International Journal for Parasitology: Parasites and Wildlife 12 (2020) 220–231

Contents lists available at ScienceDirect

International Journal for Parasitology: Parasites and Wildlife

journal homepage: www.elsevier.com/locate/ijppaw

Description of *Sarcocystis scandentiborneensis* sp. nov. from treeshrews (*Tupaia minor*, *T. tana*) in northern Borneo with annotations on the utility of COI and 18S rDNA sequences for species delineation

Paula Ortega Pérez [a], Gudrun Wibbelt [a], Annika Brinkmann [b], John A. Galindo Puentes [a], Fred Y. Y. Tuh [c], Maklarin B. Lakim [c], Andreas Nitsche [b], Konstans Wells [d], Thomas Jäkel [e,*]

[a] Department Wildlife Diseases, Leibniz Institute for Zoo and Wildlife Research, Alfred-Kowalke-Straße 17, 10315, Berlin, Germany
[b] Centre for Biological Threats and Special Pathogens, Robert Koch Institute, Nordufer 20, 13353, Berlin, Germany
[c] Sabah Parks, 88100, Kota Kinabalu, Sabah, Malaysia
[d] Department of Biosciences, Swansea University, Swansea, SA2 8PP, United Kingdom
[e] Department of Zoology, Division of Parasitology, University of Hohenheim, Emil-Wolff-Straße 34, 70599, Stuttgart, Germany

ARTICLE INFO

Keywords:
Sarcocystis
Tupaia
Phylogeny
*cox*1
COI
18S rRNA
Morphology

ABSTRACT

Sarcocystis scandentiborneensis sp. nov. was discovered in histological sections of striated musculature of treeshrews (*Tupaia minor*, *T. tana*) from Northern Borneo. Sarcocysts were cigar-shaped, 102 μm–545 μm long, and on average 53 μm in diameter. The striated cyst wall varied in thickness (2–10 μm), depending on whether the finger-like, villous protrusions (VP) were bent. Ultrastructurally, sarcocysts were similar to wall type 12 but basal microtubules extended into VPs that tapered off with a unique U-shaped, electron-dense apical structure. In phylogenetic trees of the nuclear 18S rRNA gene, *S. scandentiborneensis* formed a distinct branch within a monophyletic subclade of *Sarcocystis* spp. with (colubrid) snake-rodent life cycle. We mapped all intraspecific (two haplotypes) and interspecific nucleotide substitutions to the secondary structure of the 18S rRNA gene: in both cases, the highest variability occurred within helices V2 and V4 but intraspecific variability mostly related to transitions, while transition/transversion ratios between *S. scandentiborneensis*, *S. zuoi*, and *S. clethrionomyelaphis* were skewed towards transversions. Lack of relevant sequences restricted phylogenetic analysis of the mitochondrial Cytochrome C oxidase subunit I (COI) gene to include only one species of *Sarcocystis* recovered from a snake host (*S. pantherophisi*) with which the new species formed a sister relationship. We confirm the presence of the functionally important elements of the COI barcode amino acid sequence of *S. scandentiborneensis*, whereby the frequency of functionally important amino acids (Alanine, Serine) was markedly different to other taxa of the Sarcocystidae. We regard *S. scandentiborneensis* a new species, highlighting that structurally or functionally important aspects of the 18S rRNA and COI could expand their utility for delineation of species. We also address the question why treeshrews, believed to be close to primates, carry a parasite that is genetically close to a *Sarcocystis* lineage preferably developing in the Rodentia as intermediate hosts.

1. Introduction

The genus *Sarcocystis* comprises a group of more than 200 species of protist parasites that infect mammals, reptiles, birds, and fish. The genus *Sarcocystis* belongs to the phylum Apicomplexa and comprises species which commonly exhibit an obligatory two-host (predator-prey) life cycle. Herbivore or omnivore prey species as intermediate hosts become infected after ingesting water or food contaminated with oocysts/sporocysts, the typical stage of the coccidia that is the product of sexual development in the definitive host and is released into the environment (Odening, 1998; Dubey et al., 2016). In the intermediate host, several rounds of asexual replication in internal organs are followed by invasion of the striated musculature by motile stages (merozoites) that initiate the development of stationary cysts, so-called sarcocysts (containing cystozoites), that are characteristic of the genus. These cysts may remain viable in the musculature for the lifetime of the intermediate host until a predatory definitive host would become infected by feeding on the intermediate host, re-starting the life cycle of *Sarcocystis* in its intestine

* Corresponding author. Rice Department, 50 Phaholyothin Road, Ladyao, Chatuchak, 10900, Bangkok, Thailand.
E-mail address: thom.jaekel@t-online.de (T. Jäkel).

https://doi.org/10.1016/j.ijppaw.2020.07.003
Received 30 May 2020; Received in revised form 5 July 2020; Accepted 5 July 2020
Available online 8 July 2020

(Dubey et al., 2016).

The present study was part of an investigation into Bornean rodents (Rodentia) and non-rodent taxa (here Scandentia: treeshrews) that set out to learn more about their parasite fauna, in particular cyst-forming coccidia. Although some information exists on *Sarcocystis* spp. of commensal rodents in Southeast Asia (Jäkel et al., 1997; Paperna et al., 2004; Fong et al., 2014), little is known about treeshrews, which occupy forested habitats of that region (Emmons, 2000). The island of Borneo is particularly interesting since it is known as a global hotspot of mammalian biodiversity (Struebig et al., 2015). To date, only a single species, namely *Sarcocystis tupaia*, has been reported from treeshrews (Xiang et al., 2010). In species-rich tropical lowland forests, several treeshrews are generally sympatric, i.e. occur in both old-growth and secondary/logged forests (Wells et al., 2007). *Tupaia minor* is an arboreal species that occasionally moves on the ground (Wells et al., 2006). In contrast, *T. tana* is a terrestrial treeshrew species with some scansorial activity covering the first few meters above ground and nests located in tree hollows and plant fibers slightly above ground (Emmons, 2000; Wells et al., 2006). Treeshrews occupy a relatively low position in the food-web and frequently fall prey to a variety of tree-dwelling predators, including mammalian carnivores, like the yellow-throated marten (*Martes flavigula*) and small cats (i.e. *Pardofelis marmorata* and *Prionailurus bengalensis*), snakes, and birds like raptors or owls (Emmons, 2000). This scenario appears likely to create ideal conditions for the presence of *Sarcocystis* spp.

Here, we describe a new *Sarcocystis* sp. in the striated musculature of these hosts based on the results of our histological, ultrastructural, and molecular phylogenetic examinations. We compare this species with other known *Sarcocystis* spp. from other hosts, also including data on the structure and sequence variability of the selected genes (nuclear 18S rRNA and mitochondrial Cytochrome C oxidase subunit I [COI]) and comment on their utility for species delineation in the context of the clade of *Sarcocystis* spp. examined here.

2. Materials and methods

2.1. Ethical statement

Biological resource access and export permits were issued by the Sabah Biodiversity Centre (JKM-MBS.1000–2/2 [35], JKM-MBS.1000–2/2 [63]); access to forest field sites were approved by Sabah Parks and individual land owners. Trapping and examination of caught animals took place as part of a larger study to investigate habitat use and distribution of small mammal species across land-use gradients in Northern Borneo, Sabah, Malaysia (Wells et al., 2014). Captured animals were transferred to nearby mobile field laboratories inside their traps covered by a cotton cloth, moved from the traps into small cotton bags and further into a plastic container for subsequent anaesthesia via diethyl ether inhalation (anaesthetic grade) and killed by cervical dislocation (according to guidelines by the American Veterinary Medical Association, https://www.avma.org).

2.2. Study area, trapping of animals, and tissue sampling

Between 2012 and 2013, treeshrews (*Tupaia* spp.) and other small mammal species (rodents) were live-trapped using locally made drop-door, wire-mesh traps (ca. 280 × 140 × 140 mm) baited with mixtures of raw banana, fried banana and dried fish. Traps were set in different habitat types from mature forest to urban habitats, and frequently checked for captures. For each capture location, geographical coordinates were recorded using a GPS device (Garmin GPSmap62st, Olathe, USA). Identification of treeshrews was based on phenotypic characteristics according to Payne et al. (1998). All animal carcasses were dissected and samples of the striated musculature from diaphragm and latero-ventral abdominal wall (Musculus obliquus externus) or M. quadriceps femoris and occasionally laryngeal muscle were fixed and stored in 70% ethanol. Logistically, it was not possible to examine fresh tissues and directly fix them in glutaraldehyde for ultrastructural processing.

2.3. Light microscopy (LM) and transmission electron microscopy (TEM)

Fixed samples of musculature of approximately 1.5 cm^3 were trimmed, embedded in paraffin, sectioned at 3 µm, and stained with haematoxylin and eosin (H&E). Striated muscle tissue was generally sectioned in slices of 0.24–0.56 cm^2 and screened for sarcocysts by LM. *Sarcocystis*-positive samples were further processed for TEM. For this, 2–4 mm^3 of muscle tissue were removed from the paraffin blocks and incubated overnight in xylene (Medite GmbH, Burgsdorf, Germany) for deparaffinisation. They were subsequently washed in decreasing ethanol concentrations and again incubated overnight in phosphate-buffered saline until all ethanol was removed. Then the samples were post-fixed in 1% osmium tetroxide (Plano GmbH, Wetzlar, Germany) at 4 °C for 1h 30 min, dehydrated in increasing ethanol concentrations, and embedded in epoxy resin (Carl Roth GmbH & Co. KG., Karlsruhe, Germany). Semi-thin sections (1 µm) were stained with Richardson's stain (equal volumes of 1% methylene blue in 1% borax and 1% Azure II; Richardson et al., 1960) and examined by light microscopy. Ultrathin sections (70 nm) were double contrasted with 35 mg/ml uranyl acetate (Merck KGaA, Darmstadt, Germany) for 20 s followed by 34.42 mg/ml lead citrate (SERVA Electrophoresis GmbH, Heidelberg, Germany) for 10 min according to Reynolds (1963). Ultrathin sections were examined via electron microscopy (Tecnai BioSpirit, FEI, Netherlands).

2.4. DNA isolation, amplification, and sequencing

For DNA extraction, samples of 25 µg of striated musculature were used that had been checked for presence of sarcocysts of the novel species by LM and TEM. Samples were washed and rehydrated in distilled water for ethanol removal. Then, DNA was extracted using the commercial kit NuleoSpin® DNA RapidLyse according to instructions of the manufacturer (Macherey-Nagel, Düren, Germany).

Nuclear 18S rDNA sequences were amplified following the protocol of Prakas et al. (2014) using the primers SarAF (5′-CTGGTTGATCCTGCCAGTAG-3′) and SarDR (5′-GCAGGTTCACCTACGGAAA-3′). The amplicons were cloned using the TOPO TA Cloning® Kit for Sequencing with One Shot® TOP10 Chemically Competent *E. coli* (Invitrogen, Glasgow, UK) according to the manufacturer's instructions. Ten clones per PCR product were selected and transferred to 20 µl water. The clones were used as template in a PCR as published by Andree et al. (2010). The PCR products were purified (ExoSAP; Thermo Fisher Scientific, Waltham, USA) and Sanger sequenced using the BigDye Terminator v1.1 Cycle Sequencing Kit for Sanger sequencing (Thermo Fisher Scientific, Waltham, USA) and the primers SarAF, SarCF (5′TTAACTGTCAGAGGTGAAATTCTT-3′), SarBR (5′-GGCAAATGCTTTCGCAGTAG-3′) and SarDR (Prakas et al., 2014). Sequences were analysed using a 3130xl Genetic Analyzer (Thermo Fisher Scientific, Waltham, USA). Raw sequences were manually checked and refined using MEGA-6 (Tamura et al., 2013).

For amplification of the mitochondrial Cytochrome Coxidase subunit 1 gene (COI or *cox1*) we used the primers SF1 (5′-ATGGCGTACAACAATCATAAAGAA-3′) and SRD8 (5′-CATTGCCCATDACTACGCC-3′) published by Gjerde (2013b) with minor modifications of his protocol: the MyTaq HS master mix (Bioline, Luckenwalde, Germany) and an annealing temperature of 58 °C. PCR products were purified using Agencourt AMPure beads (Beckman Coulter, Krefeld, Germany), adapter- and barcode-ligated with the LSK109 Kit (Oxford Nanopore Technologies, Oxford, UK), and subjected to direct sequencing on a R9.4.1 flow cell (Oxford Nanopore Technologies, Oxford, UK).

2.5. Sequence alignment, and molecular and phylogenetic analysis of the 18SrRNA and COI genes

The 18S rDNA sequences of the *Sarcocystis* isolates under investigation were aligned with published Apicomplexan species available at GenBank (Clark et al., 2016) using the multiple sequence alignment algorithm 'R-Coffee' of the T-Coffee web server (version 11.00. d625267/2016-01-11/Revision-d625267-Build 507), which takes into account predicted secondary structures of RNA (Moretti et al., 2008; Tommaso et al., 2011). The alignment procedure was run several times, selecting the alignment with the smallest proportion of ambiguous positions (Tommaso et al., 2011). Pairwise analyses of transition/-transversion ratios (Ti/Tv) included the sequences of *Sarcocystis clethrionomyelaphis* (KP057504, KF309701), *S. zuoi* (JQ029112, JQ029113, KU341120), and the new isolates of the *Sarcocystis* sp. from *Tupaia minor* (GenBank accession numbers: MN733816 [E357-13] and MN733817 [E364-13]). Additional comparisons were run with published sequences of other *Sarcocystis* spp. and *Plasmodium vivax*, for which allelic and paralogous 18S rDNA sequences have been identified (Slapeta et al., 2002; Rooney, 2004). All comparisons were executed using the statistics function of the MEGA-7 software package (Kumar et al., 2016). Tajima's relative rate test (Tajima, 1993) was carried out with selected sequences to test the hypothesis that a pair of sequences has undergone different rates of evolution. Based on the R-Coffee alignment, we visually mapped to the *Toxoplasma gondii* nuclear 18S rRNA reference molecule (M97703) with its predicted secondary structure (Gagnon et al., 1996) all bp changes that were observed between the haplotypes of the new *Sarcocystis* sp. (intraspecific variation) and those that occurred between the new sequences and *S. zuoi* (KU341120) and *S. clethrionomyelaphis* (KP057504) to capture interspecific variation. The original sequences of *S. zuoi* (JQ029112, JQ029113) were not included in the latter comparison, because their posterior parts were truncated.

For phylogenetic analysis we included all relevant snake host *Sarcocystis* spp. because initial screening by BLAST (Johnson et al., 2008) had shown that the sequences under investigation were highly similar to known 18S rRNAs of this group of taxa. The 18S rDNA sequences of *S. zuoi* (KU341120; 1700 nt) and *S. zamani* (KU2445241; 1789 nt) from Thailand were excised out of a combined set of sequences after alignment with the homologous sequences of *T. gondii* (M97703) and *S. singaporensis* (KY513624). Among the four available sequences of *S. zuoi* from Thailand (KU341118-21; Watthanakaiwan et al., 2017) we restricted our analysis to KU341120, because it ranked first in the BLAST search and represented the longest nt sequence from that location. Other taxa included *S. tupaia* (FJ827486; because published sequences FJ827485-88 are identical, only one was included here; Xiang et al., 2010), the only known 18S rDNA sequence from a treeshrew to date, members of the '*S. neurona* subclade' (e.g., *S. lacertae*, *S. muris*, *S. neurona*, etc.), and *Sarcocystis* spp. with a carnivore-ruminant life cycle. Selected taxa of the Toxoplasmatinae, the fish parasite *Gouissa janae* (AY043206), and species of Eimeriidae served as outgroups. For Maximum Likelihood analysis (ML), the appropriate model for nucleotide (nt) substitution rates was determined employing MEGA-7, using the model with the lowest BIC (Bayesian Information Criterion) and AICc (Akaike Information Criterion, corrected) values. We selected the Hasegawa-Kishino-Yano model with a discrete gamma distribution rate variation among sites and accounting for a proportion of invariant sites (HKY + G + I) over five Gamma categories; site coverage cut-off was 88%–90% (1491–1377 sites of 42 taxa under analysis, respectively). The number of bootstrap replications was set to 1000. With this set up, ML computations were repeated three times. Bayesian Inference (BI) was executed from within MrBayes Version 3.2 (Ronquist et al., 2012), whereby we selected the general time-reversible substitution model with discrete gamma distribution and invariable sites (GTR + G + I; nst = 6 gamma categories) in addition to an assumed among-site rate variation between different lineages (covarion = yes; the so-called 'covariotide' model, which was found to fit best for the coccidia; Morrison et al., 2004). Bayesian Inference was computed in two replicate runs: all sequences were subjected to 1,000,000 Markow Chain Monte Carlo generations (heated chains = 4, chain temperature = 0.2, unconstrained branch length with exponential = 10) with a sampling frequency of 1000; burn-in length was set at 10%. Completion of an analysis was regarded successful, if the average standard deviation of split frequencies remained below $P = 0.01$ and the convergence diagnostic (potential scale reduction factor, PSFR) indicated a good sample of the posterior probability distribution (all values close to 1.0). Resulting trees were visualized using the program FigTree (Version 1.4.2, by Andrew Rambaut, Institute of Evolutionary Biology, University of Edinburgh, UK; http://tree.bio.ed.ac.uk/).

Analyses of nt sequences of the Cytochrome C oxidase subunit I gene (COI or *cox1*) were all executed within MEGA-7 software, including alignment of nt or amino acid (aa) sequences with MUSCLE, translation of nt into the corresponding aa sequences using the protozoan mitochondrial (mt) genetic code, application of sequence comparative statistics, and phylogenetic analyses using ML. We used 21 taxa for phylogeny reconstruction, including published COI sequences of the Toxoplasmatinae and Sarcocystinae, with *S. pantherophisi* (KU891603) as sole representative of a snake-rodent life cycle. Taxa of the Eimeriidae, Piroplasmida and Haemosporida served as out-groups. The Tamura-Nei model with gamma-distributed rates among sites and a proportion of invariant sites (TN93 + G + I) was selected as the best substitution model (lowest BIC), with three discrete gamma categories. Phylogenetic analyses were executed with and without inclusion of the 3rd codon position to assess its effect on tree topology (site coverage cut-off: 95%). Bootstrap consensus trees were constructed from 1000 replicates.

To check if the amplified COI sequences contained the barcode area of 219 aa in the enzymatically active part of the protein (six alpha-helices connected by five loops; Pentinsaari et al., 2016), we first assembled a consensus nt sequence from our identical but length-variable fragments (COI sequences E357-13 and E120-13 together comprise the full consensus; GenBank accession numbers MN732561 and MN732562, respectively). Then, the consensus sequence was translated into the corresponding protein sequence and compared with sequences of similar length of other taxa of the Sarcocystidae.

3. Results

3.1. Host

Eight adult *Tupaia minor* (four male, four female) and six *T. tana* (two male, four female) were trapped in Poring Hot Springs (Sabah, Malaysia; geo-reference: 06.04120 N 116.70328 E). Both species were captured at the same site of a primary forest edge next to the park entrance and adjacent to village housing. Seven out of the eight *T. minor* and three out of six *T. tana* showed infections with *Sarcocystis* spp. in histological sections of the striated musculature of diaphragm, abdominal or thigh musculature. Of the seven positive *T. minor*, two showed monoinfections with the new *Sarcocystis* sp. under investigation (sample IDs: E357-13, E364-13), while the other five lesser treeshrews were predominantly infected with the same species, but histologically also exhibited co-infection with other *Sarcocystis* spp. that were not further examined in the present study. In approximately 0.5 cm^2-sized sections of striated musculature from *T. minor* the following numbers of sarcocysts of the new species were counted in the examined host individuals (as depicted by their internal IDs): E351-13 (13 in diaphragm), E352-13 (6–28 in diaphragm), E357-13 (2 in diaphragm), E364-13 (2–3 in diaphragm; 7 to 11 in thigh muscle), E370-13 (9–12 in laryngeal muscle close to salivary gland; 2 to 6 in abdominal muscle), E371-13 (9 in abdominal muscle), E392-13 (3 in diaphragm). Among the three *Sarcocystis*-infected *Tupaia tana*, E70-14 (7 cysts in laryngeal muscle close to salivary

gland), E66-14 (6 in thigh) and E120-13 (6 in diaphragm), only sample E120-13 showed monoinfection with the novel *Sarcocysts* species and was used for further analysis. None of the murid rodents (33 *Sundamys muelleri*, three *Niviventer cremoriventer*, two *Maxomys whiteheadi*) trapped at the same time and area were infected with the novel *Sarcocystis* sp. observed in treeshrews.

3.2. Sarcocyst

In histological sections of diaphragm, abdominal or thigh musculature sarcocysts were located inside skeletal myofibers without any apparent inflammatory response (Fig. 1A). In the cross-sectional plane, the microscopic sarcocysts were spherical to slightly oval, measuring 22–142 µm in diameter (mean = 53.2 ± 28.9 µm [± S.D.]; n = 32, five treeshrews). Without the possibility to examine these sarcocysts in their native state we are cautious in describing their shape, but longitudinal sections (Fig. 1B) showed cigar-shaped cysts with slightly pointed ends that were between 102 µm and 545 µm long (average = 340 µm; n = 4 from two treeshrews). Most of the detected sarcocysts were mature (rarely containing metrocytes), occasionally possessed folds or invaginations, and were septated with chambers containing loosely packed cystozoites (Fig. 1C, **D**). Under the light microscope, the cyst wall (here: the primary cyst wall including the layer of ground substance underneath) appeared striated due to tightly packed, villous protrusions (VP) that could assume a brush-like appearance, in particular when they were bent down at the interface between cyst wall and host tissue (Fig. 1C). Accordingly, the thickness of the cyst wall varied, depending on whether VP were fully stretched (3.9–10.7 µm) or bent down (1.8–3.6 µm). The average thickness based on samples in H&E sections from three different hosts was 4.0 µm (±1.9; n = 87 measurements taken from 15 cysts). A conspicuous element was the very thin layer of ground substance (GS) of the cyst wall, which was about 0.3 µm thick in H&E sections (Fig. 1C and D) and 150–210 nm in ultrathin sections (Fig. 2A). At the ultrastructural level, the finger-like VP were anchored in the GS by microtubules that extended into each VP (Fig. 2B). Fully stretched VP were between 4.23 and 6.71 µm long (average 5.02 µm; n = 6 cysts) and 480–640 nm wide (300–400 nm near the tip). The VP possessed a unique electron-dense, U-shaped apical structure, which was about 250 nm wide (plane parallel to layer of GS) and 150 nm long, with a branch diameter of 74–99 nm (Fig. 2B). Additionally, the surface of the protrusions appeared to bear thorn-like structures (Fig. 2B). The banana-shaped cystozoites were relatively small, measuring 5.3 (±0.7) x 1.3 (±0.2) µm (n = 85) in H&E sections; they measured 4250 (±650) x 1410 (±120; n = 17) nm in ultrastructural preparations.

3.3. Molecular and phylogenetic analyses of the 18S rRNA and COI genes

For our molecular investigations we used muscle samples from three *Tupaia minor* and one specimen of *T. tana*. Samples of diaphragm and thigh musculature of two *T. minor and one T. tana* showed monoinfection

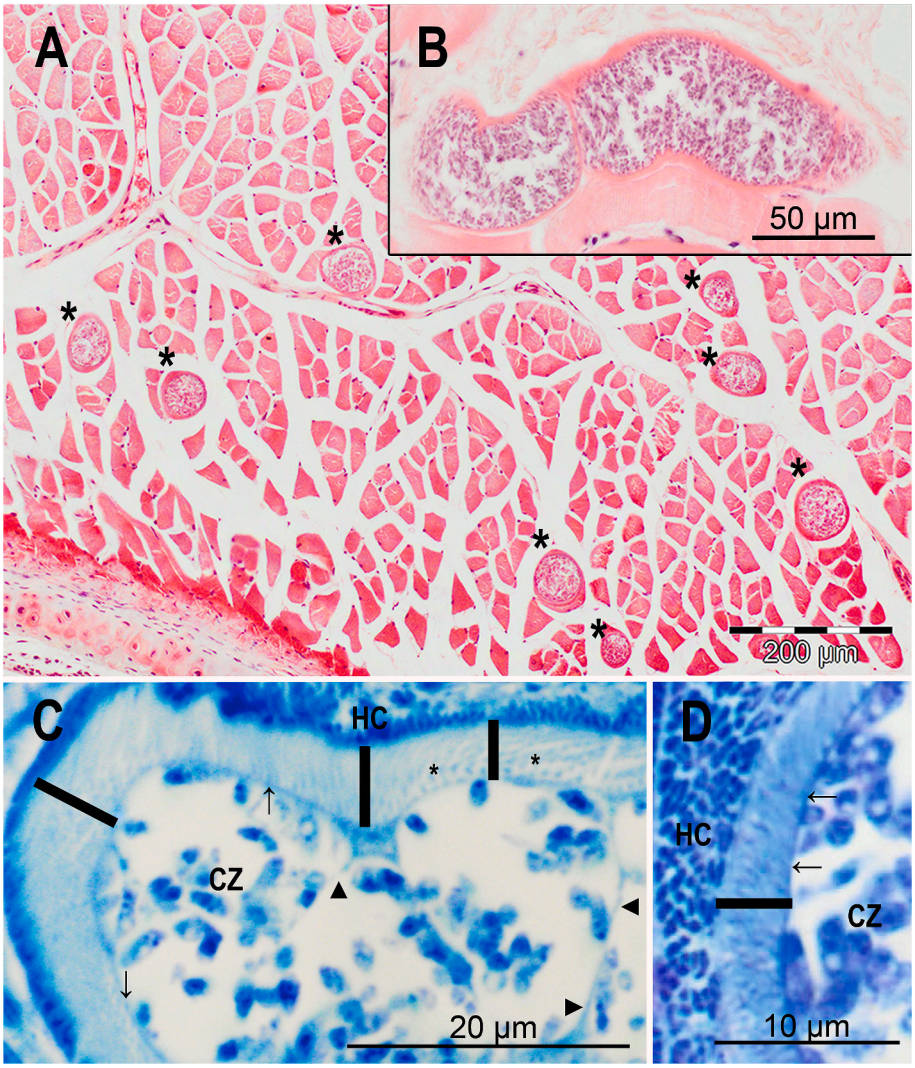

Fig. 1. Light microscopy of *Sarcocystis scandentiborneensis* sp. nov. A and B, Haematoxylin & Eosin-stained histological sections of striated musculature; C and D, Richardson's dye-stained 1.0 µm thin sections of sarcocysts. A) Tissue section of laryngeal muscle with various sarcocysts in cross section (asterisks), indicating a relatively high density of cysts in this part of musculature. B) Longitudinal section through a sarcocyst, showing a cigar-shaped appearance; however, isolated native sarcocysts, which were not available, may look different. C) Part of a longitudinal section through the tip of a sarcocyst, note the very thin ground substance (arrows) and the fine septae extending into the interior of the cyst (arrowheads); cystozoites (CZ) were loosely scattered within chambers while metrocytes were rarely seen, indicating maturity of the cyst; bars indicate the variable thickness of the cyst wall: the wall was thinner in regions where the villous protrusions were bent (right bar); note that the intense staining at the interface between host cell (HC) and parasite is part of the host cell. D) Cross-section through a sarcocyst showing cystozoites and the cyst wall (bar) including its thin ground substance (arrows).

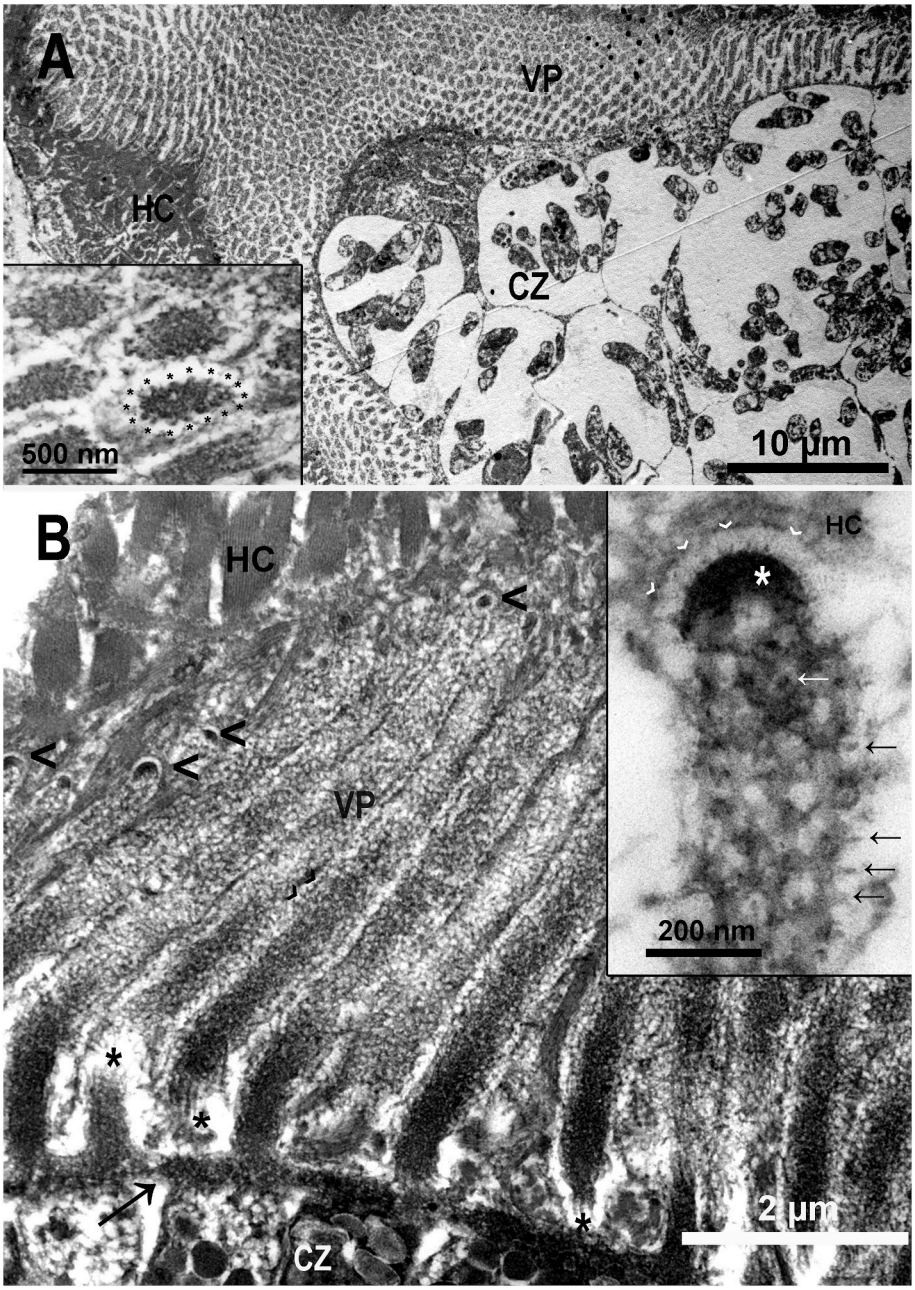

Fig. 2. Ultrastructure of *S. scandentiborneensis* sp. nov. Note, due to ethanol-fixation some ultrastructural details are poorly resolved (e.g. membranes). A) Longitudinal section through the same sample as in Fig. 1C, showing a gross view of the sarcocyst and its villous protrusions (VP) that are sectioned in different orientations. The inset shows a cross section through various VP that reveals the arrangement of microtubules in their inner core; while in this case 16 microtubules are visible (asterisks), sections through more apical portions of the VP showed lower numbers. B) Longitudinal section through the finger-like VPs that appear to be anchored in the ground substance (arrow) by microtubules (asterisks) that extend into each protrusion; note the electron-dense, U-shaped structure at each tip of the protrusions (arrowheads) and the apparently serrated surface of the VP (flat arrowheads). The inset shows a higher magnification of the apical part of a single VP with the typical U-shaped apex (asterisk), which appears to be connected with the host cell through an electron-lucent contact zone (white arrowheads); interestingly, the protrusion appears fenestrated (also visible in the main image) possessing thorn-like structures (black arrows; the white arrow indicates a cross-sectional view) that could be responsible for the serration visible at lower magnification. CZ, cystozoites; HC, host cell; VP, villous protrusion.

with the morphologically new *Sarcocytis* sp. (sample IDs: E357-13, E364-13, E120-13; verified by extensive histological screening). This enabled us to confidently relate the retrieved sequences to the corresponding sarcocyst morphology.

We isolated two full-length sequences (two haplotypes) of the nuclear 18S rRNA gene of *Sarcocystis* sp. from muscle tissue of two different specimen of *Tupaia minor*, E357-13 (1820 nt) and E364-13 (1819 nt). Initial screening of published sequences using BLAST returned a relatively high degree of similarity with a *Sarcocystis zuoi* isolate from Thailand (KU341120; 98.17% and 97.82% similarity, respectively). The life cycle of *Sarcocystis zuoi* includes a colubrid snake, *Elaphe carinata*, as definitive host and rats of the genus *Rattus* as intermediate hosts (Hu et al., 2012). Therefore, we included in our phylogenetic analyses all known sequences of *Sarcocystis* spp. from colubrid snakes, also different isolates from the same species to capture intraspecific variation. When using Maximum Likelihood (ML) or Bayesian Inference (BI; Fig. 3) for phylogeny reconstruction, the resulting tree topologies were similar: sequences E357-13 and E364-13 clustered together forming a distinct branch within a monophyletic clade, from which the basal species *S. singaporensis* branched off early, followed by *S. zamani* (both cycle between the reticulated python, *Malayopython reticulatus*, as definitive host and rats as intermediate hosts; Beaver and Maleckar, 1981) and a subclade that besides the sequences under investigation also contained *S. zuoi* and *S. clethrionomyelaphis* (which also develops in colubrid snakes; Matuschka, 1986). Branch support for the new sequences from Bornean treeshrews was high in both types of analyses (ML, bootstrap: 94–100%; BI, posterior probability: 1.0). However, tree topologies under BI and ML differed in one aspect: under ML all three *S. zuoi* sequences clustered together in a well-supported monophyletic subclade (bootstrap support: 95–100%); under BI, *S. zuoi* from China formed a subclade with sequences E357-13/E364-13 (Fig. 3; posterior probability: 0.9) from which the Thai isolate of *S. zuoi* branched off separately (posterior probability: 1.0) rendering the *S. zuoi* isolates paraphyletic. The relative divergence time of the branching node of sequences E357-13/E364-13

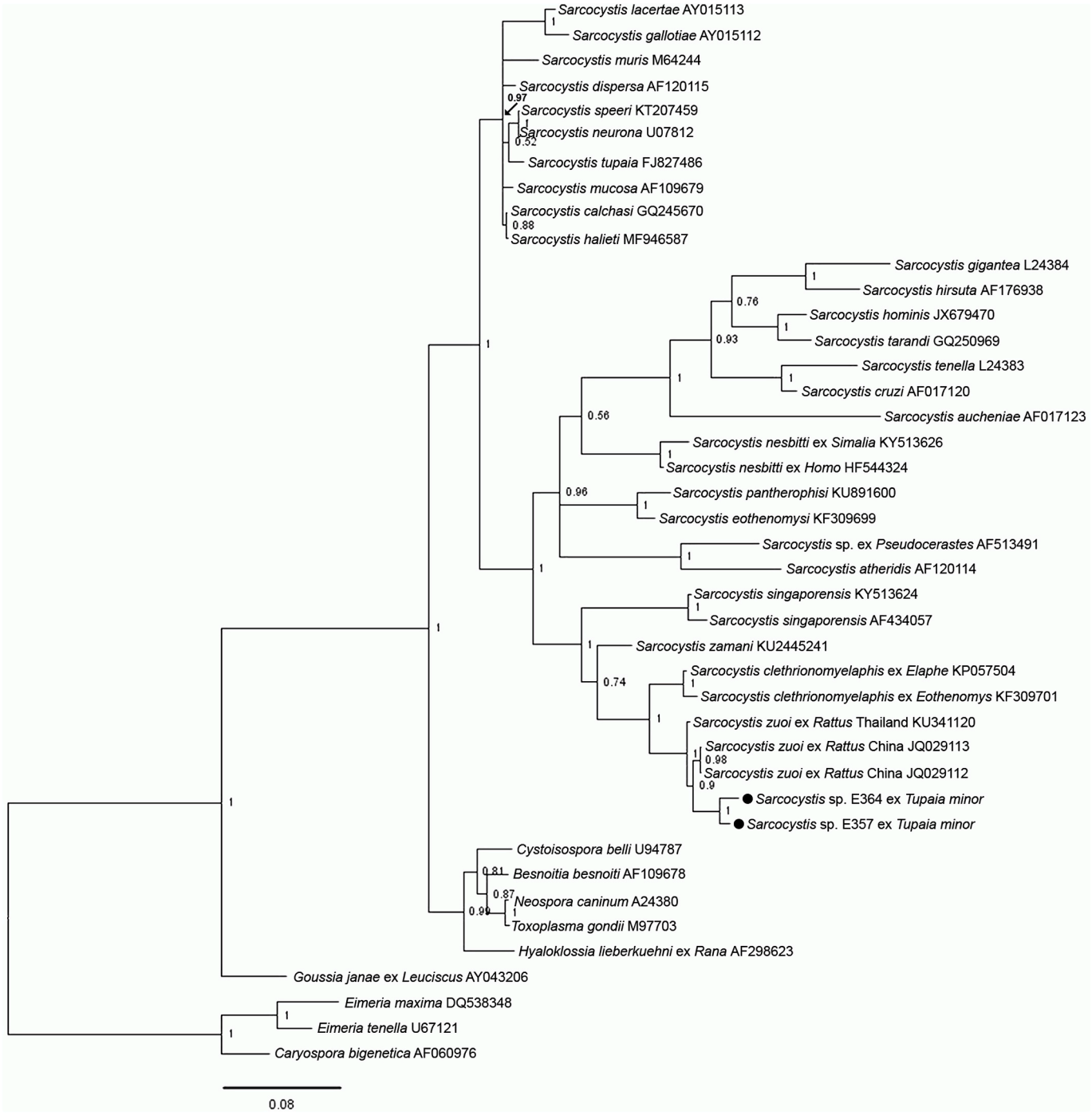

Fig. 3. Phylogenetic tree of nuclear 18S rDNA sequences of the Sarcocystidae, including the new Sarcocystis sp. from treeshrews (black symbols). Taxa of the Eimeriidae served as outgroup. Bayesian Inference was used for phylogeny reconstruction, whereby the general time-reversible substitution model (GTR + G + I) combined with an assumed among-site variation ('covariotide' model) was applied. Values for posterior probability are indicated behind nodes. Note that the new species is part of a monophyletic subclade previously tagged S1 (Wassermann et al., 2017), which includes taxa known to prefer snakes as definitive and rodents as intermediate hosts. Sequences E357-13 and E364-13 are available at GenBank under MN733816 and MN733817, respectively.

(0.0059) was about twice as long as that of *S. zuoi* (0.0027) as determined using the RelTime method based on the ML algorithm implemented in MEGA-7: this indicated that speciation time (*sensu* Hedges et al., 2015) was different in comparison to *S. zuoi.* Both, divergence between sequences E357-13/E364-13 and *S. zuoi* and within the *S. zuoi* isolates was further confirmed by our results on evolutionary rates and Ti/Tv ratios (see below). Interestingly, we identified the sequence motif 5′-AAUUCGU-3' (located in helix V7, see below) in sequences E357-13/E364-13 and *S. zuoi* (Thailand and China), an apparent synapomorphy of the '*S. zuoi* group' of species since it was only present in these sequences, not occurring in *S. clethrionomyelaphis* and other taxa which shared a similar but one nt shorter motif.

Pair-wise sequence comparison showed that the two 18S rDNA isolates of the *Sarcocystis* sp. from *T. minor* exhibited nt substitutions at 24 positions, mostly transitions, which resulted in a relatively high Ti/Tv ratio (Ti/Tv = 4.0, distance according to Timura-Nei model: 0.017). Sequence E364-13 was different to *Sarcocystis zuoi* from Thailand (KU341120) at 35 positions (Ti/Tv = 1.92; distance = 0.030), while it

exhibited 65 bp changes in comparison to *S. clethrionomyelaphis* (KP057504) (Ti/Tv = 1.83; distance = 0.075). Thus, in both interspecific comparisons Ti/Tv ratios were skewed towards transversions. Results for 18S rDNA sequence E357-13 were similar; except that ratios were even lower (1.50 and 1.43, respectively). A matrix of the Ti/Tv ratios is given in Supplementary Table S1, also including examples of other apicomplexan species. Comparison of *S. zuoi* with *S. clethrionomyelaphis* revealed 47 bp changes and a relatively low Ti/Tv ratio of 1.24 (distance = 0.044). In contrast, intraspecific variability of the two haplotypes of *S. clethrionomyelaphis* (KP057504, KF309701) was associated with a relatively high Ti/Tv ratio (3.67), similar to the pair E357-13 and E364-13. While intraspecific variation among the sequences of *S. zuoi* from China (JQ029112, JQ029113) was very low lacking transitions, the 10 and 8 bp-changes between these two and sequence KU341120 from Thailand were associated with Ti/Tv ratios of 1.50 and 0.60, respectively, resembling values of interspecific variability (although not complete, the sequences from China included the anterior, variable part of the 18S rRNA gene). In view of the observed pattern among Ti/Tv ratios (Table S1), we considered sequences E364-13 and E357-13 allelic variants of the 18S rRNA gene. Further evidence for this interpretation emerged when we tested evolutionary rates by Tajima's test, which compares two sequences with a third as outgroup (Tajima, 1993). While one could expect divergent paralogs to show significant evolutionary rate differences (Buckler et al., 1997), most of the intraspecific variation was not statistically significant, in particular between sequences E364-13 and E357-13. In contrast, interspecific comparisons revealed in most cases significantly different evolutionary rates (Supplementary Table S2).

In addition, we mapped all bp changes between sequences E357-13 and E364-13 (intraspecific) and between E364-13 and *S. zuoi/S. clethrionomyelaphis* (interspecific) to the predicted secondary structure of the 18S rRNA of *Toxoplasma gondii* (Gagnon et al., 1996) (Fig. 4). While intraspecific variation mainly occurred in helices V4, V2 and V7 (in decreasing order of frequency), interspecific variation was mainly observed in helices V2, V4, and V9, respectively. In both cases, most of the substitutions could be mapped to helices V2 and V4 (cumulatively 67% and 70%, respectively). However, a major difference between intra- and interspecific sequence variability pertained to Ti/Tv ratios of helices V2, V4, and V9, which were characterized by higher proportions of transversions in interspecific comparison (Fig. 4). Additionally, bp changes between species in these domains were mostly parsimony-informative, while intraspecific variability was mainly due to transitions that were often not parsimony-informative. Using E357-13 for interspecific comparison showed similar results.

When we excluded hypervariable helices V2 and V4 from phylogenetic analysis (which has been recommended for V4 for instance; Morrison, 2006) by using only the posterior half of sequence E357-13 (about 800 nt), phylogenetic inference by ML using truncated sequences of all taxa returned a tree topology that could not distinguish between *S. zuoi* and the species from treeshrews. A subsequent BLAST search using truncated E357-13 returned a 99.6% match with *S. zuoi* (KU341120), a value that was misleading in light of the full sequence.

We amplified five partial sequences Cytochrome Coxidase subunit 1 gene (COI), three from *Tupaia minor* (E364-13, 938 nt; E357-13, 944 nt, E351C-13, 908 nt) and two from *T. tana* (E120-13, 976 nt; E120B-13, 941 nt). Although length-variable, all sequences were 100% identical in pair-wise comparisons (one haplotype) despite origin from different host individuals and species. Initial BLAST search returned a 'best match' of 95.4% with *Sarcocystis canis* (KX721496) and *S. pantherophisi* (KU891603), which indicated that the sequences probably belonged to a previously unknown species. However, due to lack of relevant (see 18S rRNA analysis) COI sequences, we included in our phylogenetic analysis mainly taxa that ranked high in the BLAST results, furthermore *Sarcocystis* spp. with a carnivoran-ruminant life cycle, and out-group taxa for rooting the tree. Under Maximum Likelihood (ML) analysis, the *Sarcocystis* sp. from treeshrews was part of a monophyletic clade of *Sarcocystis* spp. from various hosts, which was split into a branch containing the new species and another including the rest of the taxa (Fig. 5). Notably, (identical) sequences E357-13, E364-13, and E120B-13 assumed a sister species position with regard to *S. pantherophisi* (and the partial sequence of *S. canis* that was identical with *S. pantherophisi*), which in 18S rRNA trees is an indicator-species of the S2 lineage of *Sarcocystis* spp. with snake-rodent life cycle (Wassermann et al., 2017). Interestingly, the Toxoplasmatinae formed a monophyletic clade that separated the *Sarcocystis* spp. of ruminants (as intermediate hosts) from the other *Sarcocystis* spp., rendering the Sarcocystinae paraphyletic (Fig. 5); bootstrap support for this scenario was relatively weak (62–74%). Such a tree topology has been observed previously for the tissue coccidia and was

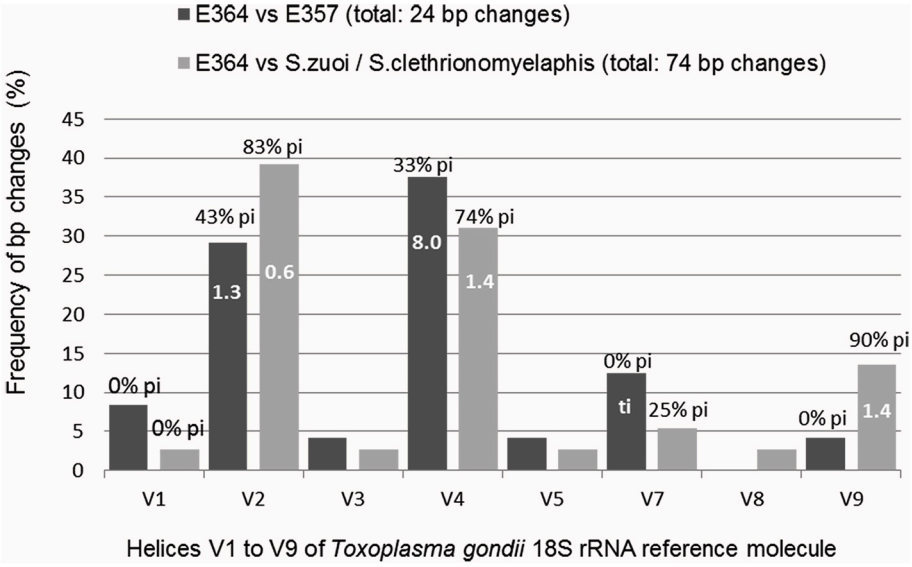

Fig. 4. Mapping (to the *Toxoplasma gondii* reference molecule M97703) of frequencies (%) of base pair changes observed in sequence comparisons of nu clear 18S rDNA within the new *Sarcocystis* sp. from treeshrews (intraspecific variation: isolates E364–13 versus E357–13) and between the new species and *Sarcocystis zuoi* and/or *S. clethrionomyelaphis* (interspecific variation: E364–13 versus *S. zuoi/clethrionomyelaphis*). Results were combined for the two latter species to simplify the graph. Here, 87.2% of 2118 alignment positions showed moderate to high levels of consistency, while sections of ambiguous alignment did not relate to the species under investigation. Due to gaps in the alignment, not all of the observed nt changes could be mapped to a homologous position of the reference molecule (i.e., 7 out of 24 bp changes in intraspecific comparison; 33 out of 74 bp changes in interspecific comparison), in which case the position of each nt relative to the helix was inferred from neighboring nt for which such position was known. Gaps were mainly due to insertions in helices V2, V4, and V9 rendering E357-13/E364-13 longer than the sequence of *T. gondii*. The percentage of parsimony-informative (pi) bp changes per helix is shown for helices V1, V2, V4, V7, and V9 above each column. Also shown is the ratio of transitions versus transversions (Ti/Tv) for selected helices.

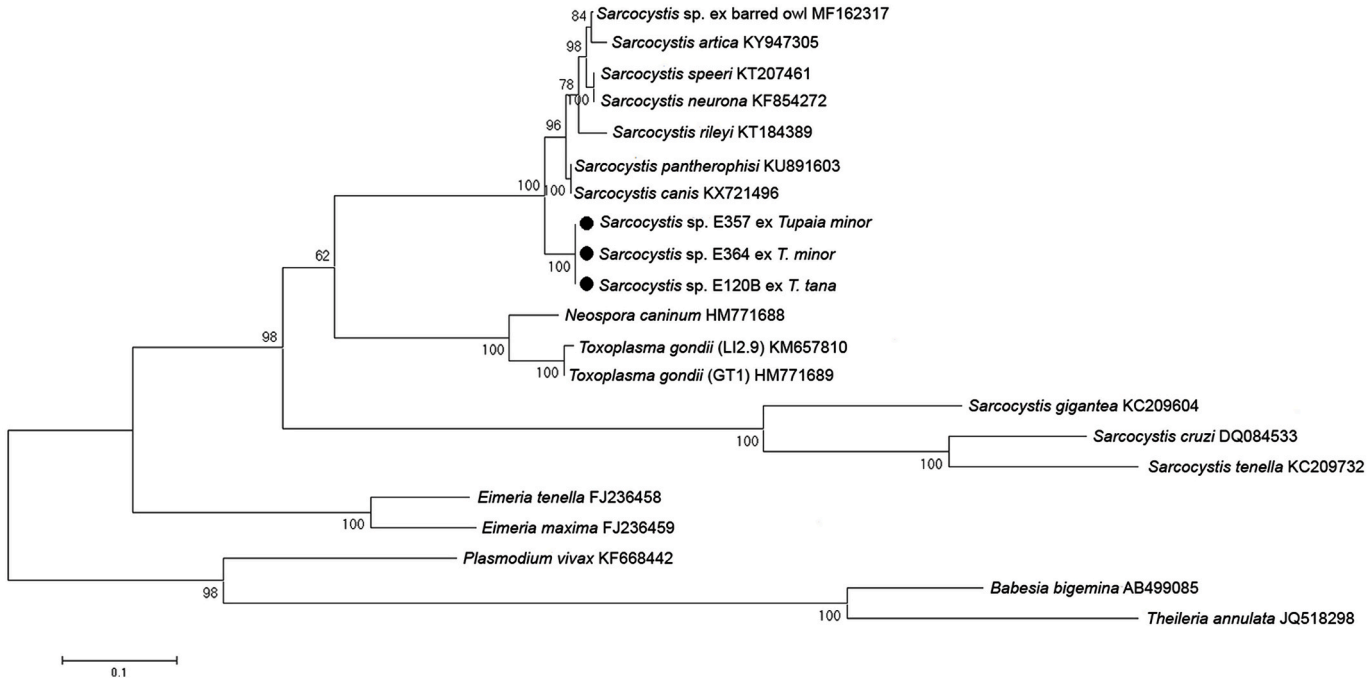

Fig. 5. Phylogenetic tree based on analysis of mitochondrial COI sequences of the Sarcocystidae including the new *Sarcocystis* sp. examined in this study (black symbols). Other taxa of the Apicomplexa served as root. Evolutionary history was inferred by the Maximum Likelihood (ML) method based on the Tamura-Nei model, whereby 619 positions were included in the final data set. All positions with less than 95% site coverage were eliminated; that is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. Bootstrap percentages (1000 iterations) are shown next to branches. COI sequences E357-13 and E120-13 (not shown in the tree) are available at GenBank (MN732561 and MN732562, respectively).

attributed to potential lack of key species (Ogedengbe et al., 2016). We were able to remedy this situation by excluding the 3rd codon position from phylogenetic analysis, which resulted in almost the same tree, but rendered the Toxoplasmatinae and Sarcocystinae as separate, monophyletic subclades of the Sarcocystidae.

Because the barcode area of COI is of global importance for comparison of species (Pentinsaari et al., 2016) and amplification of nuclear pseudogenes may complicate analysis (Gjerde, 2013a), we compared the aa composition of our translated consensus sequence (978 nt) with COI sequences of selected taxa of the Sarcocystidae. It comprised 326 aa residues (full CDS of *T. gondii*: 496 aa), of which 181 were conserved and 145 variable (113 parsimony informative) among six taxa compared (see footnote of Table 1). The sequence started with Glycine and was in large parts identical with the CDS of *T. gondii*; the nt sequence did not contain stop codons. Details of the COI protein sequence of the *Sarcocystis* sp. from treeshrews with regard to functionally important, structural elements of the barcode area are presented in Table 1. This part of the sequence is shorter, spanning 235 aa without gaps between the elements. Although a part of the anterior end of the barcode area was missing, we identified most of the 23 highly conserved aa of the six alpha-helices of COI including the loop regions (Table 1). The protein sequence of the new *Sarcocystis* sp. exhibited three aa substitutions that were not shared with any of the other species, but due to lack of sequences from closely related taxa it is unclear if these changes are species-specific. Interestingly, we observed a relatively high frequency of Serine residues (16.8% versus 12.5% of *T. gondii*., 7.5% of *S. cruzi*, and 5.0% of *S. tenella*) among the 113 parsimony-informative aa of our partial sequence, while the situation was opposite with regard to Alanine (6.2% versus 10.0%, 17.5%, and 19.2%, respectively). Interspecific differences between other aa were less pronounced. When this

Table 1
Protein sequences of the COI barcode elements as translated from nt sequence E357–13 of the *Sarcocystis* sp. isolated from *Tupaia minor* (GenBank: MN732561). Three amino acid positions different to those of other taxa of the Sarcocystidae[a] are shown in bold-italic letters; universally conserved aa are underlined. Note that not all of the 23 universally conserved aa across the tree of life (Pentinsaari et al., 2016) are present in COI of the parasite. IUPAC code for aa is used.

COI barcode element	Protein sequence	Start position in sequence	Universally conserved amino acids[b]
Helix 1	(initial 19 aa missing) GTLMSNMIRVELS	1	Y-R
Loop 1-2	S***T***GSRIICV	14	
Helix 2	ESISTYNVIITLHGLSMIFMFLMPALYAGFGNYFVPIY	23	H-FF-P-G-N-P
Loop 2-3	IGSSEVVY	61	D
Helix 3	PRINAVSYFLVPLGSILVMQSII	69	PR-N
Loop 3-4	SEFGSGMGWTMYPPLSTSLMTLN	92	WT-YPP
Helix 4	TESVDWIISGL***G***ILGISSVLGSINFLGTCLFL	115	H-G-N
Loop 4-5	GAVSGCRPNI	147	
Helix 5	LYIWAI***M***FTAIMLILTLPILTGGLVMLLMDLHT	157	D
Loop 5-6	NSEFYDSMYSGD	190	G
Helix 6	SVLYQHLFWFFGHPEVYILILPGFGVISQTLSTY	202	

[a] *Sarcocystis tenella* (GenBank KC209723; included in the barcode database BOLD under accession number JRPAA5858-15, http://www.boldsystems.org), *S. cruzi* (DQ084533), *S. singaporensis* (T.J.; unpublished sequence), *Toxoplasma gondii* (KM657810); the full CDS of *T. gondii* as published by Namasivayam (2015) was used to select the correct open reading frame.

[b] Alignment with the aa sequence of *Bos taurus* (AF493542) facilitated identification of the aa of the barcode area that are conserved across the tree of life.

type of analysis was applied to the area of the functionally important elements (235 aa), results were similar: Serine accounted for 11.9% (versus 9.8% in *T. gondii*, 7.2% in *S. cruzi*, and 6.8% in *S. tenella*) and Alanine for 2.6% of the residues (versus 4.3%, 10.2%, and 8.1%, respectively). This pattern in variation of Serine and Alanine counts between species was highly significant (Chi-square = 16.956, d.f. = 3, $P \leq 0.001$) suggesting that there exists a linkage between the frequency variation of the two aa. Another statistically highly significant relationship was observed between Alanine and Isoleucine (Chi-square = 14.554, d.f. = 3, $P = 0.002$).

3.4. Description and taxonomic summary

Sarcocystis scandentiborneensis sp. nov. (Figs. 1–5, Table 1).

Diagnosis: Microscopic sarcocysts (about 500 × 50 μm and smaller in histological sections) in skeletal muscles of treeshrews (Scandentia). Cyst wall striated with brush-like appearance, thickness variable (average 4–5 μm, maximal 10 μm), including a very thin layer of GS (about 0.3 μm) and finger-like VP, that are tightly packed, their tip revealing an electron-dense, U-shaped apex at ultrastructural level. VP contain longitudinally oriented microtubules that extend to the GS and show thorn-like structures on the surface. Loosely packed cystozoites in chambers delineated by thin septae have a maximum length of 6.7 μm in H&E-stained sections.

Taxonomic summary

Type (natural) intermediate host: *Tupaia minor* (lesser treeshrew) and *T. tana* (large treeshrew).

Definitive host: Unknown.

Type locality: Poring Hot Springs, Northern Borneo, Sabah, Malaysia.

Other locations: Unknown.

Etymology: Species named after the order of the intermediate host (Scandentia) and the geographic location of collection.

Specimens deposited: Striated muscle specimens from eleven *Tupaia* sp. individuals are deposited at the Sabah Parks Museum, Kinabalu Park, Borneo, Malaysia and the Museum für Naturkunde – Leibniz Institute for Evolution and Biodiversity Science (MfN), Berlin, Germany, including: tissue sections stained with H&E and Richardson's staining and ethanol-fixed, paraffin embedded as well as resin embedded muscle tissue, and extracted DNA stored at −80 °C. Sarcocysts in tissue samples from *Tupaia minor* E357-13 are considered the holotypus of *Sarcocystis scandentiborneensis* sp. nov. with the MfN registration number ZMB-Protoz 66. The other deposited samples serve as paratypes with the MfN registration number ZMB-Protoz 67.

Sequences deposited: Nuclear 18S rDNA: full-length sequences E357-13 (MN733816) and E364-13 (MN733817); mitochondrial COI: partial sequences E120-13 (MN732562) and E357-13 (MN732561).

ZooBank registration number (http://zoobank.org/References/): C0DCC101-235B-41E0-97 EB-45ABB74517F6.

3.5. Taxonomic remarks

The ultrastructure of the sarcocyst wall is in many cases a useful criterion for distinguishing between species of *Sarcocystis* (Dubey et al., 2016). The sarcocyst wall of the species under investigation is ultrastructurally as well as histologically distinct from the only other *Sarcocystis* sp. known to occur in treeshrews: *S. tupaia*, found in *Tupaia belangeri chinensis* from Yunnan, China (Xiang et al., 2010). Although our TEM results were not optimal due to ethanol-fixation, the ultrastructural details nevertheless permit comparison and categorization of wall morphology (Dubey et al., 2016). While *S. tupaia* possesses a smooth type 1 wall (Xiang et al., 2010), the ultrastructure of the present species clearly reveals a thick striated wall with VP that share structural elements with the tightly packed protrusions of wall type 12. Because our phylogenetic results indicate a sister taxon relationship with three previously described *Sarcocystis* species that undergo a snake-rodent life cycle (Slapeta et al., 2003; Hu et al., 2012, 2014, 2015; Wassermann et al., 2017), morphological comparison with those species is implicit: *S. singaporensis*, *S. clethrionomyelaphis*, or *S. zuoi* all possess spatula or finger-like VP, but ultrastructurally they do not match with the new species. There is also no match with another nine, relatively thick-walled snakehost *Sarcocystis* spp. as accounted for in Dubey et al. (2016) and discussed by Verma et al. (2017): *S. acanthocolubri*, *S. gongyli*, *S. lacertae*, *S. podarcicolubris*, *S. stenodactylicolubris*, *S. hoarensis*, *S. murinotechis*, *S. muriviperae*, and *S. villivillosi* (*S. zamani* cannot be included here, since it exhibits a quite thin cyst wall; Beaver and Maleckar, 1981). Furthermore, unlike *S. zuoi*, *S. clethrionomyelaphis*, and representatives of the S2 lineage of snakehost *Sarcocystis* (Wassermann et al., 2017) such as *S. pantherophisi* (Verma et al., 2017) and *S. eothenomysi* (Hu et al., 2014), cystozoites of the *Sarcocystis* sp. infecting Bornean treeshrews are markedly smaller: similar in size to species with a snake-lizard life cycle, where cystozoites are between 4 and 7 μm long (Matuschka, 1981; Abdel-Ghaffar et al., 1990; Modry et al., 2000; Morsy et al., 2012). Biologically, the full life cycle of the new species remains unknown, since we could not identify a definitive host. However, our phylogenetic results on the 18S rRNA gene suggest a snake as definitive host (see below). Regarding intermediate host specificity, our trapping results provide some insight: none of the murid rodent species captured in the same location was infected with the new *Sarcocystis* sp. from tree shews, suggesting that the Scandentia are the preferred hosts. Molecularly, the *Sarcocystis* sp. described here forms a close sister species relationship with *Sarcocystis* spp. with colubrid snake-rodent life cycle in 18S rDNA phylogenies. Our phylogenetic and molecular results indicate that the species from Borneo is distinct from *S. zuoi*; it also appears likely that *S. zuoi* reported from Thailand may be a different species, if one considers the topology of the BI tree, Ti/Tv ratios, and differences in evolutionary rates between sequences. Preliminary reconstruction of the COI gene tree suggests a sister-species relationship with *S. pantherophisi* (KU891603), but due to lack of COI sequences from closely related species this part of the tree can currently not be resolved. The question why the partial COI sequence of *S. canis* is identical to *S. pantherophisi*, including the potential taxonomic implications, is beyond the scope of this study. Considering all morphological, biological, and molecular aspects of the taxa compared, we regard *S. scandentiborneensis* a new species.

4. Discussion

Our results clearly indicate that *Sarcocystis scandentiborneensis* belongs to a different genetic lineage than *S. tupaia* with regard to its 18S rDNA sequence, expanding the S1 subclade of *Sarcocystis* spp. with snake-rodent life cycle (Wassermann et al., 2017) to include a new host order, the Scandentia. The observation that *S. scandentiborneensis* infects treeshrews showcases that taxon sampling in this group of *Sarcocystis* spp. was insufficient until today to infer pattern of host sharing and shifting (Wells and Clark, 2019), so that finding a *Sarcocystis* sp. of this genetic lineage that utilizes a non-rodent intermediate host was overdue. Because host sampling has been largely focussed on murid rodents, particularly in Southeast Asia (Kan and Pathmanathan, 1991; Jäkel et al., 1997; Paperna et al., 2004; Fong et al., 2014), future examination of other Rodentia (e.g., squirrels) or non-rodent taxa (e.g., Lagomorpha, Eulipothyphla) likely to be predated by suitable definite hosts to close transmission cycles could reveal interesting insights into the host spectrum of already known and new *Sarcocystis* spp. in the wild. The natural definitive hosts of *S. tupaia* and *S. scandentiborneensis* are unknown (Xiang et al., 2010; Dubey et al., 2016). However, it would come as no surprise if both used snakes as definitive hosts. While phylogenetic tree topology implies this for *S. scandentiborneensis*, it has to be noted that *S. tupaia* is a member of a subclade of the Sarcocystinae (in 18S rRNA trees) that also includes species like *S. lacertae* and other snakehost *Sarcocystis* spp. (Tome et al., 2013; Abe et al., 2015). In the wild, *T. minor*, *T. tana* and other treeshrews do fall prey to snakes that co-exist in their habitats in trees as well as understory of the dipterocarp forests of Borneo

(Emmons, 2000). Among the various colubrid species occurring in Borneo, the mangrove snake (*Boiga dendrophila*) has been spotted as predator of treeshrews (Munshi-South, 2005).

Accurate sequence alignment is a major pre-requisite for phylogeny reconstruction, and each multiple sequence alignment is ultimately a statement of homology (Anisimova et al., 2010). We have chosen a computerized approach to alignment (Moretti et al., 2008; Tommaso et al., 2011), because some of the length-variable sections of the 18S rDNA have been difficult to align (Morrison, 2006) and we wanted to rely on a process with a high degree of objectivity and reproducibility (Anisimova et al., 2010). Previous work has already shown that an RNA structural template-based multiple alignment can resolve well basal and crown species in phylogenetic trees of the Sarcocystidae (Wassermann et al., 2017). In combination with mapping of nt variability to the secondary structure of the rRNA and analyzing Ti/Tv ratios, we think that this approach facilitated important insights into the intra- and interspecific variability of the *Sarcocystis* lineage under investigation, without which it would have been difficult to distinguish between *S. scandentiborneensis* and *S. zuoi* at the molecular level, albeit their distinct sarcocyst morphologies. A previous analysis revealed a relatively high level of intraspecific variability of the 18S rDNA of *S. singaporensis* that did not prevent, however, the formation of a monophyletic clade of the various haplotypes (Slapeta et al., 2002). We agree with El-Sherry et al. (2013) that high intraspecific variability of nuclear 18S rDNA due to paralogous sequences complicates molecular identification of coccidian parasites. Nevertheless, this should not discourage future phylogenetic study of this gene. If one assumes that the so-called 'birth-and-death' model of evolution applies generally to the rRNA multi-gene family of the Apicomplexa (Rooney, 2004), one would expect to observe the hallmarks of that model, i.e. interspecific gene clustering and presence of pseudogenes (Nei and Rooney, 2005; Eirin-Lopez et al., 2012). While nothing is known about pseudogenes (non-functional gene copies) among the *Sarcocystis* spp. examined here, the existence of paralogs (gene duplicates) has been suggested for *S. singaporensis* (Slapeta et al., 2002). Interspecific sequence clustering of 18S rDNA has been especially observed among *Plasmodium* spp. (Rooney, 2004) and highly divergent paralogs were found in *Cryptosporidium* (Stenger et al., 2015). In view of the 'birth-and-death' model and based on our preliminary analysis of Ti/Tv ratios among apicomplexan taxa, we hypothesize that allelic variants or recent paralogs of the 18S rDNA may be generally characterized by higher Ti/Tv values or presence of transitions only, while comparison of orthologs and older paralogs (where nt substitutions have accumulated over time) would yield values similar to or lower than the average interspecific ratio determined for 56 alveolate protists (i.e., 1.85; Leander and Ramey, 2006). This hypothesis is in line with the observation that paralogs may be associated with stage-specific functions in *Plasmodium* (Mercereau-Puijalon et al., 2002). It has been shown recently that transversions have larger regulatory effects, e.g. on binding of transcription factors, than transitions in non-protein-coding DNA (Guo et al., 2017). Taken together, this could potentially render Ti/Tv ratios and other structural information of the 18S rRNA gene (for example if parsimony-informative substitutions relate to functionally constrained parts of the molecule) useful tools for studying evolution and functional diversification at species or subspecies level, provided that paralogous sequences and pseudogenes are identified as such (Buckler et al., 1997). As experienced in this study, it is necessary to use complete or nearly complete sequences of 18S rDNA for such analyses. In view of an average speciation time of eukaryotic species of about two million years (Hedges et al., 2015), one might appreciate additional information supplementing phylogenetic trees since divergence in shorter periods of time may not lead to real species but could reflect phenotypic adaptations of transitional populations or subspecies (Hedges et al., 2015).

Despite the issue regarding paralogs, phylogeny reconstruction in the phylum Apicomplexa has been for a long time solely based on the 18S rRNA gene, mainly because of its universal applicability to the tree of life (Morrison, 2009; Xie et al., 2011). For discrimination of species of the Sarcocystidae the 18S RNA gene has been generally regarded useful (Ellis and Morrison, 1995; Morrison et al., 2004), but there exist limitations since certain lineages can be resolved better than others (Morrison et al., 2004). For instance, the carnivoran-ruminant and snake-rodent *Sarcocystis* spp. (including the species described here) show higher evolutionary rates and can be better resolved than the cluster of *Sarcocystis* spp. that includes *S. neurona* (Morrison et al., 2004). This is confirmed by the present study, where what we call the '*S. neurona*-group' of species showed short branch lengths and remained largely polytomic under BI. Also, other sequences such as 28S rRNA, COI, ITS1, and the *rpoB* gene have shown mixed results in resolving this group (Gjerde et al., 2018; Kirillova et al., 2018). Thus, we believe that the problem of resolving this part of the tree also extends to other genes and may be best addressed by multigene phylogenetic reconstruction (Gadagkar et al., 2005).

Thus far, we find COI useful for delineation of *Sarcocystis* spp. of the snakehost lineages. However, the number of related taxa for which COI has been sequenced is still too small for reliably drawing conclusions on the position of *S. scandentiborneensis* in the phylogenetic tree and comparing COI with the 18S rRNA gene. Since we amplified five identical sequences of COI from five different hosts, this suggests that this gene shows no or little intraspecific variability within the given parasite population. This is in line with results from *Sarcocystis* spp. from ruminants where on average 0 to 9 nt differences occurred over a 1020 bp region (Gjerde, 2013b). Furthermore, a recent study of six *Sarcocystis* spp. infecting roe deer showed that COI intraspecific variability coincided with intraspecific variability of the 18S rRNA gene, rendering species-, not gene-specific genetic variability (Rudaitytė-Lukošienė et al., 2020). Future sampling however has to show whether low variability is true for the entire geographical distribution and host range of *S. scandentiborneensis* and related species. Species-specificity and low variability of the sequence would certainly support the concept of 'barcoding' of the animal tree of life, for which COI has been selected (Hebert et al., 2003). We can confirm that the barcode fragment of 219 amino acids (Pentinsaari et al., 2016) is part of the COI sequence of *S. scandentiborneensis* (except for a few initial aa that are missing). In view of expanding the tools for characterization of species, this could be studied in the future under the functionally important aspect of COI being located at the core of energy production within cells (Srinivasan and Avadhani, 2012), whereby the Apicomplexan Cytochrome C oxidase complex differs substantially in its protein composition from the hosts they infect (Seidi et al., 2018). For instance, our preliminary analysis suggests that there may be considerable interspecific variability of COI regarding the frequency of Alanine and Serine among taxa of the Sarcocystidae. Substitutions involving these two aa have been implicated as functionally important (Skibinski et al., 2018).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. We thank the Sabah Biodiversity Council (Majlis Biodiversity Sabah) for their support in obtaining research permits. We are grateful to all citizens in Sabah who provided access to their properties or were in other ways helpful during field work. Jorimia Molubi, Brigitte Fiala and K. Eduard Linsenmair and the Sabah Parks research department kindly provided logistic support. The excellent technical assistance of Doris Krumnow, Dagmar Viertel and Susanne Auls, as well Dr. Kristin Mühldorfer's help for the COI PCR optimization, is appreciated with much gratitude.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijppaw.2020.07.003.

References

Abdel-Ghaffar, F., Bashtar, A.R., Ashour, M.B., Sakran, T., 1990. Life cycle of *Sarcocystis gongyli* Trinci 1911 in the skink *Chalcides ocellatus ocellatus* and the snake *Spalerosophis diadema*. Parasitol. Res. 76, 444–450.

Abe, N., Matsubara, K., Tamukai, K., Miwa, Y., Takami, K., 2015. Molecular evidence of *Sarcocystis* species in captive snakes in Japan. Parasitol. Res. 114, 3175–3179.

Andree, K., Axtner, J., Bagley, M.J., Barlow, E.J., Beebee, T.J., Bennetzen, J.L., et al., 2010. Permanent genetic resources added to molecular ecology resources database 1 april 2010 –31 may 2010. Mol. Ecol. Resour. 10, 1098–1105.

Anisimova, M., Cannarozzi, G.M., Liberies, D.A., 2010. Finding the balance between the mathematical and biological optima in multiple sequence alignment. Trends Evol. Biol. 2, 39–48.

Beaver, P.C., Maleckar, J.R., 1981. *Sarcocystis singaporensis* Zaman and Colley, (1975) 1976, *Sarcocystis villivillosi* sp. n., and *Sarcocystis zamani* sp. n.: development, morphology, and persistence in the laboratory rat, *Rattus norvegicus*. J. Parasitol. 67, 241–256.

Buckler, E.S., Ippolito, A., Holtsford, T.P., 1997. The evolution of ribosomal DNA: divergent paralogues and phylogenetic implications. Genetics 145, 821–832.

Clark, K., Karsch-Mizrachi, I., Lipman, D.L., Ostell, J., Sayers, E.W., 2016. GenBank. Nucleic Acids Res. 44, D67–D72.

Dubey, J.P., Calero-Bernal, R., Rosenthal, B.M., Speer, C.A., Fayer, R., 2016. Sarcocystosis of Animals and Humans. CRC Press, Taylor & Francis Group, Boca Raton.

Eirin-Lopez, J.M., Rebordinos, L., Rooney, A.P., Rozas, J., 2012. In: Repetitive, D.N.A., Garrido-Ramos, M.A. (Eds.), The Birth-And-Death Evolution of Multigene Families Revisited, vol. 7. Genome Dyn. Basel, Karger, pp. 170–196.

Ellis, J., Morrison, D., 1995. Effects of sequence alignment on the phylogeny of *Sarcocystis* deduced from 18S rDNA sequences. Parasitol. Res. 81, 696–699.

El-Sherry, S., Ogedengbe, M.E., Hafeez, M.A., Barta, J.R., 2013. Divergent nuclear rDNA paralogs in a Turkey coccidium, *Eimeria meleagrimitis*, complicate molecular systematics and identification. Int. J. Parasitol. 43, 679–685.

Emmons, L.H., 2000. Tupai: A Field Study of Bornean Treeshrews. University of California Press, Berkely, California.

Fong, J.H., Voon, K., Ambu, S., Mak, J.W., 2014. Phylogenetic analysis and identification of *Sarcocystis* spp. found in rodents in Peninsular Malaysia. IeJSME 8, 12–17.

Gadagkar, S.R., Rosenberg, M.S., Kumar, S., 2005. Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. J. Exp. Zool. Part B. 304B, 64–74.

Gagnon, S., Bourbeau, D., Levesque, R.C., 1996. Secondary structures and features of the 18S, 5.8s and 26s ribosomal RNAs from the Apicomplexan parasite *Toxoplasma gondii*. Gene 173, 129–135.

Gjerde, B., 2013a. Characterisation of full-length mitochondrial copies and partial nuclear copies (numts) of the cytochrome b and cytochrome c oxidase subunit I genes of *Toxoplasma gondii, Neospora caninum, Hammondia heydorni* and *Hammondia triffittae* (Apicomplexa: Sarcocystidae). Parasitol. Res. 112, 1493–1511.

Gjerde, B., 2013b. Phylogenetic relationships among *Sarcocystis* species in cervids, cattle and sheep inferred from the mitochondrial cytochrome c oxidase subunit I gene. Int. J. Parasitol. 43, 579–591.

Gjerde, B., Vikøren, T., Hamnes, I.S., 2018. Molecular identification of *Sarcocystis halieti* n. sp., *Sarcocystis lari* and *Sarcocystis truncata* in the intestine of a white-tailed sea eagle (*Haliaeetus albicilla*) in Norway. Int. J. Parasitol. Parasites Wildl. 7, 1–11.

Guo, C., McDowell, I.C., Nodzenski, M., Scholtens, D.M., Allen, A.S., Lowe, W.L., Reddy, T.E., 2017. Transversions have larger regulatory effects than transitions. BMC Genom. 18, 394.

Hebert, P.D.N., Cywinska, A., Ball, S.L., deWaard, J.R., 2003. Biological identifications through DNA barcodes. Proc. Roy. Soc. Lond. B 270, 313–321.

Hedges, S.B., Marin, J., Suleski, M., Paymer, M., Kumar, S., 2015. Tree of Life reveals clock-like speciation and diversification. Mol. Biol. Evol. 32, 835–845.

Hu, J.J., Meng, Y., Guo, Y.M., Liao, J.Y., Song, J.L., 2012. Completion of the life cycle of *Sarcocystis zuoi*, a parasite from the Norway rat, *Rattus norvegicus*. J. Parasitol. 98, 550–553.

Hu, J.J., Liu, Q., Yang, Y.F., Esch, G.W., Guo, Y.M., Zou, F.C., 2014. *Sarcocystis eothenomysi* n. sp. (Apicomplexa: Sarcocystidae) from the large oriental vole *Eothenomys miletus* (Thomas) (Cricetidae: Microtinae) from anning, China. Syst. Parasitol. 89, 73–81.

Hu, J.J., Liu, T.-T., Liu, Q., Esch, G.W., Chen, J.-Q., 2015. *Sarcocystis clethrionomyelaphis* Matuschka, 1986 (Apicomplexa: Sarcocystidae) infecting the large oriental vole *Eothenomys miletus* (Thomas) (Cricetidae: Microtinae) and its phylogenetic relationships with other species of *Sarcocystis* Lankester, 1882. Syst. Parasitol. 91, 273–279.

Jäkel, T., Khoprasert, Y., Sorger, I., Kliemt, D., Seehabutr, V., Suasa-ard, K., Hongnark, S., 1997. Sarcosporidiasis in rodents from Thailand. J. Wildl. Dis. 33, 860–867.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., Madden, T.L., 2008. NCBI BLAST: a better web interface. Nucleic Acids Res. 36, W5–W9 (Web Server issue).

Kan, S.P., Pathmanathan, R., 1991. Review of sarcocystosis in Malaysia. Southeast Asian J. Trop. Med. Publ. Health 22 (Suppl. l), 129–134.

Kirillova, V., Prakas, P., Calero-Bernal, R., Gavarāne, I., Fernández-García, J.L., Martínez-González, M., Rudaitytė-Lukošienė, E., Martínez-Estéllez, M.A.H., Butkauskas, D., Kirjušina, M., 2018. Identification and genetic characterization of *Sarcocystis arctica* and *Sarcocystis lutrae* in red foxes (*Vulpes vulpes*) from Baltic States and Spain. Parasites Vectors 11, 173.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874.

Leander, B.S., Ramey, P.A., 2006. Apicomplexans: Description of the marine parasite *Rhytidocystis polygordiae* n. sp. (host: *Polygordiius* sp., Polychaeta). J. Eukaryot. Microbiol. 53, 280–291.

Matuschka, F.R., 1986. *Sarcocystis clethrionomyelaphis* n. sp. from snakes of the genus *Elaphe* and different voles of the family Arvicolidae. J. Parasitol. 72, 226–231.

Matuschka, F.R., 1981. Life cycle of *Sarcocystis* between poikilothermic hosts. Lizards are intermediate hosts for *S. podarcicolubris* sp. nov., snakes function as definitive hosts. Z. Naturforsch. 36c, 1093–1095.

Mercereau-Puijalon, O., Barale, J.-C., Bischoff, E., 2002. Three multigene families in *Plasmodium* parasites: facts and questions. Int. J. Parasitol. 32, 1323–1344.

Modry, D., Koudela, B., Slapeta, J.R., 2000. *Sarcocystis stenodactylicolubris* n. sp., a new sarcosporidian coccidium with a snake-gecko heteroxenous life cycle. Parasite 7, 201–207.

Moretti, S., Wilm, A., Higgins, D.G., Xenarios, I., Notredame, C., 2008. R-Coffee: a web server for accurately aligning noncoding RNA sequences. Nucleic Acids Res. 36 (Web Server issue).

Morrison, D.A., 2009. Evolution of the Apicomplexa: where are we now? Trends Parasitol. 25, 375–382.

Morrison, D., 2006. Phylogenetic analyses of parasites in the new millennium. Adv. Parasitol. 63, 1–124.

Morrison, D.A., Bornstein, S., Thebo, P., Wernery, U., Kinne, J., Mattsson, J.G., 2004. The current status of the small subunit rRNA phylogeny of the coccidia (Sporozoa). Int. J. Parasitol. 34, 501–514.

Morsy, K., Bashtar, A.R., Abdel-Ghaffar, F., Mehlhorn, H., Quraishy, S.A., Al-Ghamdi, A., Koura, E., Maher, S., 2012. *Sarcocystis acanthocolubri* sp. n. infecting three lizard species of the genus *Acanthodactylus* and the problem of host specificity. Light and electron microscope study. Parasitol. Res. 110, 355–362.

Munshi-South, J., 2005. *Boiga dendrophila* (mangrove snake) diet. Herpetol. Rev. 36, 188.

Namasivayam, S., 2015. Repetitive DNA and Nuclear Integrants of Organellar DNA Shape the Evolution of Coccidian Genomes. Dissertation. The University of Georgia, Athens, Georgia.

Nei, M., Rooney, A.P., 2005. Concerted and birth-and-death evolution of multigene families. Annu. Rev. Genet. 39, 121–152.

Odening, K., 1998. The present state of species-systematics in *Sarcocystis* Lankester, 1882 (Protista, Sporozoa, coccidia). Syst. Parasitol. 41, 209–233.

Ogedengbe, M.E., Ogedengble, J.D., Whale, J.C., Elliot, K., Juarez-Strada, M.A., Barta, J. R., 2016. Molecular phylogenetic analyses of tissue coccidia (Sarcocystidae; Apicomplexa) based on nuclear 18s RDNA and mitochondrial COI sequences confirms the paraphyly of the genus *Hammondia*. Parasitol. Open 2 (e2), 1–16.

Paperna, I., Peh, K.S.-H., Koh, L.P., Sodhi, N.S., 2004. Factors affecting *Sarcocystis* infection of rats on small tropical islands. Ecol. Res. 19, 475–483.

Payne, J., Francis, C.M., Phillipps, K., 1998. A Field Guide to the Mammals of Borneo. The Sabah Society, WWF Malaysia, Kota Kinabalu, Kuala Lumpur.

Pentinsaari, M., Salmela, H., Mutanen, M., Roslin, T., 2016. Molecular evolution of a widely adopted taxonomic marker (COI) across the animal tree of life. Sci. Rep. 6, 35275.

Prakas, S., Oksanen, A., Butkauskas, D., Sruoga, A., Kutkienė, L., Švažas, S., Isomursu, M., Liaugaudaitė, S., 2014. Identification and intraspecific genetic diversity of *Sarcocystis rileyi* from ducks, *Anas* spp., in Lithuania and Finland. J. Parasitol. 100, 657–661.

Reynolds, E.S., 1963. The use of lead citrate at high pH as an electron-opaque stain in electron microscopy. J. Cell Biol. 17, 208–212.

Richardson, K.C., Jarrett, L., Finke, E.H., 1960. Embedding in epoxy resin for ultrathin sectioning in electron microscopy. Stain Technol. 35, 313.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model Choice across a large model Space. Syst. Biol. 61, 539–542.

Rooney, A.P., 2004. Mechanisms underlying the evolution and maintenance of functionally heterogeneous 18S rRNA genes in Apicomplexans. Mol. Biol. Evol. 21, 1704–1711.

Rudaitytė-Lukošienė, E., Delgado de las Cuevas, G.E., Prakas, P., Calero-Bernal, R., Martínez-González, M., Strazdaitė-Žielienė, Ž., Servienė, E., Habela, M.A., Butkauskas, D., 2020. *Sarcocystis* spp. diversity in the roe deer (*Capreolus capreolus*) from Lithuania and Spain. Parasitol. Res. 119, 1363–1370.

Seidi, A., Muellner-Wong, L.S., Rajendran, E., Tjhin, E.T., Dagley, L.F., Aw, V.Y.T., Faou, P., Webb, A.I., Tonkin, C.J., van Dooren, G.G., 2018. Elucidating the mitochondrial proteome of *Toxoplasma gondii* reveals the presence of a divergent cytochrome c oxidase. eLife 7, e38131.

Skibinski, D.O.F., Ghiselli, F., Diz, A.P., Milani, L., Mullins, J.G.L., 2018. Structure-related differences between Cytochrome Oxidase I proteins in a stable heteroplasmic mitochondrial system. Genome Biol. Evol. 9, 3265–3281.

Slapeta, J.R., Kyselova, I., Richardson, A.O., Modry, D., Lukes, J., 2002. Phylogeny and sequence variability of the *Sarcocystis singaporensis* Zaman and Colley, (1975) 1976 ssrDNA. J. Parasitol. Res. 88, 810–815.

Slapeta, J.R., Modry, D., Votypka, J., Jirku, M., Lukes, J., Koudela, B., 2003. Evolutionary relationships among cyst-forming coccidia *Sarcocystis* spp. (Alveolata: Apicomplexa: Coccidea) in endemic African tree vipers and perspective for evolution of heteroxenous life cycle. Mol. Phylogenet. Evol. 27, 464–475.

Srinivasan, S., Avadhani, N.G., 2012. Cytochrome c oxidase dysfunction in oxidative stress. Free Radic. Biol. Med. 53, 1252–1263.
Stenger, B.L.S., Clark, M.E., Kvac, M., Khan, E., Giddings, C.W., Dyer, N.W., Schultz, J.L., McEvoy, J.M., 2015. Highly divergent 18S rRNA gene paralogs in a *Cryptosporidium* genotype from eastern chipmunks (*Tamias striatus*). Infect. Genet. Evol. 32, 113–123.
Struebig, M.J., Wilting, A., Metcalfe, K., Kramer-Schadt, S., 2015. Targeted conservation to safeguard a biodiversity hotspot from climate and land-cover change. Curr. Biol. 25, 372–378.
Tajima, F., 1993. Simple methods for testing molecular clock hypothesis. Genetics 135, 599–607.
Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729.
Tome, B., Maia, J.P., Harris, D.J., 2013. Molecular assessment of apicomplexan parasites in the snake *Psammophis* from North Africa: do multiple parasite lineages reflect the final vertebrate host diet? J. Parasitol. 99, 883–887.
Tommaso, P.D., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., Chang, J.-M., Taly, J.-F., Notredame, C., 2011. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. Nucleic Acids Res. 39, W13–W17.
Verma, S.K., Lindsay, D.S., Mowery, J.D., Rosenthal, B.M., Dubey, J.P., 2017. *Sarcocystis pantherophisi* n. sp., from the eastern rat snakes (*Pantherophis alleghaniensis*) as definitive hosts and interferon gamma gene knockout mice as experimental intermediate hosts. J. Parasitol. 102, 547–554.
Wassermann, W., Raisch, L., Lyons, J.A., Natusch, D.J.D., Richter, S., Wirth, M., Preeprem, P., Khoprasert, Y., Ginting, S., Mackenstedt, U., Jäkel, T., 2017. Examination of *Sarcocystis* spp. of giant snakes from Australia and Southeast Asia confirms presence of a known pathogen - *Sarcocystis nesbitti*. PloS One 12, e0187984.
Watthanakaiwan, V., Sukmak, M., Hamarit, K., Kaolim, N., Wajjwalku, W., Munagkram, Y., 2017. Molecular characterization of the ribosomal DNA unit of *Sarcocystis singaporensis, Sarcocystis zamani* and *Sarcocystis zuoi* from rodents in Thailand. J. Vet. Med. Sci. 79, 1412–1418.
Wells, K., Lakim, M.B., Pfeiffer, M., 2006. Nest sites of rodents and treeshrews in Borneo. Ecotropica 12, 141–149.
Wells, K., Kalko, E.K.V., Lakim, M.B., Pfeiffer, M., 2007. Effects of rain forest logging on species richness and assemblage composition of small mammals in Southeast Asia. J. Biogeogr. 34, 1087–1099.
Wells, K., Lakim, M.B., O'Hara, R.B., 2014. Shifts from native to invasive small mammals across gradients from tropical forest to urban habitat in Borneo. Biodivers. Conserv. 23 (9), 2289–2303.
Wells, K., Clark, N.J., 2019. Host specificity in variable environments. Trends Parasitol. 6, 452–465.
Xiang, Z., Rosenthal, B.M., He, Y., Wang, W., Wang, H., Song, J., Shen, P.-Q., Li, M.-L., Yang, Z., 2010. *Sarcocystis tupaia*, sp. nov., a new parasite species employing treeshrews (Tupaiidae, *Tupaia belangeri chinensis*) as natural intermediate hosts. Parasitol. Int. 59, 128–132.
Xie, Q., Lin, J., Qin, Y., Zhou, J., Bu, W., 2011. Structural diversity of eukaryotic 18S rRNA and its impact on alignment and phylogenetic reconstruction. Protein Cell 2, 161–170.