PeerJ Preprints NOT PEER-REVIEWED

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

Authors

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

[++] indicates shared first authorship based on equal contributions

* indicates corresponding authors

Affiliations

(1) United States Forest Service, Northern Research Station, Center for Forest Mycology Research, One Gifford Pinchot Drive, Madison, Wisconsin, USA

(2) University of Wisconsin-Madison, Department of Forest and Wildlife Ecology, Madison, Wisconsin, USA

(3) University of Wisconsin-Madison, Department of Entomology, Madison, Wisconsin, USA

(4) McGuire Center for Lepidoptera and Biodiversity, Florida Museum of Natural History, University of Florida, Gainesville, USA

(5) The Xerces Society for Invertebrate Conservation, Portland, Oregon, USA

(6) Department of Biological Sciences, Graduate Program in Ecology, Evolution and Behavior, Boise State University, Boise, Idaho, USA

*Michelle A. Jusino: USFS NRS CFMR, One Gifford Pinchot Drive, Madison, WI, 53726

michellejusino@gmail.com, mjusino@fs.fed.us

*Daniel L. Lindner: USFS NRS CFMR, One Gifford Pinchot Drive, Madison, WI, 53726

dlindner@fs.fed.us, dlindner@wisc.edu

Running title: Improved HTS of insectivore diets

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3184v1 | CC BY 4.0 Open Access | rec: 24 Aug 2017, publ: 24 Aug 2017

Abstract

DNA analysis of predator feces 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 (CO1). 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) CO1 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 of 59 taxa (41%). Furthermore, in an HTS comparison of field-collected samples, the ANML primers detected nearly four-fold 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: arthropod mock community, bat guano, dietary analysis, insectivore, next-generation sequencing, NGS

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3184v1 | CC BY 4.0 Open Access | rec: 24 Aug 2017, publ: 24 Aug 2017

Introduction

High-throughput amplicon sequencing (HTS) has become the preferred 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.* 2014a; Clare *et al.* 2014b; Mallott *et al.* 2015; Rydell *et al.* 2016; Vesterinen *et al.* 2016), birds (Crisol-Martínez *et al.* 2016; Jedlicka *et al.* 2016; Trevelline *et al.* 2016), reptiles (Kartzinel & Pringle 2015), fish (Harms-Tuohy *et al.* 2016), and arthropods (Krehenwinkel *et al.* 2016). Identification of the DNA of dietary components is accomplished by "metabarcoding", which involves extracting DNA from fecal samples, amplifying one or more barcoding loci, preparing DNA libraries, and finally sequencing, bioinformatics, and data analysis. Each of these steps involves decisions 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 *et al.* 2015; Pompanon *et al.* 2012).

The mitochondrial cytochrome oxidase C subunit 1 locus (CO1) is the most frequently used barcoding locus for identifying a wide range of taxonomic groups, including arthropods. Because CO1 has the most extensive reference library for arthropods (BOLD systems, Ratnasignham & Hebert, 2007), it is the most commonly used locus for dietary studies of insectivorous animals (Clarke *et al.* 2014). The entire CO1 barcoding region is about 658 base pairs (bp) and currently too long to be used with most HTS platforms. Therefore it is necessary to sequence shorter regions of the CO1 locus, which has proven challenging due to a lack of conserved priming sites within the CO1 region (Deagle *et al.* 2014). Therefore, novel primer pairs should be tested against as many expected target DNA sequences as possible.

Zeale *et al.* (2011) developed the ZBJ-ArtF1c/ZBJ-ArtR2c (hereafter ZBJ) primer pair for detecting arthropod prey DNA in bat guano by amplifying a 157 bp fragment of the CO1 region. In the initial study, which employed cloning and sequencing rather than HTS, the ZBJ primers

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3184v1 | CC BY 4.0 Open Access | rec: 24 Aug 2017, publ: 24 Aug 2017

amplified 37 taxa from 13 arthropod orders, but did not amplify bat CO1 DNA. The ZBJ primers were designed to target a short fragment in order 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 analyze diets of insectivorous animals, including bats (Bussche *et al.* 2016; Clare *et al.* 2014a; Clare *et al.* 2014b; 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).

The assumptions and parameters commonly employed in HTS environmental DNA analyses have a large impact on the operational taxonomic units (OTUs) that are recovered. Bioinformatics clustering algorithms can influence apparent diversity within a sample, or an entire library of samples, and trimming and filtering parameters can impact the resulting community composition (Deagle *et al.* 2013). A validation or control is needed to 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). Mock communities can be used to examine biases, starting at the sampling step and ending at the bioinformatics and community analysis steps.

Here we used a reference insect community to identify specific amplification biases associated with three commonly used primer pairs, including ZBJ, and two novel primer pairs, LCO1-1490/CO1-CFMRa (hereafter ANML) and LCO1490/CO1-CFMRb (hereafter CFMRb), for the CO1 region (Table 1). To further test primers, 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 CO1 sequences, which can serve as a standard in HTS sequencing and to help parameterize a bioinformatics pipeline. Finally, we validated the accuracy of our system of novel primers, the mock community control, and our bioinformatics pipeline by using guano from bats fed known insect diets and composite samples of morphologically identified arthropods from pitfall traps.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3184v1 | CC BY 4.0 Open Access | rec: 24 Aug 2017, publ: 24 Aug 2017

Methods and Materials

Testing of primer pairs against known insect samples

DNA was extracted from 67 arthropod taxa (Table 2) following the protocol in Lindner & Banik (2009) with modifications for insects (Supporting Information, Appendix S1). Briefly, DNA was extracted from excised leg muscles of larger insects or, for smaller insects, the thorax was punctured and the entire insect was used for extraction. Leg muscles and small insects with punctured thoraxes were placed in 100 µL of filtered cell lysis solution (CLS; Lindner & Banik 2009) and frozen at -20° C, and the extraction proceeded. Following DNA extraction, the effectiveness of the following five primer pairs in amplifying the 67 purified DNAs was evaluated: LCO1490/HCO2198 (Folmer *et al.* 1994; Hebert *et al.* 2003; hereafter CO1 L/H), ZBJ-ArtF1c/ZBJ-ArtR2c (Zeale *et al.* 2011; ZBJ), LCO1-1490/CO1-CFMRa (ANML), LCO1490/CO1-CFMRb (CFMRb), and LepF1/mLepR (Hebert *et al.* 2004; Smith *et al.* 2006; hereafter LEP). The CO1-CFMRa and CO1-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. The CO1-CFMRa and CO1-CFMRb priming sites are located in the CO1 gene approximately 180bp away from the LCO1490 priming site (Supporting Information, Figure S1). A list of the primers used is presented in Table 1.

Amplification of the extracted DNA using all primer pairs, except ZBJ, used the following reagent volumes per 15 L µL reaction: 7.88 µL DNA-free molecular grade water, 3 µL Green GoTaq 5x buffer (Promega), 0.12 µL of 20 mg/mL BSA, 0.3 µL of 10 µM dNTPs, 0.3 µL of each 10 µM primer, 0.1 uL of 5u/µL GoTaq polymerase (Promega), and 3 µL of extracted arthropod template DNA. The ZBJ primer pair was used with two different reagent regimes. One, termed the modified protocol, was the same as above except 1.0 µL of each 10 µM primer was added and the second regime was that described by original authors (Zeale *et al.* 2011). The thermocycler parameters for the CO1L/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 minutes. The LepfF-1/mLepR-5 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 minutes at 110V,

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3184v1 | CC BY 4.0 Open Access | rec: 24 Aug 2017, publ: 24 Aug 2017

stained with ethidium bromide and visualized using UV light. Presence or absence of bands was recorded for each primer pair and DNA combination. To provide reference sequences for each species the LCO1490/HCO2198 PCR products were Sanger sequenced with ABI Prism BigDye (Applied Biosystems) sequencing following the method of Lindner & Banik (2009). The resulting sequences were subjected to an NCBI BLAST search to confirm the identities of the insect species of origin.

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

The arthropod DNA present in three field-collected bat guano samples was analyzed using the ANML and ZBJ primer pairs. DNA was extracted from three *Myotis lucifugus* guano samples from three different locations in southern Wisconsin (all collections were approved by the Wisconsin Department of Natural Resources). One sample containing 100 mg of guano, approximately 10 pellets, was extracted from each site using QIAamp DNA stool Mini Kit following the procedure in Appendix S2 of the Supporting Information. The DNA was then amplified using primers modified for metabarcoding by adding an Ion Torrent Xpress trP1 adapter sequence on the reverse primer and 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 180bp. The size-selected products were then quantified on an Invitrogen Qubit 2.0 Fluorometer and brought to a concentration of 2000 pM 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 (ThermoFisher #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; ThermoFisher) with the Ion PGM Hi-Q Sequencing Kit (ThermoFisher #A25592) according to the manufacturer's protocol for 400 bp sequencing.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3184v1 | CC BY 4.0 Open Access | rec: 24 Aug 2017, publ: 24 Aug 2017

Bioinformatics

HTS data were processed using the 'DADA2' method via the AMPtk pipeline (https://github.com/nextgenusfs/amptk). Briefly, the AMPtk pipeline processes (de-multiplexes) HTS amplicon sequencing reads by: 1) identifying a valid barcode index in each read, 2) identifying forward and reverse primer sequences, 3) trimming barcode and primer sequences, 4) renaming the read based on barcode index, and 5) 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-neighbor, SWARM, etc.) 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 "inferred sequences" output of DADA2 as well as clusters the "inferred sequences" into biologically relevant OTUs using the UCLUST (Edgar 2010) algorithm employed in VSEARCH (Rognes *et al.* 2016). The resulting AMPtk OTU tables can be filtered based on spike-in mock communities (described below). Taxonomy for mtCO1 is assigned in AMPtk using a combination of global sequence alignment, UTAX (http://www.drive5.com/usearch/manual/utax_algo.html), and SINTAX (Edgar 2016) using a CO1 reference database. The current CO1 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://github.com/nextgenusfs/amptk/blob/master/docs/reference_databases.md.

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 LCO1490/HCO2198 primers as described previously. To remove intragenomic variation (Song *et al.* 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 & Banik (2009). Three clones of each arthropod taxon were subsequently Sanger sequenced to verify the presence of the CO1 insert sequence. Two of the

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3184v1 | CC BY 4.0 Open Access | rec: 24 Aug 2017, publ: 24 Aug 2017

cloned arthropods produced cloned sequence variants, and these variants (3 total) were also included in the mock community, bringing our mock community total to 46. 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 1500 pM using DNA-free, molecular grade water. Plasmids were then diluted 1:20 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 2000 pM 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 1500 pM plasmids in equal amounts. The pre-PCR combined mock community was then diluted to a 1:8000 concentration prior to amplification with ANML barcoded primers. The resulting barcoded PCR product was then visualized, size selected, quantified, and brought to 2000 pM as described before. The resulting barcoded PCR products were then prepared and sequenced on an Ion Torrent PGM and data were bioinformatically processed as described above.

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 cinerus*, were fed known diets of *Galleria mellonella*, *Tenebrio molitor* and *Antheraea polyphemus* alone and in combination (Table 4). The bats were fed each known diet for one day, and guano pellets were collected during the following 24 hours (approved by Boise State University Institutional Animal Care and Use Committee 006-AC14-018). We analyzed three known diet combinations from the *E. fuscus* individual and two known diet combinations from the *L. cinerus* 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

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3184v1 | CC BY 4.0 Open Access | rec: 24 Aug 2017, publ: 24 Aug 2017

then visualized, size selected at ≥ 150 bp using Zymo Research Select-A-Size DNA clean and concentrator spin columns, quantified, and brought to 2000 pM as described before.

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 analyzed. 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 United States 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: (1) the excised leg muscles of larger arthropods, and smaller arthropods with open thoraxes were combined and submersed in CLS and vortexed (dissected sample), or (2) the intact arthropods were added to 15mL 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.

Results

Testing of primer pairs against known insect samples

Fifty-eight of the 59 taxa (98%) amplified with the ANML (LCO1490/CO1-CFMRa) and CFMRb (LCO1490/CO1-CFMRb) primer pairs, with both pairs failing to amplify the same carabid beetle (Table 2). Fifty-two of 59 taxa (89%) amplified with the CO1L/H primer pair and 48 of 59 (81%) amplified with the LEP primer pair; the LEP pair amplified 100% of the Lepidopterans and

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3184v1 | CC BY 4.0 Open Access | rec: 24 Aug 2017, publ: 24 Aug 2017

Dipterans tested (Table 2). The ZBJ primer pair amplified 24 of the 59 (41%) taxa tested with the Zeale *et al.* (2011) protocol and 27 of the 59 (46%) taxa using our modified protocol. The ZBJ primer pair successfully amplified at least one representative from each arthropod order tested (Table 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 8 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 6 by ZBJ. The second most often detected family were mosquitoes (Family: Culicidae), with 5 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.

Development and testing of an arthropod mock community

The individual plasmid components of our post-PCR combined mock community generated read counts that ranged from 3740 to 4; the mean was 2119 and standard deviation +/- 799, with 89% (41 out of 46) yielding greater than 1500 reads (Figure 1; supplemental table 1). All mock members in the post-PCR combined community were recovered, although 3 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 2174 and standard deviation of +/- 2238, with 54% (25 of 46) yielding more than 1500 reads. Two of our mock members did not generate any sequences in the pre-PCR combined community and an additional 4 generated final read counts below 100 (range 2 to 39).

Testing of known mixed samples with mock community and our pipeline

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3184v1 | CC BY 4.0 Open Access | rec: 24 Aug 2017, publ: 24 Aug 2017

The results of the known diet HTS samples are summarized in Table 4. We detected DNA from all of the expected dietary components in all 5 of the known diet samples tested. 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 3 of 4 (75%) diet samples that included *Galleria* larvae. Finally, we detected big brown bat (*E. fuscus*) DNA in 2 of the 3 samples from big brown bats, and hoary bat (*L. cinereus*) DNA in both (2 of 2) of the samples from hoary bats (Table 4). These data were processed bioinformatically with DADA2, with and without 97% clustering 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*, 7 from Ichneumonidae, and 3 for *L. cinereus*. After clustering at 97%, we maintained the OTU number for all taxa that had one OTU before clustering, and obtained 2 OTUs for *T. molitor*, 1 OTU for Ichneumonidae, and 2 OTUs for *L. cinereus*.

The results of the pitfall trap samples are summarized based on presence or absence of families in Table 5. There appears to be no significant effect of the method in which the communities were extracted (dissected samples or macerated samples) on the efficiency of taxon recovery. Overall, in 5 samples 37 families 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 missed by HTS, 9 were probably a result of either a morphological or sequence misidentification, with the remaining 9 most likely lost through system bias.

Discussion

Through an amplification test of 5 primer pairs against a taxonomically diverse community of arthropods, we demonstrated that our ANML and CFMRb primer pairs amplified more taxa than previously described primer pairs (CO1L/H, ZBJ, and LEP) in a standard PCR. 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

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3184v1 | CC BY 4.0 Open Access | rec: 24 Aug 2017, publ: 24 Aug 2017

taxa than the ZBJ primers, 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 taxa than the ZBJ pair. We also detected chiropteran (bat) sequences in fecal samples from bats using the ANML primer pair, although the number of chiropteran DNA sequences and OTUs was insignificant compared to the overall number of sequences generated. Thus, the amplification of chiropteran DNA did not significantly impact the recovery of arthropod DNA, a feature that helps confirm the identity of the bat target species, as well as their dietary components. It is likely that the CO1 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) also allow for better taxon delineation compared to the ZBJ primers (157 bp product).

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 *et al.* 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). The prevalence of mosquitoes is usually very low in other molecular studies of bat guano that rely upon the ZBJ primers (Clare *et al.* 2014a; Clare *et al.* 2014b; Gonsalves *et al.* 2013; Rolfe *et al.* 2014), and some have gone as far as to say that mosquitoes are not important prey items for bats (Fenton 2012). Specifically, in our guano samples, the ZBJ pair was only able to detect *Aedes vexans*, while the ANML pair detected *A. vexans* plus four other Culicidae species in the same samples. Thus, the ANML primers allow for

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3184v1 | CC BY 4.0 Open Access | rec: 24 Aug 2017, publ: 24 Aug 2017

342 better estimation of the ecosystem services of bats, and perhaps other insectivores, as
343 predators of mosquitoes and other economically important pest species.

344 ***Single-copy arthropod mock community, sources of unexpected variation and some solutions***

345 While some authors have noted that HTS data are unreliable as a source to measure
346 community member abundance (Piñol *et al.* 2015), many HTS studies of environmental samples
347 continue to use abundance metrics based on read numbers. To test the validity of read number
348 as an estimate of relative abundance, we combined pre- and post-PCR mock communities in
349 equimolar amounts prior to sequencing. We predicted that if the approach is valid, read
350 numbers should be equal across taxa. Instead, even though each member of the mock
351 community amplified well in individual PCRs, we observed a large variation in read numbers for
352 the pre-PCR combined mock community, with some members being absent. In contrast, the
353 post-PCR combined mock was far less variable (Figure 1). The initial PCR introduced a large
354 amount of taxonomic bias by preferentially amplifying some taxa, as inferred from the
355 difference in variability in read numbers between the post and pre-PCR mixes of our arthropod
356 mock community. Sequencing itself also introduced bias resulting in differences in read
357 numbers between the mock members that were combined post-PCR. Some of the variation in
358 read numbers among mock community members was probably induced by mismatches in the
359 priming site, given that some members possessing three or more primer mismatches. While
360 this number of mismatches did not inhibit amplification in individual PCRs, in a competitive
361 mixed PCR the mismatches could result in an amplification bias. Differences in read numbers
362 can also be attributed to sequence characteristics such as homopolymer regions and GC
363 content. Our mock community data demonstrated that using read numbers as proxies for
364 abundance in environmental samples is problematic, especially in complex samples.

365 Because our arthropod mock community consists of single-copy cloned plasmids, we
366 expected to find only one OTU per mock member, allowing the conclusive identification of
367 spurious or chimeric sequences generated during the sequencing process. Some of these
368 chimeras are the result of simple binning errors and others are true chimeras (i.e., hybrid
369 sequences as a result of PCR and sequencing error). A critical component of chimera filtering is
370 having a curated database of reference sequences. We initially attempted to use all available

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3184v1 | CC BY 4.0 Open Access | rec: 24 Aug 2017, publ: 24 Aug 2017

CO1 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://github.com/nextgenusfs/amptk/blob/master/docs/reference_databases.md. 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 CO1.

Without the use of a mock community, final OTU counts may be greatly inflated because it is difficult to identify spurious OTUs. 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. 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.

Estimates of taxonomic richness may also be inflated by intragenomic variability in barcoding regions. Intragenomic variability is known in some of the most commonly used

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3184v1 | CC BY 4.0 Open Access | rec: 24 Aug 2017, publ: 24 Aug 2017

barcoding regions, such as the fungal ITS region (Lindner & Banik 2011; Lindner *et al.* 2013; Schoch *et al.* 2012), as well as the mitochondrial CO1 region (Song *et al.* 2008). Therefore, intragenomic variability could be a common issue with many other loci. 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). Two conditions that can cause this apparent variation in the CO1 locus, specifically, are 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). We detected intra-individual variation in the CO1 region in *Harmonia axyridis* and *Phalangium opilio* via standard cloning and sequencing, even though a limited number of clones were sequenced (i.e., two sequence variants were detected by sequencing only three clones from each of these individuals). Based on these observations from a very limited sampling of 3 clones per individual, it seems likely that individuals harbor many such variants and that individual-level variability could significantly inflate diversity estimates in HTS of the CO1 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). Many of these differences occurred in the fragment amplified by the ANML primers and thus traditional clustering would have identified them as distinct OTUs. Both variants of *H. axyridis* and *P. opilio* were included in our arthropod mock community to determine if 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 inferred sequences, the variants we observed in our single copy mock community binned together. 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.

Validation of the* ANML *primer pair and mock community

We further validated our primers and HTS system using two types of samples with known composition: (1) guano samples from bats fed known diets and (2) samples from insect traps that were identified by morphology. From the guano samples, we recovered all taxa

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3184v1 | CC BY 4.0 Open Access | rec: 24 Aug 2017, publ: 24 Aug 2017

included in the known diets and also recovered additional OTUs (Table 4). The initial total OTU estimate for our 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 7 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 5 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 missed by HTS may have been missed due to 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 missed by morphological identification. These additional taxa may have been 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.

Conclusion

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3184v1 | CC BY 4.0 Open Access | rec: 24 Aug 2017, publ: 24 Aug 2017

We demonstrated that the ANML primer pair detects a greater number of arthropod taxa than other frequently used CO1 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. However, shotgun metagenomics are currently far more expensive than amplicon sequencing and may be cost-prohibitive to most researchers, and target capture has not yet been thoroughly evaluated for community characterization of environmental samples. 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.

Acknowledgements

Funding for this work was provided by the US Forest Service, Northern Research Station and the Agricultural Experiment Station at the University of Wisconsin – Madison via Hatch Formula Funds, and NSF IOS-1121739 to AYK and IOS-1121807 to JRB. We sincerely thank James Skelton for insightful discussion that improved previous versions of this manuscript, J. Paul White and Heather Kaarakka of the Wisconsin Department of Natural Resources for assistance with coordinating field collection of guano samples, Marcelo Ferreira de Melo and Erin Green for assistance with arthropod dissections, and Eli Cinto Meija and Mitch Levenhagen for insect collection from pitfall traps.

References

Bokulich NA, Mills DA (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.

Bokulich NA, Subramanian S, Faith JJ*, et al.* (2013) Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nature methods* **10**, 57-59.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3184v1 | CC BY 4.0 Open Access | rec: 24 Aug 2017, publ: 24 Aug 2017

Boyles JG, Cryan PM, McCracken GF, Kunz TH (2011) Economic importance of bats in agriculture. *Science* **332**, 41-42.

Brandon-Mong G-J, Gan H-M, Sing K-W*, et al.* (2015) DNA metabarcoding of insects and allies: an evaluation of primers and pipelines. *Bulletin of entomological research* **105**, 717-727.

Brooks JP, Edwards DJ, Harwich MD*, et al.* (2015) The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies. *BMC microbiology* **15**, 1.

Bussche RAVD, Lee DN, Judkins ME*, et al.* (2016) Molecular dietary analysis of the endangered ozark big-eared bat (*Corynorhinus townsendii ingens*). *Acta Chiropterologica* **18**, 181-191.

Callahan BJ, McMurdie PJ, Rosen MJ*, et al.* (2016) DADA2: High-resolution sample inference from Illumina amplicon data. *Nature methods* **13**, 581-583.

Clare EL, Symondson WO, Broders H*, et al.* (2014a) The diet of *Myotis lucifugus* across Canada: assessing foraging quality and diet variability. *Molecular ecology* **23**, 3618-3632.

Clare EL, Symondson WO, Fenton MB (2014b) An inordinate fondness for beetles? Variation in seasonal dietary preferences of night-roosting big brown bats (*Eptesicus fuscus*). *Molecular ecology* **23**, 3633-3647.

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

Cleveland CJ, Betke M, Federico P*, et al.* (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.

Crisol-Martínez E, Moreno-Moyano LT, Wormington KR, Brown PH, 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 UZ, Schirmer M*, et al.* (2016) A comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA community profiling. *BMC genomics* **17**, 1.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3184v1 | CC BY 4.0 Open Access | rec: 24 Aug 2017, publ: 24 Aug 2017

513 Deagle BE, Jarman SN, Coissac E, Pompanon F, Taberlet P (2014) DNA metabarcoding and the
514 cytochrome c oxidase subunit I marker: not a perfect match. *Biology letters* **10**,
515 20140562.
516 Deagle BE, Thomas AC, Shaffer AK, Trites AW, Jarman SN (2013) Quantifying sequence
517 proportions in a DNA-based diet study using Ion Torrent amplicon sequencing: which
518 counts count? *Molecular Ecology Resources* **13**, 620-633.
519 Edgar R (2016) SINTAX: a simple non-Bayesian taxonomy classifier for 16S and ITS sequences.
520 *bioRxiv*, 074161. https://doi.org/10.1101/074161
521 Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*
522 **26**, 2460-2461.
523 Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads.
524 *Nature methods* **10**, 996-998.
525 Fenton MB (2012) Bats and white-nose syndrome. *Proceedings of the National Academy of*
526 *Sciences* **109**, 6794-6795.
527 Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of
528 mitochondrial cytochrome c oxidase Subunit 1 from diverse metazoan invertebrates.
529 *Molecular Marine Biology and Biotechnology* **3**, 294-299.
530 Gonsalves L, Bicknell B, Law B, Webb C, Monamy V (2013) Mosquito consumption by
531 insectivorous bats: does size matter? *PloS one* **8**, e77183.
532 Harms-Tuohy CA, Schizas NV, Appeldoorn RS (2016) Use of DNA metabarcoding for stomach
533 content analysis in the invasive lionfish *Pterois volitans* in Puerto Rico. *Marine Ecology*
534 *Progress Series* **558**, 181-191.
535 Hebert PD, Cywinska A, Ball SL (2003) Biological identifications through DNA barcodes.
536 *Proceedings of the Royal Society of London B: Biological Sciences* **270**, 313-321.
537 Hebert PD, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA
538 barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes*
539 *fulgerator*. *Proceedings of the National Academy of Sciences of the United States of*
540 *America* **101**, 14812-14817.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3184v1 | CC BY 4.0 Open Access | rec: 24 Aug 2017, publ: 24 Aug 2017

541 Jedlicka JA, Vo A-TE, Almeida RP (2016) Molecular scatology and high-throughput sequencing
542 reveal predominately herbivorous insects in the diets of adult and nestling Western
543 Bluebirds (*Sialia mexicana*) in California vineyards. *The Auk* **134**, 116-127.
544 Kartzinel TR, Pringle RM (2015) Molecular detection of invertebrate prey in vertebrate diets:
545 trophic ecology of Caribbean island lizards. *Molecular Ecology Resources* **15**, 903-914.
546 Krehenwinkel H, Kennedy S, Pekár S, Gillespie RG (2016) A cost-efficient and simple protocol to
547 enrich prey DNA from extractions of predatory arthropods for large-scale gut content
548 analysis by Illumina sequencing. *Methods in Ecology and Evolution* **8**, 126-134.
549 Lindahl BD, Nilsson RH, Tedersoo L*, et al.* (2013) Fungal community analysis by high-throughput
550 sequencing of amplified markers–a user's guide. *New Phytologist* **199**, 288-299.
551 Lindner DL, Banik MT (2009) Effects of cloning and root-tip size on observations of fungal ITS
552 sequences from *Picea glauca* roots. *Mycologia* **101**, 157-165.
553 Lindner DL, Banik MT (2011) Intragenomic variation in the ITS rDNA region obscures
554 phylogenetic relationships and inflates estimates of operational taxonomic units in
555 genus *Laetiporus*. *Mycologia* **103**, 731-740.
556 Lindner DL, Carlsen T, Henrik Nilsson R*, et al.* (2013) Employing 454 amplicon pyrosequencing to
557 reveal intragenomic divergence in the internal transcribed spacer rDNA region in fungi.
558 *Ecology and evolution* **3**, 1751-1764.
559 Maine JJ, Boyles JG (2015) Bats initiate vital agroecological interactions in corn. *Proceedings of*
560 *the National Academy of Sciences* **112**, 12438-12443.
561 Mallott E, Malhi R, Garber P (2015) High-throughput sequencing of fecal DNA to identify insects
562 consumed by wild Weddell's saddleback tamarins (Saguinus weddelli, Cebidae,
563 Primates) in Bolivia. *American journal of physical anthropology* **156**, 474-481.
564 Nguyen NH, Smith D, Peay K, Kennedy P (2015) Parsing ecological signal from noise in next
565 generation amplicon sequencing. *New Phytologist* **205**, 1389-1393.
566 Piñol J, Mir G, Gomez-Polo P, Agustí N (2015) Universal and blocking primer mismatches limit
567 the use of high-throughput DNA sequencing for the quantitative metabarcoding of
568 arthropods. *Molecular Ecology Resources* **15**, 819-830.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3184v1 | CC BY 4.0 Open Access | rec: 24 Aug 2017, publ: 24 Aug 2017

569 Pompanon F, Deagle BE, Symondson WO*, et al.* (2012) Who is eating what: diet assessment
570 using next generation sequencing. *Molecular ecology* **21**, 1931-1950.
571 Ratnasingham S, Hebert PD (2007) BOLD: The Barcode of Life Data System (http://www/.
572 barcodinglife. org). *Molecular ecology notes* **7**, 355-364.
573 Rognes T, Flouri T, Nichols B, Quince C, Mahé F (2016) VSEARCH: a versatile open source tool
574 for metagenomics. *PeerJ* **4**, e2584.
575 Rolfe AK, Kurta A, Clemans DL (2014) Species-level analysis of diets of two mormoopid bats
576 from Puerto Rico. *Journal of Mammalogy* **95**, 587-596.
577 Rydell J, Bogdanowicz W, Boonman A*, et al.* (2016) Bats may eat diurnal flies that rest on wind
578 turbines. *Mammalian Biology-Zeitschrift für Säugetierkunde* **81**, 331-339.
579 Schoch CL, Seifert KA, Huhndorf S*, et al.* (2012) Nuclear ribosomal internal transcribed spacer
580 (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National*
581 *Academy of Sciences* **109**, 6241-6246.
582 Smith MA, Woodley NE, Janzen DH, Hallwachs W, Hebert PD (2006) DNA barcodes reveal
583 cryptic host-specificity within the presumed polyphagous members of a genus of
584 parasitoid flies (Diptera: Tachinidae). *Proceedings of the National Academy of Sciences*
585 *of the United States of America* **103**, 3657-3662.
586 Song H, Buhay JE, Whiting MF, Crandall KA (2008) Many species in one: DNA barcoding
587 overestimates the number of species when nuclear mitochondrial pseudogenes are
588 coamplified. *Proceedings of the National Academy of Sciences* **105**, 13486-13491.
589 Trevelline BK, Latta SC, Marshall LC, Nuttle T, Porter BA (2016) Molecular analysis of nestling
590 diet in a long-distance Neotropical migrant, the Louisiana Waterthrush (*Parkesia*
591 *motacilla*). *The Auk* **133**, 415-428.
592 Vesterinen EJ, Ruokolainen L, Wahlberg N*, et al.* (2016) What you need is what you eat? Prey
593 selection by the bat *Myotis daubentonii*. *Molecular ecology* **25**, 1581-1594.
594 Williams-Guillén K, Olimpi E, Maas B, Taylor PJ, Arlettaz R (2016) Bats in the anthropogenic
595 matrix: challenges and opportunities for the conservation of Chiroptera and their
596 ecosystem services in agricultural landscapes. In: *Bats in the Anthropocene:*
597 *Conservation of Bats in a Changing World*, pp. 151-186. Springer.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3184v1 | CC BY 4.0 Open Access | rec: 24 Aug 2017, publ: 24 Aug 2017

PeerJ Preprints NOT PEER-REVIEWED

598 Zeale MR, Butlin RK, Barker GL, Lees DC, Jones G (2011) Taxon-specific PCR for DNA barcoding
599 arthropod prey in bat faeces. *Molecular Ecology Resources* **11**, 236-244.
600

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3184v1 | CC BY 4.0 Open Access | rec: 24 Aug 2017, publ: 24 Aug 2017

601 **Tables and Figures**

602

603 **Table 1:** Sequences and references, and primer pair names for the primers tested against

604 known arthropod samples.

605

Primer name	Primer sequence	Reference	Pair name
LCO1490	5'-GGTCAACAAATCATAAAGATATTGG-3'	Folmer *et al.* 1994	CO1 L/H
HCO2198	5'-TAAACTTCAGGGTGACCAAAAAATCA-3'	Folmer *et al.* 2003	
LEPF1	5'-ATTCAACCAATCATAAAGATATTGG-3'	Hebert *et al.* 2004	LEP
mLEPR	5'-CTTGTTCCAGCTCCATTTT-3'	Smith *et al.* 2006	
ZBJ-ArtF1c	5'-AGATATTGGAACWTTATATTTTATTTTTGG-3'	Zeale *et al.* 2011	ZBJ
ZBJ-ArtR2c	5'-WACTAATCAATTWCCAAATCCTCC-3'	Zeale *et al.* 2011	
LCO1490	5'-GGTCAACAAATCATAAAGATATTGG-3'	Folmer *et al.* 1994	ANML
CO1-CFMRa	5'-GGWACTAATCAATTTCCAAATCC-3'	This study	
LCO1490	5'-GGTCAACAAATCATAAAGATATTGG-3'	Folmer *et al.* 1994	CFMRb
CO1-CFMRb	5'-GGNACTAATCAATTHCCAAATCC-3'	This study	

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3184v1 | CC BY 4.0 Open Access | rec: 24 Aug 2017, publ: 24 Aug 2017

PeerJ Preprints NOT PEER-REVIEWED

Table 2: Results from testing the 5 primer pairs listed in Table 1 on known insect samples. Shading and a value of 1 indicate amplification; no shading and a value of zero indicate no amplification. Amplification was attempted on a variety of DNA concentrations for each template DNA sample before assigning a value of zero (no amplification). Arthropod mock community members are indicated with the superscript IM. An asterisk indicates that two different cloned sequence variants of an individual were added to the arthropod mock community. ** Indicates that three different cloned sequence variants of an individual were added to the arthropod mock community.

Order	Family	Identity	ANML	CFMRb	CO1 L/H	ZBJ Zeale *et al.* 2011 protocol	ZBJ modified protocol	LEP
Blattodea	Blattidae	*Periplaneta fuliginosa*[IM]	1	1	1	1	1	1
Blattodea	Ectobiidae	*Supella longipalpa*[IM]	1	1	1	0	0	1
Coleoptera	Cantharidae	*Chauliognathus pennsylvanicus*	1	1	1	1	1	1
Coleoptera	Carabidae	Carabidae sp.	0	0	0	0	0	0
Coleoptera	Cerambycidae	*Tetraopes* sp.	1	1	1	1	1	1
Coleoptera	Chrysomelidae	*Paria fragariae*[IM]	1	1	1	1	1	1
Coleoptera	Coccinellidae	*Harmonia axyridis*[IM]**	1	1	1	0	0	0
Coleoptera	Hydrophilidae	Hydrophilidae sp.[IM]	1	1	1	0	1	1
Coleoptera	Meloidae	*Epicauta* sp.	1	1	1	0	0	1
Coleoptera	Scarabaeidae	*Euphoria fulgida*[IM]	1	1	1	0	0	1
Coleoptera	Tenebrionidae	*Tenebrio molitor*	1	1	1	1	0	1
Coleoptera	unk. Coleoptera	Polyphaga sp.[IM]	1	1	1	0	0	1
Dermaptera	Forficulidae	Forficulidae sp.	1	1	0	0	0	1
Diptera	Anthomyiidae	*Delia platura*[IM]	1	1	1	1	1	1
Diptera	Bombyliidae	*Lepidophora lutea*[IM]	1	1	1	1	1	1
Diptera	Chironomidae	*Dicrotendipes* sp.[IM]	1	1	0	0	0	1
Diptera	Chironomidae	*Procladius* sp.[IM]	1	1	1	1	1	1
Diptera	Culicidae	*Aedes Albopictus*[IM]	1	1	1	0	0	1
Diptera	Culicidae	*Aedes vexans*[IM]	1	1	1	1	1	1
Diptera	Leptoceridae	*Oecetis inconspicua*	1	1	1	0	0	1
Diptera	Tipulidae	*Nephrotoma ferruginea*[IM]	1	1	1	0	1	1
Ephemeroptera	Ephemeridae	*Hexagenia limbata 1*	1	1	1	0	0	1
Ephemeroptera	Ephemeridae	*Hexagenia limbata 2*[IM]	1	1	1	0	0	1
Ephemeroptera	Heptageniidae	*Leucrocuta maculipennis*[IM]	1	1	1	1	1	1
Ephemeroptera	unk. Ephemeropotera	Ephemeroptera sp.	1	1	0	0	0	0
Hemiptera	Aphididae	*Aphis helianthi*[IM]	1	1	1	0	0	0
Hemiptera	Cicadellidae	*Osbornellus auronitens*	1	1	1	0	0	0
Hemiptera	Cicadidae	Cicadidae sp.	1	1	1	0	1	1
Hemiptera	Corixidae	Corixidae sp.	1	1	0	0	0	1
Hemiptera	Corixidae	*Sigara alternata*[IM]	1	1	1	1	1	1
Hemiptera	Pentatomidae	*Acrosternum hilare*[IM]	1	1	1	0	0	0

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3184v1 | CC BY 4.0 Open Access | rec: 24 Aug 2017, publ: 24 Aug 2017

PeerJ Preprints NOT PEER-REVIEWED

Hymenoptera	Apidae	*Apis mellifera*[IM]	1	1	1	0	0	0
Hymenoptera	Crabonidae	*Sphecius convallis*	1	1	1	0	0	1
Hymenoptera	Eucharitidae	Eucharitidae sp.[IM]	1	1	1	1	1	1
Hymenoptera	Formicidae	*Formica fusca*	1	1	1	0	0	0
Hymenoptera	Formicidae	*Formica* sp.[IM]	1	1	1	0	0	0
Hymenoptera	Tenthredinidae	*Empria takeuchii* [IM]	1	1	1	1	1	0
Lepidoptera	Crambidae	*Crambus agitatellus*[IM]	1	1	1	1	0	1
Lepidoptera	Crambidae	*Elophila obliteralis*[IM]	1	1	1	0	1	1
Lepidoptera	Crambidae	*Udea rubigalis*[IM]	1	1	1	1	1	1
Lepidoptera	Depressariidae	*Depressaria pastinacella*[IM]	1	1	1	1	1	1
Lepidoptera	Erebidae	*Hypena scabra*[IM]	1	1	1	1	1	1
Lepidoptera	Erebidae	*Hyphantria cunea*[IM]	1	1	1	0	0	1
Lepidoptera	Erebidae	*Idia aemula*[IM]	1	1	1	1	1	1
Lepidoptera	Erebidae	*Renia factiosalis*[IM]	1	1	1	1	1	1
Lepidoptera	Geometridae	*Haematopis grataria*[IM]	1	1	1	1	1	1
Lepidoptera	Noctuidae	*Agrotis ipsilon*[IM]	1	1	1	1	1	1
Lepidoptera	Tortricidae	*Choristoneura rosaceana*[IM]	1	1	1	1	1	1
Neuroptera	Chrysopidae	*Chrysopa oculata*[IM]	1	1	1	0	0	1
Neuroptera	Mantispidae	Mantispidae sp.[IM]	1	1	1	0	1	1
Opiliones	Phalangiidae	*Phalangium opilio*[IM]*	1	1	1	0	0	1
Orthoptera	Acrididae	*Melanoplus femurrubrum*[IM]	1	1	1	0	0	1
Orthoptera	Tettigoniidae	*Scudderia curvicauda*[IM]	1	1	1	0	0	1
Orthoptera	Tettigoniidae	Tettigoniidae sp.[IM]	1	1	1	1	1	1
Trichoptera	Hydropsychidae	*Potamyia flava*	1	1	0	0	0	1
Trichoptera	Hydroptilidae	*Orthotrichia* sp.[IM]	1	1	1	0	0	1
Trichoptera	Leptoceridae	*Ceraclea maculata*[IM]	1	1	1	1	1	1
Trichoptera	Leptoceridae	*Leptocerus americanus*[IM]	1	1	1	0	0	1
Trichoptera	unk. Trichoptera	Trichoptera sp.	1	1	0	0	0	0
		Negative control	0	0	0	0	0	0
		Total	**58**	**58**	**52**	**24**	**27**	**48**
		% Total	**98.31**	**98.31**	**88.14**	**40.68**	**45.76**	**81.36**

613

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3184v1 | CC BY 4.0 Open Access | rec: 24 Aug 2017, publ: 24 Aug 2017

PeerJ Preprints NOT PEER-REVIEWED

Table 3: Operational Taxonomic Units (OTUs) recovered using high-throughput amplicon sequencing (HTS) and either the ANML primers or the ZBJ primers on 3 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.

ANML	ZBJ	Phylum	Class	Order	Family	Genus	species
1	0	Arthropoda	Arachnida	Araneae	Theridiidae	*Theridion*	*Theridion frondeum*
2	0	Arthropoda	Arachnida	Trombidiformes	Limnesiidae	*Limnesia*	*Limnesia* sp.
1	0	Arthropoda	Arachnida	Trombidiformes			Trombidiformes sp.
1	0	Arthropoda	Arachnida				Arachnida sp.
1	0	Arthropoda	Insecta	Coleoptera	Coccinellidae	*Harmonia*	*Harmonia* sp.
2	0	Arthropoda	Insecta	Coleoptera	Elateridae	*Melanotus*	*Melanotus similis*
1	0	Arthropoda	Insecta	Coleoptera	Hydrophilidae	*Helocombus*	*Helocombus bifidus*
1	0	Arthropoda	Insecta	Coleoptera	Scarabaeidae		Scarabaeidae sp.
1	0	Arthropoda	Insecta	Coleoptera	Tenebrionidae	*Tenebrio*	*Tenebrio* sp.
1	0	Arthropoda	Insecta	Coleoptera			Coleoptera sp.
1	0	Arthropoda	Insecta	Diptera	Bibionidae	*Bibio*	*Bibio* sp.
1	0	Arthropoda	Insecta	Diptera	Ceratopogonidae	*Bezzia*	*Bezzia* sp.
0	2	Arthropoda	Insecta	Diptera	Chaoboridae	*Chaoborus*	*Chaoborus punctipennis*
1	1	Arthropoda	Insecta	Diptera	Chironomidae	*Ablabesmyia*	*Ablabesmyia americana*
1	0	Arthropoda	Insecta	Diptera	Chironomidae	*Ablabesmyia*	*Ablabesmyia annulata*
1	0	Arthropoda	Insecta	Diptera	Chironomidae	*Ablabesmyia*	*Ablabesmyia* sp. 1
1	0	Arthropoda	Insecta	Diptera	Chironomidae	*Ablabesmyia*	*Ablabesmyia* sp. 2
3	0	Arthropoda	Insecta	Diptera	Chironomidae	*Chironomus*	*Chironomus plumosus*
1	1	Arthropoda	Insecta	Diptera	Chironomidae	*Chironomus*	*Chironomus* sp. 1
0	2	Arthropoda	Insecta	Diptera	Chironomidae	*Chironomus*	*Chironomus* sp. 2
1	0	Arthropoda	Insecta	Diptera	Chironomidae	*Coelotanypus*	*Coelotanypus* sp.
1	0	Arthropoda	Insecta	Diptera	Chironomidae	*Conchapelopia*	*Conchapelopia* sp.
1	0	Arthropoda	Insecta	Diptera	Chironomidae	*Cryptochironomus*	*Cryptochironomus* sp. 1
2	0	Arthropoda	Insecta	Diptera	Chironomidae	*Cryptochironomus*	*Cryptochironomus* sp. 2
2	0	Arthropoda	Insecta	Diptera	Chironomidae	*Dicrotendipes*	*Dicrotendipes tritomus*
2	1	Arthropoda	Insecta	Diptera	Chironomidae	*Parachironomus*	*Parachironomus* sp. 1
1	0	Arthropoda	Insecta	Diptera	Chironomidae	*Parachironomus*	*Parachironomus* sp. 2
1	0	Arthropoda	Insecta	Diptera	Chironomidae	*Parachironomus*	*Parachironomus* sp. 3
2	0	Arthropoda	Insecta	Diptera	Chironomidae	*Polypedilum*	*Polypedilum* sp. 1
1	0	Arthropoda	Insecta	Diptera	Chironomidae	*Polypedilum*	*Polypedilum* sp. 2
2	2	Arthropoda	Insecta	Diptera	Chironomidae	*Procladius*	*Procladius* sp. 1
1	0	Arthropoda	Insecta	Diptera	Chironomidae	*Procladius*	*Procladius* sp. 2
1	0	Arthropoda	Insecta	Diptera	Chironomidae	*Procladius*	*Procladius* sp. 3
0	1	Arthropoda	Insecta	Diptera	Chironomidae	*Procladius*	*Procladius* sp. 4
1	0	Arthropoda	Insecta	Diptera	Chironomidae	*Xenochironomus*	*Xenochironomus* sp.
3	0	Arthropoda	Insecta	Diptera	Chironomidae		Chironomidae sp. 1
1	0	Arthropoda	Insecta	Diptera	Chironomidae		Chironomidae sp. 2
2	0	Arthropoda	Insecta	Diptera	Chironomidae		Chironomidae sp. 3
2	0	Arthropoda	Insecta	Diptera	Chironomidae		Chironomidae sp. 4

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3184v1 | CC BY 4.0 Open Access | rec: 24 Aug 2017, publ: 24 Aug 2017

PeerJ Preprints NOT PEER-REVIEWED

0	1	Arthropoda	Insecta	Diptera	Chironomidae		Chironomidae sp. 5
1	0	Arthropoda	Insecta	Diptera	Culicidae	*Aedes*	*Aedes abserratus*
1	0	Arthropoda	Insecta	Diptera	Culicidae	*Aedes*	*Aedes excrucians*
1	0	Arthropoda	Insecta	Diptera	Culicidae	*Aedes*	*Aedes provocans*
1	1	Arthropoda	Insecta	Diptera	Culicidae	*Aedes*	*Aedes vexans*
1	0	Arthropoda	Insecta	Diptera	Culicidae	*Culiseta*	*Culiseta melanura*
1	0	Arthropoda	Insecta	Diptera	Hybotidae	*Platypalpus*	*Platypalpus* sp.
0	1	Arthropoda	Insecta	Diptera	Limoniidae	*Shannonomyia*	*Shannonomyia lenta*
1	0	Arthropoda	Insecta	Diptera	Psychodidae	*Psychoda*	*Psychoda alternata*
1	1	Arthropoda	Insecta	Diptera	Tachinidae		Tachinidae sp.
1	0	Arthropoda	Insecta	Diptera	Tipulidae	*Nephrotoma*	*Nephrotoma ferruginea*
0	1	Arthropoda	Insecta	Diptera	Tipulidae	*Tipula*	*Tipula kennicotti*
1	1	Arthropoda	Insecta	Ephemeroptera	Caenidae	*Caenis*	*Caenis youngi*
1	0	Arthropoda	Insecta	Ephemeroptera	Palingeniidae	*Pentagenia*	*Pentagenia vittigera*
1	0	Arthropoda	Insecta	Ephemeroptera	Siphlonuridae	*Siphlonurus*	*Siphlonurus typicus*
1	0	Arthropoda	Insecta	Hemiptera	Corixidae	*Trichocorixa*	*Trichocorixa borealis*
1	0	Arthropoda	Insecta	Hemiptera	Miridae	*Lygus*	*Lygus lineolaris*
0	1	Arthropoda	Insecta	Lepidoptera	Blastobasidae	*Blastobasis*	*Blastobasis glandulella*
1	0	Arthropoda	Insecta	Lepidoptera	Depressariidae	*Antaeotricha*	*Antaeotricha leucillana*
1	0	Arthropoda	Insecta	Lepidoptera	Tortricidae	*Argyrotaenia*	*Argyrotaenia pinatubana*
1	0	Arthropoda	Insecta	Lepidoptera			Lepidoptera sp.
1	0	Arthropoda	Insecta	Megaloptera	Corydalidae	*Chauliodes*	*Chauliodes rastricornis*
1	0	Arthropoda	Insecta	Trichoptera	Hydroptilidae	*Oxyethira*	*Oxyethira serrata*
0	2	Arthropoda	Insecta				Insecta sp.
3	0	Chordata	Mammalia	Chiroptera	Vespertilionidae	*Myotis*	*Myotis lucifugus*

618

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3184v1 | CC BY 4.0 Open Access | rec: 24 Aug 2017, publ: 24 Aug 2017

PeerJ Preprints NOT PEER-REVIEWED

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

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3184v1 | CC BY 4.0 Open Access | rec: 24 Aug 2017, publ: 24 Aug 2017

625 **Table 5:** Comparison of morphological and HTS family-level identifications of arthropods
626 collected from 5 pitfall traps. Arthropods from traps 1 and 4 were dissected pre-extraction, and
627 arthropods from traps 2, 3, and 5 were macerated pre-extraction. A "+" indicates presence of a
628 family.
629

			Trap 1		Trap 2		Trap 3		Trap 4		Trap 5	
			Key	NGS	Key	NGS	Key	NGS	Key	NGS	Key	NGS
		Total Taxa	**7**	**10**	**17**	**13**	**17**	**19**	**4**	**5**	**10**	**7**
Class	**Order / subclass**	**Family**										
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						+				

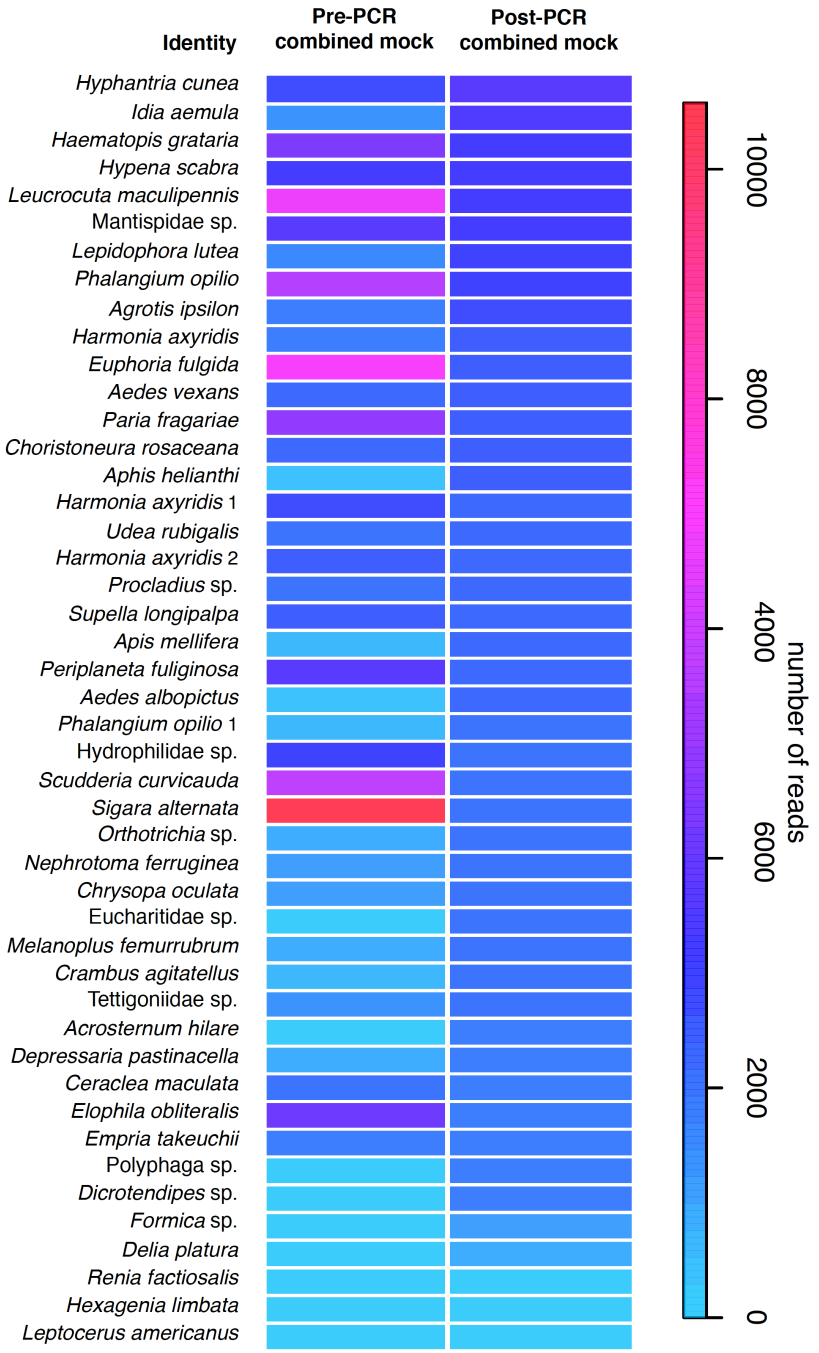
PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3184v1 | CC BY 4.0 Open Access | rec: 24 Aug 2017, publ: 24 Aug 2017

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		+								
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							+	+				

630

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3184v1 | CC BY 4.0 Open Access | rec: 24 Aug 2017, publ: 24 Aug 2017

631 **Figure 1:** Heat map of the high-throughput amplicon sequencing read numbers of the
632 arthropod mock community, equilibrated and combined both pre- and post-PCR. The post-PCR
633 combined mock community was far more even and representative of the equal amounts of
634 DNA added for each mock member than the pre-PCR combined mock community.
635



636

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3184v1 | CC BY 4.0 Open Access | rec: 24 Aug 2017, publ: 24 Aug 2017

637 **Data Accessibility**

638 The corresponding data for this paper was deposited in the NCBI SRA (SRA study SRP102878;
639 BioProject PRJNA380665), and barcoded primer information is provided in the supplemental
640 information.

641

642 **Author Contributions**

643 MAJ, MTB, and JMP wrote the paper; MAJ, MTB, JMP, and DLL designed research; MAJ, MTB,
644 JMP, AKW, and EP performed research; AYK, LX, JRB, CG, and MZP contributed samples; MAJ,
645 MTB, JMP analyzed the data; and MAJ, MTB, JMP, AKW, JRB, AYK, CG, MZP, and DLL edited
646 drafts of the paper.

647

648 **Supporting Information**

649 Appendix S1 – DNA extraction details for the CLS extraction for arthropods
650 Appendix S2 – DNA extraction details for the modified bat guano extraction
651 Table S1 – Sequences for Ion barcoded primers
652 Figure S1 – Mitochondrial cytochrome oxidase C subunit 1 locus (CO1) primer map

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3184v1 | CC BY 4.0 Open Access | rec: 24 Aug 2017, publ: 24 Aug 2017