



# Metabarcoding of eukaryotic parasite communities describes diverse parasite assemblages spanning the primate phylogeny

Jan F. Gogarten<sup>1,2</sup> | Sébastien Calvignac-Spencer<sup>1,2</sup> | Charles L. Nunn<sup>3,4</sup> |  
Markus Ulrich<sup>1</sup> | Nasrin Saiepour<sup>5</sup> | Henrik Vedel Nielsen<sup>6</sup> | Tobias Deschner<sup>7</sup> |  
Claudia Fichtel<sup>8,9</sup> | Peter M. Kappeler<sup>8,9,10</sup> | Sascha Knauf<sup>11</sup> | Nadine Müller-Klein<sup>12</sup> |  
Julia Ostner<sup>9,12,13</sup> | Martha M. Robbins<sup>7</sup> | Somboon Sangmaneedet<sup>14</sup> |  
Oliver Schülke<sup>9,12,13</sup> | Martin Surbeck<sup>7,15</sup> | Roman M. Wittig<sup>7,16</sup> | Alexander Sliwa<sup>17</sup> |  
Christina Strube<sup>18</sup> | Fabian H. Leendertz<sup>1</sup> | Christian Roos<sup>5,19</sup> | Angela Noll<sup>5</sup>

<sup>1</sup>Project Group 3: Epidemiology of Highly Pathogenic Microorganisms, Robert Koch Institute Berlin, Berlin, Germany

<sup>2</sup>Viral Evolution, Robert Koch-Institute Berlin, Berlin, Germany

<sup>3</sup>Department of Evolutionary Anthropology, Duke University, Durham, NC, USA

<sup>4</sup>Duke Global Health Institute, Duke University, Durham, NC, USA

<sup>5</sup>Primate Genetics Laboratory, German Primate Center, Leibniz Institute for Primate Research, Goettingen, Germany

<sup>6</sup>Department of Bacteria, Parasites and Fungi, Statens Serum Institut, Copenhagen, Denmark

<sup>7</sup>Department of Primatology, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

<sup>8</sup>Behavioral Ecology & Sociobiology Unit, German Primate Center, Goettingen, Germany

<sup>9</sup>Leibniz Science Campus Primate Cognition, Goettingen, Germany

<sup>10</sup>Department of Sociobiology/Anthropology, Johann-Friedrich-Blumenbach Institute for Zoology, Georg-August University, Goettingen, Germany

<sup>11</sup>Neglected Tropical Diseases Work Group, Infection Biology Unit, German Primate Center, Leibniz Institute for Primate Research, Goettingen, Germany

<sup>12</sup>Department of Behavioral Ecology, University of Goettingen, Goettingen, Germany

<sup>13</sup>Research Group Primate Social Evolution, German Primate Center - Leibniz Institute for Primate Research, Goettingen, Germany

<sup>14</sup>Department of Pathobiology, Faculty of Veterinary Medicine, Khon Kaen University, Khon Kaen, Thailand

<sup>15</sup>Department of Human Evolutionary Biology, Harvard University, Cambridge, MA, USA

<sup>16</sup>Taï Chimpanzee Project, Centre Suisse de Recherches Scientifiques, Abidjan, Ivory Coast

<sup>17</sup>Koelner Zoo, Koeln, Germany

<sup>18</sup>Institute for Parasitology, Centre for Infection Medicine, University of Veterinary Medicine Hannover, Hannover, Germany

<sup>19</sup>Gene Bank of Primates, German Primate Center, Leibniz Institute for Primate Research, Goettingen, Germany

## Correspondence

Fabian H. Leendertz, Christian Roos, and  
Angela Noll Project Group 3: Epidemiology  
of Highly Pathogenic Microorganisms,  
Robert Koch Institute Berlin, Berlin,  
Germany.  
Email: leendertzf@rki.de; CROOS@dpz.eu;  
ANoll@dpz.eu

## Abstract

Despite their ubiquity, in most cases little is known about the impact of eukaryotic parasites on their mammalian hosts. Comparative approaches provide a powerful method to investigate the impact of parasites on host ecology and evolution, though two issues are critical for such efforts: controlling for variation in methods of identifying parasites and incorporating heterogeneity in sampling effort across host species. To address these issues, there is a need for standardized methods to catalogue eukaryotic parasite diversity across broad phylogenetic host ranges. We demonstrate

**Present address**

Nadine Müller-Klein, The Institute of Evolutionary Ecology and Conservation Genomics, Ulm University, Ulm, Germany

**Funding information**

Deutsche Forschungsgemeinschaft, Grant/Award Number: 1135/3-1, DE 1135/2-1, FI 929/7-1, FOR2136, KR 3834/5-1, Ka 1082-28-1, OS201/6-1, OS201/6-2, SCHU 1554/6-1, WI 2637/3-1 and WI 2637/4-2; Max Planck Society

the feasibility of a metabarcoding approach for describing parasite communities by analysing faecal samples from 11 nonhuman primate species representing divergent lineages of the primate phylogeny and the full range of sampling effort (i.e. from no parasites reported in the literature to the best-studied primates). We detected a number of parasite families and regardless of prior sampling effort, metabarcoding of only ten faecal samples identified parasite families previously undescribed in each host ( $\bar{x} = 8.5$  new families per species). We found more overlap between parasite families detected with metabarcoding and published literature when more research effort—measured as the number of publications—had been conducted on the host species' parasites. More closely related primates and those from the same continent had more similar parasite communities, highlighting the biological relevance of sampling even a small number of hosts. Collectively, results demonstrate that metabarcoding methods are sensitive and powerful enough to standardize studies of eukaryotic parasite communities across host species, providing essential new tools for macroecological studies of parasitism.

**KEY WORDS**

comparative method, gastrointestinal parasites, next generation sequencing, parasite communities

## 1 | INTRODUCTION

Mammals host a large diversity of eukaryotic parasites (hereafter, parasites) that can have a major impact on their reproduction and survival (Charlier, van der Voort, Kenyon, Skuce, & Vercruyse, 2014; Pullan & Brooker, 2008). The comparison of mammal/parasite systems is therefore likely to unveil important ecological and evolutionary processes, from both parties' perspective. Comparative approaches, however, require that the data are obtained with standardized methods across host populations and species to ensure comparability and valid inference of (macro) ecological and evolutionary patterns (Harvey & Pagel, 1991; Nunn, 2011).

Two issues are particularly critical to consider when conducting comparative research on parasitism: variation in methods of identifying parasites and controlling for heterogeneity in sampling effort. First, with regard to identifying parasite species, morphological classification of parasites has been used for parasite classification, but collecting suitable samples is challenging and the analysis requires extensive expertise and training (Seesao et al., 2017). In some cases, morphological features for distinguishing species are not available, especially if only parasite reproductive stages (e.g. eggs, oocysts or larvae) are available (Decraemer, Brown, Karanastasi, Zheng, & De Almeida, 2001). Biochemical approaches, such as multilocus enzyme electrophoresis, provide a means of classifying particular parasites, but require well-identified reference collections (Chilton, 1999). Serological approaches such as enzyme-linked immunosorbent assays (ELISA) have also been used to assess (previous) infection with particular parasites (Xu, Sui, Cao, & Lin, 2010), but setting up these approaches is challenging and often associated with cross-reactivity with antigens from closely related parasites. All of these approaches are time and

resource intensive and differ in their sensitivity and specificity for different groups of parasites, which limit their utility for generating the large data sets needed for large-scale comparative analyses.

In terms of sampling effort, it is widely appreciated that the more a population is studied, the more parasite species will be found, and clearly, not all species are studied equally (Nunn, Altizer, Jones, & Sechrest, 2003). One solution to this bias has been to include a variable to statistically control for variation in sampling effort. Another solution is to use richness estimators, such as Chao2, to more systematically predict how parasite counts accumulate with increased sampling effort (Chao, 1987). Despite the promise of these approaches, several factors make it challenging to effectively assess and control for sampling effort in comparative studies. One issue concerns whether samples come from known individuals. Especially in the case of noninvasive samples, such as faeces, samples may come from different individuals who are not individually identifiable unless time-consuming and expensive genetic approaches are employed. This leads to biased estimates of prevalence (Miller, Schneider-Crease, Nunn, & Muehlenbein, 2018; Walther, Cotgreave, Price, Gregory, & Clayton, 1995) and makes it difficult to use the number of animals sampled as a predictor variable to control for sampling effort, though this depends on the spatial scale of the study. Similarly, the assumptions of richness estimators, such as Chao2, are commonly violated in large-scale studies from the literature; for example, different studies commonly use different methods, have different sample sizes, and will screen for different sets of parasites. The last point is especially important, because most methods for controlling sampling effort assume that all studies have the potential to find the same (complete) sets of parasites, but given different methodologies used by different research groups, this is

not how actual sampling for parasites in the wild proceeds (Cooper & Nunn, 2013).

Because of these issues, a strong need exists to develop easily applicable and standardized methods to catalogue parasite diversity in the context of comparative research (Blaxter, 2011; Dobson, Lafferty, Kuris, Hechinger, & Jetz, 2008). DNA barcoding uses sequencing of short DNA regions (usually first amplified by PCR) to assign unknown samples to a particular species based on comparison with a reference database of homologous sequences (Moritz & Cicero, 2004). With the development of high-throughput sequencing approaches, barcoding can now be used to identify many species from a single sample (i.e. metabarcoding: Taberlet, Coissac, Pompanon, Brochmann, & Willems, 2012). A major advantage of these approaches is that they generate sequence data that can be used to estimate the phylogenetic relatedness of parasite community members, which can provide information about the processes driving community assembly (e.g. phylogenetic over- or under-dispersal suggesting interactions between parasite community members: Fountain-Jones et al., 2017; Gogarten et al., 2018). Such approaches have been widely used to describe diverse communities of organisms (e.g. plants: De Barba et al., 2014; bacteria and archaea: Eren et al., 2013; Fox, Pechman, & Woese, 1977; fungi: Blaalid et al., 2013; soil nematodes: Darby, Todd, & Herman, 2013; Porazinska et al., 2009; vertebrates: Andersen et al., 2012). Despite the potential of metabarcoding, its application to parasite communities across a diversity of vertebrate hosts remains limited (Titcomb, Young, & Jerde, 2019), with most recent studies focusing on a relatively narrow set of parasites in a single host (e.g. helminths of urban rats: Tanaka et al., 2014; strongylids in wallabies: Lott, Hose, & Power, 2015; nematodes in domestic cattle: Lott et al., 2015; nematodes in wild mouse lemurs: Avelo, Medlar, Löytynoja, Laakkonen, & Jernvall, 2015).

Building on the recent development of three different PCR protocols (Krogsgaard et al., 2018), which target the 18S ribosomal RNA gene of a large array of parasites, we tested the feasibility of a standardized metabarcoding approach to describe parasite communities from multiple primate host species. For this, we collected and analysed faecal samples from 11 nonhuman primate species from five countries (Figure 1). These hosts represent major lineages of primates (Figure 1), and the species we studied have had their parasites studied to variable degrees, ranging from no previous studies to the best-studied primates in the wild (i.e. the gorilla and chimpanzee). To examine how data generated with this approach compare with more established approaches that compile existing data from the literature and other sources, we compared estimates of host parasite species richness obtained by metabarcoding with the combined results of 113 published parasite studies on these primate hosts. To assess whether metabarcoding generates biologically relevant information, we tested whether the parasite communities described using this approach are structured in a way that reflects their host phylogenetic relatedness and geographical proximity. Finally, to determine whether the metabarcoding approach detects novel parasites above and beyond what has already been described in the literature, we tested whether the overlap between these approaches increases

with sampling effort in the literature and whether estimates of parasite species richness have already saturated.

## 2 | MATERIALS AND METHODS

### 2.1 | Sample collection

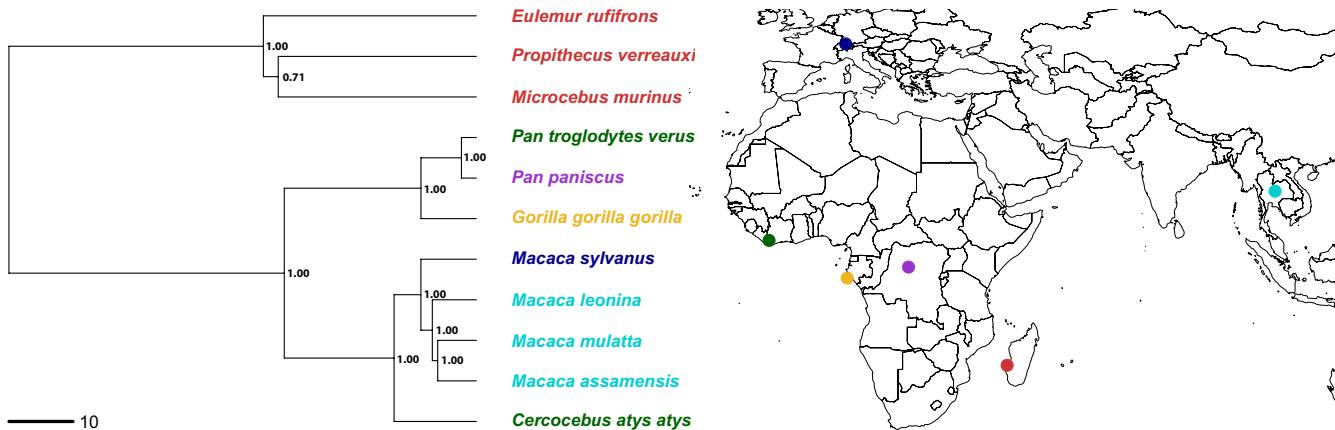
Faecal samples were collected from ten different adult individuals from each of 11 primate species at six different sites (Figure 1; *Propithecus verreauxi*, *Eulemur rufifrons* and *Microcebus murinus* at Kirindy Forest, Madagascar; *Pan paniscus* at Kokolopori, Democratic Republic of the Congo; *Gorilla gorilla gorilla* in Loango National Park, Gabon; *Macaca leonina*, *Macaca assamensis*, and *Macaca mulatta* at Phu Khieo Wildlife Sanctuary, Thailand; *Cercocebus atys atys* and *Pan troglodytes verus* at Taï National Park, Côte d'Ivoire; *Macaca sylvanus* at Affenberg Salem), associated with the German Research Foundation (DFG) research group, 'Sociality and Health in Primates'. One species (*Macaca sylvanus*) represents a semi-free ranging population, and all other populations are wild. With the exception of the gorilla samples, which came from an unhabituated population, all samples were collected from animals habituated to human observers and individuals were known to researchers, making it possible to collect from specific individuals. We collected 3 ml of fresh faeces and mixed it with 7 ml of RNAlater (Ambion) in a 15 ml tube. Samples were homogenized by shaking and subsequently stored at -20°C after a minimum of 24 hr at room temperature, with a continuous cold chain maintained until further laboratory analysis where possible.

### 2.2 | DNA extraction

DNA was extracted from faeces using the First-DNA all tissue kit (Genial) following the manufacturer's instructions. In brief, 200 mg of faecal-RNAlater homogenate was centrifuged for 5 min at 15,870 rcf in a 2 ml low-binding tube and excess liquid was discarded. We then added 1 ml of the kit's Lysis Buffer 1 and ~50 1.4 mm diameter beads (PeQlab) and homogenized samples with a TissueLyser (Qiagen: 5 rounds of 50 oscillations for 1 min followed by 2 min incubations). We then added 100 µl Lysis Buffer 2 and 20 µl proteinase K (20 mg/ml), and samples were incubated with occasional shaking for 45–60 min at 65°C. After chloroform purification, DNA was precipitated with iced isopropanol and pelleted for 15 min at 15,870 rcf. The DNA pellet was washed with 70% ethanol, dried and dissolved in 50 µl TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0). The final DNA concentration was measured using the Qubit® double-stranded high sensitivity assay kit (Life Technologies™) following the manufacturer's instructions.

### 2.3 | Amplicon generation

We used three universal primer pairs (G3F1/G3R1, G4F3/G4R3, and G6F1/G6R1) that have been shown to amplify the hyper-variable



**FIGURE 1** Primate phylogeny of species sampled, based on the phylogeny of Perelman et al. (2011), with the scale in millions of years, plotted using the R package *ape* (Paradis & Schliep, 2018). Bayesian posterior probabilities are indicated at each node. Colours correspond to field sampling sites. The map was created with the R package *rworldmap* (South, 2011)

V3-V4 and V3-V5 regions of the 18S rRNA gene from a wide range of intestinal parasites in humans (Krogsgaard et al., 2018). To simplify library preparation, we added adapter nucleotide sequences to the 5' end of these primers (Table S1). The PCR reaction was performed in a total volume of 20 µl, containing 1x Phusion High-fidelity buffer, 10 mmol of each forward and reverse primer (Metabion), 200 µM deoxynucleotide triphosphates (Thermo Fischer Scientific), 1.5 µg/µl bovine serum albumin (Sigma-Aldrich), 0.1% Triton X-100 (Sigma-Aldrich), 0.4 Unit Phusion High-fidelity DNA polymerase (Thermo Fischer Scientific), and 2 µl DNA extract (up to a maximum of 100 ng total genomic DNA). Cycling conditions were 98°C 30 s, 35 cycles of (98°C 10 s, 60°C for 30 s, 72°C 40 s), followed by a final extension at 72°C for 5 min. For each batch of ten samples, we included two nonsample controls. PCR products were visualized on 1.5% agarose gels to check for PCR performance and contamination.

PCR products were purified using Solid Phase Reversible Immobilization (SPRI; Beckman Coulter), according to the manufacturer's instructions. In brief, PCR products were mixed with SPRI beads at a DNA-to-bead ratio of 0.6–0.7. Beads were then separated from the solution with a magnet (Applied Biosystems). PCR products were then washed twice on the beads with freshly prepared 80% EtOH and remaining EtOH was evaporated. PCR products were eluted in 27 µl TET buffer (10 mM Tris-HCl, 1 mM EDTA, 0.05% Tween 20).

## 2.4 | Library preparation

For each sample, we pooled 10 ng of purified DNA of each of the three PCR products. We performed an indexing PCR in a total volume of 50 µl which contained 12 µl pooled DNA (~30 ng), 250 µM deoxynucleotide triphosphates (Thermo Fischer Scientific), 1x pfu Turbo Cx buffer and 2.5 Unit pfu Turbo Cx Hotstart DNA polymerase (Agilent Technologies). Cycling conditions were 98°C 30 s, 10 cycles of (98°C 10 s, 62°C for 30 s, 72°C 60 s), followed by a final extension at 72°C for 10 min. Subsequently, PCR products were purified using SPRI beads, as described above. We diluted final libraries to a concentration of 10 nM

and then pooled and sequenced them on two Illumina MiSeq runs, generating 2 × 250 bp (30 samples) or 2 × 300 bp (80 samples) reads.

## 2.5 | Bioinformatic analysis

Reads were sorted based on the respective primer sequences using cutadapt (Martin, 2011). Reads were then filtered using standard filtering parameters using the DADA2 workflow (i.e. maxN = 0, truncQ = 2, rm.phix = TRUE and maxEE = 2, minLen = 100; Callahan et al., 2016). We merged reads using the illumina pairedend function in OBItools (Boyer et al., 2016), with a minimum alignment score of 40. Reads that did not overlap were concatenated. Merged reads were then deduplicated using the function obiuniq in OBItools. Unique reads were then assigned to a taxon using the ecotag function in OBItools. Briefly, this alignment-based approach looks for the reference sequence showing the highest similarity to the query sequence. Then, it looks for all other reference sequences that are similarly similar and assigns the query sequence to the most recent common ancestor of all similar sequences (Boyer et al., 2016). As reference database, we used the SILVA SSU Parc Full database (Pruesse et al., 2007). We used the Obisilva function in OBItools to generate an ecoPCR database based on the SILVA reference database (Boyer et al., 2016). We then generated reference databases for each of our primer protocols using the ecoPCR function in OBItools (Boyer et al., 2016), allowing for up to 5 mismatches between the primer and a specific sequences, and a maximum length of expected PCR-product of 800 bp. We excluded all reads that were not assignable to the family, genus or species level.

## 2.6 | Comparison of primate parasite communities

To compare our findings to known eukaryotic parasite diversity, we downloaded the Global Primate Parasite Database (Nunn & Altizer, 2005; Stephens et al., 2017) and filtered for fungi, helminths and protozoa for the 11 primate hosts that we sampled (accessed on:

June 5th, 2019). In addition, we downloaded the EuPathDB (accessed on: February 1st, 2019, <https://eupathdb.org/>; Aurrecoechea et al., 2009). From these databases, we extracted a list of known parasite families. We then filtered our database to exclude reads that were not assigned to one of these parasitic families. To avoid false positives due to contamination or spillover between libraries, we used a minimum threshold of ten reads for a particular parasitic family to be considered present in a sample. Libraries prepared for other projects, not relating to parasites, that were processed in parallel and sequenced on the same MiSeq runs had no reads assignable to a parasitic taxon using this pipeline, suggesting that spillover rates between libraries was low. To explore taxonomic problems in the SILVA database, we also examined the NCBI taxonomy of the best match, which included many more parasite families.

To compare the primate parasite communities based on their membership in broad taxonomic groups (Cercopithecoidea, Hominoidea and Lemuroidea), we generated a nonmetric multidimensional scaling (NMDS) ordination plot of the detection rate of parasite families for each species (Bray–Curtis distance, stress value = 0.0758) using the *vegan* R package (Jari Oksanen et al., 2019). The Bray–Curtis index was calculated for all primate host combinations and we used adonis, a permutational multivariate analysis of variance (Anderson, 2001), to test for differences in this distance matrix, using the three broad taxonomic groups and the continent as predictors of the similarity of primate parasite communities. In addition, we generated a NMDS ordination plot using the presence–absence parasite community matrix for those samples in which parasites were detected, using both the family and the species level parasite delineations (Bray–Curtis distance, stress value<sub>family</sub> = 0.100 and stress value<sub>species</sub> = 0.106). In this case, we ran three separate adonis analyses, using the host species, the three broad taxonomic groups, or the continent as predictors of the similarity of primate parasite communities. Plots were created using the *ggplot2* R package (Wickham, 2016).

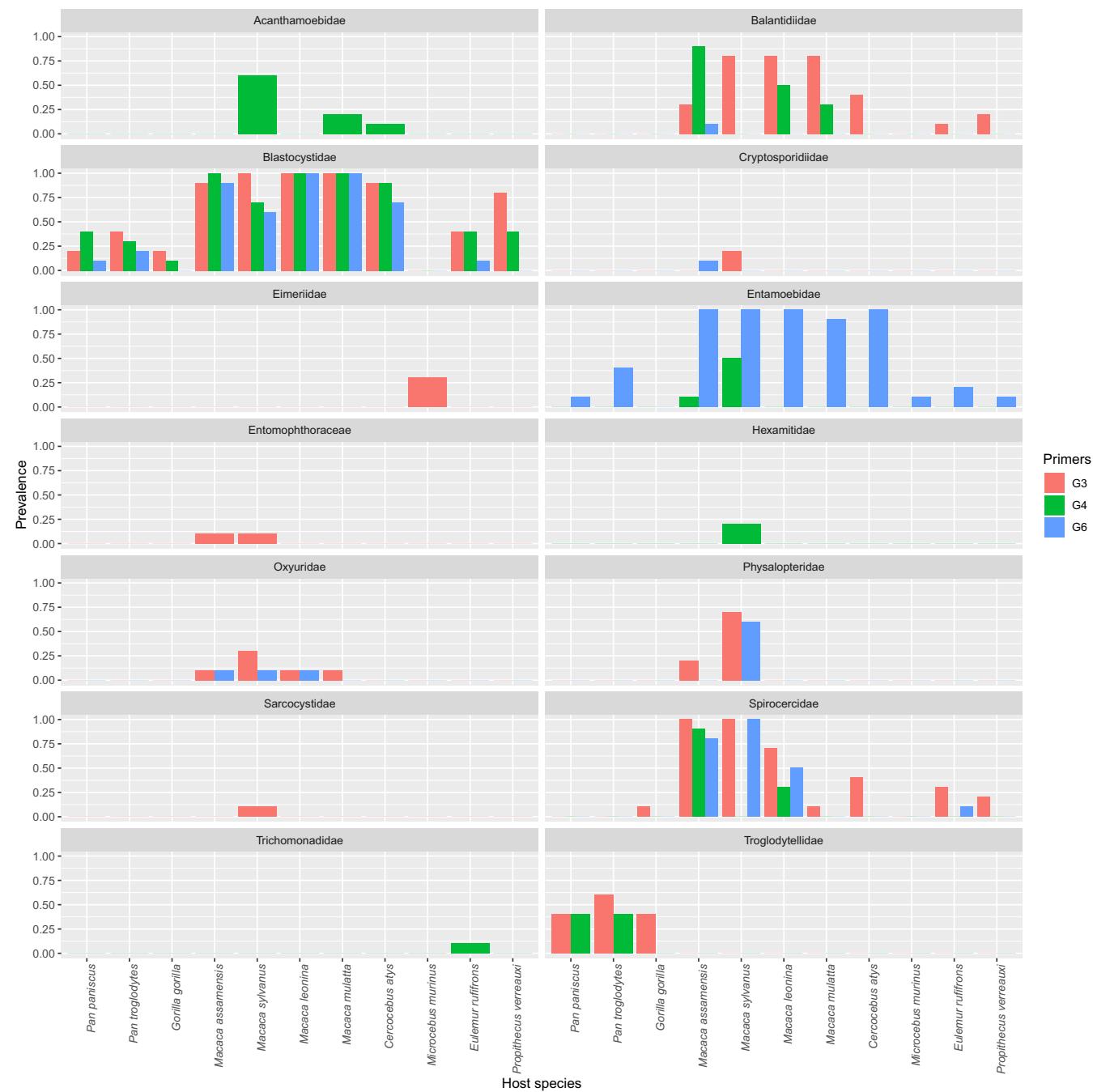
We used phylogenetic generalized least squares (PGLS) regressions to test if overlap in parasite families detected by metabarcoding and the literature increased as more studies were published on a species. Specifically, our model predicted the log-transformed number of parasite families detected by both the metabarcoding and the literature, using the log-transform of the number of publications in the literature for that host species as our predictor variable. We extracted the number of studies performed on a species and the number of parasites detected from the Global Primate Parasite Database after filtering for fungi, helminths and protozoa. PGLS includes the host phylogenetic tree underlying the data as a covariance matrix in a linear model, accounting for the phylogenetic nonindependence of primate species in our study. We used the primate phylogeny of Perelman et al. (2011). We conducted our PGLS analysis using the *ppls* function in the R package *caper* (Orme et al., 2018). We repeated this analysis removing the one semi-free ranging population from our analysis, as captivity might be expected to increase exposure to human and livestock parasites. All statistical analyses and data analyses were performed in R v3.4.1 (R Core Team, 2017).

### 3 | RESULTS

We generated 27,322,392 paired end reads from 110 samples (Table S1). After quality filtering, 3,565,202 reads were then assignable to the G3 primer pair, 1,287,066 reads to the G4 primer pair and 3,016,485 reads to the G6 primer pair (Table S2). Of these reads, 1,171,266 were assigned to a parasite family when using the SILVA taxonomy (Table S2). Of the reads assignable to a parasite family, the majority could be further assigned to the genus and species level (Figure S1). The primer pairs differed in the parasite families they detected, with different primate hosts having differences in their parasite communities (Figure 2). We detected between seven and 19 parasite families in a given host.

When using the NCBI taxonomy of the best match, we found several additional parasite families (Figure 3). Despite differences in the taxonomy associated with the different databases, these two approaches were largely complementary (Table S3). Using species aggregated parasite family detection rates, primates hosted distinct parasite communities that appeared to be structured based on the phylogenetic relationships of the hosts (Figure 4a; adonis of Bray–Curtis distances,  $F = 4.001$ ,  $R^2 = 0.501$ ,  $p = 0.001$ ) and the continent on which the hosts were found (Figure 4b;  $F = 2.199$ ,  $R^2 = 0.485$ ,  $p = 0.015$ ). Using the presence–absence parasite community matrix for samples, based on parasite families, primate host taxa again hosted distinct parasite communities (Figure S2a; adonis of Bray–Curtis distances,  $F = 7.593$ ,  $R^2 = 0.493$ ,  $p = 0.001$ ), while samples also hosted distinct parasite communities that appeared to be structured based on the phylogenetic relationships of the hosts (Figure S2b; adonis of Bray–Curtis distances,  $F = 17.471$ ,  $R^2 = 0.289$ ,  $p = 0.001$ ) and their continent of origin (Figure S2c; adonis of Bray–Curtis distances,  $F = 11.139$ ,  $R^2 = 0.282$ ,  $p = 0.001$ ). The same pattern was observed using parasite species level assignments; using the presence–absence parasite community matrix based on parasite species, primate hosts hosted distinct parasite communities (Figure S3a; adonis of Bray–Curtis distances,  $F = 8.976$ ,  $R^2 = 0.618$ ,  $p = 0.001$ ), while samples also hosted distinct parasite communities that appeared to be structured based on the phylogenetic relationships of the hosts (Figure S3b; adonis of Bray–Curtis distances,  $F = 2.491$ ,  $R^2 = 0.080$ ,  $p = 0.012$ ) and their continent of origin (Figure S3c; adonis of Bray–Curtis distances,  $F = 12.167$ ,  $R^2 = 0.395$ ,  $p = 0.001$ ).

We compared the results from the ten faecal samples for each population to what is collectively known about eukaryotic parasite diversity in these primate species as a whole, based on the Global Primate Parasite Database. We found that the molecular approach performed efficiently (Figure 5), with the overlap between the parasite families detected with metabarcoding and the published literature increasing for hosts species that were better studied for parasites in the Global Primate Parasite Database (Figure 4c, PGLS,  $B = 0.649$ ,  $p = 0.0173$ ,  $R^2 = 0.49$ ). This relationship remained after removing one captive population (PGLS,  $B = 0.597$ ,  $p = 0.0117$ ,  $R^2 = 0.57$ ).

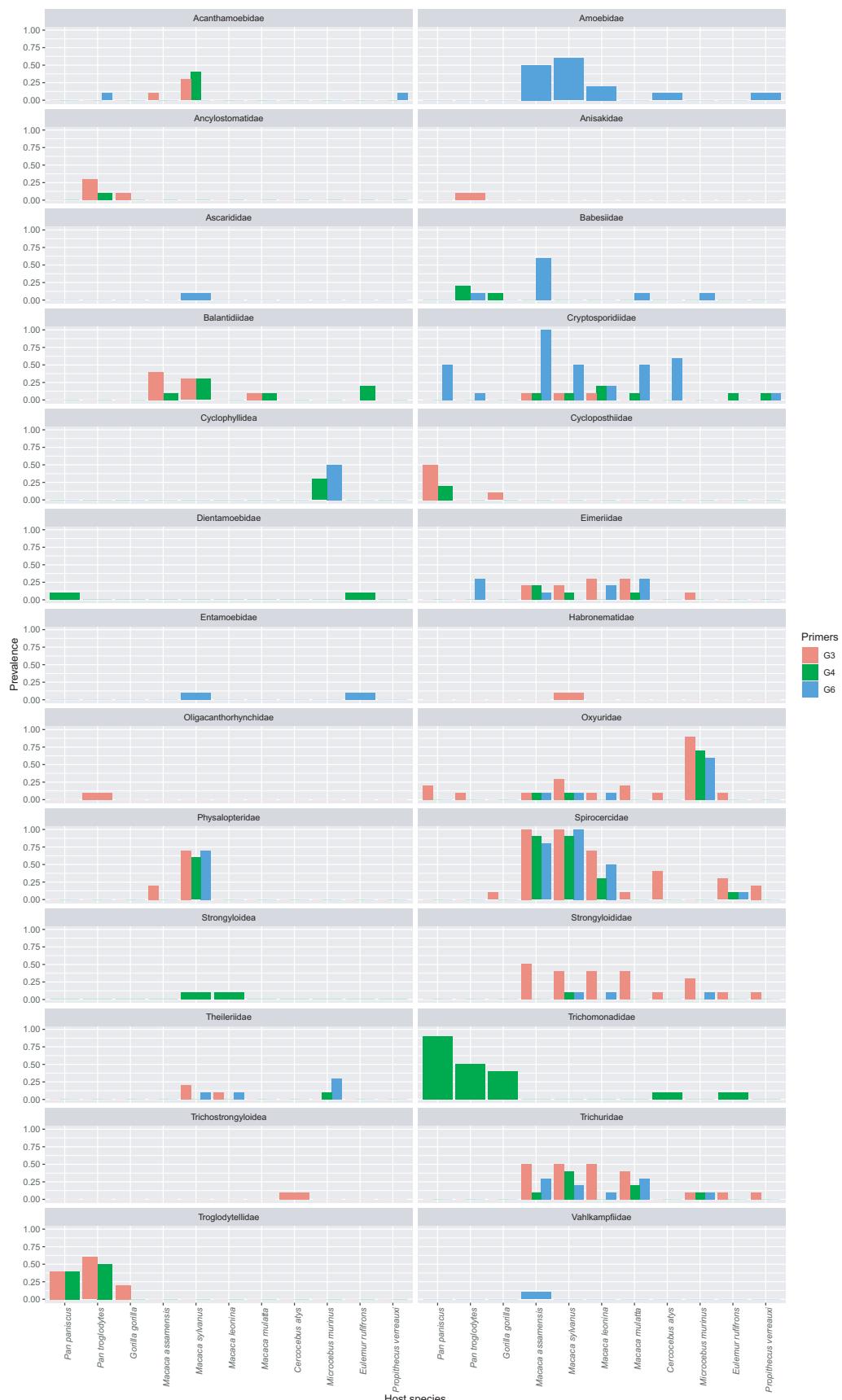


**FIGURE 2** Primate parasite detection rate (per cent of samples in which parasite family was detected) by host species, comparing the results from three different PCR primer pairs. Shown here are the results after analysing reads using the SILVA taxonomy

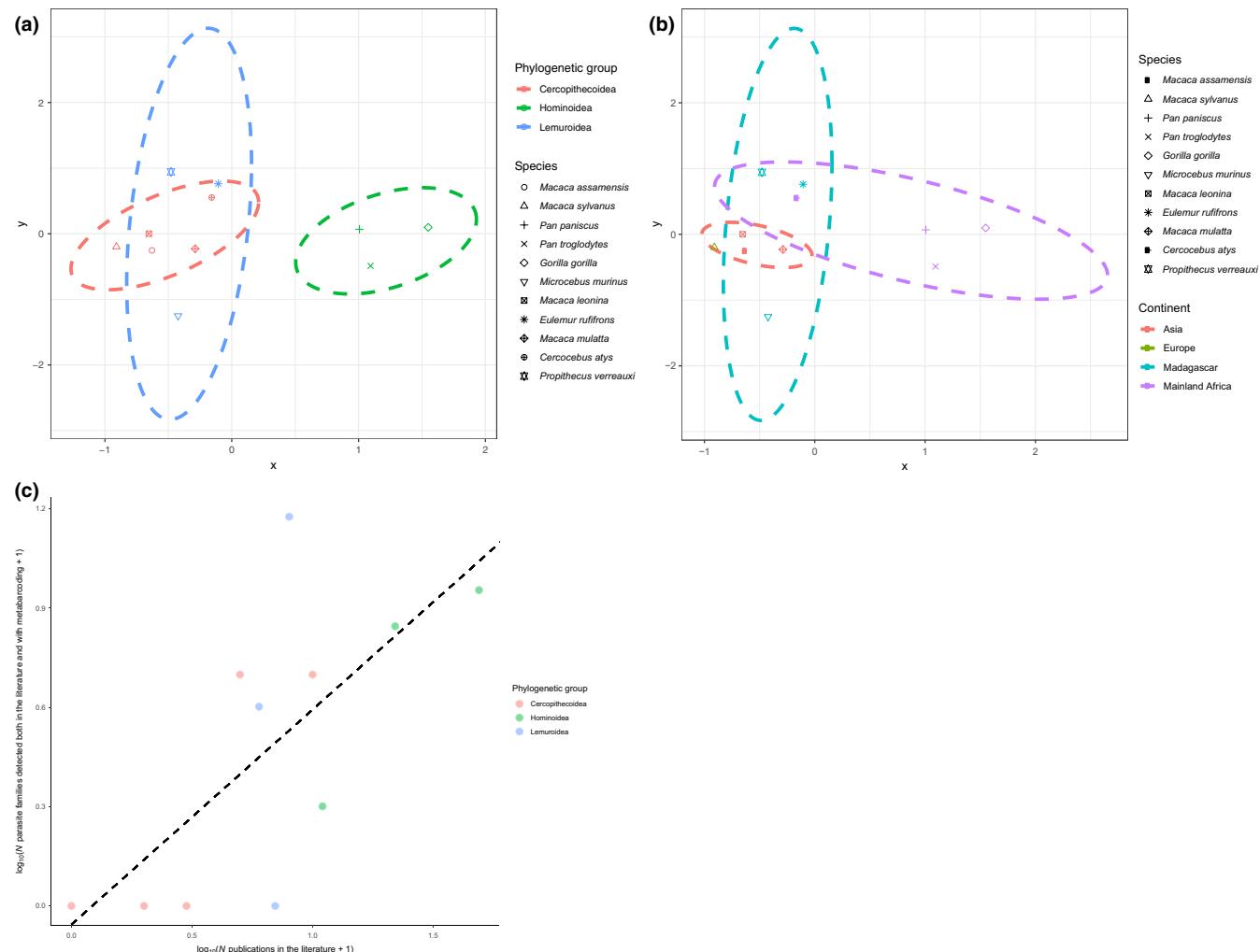
## 4 | DISCUSSION

Using a standardized metabarcoding approach on a limited number of faecal samples, we detected many parasite families in hosts from a subset of species from across the primate phylogeny. When we compared our results to published parasite data from these primate species, we found more overlap between the parasite families detected with metabarcoding and the published literature when more research had been published on that species, which lends credibility to the data generated with this approach. Surprisingly, in all host species, regardless of the prior sampling effort, metabarcoding of

just ten faecal samples identified at least one parasite family that had not been previously described in that host species. This highlights the utility of metabarcoding studies for improving assessments of parasite richness. The taxonomic richness of bacterial and fungal communities in a variety of substrates and environments is now routinely examined using DNA metabarcoding (Caporaso et al., 2011). Our results suggest that these tools have the potential to help parasitology and macroecology to move towards more standardized studies of parasite communities across a broad range of hosts. The costs for genetic approaches allowing host individuals to be determined from noninvasive samples continue to decline, suggesting



**FIGURE 3** Primate parasite detection rate (per cent of samples in which parasite family was detected) by host species, comparing the results from three different PCR primer pairs. Shown here are the results after analysing reads using the NCBI taxonomy



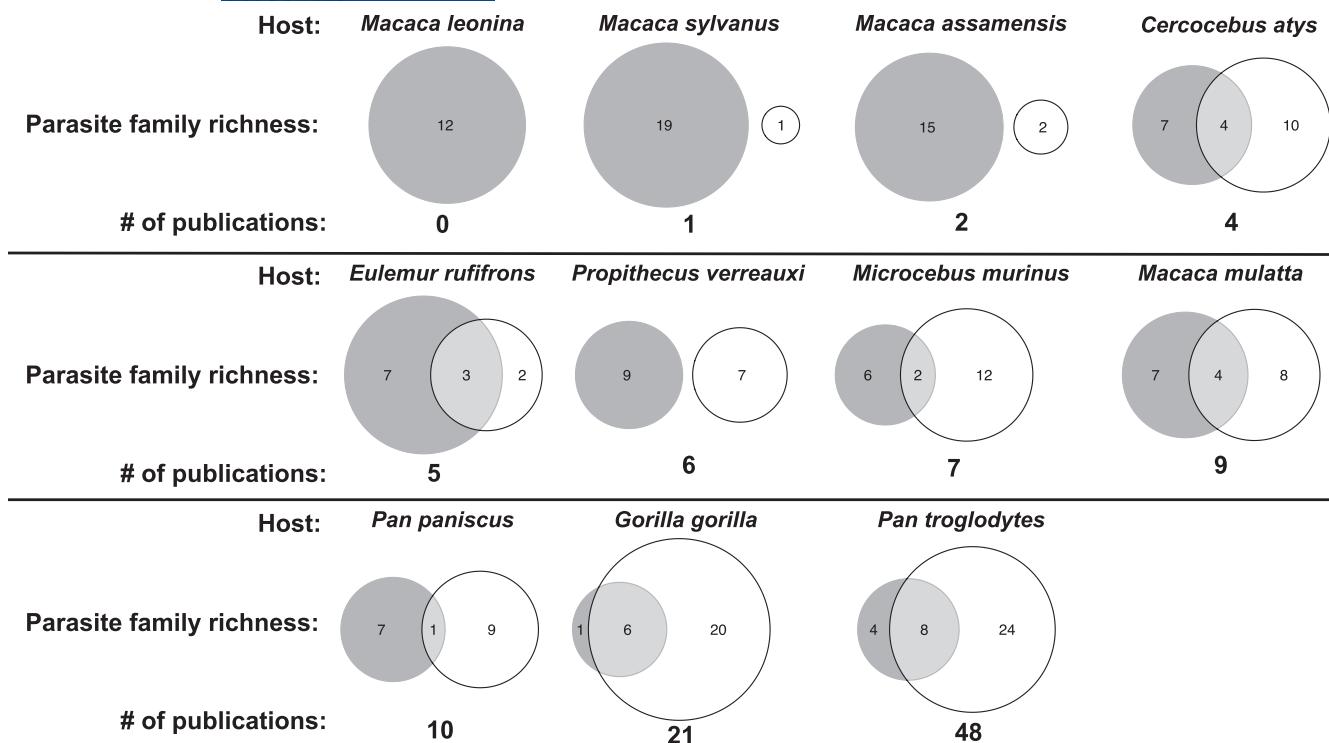
**FIGURE 4** Nonmetric multidimensional scaling (NMDS; stress value = 0.0758) ordination plots of the detection rate of parasite families for each primate species, group by (a) phylogenetic group or (b) continent, based on the combined results using the NCBI taxonomy of the best hit. Dashed ellipses indicate the 95% confidence ellipses. (c) Scatterplot showing the relationship between the log-transformed number of parasite families detected by both metabarcoding and the literature, with the log-transform of the number of publications in the literature. The colour of the points indicates the phylogenetic group of the host species. The dashed line indicates the best-fit line from the PGLS analysis

that these approaches could be coupled with parasite metabarcoding to allow for more accurate estimates of prevalence, while allowing researchers to control for variation in sampling effort between studies (Miller et al., 2018; Walther et al., 1995).

Analysis of the metabarcoding data revealed that more closely related species had more similar parasite communities. Similar patterns have been observed for symbiotic and parasitic microorganisms across a diversity of hosts (e.g. gut bacterial communities for a diversity of primates within an ecosystem: Gogarten et al., 2018; viruses and helminths of carnivores: Huang, Bininda-Emonds, Stephens, Gittleman, & Altizer, 2014), suggesting that metabarcoding of parasites is generating biologically meaningful data suitable for comparative analyses. Similarly, species from the same continent also had more similar parasite communities; biological similarity is often observed to decrease with geographical distance (Soininen, McDonald, & Hillebrand, 2007) and has been described for primate parasite communities (Davies & Pedersen, 2008). Despite

the apparent success of this approach, metabarcoding efforts rely heavily on the quality of the reference databases used to assign taxonomic information to sequences (Bohmann et al., 2014; Bush et al., 2017; López-García et al., 2018; Pedersen et al., 2015; Schnell et al., 2015; Taberlet, Bonin, Zinger, & Coissac, 2018). The majority of the sequences generated in this study could only be confidently assigned at the genus level, in large part probably due to gaps in the SILVA reference database—many parasite species have simply not been extensively barcoded, particularly across a broad diversity of host species. As reference databases improve, data sets could routinely be reexamined to refine assignments. One exciting avenue for improving reference databases is the use of museum specimens and the extensive collections of parasitologists (Salleh et al., 2017).

Beyond sparse reference databases, the SILVA database does not include the family or genus level for many of the parasites and thus misses information about the taxonomic rank. This makes it impossible for these parasites to be detected with an approach like



**FIGURE 5** Overlap in primate parasite families estimated using the published literature (white) on helminths, protozoa, and fungi compiled in the Primate Parasite Database and the current metabarcoding study (grey). The number of publications included in the primate parasite database is listed in the bottom row

OBItools, which relies on taxonomic ranks for read assignment. The fact that many additional parasite families were detected simply using the NCBI taxonomy coupled with the SILVA database highlights that improvements to the taxonomy of reference databases alone have the potential to greatly refine the information drawn from metabarcoding approaches. With this respect, efforts underway to develop a universal taxonomic framework for eukaryotic organisms are particularly promising (Berney et al., 2017). For reads assigned to a parasitic family, further analyses could be performed to examine parasite species and subspecies diversity, relatedness of parasites detected and estimates of sharing of parasites between hosts, even in the absence of informative reference databases. For example, reads from a given parasitic family could be aligned, and this alignment could be used to construct a phylogeny; species delimitation analysis could then be performed to define molecular operational taxonomic units (e.g. Generalized Mixed Yule Coalescent approaches; Fujisawa & Barraclough, 2013).

Improvements to PCR protocols also have the potential to improve metabarcoding approaches. The three PCR primer pairs that we used generated complementary information about the parasite communities in these samples, suggesting including more than one primer pair provides additional information for comparative studies of parasitism (Krogsgaard et al., 2018). As sequence reference databases improve, it will probably be possible to further improve primer design to target particular taxonomic groups of organisms (Boyer et al., 2016), in ways that increase resolution within particular taxonomic groups (De Barba et al., 2014). While our use of three different primer

pairs enabled us to capture more parasite diversity than would have been captured with a single pair, all targeted the 18S ribosomal RNA gene; other gene targets are frequently used for barcoding (e.g. cytochrome c oxidase subunit 1, internal transcribed spacer: Nassonova, Smirnov, Fahrni, & Pawłowski, 2010; Schoch et al., 2012; Taberlet et al., 2018), and combinations of targets could be explored to maximize the number of taxonomic groups detected. The results from different gene targets could be compared to verify findings and increase confidence in results. Similarly, as sequencing technologies improve, the use of longer barcoding sequences could further increase the taxonomic information recovered from each sequence (Benítez-Páez & Sanz, 2017). Not all storage methods and extraction methods can extract DNA from all organisms equally well (e.g. thick-walled eggs are often hard to break open) and optimization of extraction methods and extraction techniques may also improve the breadth of organisms detected and reduce detection biases for particular groups of organisms (Da Silva et al., 1999; Hallmaier-Wacker, Lueert, Roos, & Knauf, 2018). Meta-analyses have the potential to combine information from several studies to ask additional questions of data sets and examining the generalizability of results; facilitating such efforts is an important consideration in favour of standardization (Gilbert, Jansson, & Knight, 2014). While modifying and improving methods have the potential to reduce costs and improve the data generated in such studies, finding the balance between the need to innovate and improve systems and the need for standardization is challenging.

We considered parasites to be present or absent in a given sample, as there is considerable debate about whether the number of

reads detected with metabarcoding efforts provides reliable information about abundance of template DNA in a sample (Avramenko et al., 2015; Elbrecht & Leese, 2015; Krehenwinkel et al., 2017). Variation in 18S gene copy numbers could influence detection rates and incorporating this information might allow for improved estimates of diversity and potentially even abundance from metabarcoding efforts (Kembel, Wu, Eisen, & Green, 2012). Even with such information, additional PCR biases such as template secondary structures, primer-template mismatches and taxon-specific amplification biases all make it difficult to estimate relative abundance of parasites in a sample (Fonseca et al., 2012; Krehenwinkel et al., 2017). The use of mock communities of parasites seems to make it possible to make an assessment of these biases for particular parasites, and if abundance biases are taxon specific and predictable, correction factors could allow for more reliable abundance estimates from such data (Avramenko et al., 2015; Hallmaier-Wacker et al., 2018; Krehenwinkel et al., 2017). Such metabarcoding efforts might allow the generation not only of presence-absence data about particular parasites, but also to generate information about infection intensity as well.

Using available reference databases and technologies, our results demonstrate the potential of metabarcoding to facilitate large-scale comparative studies of parasite communities. These cost-effective tools can reduce variation in methods for identifying parasites across species and thereby allow for studies of human and wildlife parasite communities at broad spatial scales, such as those focused on parasite sharing in the context of human–livestock–wildlife contact (Ghai, Chapman, Omeja, Davies, & Goldberg, 2014; Parsons et al., 2015) or on broader ecological networks involving entire clades of mammals (Gómez, Nunn, & Verdú, 2013). Metabarcoding efforts will also enable researchers to investigate parasite communities at finer scales within a host species (Aiveló et al., 2015); for example, metabarcoding samples from across environmental gradients might identify factors probably to influence parasite richness or species composition, such as habitat degradation and temperature (Chapman, Speirs, Gillespie, Holland, & Austad, 2006; Gillespie, Chapman, & Greiner, 2005). In general, metabarcoding approaches may provide a decisive contribution to our understanding of the ecology and evolution of parasitism.

## ACKNOWLEDGEMENTS

This study was supported in part by the DFG Research Group 'Sociality and Health in Primates' (FOR2136) and the Max Planck Society (providing core funding for: Tai Chimpanzee project, Kokolopori Bonobo Project, and Loango Great Ape Project). While many funding agencies contributed to facilitating the long-term research at the these field sites, for this project we particularly grateful for additional funding for RMW through WI 2637/3-1 and WI 2637/4-2, CF through DFG FI 929/7-1 and KR 3834/5-1, for PMK through Ka 1082-28-1, for TD from the DFG DE 1135/2-1 and DFG 1135/3-1, for OS, JO and NMK through DFG SCHU 1554/6-1, DFG OS201/6-1 and DFG OS201/6-2. We thank the Ministère de l'Enseignement supérieur et de la Recherche scientifique of Côte d'Ivoire and the Office Ivoirien des Parcs et Réserves for permission to conduct the study in the Tai National Park. We thank the National Research Council of Thailand

(NRCT) and the Department of National Parks, Wildlife and Plant Conservation (DNP) for permission to conduct this study. We thank Simone Lueert for their assistance in prescreening of control samples. For valuable comments and discussions regarding this manuscript, we thank Vanessa Ezenwa. We thank the Agence Nationale des Parcs Nationaux and the Centre National de la Recherche Scientifique et Technique of Gabon for their collaboration and permission to conduct research in the Loango National Park. We thank the Ministère de Recherche Scientifique et Tecnologie, Democratic Republic of the Congo, Vie Sauvage and the Bonobo Conservation Initiative for their collaboration and permission to conduct research in Kokolopori.

## AUTHOR CONTRIBUTIONS

The research was designed by J.F.G., S.C.S., C.L.N., M.U., N.S., H.V.N., T.D., C.F., P.M.K., S.K., N.M.K., J.O., S.S., O.S., R.M.W., A.S., C.S., F.H.L., C.R. and A.N.; the research was performed by J.F.G., S.C.S., N.S., S.K., C.R. and A.N.; new reagents or analytical tools/provided samples were contributed by C.F., P.M.K., N.M.K., J.O., S.S., O.S., M.R., M.S., R.M.W., A.S., C.S. and F.H.L.; data were analysed by J.F.G., S.C.S., C.R. and A.N.; the paper was written by J.F.G., S.C.S., C.R. and A.N.; the manuscript was commented by all authors and the final version was approved by all authors.

## ORCID

Jan F. Gogarten  <https://orcid.org/0000-0003-1889-4113>

Sébastien Calvignac-Spencer  <https://orcid.org/0000-0003-4834-0509>

## DATA AVAILABILITY STATEMENT

Next generation sequencing reads are available through the Short Read Archive (SRA) BioProject accession number PRJNA564536. All additional information is available in the manuscript.

## REFERENCES

- Aiveló, T., Medlar, A., Löytynoja, A., Laakkonen, J., & Jernvall, J. (2015). Tracking year-to-year changes in intestinal nematode communities of rufous mouse lemurs (*Microcebus rufus*). *Parasitology*, 142(8), 1095–1107.
- 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.
- Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, 26(1), 32–46.
- Aurrecoechea, C., Brestelli, J., Brunk, B. P., Fischer, S., Gajria, B., Gao, X., ... Heiges, M. (2009). EuPathDB: A portal to eukaryotic pathogen databases. *Nucleic Acids Research*, 38(suppl\_1), D415–D419.
- Avramenko, R. W., Redman, E. M., Lewis, R., Yazwinski, T. A., Wasmuth, J. D., & Gillearde, J. S. (2015). Exploring the gastrointestinal "Nemabiome": Deep Amplicon sequencing to quantify the species composition of parasitic nematode communities. *PLoS ONE*, 10(12), e0143559. <https://doi.org/10.1371/journal.pone.0143559>
- Benítez-Páez, A., & Sanz, Y. (2017). Multi-locus and long amplicon sequencing approach to study microbial diversity at species level using the MinION™ portable nanopore sequencer. *GigaScience*, 6(7), gix043.
- Berney, C., Ciuprina, A., Bender, S., Brodie, J., Edgcomb, V., Kim, E., ... Audic, S. (2017). UniEuk: Time to speak a common language in protistology!. *Journal of Eukaryotic Microbiology*, 64(3), 407–411.

- Blaalid, R., Kumar, S., Nilsson, R. H., Abarenkov, K., Kirk, P., & Kauserud, H. (2013). ITS 1 versus ITS 2 as DNA metabarcodes for fungi. *Molecular Ecology Resources*, 13(2), 218–224.
- Blaxter, M. (2011). Nematodes: The worm and its relatives. *PLoS Biology*, 9(4), e1001050.
- 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 and Evolution*, 29(6), 358–367.
- 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.
- Bush, A., Sollmann, R., Wilting, A., Bohmann, K., Cole, B., Balzter, H., ... Cobbold, C. A. (2017). Connecting Earth observation to high-throughput biodiversity data. *Nature Ecology and Evolution*, 1(7), 0176.
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13(7), 581.
- Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., ... Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences*, 108(Supplement 1), 4516–4522.
- Chao, A. (1987). Estimating the population size for capture-recapture data with unequal catchability. *Biometrics*, 783–791.
- Chapman, C. A., Speirs, M. L., Gillespie, T. R., Holland, T., & Austad, K. M. (2006). Life on the edge: Gastrointestinal parasites from the forest edge and interior primate groups. *American Journal of Primatology: Official Journal of the American Society of Primatologists*, 68(4), 397–409.
- Charlier, J., van der Voort, M., Kenyon, F., Skuce, P., & Vercruyse, J. (2014). Chasing helminths and their economic impact on farmed ruminants. *Trends in Parasitology*, 30(7), 361–367. <https://doi.org/10.1016/j.pt.2014.04.009>
- Chilton, N. B. (1999). Multilocus enzyme electrophoresis: A valuable technique for providing answers to problems in parasite systematics. *International Journal for Parasitology*, 29(2), 213–253.
- Cooper, N., & Nunn, C. L. (2013). Identifying future zoonotic disease threats: Where are the gaps in our understanding of primate infectious diseases? *Evolution, Medicine, and Public Health*, 2013(1), 27–36.
- Da Silva, A. J., Bornay-Llinares, F. J., Moura, I. N., Slemenda, S. B., Tuttle, J. L., & Pieniazek, N. J. (1999). Fast and reliable extraction of protozoan parasite DNA from fecal specimens. *Molecular Diagnosis*, 4(1), 57–64.
- Darby, B., Todd, T. C., & Herman, M. A. (2013). High-throughput amplicon sequencing of rRNA genes requires a copy number correction to accurately reflect the effects of management practices on soil nematode community structure. *Molecular Ecology*, 22(21), 5456–5471.
- Davies, T. J., & Pedersen, A. B. (2008). Phylogeny and geography predict pathogen community similarity in wild primates and humans. *Proceedings of the Royal Society B: Biological Sciences*, 275(1643), 1695–1701.
- De Barba, M., Miquel, C., Boyer, F., Mercier, C., Rioux, D., Coissac, E., & Taberlet, P. (2014). DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: Application to omnivorous diet. *Molecular Ecology Resources*, 14(2), 306–323.
- Decraemer, W., Brown, D., Karanastasi, E., Zheng, J., & De Almeida, M. T. M. (2001). Interspecific differences in the fine structure of the body cuticle of Trichodoridae Thorne, 1935 (Nematoda: Diphtherophorina) and review of anchoring structures of the epidermis. *Nematology*, 3(6), 525–533.
- Dobson, A., Lafferty, K. D., Kuris, A. M., Hechinger, R. F., & Jetz, W. (2008). Homage to Linnaeus: How many parasites? How many hosts? *Proceedings of the National Academy of Sciences of the United States of America*, 105, 11482–11489.
- Elbrecht, V., & Leese, F. (2015). Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass—sequence relationships with an innovative metabarcoding protocol. *PLoS ONE*, 10(7), e0130324. <https://doi.org/10.1371/journal.pone.0130324>
- Eren, A. M., Maignien, L., Sul, W. J., Murphy, L. G., Grim, S. L., Morrison, H. G., & Sogin, M. L. (2013). Oligotyping: Differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods in Ecology and Evolution*, 4(12), 1111–1119.
- Fonseca, V., Nichols, B., Lallias, D., Quince, C., Carvalho, G., Power, D., & Creer, S. (2012). Sample richness and genetic diversity as drivers of chimera formation in nSSU metagenetic analyses. *Nucleic Acids Research*, 40(9), e66.
- Fountain-Jones, N. M., Packer, C., Troyer, J. L., VanderWaal, K., Robinson, S., Jacquot, M., & Craft, M. E. (2017). Linking social and spatial networks to viral community phylogenetics reveals subtype-specific transmission dynamics in African lions. *Journal of Animal Ecology*, 86(6), 1469–1482.
- Fox, G. E., Pechman, K. R., & Woese, C. R. (1977). Comparative cataloging of 16S ribosomal ribonucleic acid: Molecular approach to prokaryotic systematics. *International Journal of Systematic and Evolutionary Microbiology*, 27(1), 44–57.
- Fujisawa, T., & Barracough, T. G. (2013). Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent approach: A revised method and evaluation on simulated data sets. *Systematic Biology*, 62(5), 707–724. <https://doi.org/10.1093/sysbio/syt033>
- Ghai, R. R., Chapman, C. A., Omeja, P. A., Davies, T. J., & Goldberg, T. L. (2014). Nodule worm infection in humans and wild primates in Uganda: Cryptic species in a newly identified region of human transmission. *PLoS Neglected Tropical Diseases*, 8(1), e2641. <https://doi.org/10.1371/journal.pntd.0002641>
- Gilbert, J. A., Jansson, J. K., & Knight, R. (2014). The Earth Microbiome project: Successes and aspirations. *BMC Biology*, 12(1), 69. <https://doi.org/10.1186/s12915-014-0069-1>
- Gillespie, T. R., Chapman, C. A., & Greiner, E. C. (2005). Effects of logging on gastrointestinal parasite infections and infection risk in African primates. *Journal of Applied Ecology*, 42(4), 699–707. <https://doi.org/10.1111/j.1365-2664.2005.01049.x>
- Gogarten, J. F., Davies, T. J., Benjamins, J., Gogarten, J. P., Graf, J., Mielke, A., ... Calvignac-Spencer, S. (2018). Factors influencing bacterial microbiome composition in a wild non-human primate community in Taï National Park, Côte d'Ivoire. *The ISME Journal*, 12(10), 2559–2574. <https://doi.org/10.1038/s41396-018-0166-1>
- Gómez, J. M., Nunn, C. L., & Verdú, M. (2013). Centrality in primate-parasite networks reveals the potential for the transmission of emerging infectious diseases to humans. *Proceedings of the National Academy of Sciences of the United States of America*, 110(19), 7738–7741. <https://doi.org/10.1073/pnas.1220716110>
- Hallmaier-Wacker, L. K., Lueert, S., Roos, C., & Knauf, S. (2018). The impact of storage buffer, DNA extraction method, and polymerase on microbial analysis. *Scientific Reports*, 8(1), 6292. <https://doi.org/10.1038/s41598-018-24573-y>
- Harvey, P. H., & Pagel, M. D. (1991). *The comparative method in evolutionary biology* (Vol. 239). Oxford, UK: Oxford University Press.
- Huang, S., Bininda-Emonds, O. R., Stephens, P. R., Gittleman, J. L., & Altizer, S. (2014). Phylogenetically related and ecologically similar carnivores harbour similar parasite assemblages. *Journal of Animal Ecology*, 83(3), 671–680. <https://doi.org/10.1111/1365-2656.12160>
- Jari Oksanen, F., Blanchet, G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., & Wagner, H. (2019). vegan: Community Ecology Package. R package version 2.5-4. <https://CRAN.R-project.org/package=vegan>

- Kembel, S. W., Wu, M., Eisen, J. A., & Green, J. L. (2012). Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. *PLoS Computational Biology*, 8(10), e1002743.
- Krehenwinkel, H., Wolf, M., Lim, J. Y., Rominger, A. J., Simison, W. B., & Gillespie, R. G. (2017). Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. *Scientific Reports*, 7(1), 17668.
- Krogsgaard, L. R., Lee, O., Johannessen, T. B., Engsbro, A. L., Stensvold, C. R., Nielsen, H. V., & Bytzer, P. (2018). Characteristics of the bacterial microbiome in association with common intestinal parasites in irritable bowel syndrome. *Clinical and Translational Gastroenterology*, 9(6), 161.
- López-García, A., Pineda-Quiroga, C., Atxaerandio, R., Pérez, A., Hernández, I., García-Rodríguez, A., & González-Recio, O. (2018). Comparison of Mothur and QIIME for the Analysis of Rumen Microbiota Composition Based on 16S rRNA Amplicon Sequences. *Frontiers in Microbiology*, 9, 3010.
- Lott, M., Hose, G., & Power, M. (2015). Parasitic nematode communities of the red kangaroo, *Macropus rufus*: Richness and structuring in captive systems. *Parasitology Research*, 114(8), 2925–2932.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1), 10–12.
- Miller, I. F., Schneider-Crease, I., Nunn, C. L., & Muehlenbein, M. P. (2018). Estimating infection prevalence: Best practices and their theoretical underpinnings. *Ecology and Evolution*, 8(13), 6738–6747.
- Moritz, C., & Cicero, C. (2004). DNA barcoding: Promise and pitfalls. *PLoS Biology*, 2(10), e354.
- Nassonova, E., Smirnov, A., Fahrni, J., & Pawlowski, J. (2010). Barcoding amoebae: Comparison of SSU, ITS and COI genes as tools for molecular identification of naked lobose amoebae. *Protist*, 161(1), 102–115.
- Nunn, C. L. (2011). *The comparative approach in evolutionary anthropology and biology*. Chicago, IL: University of Chicago Press.
- Nunn, C. L., & Altizer, S. M. (2005). The global mammal parasite database: An online resource for infectious disease records in wild primates. *Evolutionary Anthropology: Issues, News, and Reviews*, 14(1), 1–2.
- Nunn, C. L., Altizer, S., Jones, K. E., & Sechrest, W. (2003). Comparative tests of parasite species richness in primates. *The American Naturalist*, 162(5), 597–614.
- Orme, D., Freckleton, R., Thomas, G., Petzoldt, T., Fritz, S., Isaac, N., & Pearse, W. (2018). Caper: Comparative analyses of phylogenetics and evolution in R. *R Package Version 1.0.1*.
- Paradis, E., & Schliep, K. (2018). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35(3), 526–528.
- Parsons, M. B., Travis, D., Lonsdorf, E. V., Lipende, I., Roellig, D. M. A., Kamenya, S., ... Gillespie, T. R. (2015). Epidemiology and molecular characterization of *Cryptosporidium* spp. in humans, wild primates, and domesticated animals in the Greater Gombe Ecosystem, Tanzania. *PLoS Neglected Tropical Diseases*, 9(2), e0003529.
- Pedersen, M. W., Overballe-Petersen, S., Ermini, L., Der Sarkissian, C., Haile, J., Hellstrom, M., ... Cappellini, E. (2015). Ancient and modern environmental DNA. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 370(1660), 20130383.
- Perelman, P., Johnson, W. E., Roos, C., Seuánez, H. N., Horvath, J. E., Moreira, M. A., ... Rumpler, Y. (2011). A molecular phylogeny of living primates. *PLoS Genetics*, 7(3), e1001342.
- Porazinska, D. L., Giblin-Davis, R. M., Faller, L., Farmerie, W., Kanzaki, N., Morris, K., ... Thomas, W. K. (2009). Evaluating high-throughput sequencing as a method for metagenomic analysis of nematode diversity. *Molecular Ecology Resources*, 9(6), 1439–1450.
- Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., & Glöckner, F. O. (2007). SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research*, 35(21), 7188–7196.
- Pullan, R., & Brooker, S. (2008). The health impact of polyparasitism in humans: Are we under-estimating the burden of parasitic diseases? *Parasitology*, 135(7), 783–794.
- R Core Team (2017). *R: A language and environment for statistical computing (Version 3.4.1)*. Vienna, Austria: R Foundation for Statistical Computing.
- Salleh, F. M., Ramos-Madrigal, J., Peñaloza, F., Liu, S., Sinding, M. H. S., Patel, R. P., & Shamsir, M. S. (2017). An expanded mammal mitogenome dataset from Southeast Asia. *GigaScience*, 6, 1–8.
- Schnell, I. B., Sollmann, R., Calvignac-Spencer, S., Siddall, M. E., Douglas, W. Y., Wilting, A., & Gilbert, M. T. P. (2015). iDNA from terrestrial hematophagous leeches as a wildlife surveying and monitoring tool—prospects, pitfalls and avenues to be developed. *Frontiers in Zoology*, 12(1), 24.
- Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., Levesque, C. A., ... Fungal Barcoding Consortium. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences of the United States of America*, 109(16), 6241–6246.
- Seesao, Y., Gay, M., Merlin, S., Viscogliosi, E., Aliouat-Denis, C., & Audebert, C. (2017). A review of methods for nematode identification. *Microbiological Methods*, 138, 37–49.
- Soininen, J., McDonald, R., & Hillebrand, H. (2007). The distance decay of similarity in ecological communities. *Ecography*, 30(1), 3–12.
- South, A. (2011). rworldmap: A new R package for mapping global data. *The R Journal*, 3(1), 35–43.
- Stephens, P. R., Pappalardo, P., Huang, S., Byers, J. E., Farrell, M. J., Gehman, A., ... Nunn, C. L. (2017). Global Mammal Parasite Database version 2.0. *Ecology*, 98(5), 1476. <https://doi.org/10.1002/ecy.1799>
- Taberlet, P., Bonin, A., Zinger, L., & Coissac, E. (2018). *Environmental DNA: For Biodiversity Research and Monitoring*. Oxford, UK: Oxford University Press.
- Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, 21(8), 2045–2050.
- Tanaka, R., Hino, A., Tsai, I. J., Palomares-Rius, J. E., Yoshida, A., Ogura, Y., ... Kikuchi, T. (2014). Assessment of helminth biodiversity in wild rats using 18S rDNA based metagenomics. *PLoS ONE*, 9(10), e110769. <https://doi.org/10.1371/journal.pone.0110769>
- Titcomb, G., Young, H., & Jerde, C. L. (2019). High-throughput sequencing for understanding the ecology of emerging infectious diseases at the wildlife-human interface. *Frontiers in Ecology and Evolution*, 7, 126. <https://doi.org/10.3389/fevo.2019.00126>
- Walther, B. A., Cotgreave, P., Price, R., Gregory, R., & Clayton, D. H. (1995). Sampling effort and parasite species richness. *Parasitology Today*, 11(8), 306–310. [https://doi.org/10.1016/0169-4758\(95\)80047-6](https://doi.org/10.1016/0169-4758(95)80047-6)
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York, NY: Springer.
- Xu, X., Sui, J., Cao, L., & Lin, H. (2010). Direct competitive enzyme-linked immunosorbent assay (ELISA) for rapid screening of anisakid larvae in seafood. *Journal of the Science of Food and Agriculture*, 90(5), 877–881. <https://doi.org/10.1002/jsfa.3898>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Gogarten JF, Calvignac-Spencer S, Nunn CL, et al. Metabarcoding of eukaryotic parasite communities describes diverse parasite assemblages spanning the primate phylogeny. *Mol Ecol Resour*. 2020;20:204–215. <https://doi.org/10.1111/1755-0998.13101>