* | | *Ichneumonidae sp.* | | *Eptesicus fuscus* | | *Lasiurus cinereus* | |
	DADA2	DADA2 97% Cluster	DADA2	DADA2 97% Cluster	DADA2	DADA2 97% Cluster	DADA2	DADA2 97% Cluster	DADA2	DADA2 97% Cluster	DADA2	DADA2 97% Cluster	DADA2	DADA2 97% Cluster	DADA2	DADA2 97% Cluster
EPFU1	1	1	8	1			1	1			1	1	1	1		
EPFU2	1	1	1	1	1	1	1	1	1	1	7	1	1	1		
EPFU3	1	1									1	1				
LACI1	1	1	7	1											3	2
LACI2			11	2											3	2

applied to the inferred sequence table that resulted from the DADA2 output. Without clustering, we obtained one inferred sequence for *G. mellonella*, *Antheraea polyphemus*, *E. takeuchii*, *A. ipsilon* and *E. fuscus*, but obtained 11 inferred sequences for *Tenebrio molitor*, seven from Ichneumonidae and three for *L. cinereus*. After clustering at 97%, we maintained the OTU number for all taxa that had one OTU before clustering and obtained two OTUs for *T. molitor*, one OTU for Ichneumonidae and two OTUs for *L. cinereus*.

The results of the pitfall trap samples are summarized based on the presence or absence of families in Table 5. Overall, in five samples, 37 families that were identified using conventional morphological methods were also recovered with HTS, while a further 18 families morphologically identified were not recovered with HTS, and 16 families were only recovered with HTS. Of the 18 families not detected by HTS, nine were probably a result of either a morphological or sequence misidentification, with the remaining nine most likely lost through system bias or sample degradation. We detected no significant effect of the method in which the communities were extracted (dissected vs. macerated samples) on the composition of the communities recovered (PERMANOVA, $p = 0.09$). There was also no significant effect of the identification method (key vs. HTS; PERMANOVA, $p = 0.79$). Trap number was the only significant predictor of recovered community composition (PERMANOVA, $p < 0.0001$, $r^2 = 0.99$) and explained nearly all the variation in the data indicating that extraction method and identification method contributed very little variation to the overall data set (see Supporting Information Figure S2). There was also no significant difference in the taxonomic richness recovered from the two extraction methods (dissected vs. macerated samples; $p = 0.19$) or an effect of identification method on recovered taxonomic richness (key vs. HTS; $p = 0.89$).

4 | DISCUSSION

Through an amplification test of five primer pairs against a taxonomically diverse community of arthropods, we demonstrated that our ANML primer pair amplified more taxa than previously described primer pairs (COI L/H, ZBJ and LEP; Table 2; Supporting Information Table S2). Through a direct comparison of field-collected guano samples subjected to HTS with two primer pairs, ANML and ZBJ, we demonstrated that the ANML primer pair amplified substantially more taxa than the ZBJ primer pair (Table 3), the commonly used primer pair for HTS studies examining the diets of insectivorous animals. When we used both pairs on the same environmental samples, the ANML pair yielded almost four times as many arthropod OTUs, representing three times more arthropod orders and families than the ZBJ pair. We also detected chiropteran (bat) sequences in faecal samples from bats using the ANML primer pair, with the average percentage of the total number of sequences of chiropteran origin being 0.3%, 2.3% and 25% for samples collected from *Eptesicus fuscus*, *Lasiurus cinereus* and *Myotis lucifugus*, respectively. Although we detected more Chiropteran DNA from the *Myotis lucifugus* samples than from

TABLE 5 Comparison of morphological and HTS family-level identifications of arthropods collected from five pitfall traps. Arthropods from traps 1 and 4 (underlined) were dissected pre-extraction, and arthropods from traps 2, 3 and 5 (not underlined) were macerated pre-extraction. A "+" indicates the presence of a family

Class	Order/subclass	Family	Trap 1 Key	Trap 1 HTS	Trap 2 Key	Trap 2 HTS	Trap 3 Key	Trap 3 HTS	Trap 4 Key	Trap 4 HTS	Trap 5 Key	Trap 5 HTS
		Total Taxa	7	10	17	13	17	19	4	5	10	7
Insecta	Blattodea	Ectobiidae				+						
Insecta	Coleoptera	Carabidae			+	+						
Insecta	Coleoptera	Elateridae			+							
Insecta	Coleoptera	Melyridae					+	+			+	+
Insecta	Coleoptera	Ptinidae/Anobiidae			+							
Insecta	Coleoptera	Scarabaeidae				+						
Insecta	Coleoptera	Silphidae					+	+				
Insecta	Coleoptera	Tenebrionidae			+	+						
Entognatha	Collembola						+	+				
Insecta	Diptera	Anthomyiidae			+	+	+	+				
Insecta	Diptera	Bombyliidae									+	+
Insecta	Diptera	Calliphoridae			+		+	+				
Insecta	Diptera	Cecidomyiidae					+					
Insecta	Diptera	Culicidae						+				
Insecta	Diptera	Diptera sp.						+				
Insecta	Diptera	Heleomyzidae				+						
Insecta	Diptera	Phoridae			+		+	+				
Insecta	Diptera	Scathophagidae			+	+						
Insecta	Diptera	Sciaridae				+						
Insecta	Diptera	Syrphidae									+	+
Insecta	Hemiptera	Aphididae	+	+							+	+
Insecta	Hemiptera	Cicadidae	+	+								
Insecta	Hemiptera	Cicadellidae	+	+	+		+				+	
Insecta	Hemiptera	Geocoridae			+							
Insecta	Hemiptera	Miridae	+	+								
Insecta	Hemiptera	Pentatomidae		+						+		
Insecta	Hemiptera	Psyllidae	+	+								
Insecta	Hymenoptera	Formicidae			+	+	+	+	+	+	+	+
Insecta	Hymenoptera	Braconidae						+				
Insecta	Hymenoptera	Ceraphronidae					+					
Insecta	Hymenoptera	Chalcidoidea					+					
Insecta	Hymenoptera	Crabronidae									+	
Insecta	Hymenoptera	Dryinidae		+								
Insecta	Hymenoptera	Halictidae									+	+
Insecta	Hymenoptera	Hymenoptera sp.						+				
Insecta	Hymenoptera	Ichneumonidae			+	+	+					
Insecta	Hymenoptera	Pompilidae							+	+		
Insecta	Lepidoptera	Gelechiidae					+	+				
Insecta	Lepidoptera	Tortricidae									+	+
Insecta	Neuroptera	Chrysopidae						+				
Insecta	Orthoptera	Acrididae	+									
Insecta	Orthoptera	Tettigoniidae		+								

(Continues)

TABLE 5 (Continued)

Class	Order/subclass	Total Taxa Family	Trap 1 Key	Trap 1 HTS	Trap 2 Key	Trap 2 HTS	Trap 3 Key	Trap 3 HTS	Trap 4 Key	Trap 4 HTS	Trap 5 Key	Trap 5 HTS
			7	10	17	13	17	19	4	5	10	7
Insecta	Thysanoptera	Thripidae			+							
Insecta	Thysanoptera	Hydroptillidae						+				
Arachnida	Acari		+	+	+	+	+	+	+	+		
Arachnida	Araneae	Araneae sp.				+						
Arachnida	Araneae	Gnaphosidae			+	+	+	+				
Arachnida	Araneae	Lycosidae			+	+						
Arachnida	Araneae	Pisauridae			+							
Arachnida	Araneae	Salticidae					+	+		+	+	
Arachnida	Araneae	Thomisidae								+		
Chilopoda							+	+				

Eptesicus fuscus and *Lasiurus cinereus* samples, we still detected far greater taxonomic coverage of arthropods in *M. lucifugus* samples subjected to HTS with the ANML primers compared to the ZBJ primers. Thus, the amplification of chiropteran DNA did not significantly impact the recovery of arthropod DNA and helps to confirm the identity of the bat target species, as well as their dietary components. It is likely that the COI region of other vertebrates could also be amplified by the ANML primers, thus helping to confirm the identity of the consumer in a range of systems (e.g., other mammal species, reptiles, amphibians and birds). Because they produce longer PCR products, the ANML primers (180-bp product) may also allow for better taxon delineation compared to the ZBJ primers (157-bp product).

4.1 | Improved detection of pest species

Insectivorous animals are valued as providers of pest control; however, the total economic value of this ecosystem service is difficult to estimate (Boyles, Cryan, McCracken, & Kunz, 2011; Cleveland et al., 2006; Maine & Boyles, 2015; Williams-Guillén et al., 2016). Determining the full value is dependent on the reliable detection of the pest species present in the diets of insectivorous animals. HTS can be a powerful tool for helping to build the empirical basis necessary to estimate ecosystem services, but the success of this approach depends in part on primer efficacy. Based on our analyses, the ANML primers are a major methodological improvement over existing primers, allowing for the detection of greater arthropod diversity in the environmental samples we tested, including a greater diversity of known pests such as mosquitoes (family: Culicidae). Specifically, in our environmental guano samples, the ZBJ pair was only able to detect *Aedes vexans*, while the ANML pair detected *A. vexans* plus four additional Culicidae species in the same samples (Table 3). Thus, the ANML primers allow for better estimation of the ecosystem services of bats, and perhaps other insectivores, as

predators of mosquitoes and other economically important pest species.

4.2 | Single-copy arthropod mock community, sources of unexpected variation and some solutions

While some authors have noted that HTS data are unreliable as a source to measure community member abundance (Piñol, Mir, Gomez-Polo, & Agustí, 2015), many HTS studies of environmental samples continue to use abundance metrics based on read numbers. To test the validity of read number as an estimate of relative abundance and to determine whether biases were introduced at the PCR and/or sequencing steps, we combined pre- and post-PCR mock communities in equimolar amounts prior to sequencing. We predicted that if the approach is valid, read numbers should be equal across taxa. Instead, even though each member of the mock community amplified well in individual PCRs, we observed a large variation in read numbers for the pre-PCR combined mock community, with some members being absent. In contrast, the post-PCR combined mock was far less variable (Figure 1). The initial PCR introduced a large amount of taxonomic bias by preferentially amplifying some taxa, as inferred from the difference in variability in read numbers between the post- and pre-PCR mixes of our arthropod mock community. Sequencing itself also introduced bias resulting in differences in read numbers between the mock members that were combined post-PCR. Some of the variation in read numbers among mock community members was probably induced by mismatches in the priming site, given that some members have three or more primer mismatches. While this number of mismatches did not inhibit amplification in individual PCRs, in a competitive mixed PCR, the mismatches could result in an amplification bias. Differences in read numbers can also be attributed to sequence characteristics such as homopolymer regions and GC content. Our mock community data demonstrated that using read numbers as proxies for abundance in environmental

samples is problematic, especially in complex samples, and the majority of bias is introduced at the PCR stage.

Because our arthropod mock community consists of single-copy cloned plasmids, we expected to find only one OTU per mock member, allowing the conclusive identification of spurious or chimeric sequences generated during the sequencing process. Some of these chimeras are the result of simple binning errors and others are true chimeras (i.e., hybrid sequences as a result of PCR and sequencing error). A critical component of chimera filtering is having a curated database of reference sequences. We initially attempted to use all available COI sequences in BOLD, but encountered many inconsistencies; thus, we manually curated a subset of those sequences to use for reference sequences. This curated reference database is available at https://amptk.readthedocs.io/en/latest/taxonomy.html. As additional well-documented sequences are added to the database, the ability to identify chimeric sequences will continue to improve, thus enhancing the accuracy of OTU identification in HTS of COI.

Without the use of a mock community, final OTU counts may be greatly inflated because it is difficult to identify spurious OTUs. Among other potential sources, spurious OTUs may arise from PCR- or sequencing-based chimera formation as well as errors generated by clustering algorithms. Using a widely used clustering algorithm (UPARSE; Edgar, 2013) and fine-tuned filtering parameters, our initial OTU estimate for our 46 member single-copy mock community was 70 and thus inflated by at least 52% by the generation of spurious OTUs (data not shown). Through manual inspection of the sequences, most of the spurious OTUs in the mock community were PCR-based chimeras that passed the chimera filter and were not observed in any other sample. Using our mock community as a reference, we were able to assess the efficacy of an alternative OTU picking algorithm, DADA2 (Callahan et al., 2016). Using the DADA2 algorithm followed by 97% UCLUST clustering, we were able to reduce the number of OTUs in our pre-PCR combined mock community from 70 to 43. This method is still imperfect, as one of the OTUs was attributed to sequencing error and one was a chimera, thus reducing the final number to 42. Two of our mock members were lost because they did not sequence well, and an additional two were intra-individual variants of other mock members (*Harmonia axyridis* and *Phalangium opilio*) and clustered with their "sibling" sequences after UCLUST was applied to the DADA2 output. When we used the curated reference database for chimera filtering with UCHIME in combination with the DADA2 algorithm, we were able to remove all but one spurious OTU from our mock community, demonstrating that clustering algorithms can be fine-tuned to minimize spurious OTU generation with the use of single-copy mock communities. Alternative methods for chimera detection and filtering are available and show promise (Alberdi, Aizpurua, Gilbert, & Bohmann, 2018; Zepeda-Mendoza, Bohmann, Baez, & Gilbert, 2016); however, these rely upon separately tagged PCR replicates at the cost of biological replicates, and even when using other chimera detection and filtering methods, we still recommend using a single-copy mock community to validate results.

The use of single-copy cloned plasmid DNA for mock community members is crucial because it removes cryptic sources of biological variation that might otherwise occur within the mock community. For instance, estimates of taxonomic richness may be inflated by intragenomic variability. Intragenomic variability is caused by heteroplasmy and the presence of nuclear mitochondrial pseudogenes (numts), which are pieces of mitochondrial DNA that have been incorporated into the genome (Song et al., 2008). Intragenomic variability is known in commonly used barcoding regions, including ribosomal ITS (Lindner & Banik, 2011; Lindner et al., 2013; Schoch et al., 2012) and mitochondrial COI (Song et al., 2008). The presence of this individual-level variation can lead to the inflation of taxon numbers because intragenomic variants are often misclassified as separate OTUs (Lindner & Banik, 2011; Song et al., 2008). We detected variation in the COI region in *Harmonia axyridis* and *Phalangium opilio* via standard cloning and sequencing, even though a limited number of clones were sequenced (three clones per individual). Sampling more replicates per individual could uncover many such variants and that individual-level variability could significantly inflate diversity estimates in HTS of the COI region. The *H. axyridis* variants only differed by 2.1% (14 of 658 bp), but the *P. opilio* variants were more than three percent different (3.5% or 23 of 659 bp). We included both variants of *H. axyridis* and *P. opilio* in our arthropod mock community to determine whether our bioinformatics pipeline would bin the sequence variants from the same individual into separate inferred sequences. When we applied the DADA2 algorithm without clustering, the variants separated into separate OTUs. After we applied 97% clustering to the resulting DADA2 ASVs, the variants we observed in our single-copy mock community binned together.

4.3 | Validation of the ANML primer pair and mock community

Results from two kinds of samples with known composition validated our primer choice and HTS pipeline. We recovered all taxa included in the known-diet guano samples and also recovered additional OTUs (Table 4). The initial total OTU estimate of the five known-diet samples was 42 based on UPARSE clustering, 31 inferred sequences based on DADA2 without clustering and 10 OTUs based on DADA2 with 97% clustering. Much of the taxonomic reduction in the known-diet samples after using DADA2 with clustering can be attributed to sequence variants of two taxa, *Tenebrio molitor* and Ichneumonidae sp. These two taxa yielded up to 11 and seven inferred sequences per sample with the DADA2 algorithm, respectively, before 97% clustering was applied. However, after clustering was applied, they yielded up to two OTUs per sample. The degree to which these variants represent intra-individual sequence variation, or variants among individuals, cannot be determined here, but offers an interesting topic for future investigation. The estimate with DADA2 with clustering is much closer to the expected richness of five OTUs than other estimates. Several OTUs detected from the known-diet samples were unexpected, but probably real components of the bat diet. Two of these OTUs, *E. takeuchii* and *Agrotis ipsilon*, are likely contaminants in the dietary components

because their larval forms may have been mixed into the *G. mellonella* larvae that comprised the diet. We also detected an ichneumonid parasitoid wasp, which was perhaps parasitizing one or more of the insects in the diet. The unexpected taxa could have been anticipated by sequencing a subsample of the known-dietary components prior to feeding.

HTS successfully recovered the majority of arthropods present in mixed samples from pitfall traps (Table 5). After taking into account probable morphological identification errors, approximately 80% of the taxa identified by morphology were also identified via HTS. Those taxa not detected by HTS may have been missed due to sample degradation or biases in the molecular pipeline such as PCR biases that arose in these complex communities or perhaps these taxa require more specific primers. There were also taxa that were detected with HTS but not detected by morphological identification. These additional taxa may have been consumed by or otherwise associated with the arthropods collected in the traps, misidentified during the morphological identification or may be DNA contamination of the traps or other collection equipment.

5 | CONCLUSION

We demonstrated that the ANML primer pair detects a greater number of arthropod taxa than other frequently used COI primer pairs. The use of HTS read numbers as a measure of abundance in environmental samples is problematic due to biases introduced during both PCR and HTS. These biases may be partially alleviated in the future by non-PCR-based techniques such as shotgun metagenomics and target capture techniques (Dowle, Pochon, Banks, Shearer, & Wood, 2016; Zhou et al., 2013). Failing to use appropriate positive controls for amplicon-based studies can lead to over-estimation of diversity and the persistence of "nonsense taxa." Thus, mock community controls are necessary to parameterize downstream bioinformatics, especially for diversity and community structure-related questions, and we advocate for the inclusion of a spike-in mock control in every HTS run.

AUTHOR CONTRIBUTIONS

M.A.J. and M.T.B. wrote the paper; M.A.J., M.T.B., J.M.P. and D.L.L. designed research; M.A.J., M.T.B., J.M.P., A.K.W. and E.P. performed research; A.Y.K., L.X., J.R.B., C.G. and M.Z.P. contributed samples; M.A.J., M.T.B. and J.M.P. analysed the data; and M.A.J., M.T.B., J.M.P., A.K.W., J.R.B., A.Y.K., C.G., M.Z.P. and D.L.L. edited drafts of the paper.

ORCID

Michelle A. Jusino http://orcid.org/0000-0002-3284-4254
Jonathan M. Palmer http://orcid.org/0000-0003-0929-3658
Jesse R. Barber http://orcid.org/0000-0003-3084-2973
Akito Y. Kawahara http://orcid.org/0000-0002-3724-4610
Claudio Gratton http://orcid.org/0000-0001-6262-9670
Daniel L. Lindner http://orcid.org/0000-0002-2951-4481

REFERENCES

Alberdi, A., Aizpurua, O., Gilbert, M. T., & Bohmann, K. (2018). Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods in Ecology and Evolution*, *9*, 134–147. https://doi.org/10.1111/2041-210X.12849

Bokulich, N. A., & Mills, D. A. (2013). Improved selection of internal transcribed spacer-specific primers enables quantitative, ultra-high-throughput profiling of fungal communities. *Applied and Environmental Microbiology*, *79*, 2519–2526. https://doi.org/10.1128/AEM.03870-12

Bokulich, N. A., Subramanian, S., Faith, J. J., Gevers, D., Gordon, J. I., Knight, R., … Caporaso, J. G. (2013). Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nature Methods*, *10*, 57–59. https://doi.org/10.1038/nmeth.2276

Boyles, J. G., Cryan, P. M., McCracken, G. F., & Kunz, T. H. (2011). Economic importance of bats in agriculture. *Science*, *332*, 41–42. https://doi.org/10.1126/science.1201366

Brandon-Mong, G. J., Hm, G., Sing, K. W., Lee, P. S., Lim, P. E., & Wilson, J. J. (2015). DNA metabarcoding of insects and allies: An evaluation of primers and pipelines. *Bulletin of Entomological Research*, *105*, 717–727. https://doi.org/10.1017/S0007485315000681

Brooks, J. P., Edwards, D. J., Harwich, M. D., Rivera, M. C., Fetweis, J. M., Serrano, M. G., … Strauss, J. F. (2015). The truth about metagenomics: Quantifying and counteracting bias in 16S rRNA studies. *BMC Microbiology*, *15*, 66. https://doi.org/10.1186/s12866-015-0351-6

Bussche, R. A. V. D., Lee, D. N., Judkins, M. E., Dyer, J. E., Thompson, D. M., Stark, R. C., … Fuller, B. (2016). Molecular dietary analysis of the endangered ozark big-eared bat (*Corynorhinus townsendii ingens*). *Acta Chiropterologica*, *18*, 181–191.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*, 581–583. https://doi.org/10.1038/nmeth.3869

Chase, J. M., Kraft, N. J., Smith, K. G., Vellend, M., & Inouye, B. D. (2011). Using null models to disentangle variation in community

dissimilarity from variation in α-diversity. *Ecosphere*, *2*, 1–11. https://doi.org/10.1890/ES10-00117.1

Clare, E. L., Symondson, W. O., Broders, H., Fabianek, F., Fraser, E. E., MacKenzie, A., … Menzies, A. K. (2014). The diet of *Myotis lucifugus* across Canada: Assessing foraging quality and diet variability. *Molecular Ecology*, *23*, 3618–3632.

Clare, E. L., Symondson, W. O., & Fenton, M. B. (2014). An inordinate fondness for beetles? Variation in seasonal dietary preferences of night-roosting big brown bats (*Eptesicus fuscus*). *Molecular Ecology*, *23*, 3633–3647.

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.

Cleveland, C. J., Betke, M., Federico, P., Frank, J. D., Hallam, T. G., Horn, J., … Sansone, C. G. (2006). Economic value of the pest control service provided by Brazilian free-tailed bats in south-central Texas. *Frontiers in Ecology and the Environment*, *4*, 238–243. https://doi.org/10.1890/1540-9295(2006)004[0238:EVOTPC]2.0.CO;2

R Core Team (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Crisol-Martínez, E., Moreno-Moyano, L. T., Wormington, K. R., Brown, P. H., & Stanley, D. (2016). Using next-generation sequencing to contrast the diet and explore pest-reduction services of sympatric bird species in macadamia orchards in Australia. *PloS One*, *11*, e0150159.

D'Amore, R., Ijaz, U. Z., Schirmer, M., Kenny, J. G., Gregory, R., Darby, A. C., … Hall, N. (2016). A comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA community profiling. *BMC Genomics*, *17*, 55. https://doi.org/10.1186/s12864-015-2194-9

Deagle, B. E., Jarman, S. N., Coissac, E., Pompanon, F., & Taberlet, P. (2014). DNA metabarcoding and the cytochrome c oxidase subunit I marker: Not a perfect match. *Biology Letters*, *10*, 20140562. https://doi.org/10.1098/rsbl.2014.0562

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. https://doi.org/10.1111/1755-0998.12103

Dowle, E. J., Pochon, X., Banks, J. C., Shearer, K., & Wood, S. A. (2016). Targeted gene enrichment and high-throughput sequencing for environmental biomonitoring: A case study using freshwater macroinvertebrates. *Molecular Ecology Resources*, *16*, 1240–1254. https://doi.org/10.1111/1755-0998.12488

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, *26*, 2460–2461. https://doi.org/10.1093/bioinformatics/btq461

Edgar, R. C. (2013). UPARSE: Highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, *10*, 996–998. https://doi.org/10.1038/nmeth.2604

Edgar, R. (2016) SINTAX: A simple non-Bayesian taxonomy classifier for 16S and ITS sequences. *bioRxiv*, 074161. doi: 10.1101/074161.

Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase Subunit 1 from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, *3*, 294–299.

Harms-Tuohy, C. A., Schizas, N. V., & Appeldoorn, R. S. (2016). Use of DNA metabarcoding for stomach content analysis in the invasive lionfish *Pterois volitans* in Puerto Rico. *Marine Ecology Progress Series*, *558*, 181–191. https://doi.org/10.3354/meps11738

Hebert, P. D., Cywinska, A., & Ball, S. L. (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B: Biological Sciences* 270, 313–321.

Hebert, P. D., Penton, E. H., Burns, J. M., Janzen, D. H., & Hallwachs, W. (2004). Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 14812–14817. https://doi.org/10.1073/pnas.0406166101

Jedlicka, J. A., Vo, A.-T.-E., & Almeida, R. P. (2016). Molecular scatology and high-throughput sequencing reveal predominately herbivorous insects in the diets of adult and nestling Western Bluebirds (*Sialia mexicana*) in California vineyards. *The Auk*, *134*, 116–127.

Kartzinel, T. R., & Pringle, R. M. (2015). Molecular detection of invertebrate prey in vertebrate diets: Trophic ecology of Caribbean island lizards. *Molecular Ecology Resources*, *15*, 903–914. https://doi.org/10.1111/1755-0998.12366

Krehenwinkel, H., Kennedy, S., Pekár, S., & Gillespie, R. G. (2016). A cost-efficient and simple protocol to enrich prey DNA from extractions of predatory arthropods for large-scale gut content analysis by Illumina sequencing. *Methods in Ecology and Evolution*, *8*, 126–134. https://doi.org/10.1111/2041-210X.12647

Lindahl, B. D., Nilsson, R. H., Tedersoo, L., Abarenkov, K., Carlsen, T., Kjøller, R., … Kauserud, H. (2013). Fungal community analysis by high-throughput sequencing of amplified markers–a user's guide. *New Phytologist*, *199*, 288–299. https://doi.org/10.1111/nph.12243

Lindner, D. L., & Banik, M. T. (2009). Effects of cloning and root-tip size on observations of fungal ITS sequences from *Picea glauca* roots. *Mycologia*, *101*, 157–165.

Lindner, D. L., & Banik, M. T. (2011). Intragenomic variation in the ITS rDNA region obscures phylogenetic relationships and inflates estimates of operational taxonomic units in genus *Laetiporus*. *Mycologia*, *103*, 731–740.

Lindner, D. L., Carlsen, T., Henrik Nilsson, R., Davey, M., Schumacher, T., & Kauserud, H. (2013). Employing 454 amplicon pyrosequencing to reveal intragenomic divergence in the internal transcribed spacer rDNA region in fungi. *Ecology and Evolution*, *3*, 1751–1764. https://doi.org/10.1002/ece3.586

Maine, J. J., & Boyles, J. G. (2015). Bats initiate vital agroecological interactions in corn. *Proceedings of the National Academy of Sciences*, *112*, 12438–12443. https://doi.org/10.1073/pnas.1505413112

Mallott, E., Malhi, R., & Garber, P. (2015). High-throughput sequencing of fecal DNA to identify insects consumed by wild Weddell's saddleback tamarins (Saguinus weddelli, Cebidae, Primates) in Bolivia. *American Journal of Physical Anthropology*, *156*, 474–481.

Nguyen, N. H., Smith, D., Peay, K., & Kennedy, P. (2015). Parsing ecological signal from noise in next generation amplicon sequencing. *New Phytologist*, *205*, 1389–1393. https://doi.org/10.1111/nph.12923

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., … Wagner, H. (2011) *vegan: Community ecology package*. R package version 2.4-3.

Palmer, J. P., Jusino, M. A., Banik, M. T., & Lindner, D. L. (2018). Non-biological synthetic spike-in controls and the AMPtk software pipeline improve mycobiome data. *PeerJ*, *6*, e925. https://doi.org/10.7717/peerj.4925.

Piñol, J., Mir, G., Gomez-Polo, P., & Agustí, N. (2015). Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. *Molecular Ecology Resources*, *15*, 819–830. https://doi.org/10.1111/1755-0998.12355

Pompanon, F., Deagle, B. E., Symondson, W. O., Brown, D. S., Jarman, S. N., & Taberlet, P. (2012). Who is eating what: Diet assessment using next generation sequencing. *Molecular Ecology*, *21*, 1931–1950. https://doi.org/10.1111/j.1365-294X.2011.05403.x

Ratnasingham, S., & Hebert, P. D. (2007). BOLD: The Barcode of Life Data System (http://www/. barcodinglife. org). *Molecular ecology. Notes*, *7*, 355-364.

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: A versatile open source tool for metagenomics. *PeerJ*, *4*, e2584. https://doi.org/10.7717/peerj.2584.

Rydell, J., Bogdanowicz, W., Boonman, A., Pettersson, S., Suchecka, E., & Pomorski, J. J. (2016). Bats may eat diurnal flies that rest on wind

turbines. *Mammalian Biology-Zeitschrift Für Säugetierkunde, 81*, 331–339. https://doi.org/10.1016/j.mambio.2016.01.005

Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., Levesque, C. A., … Miller, A. N. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences, 109*, 6241–6246. https://doi.org/10.1073/pnas.1117018109

Smith, M. A., Woodley, N. E., Janzen, D. H., Hallwachs, W., & Hebert, P. D. (2006). DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera: Tachinidae). *Proceedings of the National Academy of Sciences of the United States of America, 103*, 3657–3662. https://doi.org/10.1073/pnas.0511318103

Song, H., Buhay, J. E., Whiting, M. F., & Crandall, K. A. (2008). Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proceedings of the National Academy of Sciences, 105*, 13486–13491. https://doi.org/10.1073/pnas.0803076105

Trevelline, B. K., Latta, S. C., Marshall, L. C., Nuttle, T., & Porter, B. A. (2016). Molecular analysis of nestling diet in a long-distance Neotropical migrant, the Louisiana Waterthrush (*Parkesia motacilla*). *The Auk, 133*, 415–428.

Vesterinen, E. J., Ruokolainen, L., Wahlberg, N., Peña, C., Roslin, T., Laine, V. N., … Lilley, T. M. (2016). What you need is what you eat? Prey selection by the bat *Myotis daubentonii*. *Molecular Ecology, 25*, 1581–1594.

Williams-Guillén, K., Olimpi, E., Maas, B., Taylor, P. J., & Arlettaz, R. (2016). Bats in the anthropogenic matrix: Challenges and opportunities for the conservation of Chiroptera and their ecosystem services in agricultural landscapes. In: *Bats in the Anthropocene: Conservation of Bats in a Changing World*, pp. 151–186. Springer.

Zeale, M. R., Butlin, R. K., Barker, G. L., Lees, D. C., & Jones, G. (2011). Taxon-specific PCR for DNA barcoding arthropod prey in bat faeces. *Molecular Ecology Resources, 11*, 236–244. https://doi.org/10.1111/j.1755-0998.2010.02920.x

Zepeda-Mendoza, M. L., Bohmann, K., Baez, A. C., & Gilbert, M. T. (2016). DAMe: A toolkit for the initial processing of datasets with PCR replicates of double-tagged amplicons for DNA metabarcoding analyses. *BMC Research Notes, 9*, 255. https://doi.org/10.1186/s13104-016-2064-9

Zhou, X., Li, Y., Liu, S., Yang, Q., Su, X., Zhou, L., … Huang, Q. (2013). Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. *Gigascience, 2*, 4. https://doi.org/10.1186/2047-217X-2-4

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Jusino MA, Banik MT, Palmer JM, et al. An improved method for utilizing high-throughput amplicon sequencing to determine the diets of insectivorous animals. *Mol Ecol Resour*. 2019;19:176–190. https://doi.org/10.1111/1755-0998.12951