Received: 9 December 2023 | Revised: 2 May 2024 | Accepted: 13 May 2024

DOI: 10.1111/1755-0998.13980

MOLECULAR ECOLOGY RESOURCES WILEY

RESOURCE ARTICLE

CRISPR-Cas9-mediated host signal reduction for 18S metabarcoding of host-associated eukaryotes

Leah A. Owens[1] | Mary I. Thurber[2] | Tony L. Goldberg[1]

[1]Department of Pathobiological Sciences, School of Veterinary Medicine, University of Wisconsin-Madison, Madison, Wisconsin, USA

[2]Henry Vilas Zoo, Madison, Wisconsin, USA

Correspondence
Leah A. Owens and Tony L. Goldberg, Department of Pathobiological Sciences, School of Veterinary Medicine, University of Wisconsin-Madison, Madison, WI, USA.
Email: leah.owens@wisc.edu, tony.goldberg@wisc.edu

Funding information
Division of Intramural Research, Grant/Award Number: 1R01AG049395-01, 1R21AI163592-01 and T32AI007414

Handling Editor: Kayla King

Abstract

Metabarcoding-based methods for identification of host-associated eukaryotes have the potential to revolutionize parasitology and microbial ecology, yet significant technical challenges remain. In particular, highly abundant host reads can mask the presence of less-abundant target organisms, especially for sample types rich in host DNA (e.g., blood and tissues). Here, we present a new CRISPR-Cas9-mediated approach designed to reduce host signal by selective amplicon digestion, thus enriching clinical samples for eukaryotic endosymbiont sequences during metabarcoding. Our method achieves a nearly 76% increased efficiency in host signal reduction compared with no treatment and a nearly 60% increased efficiency in host signal reduction compared with the most commonly used published method. Furthermore, the application of our method to clinical samples allows for the detection of parasite infections that would otherwise have been missed.

KEYWORDS
CRISPR-Cas9, eukaryotic microbiome, metabarcoding, parasitology

1 | INTRODUCTION

Metagenomic barcoding (metabarcoding) provides a high throughput alternative to traditional methods for reconstructing communities of host-associated organisms (Forsman et al., 2022). Substantial progress has been made in methods for metabarcoding bacteria and archaea (i.e., the 'microbiome') (Hamady & Knight, 2009) and fungi (i.e., the 'mycobiome') (Tedersoo et al., 2022), but similar progress has lagged for eukaryotic endosymbionts (defined here as all non-fungal eukaryotes residing within vertebrate hosts, spanning the continuum of parasites to commensals and including micro- and macro-organisms) (Laforest-Lapointe & Arrieta, 2018). One critical reason for this lag is that eukaryotic endosymbionts share highly similar DNA sequences with their eukaryotic hosts but usually at much lower concentration, leading to host signal interference (Lundberg et al., 2013; Sakai & Ikenaga, 2013). Polymerase chain reaction (PCR) primers designed to broadly recognize eukaryotic endosymbionts (especially metazoans, such as helminths) also often bind to and amplify host DNA (i.e., non-specific or off-target amplification) (Belda et al., 2017; Vestheim & Jarman, 2008). Primers that recognize both host and target sequences generally detect only 10^{-3} ng parasite DNA for every ng host DNA present (Sow et al., 2019). For example, spleen tissue from mice experimentally infected via tail vein injection with *Leishmania donovoni* harboured an average of 200 promastigotes per 0.2 mg spleen tissue, resulting in an average ng parasite DNA: ng host DNA ratio of 10^{-5} (Nicolas et al., 2002; Titus et al., 1985). One 'brute force' solution to this problem is ultra-deep sequencing—in other words, sequencing amplicons to great enough depth to compensate for host signal overabundance—but this approach is inefficient, costly and biased against detecting

low-abundance organisms (Alberdi et al., 2018; Belda et al., 2017). Using metabarcoding to reconstruct eukaryotic endosymbiont assemblages from faeces is commonplace, but faecal matter is so dominated by bacterial DNA that it can also interfere with detection of eukaryotes, even using primers that appear to be eukaryote-specific (Feehery et al., 2013; Jiang et al., 2020).

A reliable and efficient eukaryotic endosymbiont metabarcoding method should include a host-blocking element to enrich resulting sequences for eukaryotic endosymbiont reads in any sample type with high host DNA content (O'Rorke et al., 2012). We refer to this process as 'host signal reduction' (HSR). Published HSR methods, including restriction enzyme digestion (Flaherty et al., 2018), peptide nucleic acid (PNA) clamps (Terahara et al., 2011), blocking oligonucleotides (Vestheim et al., 2011) and nested blocking primers (Mayer et al., 2020), each have advantages and disadvantages. The restriction enzyme approach, in which primers are designed such that only host amplicons contain a restriction enzyme recognition site, allowing for selective cleavage of off-target amplicons prior to sequencing (Flaherty et al., 2021), is effective, but suitable restriction sites with flanking PCR primer sites are rare or sometimes non-existent. Selective inhibition of off-target amplification during PCR is the most commonly published host signal reduction technique (Mamanova et al., 2010) and can be achieved using PNA clamps or various blocking oligonucleotides (Troedsson et al., 2008; von Wintzingerode et al., 2000). Such methods have been used in published eukaryotic endosymbiont metabarcoding studies (Hino et al., 2016; Lappan et al., 2019; Mann et al., 2020), but efficacy can be low, particularly in samples with high host biomass (Lundberg et al., 2013). Nested blocking primers were recently published for plant systems (Mayer et al., 2020) but have yet to be adapted for eukaryotic endosymbiont metabarcoding and may suffer the same drawbacks as PNA clamps and blocking oligos.

CRISPR-Cas9 (CC9)-mediated removal of highly abundant off-target nucleic acids is regularly used in other sequencing-based approaches, such as chromatin structure studies (Wu et al., 2016), cancer screening (Gu et al., 2016) and plant microbiome profiling (Song & Xie, 2020). CC9 is a promising method for host signal blocking in eukaryotic endosymbiont metabarcoding because CRISPR-Cas9 nuclease activity is highly specific (Wu et al., 2014), reagents are readily available and relatively inexpensive, and the reaction components are modular such that different hosts or read types (e.g., dietary or environmental sequences) can be eliminated depending on experimental requirements (Lin et al., 2021). In fact, CC9 can be designed to target a range of organisms, from a single eukaryotic lineage to broad taxonomic groups, such as entire clades of eukaryotes, simply by altering the guide RNA sequence. To our knowledge, however, CC9 has not been applied to HSR in the context of eukaryotic endosymbiont metabarcoding.

Here, we assess the most commonly published HSR protocol for eukaryotic endosymbiont metabarcoding, off-target PCR inhibition, and demonstrate the need for a more effective approach. We design such a method based on a recombinant *Streptococcus pyogenes* CC9 system, in which vertebrate sequences are selectively targeted for cleavage and removal by host-specific short guide RNAs (sgRNAs) while leaving amplicons of interest intact for sequencing and analysis. Using in silico analyses, in vitro digests and samples from experimentally infected animals, we show that our method is more effective than published HSR methods across various sample types. Finally, we compare the efficacy of eukaryotic endosymbiont metabarcoding for the detection of known parasite infections and show that CC9 host signal reduction is necessary to detect haemoparasites in blood samples from naturally infected hosts.

2 | MATERIALS AND METHODS

2.1 | Sample collection, characterization and DNA extraction

Chimpanzee samples—We used archived whole blood, plasma, serum, faeces and solid tissues (brain, liver lung, spleen and colon) from western chimpanzees (*Pan troglodytes verus*) in Sierra Leone, collected as part of a previous study (Owens et al., 2021). We used only surplus materials and did not collect any samples solely for the purpose of this research. The care and sampling of this population of chimpanzees is officially sanctioned by the Government of Sierra Leone, and samples were shipped to the USA with the official permission of the Government of Sierra Leone under Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) permit number: 17US19807C/9. All samples were fresh frozen, stored at −80°C, shipped frozen on dry ice and stored at −80°C upon arrival. For DNA extraction, we thawed faeces and blood/blood products on ice and subsampled solid tissues using a sterile 6-mm biopsy punch (Integra Life Sciences, Princeton, NJ, USA) while still frozen. We homogenized faecal samples by vortexing prior to transferring to bead beating tubes for DNA extraction using the DNeasy PowerLyzer PowerSoil Kit (Qiagen, Hilden, Germany), according to the manufacturer's directions, eluted genomic DNA in C6 buffer and stored at −20°C. We extracted DNA from blood/blood products and tissue samples using the Qiagen DNeasy Blood and Tissue Kit following the manufacturer's instructions, eluted in buffer AE and stored at −20°C.

Single host and parasite samples—For in vitro CRISPR-Cas9 digests of amplicons, we used in-house archives of surplus genomic DNA retained from prior studies (Owens et al., 2023). Genomic DNA was previously extracted from single hosts and parasites using the Qiagen DNeasy Blood and Tissue Kit according to the manufacturer's instructions, eluted in buffer AE and stored at −80°C.

Dog samples—We obtained fresh, heparinized blood from domestic dogs (*Canis lupis familiaris*) infected with live *Dirofilaria immitis* strain 'Missouri' microfilariae from a commercial source (BEI resources, Manassas, VA, USA; Catalog # NR-48907). We assessed microfilarial numbers by adding 20 μL of whole heparinized blood immediately after arrival in our laboratory to a glass slide with two drops of 2% formalin; then, we enumerated microfilariae using phase optics at ×10 magnification. We examined samples in triplicate and

calculated microfilarial load as number of microfilariae per 20 μL of blood averaged across the three replicates. After counting microfilariae, we immediately extracted DNA from blood using the Qiagen DNeasy Blood and Tissue Kit according to the manufacturer's instructions, eluted in buffer AE and stored at −20°C.

Red colobus samples—We used archived whole blood samples from red colobus (*Procolobus rufomitratus*) in Uganda, collected as part of a previous study (Thurber et al., 2013). We used only surplus materials and did not collect any samples solely for the purpose of this research. All animal procedures in the original study were approved by the Uganda National Council for Science and Technology, the Uganda Wildlife Authority and the animal use committees of the University of Wisconsin-Madison, USA and McGill University, Canada. Samples were shipped following International Air Transport Association (IATA) regulations under the Ugandan Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) permit 002290. Blood was frozen immediately after collection in liquid nitrogen for storage and transport to the United States where it was stored at −80°C until use. For the original study, samples were assessed for the presence of *Hepatocystis* parasites using genus-specific PCR followed by 454 pyrosequencing of a 420 base pair region of the *cytochrome b* gene (Thurber et al., 2013). We used the same blood samples for DNA extraction using the Qiagen DNEasy Blood and Tissue Kit according to the manufacturer's instructions, eluted in buffer AE and stored at −20°C.

2.2 | 18S V4 metabarcoding with PNA clamp

The most commonly published method for blocking host signal in metabarcoding of the 18S small subunit (SSU) ribosomal RNA (rRNA) hypervariable 4 region (V4) (18S V4 hereafter) is the use of a peptide nucleic acid (PNA) mammal-blocking primer to inhibit host amplification during PCR (Mann et al., 2020; Vestheim et al., 2011). Using samples from chimpanzees as starting material, we followed the published protocol with a few minor modifications. Specifically, primers used to amplify 18S V4 were based on published pan-eukaryotic sequences E572F and E1009R (Comeau et al., 2011), which we modified to replace individual barcodes with overhang adapters (underlined) compatible with the Nextera library preparation system (Illumina, San Diego, CA, USA): F 5′-TCGTCGGCAG CGTCAGATGTGTATAAGAGACAG**CYGCGGTAATTCCAGCTC**-3′ and R 5′-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**AYG GTATCTRATCRTCTTYG**-3′ (amplicon specific region in bold). We used the PNA mammal-blocking primer (PNA Bio; Thousand Oaks, CA, USA): 5′-TCTTAATCATGGCCTCAGTT-3′ as described previously (Mann et al., 2020) and conditions for amplicon PCR with and without blocking primer based on those detailed in Mann et al. We cleaned resulting PCR products using AMPure XP beads (Agencourt, Beverley, MA, USA) according to the manufacturer's instructions and used 5 μL of clean template in a 25-μl PCR with the Illumina Nextera XT Index Kit v2, KAPA HiFi HotStart ReadyMix (Roche Diagnostics, Indianapolis, IN, USA) and an annealing temperature of 55°C for 10 cycles. We cleaned Indexed libraries using Agencourt AMPure XP beads and quantified using a Qubit fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). We sequenced libraries on an Illumina MiSeq instrument using paired-end 300 × 300 cycle V3 chemistry.

2.3 | Short guide RNA design and in silico screening

We assessed host signal reduction using CRISPR Cas-9 in vitro digestion, which cleaves DNA in a sequence-specific manner. In this system, a short guide RNA (sgRNA) of 20 base pairs, including a 3′ seed sequence of 6 base pairs, is loaded onto the recombinant *Streptococcus pyogenes* Cas9 ribonucleoprotein complex, to form a functional endonuclease. The sgRNA targets the entire complex to a specific sequence by binding a complimentary site on the DNA targeted for cleavage. If a protospacer adjacent motif (PAM) 'NGG' lies immediately adjacent to the 3′ end of the seed sequence, it will be cleaved by the active site of the Cas9 endonuclease complex and result in two smaller DNA fragments (Cao et al., 2016). When electrophoresed on an agarose gel, the larger uncleaved DNA will separate from the smaller cleaved DNA fragments and can be visualized. Finally, the larger band may be excised from the gel and extracted for downstream use, leaving behind the unwanted, cleaved DNA.

Our goal was to design sgRNAs to specifically recognize and cleave 'off-target' host (vertebrate) DNA while leaving 'target' endosymbiont (helminth, protozoan) DNA intact for sequencing. We used two concurrent approaches to design sgRNA sequences to recognize vertebrate host 18S V4: (1) the ARB 7.0 software package (Ludwig et al., 2004) with the SILVA SSU rRNA 132 Non-redundant Reference (RefNR) database (Quast et al., 2013) and (2) The Broad Institute's online CRISPick tool (https://portals.broadinstitute.org/gppx/crispick/public) (Doench et al., 2016) using human (*Homo sapiens*, NCBI RefSeq GCF_000001405.40), house mouse (*Mus musculus*, NCBI RefSeq GCF_000001635.26), domestic dog (*Canis lupus familiaris*, NCBI RefSeq GCF_000002285.5) and chimpanzee (*Pan troglodytes*, NCBI RefSeq GCF_002880755.1) genomes as input. We screened 50 candidate sgRNA sequences generated from each of these tools ($n = 100$ total) using SILVA TestProbe (Klindworth et al., 2013) in silico hybridization to the SILVA 138.1 RefNR database with maximum stringency (no mismatches between sgRNA sequence and DNA target) or allowing for a single mismatch outside of the 6-base pair seed sequence (Table 1). Resulting coverage metrics were used to choose the six sgRNA sequences that targeted the highest number of vertebrates and lowest number of eukaryotic endosymbionts for further testing: arb321, arb326 and arb615 were designed in the arb software suite and CA149, CA172 and PT7.1 were designed using CRISPick. Alignments of sgRNAs with host sequences and digest maps were visualized using CLC Genomics Workbench v.20.2.4 (Qiagen). We checked host DNA sequences targeted by the sgRNAs to ensure they include a protospacer adjacent motif (PAM) 'NGG' required by the *Streptococcus pyogenes* Cas9 enzyme for cleavage.

TABLE 1 sgRNA sequences and characteristics.

ID	Target/sgRNA Seq	Orientation	PAM Seq	GC %	Seed seq	Host specificity[a]
arb321[b]	AACTGAGGCCATGATTAAGA[b]	Sense	GGG	45	TTAAGA	Mammals
arb326	AGGCCATGATTAAGAGGGA	Sense	CGG	40	GAGGGA	Mammals
arb615	GCAGCTAGGAATAATGGAAT	Sense	AGG	55	TGGAAT	Mammals, Birds, Fish
PT7.1	ATTCTTGGACCGGCGCAAGA	Sense	CGG	40	GCAAGA	Vertebrates
CA149	CTCAGCTAAGAGCATCGAGG	Antisense	GGG	60	ATCGAGG	Mammals, Birds, Fish
CA172	TCTTAGCTGAGTGTCCCGCG	Sense	GGG	55	CCCGCG	Mammals, Birds, Fish

Abbreviations: PAM, protospacer adjacent motif; seq, sequence; sgRNA, short guide RNA.

[a]Specificity to host groups determined by SILVA TestProbe in silico hybridization data.

[b]Sequence identical to V4 mammal-blocking PNA oligo used in Mann et al., 2020.

2.4 | CRISPR-Cas9 in vitro digestion of representative organisms

All reagents for CC9 treatment of amplicons were components of the Alt-R CRISPR-Cas9 system (Integrated DNA Technologies [IDT] Coralville, IA, USA), based on recombinant *Streptococcus pyogenes* Cas9 nuclease, including Alt-R® S.p. Cas9 Nuclease V3, Alt-R® CRISPR-Cas9 tracrRNA and Alt-R® CRISPR-Cas9 crRNA. crRNA is the component containing the specific targeting sequence that, when complexed with tracrRNA, forms the functional sgRNA (see Table 1 for sequences). Digest reactions were performed following the IDT 'Alt-R CRISPR-Cas9 system – *in vitro* cleavage of target DNA with RNP complex' protocol version 2.2 using recommendations for PCR product templates of 500–2000 base pair lengths and 2–5 nM final DNA concentration per reaction.

CC9 cleavage and sgRNA specificity were initially assessed in vitro using a panel of genomic DNA samples extracted from single representative vertebrate hosts ($n=5$) and eukaryotic endosymbionts ($n=6$). Representative host organisms tested included as follows: Mammal—*Ursus maritimus* (polar bear), Amphibian—*Lithobates chiricahuensis* (leopard frog), Bird—*Gallus gallus* (chicken), Reptile—*Varanus varius* (monitor lizard) and Fish—*Salmo trutta* (brown trout). Representative eukaryotic endosymbiont organisms tested included: Protozoan—*Entamoeba histolytica* (amoeba), Protozoan—*Trypanosoma brucei* (flagellate), Microsporidian—*Encephalitozoon cuniculi*, Acanthocephalan—*Echinorhynchus salmonis* (spiny-headed worm), Platyhelminth—*Schistosoma mansoni* (fluke) *and* Nematode—*Ascaris suum* (roundworm). For this initial experiment, we chose four sgRNAs to constitute the minimum representative set of sgRNAs that would include all of the host specificity groups (Mammals, Mammals/Birds/Fish and Vertebrates), both sequence orientations (Sense and Antisense) and both design tools (arb and CRISPick). 18S V4 amplicon PCR was performed as described above, and resulting amplicons were used in Alt-R CRISPR-Cas9 digest reactions with sgRNAs arb321, arb615, PT7.1 and CA149. Cleavage products were separated by gel electrophoresis on 1.5% agarose gels containing 0.02 μg/mL ethidium bromide, visualized under ultraviolet light and documented using a GelDoc XR imager (Bio-Rad, Hercules, CA, USA). Successful cleavage was indicated by the presence of bands of between approximately 150 and 500 base pairs, which were discernably smaller than the full 18S V4 amplicon of approximately 700 base pairs.

2.5 | Comparison of host signal reduction methods

We compared the efficacy of two HSR methods for improving eukaryotic endosymbiont metabarcoding: the commonly published PNA blocking method and our newly designed CC9 method. Since the PNA oligo works during PCR to block host amplification and CC9 works after PCR to digest host sequences, both methods could theoretically be used together, so we included a dual protocol to investigate whether both methods might synergize. Thus, we performed 18S V4 library preparation in conjunction with four different protocols: (1) CC9 digestion of amplicons using sgRNA arb321, (2) published V4 PNA mammal-blocking oligo described above added to the amplicon PCR, (3) both CC9/sgRNA 321 digestion and PNA mammal-blocking oligo and (4) untreated control (no PNA, no CC9). PCR templates consisted of genomic DNA extracted from chimpanzee blood, liver, lung, colon and faecal samples ($n=3$ each). 18S V4 library preparation, 18S V4 library preparation with PNA blocker and CC9 digests were performed as described above. For CC9 digested amplicons, uncleaved products (bands corresponding to undigested target amplicons) were excised from agarose gels using sterile razor blades and DNA was extracted from the gel matrix using the ZymoClean Gel DNA Recovery Kit (Zymo, Irvine, CA, USA) according to the manufacturer's instructions.

2.6 | Optimization of CRISPR-Cas9 digest

We examined ratios of ribonucleoprotein complex (RNP) to host target DNA of 0.75:1, 1:1 and 1.25:1. CC9 treatment was also tested at two different places in the metabarcoding protocol: (1) after the initial amplification PCR prior to indexing (requiring one digest reaction per sample) or (2) after the second PCR (requiring one digest reaction total for the combined pool of samples). For evaluation of the effect of sgRNA targeting sequence on CC9 digest efficiency, we performed metabarcoding on chimpanzee blood samples ($n=3$) using all six newly designed sgRNAs in six separate reactions. We amplified 18S V4 from each sample and divided the PCR products into seven equal parts (one for each sgRNA and one for an untreated control) prior to library preparation, followed by sequencing and quantification of host read abundance under each condition. The top three sgRNAs (arb326, CA149 and PT7.1) were then tested in the same manner on a larger set of chimpanzee blood samples ($n=31$).

2.7 | Detection of known parasite infections in mammal blood samples

To test the effect of HSR and CC9 on the detection of eukaryotic parasites in a verified infection, we performed eukaryotic endosymbiont metabarcoding on dog blood samples containing a mean of 57.8 *Dirofilaria immitis* microfilariae per 20 μL whole blood. We prepared sequencing libraries using CC9 digestion with each of the 6 newly designed sgRNAs, amplification with a PNA blocking oligo or untreated control prior to sequencing and quantified host read abundance under each condition.

For metabarcoding of naturally infected hosts, we used whole blood samples from wild red colobus that were characterized by PCR and *cytochrome b* amplicon sequencing as part of a concluded study (see sample information above for details) (Thurber et al., 2013). Most samples ($n = 16$ of 19) had been found to contain one of two distinct lineages of the apicomplexan parasite *Hepatocystis*: Species A in 12 of 16 infected hosts and Species B in 4 of 16 infected hosts (Thurber et al., 2013). We used aliquots of these same blood samples for genomic DNA extraction, 18S amplicon library preparation, treatment with CC9 digest (with sgRNA CA149) or untreated control, sequencing and quantification of host read abundances.

2.8 | Sequence data processing and analyses

We performed bioinformatics using QIIME 2 2020.6 (Bolyen et al., 2019). We demultiplexed and quality-filtered raw sequencing reads with the q2-demux plugin followed by denoising with DADA2 (q2-dada2 plugin) (Callahan et al., 2016). We aligned

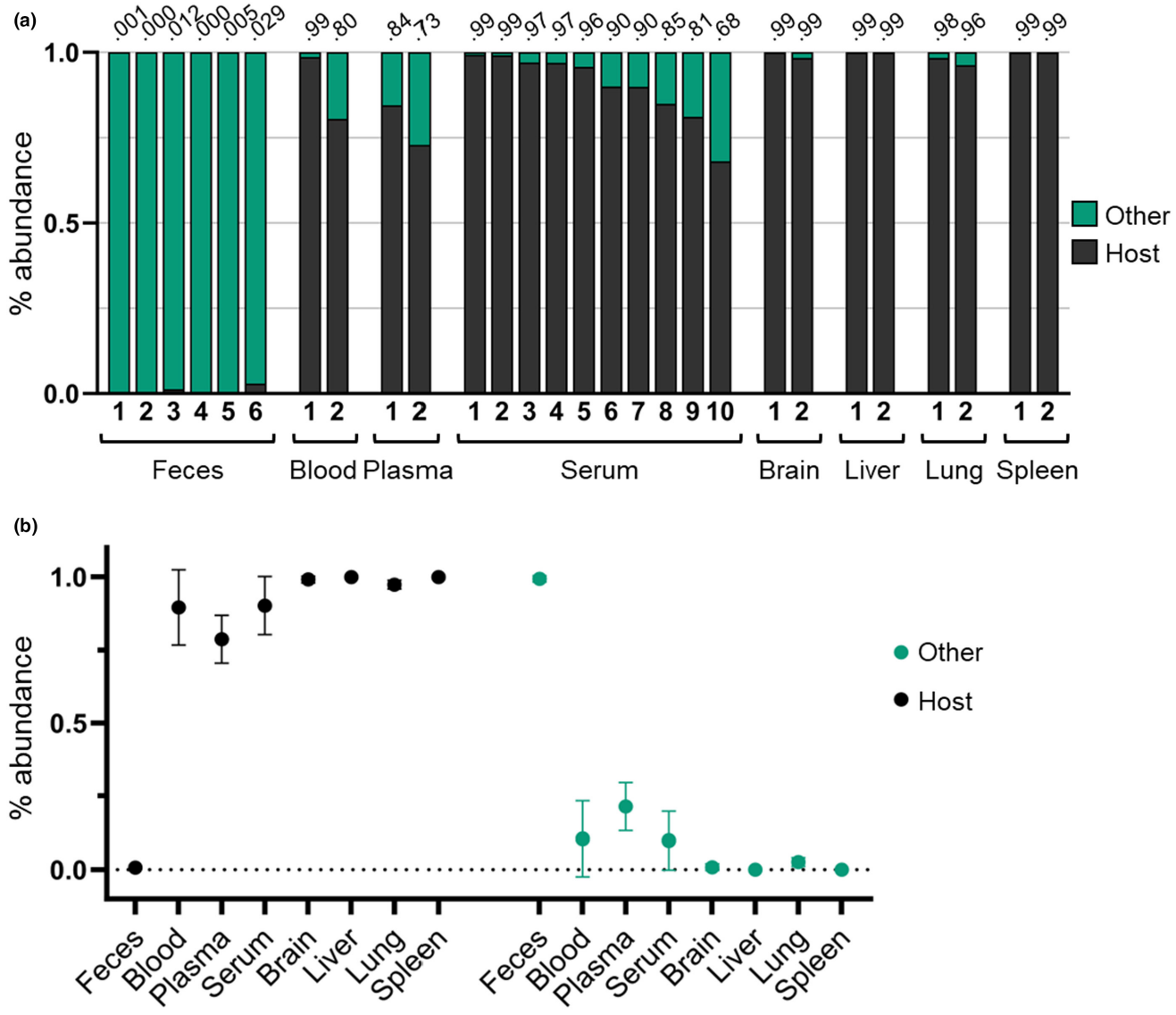

FIGURE 1 18S metabarcoding with PNA mammal blocker in chimpanzee samples. (a) Per cent relative abundance after quality filtering is shown for host reads (Host) and all other reads (Other). Numbers above bars represent percentage abundance of host reads. (b) Mean relative abundance after quality filtering ±SEM is shown for host reads (Host) and all other reads (Other). See Table S1 for source data.

resulting amplicon sequence variants (ASVs) with mafft using the q2-alignment plugin (Katoh et al., 2019) and constructed a phylogenetic tree with fasttree2 using the q2-phylogeny plugin (Price et al., 2010). Taxonomy was assigned to ASVs using the q2-feature-classifier (Bokulich et al., 2018) classify-sklearn naïve Bayes taxonomy classifier against the PR2 4.13.0 18S rRNA database (Guillou et al., 2013). Prism v.8.4.3 (GraphPad Software, Inc., La Jolla, CA, USA) was used for plotting data and conducting statistical analyses.

3 | RESULTS

3.1 | High host read abundance in 18S V4 metabarcoding data using a PNA clamp

18S V4 metabarcoding (Comeau et al., 2017; Mann et al., 2020) using DNA extracted from chimpanzee samples as input ($n = 28$) and including the published mammal-blocking PNA clamp in every amplification (Mann et al., 2020) yielded a wide range of host signal relative abundances (Figure 1a). The per cent abundance of host reads obtained was low in faecal samples (overall mean <1%) but high in all other sample types tested, including blood, plasma, serum, brain, liver, lung and spleen (overall mean = 93.5%; Table S1). Of non-faecal samples, plasma samples contained the lowest relative abundance of host reads (mean = 78.6%) and spleen samples contained the highest (mean = 99.9%; Figure 1b).

3.2 | Short guide RNA design for universal eukaryotic endosymbiont enrichment

We designed six candidate vertebrate host-specific sgRNAs targeting 18S V4 (Figure 2a), including one fortuitously identical to the published 18S V4 mammal-blocking PNA oligo used above (arb321; Table 1) (Mann et al., 2020). Target sites are located centrally in 18S V4 (Figure 2b) such that the digestion products can be differentiated from uncleaved amplicons based on size (Figure 2c).

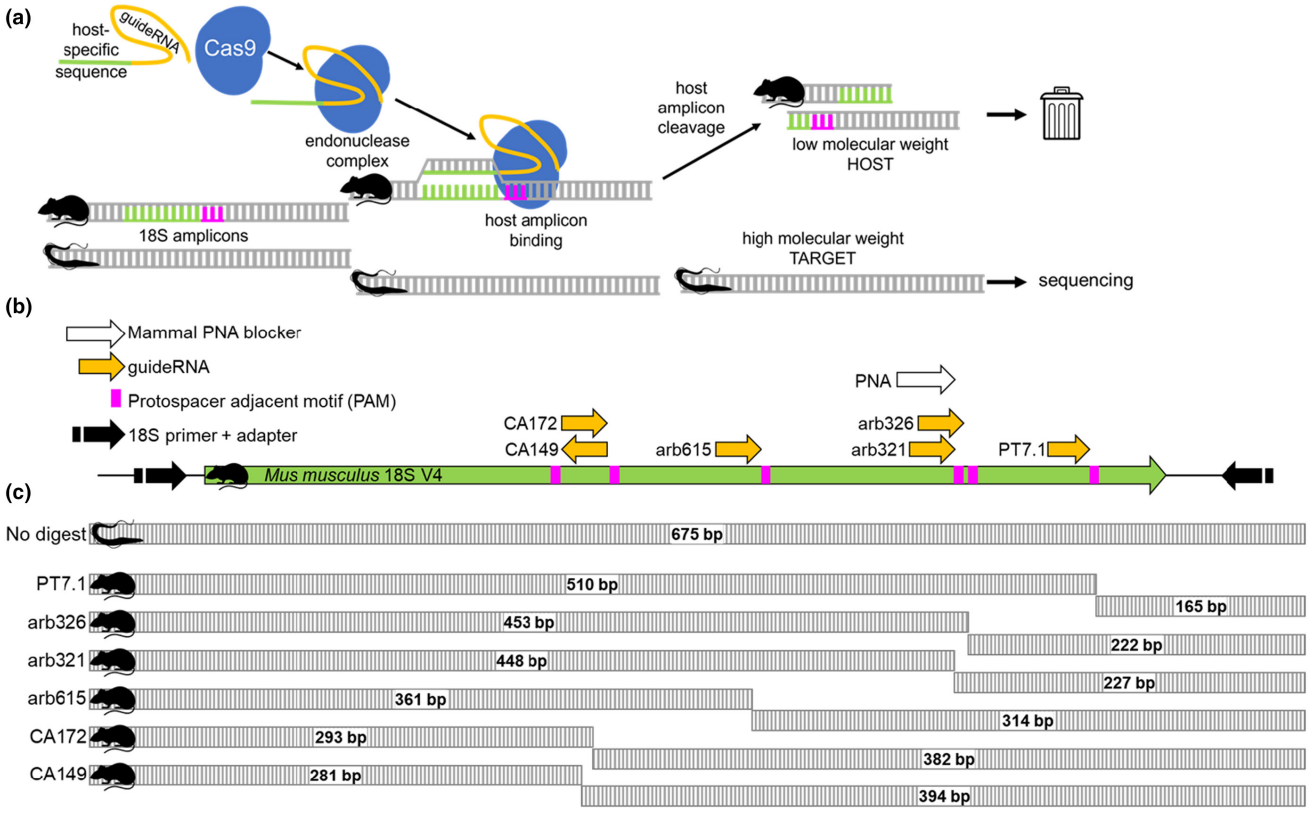

FIGURE 2 Overview of CRISPR-Cas9 host digestion method. (a) Schematic of steps in CRISPR-Cas9 in vitro digestion of host amplicons, but not target (protozoan) amplicons. (b) Map of representative mammal 18S rRNA gene (green region) from the house mouse *Mus musculus* (GenBank NR_003278) with locations of 18S amplicon primers (black arrows), newly designed short guide RNA (sgRNA) sequences (yellow arrows) and published PNA mammal blocker (white arrow). Protospacer adjacent motifs (PAMs) within the host 18S sequence are shown in pink. sgRNAs must bind next to a PAM sequence, and binding determines the location of cleavage by the Cas9 ribonucleoprotein complex. (c) Schematic of digestion products of host and target amplicons using sgRNAs with various complementarity sites. Topmost fragment (no digest) represents a target (protozoan) 18S V4 amplicon which is not recognized by the CC9 complex and remains full-length. The six bottommost fragments represent host (mouse) 18S V4 fragments recognized by the CC9 complex and cleaved. Labels to the left are sgRNA names. See Table 1 for sgRNA and PAM sequences.

Using in silico hybridization to the SILVA 138 RefNR database (Quast et al., 2013), we found all six candidates to have similar mammalian complementarity (Figure 3), with each hybridizing to 50% or more of mammalian sequences (mean = 66.4%) with no mismatches and 60% or more when allowing for a single mismatch outside of the seed sequence (mean = 76.4%). sgRNAs arb321 and arb326 were effective for mammalian hosts, but several gRNAs additionally recognized non-mammalian vertebrate groups, making them useful for a wider variety of hosts: arb615, CA149 and CA172 recognized mammal, bird and fish sequences, while PT7.1 recognized all vertebrates (Table 1). All six sgRNA oligos failed to hybridize to any parasite/endosymbiont group, with the sole exception of *Trichinella pseudospiralis* (mean = 17.8%; Figure 3) due to high 18S sequence similarity between *Trichinella* and mammals (mean = 45.5% DNA identity for all sgRNA target regions combined in *Trichinella pseudospiralis* AY851258; Table S2). We note that, despite the high sequence complementarity in this region, no available *Trichinella pseudospiralis* 18S V4 sequences contain the correct PAM sequence 'NGG' for CC9 cleavage (Table S2).

3.3 | CRISPR-Cas9 in vitro digestion selectively cleaves target organisms

In vitro digests of 18S V4 amplicons from single representative vertebrate hosts and eukaryotic endosymbionts corresponded to SILVA TestProbe predicted coverages (Figure 3) and fragment sizes (Figure 2b). For example, CC9 digestion with the 'mammal' arb321 sgRNA resulted in cleavage of mammal samples, but not amphibian, reptile, bird or fish samples, whereas digestion with the 'vertebrate' PT7.1 sgRNA resulted in cleavage of all five host samples including mammal, amphibian, reptile, bird and fish (Figure 4, left panel). All eukaryotic endosymbiont amplicons, including protozoans ($n = 2$), microsporidians ($n = 1$) and helminths ($n = 3$), were unaffected by CC9 digestion using any sgRNA (Figure 4, right panel).

3.4 | Evaluating host signal reduction methods

18S V4 metabarcoding using DNA extracted from chimpanzee samples as input ($n = 15$) with PNA blocker, CC9 digest with sgRNA

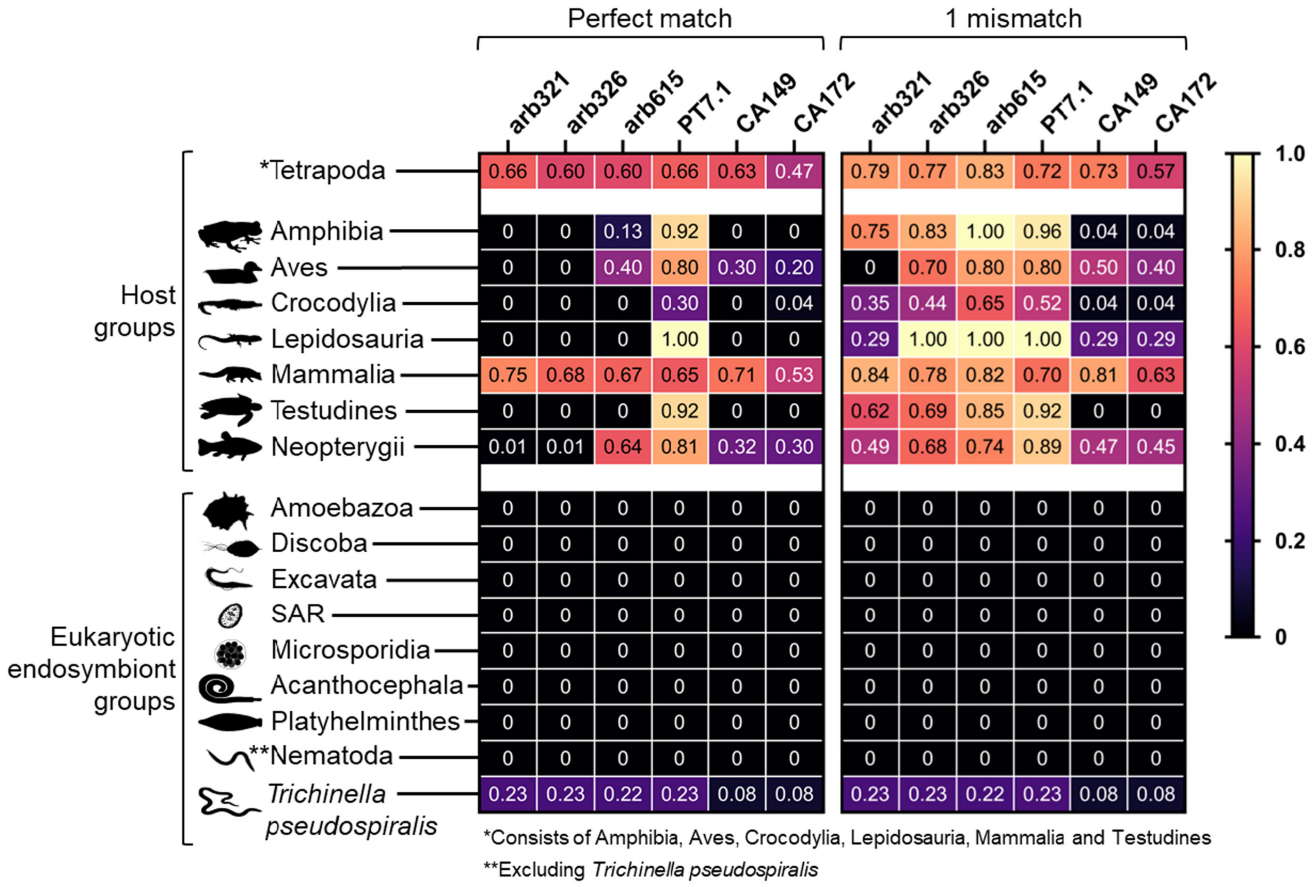

FIGURE 3 Short guide RNA in silico complementarity to host and eukaryotic endosymbiont groups. Per cent coverage of the SILVA 138 Ref NR database is shown with numbers and colour scale. Left panel, SILVA TestProbe with the most stringent settings (no mismatches, no N's considered as matches). Right panel, SILVA TestProbe allowing for a single mismatch outside of the conserved seed sequence. Taxonomic groups containing non-target 'Host' groups and target 'Eukaryotic endosymbiont' groups are shown with representative organism icons to the left of the heatmap. Tetrapoda* includes the 'Host' groups Amphibia, Aves, Crocodylia, Lepidosauria, Mammalia and Testudines. Nematoda** includes all nematode accessions other than *Trichinella pseudospiralis*. See Table 1 for sgRNA sequences.

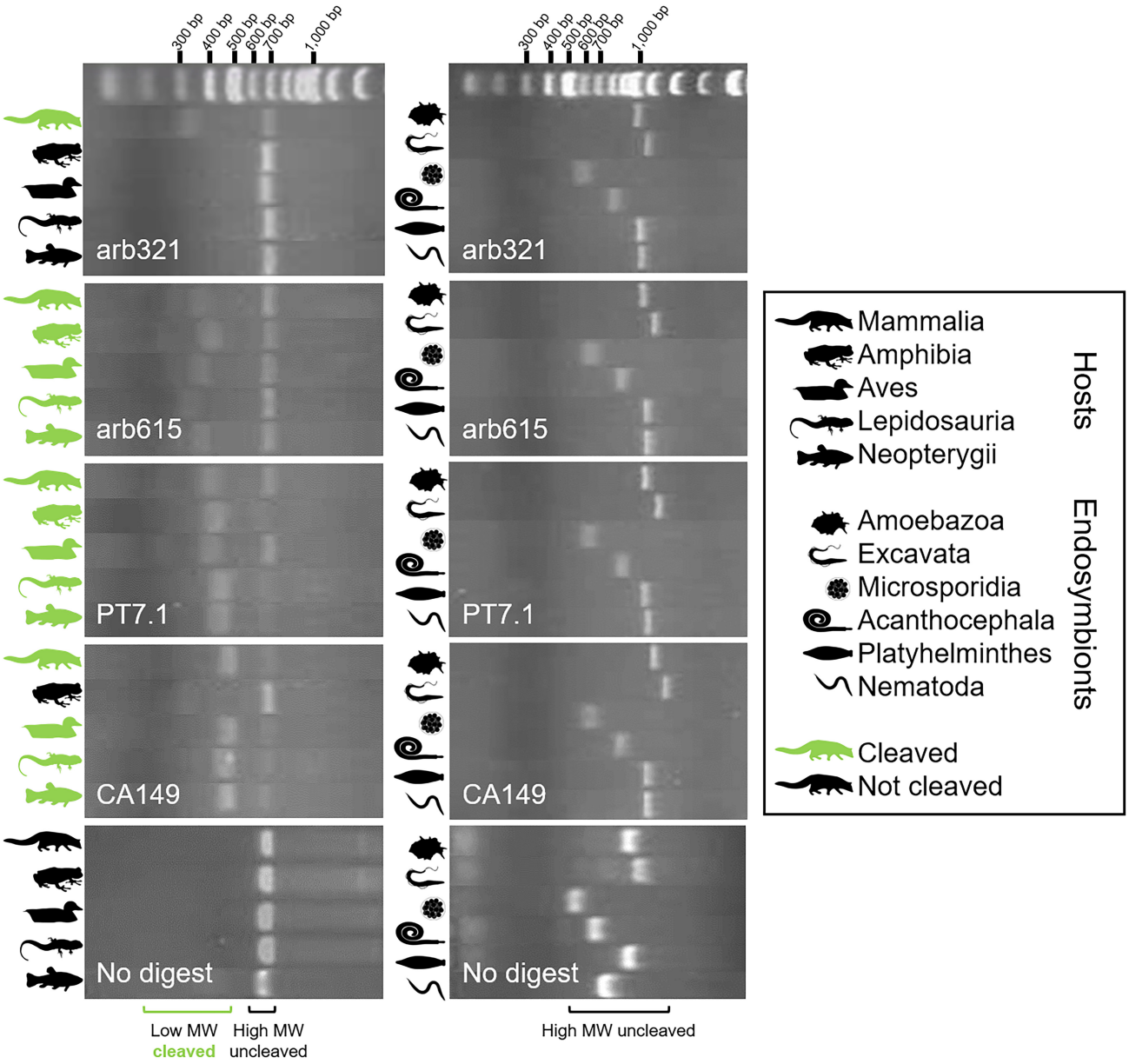

FIGURE 4 In vitro CRISPR-Cas9 digests of host and eukaryotic endosymbiont 18S V4 amplicons. Gel electrophoresis images show CRISPR-Cas9 digestion products or no digest controls (bottommost panels) of 18S V4 DNA amplified from vertebrate hosts (left panel) and eukaryotic endosymbiotic organisms (right panel) with the name of the guideRNA at the bottom left of each image. Sources of substrate DNA are shown as organism icons. Black icons represent organisms not cleaved by CRISPR-Cas9 digest with the specified guideRNA (or no digest control), and green icons represent organisms cleaved by CRISPR-Cas9 with the specified guideRNA. Organisms used for digest were: Mammalia—*Ursus maritimus* (polar bear), Amphibia—*Lithobates chiricahuensis* (leopard frog), Aves—*Gallus gallus* (chicken), Lepidosauria—*Varanus varius* (monitor lizard), Neopterygii—*Salmo trutta* (brown trout), Amoebazoa—*Entamoeba histolytica*, Excavata—*Trypanosoma brucei*, Microsporidia—*Encephalitozoon cuniculi*, Acanthocephala—*Echinorhynchus salmonis*, Platyhelminthes—*Schistosoma mansoni*, Nematoda—*Ascaris suum*. Topmost row is a DNA size standard. Note that 18S V4 amplicon length is variable among eukaryotic endosymbionts and that no eukaryotic endosymbiont amplicons were digested using any of the guideRNAs tested.

arb321, both PNA and CC9 digest, and no host signal reduction demonstrated CC9 digest to be the most effective method for enriching target read abundance for all sample types (blood, liver, lung, colon and faecal samples; Figure 5a; Table S3). Faecal samples yielded consistently low levels of host reads and were therefore not analysed further. In tissue samples (blood, liver, lung and colon), the overall percentage change in target (non-host) reads compared with no-treatment control was significantly higher for CC9 treatment (mean 58.7% increase in target reads, SEM 3.6%, range 37.2%–79.9%) compared with PNA (mean 1.5%, SEM 1.3%, range −7.1%–12.6%;

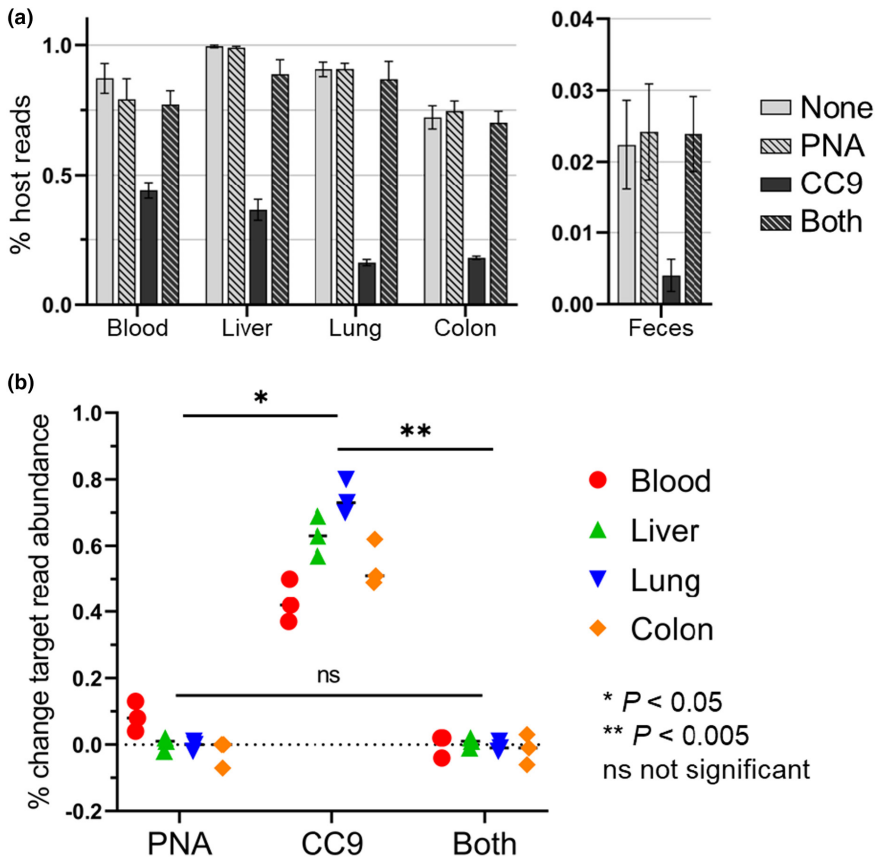

FIGURE 5 Methods comparison: Host signal reduction with mammal-blocking PNA oligo compared with CRISPR-Cas9 amplicon digest in 18S V4 metabarcoding. (a) Per cent abundance of host reads after quality filtering for five DNA samples metabarcoded under four conditions (triplicate mean): no host signal reduction used (None), published mammal-blocking PNA oligo added to amplicon PCR (PNA), CRISPR-Cas9/sgRNA arb321 digest of amplicons (CC9), and mammal-blocking PNA oligo added to amplicon PCR plus subsequent CRISPR-Cas9/sgRNA arb321 digest of amplicons (both). Note scale difference in tissues versus faecal sample. (b) Results from (a) displayed as per cent change in target (non-host) read abundance as compared with no-treatment control for all non-faecal samples. PNA, published mammal-blocking PNA oligo added to amplicon PCR; CC9, CRISPR-Cas9 digest of amplicons; both, mammal-blocking PNA oligo added to amplicon PCR plus subsequent CRISPR-Cas9 digest of amplicons. CC9 treatment is significantly different from PNA (paired *t*-test: $t = 6.94$, df = 3, $p = .0061$) and both (paired *t*-test: $t = 8.89$, df = 3, $p = .0030$). See Table S3 for source data.

paired *t*-test: $t = 6.94$, df = 3, $p = .0061$) or combination treatment (mean −0.2%, SEM 0.7%, range −5.6%–2.9%; paired *t*-test: $t = 8.89$, df = 3, $p = .0030$; Figure 5b).

3.5 | Optimization of CRISPR-Cas9 digest

We optimized the parameters of the CC9 digest by varying the ratio of ribonucleoprotein complex to target DNA PAM sequence and found that a ratio of 1:1 was most effective at lowering host signal (Figure 6a). To confirm the identity of the low molecular weight (MW) bands resulting from CC9 digest of mixed samples (containing both host and parasite DNA), we compared host read abundance in the higher- and lower-MW bands to show that the cleaved products are indeed of host origin (Figure 6b). We also evaluated the application of the CC9 digest before and after indexing PCR. There was no significant difference in digest efficiency for CC9 treatment applied to each individual amplicon prior to library preparation compared with CC9 applied to a library pool (paired *t*-test: $t = 0.38$, df = 30, $p = .18$; Figure 6c). Because the application of the digest after indexing is simpler and cheaper, we used this variation of the HSR protocol for all subsequent metabarcoding experiments.

18S V4 metabarcoding using newly designed sgRNAs (each in a separate sequencing library using the same starting material) demonstrated all sgRNAs to reduce host signal compared with untreated controls, with vertebrate sgRNA PT7.1 having the lowest abundance and mammal/bird/fish sgRNA CA172 having the highest (Figure 6d; Table S4). Further testing using the three top-performing sgRNAs (arb326, CA149 and PT7.1) showed that digestion with any of the three sgRNAs significantly reduced host reads compared with no-treatment controls (arb326 compared with none, paired *t*-test: $t = 282.2$, df = 30, $p < .0001$; CA149 compared with none, paired *t*-test: $t = 123.6$, df = 30, $p < .0001$; PT7.1 compared with none, paired *t*-test: $t = 370.3$, df = 30, $p < .001$). There was also a small, but significant difference in signal

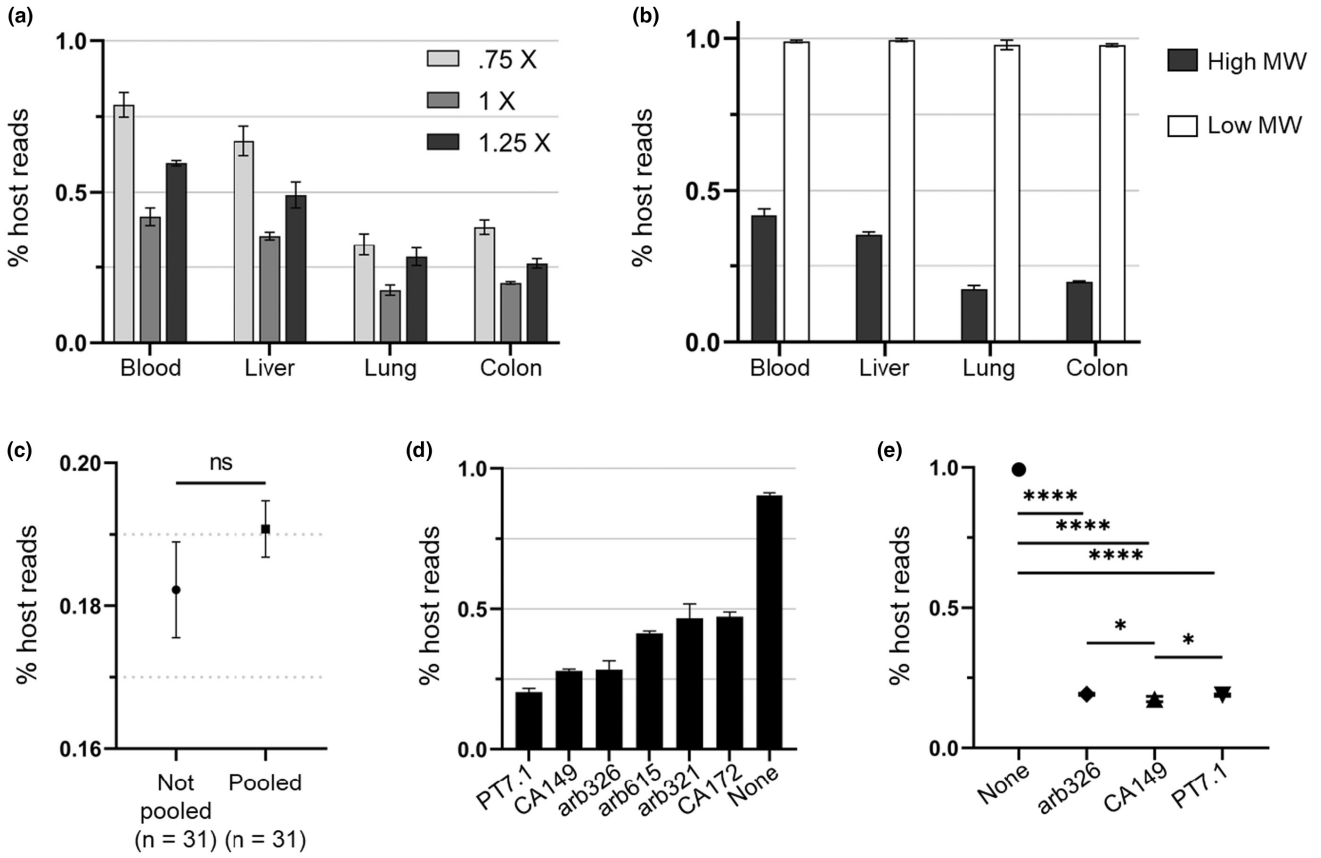

FIGURE 6 Characterization and optimization of CRISPR-Cas9-mediated host signal reduction in 18S V4 metabarcoding of chimpanzee blood and tissue samples. (a) CRISPR-Cas9 (CC9) reaction optimization. Per cent host read abundance (triplicate mean ± SEM) after quality filtering using varying ribonucleoprotein complex (RNP) to DNA target sequence ratios, where 1× represents a 1:1 ratio. (b) Identity of high and low molecular weight (MW) CC9 cleavage products. Per cent host read abundance (triplicate mean ± SEM) after quality filtering is shown for high and low MW bands extracted after separation by gel electrophoresis. (c) Comparison of CC9 digest before and after indexing PCR. Mean per cent host read abundance ±SEM after quality filtering is shown for CC9 digest (sgRNA PT7.1) applied to each amplicon prior to library preparation (not pooled) or to a single pool of amplicons after library preparation (Pooled). ns, not significant (paired *t*-test: $t = 1.38$, df = 30, $p = .18$). (d) Effect of short guide RNA (sgRNA) sequence on blood sample 18S V4 metabarcoding. Per cent host read abundance (triplicate mean ± SEM) after quality filtering is shown for 18S V4 amplicons that were not treated with any host signal reduction method (None) or digested with CRISPR-Cas9 using the specified sgRNA prior to library preparation. See Table S4 for source data. (e) Comparison of sgRNAs in blood sample metabarcoding. Mean per cent host reads abundance ±SEM after quality filtering is shown for three guideRNAs compared with no digest control. $^{*}p < .05$, $^{****}p < .0001$, all comparisons not shown are insignificant (paired *t*-test, df = 30 in all comparisons). See Table S4 for source data.

reduction among the three sgRNAs, with CA149 being most effective (CA149 compared with arb326, paired *t*-test: $t = 2.10$, df = 30, $p = .049$; CA149 compared with PT7.1, paired *t*-test: $t = 2.52$, df = 30, $p = .021$; Figure 6e; Table S4).

3.6 | CRISPR-Cas9 digest validation using known parasite infections of mammals

3.6.1 | *Dirofilaria immitis* in experimentally infected dogs

18S V4 metabarcoding of experimentally infected dog blood samples containing *Dirofilaria immitis* microfilariae (mean 57.8 microfilariae per 20 μL whole blood) demonstrated CC9 digestion to be more effective at host signal reduction than PNA blocking oligo or no treatment (Figure 7). Specifically, CC9-digested samples yielded a higher abundance of *Dirofilaria immitis* reads (mean of 6 sgRNAs = 52.41%, SEM = 3.28%, range: 40.06%–62.81%) than did PNA blocking oligo treatment (6.08%) or untreated control (9.77%). Intriguingly, CC9-digested samples also recovered reads from fungi and dietary items that were not detected by the other methods (Figure 7; Table S5).

3.6.2 | *Hepatocystis* in naturally infected red colobus

Data from wild red colobus blood samples demonstrated that, in untreated libraries, almost all reads were of host origin (mean = 99.9%) and no haemoparasites were detected. By contrast, CC9/sgRNA

CA149 treated libraries from the same samples had, on average, only 42.6% host reads and haemoparasites were detected in 17 of 19 samples (Figure 8; Table S6). These findings mirrored previous results from *Hepatocystis*-specific PCR and *cytochrome b* pyrosequencing of these same samples (Thurber et al., 2013), in which the same two species/lineages of *Hepatocystis* were detected: Species A in 13 of the 17 infected samples and Species B in 5 of the 17 infected samples (Table 2). One sample was positive by metabarcoding that was negative by PCR. Per cent agreement was low between PCR and metabarcoding without HSR treatment (Cohen's Kappa test: $\kappa = 0.0$, 95% CI from 0.0 to 0.0) and high between PCR and metabarcoding with CC9 digest (Cohen's Kappa test: $\kappa = 0.855$, 95% CI from 0.581 to 1.000). Overall application of CC9 digest increased agreement with PCR sixfold compared with no treatment (Table S7).

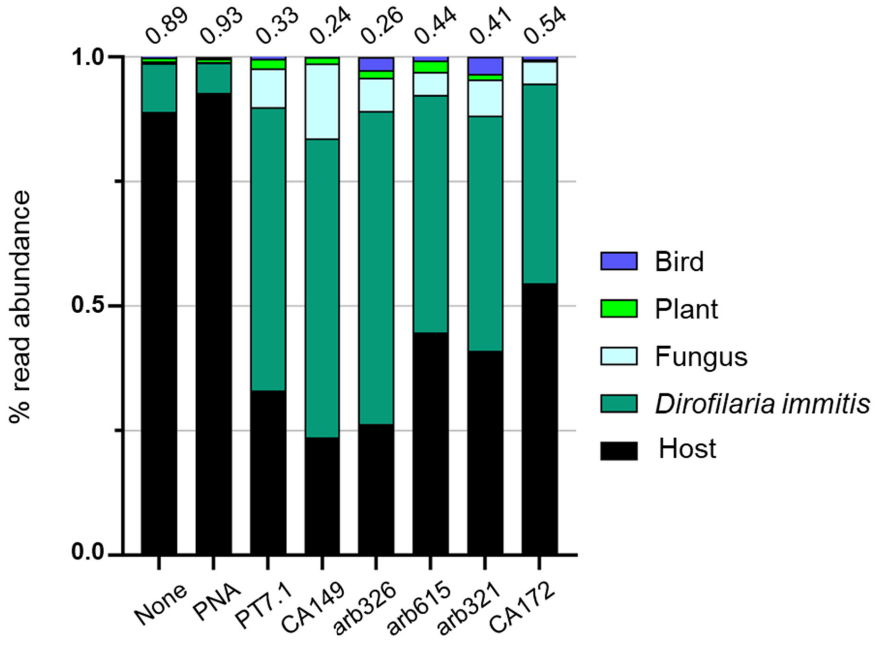

FIGURE 7 Effect of host signal reduction method on the detection of a known parasite infection. Dog blood infected with *Dirofilaria immitis* microfilariae was used as starting material for DNA extraction and 18S metabarcoding. Amplicons were untreated for host signal reduction (None), amplified with a PNA mammal blocker (PNA) or digested with CRISPR-Cas9 using the specified short guide RNAs (*X*-axis). Per cent abundance after quality filtering is shown, and numbers above bars represent total percentage host reads. See Table S5 for source data.

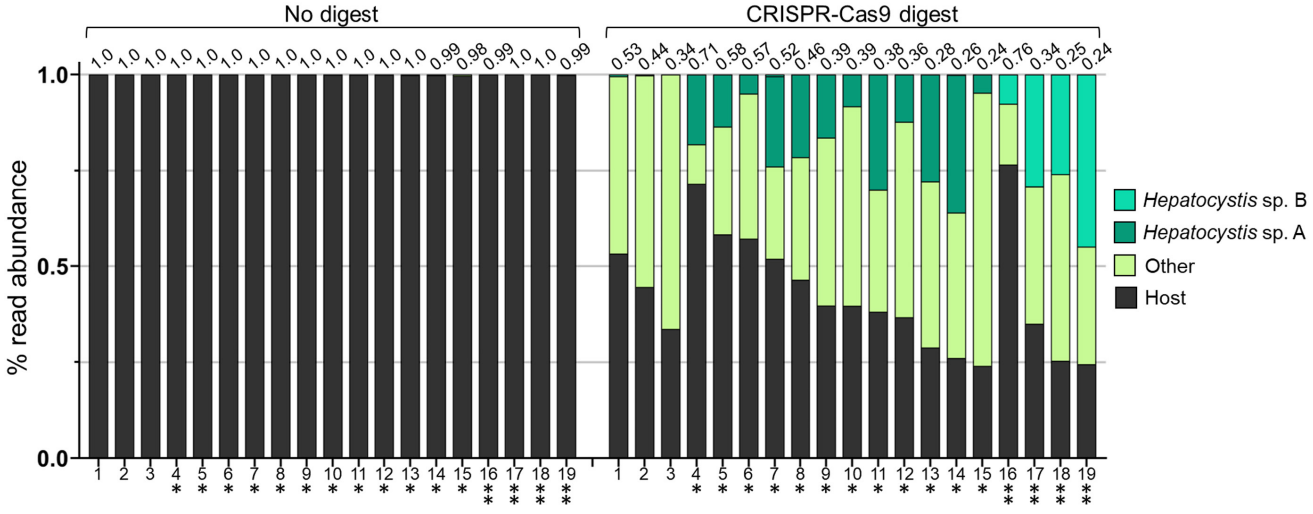

FIGURE 8 Effect of CRISPR-Cas9 host signal reduction on the detection of haemoparasite infection in wild red colobus blood samples. Metabarcoding data are shown as per cent read abundance after quality filtering for undigested (left panel) and CRISPR-Cas9 (sgRNA CA149) digested (right panel) amplicons using 19 samples. Reads are categorized as host, *Hepatocystis* spp., and all other reads (Other). Numbers above bars represent total % host reads per sample. No *Hepatocystis* spp. positives were detected by metabarcoding in undigested samples. Samples marked with a single asterisk were positive by genus-specific PCR/*cytochrome b* sequencing in a previous study for *Hepatocystis* sp. (A), and samples marked with a double asterisk were positive by genus-specific PCR/*cytochrome b* sequencing in the same previous study for *Hepatocystis* sp. (B). See Table S6 for source data.

TABLE 2 *Hepatocystis* detection by polymerase chain reaction (PCR) versus metabarcoding with and without CRISPR-Cas9 (CC9) digestion.

	PCR		Metabarcoding, no treatment		Metabarcoding, CC9 digest	
	Positive/Negative		% reads post-quality filtering		% reads post-quality filtering	
ID #	*Hepatocystis* sp. A	*Hepatocystis* sp. B	*Hepatocystis* sp. A	*Hepatocystis* sp. B	*Hepatocystis* sp. A	*Hepatocystis* sp. B
1	Negative	Negative	0	0	0.002	0
2	Negative	Negative	0	0	0	0
3	Negative	Negative	0	0	0	0
4	Positive	Negative	0	0	0.182	0
5	Positive	Negative	0	0	0.135	0
6	Positive	Negative	0	0	0.049	0
7	Positive	Negative	0	0	0.235	0.005
8	Positive	Negative	0	0	0.215	0
9	Positive	Negative	0	0	0.164	0
10	Positive	Negative	0	0	0.083	0
11	Positive	Negative	0	0	0.302	0
12	Positive	Negative	0	0	0.123	0
13	Positive	Negative	0	0	0.278	0
14	Positive	Negative	0	0	0.36	0
15	Positive	Negative	0	0	0.047	0
16	Negative	Positive	0	0	0	0.076
17	Negative	Positive	0	0	0	0.291
18	Negative	Positive	0	0	0	0.26
19	Negative	Positive	0	0	0	0.45

4 | DISCUSSION

Here, we show that a newly designed method using CRISPR-Cas9 and vertebrate host-targeted short guide RNAs was more effective at host signal reduction than PNA blocking or no treatment. Furthermore, in samples known from prior analyses to contain parasites, eukaryotic endosymbiont reads were rare or not detectable in samples treated with a PNA-blocking primer or not treated with any HSR method. However, when the new CC9 method was applied to these same samples, the parasites were detected at high read intensities. The new CC9 method also yielded reads matching two lineages of *Hepatocystis* previously characterized in red colobus using genus-specific PCR and *cytochrome b* pyrosequencing (Thurber et al., 2013).

PNA blocking and CRISPR-Cas9 digestion methods differ with respect to ease and cost. As of the time of writing, one company manufactures custom PNA oligos (PNA Bio, Newbury Park, CA, USA), but there many commercial sources for CRISPR-Cas9 reagents and custom guide RNAs (here we used Integrated DNA Technologies). The two methods cost approximately the same, and lead times for obtaining reagents are also comparable. PNA blockers are added to the amplification PCR directly, whereas CRSIPR-Cas9 requires an additional digest and gel extraction for size selection, but this additional digest may be performed after indexing and library pooling. Thus, only a single digest and a single gel extraction are required per sequencing run, minimizing the time and effort required.

The utility of the CC9 HSR method depends on the specificity of sgRNAs (Cho et al., 2014; Doench et al., 2016). We attempted to maximize specificity by designing sgRNAs using several complementary approaches and screening a large pool of 100 candidate oligos to identify six final sgRNA sequences. We then rigorously evaluated these six oligos in silico and in laboratory experiments using genomic DNA from individual eukaryotic organisms and from clinical samples infected with eukaryotic parasites. The consistency of our results across these conditions strongly suggests that the CC9 method is specific, effective and robust. We note, however, that 8%–23% of sequences from the nematode parasite *Trichinella pseudospiralis* were highly similar to the mammalian 18S V4 region CC9 recognition sites, although no *Trichinella pseudospiralis* sequences contain a perfect PAM (NGG), which is required for Cas9 cleavage. An alternative CRISPR system than described here with a different PAM site could therefore introduce problematic cross-reactivity. If *Trichinella* is suspected, we recommend in silico analysis to verify host complementarity prior to choosing a particular sgRNA and/or CRISPR system.

A distinct advantage of our method is that it does not depend on the PCR primers used to amplify the 18S V4 region, as long as those primers flank the site of sgRNA complementarity. Therefore, any amplicon including the 18S V4 region is compatible with all sgRNA oligos presented here. We note that we recently published a new set of eukaryotic endosymbiont metabarcoding primers that out-performs all other published primer sets in terms of taxonomic breath, on-target

amplification and unbiased reconstruction of eukaryotic communities (Owens et al., 2023). We have examined this primer set in conjunction with the CC9 protocol described herein, and in combination the two methods achieve a similar reduction in host signal as this study (82% less host reads compared with no treatment and 74% compared with PNA clamp in blood samples; unpublished data). Also, because 18S V4 has the highest entropy of the hypervariable regions constituting 18S (Bradley et al., 2016; Pinol et al., 2019) and, thus, the highest taxonomic resolution, we expect our sgRNAs designs to stay relevant for as long as this locus is used for eukaryotic endosymbiont metabarcoding.

Overall, we have shown that CRISPR-Cas9 digestion of amplicons reduces host signal sufficiently to allow for the detection of rare eukaryotic endosymbionts and thus to increase the sensitivity and efficiency of eukaryotic endosymbiont metabarcoding. Our new method should help advance the fields of parasitology and eukaryotic community ecology, similar to how 16S prokaryote metabarcoding has facilitated the study of the microbiome.

AUTHOR CONTRIBUTIONS

L.A.O. designed the study, performed all the laboratory work, and collected and analysed the data. M.I.T. and T.L.G. obtained and analysed the samples as part of a previous study. L.A.O. and T.L.G. wrote the manuscript. All authors made substantive intellectual contributions, revised the manuscript and approved the final draft.

ACKNOWLEDGEMENTS

We gratefully acknowledge Lyric Bartholomay for assistance in sample acquisition, and we thank the NIH/NIAID Filariasis Research Reagent Resource Center (www.filariasiscenter.org) for providing microfilaremic dog blood samples. We thank the Uganda Wildlife Authority, the Uganda National Council for Science and Technology, and the Tacugama Chimpanzee Sanctuary in Sierra Leone for granting permission to conduct the original studies from which we obtained excess samples for this work. This research was funded by the National Institutes of Health through grants 1R01AG049395-01, 1R21AI163592-0 and T32AI007414 (University of Wisconsin-Madison Parasitology and Vector Biology Program), and by the University of Wisconsin-Madison through the John D. MacArthur Professorship Chair.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Raw sequence reads are deposited in the NCBI BioSample Database (BioProject PRJNA1016069) under accession nos.: SAMN37367571 – SAMN37367614.

BENEFIT-SHARING STATEMENT

Benefits from this research accrue from the sharing of our data and results on public databases as described above.

ORCID

Leah A. Owens https://orcid.org/0000-0002-4232-3516

REFERENCES

Alberdi, A., Aizpurua, O., Gilbert, M. T. P., & Bohmann, K. (2018). Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods in Ecology and Evolution*, *9*(1), 134–147.

Belda, E., Coulibaly, B., Fofana, A., Beavogui, A. H., Traore, S. F., Gohl, D. M., Vernick, K. D., & Riehle, M. M. (2017). Preferential suppression of *Anopheles gambiae* host sequences allows detection of the mosquito eukaryotic microbiome. *Scientific Reports*, *7*(1), 3241. https://doi.org/10.1038/s41598-017-03487-1

Bokulich, N. A., Kaehler, B. D., Rideout, J. R., Dillon, M., Bolyen, E., Knight, R., Huttley, G. A., & Gregory Caporaso, J. (2018). Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome*, *6*(1), 90. https://doi.org/10.1186/s40168-018-0470-z

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., … Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, *37*(8), 852–857. https://doi.org/10.1038/s41587-019-0209-9

Bradley, I. M., Pinto, A. J., & Guest, J. S. (2016). Design and evaluation of Illumina MiSeq-compatible, 18S rRNA gene-specific primers for improved characterization of mixed phototrophic communities. *Applied and Environmental Microbiology*, *82*(19), 5878–5891. https://doi.org/10.1128/Aem.01630-16

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

Cao, J., Wu, L., Zhang, S. M., Lu, M., Cheung, W. K., Cai, W., Gale, M., Xu, Q., & Yan, Q. (2016). An easy and efficient inducible CRISPR/Cas9 platform with improved specificity for multiple gene targeting. *Nucleic Acids Research*, *44*(19), e149. https://doi.org/10.1093/nar/gkw660

Cho, S. W., Kim, S., Kim, Y., Kweon, J., Kim, H. S., Bae, S., & Kim, J. S. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Research*, *24*(1), 132–141. https://doi.org/10.1101/gr.162339.113

Comeau, A. M., Douglas, G. M., & Langille, M. G. (2017). Microbiome helper: A custom and streamlined workflow for microbiome research. *mSystems*, *2*(1), e00127-16. https://doi.org/10.1128/mSystems.00127-16

Comeau, A. M., Li, W. K., Tremblay, J. E., Carmack, E. C., & Lovejoy, C. (2011). Arctic Ocean microbial community structure before and after the 2007 record sea ice minimum. *PLoS One*, *6*(11), e27492. https://doi.org/10.1371/journal.pone.0027492

Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., Virgin, H. W., Listgarten, J., & Root, D. E. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology*, *34*(2), 184–191. https://doi.org/10.1038/nbt.3437

Feehery, G. R., Yigit, E., Oyola, S. O., Langhorst, B. W., Schmidt, V. T., Stewart, F. J., Dimalanta, E. T., Amaral-Zettler, L. A., Davis, T., Quail, M. A., & Quail, M. A. (2013). A method for selectively enriching microbial DNA from contaminating vertebrate host DNA. *PLoS One*, *8*(10), e76096.

Flaherty, B. R., Barratt, J., Lane, M., Talundzic, E., & Bradbury, R. S. (2021). Sensitive universal detection of blood parasites by

selective pathogen-DNA enrichment and deep amplicon sequencing. *Microbiome*, *9*(1), 1. https://doi.org/10.1186/s40168-020-00939-1

Flaherty, B. R., Talundzic, E., Barratt, J., Kines, K. J., Olsen, C., Lane, M., Sheth, M., & Bradbury, R. S. (2018). Restriction enzyme digestion of host DNA enhances universal detection of parasitic pathogens in blood via targeted amplicon deep sequencing. *Microbiome*, *6*(1), 164. https://doi.org/10.1186/s40168-018-0540-2

Forsman, A. M., Savage, A. E., Hoenig, B. D., & Gaither, M. R. (2022). DNA Metabarcoding across disciplines: Sequencing our way to greater understanding across scales of biological organization. *Integrative and Comparative Biology*, *62*(2), 191–198. https://doi.org/10.1093/icb/icac090

Gu, W., Crawford, E. D., O'Donovan, B. D., Wilson, M. R., Chow, E. D., Retallack, H., & DeRisi, J. L. (2016). Depletion of abundant sequences by hybridization (DASH): Using Cas9 to remove unwanted high-abundance species in sequencing libraries and molecular counting applications. *Genome Biology*, *17*, 41. https://doi.org/10.1186/s13059-016-0904-5

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., de Vargas, C., Decelle, J., Del Campo, J., Dolan, J. R., Dunthorn, M., Edvardsen, B., Holzmann, M., Kooistra, W. H., Lara, E., Le Bescot, N., Logares, R., … Christen, R. (2013). The protist ribosomal reference database (PR2): A catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Research*, *41*(Database issue), D597–D604. https://doi.org/10.1093/nar/gks1160

Hamady, M., & Knight, R. (2009). Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. *Genome Research*, *19*(7), 1141–1152. https://doi.org/10.1101/gr.085464.108

Hino, A., Maruyama, H., & Kikuchi, T. (2016). A novel method to assess the biodiversity of parasites using 18S rDNA Illumina sequencing; parasitome analysis method. *Parasitology International*, *65*(5), 572–575.

Jiang, P., Lai, S., Wu, S., Zhao, X. M., & Chen, W. H. (2020). Host DNA contents in fecal metagenomics as a biomarker for intestinal diseases and effective treatment. *BMC Genomics*, *21*(1), 348. https://doi.org/10.1186/s12864-020-6749-z

Katoh, K., Rozewicki, J., & Yamada, K. D. (2019). MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, *20*(4), 1160–1166. https://doi.org/10.1093/bib/bbx108

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., & Glockner, F. O. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research*, *41*(1), e1. https://doi.org/10.1093/nar/gks808

Laforest-Lapointe, I., & Arrieta, M. C. (2018). Microbial eukaryotes: A missing link in gut microbiome studies. *mSystems*, *3*(2), e00201-17. https://doi.org/10.1128/mSystems.00201-17

Lappan, R., Classon, C., Kumar, S., Singh, O. P., de Almeida, R. V., Chakravarty, J., Kumari, P., Kansal, S., Sundar, S., & Blackwell, J. M. (2019). Meta-taxonomic analysis of prokaryotic and eukaryotic gut flora in stool samples from visceral leishmaniasis cases and endemic controls in Bihar state India. *PLoS Neglected Tropical Diseases*, *13*(9), e0007444. https://doi.org/10.1371/journal.pntd.0007444

Lin, W., Tian, T., Jiang, Y., Xiong, E., Zhu, D., & Zhou, X. (2021). A CRISPR/Cas9 eraser strategy for contamination-free PCR end-point detection. *Biotechnology and Bioengineering*, *118*(5), 2053–2066. https://doi.org/10.1002/bit.27718

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, Buchner, A., Lai, T., Steppi, S., Jobb, G., Förster, W., Brettske, I., Gerber, S., Ginhart, A. W., Gross, O., Grumann, S., Hermann, S., Jost, R., König, A., … Schleifer, K. H. (2004). ARB: A software environment for sequence data. *Nucleic Acids Research*, *32*(4), 1363–1371. https://doi.org/10.1093/nar/gkh293

Lundberg, D. S., Yourstone, S., Mieczkowski, P., Jones, C. D., & Dangl, J. L. (2013). Practical innovations for high-throughput amplicon sequencing. *Nature Methods*, *10*(10), 999–1002. https://doi.org/10.1038/nmeth.2634

Mamanova, L., Coffey, A. J., Scott, C. E., Kozarewa, I., Turner, E. H., Kumar, A., Howard, E., Shendure, J., & Turner, D. J. (2010). Target-enrichment strategies for next-generation sequencing. *Nature Methods*, *7*(2), 111–118.

Mann, A. E., Mazel, F., Lemay, M. A., Morien, E., Billy, V., Kowalewski, M., Di Fiore, A., Link, A., Goldberg, T. L., Tecot, S., Baden, A. L., Gomez, A., Sauther, M. L., Cuozzo, F. P., Rice, G. A. O., Dominy, N. J., Stumpf, R., Lewis, R. J., Swedell, L., … Wegener Parfrey, L. (2020). Biodiversity of protists and nematodes in the wild nonhuman primate gut. *ISME Journal*, *14*(2), 609–622. https://doi.org/10.1038/s41396-019-0551-4

Mayer, T., Mari, A., Almario, J., Murillo-Roos, M., Abdullah, M., Dombrowski, N., Dombrowski, N., Hacquard, S., Kemen, E. M., & Agler, M. T. (2020). Obtaining deeper insights into microbiome diversity using a simple method to block host and non-targets in amplicon sequencing. *bioRxiv*, https://doi.org/10.1101/2020.10.05.322305

Nicolas, L., Prina, E., Lang, T., & Milon, G. (2002). Real-time PCR for detection and quantitation of leishmania in mouse tissues. *Journal of Clinical Microbiology*, *40*(5), 1666–1669. https://doi.org/10.1128/JCM.40.5.1666-1669.2002

O'Rorke, R., Lavery, S., & Jeffs, A. (2012). PCR enrichment techniques to identify the diet of predators. *Molecular Ecology Resources*, *12*(1), 5–17. https://doi.org/10.1111/j.1755-0998.2011.03091.x

Owens, L. A., Colitti, B., Hirji, I., Pizarro, A., Jaffe, J. E., Moittie, S., Bishop-Lilly, K. A., Estrella, L. A., Voegtly, L. J., Kuhn, J. H., Suen, G., Deblois, C. L., Dunn, C. D., Juan-Sallés, C., & Goldberg, T. L. (2021). A *Sarcina* bacterium linked to lethal disease in sanctuary chimpanzees in Sierra Leone. *Nature Communications*, *12*(1), 763. https://doi.org/10.1038/s41467-021-21012-x

Owens, L. A., Friant, S., Martorelli Di Genova, B., Knoll, L. J., Contreras, M., Noya-Alarcon, O., Dominguez-Bello, M. G., & Goldberg, T. L. (2023). VESPA: An optimized protocol for accurate metabarcoding-based characterization of vertebrate eukaryotic endosymbiont and parasite assemblages. *Nature Communications*, *15*, 402.

Pinol, J., Senar, M. A., & Symondson, W. O. C. (2019). The choice of universal primers and the characteristics of the species mixture determine when DNA metabarcoding can be quantitative. *Molecular Ecology*, *28*(2), 407–419. https://doi.org/10.1111/mec.14776

Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS One*, *5*(3), e9490. https://doi.org/10.1371/journal.pone.0009490

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glockner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, *41*(Database issue), D590–D596. https://doi.org/10.1093/nar/gks1219

Sakai, M., & Ikenaga, M. (2013). Application of peptide nucleic acid (PNA)-PCR clamping technique to investigate the community structures of rhizobacteria associated with plant roots. *Journal of Microbiological Methods*, *92*(3), 281–288. https://doi.org/10.1016/j.mimet.2012.09.036

Song, L., & Xie, K. (2020). Engineering CRISPR/Cas9 to mitigate abundant host contamination for 16S rRNA gene-based amplicon sequencing. *Microbiome*, *8*(1), 80. https://doi.org/10.1186/s40168-020-00859-0

Sow, A., Brevault, T., Benoit, L., Chapuis, M. P., Galan, M., Coeur d'acier, A., Delvare, G., Sembène, M., & Haran, J. (2019). Deciphering host-parasitoid interactions and parasitism rates of crop pests using DNA metabarcoding. *Scientific Reports*, *9*(1), 3646. https://doi.org/10.1038/s41598-019-40243-z

Tedersoo, L., Bahram, M., Zinger, L., Nilsson, R. H., Kennedy, P. G., Yang, T., Anslan, S., & Mikryukov, V. (2022). Best practices in metabarcoding of fungi: From experimental design to results. *Molecular Ecology*, *31*(10), 2769–2795. https://doi.org/10.1111/mec.16460

Terahara, T., Chow, S., Kurogi, H., Lee, S. H., Tsukamoto, K., Mochioka, N., Tanaka, H., & Takeyama, H. (2011). Efficiency of peptide nucleic acid-directed PCR clamping and its application in the investigation of natural diets of the Japanese eel leptocephali. *PLoS One*, *6*(11), e25715. https://doi.org/10.1371/journal.pone.0025715

Thurber, M. I., Ghai, R. R., Hyeroba, D., Weny, G., Tumukunde, A., Chapman, C. A., Wiseman, R. W., Dinis, J., Steeil, J., Greiner, E. C., Friedrich, T. C., O'Connor, D. H., & Goldberg, T. L. (2013). Co-infection and cross-species transmission of divergent hepatocystis lineages in a wild African primate community. *International Journal for Parasitology*, *43*(8), 613–619. https://doi.org/10.1016/j.ijpara.2013.03.002

Titus, R. G., Marchand, M., Boon, T., & Louis, J. A. (1985). A limiting dilution assay for quantifying *Leishmania major* in tissues of infected mice. *Parasite Immunology*, *7*(5), 545–555. https://doi.org/10.1111/j.1365-3024.1985.tb00098.x

Troedsson, C., Lee, R. F., Stokes, V., Walters, T. L., Simonelli, P., & Frischer, M. E. (2008). Development of a denaturing high-performance liquid chromatography method for detection of protist parasites of metazoans. *Applied and Environmental Microbiology*, *74*(14), 4336–4345.

Vestheim, H., Deagle, B. E., & Jarman, S. N. (2011). Application of blocking oligonucleotides to improve signal-to-noise ratio in a PCR. *Methods in Molecular Biology*, *687*, 265–274. https://doi.org/10.1007/978-1-60761-944-4_19

Vestheim, H., & Jarman, S. N. (2008). Blocking primers to enhance PCR amplification of rare sequences in mixed samples – a case study on prey DNA in Antarctic krill stomachs. *Frontiers in Zoology*, *5*, 12. https://doi.org/10.1186/1742-9994-5-12

von Wintzingerode, F., Landt, O., Ehrlich, A., & Göbel, U. B. (2000). Peptide nucleic acid-mediated PCR clamping as a useful supplement in the determination of microbial diversity. *Applied and Environmental Microbiology*, *66*(2), 549–557.

Wu, J., Huang, B., Chen, H., Yin, Q., Liu, Y., Xiang, Y., Zhang, B., Liu, B., Wang, Q., Xia, W., Li, W., Li, Y., Ma, J., Peng, X., Zheng, H., Ming, J., Zhang, W., Zhang, J., Tian, G., … Xie, W. (2016). The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature*, *534*(7609), 652–657. https://doi.org/10.1038/nature18606

Wu, X., Kriz, A. J., & Sharp, P. A. (2014). Target specificity of the CRISPR-Cas9 system. *Quantitative Biology*, *2*(2), 59–70. https://doi.org/10.1007/s40484-014-0030-x

SUPPORTING INFORMATION

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

How to cite this article: Owens, L. A., Thurber, M. I., & Goldberg, T. L. (2024). CRISPR-Cas9-mediated host signal reduction for 18S metabarcoding of host-associated eukaryotes. *Molecular Ecology Resources*, *00*, e13980. https://doi.org/10.1111/1755-0998.13980