RESOURCE ARTICLE

WILEY | MOLECULAR ECOLOGY RESOURCES

Carrion fly-derived DNA metabarcoding is an effective tool for mammal surveys: Evidence from a known tropical mammal community

Torrey W. Rodgers[1,2] (iD) | Charles C. Y. Xu[3,4,5] (iD) | Jacalyn Giacalone[6] | Karen M. Kapheim[2,7] | Kristin Saltonstall[2] | Marta Vargas[2] | Douglas W. Yu[3,8] | Panu Somervuo[9] | W. Owen McMillan[2*] | Patrick A. Jansen[2,10*]

[1]Department of Wildland Resources, Utah State University, Logan, UT, USA

[2]Smithsonian Tropical Research Institute, Balboa, Panama

[3]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, China

[4]Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, The Netherlands

[5]Redpath Museum and Department of Biology, McGill University, Montreal, QC, Canada

[6]College of Science and Mathematics, Montclair State University, Montclair, NJ, USA

[7]Department of Biology, Utah State University, Logan, UT, USA

[8]School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich, Norfolk, UK

[9]Department of Biosciences, University of Helsinki, Helsinki, Finland

[10]Department of Environmental Sciences, Wageningen University, Wageningen, The Netherlands

Correspondence
Torrey W. Rodgers, Department of Wildland Resources, Utah State University, Logan, UT, USA.
Email: torrey.w.rodgers@gmail.com

Abstract

Metabarcoding of vertebrate DNA derived from carrion flies has been proposed as a promising tool for biodiversity monitoring. To evaluate its efficacy, we conducted metabarcoding surveys of carrion flies on Barro Colorado Island (BCI), Panama, which has a well-known mammal community, and compared our results against diurnal transect counts and camera trapping. We collected 1,084 flies in 29 sampling days, conducted metabarcoding with mammal-specific (16S) and vertebrate-specific (12S) primers, and sequenced amplicons on Illumina MiSeq. For taxonomic assignment, we compared BLAST with the new program PROTAX, and we found that PROTAX improved species identifications. We detected 20 mammal, four bird, and one lizard species from carrion fly metabarcoding, all but one of which are known from BCI. Fly metabarcoding detected more mammal species than concurrent transect counts (29 sampling days, 13 species) and concurrent camera trapping (84 sampling days, 17 species), and detected 67% of the number of mammal species documented by 8 years of transect counts and camera trapping combined, although fly metabarcoding missed several abundant species. This study demonstrates that carrion fly metabarcoding is a powerful tool for mammal biodiversity surveys and has the potential to detect a broader range of species than more commonly used methods.

KEYWORDS
Barro Colorado Island, biodiversity, camera trapping, environmental DNA, transect counts

*These authors jointly advised this work.

1 | INTRODUCTION

Due to the rapid decline in biodiversity worldwide, there is an urgent need for more efficient techniques to survey and monitor biodiversity. Traditional surveys based on direct observation are costly and time-consuming, detection rates can vary across observers, and rare or cryptic species are often overlooked. Camera trapping has emerged as a more efficient survey method, especially for large-bodied vertebrates (Beaudrot et al., 2016), but it has limited ability to detect small-bodied, arboreal and volant species. Moreover, camera traps are expensive for use at large scales and are subject to damage and theft. An efficient technique capable of detecting a wide range of vertebrate species with limited bias would greatly increase our ability to monitor vertebrate diversity, and to gauge the effectiveness of conservation measures such as nature reserves.

One potential method has emerged from the field of environmental DNA (eDNA) sampling, which uses trace amounts of DNA from the environment for species detection (Bohmann et al., 2014). This approach, particularly when used in conjunction with next-generation metabarcode sequencing, shows great potential for efficient monitoring of biodiversity (Ji et al., 2013; Taberlet, Coissac, Hajibabaei, & Rieseberg, 2012; Yu et al., 2012). To date, the majority of vertebrate eDNA research has focused on aquatic species because eDNA is easy to collect from water (Thomsen et al., 2012). Collecting eDNA from terrestrial vertebrates is more difficult. Researchers have used soil (Andersen et al., 2012), browsed twigs (Nichols, Konigsson, Danell, & Spong, 2012), prey carcasses (Wheat, Allen, Miller, Wilmers, & Levi, 2016) and drinking water (Rodgers & Mock, 2015) for eDNA detection of terrestrial species, but these approaches have limited scope. So far, the most promising approach for use across a broad range of species is mass trapping and metabarcoding of invertebrates that feed on vertebrates (Bohmann, Schnell, & Gilbert, 2013; Calvignac-Spencer, Leendertz, Gilbert, & Schubert, 2013). Such invertebrate "samplers" tested to date include leeches (Schnell et al., 2015), mosquitos (Logue et al., 2016), dung beetles (Gillett, Johnson, Barr, & Hulcr, 2016) and carrion flies (Calvignac-Spencer, Merkel et al., 2013; Lee, Gan, Clements, & Wilson, 2016; Lee, Sing, & Wilson, 2015; Schubert et al., 2014). We focus on the latter.

Carrion flies are species from the families Calliphoridae and Sarcophagidae that feed and oviposit upon dead animals, open wounds and faeces. When carrion flies feed, they ingest vertebrate DNA. Carrion flies are ideal candidates for eDNA surveys because they are easy to trap, are ubiquitous worldwide and are believed to feed opportunistically on vertebrates of all sizes, including terrestrial species, volant species and species occupying the forest canopy. In a remarkable study, Calvignac-Spencer, Merkel et al. (2013) used Sanger sequencing to detect 20 different mammal species from just 115 flies collected in Côte d'Ivoire and Madagascar. They also used Roche GS FLX ("454") pyrosequencing to metabarcode a subset of samples and were able to detect the majority of species that had been detected with Sanger sequencing, plus several others.

Although carrion fly metabarcoding is clearly a promising method, there is still a need to quantify its effectiveness relative to conventional methods. There is also a need to test new methods for assigning taxonomies to sequence data. This step may be particularly error prone in eDNA studies, because target amplicons for eDNA are short due to the need to amplify degraded DNA and thus have lower taxonomic information content. Also, because reference sequence databases are incomplete, taxonomic assignment software has a bias towards overconfidence, meaning that an operational taxonomic unit (OTU) sequence may be assigned to a species that is in the database and fairly similar, when the correct species is not in the database (Somervuo et al., 2017). Correct species-level identification is especially important when applied to vertebrates, as incorrect assignment of an OTU to an endangered species, as opposed to a less endangered congener (or vice versa), can have a large impact on conservation decision making.

With these goals in mind, we conducted a field test on Barro Colorado Island (BCI) in Panama, where the vertebrate fauna is well documented, particularly for mammals. We collected carrion flies for metabarcoding and compared results with data sets from annual, diurnal transect counts (distance sampling) and semi-continuous camera trapping of the mammal community for 8 years leading up to, and concurrent with, fly collection. In addition, we compared taxonomic assignments between the most commonly used method, BLAST (Altschul, Gish, Miller, Myers, & Lipman, 1990) and the new method PROTAX (Somervuo et al., 2017), which uniquely takes into account the possibility that an OTU sequence belongs to a species that is not in the reference database, thus avoiding overconfident assignments. We find that metabarcoding of carrion flies is an effective, but imperfect, method for surveying mammal communities, and PROTAX substantially outperformed BLAST for taxonomic assignment.

2 | MATERIALS AND METHODS

2.1 | Study site

Field work was conducted on Barro Colorado Island, a 1,560-ha island in the Panama Canal waterway (Figure 1). BCI (9°10′N, 79°51′W) sits within Gatun Lake, an artificial body of water created in 1912 by the damming of the Chagres River to create the Panama Canal, and is part of the protected 54-km^2 Barro Colorado Nature Monument. BCI has 108 known mammal species including 74 bats, and 34 nonvolant species, (Glanz, 1982 and current expert information). However, some of these species, such as jaguar (*Panthera onca*), puma (*Puma concolor*) and jaguarundi (*Herpailurus yagouaroundi*), are only infrequent visitors to the island.

2.2 | Fly collection

Flies were collected between 10 February 2015 and 5 May 2015 in three trapping sessions totalling 620 fly-trap days. All flies col-
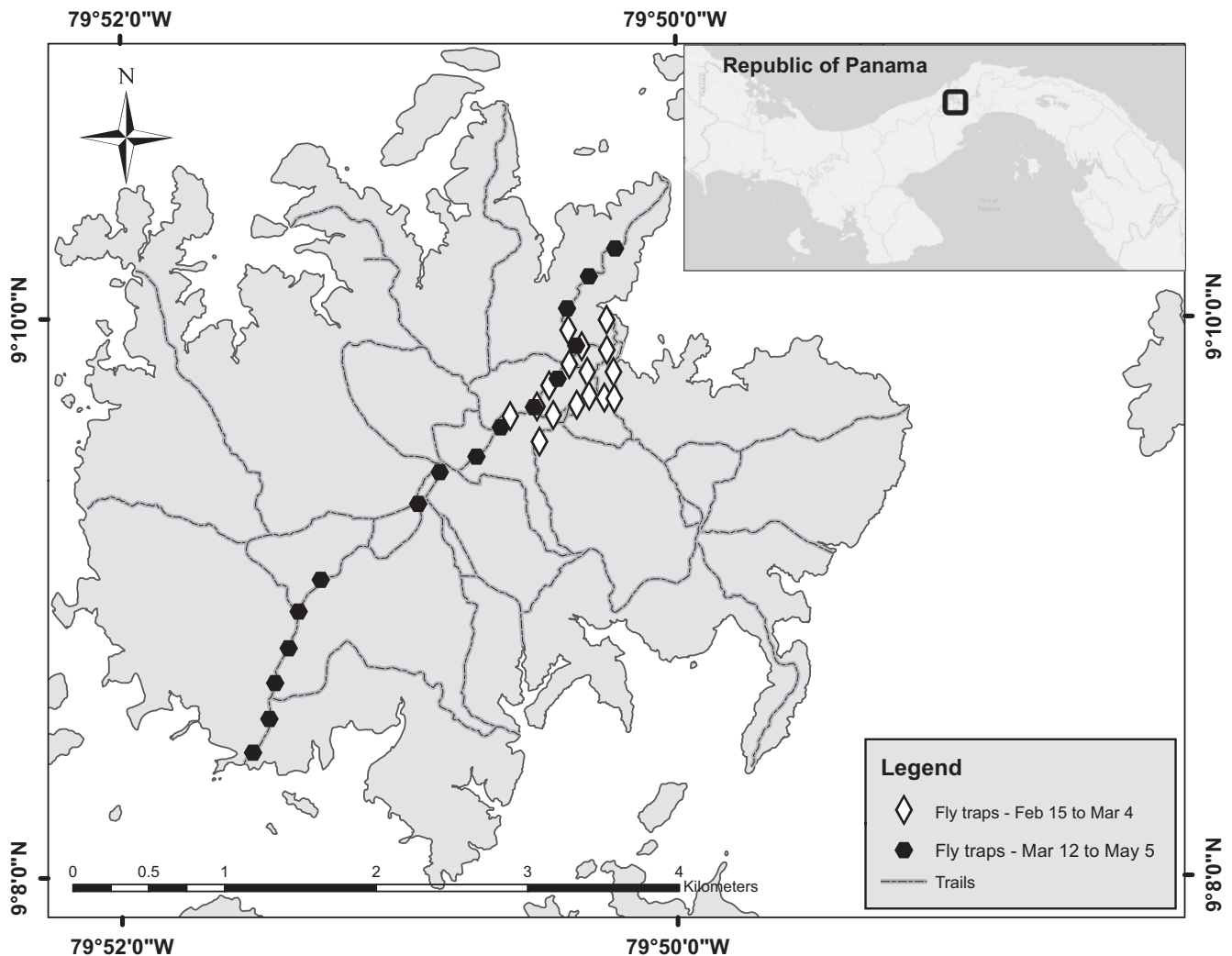
FIGURE 1 Location of carrion fly trapping on Barro Colorado Island, Panama

lected were used in the metabarcoding analysis. First, flies were collected from 10 February to 14 February using a variety of trap types within 1,000 m of the laboratories on BCI, to determine the best methods for sampling. Sampling methods tested included netting flies above covered bowls of pork, and the trap types described in Calvignac-Spencer, Merkel et al. (2013). Based on results from this testing, we used a trap type modified from www.b lowflies.net/collecting.htm for all subsequent trapping. Traps were baited with raw pork hung in a cup surrounded by fine cloth netting to keep flies from landing directly on the bait, and bait was replaced every 2–3 days. From 15 February to 23 February, flies were then collected from traps placed along trails at 16 trap locations, with one trap per location, spaced approximately 200 m apart in a nonuniform grid (Figure 1). Second, from 28 February to 4 March, traps were placed along trails at the same 16 trap locations as session 1, but with two traps per location. Third, from 12 March 2015 to 5 May 2015, flies were collected from traps placed in a transect crossing the island with 16 trap locations, each containing two traps, placed every 250 m along a trial in a roughly straight line (Figure 1), for 10 sampling days. In all three collection efforts,

flies were removed from traps once or twice daily and placed in a −40°C freezer within 2 hr of collection, because DNA degradation causes detection success to decline 24 hr after fly feeding (Lee et al., 2015).

2.3 | Library preparation and sequencing

DNA was extracted from flies using the GeneMATRIX Stool DNA Purification Kit (Roboklon, Berlin, Germany). To reduce extraction costs, up to 16 flies were pooled for each extraction. Flies were first cut into several pieces with sterile scissors and placed into 2.5 ml Polypropylene 96 Deep Well Plates along with stainless steel, 5/32" grinding balls (OPS Diagnostics, Lebanon, NJ, USA). Up to four flies and 100 μl of lysis buffer per fly were added to each well. Plates were shaken on a TissueLyser II (Qiagen, Germantown, MD, USA) until fly tissue was homogenized in the lysis buffer. An equal volume of the resulting homogenate per fly was then pooled in a total volume of 160 μl and added to the extraction kit bead tube. Extractions then proceeded following manufacturer's recommendations. For flies collected in the first trapping session, each extraction included 16

flies. For flies collected in the second and third trapping sessions, each extraction contained all flies collected at the same trap location on the same day, up to a maximum of 16 flies. To examine whether pooling had an effect on species detection, 24 flies were also homogenized and extracted individually, and 20 µl of homogenate from each of these same 24 flies was also pooled and extracted in two samples containing 12 flies each. Single-fly and pooled samples were all sequenced at the same depth. All extractions included blank controls that were included in the PCR step to test for contamination.

PCR was performed on all samples using a mammal-specific primer set (16Smam1, 16Smam2; Boessenkool et al., 2012) targeting 130–138 bp (including primers) of the mitochondrial 16S rRNA locus, and a pan-vertebrate primer set (12SV5F, 12SV5F; Riaz et al., 2011) targeting 140–143 bp (including primers) of the mitochondrial 12S locus. Reactions also included human blocking primers (16Smam_blkhum3; Boessenkool et al., 2012; 12S_V5_blkhum; Calvignac-Spencer, Merkel et al., 2013) and *Sus* blocking primers (16Smam_blkpig, 12S_V5_blkpig; Calvignac-Spencer, Merkel et al., 2013) to decrease competition from contaminating DNA from pork bait. Primers also included a 5' addition of a 33 bp Illumina-specific sequence for addition of adapters in a second round of PCR. A minimum of two PCR replicates per sample were performed in 10 µl volumes. For 16S, reactions included 0.2 uM of each primer, 1 µM of human blocking primer (5x) 4 µM of *Sus* blocking primer (20X), 200 µM dNTP, 4 mM MgCl2, 1X PCR buffer, 1.25 U Platinum® Taq polymerase (Invitrogen) and 3 µl of template DNA. For 12S, volumes were the same, except MgCl$_2$ was reduced to 2.5 mM. Cycling conditions were 10 min at 95°C, followed by 42 cycles (16S) or 47 cycles (12S) of 30 s at 95°C, 30 s at 64°C, and 1 min at 72°C, with a final extension of 10 min at 72°C. Following the initial PCR, sample replicates were pooled, and a second PCR was conducted to add Illumina flow cell binding sequences and unique 8-bp sample-specific indexes to each end of each amplicon. Reactions included 1 µl each of the forward and reverse Illumina index primers, 2.5 µl of 10X QTaq buffer, 1.5 mM MgCl$_2$ and 1.25 units of Qiagen Qtaq. Cycling conditions were 3 min at 94°C, followed by six cycles of 45 s at 94°C, 60 s at 50°C and 1 min at 72°C, with a final extension of 10 min at 72°C. All PCR reactions were prepared in a UV-sterilized laminar flow hood and included no-template negative controls to test for contamination. None of the extraction negative controls and no-template PCR negative controls showed any sign of amplification when run on an agarose gel, so they were not included in sequencing libraries. Samples were purified and normalized using SequalPrep™ Normalization Plates (Thermo-Fisher Scientific, Waltham, MA, USA). All samples were pooled in equimolar concentration, and pools were concentrated using standard Ampure XP beads (Beckman Coulter, Indianapolis, IN, USA). The concentrated pool was quantified using a Qubit fluorometer and quality checked using a Bioanalyzer High Sensitivity DNA Kit (Agilent, Santa Clara, CA, USA). The library was sequenced on an Illumina MiSeq as 30% of a 500 cycle 2 × 250 Paired-end sequencing run.

2.4 | Sequence filtering

After demultiplexing, paired ends were merged for each sample using PEAR v. 0.9.6 (Zhang, Kobert, Flouri, & Stamatakis, 2014) with default parameters, except minimum overlap (-v) = 100, minimum length (-n) = 100 and quality score threshold (-q) = 15. We then separated 12S and 16S sequences and removed primers and remaining adapters with CUTADAPT v. 1.3 (Martin, 2011), keeping sequences with a minimum length (-m) of 10. Sequences had been previously trimmed to minimum 100. Mean ± 1 *SE* length after trimming was 109.78 ± 0.13 for 12S and 93.70 ± 0.08 for 16s. Sequences were subsequently pooled and filtered for chimeras with *the identify_chimeric_seqs.py* script and the *usearch61* method within the QIIME environment (Caporaso et al., 2010; Edgar, 2010). We used the *pick_de_novo_otus.py* script with the *uclust* method to cluster sequences into OTUs (Operational Taxonomic Units) at 97% similarity, create a table of OTU frequency in each sample, and pick representative sequences (Caporaso et al., 2010; Edgar, 2010). We removed OTUs with fewer than 20 sequences using the *filter_otus_from_otu_table.py* script (Caporaso et al., 2010). A threshold of 20 sequence reads was chosen because all OTUs with fewer than 20 reads had no similarity to any vertebrate species and were thus likely artefactual. To remove OTUs from contaminants such as the pork bait, human DNA and bacteria, we conducted an initial BLAST search and removed all OTUs with a top hit for *Sus*, *Homo* or bacteria. For 12S, we used this initial BLAST search to separate OTUs into mammal, bird and reptile groups for downstream taxonomic assignment.

2.5 | Taxonomic assignment

For mammalian taxonomic assignment, we compared two methods: the most commonly used method BLAST (Altschul et al., 1990) and the new probabilistic taxonomic placement method PROTAX (Somervuo et al., 2017). For nonmammals, only BLAST was used. BLAST searches were conducted against the National Center for Biotechnology Information (NCBI) nonredundant database with default settings except for *e*-value ≤1e-6, and per cent identity ≥95%. We processed the BLAST output in two ways for each OTU: selecting the top hit, and using the lowest common ancestor (LCA) algorithm in MEGAN 5.11.3 (Huson, Auch, Qi, & Schuster, 2007; default settings except top per cent = 5), a software program designed to nonsubjectively "choose" the appropriate taxonomic assignment among multiple BLAST hits.

For PROTAX, we generated probabilities of taxonomic placement for each OTU at four taxonomic ranks (order, family, genus and species) (Somervuo, Koskela, Pennanen, Henrik Nilsson, & Ovaskainen, 2016). PROTAX is a statistical wrapper that processes the output of one or more other taxonomic assignment methods and takes into account the uncertainty of taxonomic assignment contributed by species that do not have reference sequences. Briefly, this is achieved by training a model against a reference data set that includes all available reference sequences for the taxon of interest, plus a full Linnaean taxonomy for that taxon which includes all named species, including those that do not have reference

sequences. By taking into account the taxonomic information content of both the gene sequence being assigned and the reference sequence database, PROTAX removes the inherent "over-assignment" bias of other taxonomic assignment software. This makes PROTAX especially suited for markers that have highly incomplete reference databases, such as 16S and 12S.

For the reference sequence databases, we downloaded all available mammalian mitochondrial 16S and 12S sequences from Gen-Bank, randomly truncated each species to a maximum of 10 sequences via a Fisher–Yates shuffle, removed ambiguous bases and used ecoPCR within OBITOOLS 1.2.6 (Boyer et al., 2016; Ficetola et al., 2010) to extract the amplicon regions used in this study (Clark, Karsch-Mizrachi, Lipman, Ostell, & Sayers, 2016). This resulted in 5,243 16S sequences representing 1,733 species and 5,395 12S sequences representing 1,935 species. For the taxonomic database, we downloaded full ranks for the 6,711 species in the NCBI Mammalia taxonomy. We are aware that the NCBI taxonomy is not complete for the Mammalia, but the important benefit is that the names are consistent between the sequence and taxonomic databases, and even this incomplete Mammalia taxonomy has ca. 3.8X and 3.4X the number of species as represented in the 16S and 12S reference sequence databases, respectively, which allows us to contrast PROTAX with BLAST.

To start, all OTU representative sequences were pairwise compared to the reference sequences using LAST version 744 (Kielbasa, Wan, Sato, Horton, & Frith, 2011), and the PROTAX model used the maximum and second-best similarities. We parameterized three PROTAX models, one unweighted model in which all mammal species were given an equal prior weight of being in the sampling location, and two weighted models in which species independently known to be present in Panama (Panama-weighted) or BCI (BCI-weighted) were given a 90% prior probability of being in the sampling location, with all other species being given a prior probability of 10%. The accuracy and bias of the trained PROTAX model at each taxonomic rank were estimated by plotting the cumulative predicted probabilities against the cumulative number of cases in which the outcome with the highest probability was correct when training the model.

For all methods (BLAST top hit, BLAST plus MEGAN, and PROTAX unweighted and weighted), we estimated accuracy by calculating the percentage of OTUs assigned to a genus or species known from BCI (rate of correct assignment) and the percentage of OTUs assigned to a genus or species known to not be present on BCI (rate of clear false positives). We also counted the number of mammal species known from BCI identified with each method. Given that we only used wild-caught flies, we cannot directly test the probability that an OTU was assigned incorrectly to a species known from BCI. However, the PROTAX output does give us an estimated probability of correct assignment for each OTU at each rank.

2.6 | Transect counts and camera trapping

We compared results from fly metabarcoding with results of two traditional survey methods. First, diurnal mammal transect counts were carried out yearly from 2008 to 2015 during January and/or February of each year. Transects covered all trails on BCI (Figure 1) each year, with a mean distance of 120 km walked per year. An average walking rate of 1 km/hr was maintained. Record of each mammal sighting included date, time, location, species and number of individuals. All censuses were conducted between 06:40 and 12:00 and were thus unlikely to detect nocturnal animals. In 2015, the year fly sampling was conducted, 151.1 km was surveyed from 24 January to 22 February. Transect counts started before, but overlapped with fly and camera trap sampling; however, the presence of mammals on BCI does not vary throughout the year, so this slight difference in timing should not affect the methods comparison.

Second, camera traps were operated continuously across BCI from 2008 to 2015. A mean of 25 (range 19–34) trail cameras (PC900 and RC55 Reconyx Inc., Holmen, Wisconsin) were mounted at knee height, with a spacing of 500–1,000 m. Thus, they were unlikely to capture arboreal or volant species. Additional cameras were deployed for shorter periods. Cameras were checked at least once every 6–7 months and replaced or repaired if no longer functioning. This effort resulted in a total of 39,151 total camera days. For the period concurrent with fly collection, February 10 2015 to May 5 2015, 26 cameras were active for the entire period, resulting in a total effort of 1,967 camera days.

2.7 | Methods comparison

We compared the total number of mammal species detected by fly metabarcoding with the total number of species detected by concurrent (2015) and long-term (2008–2015) transect counts and camera trapping. In addition, we fit species accumulation curves for each of the 2015 surveys using the *specaccum* function in the R package VEGAN 2.4-1 (Oksanen et al., 2016) with *method* = random and 10,000 permutations.

3 | RESULTS

3.1 | Transect counts and camera trapping

In the 2015 transect counts, we detected 13 mammal species, whereas in 2008–2015 transect counts, we detected 17 mammal species. With concurrent 2015 camera trapping, we detected 17 mammal species, whereas in 2008–2015 camera trapping, we detected 26 mammal species. Species detected by the two methods overlapped only partially, so that the 2015 data sets combined detected a total of 22 mammal species, and the 2008–2015 data sets combined detected a total of 30 mammal species (Table 2).

3.2 | Fly metabarcoding

3.2.1 | Fly collection, extraction and sequencing

A total of 1084 flies were collected and pooled into 102 pooled samples and 24 single-fly samples. We obtained 2,780,574 initial

reads, which were reduced to 2,288,009 after quality filtering (1,504,440 16S; 783,569 12S). OTU clustering, removal of OTUs with fewer than 20 reads, and removal of OTUs assigned to human, *Sus* (pork bait), and bacteria resulted in a final set of 54 OTUs for 16S and 63 OTUs for 12S (49 mammal, 12 bird and two reptile OTUs; OTU sequences in Appendix S1). After quality filtering and removal of human, *Sus* and bacterial reads, mean number of reads per OTU was 8,401 (Range = 20–336,679; *SD* = 46,578) for 16S and 4,585 (Range = 20–78,841; *SD* = 13,029) for 12S, and mean number of reads per sample was 2,525 (Range = 1–14,716; *SD* = 3,098) for 16S and 2,218 (Range = 1–25,575; *SD* = 4,582) for 12S for pooled samples. For single-fly samples, mean number of

reads per sample was 3,744 (Range = 1–11,173; *SD* = 3,195) for 16S and 87 (Range = 0–2,069, *SD* = 422) for 12S.

3.2.2 | Mammal species detection

The number of mammal species detected differed between assignment methods (Figure 2a). BCI-weighted PROTAX resulted in the detection of 20 total mammal species (Table 1), the most of any method (16 species from 53 OTUs with 16S, 13 species from 37 OTUs with 12S and nine species with both markers). Assignment of multiple OTUs to the same species is expected in metabarcoding as the sequence-clustering step typically applies a single similarity
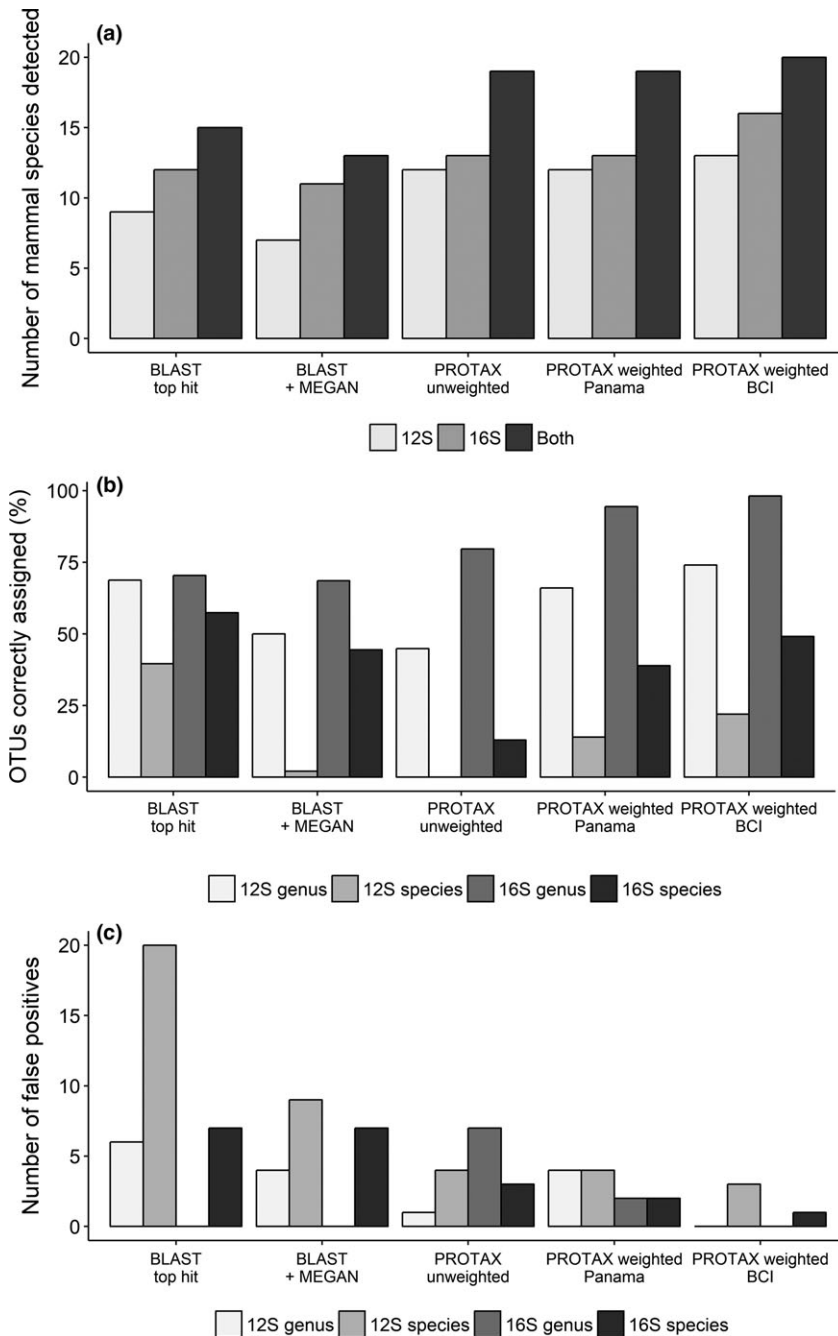

FIGURE 2 Comparison of taxonomic placement methods for assigning mammal OTUs to species from metabarcoding of carrion flies collected on Barro Colorado Island (BCI). (a) Number of mammal species detected from BCI using each taxonomic method. (b) Percentages of OTUs assigned to genera or species known to occur on BCI. (c) Numbers of OTUs assigned to genera or species known not to exist on BCI (clear false positives)

TABLE 1 Twenty mammal species detected from metabarcoding of carrion flies on Barro Colorado Island Panama in 2015, along with PROTAX (BCI-weighted) estimated probabilities of correct assignment at genus and species rank, and the percentage of fly pool samples that each species was detected from

Species name	Common name	Genus prob.	Species prob.	% of samples
Alouatta palliata	Mantled howler monkey	.996	.996	98
Ateles geoffroyi	Geoffroy's spider monkey	.983	.983	85
Tamandua Mexicana	Northern tamandua	1.000	.997	29
Cebus capucinus	White-faced capuchin	.997	.979	26
Caluromys (derbianus)[a]	Derby's woolly opossum	.993	NA	22
Saguinus (geoffroyi)[a]	Geoffroy's tamarin	.400	NA	21
Sciurus granatensis	Red-tailed squirrel	.994	.994	21
Proechimys (semispinosus)[a]	Tome's spiny-rat	.457	NA	20
Marmosa robinsoni	Robinson's mouse opossum	.982	.98	19
Procyon (cancrivorus)[a]	Crab-eating raccoon	.934	NA	16
Potos flavus	Kinkajou	.979	.978	11
Didelphis marsupialis	Common opossum	.984	.957	11
Pecari tajacu	Collared peccary	.979	.978	6
Leopardus (pardalis)[a]	Ocelot	.817	NA	3
Pteronotus (gymnonotus or parnellii)[a]	Big naked-backed bat or Parnell's mustached bat	.230	NA	3
Canis (latrans or familiaris)[a]	Domestic dog or coyote	.942	NA	1
Bradypus variegatus	Brown-throated sloth	1.000	.998	1
Choloepus hoffmanni	Hoffmann's two-toed sloth	1.000	.999	1
Dasyprocta punctata	Agouti	.958	.924	1
Tonatia (saurophila)[a]	Stripe-headed round-eared bat	.894	NA	1

[a]PROTAX assignment only at genus level, thus species probabilities were not estimated. Species names in parentheses are the species of the identified genus that occurs on BCI.

threshold across all OTUs, which can result in species splitting. Fifteen of these 20 species had PROTAX probabilities of >0.9, but several had somewhat low PROTAX probabilities despite their known presence on BCI (Table 1). Panama-weighted and unweighted PROTAX resulted in detection of 19 total mammal species (13 with 16S and 12 with 12S).

BLAST top hit and BLAST plus MEGAN both resulted in detection of fewer mammal species (15 and 13 respectively). The percentage of OTUs assigned to a species known from BCI was higher for PROTAX than for BLAST (Figure 2b). Rates of mammal OTUs assigned to species not present on BCI (clear false positives) were generally higher for BLAST than for PROTAX. Some clear false positives were still present with all assignment methods, although no clear false positives at the genus level occurred with BCI-weighted PROTAX (Figure 2c).

3.2.3 | Effect of fly pooling

From the 24 flies that were extracted both singly and in pools, four total mammal species were detected from the pooled samples, whereas 10 total mammal species were detected from the same 24 flies when extracted singly. From the single-fly extractions, a range of 1–4 species were detected from each fly.

3.2.4 | Nonmammal species detection

The 12S primers also detected several birds and one lizard. We were only able to assign one OTU to the species level, wattled jacana (*Jacana jacana*). Two OTUs were assigned to birds at the genus level, one to an antshrike species (genus *Thamnophilus*, likely *atrinucha* or *doliatus*; both known from BCI), and one to a trogon species (genus *Trogon*, of which five species are known from BCI). Several OTUs were assigned the family Anatidae; however, these were a 100% match to many species within that family, and six species from Anatidae are known from BCI. We also assigned one reptile OTU to the whiptail lizards (family Teiidae) which is likely either *Ameiva festiva* or *Ameiva leptophrys*, the two Teiidae species known from BCI.

3.3 | Methods comparison

In 2015, we detected a greater number of mammal species in 29 days of carrion fly metabarcoding than were detected by either 29 days of transect counts or 84 calendar days of camera trapping carried out concurrently (metabarcoding = 20 species; camera trapping = 17 species, transect counts = 13 species, Figure 3). We detected more mammal species with fly metabarcoding in 2015 than with 8 years of transect counts from 2008 to 2015, but fewer than with 8 years of
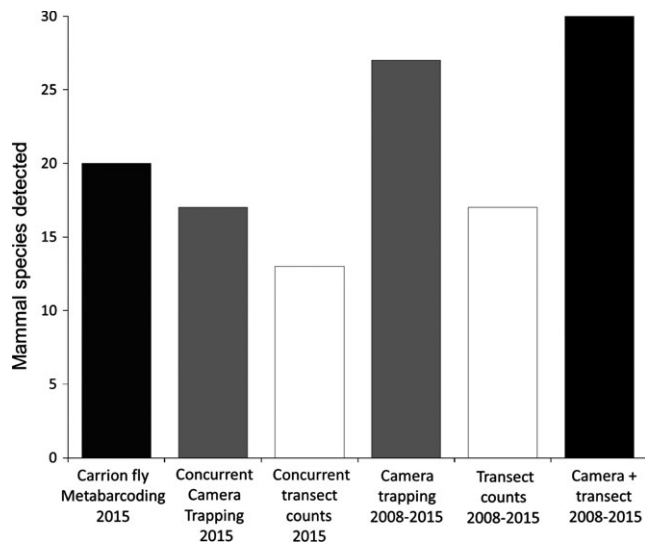
FIGURE 3 Number of mammal species detected by alternative sampling methods on Barro Colorado Island, Panama

FIGURE 4 Species accumulation curves for three different survey methods used to sample mammal diversity on Barro Colorado Island, Panama in 2015

camera trapping (transect counts 2008–2015 = 17; camera trapping 2008–2015 = 26). Using all three methods combined, we detected a total of 27 mammal species in 2015, and 34 mammal species from 2008 to 2015. Visual inspection of the species accumulation curves from 2015 (Figure 4) suggests that additional carrion fly sampling effort, or greater sequencing depth, would likely have resulted in a greater number of species detections, whereas the transect count and camera-trap data sets were nearing asymptotes.

Of the 20 species detected with fly metabarcoding (Table 1), four were not detected by camera traps or transect counts from 2008 to 2015. These included Derby's woolly opossum (*Caluromys derbianus*), a species from the genus *Canis* that was a 100% match to reference sequences of domestic dog (*C. domesticus)* and coyote (*C. latrans*), and two bat species. Conversely, metabarcoding did not detect three species commonly detected with transect counts or camera trapping: paca (*Agouti paca*), white-nosed coati (*Nasua narica*) and red brocket deer (*Mazama americana*; Table 2). Moreover, the Central American agouti (*Dasyprocta punctata*), a large rodent that is by far the most common mammal on BCI, was detected in just one metabarcoding sample, and only by the 12S marker.

4 | DISCUSSION

In total, we detected a greater number of mammal species with carrion fly metabarcoding than with transect counts or camera trapping carried out during the same general time frame with similar levels of effort. Thus, carrion fly metabarcoding appears to be an effective, albeit imperfect, method for surveys of mammalian diversity. Of the 20 species detected with fly metabarcoding (Table 1), four were not detected by camera traps or transect counts. Two were bat species that are unlikely to be detected by camera traps due to their size and are unlikely to be detected by diurnal transect counts. The third, Derby's woolly opossum (*Caluromys derbianus*), is arboreal and
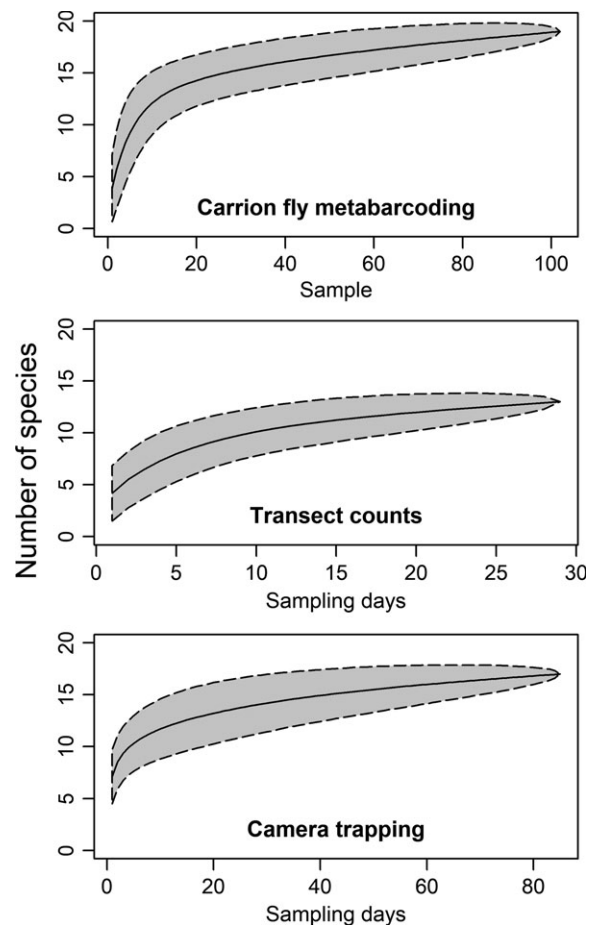
nocturnal, but it has been commonly observed on BCI at night feeding on nectar of canopy flowers. The fourth and most unexpected detection was a species from the genus *Canis*, which was a 100% match to reference sequences of both domestic dog (*C. domesticus*) and coyote (*C. latrans*). Domestic dogs are not permitted on BCI, but they are present on the mainland <1 km away at the nearest point. Coyotes have never been detected on BCI but have been expanding in Panama (Bermudez, Gonzalez, & Garcia, 2013) and have recently been photographed by camera traps in the adjacent Soberania National Park (www.teamnetwork.org). It is possible that a fly fed on a canid on the mainland and then flew to BCI. Little is known about flight and foraging distances for carrion flies; however, Lee (2016) detected a blowfly movement distance of 3 km in tropical Malaysia. It is also possible that ex situ contaminating DNA from a domestic dog was introduced in the laboratory or was present in PCR reagents, although this seems unlikely.

Although fly metabarcoding detected more mammal species than the traditional techniques, a clear shortcoming was failure to detect three abundant species that were commonly detected by transect counts and camera trapping. This included a common rodent (*Agouti paca*), the most common carnivore (*Nasua narica*), and a common ungulate (*Mazama americana*). Metabarcoding did, however, detect

TABLE 2 Mammal species detected by carrion fly metabarcoding, camera trapping and diurnal transect counts on Barro Colorado Island, Panama. For metabarcoding, values represent number of samples in which a species was detected. For camera trapping and transect counts, values are number of individuals detected

Species	Common name	Metabarcoding		Camera trapping		Transect counts		Habit[a]	
		16s	12s	2015	2008–2015	2015	2008–2015		
Order *Primates*									
Alouatta palliata	Mantled howler monkey	99	67		6	315	1,996	D	A
Ateles geoffroyi	Geoffroy's spider monkey	87	17		4	32	136	D	A
Cebus capucinus	White-faced capuchin	27	2	4	147	34	429	D	A
Saguinus geoffroyi	Geoffroy's tamarin	21				17	94	D	A
Order *Carnivora*									
Potos flavus	Kinkajou		11		1			N	A
Procyon cancrivorus	Crab-eating raccoon	2	14	8	52			N	T
Leopardus pardalis	Ocelot	3		329	5,801		6	C	T
Canis sp.	Domestic dog or coyote	1	1						
Nasua narica	White-nosed coati			1,064	7,358	18	178	D	T
Puma concolor	Puma				6			N	T
Panthera onca	Jaguar				9			C	T
Eira barbara	Tayra			1	54		4	D	T
Herpailurus yagouaroundi	Jaguarondi			3	5			C	T
Order *Rodentia*									
Dasyprocta punctata	Agouti		1	2,975	72,719	302	1,821	D	T
Sciurus granatensis	Red-tailed squirrel	10	15	20	1,769	68	735	D	ST
Proechimys semispinosus	Tome's spiny-rat	5	18	114	1,074	1	2	N	T
Agouti paca	Paca			108	2,412			N	T
Heteromys desmarestianus	Desmarest's spiny pocket mouse				44			N	T
Coendou rothschildi	Rothschild's porcupine				3			N	A
Diplomys labilis	Rufous Tree Rat						4	N	A
Hydrochoerus hydrochaeris	Capybara				18			N	T
Order *Didelphimorphia*									
Didelphis marsupialis	Common opossum	1	11	57	1,065			N	ST
Marmosa robinsoni	Robinson's mouse opossum		19	1	7			N	ST
Caluromys derbianus	Derby's woolly opossum	2	21					N	ST
Order *Pilosa*									
Choloepus hoffmanni	Hoffmann's two-toed sloth	1				1	2	C	A
Bradypus variegatus	Brown-throated sloth	1				1	2	C	A
Tamandua mexicana	Northern tamandua		30	20	359	2	27	D	ST
Order *Artiodactyla*									
Pecari tajacu	Collared peccary	6		1,040	12,731	25	383	C	T
Mazama temama	Central American red brocket			304	4,794	9	48	D	T
Order *Cingulata*									
Dasypus novemcinctus	Nine-banded armadillo			5	169			N	T
Cabassous centralis	Northern naked-tailed armadillo				2			N	T
Order *Chiroptera*									
Tonatia saurophila	Stripe-headed round-eared bat	1						N	V
Pteronotus gymnonotus or parnellii	Big naked-backed bat or Parnell's mustached bat	3						N	V

(Continues)

TABLE 2 (Continued)

Species	Common name	Metabarcoding 16s	12s	Camera trapping 2015	2008–2015	Transect counts 2015	2008–2015	Habit[a]	
Order *Perissodactyla*									
Tapirus bairdii	Baird's tapir			2	54	1		N	T

[a]Habit: D = diurnal, C = cathemeral, N = nocturnal, A = arboreal, ST = semi-terrestrial, T = terrestrial, V = volant.

two other rodent species, red-tailed squirrel (*Sciurus granatensis*) and Tome's spiny-rat (*Proechimys semispinosus*), and two confamilial carnivores, crab-eating raccoon (*Procyon cancrivorus*) and kinkajou (*Potos flavus*; Table 1).

There are three potential explanations for why we were unsuccessful in detecting these three common species with metabarcoding. The first possible explanation is that reference sequences for those species are missing from the reference database. Only 59% and 62% of mammal species and 82% and 85% of genera known from BCI are represented in our 16S and 12S reference databases, respectively. However, all three undetected species were present in our reference databases to varying extents: paca was represented in our database by 12S, coati was represented by both markers, albeit only at the genus level (South American coati; *Nasua nasua*), and red brocket deer was represented by both markers at the species level. Given that PROTAX was able to place 98% of 16S OTUs, and 74% of 12S OTUs to at least the genus level (Figure 2), missing reference sequences are not a strong explanation for our failure to detect these species. However, in less-studied areas, incomplete reference databases are more likely to hamper species assignment, and thus, investment in building reference databases should continue.

The second potential explanation for why some common species were not detected with metabarcoding is that there were mismatches in our 16S and 12S primers with binding sites for these species. This could result in failed PCR amplification, even if DNA from these species was present in fly samples. For paca, all nine reference sequences on GenBank have one mismatch in the forward primer, 6 bp from the 3' end. One mismatch is unlikely to completely prevent amplification but could reduce PCR efficiency, especially in mixed samples with other targets. For red brocket deer, no primer mismatches are observed for 12S. For 16S, the forward primer has one mismatch with all GenBank reference sequences, but at the far 5' end. We could not evaluate primer mismatches for white-nosed coati at the species level, but both 16S and 12S primer sets have one 5' mismatch to the congener *Nasua nasua*. Further, currently available GenBank reference sequences may not account for local sequence variants with primer mismatches.

Lastly, it is likely that fly feeding preferences may contribute to sampling bias. Although carrion flies feed on carcasses of dead animals, we suspect that the main source of eDNA is faeces. Primates were the most commonly detected species, with mantled howler monkey (*Alouatta palliata*) and Geoffroy's spider monkey (*Ateles geoffroyi*) identified in 98% and 85% of samples, respectively. Monkeys, and particularly howler monkeys which feed on leaves, produce an abundant quantity of soft scat that can be easily consumed by flies,

while rodents such as agoutis and pacas, and ungulates such as red brocket deer, have smaller and harder scats that may not be as attractive to flies. It is notable that some rodent species were detected nonetheless, but this could have been from carrion, and not from faeces. Likewise, Calvignac-Spencer, Merkel et al. (2013) were able to detect forest ungulates with carrion flies. Further research into the source of carrion fly-derived eDNA and how this affects detection is needed. Fly metabarcoding does seem to work especially well for primates, as all four primate species present on BCI were detected, and Calvignac-Spencer, Merkel et al. (2013) also detected many primates with carrion fly metabarcoding in Cote d'Ivoire and Madagascar.

For species assignment, PROTAX, especially weighted PROTAX, outperformed other assignment methods. With weighted PROTAX, we were able to assign nearly all OTUs to genus, and most to species, especially with 16S. For those OTUs that we could only place at the genus level, we could assign all but two to species based on prior knowledge of the vertebrate community on BCI, as only one member of each genus is present on the island. Weighted PROTAX resulted in fewer clearly false positives (species not known from BCI) than either BLAST or unweighted PROTAX (Figure 2c). However, even weighted PROTAX produced a few false-positive assignments at the species level. In cases where it is important to be conservative with species assignment, and eliminate or minimize false assignments, selecting a high probability threshold (e.g., 0.9 or 0.95) may be desirable. All but one of our false-positive assignments from weighted PROTAX at the species level, and all at the genus level, had PROTAX probabilities of <0.9 (Appendix S2). Assigning such a cut-off value, however, is a trade-off between minimizing false positives and false negatives. We detected several species known from BCI at assignment probabilities of <0.9 (Table 1), and so with a higher probability threshold we would have considered those species as undetected even though they are present. The results from our weighted data sets suggest that an effective way to simultaneously reduce false-positive (overly confident) and false-negative (overly conservative) assignments in PROTAX is to use expert knowledge to assign high prior probabilities to species known to exist in the region (Figure 2).

As evidenced by the 24 flies that we analysed individually and in pools, pooling reduced the number of species detected. When flies were extracted individually, we detected 10 species, including ocelot (*Leopardus pardalis*), Hoffmann's two-toed sloth (*Choloepus hoffmanni*), collared peccary (*Pecari tajacu*), kinkajou, crab-eating raccoon, red-tailed squirrel and all four primates. In the two pooled samples from these same 24 flies, we only detected collared peccary, kinkajou and the two most commonly detected primate species.

Hoffmann's two-toed sloth was only detected in one of the single-fly extractions and was not detected in any of the pooled samples. Thus, if we had not extracted some flies singly, this species would have been missed entirely. Pooling is more likely to affect detection of rare species, as their DNA may be outcompeted during PCR by more abundant DNA of common species. It is also likely that species with more primer mismatches will be outcompeted by species with greater PCR efficiency if DNA of two such species exist in the same pool. Pooling of flies, however, allows for far fewer DNA extractions and individual PCRs, which substantially reduces labour and cost (up to 16 fold in this example), which may allow many more flies to be processed, ultimately increasing the number of species detected. If a pooling strategy is employed, and there are particular target species of concern for which detection is essential, it may be desirable to include species-specific primers in addition to general primers (Schubert et al., 2014). It is also possible that the high number of PCR cycles we employed led to "PCR runaway," in which common amplicons became exponentially abundant in pooled samples at the expense of less-common amplicons. Thus, we advise future studies to optimize the number of PCR cycles with quantitative PCR prior to metabarcoding (Murray, Coghlan, & Bunce, 2015). Finally, the discrepancy in species detection between pooled and single-fly samples could have been the result of insufficient sequencing depth. Because single-fly and pooled samples were sequenced at the same depth, it is possible that for pooled samples, sequencing depth was not sufficient to detect all species in the sample.

The use of two different markers (12S and 16S) resulted in the detection of more species than either marker alone. Thus, we recommend using multiple markers to improve species detection. The inclusion of the 12S marker allowed us to detect more mammal species and also allowed detection of birds and reptiles. This marker, however, appears to have relatively poor information content for discriminating birds and reptiles at the species level. Thus, if the goal is to detect nonmammalian vertebrates, it may be preferable to employ additional group-specific markers. Multiple markers optimized for different groups or families could be run simultaneously, which should increase detection in each group. For mammals, 12S species assignment were generally lower confidence than with 16S (Appendix S2), and a greater proportion of 12S OTUs could not be assigned at the genus or species level (Figure 2b). The standard cytochrome oxidase I (COI) barcode region is by far the most represented in reference databases, but lack of conserved sites in the coding region of COI make primer design for amplification of short fragments from degraded DNA difficult for a wide range of taxa (Deagle, Jarman, Coissac, Pompanon, & Taberlet, 2014). Thus, mitochondrial gene regions such as 16S and 12S are more appropriate for metabarcoding studies targeting degraded DNA.

Choice of bait, trap configuration and trap distribution may have an impact on carrion fly metabarcoding results. We chose to use pork for bait because we wanted to target flies that feed on mammals, and because a *Sus* blocking primer has been previously designed for both markers we used (Calvignac-Spencer, Merkel et al., 2013). We made an effort to keep bait separate from flies during

sampling, but even with the blocking primers, *Sus* DNA was still commonly amplified. *Sus* OTUs accounted for 68% of total 16S reads and 43% of total 12S reads in our data set. In a small exploratory study (unpublished data), we found that increasing *Sus* blocking primer concentrations to 20× as opposed to the 5× used by Calvignac-Spencer, Merkel et al. (2013) led to increased detection of non-*Sus* DNA. If pork is used as bait, we recommend taking as much care as possible to reduce contact of flies with the bait to reduce contamination, and ensure enough sequencing depth so that non-*Sus* amplicons are still sequenced even if they are in the minority. Use of other baits such as chicken or fish, or nonbiological commercial fly baits may mitigate these concerns, but the 12S marker will still amplify chicken or fish DNA, and the 16S marker will also amplify fish DNA (Cannon et al., 2016). Finally, most of the fly samples we collected were from a relatively clustered area (Figure 1) whereas transect counts and camera trapping were more widely distributed throughout the island. It is possible that more widely distributed fly sampling may have further improved species detection.

For maximal species detection, metabarcoding, camera trapping and transect counts could be used simultaneously, as they are complementary. All three techniques combined detected more species than any one of the techniques alone. Conveniently, conducting fly collection and transect counts simultaneously would add little additional field time or cost. During this study, we visually observed two common species missed by metabarcoding, white-nosed coati and red brocket deer, as well as numerous others, while collecting flies (unpublished data). Also, if the goal was to survey biodiversity of both vertebrates and invertebrates, fly traps could be paired with other types of insect traps, for example pitfall traps, and fly DNA and bulk DNA from traps could be amplified with invertebrate primers and included in sequencing runs (Ji et al., 2013; Yu et al., 2012).

5 | CONCLUSIONS

This field test confirms that carrion fly-derived DNA metabarcoding is a powerful tool for mammal biodiversity surveys, already on par with other commonly used methods. A relatively small effort (29 days of fly sampling conducted by one individual) detected the majority of the nonvolant mammal species resident on BCI, and a greater number of mammal species than were detected by camera trapping or diurnal transect counts with similar sampling effort. Nonetheless, carrion fly metabarcoding failed to detect three abundant species that were easily detected with other methods, and metabarcoding might not provide reliable estimates of species relative abundance (Schnell et al., 2015). Thus, we do not advocate that carrion fly metabarcoding replace current methods, but rather add to the tools available for biodiversity surveys. Our results suggest some methodological modifications that will likely increase the detection power of carrion fly metabarcoding, including more complete reference sequence databases, optimization of PCR conditions, multiple custom markers, greater sampling

effort and greater sequencing depth. Individual fly metabarcoding could also increase detection power, but at a larger cost. Nevertheless, we conclude that carrion fly metabarcoding is an important tool, with the potential to greatly improve researchers' ability to efficiently monitor biodiversity. As sequencing costs drop further and as reference databases of complete mitochondrial genomes becomes readily available (Tang et al., 2014), this method promises to allow for the rapid characterization and monitoring of vertebrate communities.

DATA ACCESSIBILITY

Raw sequence data are available on the NCBI Sequence Read Archive under Accession number PRJNA382243. OTU sequences are available in the supplementary information. Scripts for the PRO-TAX analysis are available on Dryad under https://doi.org/10.5061/dryad.bj5k0.

AUTHOR CONTRIBUTIONS

T.W.R. designed the study, carried out all fly collection field work and laboratory work, coordinated the data analysis, and wrote the first draft of the manuscript. C.C.Y.X. performed the PROTAX analysis; J.G. collected the camera trap and transect count data; K.M.K. conducted the sequence filtering and OTU generation bioinformatics work; K.S. helped with design and implementation of laboratory work; M.V. helped with laboratory work management and conducted MiSeq sequencing; D.W.Y. supervised the PROTAX analysis; P.S. provided scripts and helped with the PROTAX analysis; and P.A.J. helped with field study design. W.O.M. and P.A.J. jointly advised this work. All authors contributed to manuscript writing and editing.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. [Article]. *Journal of Molecular Biology*, 215(3), 403–410. https://doi.org/10.1006/jmbi.1990.9999

Andersen, K., Bird, K. L., Rasmussen, M., Haile, J., Breuning-Madsen, H., Kjaer, K. H., . . . Willerslev, E. (2012). Meta-barcoding of 'dirt' DNA from soil reflects vertebrate biodiversity. *Molecular Ecology*, 21(8), 1966–1979. https://doi.org/10.1111/j.1365-294x.2011.05261.x

Beaudrot, L., Ahumada, J. A., O'Brien, T., Alvarez-Loayza, P., Boekee, K., Campos-Arceiz, A., . . . Andelman, S. J. (2016). Standardized assessment of biodiversity trends in tropical forest protected areas: The end is not in sight. *PLoS Biology*, 14(1), e1002357. https://doi.org/10.1371/journal.pbio.1002357

Bermudez, C. S. E., Gonzalez, D. D., & Garcia, S. G. (2013). Ticks (Acari: Ixodidae, Argasidae) of Coyotes in Panama [Article]. *Systematic and Applied Acarology*, 18(2), 112–115.

Boessenkool, S., Epp, L. S., Haile, J., Bellemain, E., Edwards, M., Coissac, E., . . . Brochmann, C. (2012). Blocking human contaminant DNA during PCR allows amplification of rare mammal species from sedimentary ancient DNA. *Molecular Ecology*, 21(8), 1806–1815. https://doi.org/10.1111/j.1365-294x.2011.05306.x

Bohmann, K., Evans, A., Gilbert, M. T. P., Carvalho, G. R., Creer, S., Knapp, M., . . . de Bruyn, M. (2014). Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology & Evolution*, 29(6), 358–367. https://doi.org/10.1016/j.tree.2014.04.003

Bohmann, K., Schnell, I. B., & Gilbert, M. T. P. (2013). When bugs reveal biodiversity. *Molecular Ecology*, 22(4), 909–911. https://doi.org/10.1111/mec.12221

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2016). obitools: A unix-inspired software package for DNA metabarcoding. *Molecular Ecology Resources*, 16(1), 176–182. https://doi.org/10.1111/1755-0998.12428

Calvignac-Spencer, S., Leendertz, F. H., Gilbert, M. T. P., & Schubert, G. (2013). An invertebrate stomach's view on vertebrate ecology Certain invertebrates could be used as "vertebrate samplers" and deliver DNA-based information on many aspects of vertebrate ecology. *BioEssays*, 35(11), 1004–1013. https://doi.org/10.1002/bies.201300060

Calvignac-Spencer, S., Merkel, K., Kutzner, N., Kuhl, H., Boesch, C., Kappeler, P. M., . . . Leendertz, F. H. (2013). Carrion fly-derived DNA as a tool for comprehensive and cost-effective assessment of mammalian biodiversity. *Molecular Ecology*, 22(4), 915–924. https://doi.org/10.1111/mec.12183

Cannon, M. V., Hester, J., Shalkhauser, A., Chan, E. R., Logue, K., Small, S. T., & Serre, D. (2016). In silico assessment of primers for eDNA studies using PrimerTree and application to characterize the biodiversity surrounding the Cuyahoga River. *Scientific Reports*, 6, 1–11. https://doi.org/10.1038/srep22908

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., . . . Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, 7(5), 335–336. https://doi.org/10.1038/nmeth.f.303

Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2016). GenBank. *Nucleic Acids Research*, 44(D1), D67–D72. https://doi.org/10.1093/nar/gkv1276

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(9), 1–4. https://doi.org/10.1098/rsbl.2014.0562

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26(19), 2460–2461.

Ficetola, G. F., Coissac, E., Zundel, S., Riaz, T., Shehzad, W., Bessière, J., . . . Pompanon, F. (2010). An In silico approach for the evaluation of DNA barcodes. *Bmc Genomics*, 11(1), 434. https://doi.org/10.1186/1471-2164-11-434

Gillett, C. P. D. T., Johnson, A. J., Barr, I., & Hulcr, J. (2016). Metagenomic sequencing of dung beetle intestinal contents directly detects and identifies mammalian fauna. *bioRxiv*. https://doi.org/10.1101/074849

Glanz, W. E. (1982). The terrestrial mammal fauna of Barro Colorado Island: Censuses and long-term changes. In E. G. Leigh, A. S. Rand, & D. M. Windsor (Eds.), *The ecology of a tropical forest: Seasonal rhythms*

and long-term changes (pp. 455–468). Washington, DC: Smithsonian Institution Press.

Huson, D. H., Auch, A. F., Qi, J., & Schuster, S. C. (2007). MEGAN analysis of metagenomic data. *Genome Research*, 17(3), 377–386. https://doi.org/10.1101/gr.5969107

Ji, Y., Ashton, L., Pedley, S. M., Edwards, D. P., Tang, Y., Nakamura, A., ... Yu, D. W. (2013). Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters*, 16(10), 1245–1257. https://doi.org/10.1111/ele.12162

Kielbasa, S. M., Wan, R., Sato, K., Horton, P., & Frith, M. C. (2011). Adaptive seeds tame genomic sequence comparison. *Genome Research*, 21(3), 487–493. https://doi.org/10.1101/gr.113985.110

Lee, P. S. (2016). *A blowfly-derived dna approach to assess diversity of tropical mammals*. PhD thesis, University of Malaysia. Retrieved from http://studentsrepo.um.edu.my/6715/

Lee, P. S., Gan, H. M., Clements, G. R., & Wilson, J. J. (2016). Field calibration of blowfly-derived DNA against traditional methods for assessing mammal diversity in tropical forests. *Genome*, 59(11), 1008–1022. https://doi.org/10.1139/gen-2015-0193

Lee, P. S., Sing, K. W., & Wilson, J. J. (2015). Reading mammal diversity from flies: The persistence period of amplifiable mammal mtDNA in blowfly guts (*Chrysomya megacephala*) and a new DNA mini-barcode target. *PLoS ONE*, 10(4), 1–12. https://doi.org/10.1371/journal.pone.0123871

Logue, K., Keven, J. B., Cannon, M. V., Reimer, L., Siba, P., Walker, E. D., ... Serre, D. (2016). Unbiased characterization of anopheles mosquito blood meals by targeted high-throughput sequencing. *PLOS Neglected Tropical Diseases*, 10(3), e0004512. https://doi.org/10.1371/journal.pntd.0004512

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal*, 17(1), 10–12. https://doi.org/10.14806/ej.17.1.200

Murray, D. C., Coghlan, M. L., & Bunce, M. (2015). From benchtop to desktop: Important considerations when designing amplicon sequencing workflows. *PLoS ONE*, 10(4), e0124671. https://doi.org/10.1371/journal.pone.0124671

Nichols, R. V., Konigsson, H., Danell, K., & Spong, G. (2012). Browsed twig environmental DNA: Diagnostic PCR to identify ungulate species. *Molecular Ecology Resources*, 12(6), 983–989. https://doi.org/10.1111/j.1755-0998.2012.03172.x

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., ... Wagner, H. (2016). *vegan: Community Ecology Package*. R package version 2.4-1.

Riaz, T., Shehzad, W., Viari, A., Pompanon, F., Taberlet, P., & Coissac, E. (2011). ecoPrimers: Inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Research*, 39(21), e145. https://doi.org/10.1093/nar/gkr732

Rodgers, T. W., & Mock, K. E. (2015). Drinking water as a source of environmental DNA for the detection of terrestrial wildlife species. *Conservation Genetics Resources*, 7(3), 693–696. https://doi.org/10.1007/s12686-015-0478-7

Schnell, I. B., Sollmann, R., Calvignac-Spencer, S., Siddall, M. E., Yu, D. W., Wilting, A., & Gilbert, M. T. P. (2015). iDNA from terrestrial haematophagous leeches as a wildlife surveying and monitoring tool - prospects, pitfalls and avenues to be developed. *Frontiers in Zoology*, 12, 1–14. https://doi.org/10.1186/s12983-015-0115-z

Schubert, G., Stockhausen, M., Hoffmann, C., Merkel, K., Vigilant, L., Leendertz, F. H., & Calvignac-Spencer, S. (2014). Targeted detection of mammalian species using carrion fly-derived DNA. *Molecular Ecology Resources*, 15(2), 285–294. https://doi.org/10.1111/1755-0998.12306

Somervuo, P., Koskela, S., Pennanen, J., Henrik Nilsson, R., & Ovaskainen, O. (2016). Unbiased probabilistic taxonomic classification for DNA barcoding. *Bioinformatics*, 32(19), 2920–2927. https://doi.org/10.1093/bioinformatics/btw346

Somervuo, P., Yu, D. W., Xu, C., Ji, Y., Hultman, J., Wirta, H., & Ovaskainen, O. (2017). Quantifying uncertainty of taxonomic placement in DNA barcoding and metabarcoding. *Methods in Ecology and Evolution*, 8(4), 398–407. https://doi.org/10.1111/2041-210x.12721

Taberlet, P., Coissac, E., Hajibabaei, M., & Rieseberg, L. H. (2012). Environmental DNA. *Molecular Ecology*, 21(8), 1789–1793. https://doi.org/10.1111/j.1365-294x.2012.05542.x

Tang, M., Tan, M., Meng, G., Yang, S., Su, X., & Liu, S. (2014). Multiplex sequencing of pooled mitochondrial genomes – A crucial step toward biodiversity analysis using mito-metagenomics. *Nucleic Acids Research*, 42, https://doi.org/10.1093/nar/gku917

Thomsen, P. F., Kielgast, J., Iversen, L. L., Wiuf, C., Rasmussen, M., Gilbert, M. T. P., ... Willerslev, E. (2012). Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular Ecology*, 21(11), 2565–2573. https://doi.org/10.1111/j.1365-294x.2011.05418.x

Wheat, R. E., Allen, J. M., Miller, S. D. L., Wilmers, C. C., & Levi, T. (2016). Environmental DNA from residual saliva for efficient noninvasive genetic monitoring of brown bears (*Ursus arctos*). *PLoS ONE*, 11(11), e0165259. https://doi.org/10.1371/journal.pone.0165259

Yu, D. W., Ji, Y., Emerson, B. C., Wang, X., Ye, C., Yang, C., & Ding, Z. (2012). Biodiversity soup: Metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution*, 3(4), 613–623. https://doi.org/10.1111/j.2041-210x.2012.00198.x

Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2014). PEAR: A fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, 30(5), 614–620. https://doi.org/10.1093/bioinformatics/btt593

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.