

RESEARCH

Open Access



Amplicon sequencing allows differential quantification of closely related parasite species: an example from rodent *Coccidia* (*Eimeria*)

Susana C. M. Ferreira^{1,2*}, Víctor Hugo Jarquín-Díaz^{2,3,4,5,6} and Emanuel Heitlinger^{2,3}

Abstract

Background Quantifying infection intensity is a common goal in parasitological studies. We have previously shown that the amount of parasite DNA in faecal samples can be a biologically meaningful measure of infection intensity, even if it does not agree well with complementary counts of transmission stages (oocysts in the case of *Coccidia*). Parasite DNA can be quantified at relatively high throughput using quantitative polymerase chain reaction (qPCR), but amplification needs a high specificity and does not simultaneously distinguish between parasite species. Counting of amplified sequence variants (ASVs) from high-throughput marker gene sequencing using a relatively universal primer pair has the potential to distinguish between closely related co-infecting taxa and to uncover the community diversity, thus being both more specific and more open-ended.

Methods We here compare qPCR to the sequencing-based amplification using standard PCR and a microfluidics-based PCR to quantify the unicellular parasite *Eimeria* in experimentally infected mice. We use multiple amplicons to differentially quantify *Eimeria* spp. in a natural house mouse population.

Results We show that sequencing-based quantification has high accuracy. Using a combination of phylogenetic analysis and the co-occurrence network, we distinguish three *Eimeria* species in naturally infected mice based on multiple marker regions and genes. We investigate geographical and host-related effects on *Eimeria* spp. community composition and find, as expected, prevalence to be largely explained by sampling locality (farm). Controlling for this effect, the novel approach allowed us to find body condition of mice to be negatively associated with *Eimeria* spp. abundance.

Conclusions We conclude that amplicon sequencing provides the underused potential for species distinction and simultaneous quantification of parasites in faecal material. The method allowed us to detect a negative effect of *Eimeria* infection on the body condition of mice in the natural environment.

Keywords *Eimeria*, Parasite infection, Microbiome, Parasite quantification, Amplicon sequencing, qPCR, Body condition

*Correspondence:

Susana C. M. Ferreira

susana.ferreira@univie.ac.at; susanaferreira@posteo.net

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Amplification and sequencing of marker gene fragments, termed “amplicon sequencing”, is widely used in studies of the bacterial microbiome [1]. Similarly, in biodiversity assessment, amplicon sequencing is used to study the biodiversity of eukaryotes [2, 3]. Commonly applied to bacteria in intestinal systems and eukaryotes in terrestrial and aquatic systems, it is surprising how rarely amplicon sequencing is used for intestinal parasites. Amplicon sequencing identified *Eimeria* species in chickens [4, 5] and in other livestock and wildlife [6], but fewer studies have simultaneously estimated parasite abundance from the same data (but see [7–9] for other parasites). While a main focus of bacterial microbiome studies has been on intestinal “ecosystems” within human and animal hosts, sequencing of symbiotic eukaryotes from intestinal contents or faecal samples is less frequently used for detection of (e.g. parasite) taxa and simultaneous differential quantification (assessment of the abundance of multiple species at once).

The intensity of infection with intestinal parasites is classically estimated by counting parasite (transmission) stages present in the faeces. Although automation has progressed [10, 11], classical “coprological methods” are laborious, making it difficult to quantify parasites in high numbers of samples. They also require high technical and taxonomic expertise to distinguish between closely related species. This means that for samples from natural populations, in which co-infections, often with closely related species, are common, differential quantification of multiple species is almost impossible. For this reason, quantitative polymerase chain reaction (qPCR) has become increasingly popular for the quantification of helminths [12, 13] and unicellular parasites, such as *Coccidia* [14]. DNA-based quantification often shows a considerable discrepancy from counting of transmission stages, but both measures can be complementary in their biological meaning for the host-parasite system, as recently shown for *Coccidia* [15]. Quantitative PCR, however, requires primer pairs specific to the taxon of interest meaning that quantification has to be targeted and is not easily possible for multiple parasite species in a differential manner.

Multiple potential methodological issues might deter parasitologists from embracing amplicon sequencing for parasite quantification. Universal primer pairs for eukaryotic parasites are not commonly agreed on, in contrast to well-established primers and marker fragments for bacterial microbiome studies [16]. While the ribosomal RNA (rRNA) cluster (18S and 28S ribosomal subunits) in eukaryotes provides suitable targets [17], those have not been validated against the eukaryotic microbiome and parasites. The testing of multiple amplicons might

thus be necessary to help the establishment of universal primers for the simultaneous detection of a broad range of taxa. Microfluidic PCR systems compartmentalise and parallelise PCRs in tiny volumes and have been proposed as suitable for easing the work with multiple amplicons [18, 19]. Universal amplification of many parasite taxa not only allows simultaneous differential quantification of multiple (e.g. closely related) target taxa, but is also necessary to provide a background, against which “normalisation” estimates relative abundance. Again, it might be necessary to test multiple approaches for normalisation to optimise quantification with regard to the aims of a study [20–22]. Taxonomic annotation, with the identification of morphologically described species as an ideal commonly aspired, provides the final set of challenges. While modern analysis approaches identify probabilistically likely “amplified sequence variants” (ASVs) [23], their annotation with species names (multiple ASVs might be presenting different species, within species diversity or residual sequencing error) is hampered by imperfect representation of correctly annotated reference sequences in databases [24, 25]. These challenges might seem daunting, but they can be addressed for parasites by careful evaluation of amplicon sequencing against more established techniques.

We here work with *Eimeria* species infecting the house mouse to assess the viability of an amplicon sequencing approach for differential quantification. Three species of *Eimeria* naturally infect the house mouse: *Eimeria ferrisi*, *E. falciformis* and *E. vermiformis* [26, 27]. All species of the genus have a direct life cycle with a predictable genetically determined progression [28]. *Eimeria* spp. have a relatively high prevalence of around 30% (depending on diagnostic method) in natural populations of mice and co-infections have been reported [27]. Laboratory infections of mice with *Eimeria* can easily be conducted and monitored for individual animals, which show a self-limiting infection [15, 29, 30] causing diarrhoea, lack of appetite and weight loss as quantifiable effects [31]. These pathogenic effects are stronger for *E. falciformis* than for *E. ferrisi* [32, 33]. In natural infections, such differences have so far not been detectable.

In this study we used an experimental infection of mice with *E. ferrisi* to benchmark the accuracy of detection and precision of sequencing-based quantification against qPCR. We test whether an amplification in a microfluidics-based device provides a similar quality of quantification to standard PCR and use multiple amplicons in mice sampled in the natural environment to compare marker regions for their taxonomic resolution in very closely related *Eimeria* species. We showcase the methods we established in an assessment of *Eimeria* epidemiology in mice in the natural environment.

Methods

Study design

We evaluate the accuracy and usefulness of amplicon sequencing-based parasite quantification in two settings. Firstly, we use laboratory infections of house mice with *E. ferrisi* to establish sensitivity, specificity and quantitative precision of the method against qPCR measurements. The underlying laboratory infection experiments presented here are the same as those used in our previous work comparing qPCR and classical oocyst counting [15]. Secondly, we test differential quantification in mice naturally infected with three different *Eimeria* species. The samples from the natural environment are largely the same as those we analysed in previous work [27, 30]. Not all samples were available for each method, as some mice had not shed sufficient faeces the quantity of extracted DNA not was in some cases not sufficient for all methods. In a few samples, other issues arose during the pre-processing of the samples. We report sample sizes for each analysis.

Animal husbandry

We obtained in total 22 house mice (*Mus musculus*) of the “wild-derived” inbred strains SCHUNT, STRA, BUSNA and PWD as well as F1 inter-strain crosses [34] from the Institute of Vertebrate Biology of the Czech Academy of Science in Studenec (licence: 61974/2017-MZE-17214). We acclimatised mice to the animal experiment facilities of Humboldt University for at least 1 week before infection. We housed mice in individual cages equipped with tunnels and bedding material for behavioural enrichment and provided them with food and water ad libitum during the experiment.

Parasite inoculum and experimental infection

For infection, we used the Brandenburg64 isolate of *E. ferrisi*, which had been isolated from the faeces of a wild *M. musculus domesticus* mouse captured in Brandenburg, Germany in 2016 and identified by microscopical description and molecular amplification of the 18S rRNA and cytochrome *c* oxidase (*COI*) markers [27]. We obtained oocysts by continuous passage in NMRI (Naval Medical Research Institute) mice and sporulated them as described previously [33]. We infected mice orally with 150 sporulated oocysts in 100 µl of phosphate-buffered saline (PBS 1×, pH 7.4) and monitored them for 11 days. We recorded the weight daily, and a weight loss of 18% was defined as the humane endpoint at which animals had to be sacrificed (experiment license: 0431/17). At the end of the experiment, we euthanised mice (for which the humane endpoint had not been reached before) by cervical dislocation. We collected an average of 0.12 g of faeces (3–4 faecal pellets) from individual mice daily and

stored it after flash-freezing in liquid nitrogen at –80 °C until extraction of DNA.

Sampling of mice in the natural environment

We trapped 672 mice using live traps in 182 farms or houses between 2015 and 2018. The study area ranges from 51.68° to 53.29° latitude (a 200-km-wide area) and from 12.52° to 14.32° longitude (a 140-km-long area). Each year mice were trapped in September to reduce potential seasonal variation. A median of two mice per locality were captured. Mice were individually isolated in cages and euthanised by cervical dislocation within 24 h after capture (animal experiment permit no. 2347/35/2014). Individual mice were measured (body length from nose to anus), weighed and dissected. Faecal pellets were collected from the cages the mice had been housed in overnight and stored at –80 °C after shock-freezing in liquid nitrogen.

DNA extraction

We extracted genomic DNA (gDNA) from faeces collected in the infection experiment and in the wild using the NucleoSpin®Soil kit (Macherey–Nagel GmbH & Co. KG, Düren, Germany) following the manufacturer’s protocol with the following modifications: we performed mechanical lysis of the sample in the Precellys®24 high-speed benchtop homogeniser (Bertin Technologies, Aix-en-Provence, France) using two cycles of disruption at 6000 rpm for 30 s, with 15-s delay between cycles. For each sample, we repeated extraction once to maximise the DNA yield, and nucleic acids were eluted with 40 µl of TE buffer. We assessed the quality and integrity of the DNA using a full-spectrum spectrophotometer (NanoDrop 2000c; Thermo Fisher Scientific, Waltham, MA USA). We quantified concentrations of double-stranded DNA (dsDNA) using a Qubit® Fluorometer and the dsDNA BR (broad-range) Assay Kit (Thermo Fisher Scientific). We adjusted DNA extracts to a final concentration of 50 ng/µl with nuclease-free water (Carl Roth GmbH & Co. KG) and stored them at –80 °C until further processing.

Real-time qPCR

As a reference measurement, we quantified *Eimeria* DNA by qPCR amplification of a 140-base-pair (bp) fragment of the mitochondrially encoded *COI* gene (*COI*) using *Eimeria*-specific primers Eim_COI_qX_F 5′-TGT CTATCACTTGGGCTATTGT-3′ and Eim_COI_qX_R 5′-GGA TCACCGTTAAATGAGGCA-3′. Each reaction contained 1× iTaq™ Universal SYBR® Green Supermix (Bio-Rad Laboratories, Hercules, CA, USA), 400 nM forward and reverse primers and 50 ng template gDNA in a total reaction volume of 20 µl. We performed the

reactions in either the ABI 7300 Real-Time PCR System (Applied Biosystems, Thermo Fisher Scientific, Foster City, CA, USA) or the MasterCycler® RealPlex2 machine (Eppendorf, Hamburg, Germany). For both PCR systems, cycling conditions consisted of an initial denaturation at 95 °C for 2 min; 40 cycles of denaturation at 95 °C for 15 s, annealing at 55 °C for 15 s and extension at 68 °C for 20 s. Data were collected at the end of each cycle. Melting curve analysis was included to discard primer dimer formation and non-specific amplification: after the last amplification cycle, the temperature was increased from 65 °C to 95 °C with 0.5 °C increments and 3 s per step. We performed amplifications in triplicate, and each run included a non-template control (NTC). We analysed melting curves blindly for the presence of distinct “*Eimeria* products” and PCR artefacts. We labelled samples with all three replicates showing the melting temperature (T_m) in the range of $74.1\text{ °C} \pm 1.78\text{ °C}$ (observed for positive controls) as “qPCR-positive”, samples with only one or two of the triplicates showing a correct peak, were designated as negative samples. For qPCR-negative samples, we set the estimated DNA quantity to zero. For positive samples, we predicted genome copies from a linear model fitted for a standard curve obtained from a known number of oocysts [15]. We adjusted these predictions by dividing them by the amount of starting material the respective DNA was extracted from to obtain an estimate of “genome copies per gram of faeces”.

Library preparation and sequencing

In total, 236 faecal DNA preparations from the laboratory experiment and 672 from mice sampled in the natural environment were used for a multimarker amplification using the microfluidics-based PCR system Fluidigm Access Array 48×48 (Fluidigm, San Francisco, CA, USA). Samples were randomised in their order and amplified in parallel with non-template negative controls using a microfluidic PCR. This makes it possible to amplify multiple fragments (amplicons) for different marker genes (primer pairs in Additional file 1). In an alternative approach, for validation of the microfluidic amplification and sequencing of multiple amplicons, 210 laboratory samples were amplified using the primer pair 5′-GAATTGACGGAAGGGCACC-3′ and 5′-AAGGGC ATCACAGACCTGTTAT-3′, targeting the V6-V7 region of the 18S rRNA gene, in standard reaction volumes (96 well microtiter plates). This primer pair has been previously tested to amplify gastrointestinal eukaryotes and showed good coverage of Apicomplexa [9].

For both PCR setups library preparation was integrated into the amplification procedure and was performed according to the protocol Access Array Barcode Library for Illumina Sequencers (single direction indexing) as

described by the manufacturer (Fluidigm, San Francisco, CA, USA). The amplicons were quantified (Qubit fluorometric quantification dsDNA High Sensitivity Kit, Thermo Fisher Scientific, Waltham, MA, USA) and pooled in equimolar concentration. The final library was purified using Agencourt AMPure XP Reagent beads (Beckman Coulter Life Sciences, Krefeld, Germany). The quality and integrity of the library were confirmed using the Agilent 2200 TapeStation with D1000 ScreenTapes (Agilent Technologies, Santa Clara, CA, USA). Sequences were generated at the Berlin Center for Genomics in Biodiversity Research (BeGenDiv) on the Illumina MiSeq platform (Illumina, San Diego, CA, USA) using v2 chemistry with 500 cycles (one run each for 210 and 236 laboratory samples with single and multiple amplicon products, respectively, and four runs for 672 wild mouse samples with multiple amplicon products). All sequencing raw data can be accessed through the BioProject PRJNA548431 in the NCBI Short Read Archive (SRA).

Identification and quality screening of ASVs

We used the R packages dada2 [23] and MultiAmplicon [35] to filter, sort, merge, denoise and remove chimaeras for each run separately and for each amplicon. Additionally, we used the package DECONTAM [36] to remove contaminants and sequencing errors using “prevalence” and “frequency” methods (method=“combined”). We removed ASVs that have less than 1% prevalence, less than 0.005% relative abundance [37] and samples with less than a total sum of 100 reads. Filtering was done individually for each amplicon in the multi amplicon datasets and then all amplicon’s products were collated into one “phyloseq” object using the function “merge_phyloseq” implemented in the package “phyloseq” [38]. For the infection experiment, this resulted in 200 samples for the standard PCR and 218 samples for the microfluidic PCR. 619 samples for the microfluidic PCR on wild mouse samples were available for further sequence analysis after the same procedure.

Taxonomic annotation of ASVs

A first taxonomic assignment of the resulting ASVs was performed based on the amplicon target with the assign-Taxonomy function from dada2, using the RDP classifier [39]. 18S and 16S rRNA sequences were classified against SILVA 138.1 SSU Ref NR 99, 28S rRNA against SILVA 138.1 LSU Ref NR 99 databases [40] and ITS rRNA against the UNITE database [41]. Sequences and taxonomies from all other targeted regions which do not have a publicly available curated database were downloaded from NCBI, sequences with more than 5 degenerated bases and with lengths less than 300 bases were removed using RESCRIPt [42]. All databases were dereplicated

using RESCRIPt [42]. For the present study, we focus on ASVs annotated as *Eimeria* in the amplicons targeting 18S rRNA and 28S rRNA genes.

To refine the taxonomic annotation, we constructed phylogenetic trees. To do so, we constructed alignments using the function AlignSeqs of the package DECIPHER [43] with “iterations=20” and “refinements=20”. We used maximum likelihood models and ModelFinder [44] for selecting the most appropriate evolution model, as implemented in iqtree2 [45]. The selected best-fitting model was TN+G4+I. We assessed branch support with an ultrafast bootstrap approximation (UFBoot) with 5000 replicates [46]. For the 18S rRNA gene, we included sequences of *Eimeria* species previously detected in house mice and other rodents [27], all 18S rRNA *Eimeria* ASVs recovered from this study and 18S rRNA sequences from the outgroup *Isospora* sp. ex *Talpa europaea*. Trees were visualised with iTOL version 6.7.4 [47].

The consensus tree resulted in poorly supported clades. Thus we constructed separate phylogenetic trees for all ASVs from each amplicon targeting the 18S rRNA gene, together with the *Eimeria* reference and outgroup sequences. We assigned a species name to each ASV that shared the most recent common ancestry with the reference sequences of *E. ferrisi*, *E. falciformis* and *E. vermiciformis* in the rooted tree using the ape [48] and ggtree [49] packages. Bootstrap support values for each species clade were recorded in each tree, and indicate the reliability of species assignment for the respective amplicon's ASVs (Additional file 2). One 18S rRNA ASV remained unassigned. We performed a BLAST search for each 28S rRNA ASV on the *E. falciformis* reference genome (ASM227181v1) [50]. Due to the lack of other reference sequences for the *Eimeria* 28S rRNA gene, we constructed a phylogenetic tree without those, including the eight 28S rRNA gene ASVs from the present study and the corresponding region of the *E. falciformis* genome to the 28S rRNA ASVs. For this tree, we found the substitution model TPM3u+G4+I to provide the best fit and used it in phylogenetic reconstruction. Branch support was estimated after 5000 replicates. All alignments and phylogenetic trees with respective GenBank accession codes, including a tree with only reference sequences for comparison, are in Additional file 3.

To transfer taxonomic annotation to ASVs without phylogenetic clustering with reference sequence (i.e. when reference sequence was unavailable for the amplified region, as for 28S rRNA or when assignment is poorly supported in the phylogenetic analysis) we created a co-abundance network with all *Eimeria* ASVs allowing only significant Pearson correlation coefficients after adjusting for multiple testing using the Benjamini–Hochberg method.

Normalisation

Normalisation contextualises the abundance of *Eimeria* sequences in the samples using the overall sequencing read count for other taxa (fungi and Eukaryotic parasites in case of 18S rRNA gene, also bacteria and archaea in case of 16S rRNA gene, as used in simultaneous microbiome PCRs on mice sampled in the natural environment). We evaluated different normalisation techniques by comparing correlation coefficients (Pearson's rho) of sequencing-derived abundance with qPCR abundance. We tested the following normalisations: (1) total sum scaling (TSS), (2) trimmed mean by M-value (TMM) implemented in “microbial” [51], (3) centred log-ratio (CLR) as implemented in the “microbiome” package [52] and (4) rarefaction, as implemented in “phyloseq” [38].

We tested the significance of the differences between correlations with back-transformed averaged Fisher's Zs [53] based on dependent groups with overlapping correlations, implemented with the package “cocor” [54]. We report *P*-values adjusted to multiple testing using the Benjamini–Hochberg method [55]. Based on the results of this comparison (Additional file 4: Table S1, Figures S1, S2), we decided to apply TSS normalisation for further analysis.

Eimeria detection and quantification

We analysed the sensitivity and specificity of *Eimeria* detection against our reference method, qPCR, in experimentally infected laboratory mice. Sensitivity is defined here as the proportion of amplicon sequencing-positive samples in qPCR-positives (proportion of “true positives”). Specificity is the proportion of qPCR-negative samples in sequencing-negative samples (proportion of “true negatives”).

For the analysis of quantitative precision, we excluded samples with no *Eimeria* detection in qPCR or amplicon sequencing. We analysed whether the abundance of all ASVs annotated to the genus *Eimeria* within a sample correlates with *Eimeria* genome copies/ng DNA using Pearson correlation on log-transformed values.

We used linear mixed-effects models (lme4) to test whether the abundance of individual ASVs predicts qPCR-derived abundance (*Eimeria* genome copies). We include days post-infection (dpi) as a random effect to control for increasing DNA abundance during the infection. All lme4 were performed with the package lme4 [56]. We tested the significance of random effects with the ranova function of the package “lmerTest” [57] and fixed effects with likelihood ratio tests (LRT; compared against an F-distribution) using the anova function to compare between the full model and a model with the predictor removed. The global goodness of fit was examined as the LRT against intercept-only models.

Model assumptions were investigated with diagnostic residual plots. The variance explained (R^2) was calculated as in Nakagawa and Schielzeth [58].

Analysis of *Eimeria* community structure

Adjusted prevalence and respective 95% confidence intervals were calculated with Sterne's exact method [59] implemented in the package "epiR" [60]. We analysed the variation in *Eimeria* spp. community composition by the marginal effects of locality of capture, year, host sex and host body mass index (BMI) and controlled for the effect of different sequencing depths by including a categorical factor indicating the sequencing run for each sample. For that, we used permutational multivariate analysis of variance (PERMANOVA) on a matrix of Jaccard similarity coefficient implemented with the package vegan [61], function `adonis2`, parameter `by="margin"`. BMI was calculated as body weight divided by squared body length. We then tested the associations between BMI and *Eimeria* spp. infection status and abundance by applying mixed-effects linear models with either the infection status or the abundance of infection of *E. ferrisi*, *E. falciiformis* and *E. vermiformis* as fixed effect and locality as random effect.

Results

High precision of amplicon sequencing-based quantification

Sequencing of an 18S rRNA gene fragment amplified with a standard PCR approach in faecal samples of mice experimentally infected with *E. ferrisi* resulted in four ASVs in the genus *Eimeria*. Sequencing of the same amplicon, prepared with a microfluidic PCR protocol, resulted in two ASVs for *Eimeria* that correspond to the two most abundant ASVs produced with the standard PCR approach. The abundance of the two shared ASVs was highly correlated (ASV1: Pearson's $\rho=0.97$, $df=180$, $P<0.001$; ASV2: Pearson's $\rho=0.92$, $df=180$, $P<0.001$, Fig. 1). This indicates consistency in *Eimeria* quantification between standard and microfluidic PCR.

As "background amplification", i.e. all amplified taxa, the standard PCR amplification targeting the 18S rRNA gene resulted in 57 sequenced taxa (four of these annotated as *Eimeria*), across 200 samples with a median sampling depth of 6245 reads and *Eimeria* ASV median sampling depth of 109 reads. The microfluidic PCR amplification was successful for 13 amplicons, resulting in 597 sequenced taxa (two of these annotated as *Eimeria*) across 218 samples with a median sampling depth of 7126 reads (17 median *Eimeria* sequencing depth). Out of these, the same amplicon targeted in the standard PCR amplification resulted in 43 sequenced taxa, including the two *Eimeria* ASVs, across 214

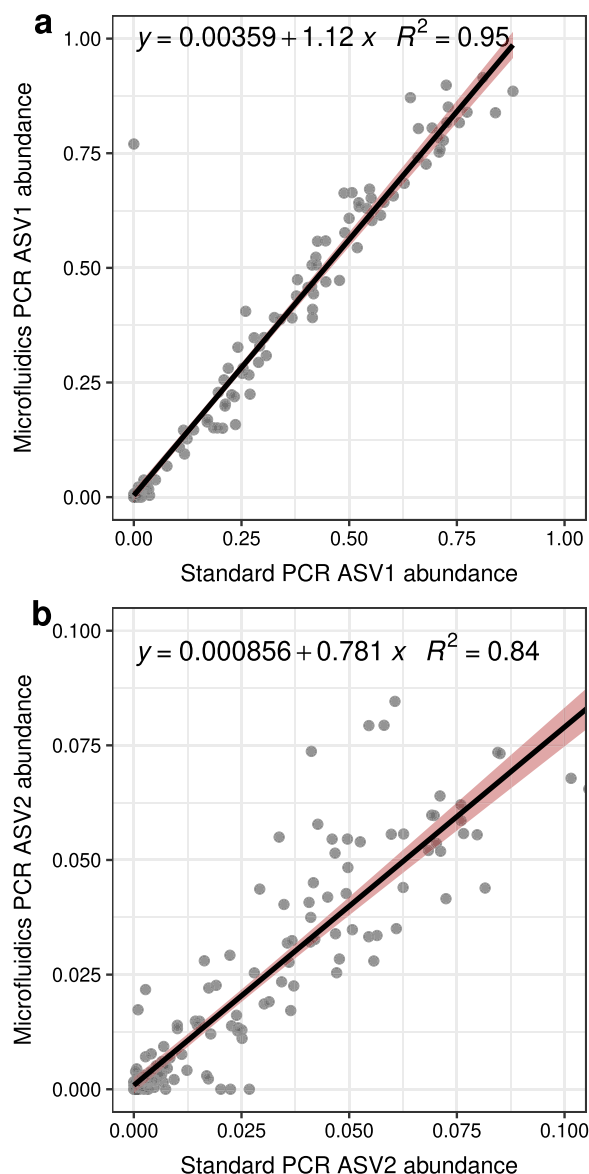


Fig. 1 The same amplicon sequence variants (ASVs) annotated as *Eimeria*, sequenced from standard PCR and microfluidic PCR approaches are proportional. **a** ASV1, the most abundant ASV in both approaches, and **b** ASV2. Linear equation and coefficient of determination (R^2) are reported for each curve. Regression lines are in black and shaded areas represent 95% confidence intervals. Abundance is expressed in relative abundance after total sum scaling

samples, with a median sampling depth of 2694 reads, and *Eimeria* ASV median sampling depth of 545 reads.

We tested different normalisations commonly applied in microbiome studies for their performance on our *Eimeria* ASV dataset. To do so we computed (Pearson) correlations between qPCR-based quantification (in genome copies per ng/DNA) and ASV abundance (summed over different ASVs). We compared the

correlations for each normalised dataset against unnormalised data (sequencing read counts). Tested normalisations include TSS, TMM, CLR and rarefaction. The applied normalisations did not greatly impact the resulting correlations. However, TSS and rarefaction demonstrated the best agreement with qPCR-based abundance (Additional file 4: Figures S1, S2, Table S1). We thus decided to report data normalised with TSS.

Generally, *Eimeria* ASV abundance (summed of the two or four ASVs, respectively) was highly correlated with qPCR-derived abundance (genome copies/ng of DNA) for both sequencing approaches (standard PCR: $\rho=0.93$, $t=31.987$, $df=150$, $P < 0.001$; microfluidic PCR: $\rho=0.89$, $t=21.398$, $df=126$, $P < 0.001$; Fig. 2, Additional file 4: Figures S1, S2). We further tested the predictive value of each individual ASV abundance against qPCR-derived abundance using generalised linear mixed modelling (log-likelihood ratio tests (LRT) for standard PCR: $\chi=82.4$, $df=4$, $P < 0.001$; microfluidic PCR: $\chi=47.115$, $df=2$, $P < 0.001$, Table 1). ASV3 and ASV4 were not found to be predictive for qPCR-measured DNA abundance and are thus likely residual “noise” from sequencing errors. In contrast, ASV1 and ASV2 were significantly associated with qPCR-derived abundance for both amplification approaches. Overall, the variance explained by ASVs and dpi effects is very high (standard PCR conditional $R^2=0.89$; microfluidic PCR conditional $R^2=0.79$). ASVs explain much of the variance in the model (standard PCR marginal $R^2=0.66$; microfluidic PCR marginal $R^2=0.52$), even when controlling for dpi, which explains a large proportion of the variance in *Eimeria* abundance, as there is a clear progression of the infection through time (LRT for dpi as random effect in standard PCR: $\chi=42.477$, $df=1$, $P < 0.001$; microfluidic PCR: $\chi=8.694$, $df=1$, $P=0.003$, Additional file 4: Figure S3). We can regard DNA quantification provided by amplicon sequencing as highly precise.

Amplicon sequencing-based detection has imperfect specificity and sensitivity

We compared the sensitivity and specificity, of the merged ASVs annotated to *Eimeria* and individual ASVs. As for quantitative precision, we again compare standard amplification in microtiter plates with microfluidic amplification using the melting curve analysis in our qPCR as the “gold standard” for reference (Table 2). We found 13 false positives after standard PCR, but only three false positives with microfluidic PCR amplification. These have relatively low abundance when compared to the mean average of true-positive samples (mean \pm standard deviation standard PCR: 0.001 ± 0.0009 vs 0.28 ± 0.27 ; microfluidic PCR: 0.26 ± 0.45 vs 0.35 ± 0.30). In contrast, we found 10 false negatives with standard PCR, but 35

with microfluidic PCR. This means that PCR amplification with standard methods had a relatively high sensitivity, but lower specificity while microfluidic amplification had inversely relatively low sensitivity but higher specificity. Both methods were imperfect in the detection of *Eimeria* infection compared against a highly sensitive qPCR.

Amplicon sequencing distinguishes and quantifies three *Eimeria* species in wild house mice

We sequenced faecal samples of wild-caught mice using amplification of multiple amplicons on a microfluidic device. In the background, 34 amplicons were successfully amplified, resulting in 2880 taxa sequenced across 619 samples, with a mean sampling depth of 15056 reads. Out of these, we retrieved 37 *Eimeria* ASVs from 10 amplicons targeting the 18S rRNA gene and one targeting the 28S rRNA gene amplicon. *Eimeria* ASV median sampling depth was 273 reads. 163 mice (adjusted prevalence: 28.1%, 95% CI 23.5–33.1) showed an infection based on amplicon sequencing. The distribution of *Eimeria* ASVs per amplicon is represented in Additional file 5: Figure S4a. Assigning species annotation to ASVs will allow species quantification based on combined ASV abundance per species, but is highly dependent on the correctness of taxonomic annotation.

We thus improved taxonomic annotation, relative to established sequence-similarity-based methods, using a phylogenetic approach. This allowed us to resolve annotations for all but one of the 18S rRNA gene ASV with reference sequences of three *Eimeria* species known to infect house mice (Fig. 3). As there are no house mouse *Eimeria* reference sequences for the 28S rRNA gene, other than *E. falciformis*, phylogenetic analysis clustered the ASVs without allowing assignment of a species name at this point (Fig. 4). Two ASVs cluster with *Eimeria falciformis*, whereas six ASVs cluster together, suggesting the presence of two species. We then created a co-abundance network with all *Eimeria* ASVs showing that ASVs form three separate abundance clusters. Each of these abundance clusters contains ASVs annotated to one *Eimeria* species. Based on the position of the phylogenetically unassigned ASVs, it was then possible to assign them to a species: The unassigned ASVs from the 28S rRNA amplicon are divided between the *E. ferrisi* (six ASVs) and *E. falciformis* (two ASVs) clusters (Figs. 4, 5). The phylogenetically unassigned 18S rRNA gene ASV can be assigned to *E. falciformis* by this co-occurrence network. Overall, we were able to assign species to ASVs irrespective of the presence of reference sequences with a combination of phylogenetic analysis and a co-occurrence network. We identified 124 mice (adjusted

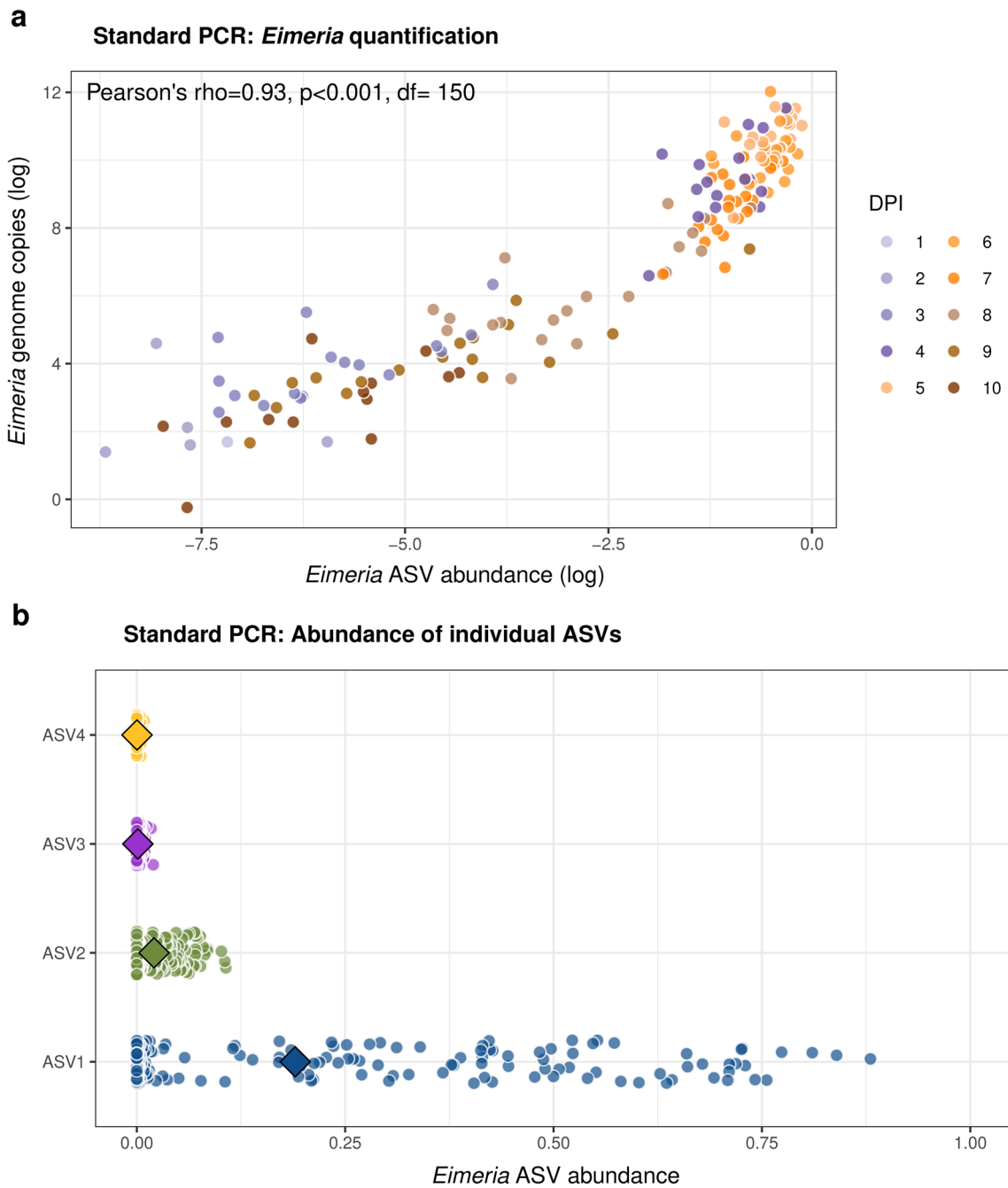


Fig. 2 Precise *Eimeria* quantification with amplicon 18S rRNA gene fragment sequencing. **a** Relationship between *Eimeria* DNA measured with qPCR and ASVs from sequenced standard PCR. **b** Abundance distribution for each individual *Eimeria* ASV sequenced from standard PCR amplification. The mean for each ASV abundance is depicted as a diamond. dpi: days post-infection. Abundance is expressed as relative abundance after total sum scaling

prevalence: 19.5%, 95% CI 15.3–24.1%) infected with *E. ferrisi*, 58 mice (adjusted prevalence: 9.4%, 95% CI 7.2–11.9%) infected with *E. falciformis* and nine mice

(adjusted prevalence: 1.4%, 95% CI 0.7–2.7%) infected with *E. vermiformis*. The distribution of *Eimeria* ASVs per species is presented in Additional file 5, Figure S4b.

Table 1 The relationship of individual ASV abundance with qPCR-derived *Eimeria* genome copies/ng while controlling for days post-infection as a random effect

		Estimate	SE	t-value	P-value	F-value	df	P-value
(a)								
Standard PCR	Intercept	4.119	0.470	8.756	< 0.001*			
	ASV1	8.894	0.840	10.592	< 0.001*	34.665	1	< 0.001*
	ASV2	28.29	4.426	6.391	< 0.001*	4.567	1	< 0.001*
	ASV3	-12.064	26.970	-0.447	0.655	0.200	1	0.655
	ASV4	20.027	59.011	0.339	0.735	0.115	1	0.735
(b)								
Microfluidic PCR	Intercept	4.961	0.476	10.412	< 0.001*	-	-	-
	ASV1	6.389	0.785	8.134	< 0.001*	66.169	1	< 0.001*
	ASV2	27.096	5.555	4.878	< 0.001*	23.797	1	< 0.001*

(a) ASVs sequenced from standard amplification, n (samples) = 152, groups (dpi) = 10. (b) ASVs sequenced from microfluidic amplification, n (samples) = 128, groups (dpi) = 10. ASV abundance is the relative abundance after total sum scaling for each amplicon

SE, standard error; *Statistically significant

Table 2 Detecting *Eimeria* sensitivity and specificity for sequencing of *Eimeria* ASVs produced with standard PCR and microfluidic PCR. *Eimeria* detection with qPCR (genome copies per nanograms of DNA) is the reference standard

	All <i>Eimeria</i> ASVs (%)	ASV1 (%)	ASV2 (%)
Sequenced with standard PCR			
Sensitivity	93.8	83.3	92.6
Specificity	64.9	89.2	64.9
Sequenced with microfluidic PCR			
Sensitivity	78.5	69.9	73.6
Specificity	94.1	98.0	94.1

All *Eimeria* ASVs: the sum of all ASVs annotated to *Eimeria* for each sample

***Eimeria ferrisi* and *Eimeria falciformis* infection is associated with reduced body condition**

To test the usability of amplicon sequencing-derived quantification, we then asked two questions: (i) whether the occurrence of different species is clustered within sampling localities and b) whether the body condition of mice is associated with *Eimeria* spp. infection. BMI was not available for 12 mice and was therefore removed for analysis. For these questions, we first analysed a multivariate response and found a strong effect of mice capture location on *Eimeria* spp. composition (N=161, Permanova on Jaccard index: R²=0.47, pseudo-F=1.340, P<0.001, Additional file 5: Table S2), meaning that about half of the occurrences of infection with a particular species were explained by specific parasites circulating in a local population. This is an expected result from an epidemiological point of view, but also a confirmation of the detection accuracy of amplicon sequencing in a natural population. More

strikingly, in a biological sense, we also found a smaller but still significant association of BMI (R²=0.02, pseudo-F=3.752, P=0.002, Additional file 5: Table S2) on *Eimeria* species composition, meaning that about 2% of the variation in *Eimeria* spp. composition was associated with the body condition of the mice. We controlled for the effect of different sampling depths due to samples being sequenced in different sequencing runs and found a small and not significant effect (R²=0.003, pseudo-F=0.557, P=0.849).

We further investigated this finding with two models using BMI as a (univariate) response variable and (a) the infection status of *Eimeria* spp. as predictors (infected and uninfected, N=607, LRT: χ=10.064, df=3, P=0.018, Table 3a) and (b) the intensity of *Eimeria* spp. as predictors (N=161, LRT: χ=16.145, df=3, P=0.001, Table 3b). We controlled for location effects in these models (which were, as expected, significant: random effect for infection status model: LRT=35.269, df=1, P<0.001, intensity model: LRT=10.336, df=1, P=0.001). We found that animals infected with *E. falciformis* have a lower BMI when compared to uninfected animals (Fig. 6a, Table 3a), and the intensity of infection with *E. ferrisi* and *E. falciformis* abundance have a negative effect on BMI (Fig. 6b–c, Table 3b). Overall, the *Eimeria* spp. infection status and intensity of infection explained a substantial proportion of BMI variation (infection status model: marginal R²=0.02, conditional R²=0.18; infection intensity model: marginal R²=0.09, conditional R²=0.40). This demonstrates the usability of amplicon sequencing-derived detection and intensity estimates for biological questions.

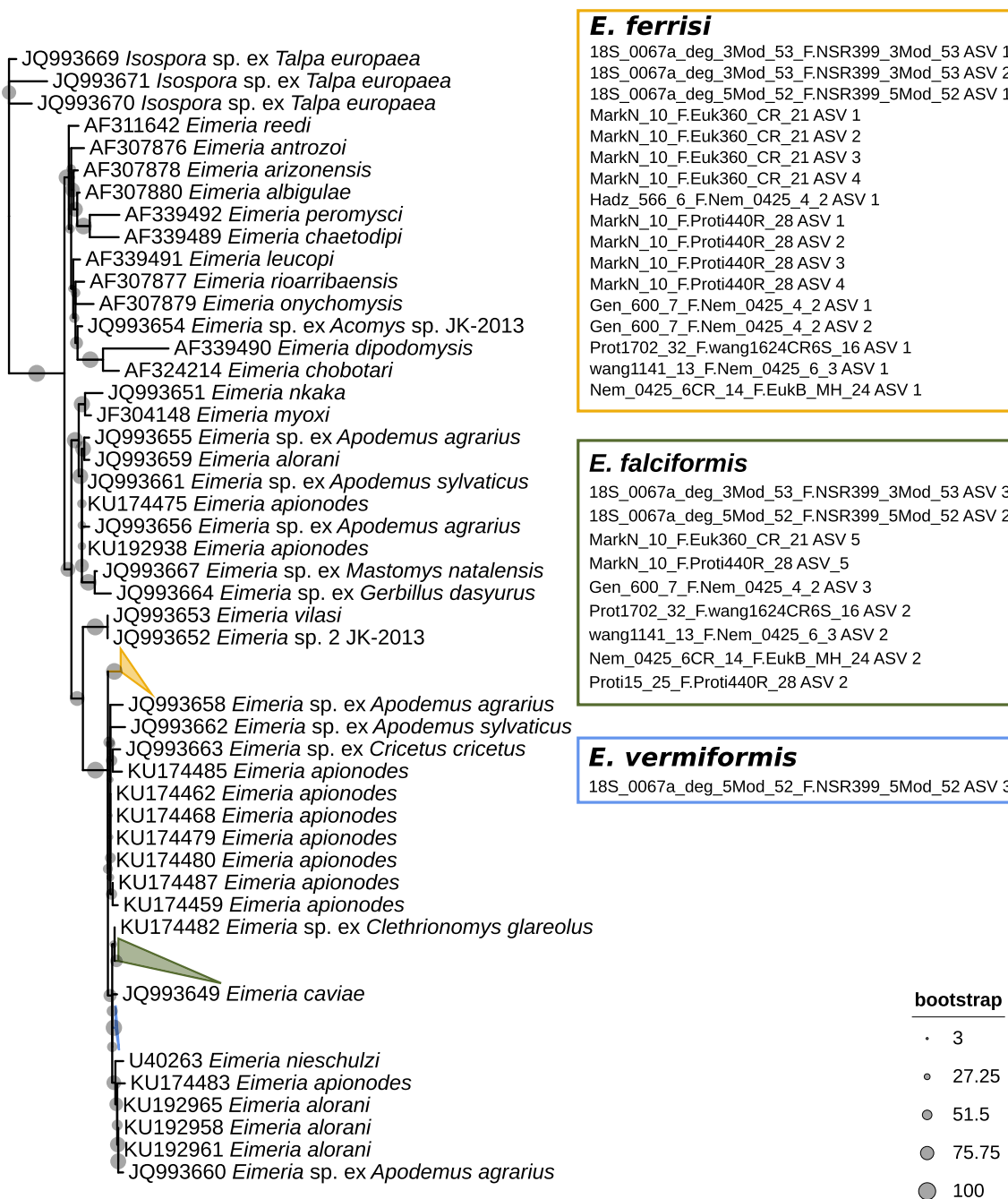


Fig. 3 Phylogenetic analysis of *Eimeria* spp. for 18S rRNA gene reference sequences. Clades containing all reference sequences for *E. ferrisi* (yellow), *E. falciformis* (green) and *E. vermiformis* (blue) are collapsed and coloured. Coloured boxes represent the species assignment of ASVs based on phylogenetic clustering

Discussion

Quantification of Coccidia based on DNA abundance is a suitable addition or alternative to classical coprological counts of transmission stages for Coccidia [15]. Assessment of DNA abundance using qPCR, however, does

not allow scientists to distinguish different parasite species simultaneously. Sequencing of PCR products might provide an alternative to permit differential quantification of multiple species at once, but such methods are not well established and have seen relatively little use by

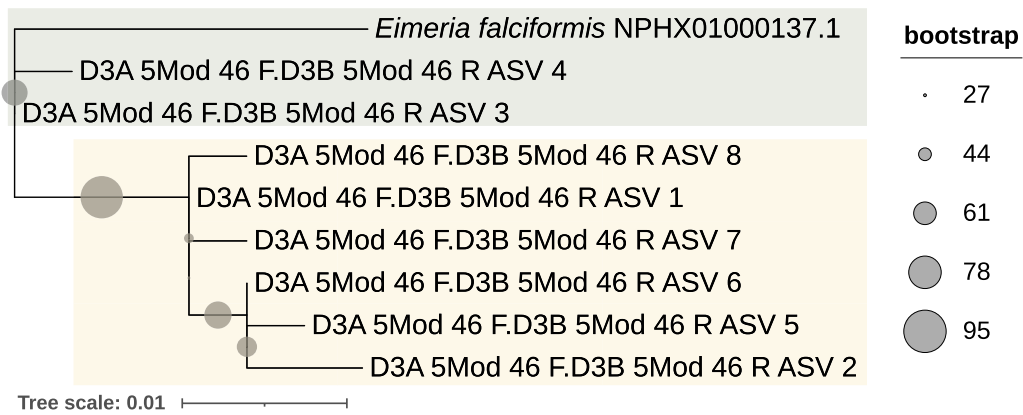


Fig. 4 Phylogenetic analysis for *Eimeria* spp. based on the 28S rRNA gene. Shaded tips represent *E. ferrisi* (yellow) and *E. falciformis* (green) annotation according to ASV co-occurrence network

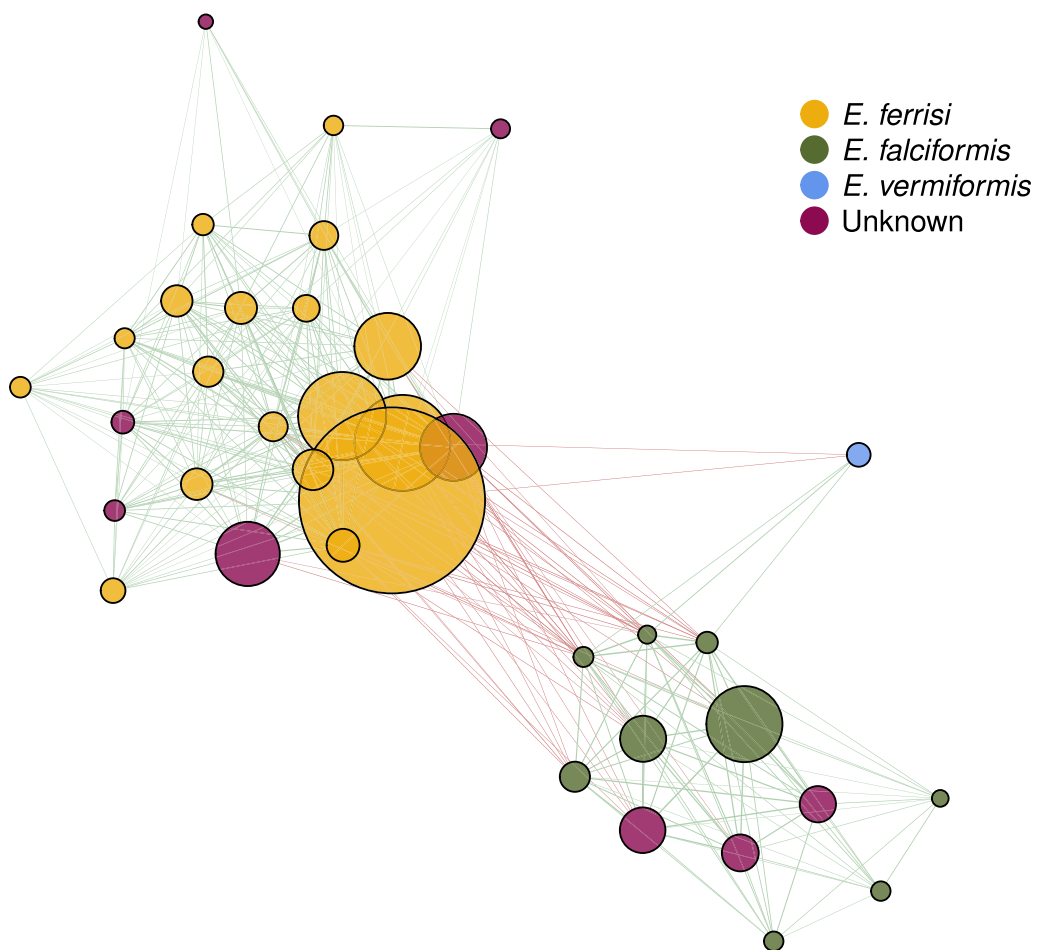


Fig. 5 Co-occurrence network of ASVs annotated as three *Eimeria* species based on phylogenetic analysis. Nodes represent all ASVs annotated to *Eimeria* sequenced with a multiple amplicon approach. Node size reflects the frequency and relative abundance of each ASV: the relative abundance after total sum scaling within each amplicon, and summed across amplicons for each sample. Edges represent significant Pearson correlations after adjusting for multiple testing. Green edges mark positive correlations and red edges mark negative correlations

Table 3 Generalised linear mixed effect models investigating (a) the infection status of *Eimeria* spp. (infected vs uninfected) and (b) the effects of *Eimeria* spp. abundance on the body mass index of house mice from a natural population

	Estimate	SE	t-value	P-value
(a) <i>Eimeria</i> spp. presence effects on body mass index, N=607, localities=171				
Intercept	0.00252	0.00002	132.173	<0.001 *
<i>E. ferrisi</i>	-0.00001	0.00003	-0.423	0.672
<i>E. falciformis</i>	-0.00012	0.00005	-2.370	0.018 *
<i>E. vermiformis</i>	-0.00012	0.00011	-1.015	0.311
(b) <i>Eimeria</i> spp. infection intensity effects on body mass index, N=163, localities=67				
Intercept	0.00258	0.00004	62.904	<0.001 *
<i>E. ferrisi</i>	-0.00005	0.00002	-3.132	0.002 *
<i>E. falciformis</i>	-0.00011	0.00003	-3.189	0.002 *
<i>E. vermiformis</i>	-0.00025	0.00033	-0.750	0.446

The location where mice were captured is included as a random effect. Abundance is the relative abundance after total sum scaling within each amplicon, and summed across amplicons for each sample

SE, standard error; *Statistically significant

parasitologists. The general applicability of amplification of universal marker regions and DNA sequencing makes this differential quantification of multiple species attractive for parasitologists working on any parasite shedding DNA into host faeces.

Hinsu et al.[5] reported higher sensitivity for single amplicon sequencing of the 18S rRNA gene than for species-specific qPCR in chicken sampled in farms and reported the detection of multiple *Eimeria* species. As discussed by the authors, an alternate explanation of a higher false-positive rate in the amplicon sequencing approach cannot be excluded. In this regard, using experimentally infected animals is more suitable to evaluate the detection and quantification of amplicon sequencing. In our previous work, we have shown that the qPCR accurately quantifies both *E. ferrisi* and *E. falciformis*, with about 0.6% variation in standard curves attributed to species differences [15]. This means our qPCR is equally applicable to the *Eimeria* species with high prevalence in house mice. Compared against qPCR, we show that *Eimeria* DNA can be accurately quantified using amplicon sequencing, while sensitivity and specificity of detection are—when used with caution, tolerably—imperfect. We tested two different amplification procedures and sequenced the resulting products. Both approaches resulted in identical ASVs, meaning that differences between the two methods can be partially attributed to differences in the depth of sequencing: sensitivity of detection is higher when sequencing more deeply, whereas specificity is higher at more shallow

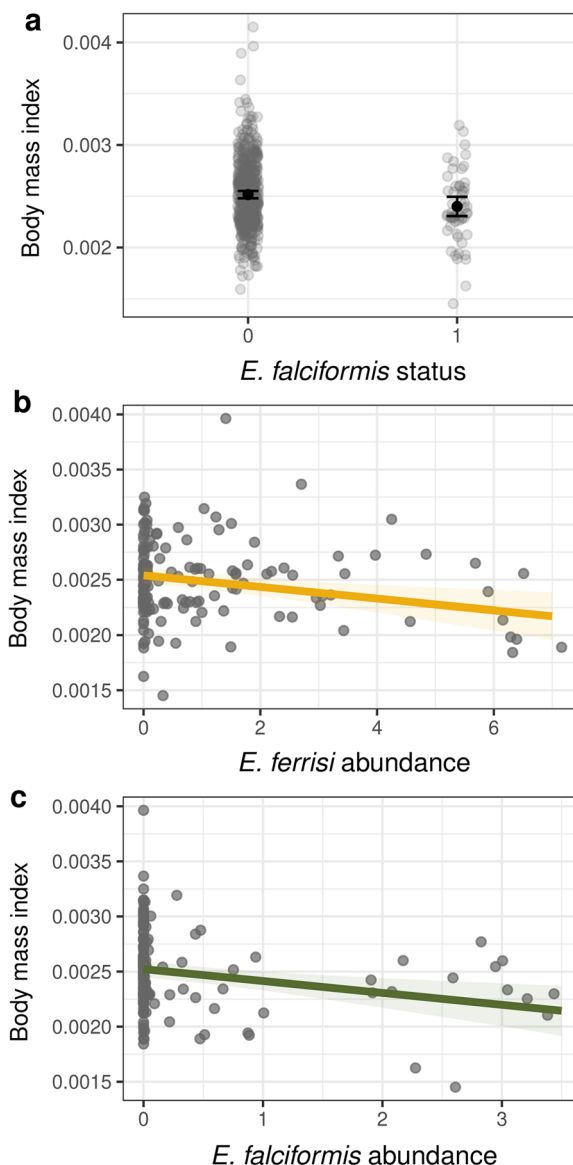


Fig. 6 *Eimeria* infection and intensity predict poor body condition in mice. **a** *Eimeria falciformis* infection status, **b** *Eimeria ferrisi* abundance and **c** *Eimeria falciformis* abundance association with body mass index. Grey points represent the relationship (raw data) between body mass index and *Eimeria* spp. infection status or abundance. The mean predicted effect from a linear mixed model and associated 95% confidence intervals are represented as black points. Lines with respective coloured shades represent the predicted mean effect corresponding 95% confidence intervals. Abundance is the *Eimeria* spp. ASV relative abundance after total sum scaling within each amplicon, and summed across amplicons for each sample

sequencing depth. This is analogous to classical coprological techniques, in which the assessed volume of the sample determines sensitivity. Because false positives are likely resulting from cross-contamination, common in amplicon sequencing [36, 62], the microfluidic

amplification in closed chambers might provide additional specificity here. For both amplification schemes, however, the sensitivity and specificity of sequencing are considerably lower than the 100% reported in coprological trials with spiked oocysts [63]. There is, however, a lower limit of detection for coprological techniques [64], which are often hit in samples from natural populations. Specificity in the sense of avoiding false positives is hence dependent on amplification procedure and sequencing depth, but we argue that this is tolerable in applications on real-world samples. It is tolerable, because specificity in the sense of being able to detect and quantify different species differentially is more important. Additionally, quantitative precision in the comparison between the two amplification methods (~97%) and between both sequencing-based assessments and qPCR (~93%) was very high. To put this in perspective, this precision is higher than that between different flotation-counting techniques for chicken *Eimeria* [63].

We show that dedicated taxonomical annotation via phylogenetic analysis of ASVs can identify and differentiate *E. ferrisi*, *E. falciformis* and *E. vermiformis* in mice from a natural population. These species are so closely related, that they are difficult to distinguish even with classical, non-quantitative sequencing approaches [26, 27]. It is hence necessary to use a modern approach to sequence processing, inferring probabilistically plausible sequence variants (ASVs), at a resolution of a single nucleotide difference [23] instead of operational taxonomic units (OTUs) collapsed at similarity thresholds. At single nucleotide resolution we detect intraspecific variability of even the relatively conserved 18S rRNA regions we amplified for *E. ferrisi* from our laboratory infection. We show that this variation is credible in two out of four ASVs and not arising from sequencing error or “chimeric” PCR artefacts [65], as abundance trajectories are consistent within individual mice. The two credible ASVs might result from intragenomic variation within single cells [66], or from variation in different merozoites, as suggested for *Sarcocystis* spp. [67, 68]. Interestingly, we detected the less abundant of the credible ASVs for *E. ferrisi* consistently earlier during controlled laboratory infection, meaning that the 18S rRNA gene variant might be associated with a genetic variant with faster development, such as those selected for precocious lines [69, 70]. By applying a phylogenetically guided taxonomic annotation in natural samples, we detect the three *Eimeria* species previously described in our study area [27] with multiple amplified fragments of the 18S rRNA gene. To unify taxonomic annotation across those amplicons, we use the correlation of abundance across different samples, a concept well established in metagenomic studies [71], but much simpler in its application to ASVs. In

our wild-derived samples, the occurrence of infection with a particular *Eimeria* species in an individual host is explained to a large extent by the presence of the parasite species in the local host population. This is expected from local epidemiology, i.e. from the persistence of infections in the population [72, 73]. Controlling for this “locality effect”, we find the body condition of the mice to be negatively associated with the presence of *E. falciformis*, and *within infected mice*, negatively associated with *E. ferrisi* and *E. falciformis* abundance measured by amplicon (DNA) sequencing. Such correlation of body condition with *Eimeria* infection has been reported in other mammals [74, 75], but we were unable to detect it in our system with the quantification of oocysts or DNA in tissue [30]. It is plausible that *Eimeria* infection causes poor body condition considering the noticeable weight loss usually seen in laboratory infection [33]. The observed nonsignificant effect of *E. ferrisi* infection, *E. vermiformis* infection and intensity on body condition might be due to poor statistical power (small effect and low prevalence, particularly for *E. vermiformis*) or due to the difficulties with the accuracy (sensitivity and specificity) of detection (particularly for *E. ferrisi* in the qualitative analysis). It was, however, possible to detect a negative effect of *E. ferrisi* intensity on BMI, this effect was smaller than that of *E. falciformis* but still significant. The negative association between body condition and the presence of *E. falciformis*, despite its lower prevalence than *E. ferrisi*, suggests a stronger effect than the latter that also fits observations from laboratory infection [32]. This demonstrates the potential of amplicon sequencing to address epidemiological and ecological questions previously hard to address in wildlife systems. We recommend that such effects are best analysed in a quantitative way relating them to infection intensity derived from sequencing counts.

Amplicon sequencing generates sparse (meaning many samples without detection of a particular taxon) and proportional (meaning that taxa abundance estimates impact each other, as they are obtained as a fraction of the overall sequencing reads) [76–78]. This results in a negative correlation bias, as an increase of a taxon is accompanied by a decrease in the remaining taxa. Normalisation techniques accounting for this [22, 78] are increasingly important if the analysed taxa make up a large proportion of the overall sequencing reads. We see this in our results, as the choice of normalisation is more important for quantitative precision in the less deeply sequenced microfluidic amplification, in which *Eimeria* sequences make up a larger proportion compared to the more deeply sequenced classical amplification. More importantly, suitable selection of the PCR primers to allow both good on-target (here: *Eimeria*)

and wide taxonomically off-target reference for relative quantification, and sequencing at a suitable sequencing depth help to avoid such challenges. Many factors can affect the results of next-generation sequencing studies, including experimental design and analytical methods; therefore, meticulous consideration of the bioinformatics and statistical methodology is crucial. The optimal primer choice and sequencing depth are almost impossible to establish a priori, but we interpret our results as an encouragement to dare conducting such experiments in Coccidia and other parasites, as results will be robust across a wide range of parameters and can be corrected statistically for imperfect choices of molecular methods.

Conclusion

Amplicon sequencing provides a unified, scalable methodology to study parasites (or generally, eukaryotic symbionts), the bacterial microbiome, and hosts (e.g. if specific amplicons are used for simultaneous genotyping). “Universal” primer pairs, mostly targeting the 18S rRNA gene, are available or can be designed [17]. More work on amplicons sequencing is needed to evaluate how different primer pairs amplify and quantify Coccidia, other parasites and eukaryotic symbionts in faecal samples from mammals and other hosts. We have shown here that amplicon sequencing can differentially quantify *Eimeria* over a wide range of amplicon (primer pair) choices. This sequencing-based DNA quantification of Coccidia has good precision relative to qPCR-based quantification, and there is no reason to suspect this would be different for other parasites. We have previously shown that DNA-based quantification and classical coprological counts of transmission stages are complementary [15]. DNA intensity should be evaluated with respect to othe0.00011r biologically relevant measures in more host–parasite systems and amplicons sequencing is a suitable quantitative tool for this.

Abbreviations

ASVs	Amplified sequence variants
<i>COI</i>	Cytochrome c oxidase
CLR	Centred log-ratio
OTU	Operational taxonomic units
PERMANOVA	Permutational multivariate analysis of variance
PCR	Polymerase chain reaction
qPCR	Quantitative polymerase chain reaction
rRNA	Ribosomal RNA
TSS	Total sum scaling

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13071-023-05800-6>.

Additional file 1: Primer pairs used for multi-amplicon sequencing in the experimental (laboratory) and natural population datasets.

Additional file 2: Sequence alignments and respective phylogenetic trees.

Additional file 3: *Eimeria* taxonomic annotation to species level based on phylogenetic analysis. A phylogenetic tree was produced for all *Eimeria* ASVs from each amplicon (10 amplicons targeting the 18S rRNA gene), reference sequences for *Eimeria* spp of house mice and other rodents. Bootstrap_MRCA_clade: Bootstrap values from the clade of each species (reference species sharing the most recent common ancestor) retrieved from the respective amplicon phylogenetic tree.

Additional file 4: Table S1. Analysing the effects of normalisation techniques in the quantification of *Eimeria*. Comparison of two overlapping Pearson correlations based on dependent groups. Reference correlation is between *Eimeria* genome copies and *Eimeria* merged ASV abundance not normalised. rarefaction: random resampling of the OTU table to the minimum library depth. In green are the significantly improved and in red are significantly worsened correlations in relation to the overlapping correlation between *Eimeria* genome copies per gram of faeces and *Eimeria* ASV counts. P-adj: P-values corrected for multiple testing with Benjamini–Hochberg. **Figure S1.** Comparison of normalisation methods for quantification of *Eimeria* with amplicon sequencing (18S rRNA gene) using a standard PCR. dpi = days post-infection, df = degrees of freedom. **Figure S2.** Comparison of normalisation methods for quantification of *Eimeria* with amplicon sequencing (18S rRNA gene) using microfluidic PCR. dpi = days post-infection, df = degrees of freedom. **Figure S3.** The different amplified *Eimeria* ASVs consistently reflect infection dynamics. *Eimeria* ASV abundance throughout days post-infection measured as (a) ASV1 sequenced from standard PCR amplification; (b) ASV2 sequenced from standard PCR amplification; (c) ASV1 sequenced from microfluidic PCR amplification; (d) ASV1 sequenced from microfluidic PCR amplification. Abundance is measured as relative abundance after a total sum scaling normalisation, within the respective amplicon. Each dot represents one sample, samples from the same individual are connected with grey lines. The mean for each day post-infection is depicted as a diamond. Colours represent the different days post-infection. Note the difference in the scale of the y-axis in (a) and (c) vs (b) and (d).

Additional file 5: Table S2. Summary results of PERMANOVA based on the overall variation of *Eimeria* spp. composition, using the Jaccard similarity coefficient. Significant predictors are in bold. **Figure S4:** The distribution of *Eimeria* ASV abundance in mice sampled in the natural environment. (a) ASVs coloured by amplicon; (b) ASVs coloured by *Eimeria* species. Samples are ordered in both axes based on the first axis of a principal coordinate analysis with Bray–Curtis dissimilarities. Abundance is the relative abundance after total sum scaling within each amplicon, and summed across amplicons for each sample.

Acknowledgements

The authors acknowledge Ludovít Ďureje and Jaroslav Piálek and his team (Institute of Vertebrate Biology, AS CR, Brno, Department of Population Biology in Studenec) for providing the mice employed during the infection experiment and help with catching the house mice. We acknowledge the Open Access Publication Fund of Humboldt University Berlin.

Author contributions

SCMF and EH organised and planned the study. VHJD carried out the infection experiment and collected data required for the study. SCMF performed the analysis. SCMF and EH wrote the manuscript, with contributions and feedback from VHJD. All authors read and approved the final manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. Open Access Publication Fund of Humboldt-Universität zu Berlin. This work was supported by the German Foundation of Scientific Research (DFG) (Grant Number: 285969495/HE 7320/2-1; Grant Number: 440909536 [SCMF]) and the German Academic Exchange Service (DAAD) (VHJD scholarship holder during PhD studies) and the Research Training Group 2046 “Parasite Infections: From Experimental Models to Natural Systems” (VHJD associated PhD student/EH Senior Researcher). VHJD is funded by the Deutsche Forschungsgemeinschaft (DFG) grant Integrative evolutionäre und ökologische Analyse von Antibiotikaresistenzen: Auftreten und Verbreitung vom bakteriellen Genom bis zur geographischen Landschaft FO 1279/6-1|HE 7320/5-1|KR 4266/4-1.

Availability of data and materials

The datasets supporting the conclusions of this article are in https://github.com/ferreira-scm/Eimeria_AmpSeq.git. All sequencing raw data can be accessed through the BioProject PRJNA548431 in the NCBI Short Read Archive (SRA).

Declarations

Ethics approval and consent to participate

Experiment license 0431/17. On the basis of §8 of the Animal Protection Act Tierschutzgesetzes), the State Office for Health and Social Affairs–Berlin (Landesamt für Gesundheit und Soziales–Berlin [LAGeSo]) hereby grant permission to Prof. Dr. Emanuel Heitlinger to carry out scientific experiments on live vertebrates: a total of 1080 mice, *M. musculus*. Permit to capture house mice: capture permit No. 2347/35/2014.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Division of Computational Systems Biology, Center for Microbiology and Ecological Systems Science, University of Vienna, Djerassipl. 1, 1030 Vienna, Austria. ²Institute for Biology, Department of Molecular Parasitology, Humboldt-Universität zu Berlin (HU), Philippstr. 13, Haus 14, 10115 Berlin, Germany. ³Leibniz-Institut Für Zoo- Und Wildtierforschung (IZW) im Forschungsverbund Berlin E.V., Alfred-Kowalke-Straße 17, 10315 Berlin, Germany. ⁴Experimental and Clinical Research Center, a cooperation between the Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association and the Charité - Universitätsmedizin Berlin, Berlin, Germany. ⁵Experimental and Clinical Research Center, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität Zu Berlin, Lindenberger Weg 80, 13125 Berlin, Germany. ⁶Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany.

Received: 24 March 2023 Accepted: 3 May 2023

Published online: 17 June 2023

References

1. Knight R, Vrbanac A, Taylor BC, Aksenov A, Callewaert C, Debelius J, et al. Best practices for analysing microbiomes. *Nat Rev Microbiol*. 2018;16:410–22. <https://doi.org/10.1038/s41579-018-0029-9>.
2. Blaxter M. Counting angels with DNA. *Nature*. 2003;421:122–3. <https://doi.org/10.1038/421122a>.
3. Cordier T, Alonso-Sáez L, Apothéloz-Perret-Gentil L, Aylagas E, Bohan DA, Bouchez A, et al. Ecosystems monitoring powered by environmental genomics: a review of current strategies with an implementation road-map. *Mol Ecol*. 2021;30:2937–58. <https://doi.org/10.1111/mec.15472>.
4. Hauck R, Carrisosa M, McCrear BA, Dormitorio T, Macklin KS. Evaluation of next-generation amplicon sequencing to identify *Eimeria* spp. of Chickens. *Avian Dis*. 2019;63:577–83. <https://doi.org/10.1637/aviandiseases-D-19-00104>.
5. Hinsu AT, Thakkar JR, Koringa PG, Vrba V, Jakhesara SJ, Psifidi A, et al. Illumina Next Generation Sequencing for the Analysis of *Eimeria* Populations in Commercial Broilers and Indigenous Chickens. *Front Vet Sci*. 2018. <https://doi.org/10.3389/fvets.2018.00176>.
6. Zahedi A, Liu D, Yang R, Austen JM, Potter A, Ryan U. Next-generation sequencing amplicon analysis of the genetic diversity of *Eimeria* populations in livestock and wildlife samples from Australia. *Parasitol Res*. 2023;122:615–24. <https://doi.org/10.1007/s00436-022-07764-5>.
7. Avramenko RW, Redman EM, Lewis R, Yazwinski TA, Wasmuth JD, Gil-leard JS. Exploring the Gastrointestinal “Nemabiome”: Deep Amplicon Sequencing to Quantify the Species Composition of Parasitic Nematode Communities. *PLoS ONE*. 2015;10:e0143559. <https://doi.org/10.1371/journal.pone.0143559>.
8. Glidden CK, Koehler AV, Hall RS, Saeed MA, Coppo M, Beechler BR, et al. Elucidating cryptic dynamics of Theileria communities in African buffalo using a high-throughput sequencing informatics approach. *Ecol Evol*. 2020;10:70–80. <https://doi.org/10.1002/ece3.5758>.
9. Heitlinger E, Ferreira SCM, Thierer D, Hofer H, East ML. The intestinal eukaryotic and bacterial biome of spotted hyenas: the impact of social status and age on diversity and composition. *Front Cell Infect Microbiol*. 2017;7:262. <https://doi.org/10.3389/fcimb.2017.00262>.
10. He P, Chen Z, He Y, Chen J, Hayat K, Pan J, et al. A reliable and low-cost deep learning model integrating convolutional neural network and transformer structure for fine-grained classification of chicken *Eimeria* species. *Poult Sci*. 2023;102:102459. <https://doi.org/10.1016/j.psj.2022.102459>.
11. Adams DS, Ruiz-Jimenez F, Fletcher OJ, Gall S, Crespo R. Image analysis for *Eimeria* oocyst counts and classification. *J Appl Poult Res*. 2022;31:100260. <https://doi.org/10.1016/j.japr.2022.100260>.
12. Papaiakevou M, Gasser RB, Littlewood DTJ. Quantitative PCR-based diagnosis of soil-transmitted helminth infections: faecal or fickle? *Trends Parasitol*. 2019;35:491–500. <https://doi.org/10.1016/j.pt.2019.04.006>.
13. Zendejas-Heredia PA, Colella V, Hii SF, Traub RJ. Comparison of the egg recovery rates and limit of detection for soil-transmitted helminths using the Kato-Katz thick smear, faecal flotation and quantitative real-time PCR in human stool. *PLoS Negl Trop Dis*. 2021;15:e0009395. <https://doi.org/10.1371/journal.pntd.0009395>.
14. Velkers FC, Blake DP, Graat EAM, Vernooij JCM, Bouma A, de Jong MCM, et al. Quantification of *Eimeria acervulina* in faeces of broilers: comparison of McMaster oocyst counts from 24h faecal collections and single droppings to real-time PCR from cloacal swabs. *Vet Parasitol*. 2010;169:1–7. <https://doi.org/10.1016/j.vetpar.2010.01.001>.
15. Jarquín-Díaz VH, Balard A, Ferreira SCM, Mittné V, Murata JM, Heitlinger E. DNA-based quantification and counting of transmission stages provides different but complementary parasite load estimates: an example from rodent coccidia (*Eimeria*). *Parasit Vectors*. 2022;15:1–13. <https://doi.org/10.1186/s13071-021-05119-0>.
16. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res*. 2013. <https://doi.org/10.1093/nar/gks808>.
17. Hadziavdic K, Lekang K, Lanzen A, Jonassen I, Thompson EM, Troedsson C. Characterization of the 18S rRNA gene for designing universal eukaryote specific primers. *PLoS ONE*. 2014. <https://doi.org/10.1371/journal.pone.0087624>.
18. Brown SP, Ferrer A, Dalling JW, Heath KD. Don't put all your eggs in one basket: a cost-effective and powerful method to optimize primer choice for rRNA environmental community analyses using the Fluidigm Access Array. *Mol Ecol Resour*. 2016. <https://doi.org/10.1111/1755-0998.12507>.
19. Schriefer AE, Clifton PF, Hibberd MC, Sawyer C, Brown-Kennerly V, Burcea L, et al. A multi-amplicon 16S rRNA sequencing and analysis method for improved taxonomic profiling of bacterial communities. *J Microbiol Methods*. 2018;154:6–13. <https://doi.org/10.1016/j.mimet.2018.09.019>.
20. Morton JT, Marotz C, Washburne A, Silverman J, Zaramela LS, Edlund A, et al. Establishing microbial composition measurement standards with

- reference frames. *Nat Commun.* 2019;10:2719. <https://doi.org/10.1038/s41467-019-10656-5>.
21. Roche KE, Mukherjee S. The accuracy of absolute differential abundance analysis from relative count data. *PLoS Comput Biol.* 2022;18:1–25. <https://doi.org/10.1371/journal.pcbi.1010284>.
 22. Weiss S, Xu ZZ, Peddada S, Amir A, Bittinger K, Gonzalez A, et al. Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome.* 2017. <https://doi.org/10.1186/s40168-017-0237-y>.
 23. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods.* 2016. <https://doi.org/10.1038/nmeth.3869>.
 24. Edgar RC. Accuracy of taxonomy prediction for 16S rRNA and fungal ITS sequences. *PeerJ.* 2018;6:e4652. <https://doi.org/10.7717/peerj.4652>.
 25. Edgar R. Taxonomy annotation and guide tree errors in 16S rRNA databases. *PeerJ.* 2018;6:e5030. <https://doi.org/10.7717/peerj.5030>.
 26. Jarquín-Díaz VH, Balard A, Mácová A, Jost J, Roth von Szepesbela T, Bertold K, et al. Generalist *Eimeria* species in rodents: multilocus analyses indicate inadequate resolution of established markers. *Ecol Evol.* 2020;10:1378–89. <https://doi.org/10.1002/ece3.5992>.
 27. Jarquín-Díaz VH, Balard A, Jost J, Kraft J, Dikmen MN, Kvičeroš J, et al. Detection and quantification of house mouse *Eimeria* at the species level—challenges and solutions for the assessment of coccidia in wildlife. *Int J Parasitol Parasites Wildl.* 2019;10:29–40. <https://doi.org/10.1016/j.ijppaw.2019.07.004>.
 28. Chapman HD, Barta JR, Blake D, Gruber A, Jenkins M, Smith NC, et al. A selective review of advances in coccidiosis research. *Adv Parasitol.* 2013;83:93–171. <https://doi.org/10.1016/B978-0-12-407705-8.00002-1>.
 29. Ehret T, Spork S, Dieterich C, Lucius R, Heitlinger E. Dual RNA-seq reveals no plastic transcriptional response of the coccidian parasite *Eimeria falciformis* to host immune defenses. *BMC Genomics.* 2017. <https://doi.org/10.1186/s12864-017-4095-6>.
 30. Balard A, Jarquín-Díaz VH, Jost J, Martincová I, Ďureje L, Piálek J, et al. Intensity of infection with intracellular *Eimeria* spp. and pinworms is reduced in hybrid mice compared to parental subspecies. *J Evol Biol.* 2020;33:435–48. <https://doi.org/10.1111/jeb.13578>.
 31. Schito ML, Barta JR, Chobotar B. Comparison of four murine *Eimeria* species in immunocompetent and immunodeficient mice. *J Parasitol.* 1996;82:255–62.
 32. Al-khlifeh E, Balard A, Jarquín-Díaz VH, Weyrich A, Wibbelt G, Heitlinger E. *Eimeria falciformis* BayerHaberKorn1970 and novel wild derived isolates from house mice: differences in parasite lifecycle, pathogenicity and host immune reactions. *BioRxiv.* 2019. <https://doi.org/10.1101/611277>.
 33. Balard A, Jarquín-Díaz VH, Jost J, Mittné V, Böhning F, Ďureje L, et al. Coupling between tolerance and resistance for two related *Eimeria* parasite species. *Ecol Evol.* 2020;10:13938–48. <https://doi.org/10.1002/ece3.6986>.
 34. Piálek J, Vyskočilová M, Bímová B, Havelková D, Piálková J, Duřková P, et al. Development of unique house mouse resources suitable for evolutionary studies of speciation. *J Hered.* 2008;99:34–44. <https://doi.org/10.1093/jhered/esm083>.
 35. Heitlinger E. MultiAmpliconv0.1. 2019. Retrieved from: <https://derede.github.io/MultiAmplicon/index.html>.
 36. Davis NM, DIM P, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome.* 2018;6:1–14. <https://doi.org/10.1186/s40168-018-0605-2>.
 37. Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, et al. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat Methods.* 2013;10:57–9. <https://doi.org/10.1038/nmeth.2276>.
 38. McMurdie PJ, Holmes S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE.* 2013;8:e61217. <https://doi.org/10.1371/journal.pone.0061217>.
 39. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol.* 2007;73:5261–7. <https://doi.org/10.1128/AEM.00062-07>.
 40. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 2012;41:D590–6. <https://doi.org/10.1093/nar/gks1219>.
 41. Nilsson RH, Larsson K-H, Taylor AFS, Bengtsson-Palme J, Jeppesen TS, Schigel D, et al. The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Res.* 2019;47:D259–64. <https://doi.org/10.1093/nar/gky1022>.
 42. Robeson MS, O'Rourke DR, Kaehler BD, Ziemski M, Dillon MR, Foster JT, et al. RESCRIPt: Reproducible sequence taxonomy reference database management. *PLoS Comput Biol.* 2021. <https://doi.org/10.1371/journal.pcbi.1009581>.
 43. Wright ES. Using DECIPHER v2.0 to analyze big biological Sequence Data in R. *Version 2.* 24. 0. R J. 2016;8:352–9.
 44. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jeremiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods.* 2017;14:587–9. <https://doi.org/10.1038/nmeth.4285>.
 45. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol.* 2020;37:1530–4. <https://doi.org/10.1093/molbev/msaa015>.
 46. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol.* 2018;35:518–22. <https://doi.org/10.1093/molbev/msx281>.
 47. Letunic I, Bork P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics.* 2007;23:127–8. <https://doi.org/10.1093/bioinformatics/btl529>.
 48. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R, version 5. 6. 2. *Bioinformatics.* 2019;35:526–8.
 49. Xu S, Li L, Luo X, Chen M, Tang W, Zhan L, et al. Ggtree: A serialized data object for visualization of a phylogenetic tree and annotation data. *Version 3. 4. 4.* IMeta. 2022. <https://doi.org/10.1002/mt2.56>.
 50. Heitlinger E, Spork S, Lucius R, Dieterich C. The genome of *Eimeria falciformis*—reduction and specialization in a single host apicomplexan parasite. *BMC Genomics.* 2014;15:1–17. <https://doi.org/10.1186/1471-2164-15-696>.
 51. Guo K, Gao P. microbial: Do 16S Data Analysis and Generate Figures. R Package Version 0.0.20 2021. <https://CRAN.R-project.org/package=microbial>.
 52. Lahti L, Shetty S, et al. Tools for microbiome analysis in R. *Version.* 2017;1:28.
 53. Hittner JB, May K, Silver NC. A Monte Carlo evaluation of tests for comparing dependent correlations. *J Gen Psychol.* 2003;130:149–68. <https://doi.org/10.1080/00221300309601282>.
 54. Diedenhofen B, Musch J. cocor: A comprehensive solution for the statistical comparison of correlations. *PLoS ONE.* 2015;10:e0121945. <https://doi.org/10.1371/journal.pone.0121945>.
 55. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol.* 1995;57:289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
 56. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Softw.* 2015. <https://doi.org/10.18637/jss.v067.i01>.
 57. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest package: tests in linear mixed effects models. *J Stat Softw.* 2017. <https://doi.org/10.18637/jss.v082.i13>.
 58. Nakagawa S, Schielzeth H. A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods Ecol Evol.* 2013;4:133–42. <https://doi.org/10.1111/j.2041-210x.2012.00261.x>.
 59. Sterne TE. Some remarks on confidence or fiducial limits. *Biometrika.* 1954;41:275–8. <https://doi.org/10.2307/2333026>.
 60. Stevenson M, Sergeant E. epiR: Tools for the analysis of epidemiological data. R package version 2.0.60 2018. <https://CRAN.R-project.org/package=epiR>.
 61. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlenn D, et al. vegan: Community Ecology Package. R Package Version. 2019;2:5–6.
 62. Ballenghien M, Faivre N, Galtier N. Patterns of cross-contamination in a multispecies population genomic project: detection, quantification, impact, and solutions. *BMC Biol.* 2017;15:25. <https://doi.org/10.1186/s12915-017-0366-6>.
 63. Bortoluzzi C, Paras KL, Applegate TJ, Verocai GG. Comparison between McMaster and Mini-FLOTAC methods for the enumeration of *Eimeria maxima* oocysts in poultry excreta. *Vet Parasitol.* 2018;254:21–5. <https://doi.org/10.1016/j.vetpar.2018.02.039>.

64. Daş G, Klausner S, Stehr M, Tuchscherer A, Metges CC. Accuracy and precision of McMaster and Mini-FLOTAC egg counting techniques using egg-spiked faeces of chickens and two different flotation fluids. *Vet Parasitol.* 2020;283:109158. <https://doi.org/10.1016/j.vetpar.2020.109158>.
65. Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ. At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. *Appl Environ Microbiol.* 2005;71:7724–36. <https://doi.org/10.1128/AEM.71.12.7724-7736.2005>.
66. Sun D-L, Jiang X, Wu QL, Zhou N-Y. Intragenomic heterogeneity of 16S rRNA genes causes overestimation of prokaryotic diversity. *Appl Environ Microbiol.* 2013;79:5962–9. <https://doi.org/10.1128/AEM.01282-13>.
67. Gjerde B, Hilali M, Mawgood SA. Molecular characterisation of three regions of the nuclear ribosomal DNA unit and the mitochondrial *cox1* gene of *Sarcocystis fusiformis* from water buffaloes (*Bubalus bubalis*) in Egypt. *Parasitol Res.* 2015;114:3401–13. <https://doi.org/10.1007/s00436-015-4566-0>.
68. Sudan V, Shanker D, Paliwal S, Kumar R, Singh A. Phylogenetics of *Sarcocystis fusiformis* isolates based on 18S rRNA and *cox 1* genes. *Microb Pathog.* 2021;159:105144. <https://doi.org/10.1016/j.micpath.2021.105144>.
69. McDonald V, Shirley MW. The endogenous development of virulent strains and attenuated precocious lines of *Eimeria tenella* and *E. necatrix*. *J Parasitol.* 1987;73:993–7.
70. McDonald V, Ballingall S. Further investigation of the pathogenicity, immunogenicity and stability of precocious *Eimeria acervulina*. *Parasitology.* 1983;86:361–9.
71. Wu Y-W, Ye Y. A Novel Abundance-Based Algorithm for Binning Metagenomic Sequences Using l-tuples. *J Comput Biol.* 2011;18:523–34. <https://doi.org/10.1089/cmb.2010.0245>.
72. Svensson C, Hooshmand-Rad P, Pehrson B, Tömquist M, Ugglå A. Excretion of *Eimeria* Oocysts in Calves during their first three weeks after turnout to pasture. *Acta Vet Scand.* 1993;34:175–82. <https://doi.org/10.1186/BF03548207>.
73. von Samson-Himmelstjerna G, Epe C, Wirtherle N, von der Heyden V, Welz C, Radeloff J, et al. Clinical and epidemiological characteristics of *Eimeria* infections in first-year grazing cattle. *Vet Parasitol.* 2006;136:215–21. <https://doi.org/10.1016/j.vetpar.2005.11.022>.
74. Hakkarainen H, Huhta E, Koskela E, Mappes T, Soveri T, Suorsa P. *Eimeria* parasites are associated with a lowered mother's and offspring's body condition in island and mainland populations of the bank vole. *Parasitology.* 2007;134:23–31. <https://doi.org/10.1017/S0031182006001120>.
75. Ptáček M, Kyriánová IA, Nápravníková J, Ducháček J, Husák T, Chay-Canul AJ, et al. Do live weight, body condition score, back muscle or back-fat reserves create the suspicion of goats infected with *Eimeria* or Trichostromylids? *Anim Open Access J MDPI.* 2021;11:3591. <https://doi.org/10.3390/ani11123591>.
76. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome datasets are compositional: and this is not optional. *Front Microbiol.* 2017;8:1–6. <https://doi.org/10.3389/fmicb.2017.02224>.
77. Kurtz ZD, Müller CL, Miraldi ER, Littman DR, Blaser MJ, Bonneau RA. Sparse and compositionally robust inference of microbial ecological networks. *PLoS Comput Biol.* 2015;11:1–25. <https://doi.org/10.1371/journal.pcbi.1004226>.
78. Lin H, Peddada SD. Analysis of microbial compositions: a review of normalization and differential abundance analysis. *NPJ Biofilms Microbiomes.* 2020. <https://doi.org/10.1038/s41522-020-00160-w>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

