Research Article

Author for correspondence:
Joel L. N. Barratt, E-mail: jbarratt@cdc.gov; joelbarratt43@gmail.com and Sarah G. H. Sapp, E-mail: xyz6@cdc.gov

Machine learning-based analyses support the existence of species complexes for *Strongyloides fuelleborni* and *Strongyloides stercoralis*

Joel L. N. Barratt[1,2] (ID) and Sarah G. H. Sapp[1] (ID)

[1]Division of Parasitic Diseases and Malaria, Centers for Disease Control and Prevention, Parasitic Diseases Branch, Atlanta, USA and [2]Oak Ridge Associated Universities, Oak Ridge, Tennessee, USA

Abstract

Human strongyloidiasis is a serious disease mostly attributable to *Strongyloides stercoralis* and to a lesser extent *Strongyloides fuelleborni*, a parasite mainly of non-human primates. The role of animals as reservoirs of human-infecting *Strongyloides* is ill-defined, and whether dogs are a source of human infection is debated. Published multi-locus sequence typing (MLST) studies attempt to elucidate relationships between *Strongyloides* genotypes, hosts, and distributions, but typically examine relatively few worms, making it difficult to identify population-level trends. Combining MLST data from multiple studies is often impractical because they examine different combinations of loci, eliminating phylogeny as a means of examining these data collectively unless hundreds of specimens are excluded. A recently-described machine learning approach that facilitates clustering of MLST data may offer a solution, even for datasets that include specimens sequenced at different combinations of loci. By clustering various MLST datasets as one using this procedure, we sought to uncover associations among genotype, geography, and hosts that remained elusive when examining datasets individually. Multiple datasets comprising hundreds of *S. stercoralis* and *S. fuelleborni* individuals were combined and clustered. Our results suggest that the commonly proposed 'two lineage' population structure of *S. stercoralis* (where lineage A infects humans and dogs, lineage B only dogs) is an over-simplification. Instead, *S. stercoralis* seemingly represents a species complex, including two distinct populations over-represented in dogs, and other populations vastly more common in humans. A distinction between African and Asian *S. fuelleborni* is also supported here, emphasizing the need for further resolving these taxonomic relationships through modern investigations.

Introduction

Strongyloidiasis affects approximately 370 million people in more than 70 countries, mostly in tropical and sub-tropical regions (Olsen *et al.*, 2009; Bisoffi *et al.*, 2013). *Strongyloides* infections in humans are typically caused by two species; predominantly *Strongyloides stercoralis* Stiles and Hassall 1902, with a smaller proportion caused by *Strongyloides fuelleborni* von Linstow, 1905. Most *Strongyloides* species exhibit high host specificity (Speare, 1989) but *S. stercoralis* infections have been reported in humans, dogs, cats and non-human primates. In particular, the role of dogs as reservoirs of zoonotic *S. stercoralis* has been a matter of contention historically and remains unclear today. Originally thought to infect humans exclusively, a nematode indistinguishable from *S. stercoralis* was identified in a dog in China in the early 1900s (Fuelleborn, 1914), raising questions as to whether *S. stercoralis* was responsible for both human and canine strongyloidiasis. The taxonomic separation of *Strongyloides* infecting humans and dogs was suggested on the basis of experimental and epidemiological evidence (Brumpt, 1922; Augustine, 1940; Jaleta *et al.*, 2017), but this distinction has not been adopted. Even today, the epidemiology of canine strongyloidiasis remains poorly understood, and strongyloidiasis is generally considered a rare disease of dogs outside of East and Southeast Asia and the USA (Kreis, 1932; Galliard, 1951*a*). Reports of *S. stercoralis* or *S. stercoralis*-like species in domestic cats further complicates the question surrounding possible zoonotic reservoirs and *Strongyloides* species diversity (Chandler, 1925; Thamsborg *et al.*, 2017; Wulcan *et al.*, 2019).

Strongyloides stercoralis is currently thought to possess a population structure consisting of lineages A and B (Jaleta *et al.*, 2017; Nagayasu *et al.*, 2017; Barratt *et al.*, 2019*b*). Humans and dogs are considered permissive hosts of *S. stercoralis* lineage A which is genetically characterized by possession of hyper-variable region (HVR)-IV haplotype A in most isolates, with the exception of some recently described Chinese *S. stercoralis* (Zhou *et al.*, 2019). On the other hand, lineage B seemingly infects only dogs and invariably possesses HVR-IV haplotype B. However, fragmentary experimental evidence from early investigators suggests a more complicated population structure than the recently proposed 'two lineage' model implies (Nagayasu *et al.*, 2017).

The less-common zoonotic species, *S. fuelleborni* (subsp. *fuelleborni*), is a non-human primate specialist (Pampiglione and Ricciardi, 1971, 1972; Hira and Patel, 1980; Nutman, 2017;

CAMBRIDGE
UNIVERSITY PRESS

Thanchomnang *et al.*, 2017). Its global population structure has not been extensively studied and it is not known whether the ability to infect humans varies among populations. This variation might contribute to the relative geographic restriction of *S. fuelleborni* (subsp. *fuelleborni*) infections in humans, which are nearly exclusive to sub-Saharan Africa despite the broader occurrence of *S. fuelleborni* (subsp. *fuelleborni*) in primates in other parts of the Old World. The enigmatic *S. fuelleborni* subsp. *kellyi* Viney, Ashford, and Barnish 1991 has only been reported from humans in Papua New Guinea, and its relation to primate *S. fuelleborni* is ambiguous, possibly representing another species altogether (Dorris *et al.*, 2002).

The broad geographic range, complicated taxonomic history, and possible differences in host permissibility raise questions about whether *S. fuelleborni* represents a complex of species with varying degrees of transmissibility to humans (Hung and Höppli, 1923; Sandground, 1925; Augustine, 1940; Premvati, 1959; Little, 1966*a*). Current taxonomic nomenclature does not distinguish between *S. fuelleborni* from Asian and African primates (Ashford and Barnish, 1989; Hasegawa *et al.*, 2010). Like for *S. stercoralis,* modern genetic approaches could greatly aid in reevaluating the diversity, host and geographic associations of this species.

Hasegawa *et al.* (2009) proposed HVRs I to IV of the 18S rDNA as useful markers for *Strongyloides* species diagnosis, especially HVR-IV, based on the observation that 'its nucleotide arrangements are mostly species specific'. HVR-I and HVR-IV are now routinely used in *Strongyloides* genotyping surveys (Barratt *et al.*, 2019*b*; Zhou *et al.*, 2019). The mitochondrial cytochrome c oxidase subunit 1 (*cox1*) locus of *Strongyloides* spp. is hypervariable and has been used to investigate genetic variation within and among *Strongyloides* populations by phylogenetic analysis and/or sequence clustering (Jaleta *et al.*, 2017; Frias *et al.*, 2018; Barratt *et al.*, 2019*b*). When used in various combinations, these three multi-locus sequence typing (MLST) markers (*cox1*, HVR-I and HVR-IV) have demonstrable genotyping utility. However, *Strongyloides* genotyping surveys undertaken to date have used various combinations of these three loci, with some examining only one (Frias *et al.*, 2018), or two loci (Sato *et al.*, 2007; Hasegawa *et al.*, 2009; Thanchomnang *et al.*, 2017). In two recent surveys all three loci were examined (Jaleta *et al.*, 2017; Zhou *et al.*, 2019), and subsequently these three loci were used in a next-generation sequencing approach to attempt to identify all *Strongyloides* genotypes present in a single sample, directly from faecal DNA extracts (Barratt *et al.*, 2019*b*; Beknazarova *et al.*, 2019). These studies, which aimed to elucidate relationships between *Strongyloides* genotypes, hosts and geographic distributions, have led to an abundance of MLST data in public databases, representing hundreds of individual worms.

Individual published MLST studies usually analysed relatively small numbers of worms (often less than 100) from small numbers of hosts and locations, making it difficult to identify clear patterns within these populations. In addition, integrating MLST data from these various studies into a single large dataset is also impractical because the studies sometimes examine overlapping, but different combinations of loci. This means that phylogenetic and other analytic approaches cannot be utilized effectively to explore such datasets in an integrative manner.

A recently described unsupervised machine learning (ML) procedure that calculates a distance statistic from MLST data for downstream clustering, even for datasets composed of different loci and combinations thereof, produced at different timepoints, may offer a solution to this problem. This approach constitutes a novel population genetics tool comprising an ensemble of two ML algorithms (Barratt *et al.*, 2019*a*). As input, this method requires a set of user-defined haplotypes from large population datasets. An advantage of this method over traditional sequence analysis approaches (e.g. phylogeny) is its ability to calculate distances from MLST data, even when the genotype of some specimens in the dataset is not completely defined (e.g. in the absence of data for a particular marker). Such an approach could facilitate the integration of MLST data for *Strongyloides* from multiple studies into a single analysis, even though the marker combinations vary among studies. Additionally, this method has an advantage that it can address the challenge of heterozygosity that might be encountered at nuclear loci in sexually reproducing eukaryotes (Barratt *et al.*, 2019*a*). Thus, data for heterozygous individuals could also be retained for analysis.

In the present study, we applied this ML method to all publicly available MLST datasets for *S. stercoralis* and *S. fuelleborni*. We integrated these data to increase our ability to detect novel population-level associations. We hypothesized that the use of this method would enable the identification of relationships among *Strongyloides* genotypes, hosts and geographic distributions, that were not evident using smaller datasets individually. We propose that this approach could aid in resolving taxonomic questions and controversies surrounding *Strongyloides* species.

Materials and methods

Data selection

All available sequences of *S. stercoralis* and *S. fuelleborni* 18S HVR-I, 18S HVR-IV ($n = 218$) and mitochondrial *cox1* ($n = 789$) were obtained from GenBank (accessed August of 2019). *Strongyloides* sp. *cox1* sequences obtained from Bornean slow lorises (*Nycticebus borneanus*, $n = 18$) were also included for analysis, as they share a relatively close phylogenetic relationship with *S. stercoralis* and *S. fuelleborni* despite representing a distinct group (Frias *et al.*, 2018). In cases where the HVR-I, HVR-IV and/or *cox1* sequences were available from the same individual worm, the complete or partial genotype of these worms was recorded (Supplementary File S1, Tabs B and E, column F) after the genotyping approach by Barratt *et al.* (2019*b*) and Jaleta *et al.* (2017). In instances where specimens were genotyped using polymerase chain reaction (PCR) products amplified directly from stool (Barratt *et al.*, 2019*b*; Beknazarova *et al.*, 2019), only specimens obtaining a single haplotype for each locus were included in this analysis. The rationale for this criterion was that for amplicons sequenced directly from stool where multiple haplotypes were detected, the underlying genotype of individual worms cannot be elucidated. Sequences in GenBank possessing International Union of Pure and Applied Chemistry (IUPAC) ambiguity codes were unphased into their underlying haplotypes only if a sequence supporting the existence of both possible haplotypes was available in GenBank. If only one of two possible haplotypes were supported by other haplotypes in GenBank for such a sequence, the IUPAC code was changed in favour of the supported haplotype. If there was no support for either of the underlying haplotypes for a sequence in GenBank possessing IUPAC codes, this part of the sequence was trimmed off from the ambiguous base onwards, in the direction that maximized the final length of the sequence. This analysis was inclusive of all *S. stercoralis* and *S. fuelleborni* sequences available in GenBank that met these criteria, irrespective of their host origin, and included specimens from humans, non-human primates, dogs and cats (Supplementary File S1, Tabs B and E, column D). Due to the limited sequence data available for *S. fuelleborni kellyi*, it was excluded from this analysis.

Genotypes were also constructed for five *S. fuelleborni* specimens from Japanese macaques, using all three MLST markers (i.e. HVR-I, HVR-IV and *cox1*). These genotypes were constructed from five *cox1* sequences generated in one study

(Hasegawa *et al.*, 2010), and 18S rDNA sequences published in different studies presumably from different worms (Sato *et al.*, 2007; Hasegawa *et al.*, 2009). Therefore, 'synthetic' genotypes were constructed from these sequences to represent *S. fuelleborni* from the five Japanese macaques (*Macaca fuscata*). The synthetic genotypes were constructed in this way based on the observation that each of these 18S rDNA sequences was generated from worms collected from different Japanese macaque populations, yet all sequences are identical (GB: AB272235.1, AB453317.1, AB453318.1, AB453319.1). This supports that they represent the haplotypes found in all Japanese macaques (Sato *et al.*, 2007; Hasegawa *et al.*, 2009). However, because these genotypes were based on human inference, classification was performed twice on the *S. fuelleborni* dataset; once including and once excluding these five 'synthetic' Japanese *S. fuelleborni* specimens.

Extracting cox1 and 18s haplotypes from published Strongyloides genomes

Genome sequences of *S. stercoralis* (Kikuchi *et al.*, 2016) (36 genomes in total) were downloaded from the GenBank SRA database (Supplementary File S1, Tab B, column E) and the raw Illumina reads were subjected to a workflow designed in Geneious Prime (version 11: www.geneious.com). This workflow performed removal of Illumina adapter sequences (using BBDuk – default parameters) and filtering for quality (minimum PHRED score: 20 and minimum read length: 50 bases). Trimmed reads were mapped to references of HVR-I (GenBank: AF279916.2), HVR-IV (GenBank: AF279916.2), and *cox1* (GenBank: MK463927.1) using Geneious mapper (default parameters) and the consensus of each mapped assembly was extracted for inclusion in this analysis.

Haplotype definitions

Haplotypes were identified in each specimen using a recently described genotyping system (Jaleta *et al.*, 2017; Barratt *et al.*, 2019b) with some modifications (Fig. 1). Briefly, for the purposes of this study, sequences of the ~434 bp fragment of HVR-I were divided into four segments, sequences of the ~260 bp fragment of HVR-IV were divided into three segments, and a 217 bp region of *cox1* defined elsewhere (Beknazarova *et al.*, 2019; Barratt *et al.*, 2019b) was divided into nine segments (A1 to A3, B1 to B3 and C1 to C3), so that each of these sub-segments was considered a distinct locus. The rationale for dividing the sequences into these segments is based on extensive testing of the ML approach for performance optimization. These tests indicated that performance is best when the number of haplotypes at a given locus is between 10 and 20, and preferably less than 30 (https://github.com/Joel-Barratt/Eukaryotyping). The reasons for this are explained in detail in Supplementary File S2, pages 4–7. The sequence of each haplotype after the division of the three MLST markers into segments (resulting in a final panel of 16 markers) is provided in Supplementary File S2, Appendix (part A).

Assessment of population structure

Genotypes were assigned to each specimen using a Geneious workflow developed by the study team. Firstly, the *Strongyloides* sequences from individual worms were merged into a sequence list and then compared to a BLASTN database containing the FASTA sequences provided in Supplementary File S2, Appendix part A. The results were exported from Geneious in text format (one file for each specimen) and this result was converted to the format shown in Supplementary file S1 (Tabs A and D). The ML procedure (Barratt *et al.*, 2019a) was applied to the

resultant *Strongyloides* dataset using the scripts and instructions available here: https://github.com/Joel-Barratt/Eukaryotyping. Briefly, the haplotype data sheets provided in Supplementary file S1 (Tabs A and D) were exported as .txt files; these files were used directly as the input for the R scripts. This input included 138 specimens from *S. fuelleborni* and *Strongyloides* sp. 'loris', and 764 from *S. stercoralis*.

The ML procedure performs an unsupervised similarity-based classification task; it assesses whether any two specimens are related or unrelated on the basis of their genotype. The algorithms underpinning this method do not require that the genotype of every specimen be defined in the same manner; specimens are not required to be sequenced at the same markers (Barratt *et al.*, 2019a), although, realistic minimum data requirements must be set by the user prior to analysis (https://github.com/Joel-Barratt/Eukaryotyping). For example, it would be unrealistic to expect that a specimen would be accurately classified in the event that only 20 bases of a single locus were available for this specimen. In this study, specimens were only retained for classification if they met at least one of two minimum data availability criteria. Firstly (1), when *cox1* was the only sequence available for a specimen, or when data was not obtained for Part B of HVR-IV, the availability of nine out of the nine *cox1* segments was required for a specimen to be included in the analysis. Secondly (2), if the only *cox1* sequence available for a specimen was truncated (i.e. partially overlapping with the 217 bp region of *cox1* being analysed here), the specimen could still be analysed provided its *cox1* sequence overlapped with this 217 bp region by a minimum of seven out of nine segments. However, specimens with a truncated *cox1* sequence were only retained in this analysis if part B of HVR-IV was also available for that specimen. This requirement considers the pertinent observations of Hasegawa *et al.* (2009), regarding HVR-IV, indicating that 'its nucleotide arrangements are mostly species specific'. It also considers that Part B of HVR-I can differentiate lineages A and B of *S. stercoralis* (Nagayasu *et al.*, 2017; Barratt *et al.*, 2019b). Therefore, part B of HVR-IV was set as a supplementary data requirement to compensate for information lost in the event of a truncated *cox1* sequence. Data for HVR-I and/or additional parts of HVR-IV were considered in the analysis if available for a specimen, but data for these loci were not an absolute requirement for retaining specimens in this analysis.

The classification was performed for *S. stercoralis* and *S. fuelleborni* separately using an epsilon value of 0.05 for Plucinski's naïve Bayes classifier (refer to: https://github.com/Joel-Barratt/Eukaryotyping), and the resultant pairwise distance matrices (a pairwise matrix is the standard output of our ML method – Supplementary File S1, Tabs C and F) were clustered. Clustering was performed using the agglomerative nested approach in the 'agnes' R package, utilizing Ward's clustering method as described here (Barratt *et al.*, 2019b). The 'ggtree' R package was then used to generate cluster dendrograms. Images of relevant hosts were obtained from PhyloPic (http://phylopic.org) or prepared in-house at the Centers for Disease Control and Prevention (CDC) for annotation of dendrograms. Images were rendered using the freely available GNU Image manipulation program (https://www.gimp.org).

Rationale for study design

Strongyloides cox1 sequences are considered hypervariable; this variability relates to single nucleotide polymorphisms (SNPs). As *cox1* is also encoded on the mitochondrion, it is not subject to heterozygosity. These features make *cox1* conducive to an analysis by phylogenetic methods or clustering based on sequence similarity (Jaleta *et al.*, 2017; Barratt *et al.*, 2019b; Beknazarova
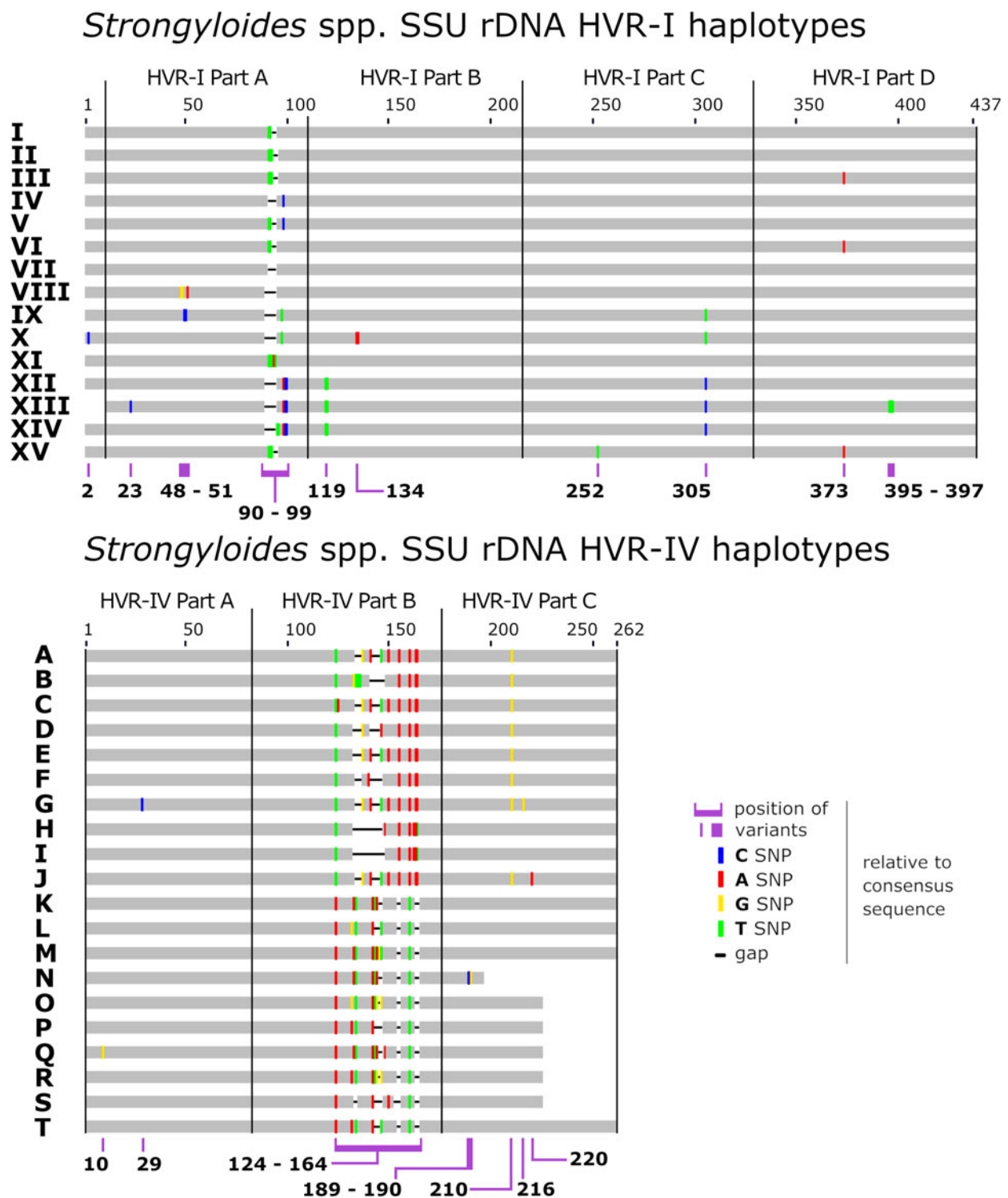
Fig. 1. Schematic of the *Strongyloides* genotyping scheme referenced here. Graphical representation a *Strongyloides* sp. genotyping scheme after the description of Barratt *et al.* (2019*b*) and Jaleta *et al.* (2017). This scheme was expanded here to include haplotype XV of 18S HVR-I (GenBank Accession: MT436714), identified in genomes sequenced from Japanese humans by Kikuchi *et al.* (2016). For the purposes of this study, note that the haplotypes originally defined after the description of Barratt and colleagues were split into smaller segments as shown here.

et al., 2019; Zhou *et al.*, 2019). Alternatively, the 18S rDNA locus may be heterozygous in some individuals of *S. stercoralis* (see Zhou *et al.*, 2019), which is an important confounding factor for phylogenetic methods. Within the *Strongyloides* genus, the 18S rDNA locus is less diverse than *cox1* but it contains nucleotide insertions and deletions (indels) that represent important variants for differentiating *Strongyloides* species and genotypes. Most phylogenetic algorithms typically treat indels as missing data (Truszkowski and Goldman, 2016; Donath and Stadler,

2018), and variability in indel handling between various multiple sequence aligners prior to phylogeny can produce markedly different results (Golubchik *et al.*, 2007; Ashkenazy *et al.*, 2014). For this reason, algorithms such as Gblocks are often used to identify and remove regions of the alignment containing gaps, making this data more amenable to phylogenetic applications (Castresana, 2000). Unfortunately, this practice would result in a loss of meaningful information when analyzing *Strongyloides* 18S rDNA HVRs. Furthermore, *Strongyloides* genotyping surveys

published in the last decade display only some consistency in the combinations of the three widely used genotyping loci examined (Hasegawa *et al.*, 2009, 2010; Schär *et al.*, 2014; Laymanivong *et al.*, 2016; Jaleta *et al.*, 2017; Thanchomnang *et al.*, 2017, 2019; Frias *et al.*, 2018; Barratt *et al.*, 2019b; Beknazarova *et al.*, 2019). This makes comparative analyses of distinct datasets difficult when using phylogenetic and/or sequence clustering methods. These limitations represent the main impetus for the design of this study, specifically; the ML procedure that was designed to overcome these challenges (Barratt *et al.*, 2019a) (Supplementary File S2).

Results

Data filtering and detection of novel 18s rDNA haplotypes

After filtering the combined MLST dataset according to the minimum data availability requirements, data from 704 of the original 764 genotyped *S. stercoralis* specimens were retained. For *S. fuelleborni*, data from 133 of the 138 genotyped specimens were retained including all sequences from *Strongyloides* sp. 'Loris'. These specimens had varying combinations of HVR-I, HVR-IV and *cox1* data available; the markers available for each individual specimen that was retained for analysis are provided in Supplementary File S1, Tabs B and E, column F.

A novel *S. stercoralis* 18S rDNA HVR-I haplotype was identified in previously published genome datasets (Kikuchi *et al.*, 2016). This 18S rDNA haplotype was detected in specimens from Okinawa, Japan and was assigned to haplotype XV, which contains a unique SNP at position 252 relative to the alignment shown in Fig. 1. The sequence of haplotype XV is identical to that of HVR-I haplotype III at all other base positions. The *Strongyloides* sp. genotyping scheme previously described (Jaleta *et al.*, 2017; Barratt *et al.*, 2019b; Beknazarova *et al.*, 2019), was thus expanded to include this novel type (GenBank Accession: MT436714). The typing scheme used here (Barratt *et al.*, 2019b; Beknazarova *et al.*, 2019), differs from that described by Zhou *et al.* (2019), who described HVR-IV haplotype C, which is identical to haplotype E reported by Beknazarova *et al.* (2019) (Fig. 1).

The population structure of Strongyloides stercoralis

The final filtered *S. stercoralis* dataset includes sequences from specimens representing all inhabited continents. However, *S. stercoralis* sequences from specimens collected in Japan and some Southeast Asian countries were more highly represented (Supplementary File S2, page 13). MLST data from *S. stercoralis* collected from humans were more highly represented than data collected from dogs (Table 1, Fig. 2). Following classification using ML, the dataset was divided into 7 distinct genetic clusters, each representing a proposed sub-population of *S. stercoralis* (Fig. 2). The vast majority of *S. stercoralis* from dogs were assigned to one of two populations based on the MLST data available, represented by clusters 4 and 5 (129 of 146 *S. stercoralis* from dogs were assigned to these clusters; 88.4%). Most *S. stercoralis* specimens of human origin (550/554, 99.3%) were assigned to clusters 1, 2, 3, 6 or 7 based on their genotype (Table 1). All specimens assigned to cluster 5 (92/92) were derived from dogs as were 86% of specimens (37/43) assigned to cluster 4. Two specimens assigned to cluster 4 were obtained from cats and only 4 specimens assigned to this cluster were from humans (90.7% of cluster 4 included *S. stercoralis* collected from dogs or cats). Overall, 95.5% of *S. stercoralis* assigned to clusters 4 and 5 based on their MLST genotype had been isolated from dogs. Given the geographical sampling bias towards specimens collected in Southeast Asia and Japan (Supplementary File S2, pages 13 and 14), no

Table 1. Frequency of *S. stercoralis* from dogs and humans assigned to each genetic cluster and their respective χ2 *P* values

	Humans	Dogs	P-value	Over-represented host in each cluster
Cluster_1	138	2	0.000	**Humans**
Cluster_2	56	1	0.000	**Humans**
Cluster_3	236	3	0.000	**Humans**
Cluster_4	4	37	0.000	**Dogs**
Cluster_5	0	92	0.000	**Dogs**
Cluster_6	88	2	0.000	**Humans**
Cluster_7	32	9	0.000	**Humans**
TOTAL	554	146		

Note: *Strongyloides stercoralis* from humans were rarely assigned to cluster 4 and never to cluster 5. *Strongyloides stercoralis* from dogs, were vastly more common in clusters 4 and 5 (88.4% of cases), though 6.2% *S. stercoralis* from dogs were assigned to cluster 7. Boxes containing numeric values are shaded according to their frequency, with the highest frequencies shown in black and the lowest frequency (zero) in white.

attempt was made to statistically validate possible associations between genetic cluster assignment and the geographic origin of *S. stercoralis* specimens. An exception to this is genetic cluster 5, where 90 of 92 specimens assigned to this cluster (98%) were collected from dogs in Cambodia or Myanmar.

The population structure of Strongyloides fuelleborni

The *S. fuelleborni* dataset retained for classification includes specimens collected mostly from parts of Southeast Asia and Africa, one specimen from India, and five specimens from Japan. Overall, the study population included 9 specimens from humans and 124 specimens from non-human primates. Long-tailed macaques (*Macaca fascicularis*) were the most highly sampled host among the specimens included in this analysis. The *S. fuelleborni* dataset also included 18 specimens obtained from Bornean slow lorises (*Nycticebus borneanus*). For precise numbers of specimens retained for analysis from each host and country refer to Supplementary File S2, Tables S4 to S7. The genotypes constructed for the five *S. fuelleborni* specimens from Japanese macaques possess all markers (i.e. HVR-I, HVR-IV and *cox1*) but these genotypes were generated from five *cox1* sequences from one study (Hasegawa *et al.*, 2010), and 18S rDNA sequences obtained in different studies from different worms. Regarding these 'synthetic' genotypes from *S. fuelleborni* infecting Japanese macaques, classification was performed twice on the *S. fuelleborni* dataset; once including and once excluding these five 'synthetic' specimens. Irrespectively, the presence or absence of these specimens during classification had very little impact on the *S. fuelleborni* population structure (Supplementary File S2, Appendix part D).

Following classification, the *S. fuelleborni* specimens (plus *Strongyloides* sp. from slow lorises) were divided amongst 7 clusters, each representing a distinct population (Fig. 3). Clusters 1 and 2 as well as clusters 6 and 7, showed no clear association with a particular primate host species. Cluster 5 is exclusively occupied by the *Strongyloides* sp. from lorises, though the distinctness of this group was previously reported (Frias *et al.*, 2018). Cluster 4 is occupied by only a small number of inferred genotypes from one host species (Japanese macaques), and the specimens obtained from long-tailed macaques assigned to cluster 3 come from a single study surveying only this primate species in Thailand and Laos (Thanchomnang *et al.*, 2017). Therefore, the possibility that this is a geographic association as opposed to a host association cannot be excluded. Generally, the data support
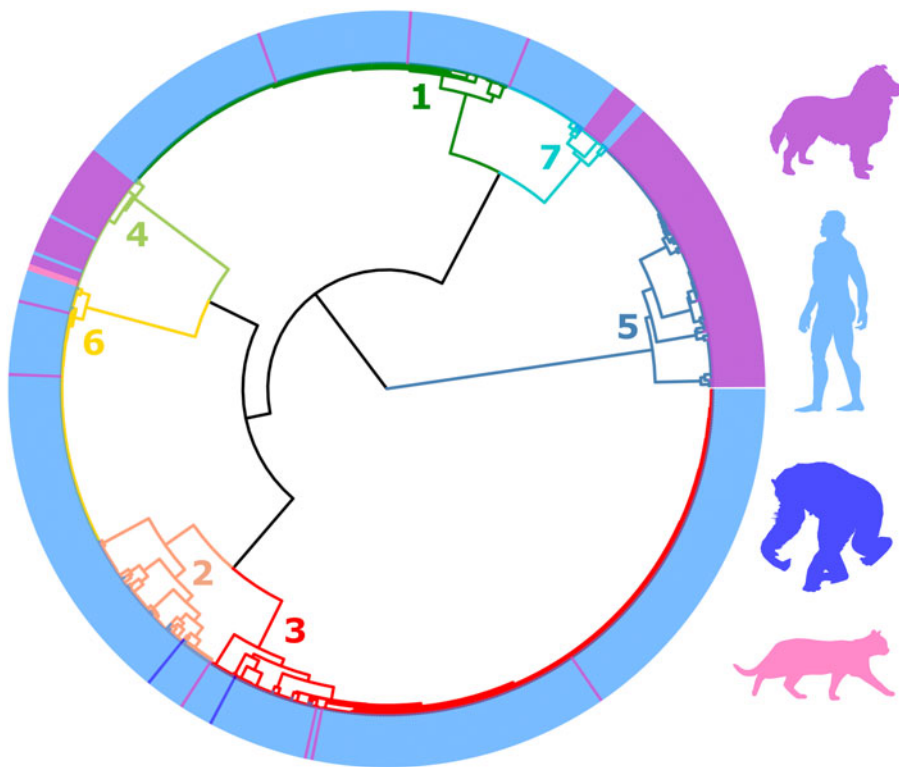
Fig. 2. Dendrogram of clustered distances generated from genotyped *Strongyloides stercoralis* using ML. This dendrogram was divided into seven distinct clusters. Branches are coloured according to their cluster number and are simply used to differentiate branches. Coloured peripheral bars indicate the host species from which the genotyped *S. stercoralis* were collected from; either a dog, human, chimpanzee or cat. The colour of these respective silhouettes matches the peripheral bars. For the identity of each specimen in this dendrogram refer to Supplementary File S2, Appendix part B. Cluster 5 is exclusively occupied by *S. stercoralis* infecting dogs. Specimens assigned to Cluster 4 were more frequently collected from dogs, though were also found in humans and cats. *S. stercoralis* from chimpanzees were assigned to clusters 2 and 3, though these were less frequently observed in these samples. *S. stercoralis* from clusters 1, 2, 3, 6 and 7 were more frequently collected from human samples than other hosts. The geographic origin of sequenced specimens included in this dendrogram is shown in Supplementary File S1, Tab B.

a population structure for *S. fuelleborni* based on associations among genotype and geography; clusters 1 and 2 including *S. fuelleborni* from Africa, clusters 3 and 4 including *S. fuelleborni* from Southeast Asia and Japan (respectively), and clusters 6 and 7 including *S. fuelleborni* from Malaysian Borneo, which is suggestive of allopatric speciation.

Discussion

The ability to integrate large sequence datasets from multiple studies composed of specimens with varying combinations of MLST markers amplified, and specimens that are heterozygous, represents an important advantage of the ML procedure employed here over traditional methods utilized in the field of population genetics. Furthermore, the way in which haplotypes are defined using our ML procedure avoids problems associated with the treatment of indels that are inherent in many phylogenetic methods; indels (alignment gaps) are often treated as missing data in phylogenetic analyses and are excluded altogether (Truszkowski and Goldman, 2016; Donath and Stadler, 2018). This exclusion of alignment gaps is problematic for the analysis of *Strongyloides* HVRs because these loci contain informative indels that differentiate haplotypes. Indels do not represent a problem for the ML approach used here which represents another advantage of this approach over phylogenetic methods. This method facilitated the inclusion of hundreds of genotyped *S. stercoralis* specimens from multiple studies in a single analysis, despite that many published genotyping studies examined different, yet overlapping, combinations of markers. Therefore, this analysis provides a broader view of potential host-associations and geographic patterns that exist in populations of these worms than could be provided in previous studies given the data analysis challenges mentioned above. The large number of specimens retained for analysis here meant that any observable trends are supported with greater statistical power than could be achieved within the smaller genotyping studies. Consequently, this ML procedure represents a powerful alternative to traditional phylogenetic

methods. Using this approach, we provide evidence supporting that *S. stercoralis* represents a species complex and that African and s.e. Asian *S. fuelleborni* are distinct. These results are supported by a synthesis of scientific literature, focusing on observed patterns relating to host preference and geography reported on the basis of biological and experimental evidence published by pioneering investigators who were predominantly active in the early to middle 20th century. This synthesis highlights the agreement between our current analysis and the observations of these classic parasitologists who lacked the molecular tools required to help resolve these important taxonomic questions.

Since Fuelleborn's initial discovery of a *Strongyloides* infection in a Chinese dog, the status of dog-derived isolates as a valid species or as a variant of human *S. stercoralis,* and thus its zoonotic potential, has remained a subject of debate (Fuelleborn, 1914). While many authors regarded canine and human isolates as more or less morphologically identical strains with differing host specificity within the same species *S. stercoralis* (Hung and Höppli, 1923; Kreis, 1932), others recognized the potential existence of 'geographic races' or even suggested the separation of canine *Strongyloides* on experimental and epidemiological grounds in spite of negligible morphological differences (Brumpt, 1922; Augustine and Davey, 1939). Brumpt (1922) argued for the establishment of a new species *Strongyloides canis* on the basis of geographic disparities in prevalence, barriers in experimental cross-transmission, and reported developmental differences *in vitro*. As most early 20th century authors were averse to applying for specific status without important morphological differences (Chandler, 1925; Sandground, 1925; Goodey, 1926; Kreis, 1932), *S. canis* was never recognized as a valid name. However, the modern availability of genetic data – including our results – and the resulting shift in taxonomic dogmas allows this question to be revisited.

The body of experimental evidence demonstrates varying abilities of human-derived *S. stercoralis* from different geographic regions to establish infections in dogs. This could correspond with the 'spectrum' between human- and apparently
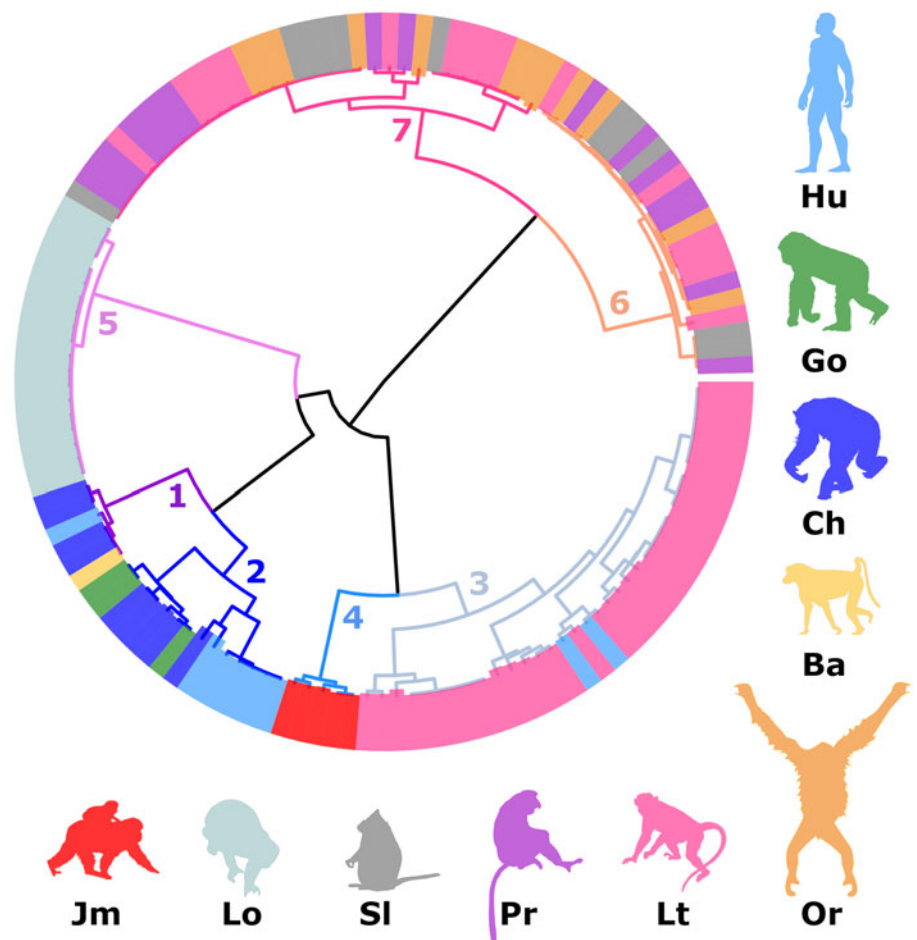
Fig. 3. Dendrogram of clustered distances generated from genotyped specimens of *Strongyloides fuelleborni* and *Strongyloides* sp. from lorises using ML. This dendrogram was divided into seven distinct clusters. Branches are coloured according to their cluster number and the coloured peripheral bars indicate the host species from which the *S. stercoralis* specimens/ sequences were derived, either a human (Hu), gorilla (Go), chimpanzee (Ch), baboon (Ba), orangutan (Or), Long-tailed macaque (Lt), proboscis monkey (Pr), Silvered leaf monkey (Sl), Slow loris (Lo), or Japanese macaque (Jm). Black branches visualize the relationship between clusters. For the identity of each specimen in this dendrogram refer to Supplementary File S1, Appendix part C. Clusters 1 & 2 include specimens from Africa. Cluster 3 includes specimens from Thailand and Laos, and a single specimen from an Indian human. Cluster 4 includes specimens collected in Japan. Clusters 5, 6 and 7 include specimens from Malaysian Borneo. This dendrogram supports that clusters 1 and 2 include *S. fuelleborni* genotypes that infect humans more commonly than the *Strongyloides* in other clusters.

canine-restricted genetic lineages presented here (Fig. 2). In Fuelleborn's investigations (Fuelleborn, 1914, 1927), dogs developed only short duration infections (2–3 weeks) with human-derived Chinese *S. stercoralis* and were refractory to infection with East African *S. stercoralis.* Sandground successfully infected dogs with a USA human-derived strain, though the majority of dogs eventually self-cured after about 3–10 weeks and were refractory towards repeated exposures (Sandground, 1928; Galliard, 1939, 1951b). Galliard reported ease in infecting dogs with human *Strongyloides* from Vietnam (Galliard, 1939), but that dogs were more or less refractory to isolates from the West Indies and North Africa (Galliard, 1951a). Also, in his experiments, dogs that were not successfully infected with the African strain developed severe infections after exposure to the Vietnamese strain.

Sandosham (1952) reported short-term, unstable infection in a dog repeatedly inoculated with many larvae derived from former prisoners of war in Thailand, concluding that dogs could not have been the reservoir host for *S. stercoralis* in that setting despite their supposed abundant presence in the prison camps (Sandosham, 1952). Many years later, attempts to infect laboratory dogs with *S. stercoralis* derived from a human patient infected in Southeast Asia yielded mixed results; in one dog, a chronic infection lasting at least 15 months was established, but in four others, fecal larval counts peaked at 3 weeks and dropped off drastically after that (Grove and Northern, 1982). A similar pattern of unstable, transient patent infections has been noted previously in human volunteers infected with other animal *Strongyloides* species, such as *Strongyloides procyonis* (raccoons) and *Strongyloides ransomi* (swine), further suggesting that host adaptation strongly influences the duration of infection and larval output (Freedman, 1991). Reciprocally, Augustine and Davey (1939) failed to infect a human volunteer, guinea pigs and cats

with filariform larvae reared from a naturally-infected dog in Massachusetts but readily infected both young and aged dogs, which developed long-lasting patent infections.

Clearly, historical workers had differing levels of success infecting dogs, and one varying factor was the geographic origin of isolates. Though fragmentary, it is interesting to compare the geographic origins of parasites used in dog infection trials to the emerging picture of the *S. stercoralis* global population structure shown here. For example, dogs were found to be refractory or developing very short-lived infections when exposed to human-derived isolates of *S. stercoralis* from China, East Africa, North Africa and the West Indies (Fuelleborn, 1914; Galliard, 1950; Galliard and Berdonneau, 1953). Notably, the lineages of *S. stercoralis*, represented by clusters 1, 2, 3 and 6 were almost all obtained from human samples and known to occur in these locations (Fig. 2). Intermediate results (e.g. short-lived infections, low larval outputs, and/or resistance to reinfection in most dogs, with occasional chronic infections) were seen with North American and some Southeast Asian *S. stercoralis* strains (Sandground, 1925; Faust, 1933; Grove and Northern, 1982; Genta, 1989); these locations are represented by lineages of *S. stercoralis* assigned to genetic clusters 4 and 7, where we observe genotypes that are seemingly capable of infecting both hosts, yet with the majority of specimens assigned to these clusters collected from dogs and humans, respectively (Fig. 2, Table 1). The sole trial where dogs were readily infected with no immunity to reinfection used a human-derived *S. stercoralis* from Vietnam (Galliard, 1939), which is in close geographic proximity to the origin of roughly half of all specimens assigned to cluster 7 (*n* = 20, Myanmar and Thailand). In summary, the observed historical variations show some overlap with both the locations and host frequencies in our observed *S. stercoralis* clusters.

Our analysis suggests a disparity between where human and canine *S. stercoralis* infections are occurring; infections caused by *S. stercoralis* assigned to cluster 5 are almost exclusive to Southeast Asia (98%). Our analysis also indicates that different genotypes of *S. stercoralis* are more frequently reported in specific hosts (dogs are over-represented hosts in clusters 4 and 5), based on a χ^2 analysis (Table 1). These observations were also extensively discussed by historical authors who commented on the marked disparity in the occurrence of canine and human *S. stercoralis* infections i.e. in places where dog *S. stercoralis* infections occur commonly, human infections are rare, and vice versa. Outside of East and Southeast Asia, natural infections in dogs appeared to only occur in North America, and reports of dog infections were exceedingly rare to absent in areas where prevalence was high in humans (Brumpt, 1922; Faust, 1933; Augustine, 1940; Galliard, 1950; Sandosham, 1952; Genta and Grove, 1989). For example, not a single infection was detected in 528 dog faecal specimens during a survey in a highly endemic region of Colombia, where at least 14% of humans were *S. stercoralis*-positive (Faust and Giraldo, 1960); even if some infections were missed due to the use of flotation techniques, infection would still be very rare in dogs *vs* humans. Recent case reports and molecular evidence show that natural canine infections also occur in Europe and Australia (Paradies *et al.*, 2017; Basso *et al.*, 2019; Beknazarova *et al.*, 2019; Barratt *et al.*, 2019b). However, an important caveat is that results based solely on microscopy of feces must be interpreted with caution as free-living nematode larvae, common contaminants of specimens collected from the ground can easily be confused with *Strongyloides* spp. (Speare, 1989). Direct epidemiologic evidence for cross-transmissibility is limited, in part due to the apparent variability in host adaptation and geographic occurrence of these strains. In one instance, a kennel worker in the USA with no foreign travel acquired *Strongyloides* from dogs under his care; this strain was successfully used to infect laboratory dogs (Georgi and Sprinkle, 1974). A small study in an area of Japan considered endemic found no association in *S. stercoralis* infection status among owners and dogs (Takano *et al.*, 2009). Overall, there are still many knowledge gaps in our understanding of canine strongyloidiasis as a zoonosis and its epidemiology, and interpretation will require further characterization of strains across a diverse geographic range.

While establishing the specific identity of canine *Strongyloides* will require further investigation, based on historical reports and our novel analysis, it appears that different *S. stercoralis* lineages display differences in infection frequency for canine and human hosts – which may exclude or drastically limit cross-transmission potential in some cases (Table 1). Given the epidemiological implications of this genetic variability and the now large body of evidence across many independent studies, it now seems appropriate to treat *S. stercoralis* as a species complex. For example, '*S. stercoralis* sensu lato' could be useful for referring to all isolates, and '*S. stercoralis* sensu stricto' could be considered for the human-genotypes represented in 6 of the 7 clusters associated with HVR-IV haplotype A (Fig. 2, excluding cluster 5). This would also include specimens assigned to genetic cluster 4 which seems associated with the 18S rDNA genotype VI + A, and likely represents a canine adapted lineage of *S. stercoralis* sensu stricto. Subspecific ranks (e.g. *S. stercoralis canis*) were suggested nearly a century ago by Chandler for *S. stercoralis* of different animal hosts (Chandler, 1925). This seems reasonable, given the exclusive finding in canine hosts (100% of cluster 5), the restricted geographic range of *S. stercoralis* assigned to cluster 5 (98% were from Southeast Asia), and the fact this lineage (associated with HVR-IV haplotype B) naturally clusters as an outgroup to all other *S. stercoralis* (Fig. 2).

An important outstanding question is whether the canine infections in certain human host-dominated populations represent transient, spurious, or incidental infections, or if there exists a substructure that is not clearly apparent due to under-sampling of dog-derived genotypes from specific locations. Based on our analysis, dog-derived genotypes assigned to clusters 1–3 and 6 likely represent transient infections with strains originating from humans; this is supported by the experiments of Fuelleborn, Sanground and Galliard who infected dogs with human-derived *S. stercoralis* which often resulted in transient infections (Fuelleborn, 1914, 1927; Sandground, 1928; Galliard, 1939, 1951a, 1951b). However, for cluster 7 it seems possible that some additional population substructure exists. Cluster 7 contains a small 'sub-cluster' composed of eight *cox1* sequences derived exclusively from *S. stercoralis* from Japanese dogs. Sequencing of HVR-I and/or HVR-IV from these worms, and the sampling of worms from additional dogs in the same geographic area that these specimens were collected, would provide greater clarity as to what this smaller 'sub-cluster' of Cluster 7 might represent.

The identity of *Strongyloides* spp. infecting cats is a more complex question, and it appears multiple species are involved, including *S. stercoralis*. Chandler first reported *S. stercoralis* in cats in India (Chandler, 1925), and some attempts to infect cats with *S. stercoralis* from dogs and humans have been successful; though infections were usually short-lived, and it appears that cats are competent but abnormal hosts for *S. stercoralis* (Sandground, 1928; Desportes, 1944; Wulcan *et al.*, 2019). *Strongyloides felis* has been described from India and Australia, and is somewhat similar morphologically to *S. stercoralis* although no molecular data are available (Speare and Tinsley, 1986). *Strongyloides planiceps* occasionally infects cats but is dissimilar in its life cycle and morphologic characteristics to *S. stercoralis* (Thamsborg *et al.*, 2017). A fourth species, *Strongyloides tumefaciens,* is of interest as it has been reported as causing colonic nodules, an unusual clinical presentation for *Strongyloides* (Thamsborg *et al.*, 2017). However, the original description was incomplete, and a recent study cast doubt on the validity of *S. tumefaciens*. *Strongyloides* sp. extracted from nodules of necropsied cats were found to have *cox1* sequences that matched *S. stercoralis*. These sequences were assigned to genetic cluster 4 in this study (Fig. 2), suggesting that *S. tumefaciens* infections could simply be an unusual pathological presentation of *S. stercoralis* (Wulcan *et al.*, 2019). There is clearly a need for further investigation into *Strongyloides* spp. of domestic cats, especially in regard to their zoonotic potential. The analytic approach presented here, along with morphological and biological characterization, should prove beneficial in answering these questions.

The other topic of interest in our study was the identity and population structure of *S. fuelleborni* as it relates to human and other primate infections. The validity and number of species infecting primates have been long debated. Genetic analyses, including the approach used here, have proven valuable in reconciling some of these viewpoints, as morphological comparisons have confounded historical investigations of *Strongyloides* taxonomy. This challenge mainly owes to the variability of characters and their interpretation – importantly, how much relative weight was given to which characteristics for species designations. For example, both Chandler and Little felt that morphologic features of the parasitic female were most reliable (Chandler, 1925; Little, 1966a), whereas Looss and Goodey placed more importance on the free-living adult stages (Looss, 1911; Goodey, 1926). These discrepancies, along with the broad geographic and host range from which *Strongyloides* specimens were derived in these investigations, lead to differing conclusions and opinions regarding the number of species from primates.

Formerly, two species names were used for *Strongyloides* derived from Old World primates – *S. fuelleborni* Von Linstow, 1905, described from *Pan troglodytes* and *Papio cynocephalus* from Africa, and *S. simiae* Hung and Höppli, 1923 from Asian *Macaca* sp. The latter species was established on the basis of a smaller oesophagus to total body length ratio in comparison to *S. fuelleborni*, a lack of prominent constriction behind the vulva of the free-living female, and a 'finely-striated cuticle' (the authors considered *S. fuelleborni* to have a smooth cuticle) (Hung and Höppli, 1923). However, other authors report that *S. fuelleborni* and many, if not all other *Strongyloides* spp., indeed has a striated cuticle though sometimes very difficult to observe (Sandground, 1925; Grove, 1989). As such, the validity of *S. simiae* has been scrutinized by several authors, most of whom regarded the morphological evidence for specific status insufficient or too highly variable for unequivocal species discrimination (Sandground, 1925; Premvati, 1959; Little, 1966a). *Staphylococcus simiae* became a junior synonym of *S. fuelleborni,* which is the name currently applied to all egg-passing *Strongyloides* from Old World apes and monkeys (Grove, 1989).

Some biological evidence suggests there may still be valid differences between '*S. simiae*' from *Macaca* spp. and *S. fuelleborni.* Experimental work by Augustine revealed that crosses between strains derived from *Macaca* and strains derived from *Pan* failed to produce offspring – suggesting perhaps that *S. simiae* is indeed a separate species, but nearly impossible to distinguish morphologically from *S. fuelleborni* (Augustine, 1940). It was also observed that *Cebus*-derived strains (later designated *S. cebus* – a species currently considered valid) also failed to cross with the other primate strains, supporting *S. cebus* as a distinct species despite morphologic overlap with *S. fuelleborni* in several characters (Augustine, 1940; Premvati, 1959). Another potential point of interest is that Von Linstow's original description of *S. fuelleborni* was based on parasites from two African host species (*Pan troglodytes* and *Papio cynocephalus*); though no major morphological differences were observed, this could have created inherent variability and lead to assumptions on a lack of host specificity (Von Linstow, 1905). While it has been established that New World and Old World-primate-derived *Strongyloides* (i.e. *S. cebus* and *S. fuelleborni*) are not capable of cross-infection (Faust, 1931), no experiments have been conducted to compare the host specificity within African and Asian *S. fuelleborni* from different primate hosts.

The distinction between *S. fuelleborni* lineages on the basis of geography is supported by our analysis, where specimens obtained from Africa (Fig. 3, clusters 1 and 2) are clearly distinct from Asian genotypes. The finer division of Asian *S. fuelleborni* into multiple lineages is also supported (Fig. 3). *Strongyloides fuelleborni* from Bornean monkeys (proboscis monkeys, long-tailed macaques, silvered leaf monkeys) and orangutans form an outgroup to other genotypes (Fig. 3, clusters 6 and 7), and are also distinct from an undescribed *Strongyloides* sp. collected from lorises as part of the same Bornean survey (Frias *et al.,* 2018). We also note that the host range of *S. fuelleborni* from Malaysian Borneo (Fig. 3, clusters 6 and 7) and mainland Southeast Asia (Fig. 3, cluster 3) overlap; both genotypes were found in long-tailed macaques, yet the genetic distinction between Bornean and mainland Southeast Asian genotypes is clear. Japanese macaques (Fig. 3, cluster 4) harbour a lineage of *S. fuelleborni* that shares an affinity with isolates from mainland Southeast Asia and Southern Asia (Fig. 3, cluster 3), but these two clusters may still represent distinct groups. We note, however, that the specimens from Japanese macaques were based on an inferred genotype so the analysis was performed a second time, excluding data from Japanese macaques. The structure of the resulting dendrograms remained virtually unchanged following exclusion of this data, indicating that the distinction between *S. fuelleborni* from Malaysian Borneo, Africa and mainland Southeast

Asia/Southern Asia remains supported (Supplementary File S2, Appendix part D). Further sampling of primate *Strongyloides* across these areas of interest and from other primate species, will help to resolve the true picture of species diversity.

Similar to how the genetic differences observed in host varieties raised important questions on the most appropriate naming of *S. stercoralis* across hosts, the geographic differences we uncovered for *S. fuelleborni* may also prompt a possible nomenclatural revision. Prior to the availability of molecular analyses, Ashford and Barnish stated that 'In the event of *S. fuelleborni* is as thus conceived including more than one species, the name *S. simiae* is available for the parasites of Asian primates' (Ashford and Barnish, 1989). In the 'molecular era', Hasegawa *et al.* (2009, 2010) also remarked on the considerable diversity of *S. fuelleborni* across Asian and African varieties, suggesting that subspecific designations might be indicated for varieties they investigated. Though not possible to definitively confirm the specific status and assign '*S. simiae*' or suggest another name for Asian isolates from this work alone, this opens the possibility for such if further characterization continues to support that conclusion. It is also possible that *Macaca* spp. and other primates may harbour multiple, possibly cryptic, *Strongyloides* species. Importantly, genetic differences between and among African and Asian primate *S. fuelleborni* may lead to corresponding differences in zoonotic potential, which could explain the nearly exclusive restriction of human *S. fuelleborni* infections to sub-Saharan Africa. If this is the case, recognizing separate species would be useful if further investigation supports this. Presently, *S. fuelleborni* genotyping datasets are smaller and less complete compared to those available for *S. stercoralis*. Additional sequence data from morphologically characterized African and Asian *S. fuelleborni* would be required from a range of hosts before any taxonomic changes are formalised.

Continuing genetic analyses will aid in resolving a major outstanding question on whether the occurrence of *S. stercoralis* in animals and humans represents zoonotic spillover (animal to human transmission) or spillback (human to animal transmission). Apart from occasional *S. stercoralis* infections in great apes living in proximity to infected humans, bona fide *S. stercoralis* infections, or infections with similar species are very seldom detected in herbivores or the omnivorous primates (Grove, 1989). Therefore, it seems that *S. stercoralis* would have originally evolved in canids or allied taxa (e.g. within the suborder Caniformia) rather than in hominids.

Finally, historic experimentation reveals that two *Strongyloides* species described from procyonids, *S. procyonis* and *S. nasua,* show interesting patterns analogous to observations of *S. stercoralis* of dogs and humans. The raccoon-associated species *S. procyonis* bears close morphological similarities to *S. stercoralis* (see Little, 1966a). In one trial, a short-lived, transient patent *S. procyonis* infection was successfully established in a human volunteer; one dog also developed a moderate-duration patent infection (3.5 months) (Little, 1966b). Raccoons may also be susceptible to human *S. stercoralis* – a young raccoon developed a patent infection of moderate duration (92 days) following exposure to filariform larvae collected from an infected human (Johnson, 1962). *Strongyloides nasua* of coati (*Nasua* spp.) – is also highly similar to *S. stercoralis* and some authors regard *Strongyloides nasua* as a synonym for *S. stercoralis* (Moraes *et al.,* 2019). Also, a patent infection was established in a white-nosed coati (*Nasua narica*) using human-derived *S. stercoralis* (Sandground, 1926). The observation of morphological overlap and cross-transmissibility among these species of *S. stercoralis*-like parasites (*S. procyonis, S. nasua, S. stercoralis* of dogs, and *S. stercoralis* of humans) supports the idea that the common ancestor of the *S. stercoralis*-like parasites originated in some ancestral Caniformia carnivore, that adapted to infecting humans sometime

during the domestication of dogs. This idea was also suggested by Nagayasu *et al.* (2017) on the basis of *S. procyonis* being the closest extant relative to *S. stercoralis*, which is also supported by phylogenies constructed from 18S rDNA sequences by Hino *et al.* (2014). Further genetic characterization of the procyoid parasites (*S. procyonis, S. nasua*) and other currently undescribed *Strongyloides* of other caniforms, including both true canids and arctoid mammals (e.g. bears, mustelids), will be necessary to investigate this evolutionary hypothesis.

It follows from the evidence above that *S. fuelleborni,* rather than *S. stercoralis*, is probably best regarded as the true 'human *Strongyloides*' that has coevolved with our species, given the primate origin of the former, and that *S. stercoralis* probably represents an evolutionary spillover event. The inability to morphologically distinguish distinct genetic lineages of *S. stercoralis*, the relationship between *S. stercoralis* and other *Strongyloides* infecting mammals of the order Canifornia, in conjunction with the clustering of canine *S. stercoralis* (cluster 5) as an outgroup in this analysis, support the hypothesis of Nagayasu *et al.* (2017) who proposed that the origin of human-infecting *S. stercoralis* was related to the domestication of dogs circa 14,000–6500 years before present (MacHugh *et al.*, 2017). However, our data indicate that some genotypes have seemingly adapted to the human host to a point where they rarely infect dogs (Fig. 2, excluding clusters 4 & 5). Therefore, these rare cases of dog infection caused by clearly human-adapted genotypes (e.g. Fig. 2, clusters 1–3 & 6) possibly represent modern and ongoing zoonotic 'spillback'. This highlights that generalizations regarding the role of dogs in human strongyloidiasis are difficult to make without understanding the population structure of this species complex.

Concluding remarks

Our analysis indicates that the 'two lineage' population structure supposing one zoonotic lineage (genotype A) and a second dog-specific lineage (genotype B) represents an over-simplification of the *S. stercoralis* population structure. Our findings show that a gradient of host permissibility is supported, where one genotype (lineage B; associated with cluster 5, Fig. 2), is dog-specific, producing a natural and distinct outgroup to all of lineage A (Fig. 2, lineage A includes all clusters except 5). *Strongyloides stercoralis* assigned to clusters 1, 2, 3 and 6 support a 'spillback' model; otherwise more dog-derived *S. stercoralis* specimens would be expected in these clusters. While available data for *S. fuelleborni* was sparser and less complete in comparison to *S. stercoralis*, the distinction between African and Asian genotypes seems to support allopatric or vicariant speciation. The separation of Asian genotypes into at least two separate lineages (Fig. 3, clusters 6 and 7 from Malaysian Borneo and clusters 3 and 4 from mainland Asia and Japan), may also be warranted. Close examination of experimental infection and reproductive studies performed by historic investigators generally support the aforementioned trends, as evidenced in our detailed discussion and review of published literature.

The associations between host, geography and genotype reported here, tend to corroborate both early and modern observations that support the proposal that *S. stercoralis* represents a species complex (*S. stercoralis* sensu lato), and referring to lineage B (Fig. 2, cluster 5) provisionally by a different designation – for example, *Strongyloides stercoralis* dog genotype, or *S. stercoralis canis* as proposed by Chandler (1925). Results from further morphological and molecular studies as well as modern *in vitro* and *in vivo* biological comparative studies of defined isolates might provide support for *S. stercoralis* from dogs (specifically, lineage B/cluster 5) to be renamed as *S. canis* as originally suggested by Brumpt (1922) and later by Augustine (1940).

It appears that *Strongyloides fuelleborni* is the most appropriate name for species infecting African primates, as this species was originally described from African chimpanzees and baboons (Von Linstow, 1905). However, the genetic distinctions noted here and elsewhere (Fig. 3) (Barratt *et al.*, 2019*b*) support that future taxonomic revision is needed (Hung and Höppli, 1923) for *Strongyloides* currently designated as '*S. fuelleborni*' infecting Asian primates, although characterization of additional parasite material is needed. It is also possible that African and Asian *S. fuelleborni* groups represent a species complexes of their own, further emphasizing the need for resolving taxonomic relationships among these lineages.

To address the remaining unanswered questions raised here regarding the statuses of *S. stercoralis* and *S. fuelleborni* as species complexes, we first propose that additional sampling and genotyping is required. This sampling should be focused particularly in areas where there is an ongoing dog and human *S. stercoralis* sensu lato transmission. Given the strong sampling bias towards *S. stercoralis* from Southeast Asian countries and Japan, sampling of specimens from other endemic areas would be of great value; data from Africa, the Middle East and South America are particularly sparse. A wider sampling of *S. fuelleborni* from African humans and other primates is also required. This is especially pertinent to Japanese *S. fuelleborni* for which very few sequences are currently available. In addition to experimental infections to evaluate host specificity, *in vitro* fertilization/mating experiments such as those performed by Augustine (1940), using genetically characterized isolates could provide critical insight into reproductive isolation and therefore species statuses of *Strongyloides* populations of interest. Finally, sequencing of additional *Strongyloides* loci (e.g. as per Nagayasu *et al.*, 2017) could provide increased resolution of genetic relationships when analyses like the one described here are performed.

References

Ashford RW and Barnish G (1989) Strongyloides fuelleborni. In Grove DI (ed.), *Strongyloidiasis: A Major Roundworm Infection of Man*. Oxfordshire, UK: Taylor & Francis, pp. 271–286.

Ashkenazy H, Cohen O, Pupko T and Huchon D (2014) Indel reliability in indel-based phylogenetic inference. *Genome Biology and Evolution* **6**, 3199–3209.

Augustine DL (1940) Experimental studies on the validity of species in the genus *Strongyloides*. *American Journal of Hygiene. Section D. Helminthology* **32**, 24–32.

Augustine DL and Davey DG (1939) Observations on a natural infection with *Strongyloides* In the dog. *Journal of Parasitology* **25**, 117–119.

Barratt JLN, Park S, Nascimento FS, Hofstetter J, Plucinski M, Casillas S, Bradbury RS, Arrowood MJ, Qvarnstrom Y and Talundzic E (2019*a*)

Genotyping genetically heterogeneous *Cyclospora Cayetanensis* infections to complement epidemiological case linkage. *Parasitology* 146, 1275–1283.

Barratt JLN, Lane M, Talundzic E, Richins T, Robertson G, Formenti F, Pritt B, Verocai G, Nascimento de Souza J, Mato Soares N, Traub R, Buonfrate D and and Bradbury RS (2019*b*) A global genotyping survey of *Strongyloides Stercoralis* and *Strongyloides Fuelleborni* using deep amplicon sequencing. *PLoS Neglected Tropical Diseases* 13, e0007609.

Basso W, Grandt L-M, Magnenat A-L, Gottstein B and Campos M (2019) *Strongyloides Stercoralis* infection in imported and local dogs in Switzerland: from clinics to molecular genetics. *Parasitology Research* 118, 255–266.

Beknazarova M, Barratt JLN, Bradbury RS, Lane M, Whiley H and Ross K (2019) Detection of classic and cryptic *Strongyloides* Genotypes by deep amplicon sequencing: a preliminary survey of dog and human specimens collected from remote Australian communities. *PLoS Neglected Tropical Diseases* 13, e0007241.

Bisoffi Z, Buonfrate D, Montresor A, Requena-Méndez A, Muñoz J, Krolewiecki AJ, Gotuzzo E, Mena MA, Chiodini PL, Anselmi M, Moreira J and Albonico M (2013) *Strongyloides Stercoralis*: a plea for action. *PLoS Neglected Tropical Diseases* 7, e2214.

Brumpt E (1922) *Précis de Parasitologie*. Paris, France: Masson et Cie.

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* 17, 540–552.

Chandler AC (1925) The species of *Strongyloides* (Nematoda). *Parasitology* 17, 426–433.

Desportes C (1944) Sur *Strongyloides Stercoralis* (Bavay 1876) et sur les *Strongyloides* de primates. *Annales de Parasitologie Humaine et Comparée* 20, 160–190.

Donath A and Stadler PF (2018) Split-inducing indels in phylogenomic analysis. *Algorithms for Molecular Biology* 13, 12.

Dorris M, Viney ME and Blaxter ML (2002) Molecular phylogenetic analysis of the genus *Strongyloides* and related nematodes. *International Journal for Parasitology* 32, 1507–1517.

Faust EC (1931) Infection experiments in monkeys with human, macaque and Ateles Strains of Strongyloides. *Proceedings of the Society for Experimental Biology And Medicine* 28, 919–920.

Faust EC (1933) Experimental studies on human and primate species of *Strongyloides*. II. the development of *Strongyloides* in the experimental host. *American Journal of Hygiene* 18, 114–132.

Faust EC and Giraldo LE (1960) Parasitological surveys in Cali, Departamento del Valle, Colombia VI. Strongyloidiasis in Barrio Siloé, Cali, Colombia. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 54, 556–563.

Freedman DO (1991) Experimental infection of human subjects with *Strongyloides* species. *Reviews of infectious diseases* 13, 1221–1226.

Frias L, Stark DJ, Lynn MS, Nathan SK, Goossens B, Okamoto M and MacIntosh AJJ (2018) Lurking in the dark: cryptic *Strongyloides* in a bornean slow loris. *International Journal for Parasitology. Parasites and Wildlife* 7, 141–146.

Fuelleborn F (1914) Untersuch. Uber den Infektionsweg bei *Strongyloides* Und *Ankylostomum* Und die Biologie dieser Parasiten. *Arch. Schiffs-u Tropen-Hyg* 18, 26–80.

Fuelleborn F (1927) Über das Verhalten der Larven von *Strongyloides Stercoralis*, Hakenwürmern und *Ascaris lumbricoides* im Körper des Wirtes. *Arch. Schiffs.-u. Tropenhyg* 31, 1–56.

Galliard H (1939) Recherches sur la strongyloïdose au Tonkin. Rôle des animaux domestiques dans l'étiologie de l'infestation humaine. *Annales de Parasitologie Humaine et Comparée* 17, 533–541.

Galliard H (1950) Recherches sur l'infestation expérimentale a *Strongyloides Stercoralis* au Tonkin (1e note). *Annales de Parasitologie Humaine et Comparée* 25, 441–473.

Galliard H (1951*a*) Recherches sur l'infestation expérimentale a *Strongyloides Stercoralis* au Tonkin (2e note). *Annales de Parasitologie Humaine et Comparée* 26, 67–84.

Galliard H (1951*b*) Recherches sur l'infestation expérimentale a *Strongyloides Stercoralis* Au Tonkin (3e note). *Annales de Parasitologie Humaine et Comparée* 26, 201–227.

Galliard H and Berdonneau R (1953) Strongyloïdose expérimentale chez le chien. Effets de la cortisone. Résultats du test de Thorn à l'hormone corticotrope (ACTH). *Annales de Parasitologie Humaine et Comparée* 28, 163–171.

Genta RM (1989) *Strongyloides Stercoralis*: loss of ability to disseminate after repeated passage in laboratory beagles. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 83, 539–541.

Genta RM and Grove DI (1989). Strongyloides stercoralis infections in animals. In Grove DI (ed.), *Strongyloidiasis: A Major Roundworm Infection of Man*. London: Taylor & Francis, pp. 251–270.

Georgi JR and Sprinkle CL (1974) A case of human strongyloidosis apparently contracted from asymptomatic colony dogs. *The American journal of tropical medicine and hygiene* 23, 899–901.

Golubchik T, Wise MJ, Easteal S and Jermiin LS (2007) Mind the gaps: evidence of bias in estimates of multiple sequence alignments. *Molecular Biology and Evolution* 24, 2433–2442.

Goodey T (1926) Observations on *Strongyloides Fülleborni* von Linstow, 1905, with some remarks on the genus *Strongyloides*. *Journal of Helminthology* 4, 75–86.

Grove DI (1989) *Strongyloidiasis: A Major Roundworm Infection of Man*. London: Taylor & Francis.

Grove DI and Northern C (1982) Infection and immunity in dogs infected with a human strain of *Strongyloides stercoralis*. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 76, 833–838.

Hasegawa H, Hayashida S, Ikeda Y and Sato H (2009) Hyper-variable regions in 18S rDNA of *Strongyloides* Spp. as markers for species-specific diagnosis. *Parasitology Research* 104, 869–874.

Hasegawa H, Sato H, Fujita S, Nguema PPM, Nobusue K, Miyagi K, Kooriyama T, Takenoshita Y, Noda S, Sato A, Morimoto A, Ikeda Y and Nishida T (2010) Molecular identification of the causative agent of human strongyloidiasis acquired in Tanzania: dispersal and diversity of *Strongyloides* Spp. and their hosts. *Parasitology International* 59, 407–413.

Hino A, Tanaka T, Takaishi M and Fuji Y (2014) Karyotype and reproduction mode of the rodent parasite *Strongyloides venezuelensis*. *Parasitology* 141, 1736–1745.

Hira PR and Patel BG (1980) Human strongyloidiasis due to the primate species *Strongyloides fülleborni*. *Tropical and Geographical Medicine* 32, 23–29.

Hung SL and Höppli R (1923) Morphologische und histologische Beiträge zur *Strongyloides*-Infection der Tiere. *Arch. f. Schiffs-u. Tropen-Hyg* 27, 118–129.

Jaleta TG, Zhou S, Bemm FM, Schär F, Khieu V, Muth S, Odermatt P, Lok JB and Streit A (2017) Different but overlapping populations of *Strongyloides Stercoralis* in dogs and humans – dogs as a possible source for zoonotic strongyloidiasis. *PLoS Neglected Tropical Diseases* 11, e0005752.

Johnson AD (1962) Experimental infection of the raccoon with *Strongyloides stercoralis*. *Journal of Parasitology* 48, 332.

Kikuchi T, Hino A, Tanaka T, Aung MPPTHH, Afrin T, Nagayasu E, Tanaka R, Higashiarakawa M, Win KK, Hirata T, Htike WW, Fujita J and Maruyama H (2016) Genome-wide analyses of individual *Strongyloides Stercoralis* (Nematoda: Rhabditoidea) provide insights into population structure and reproductive life cycles. *PLOS Neglected Tropical Diseases* 10, e0005253.

Kreis HA (1932) Studies on the genus *Strongyloides* (Nematodes). *American Journal of Hygiene* 16, 450–491.

Laymanivong S, Hangvanthong B, Insisiengmay B, Vanisaveth V, Laxachack P, Jongthawin J, Sanpool O, Thanchomnang T, Sadaow L, Phosuk I, Rodpai R, Maleewong W and Intapan PM (2016) First molecular identification and report of genetic diversity of *Strongyloides Stercoralis*, a current major soil-transmitted helminth in humans from Lao People's Democratic Republic. *Parasitology Research* 115, 2973–2980.

Little MD (1966*a*) Comparative morphology of six species of *Strongyloides* (Nematoda) and redefinition of the genus. *Journal of Parasitology* 52, 69–84.

Little MD (1966*b*) Seven new species of *Strongyloides* (Nematoda) from Louisiana. *Journal of Parasitology* 52, 85–97.

Looss A (1911) The anatomy and life history of *Ancylostoma Duodenale* Dub. A monography. Part II. The development in the free state. *Records of the School of Medicine, Egyptian Ministry of Education* 4, 163–613.

MacHugh DE, Larson G and Orlando L (2017) Taming the past: ancient DNA and the study of animal domestication. *Annual Review of Animal Biosciences* 5, 329–351.

Moraes MFD, da Silva MX, Tebaldi JH and Lux Hoppe EG (2019) Parasitological assessment of wild ring-tailed coatis (*Nasua nasua*) from the Brazilian Atlantic rainforest. *International Journal for Parasitology: Parasites and Wildlife* 9, 154.

Nagayasu E, Htwe MPPTH, Hortiwakul T, Hino A, Tanaka T, Higashiarakawa M, Olia A, Taniguchi T, Win SMT and Ohashi I (2017) A possible origin population of pathogenic intestinal nematodes, *Strongyloides Stercoralis,* Unveiled by molecular phylogeny. *Scientific Reports* 7, 4844.

Nutman TB (2017) Human infection with *Strongyloides Stercoralis* and other related *Strongyloides* species. *Parasitology* 144, 263–273.

Olsen A, van Lieshout L, Marti H, Polderman T, Polman K, Steinmann P, Stothard R, Thybo S, Verweij JJ and Magnussen P (2009) Strongyloidiasis – the most neglected of the neglected tropical diseases? *Transactions of the Royal Society of Tropical Medicine and Hygiene* 103, 967–972.

Pampiglione S and Ricciardi ML (1971) The presence of *Strongyloides Fülleborni* von Linstow, 1905, in man in Central and East Africa. *Parassitologia* 13, 257–269.

Pampiglione S and Ricciardi ML (1972) Experimental infestation with human strain *Strongyloides Fuelleborni* in man. *Lancet (London, England)* 1, 663–665.

Paradies P, Iarussi F, Sasanelli M, Capogna A, Lia RP, Zucca D, Greco B, Cantacessi C and Otranto D (2017) Occurrence of strongyloidiasis in privately owned and sheltered dogs: clinical presentation and treatment outcome. *Parasites & Vectors* 10, 345.

Premvati G (1959) Studies on *Strongyloides* of primates: V. Synonymy of the species in monkeys and apes. *Canadian Journal of Zoology* 37, 75–81.

Sandground JH (1925) Speciation and specificity in the nematode genus *Strongyloides. Journal of Parasitology* 12, 59–80.

Sandground JH (1926) Biological studies on the life-cycle in the genus *Strongyloides* Grassi, 1879. *American Journal of Epidemiology* 6, 337–388.

Sandground JH (1928) Some studies on susceptibility, resistance, and acquired immunity to infection with *Strongyloides Stercoralis* (Nematoda) in dogs and cats. *American Journal of Hygiene* 8, 507–538.

Sandosham AA (1952) An investigation into the association of creeping eruption with *Strongyloides* infection contracted in the Far East. *Journal of Helminthology* 26, 1–24.

Sato H, Torii H, Une Y and Ooi H-K (2007) A new rhabditoid nematode species in Asian sciurids, distinct from *Strongyloides robustus* in North American sciurids. *Journal of Parasitology* 93, 1476–1486.

Schär F, Guo L, Streit A, Khieu V, Muth S, Marti H and Odermatt P (2014) *Strongyloides Stercoralis* genotypes in humans in Cambodia. *Parasitology International* 63, 533–536.

Speare R (1989) Identification of species of Strongyloides. In Grove DI (ed.), *Strongyloidiasis: A Major Roundworm Infection of Man.* London, New York, Philadelphia: Taylor & Francis, pp. 11–84.

Speare R and Tinsley DJ (1986) *Strongyloides felis*: an" old" worm rediscovered in Australian cats. *Australian Veterinary Practitioner* 16, 10–18.

Takano Y, Minakami K, Kodama S, Matsuo T and Satozono I (2009) Cross infection of *Strongyloides* Between humans and dogs in the Amami Islands, Japan. *Tropical Medicine and Health* 37, 149–152.

Thamsborg SM, Ketzis J, Horii Y and Matthews JB (2017) *Strongyloides* Spp. infections of veterinary importance. *Parasitology* 144, 274–284.

Thanchomnang T, Intapan PM, Sanpool O, Rodpai R, Tourtip S, Yahom S, Kullawat J, Radomyos P, Thammasiri C and Maleewong W (2017) First molecular identification and genetic diversity of *Strongyloides Stercoralis* and *Strongyloides Fuelleborni* in human communities having contact with long-tailed macaques in Thailand. *Parasitology Research* 116, 1917–1923.

Thanchomnang T, Intapan PM, Sanpool O, Rodpai R, Sadaow L, Phosuk I, Somboonpatarakun C, Laymanivong S, Tourtip S and Maleewong W (2019) First molecular identification of *Strongyloides Fuelleborni* in long-tailed macaques in Thailand and Lao People's Democratic Republic reveals considerable genetic diversity. *Journal of Helminthology* 93, 608–615.

Truszkowski J and Goldman N (2016) Maximum likelihood phylogenetic inference is consistent on multiple sequence alignments, with or without gaps. *Systematic Biology* 65, 328–333.

Von Linstow OFB (1905) *Strongyloides Fülleborni* N.sp. *Centralblatt für Bakteriologie und Parasitenkunde* 38, 532–534.

Wulcan JM, Dennis MM, Ketzis JK, Bevelock TJ and Verocai GG (2019) *Strongyloides* Spp. in cats: a review of the literature and the first report of zoonotic *Strongyloides Stercoralis* in colonic epithelial nodular hyperplasia in cats. *Parasites & Vectors* 12, 349.

Zhou S, Fu X, Pei P, Kucka M, Liu J, Tang L, Zhan T, He S, Chan YF, Rödelsperger C, Liu D and Streit A (2019) Characterization of a non-sexual population of *Strongyloides Stercoralis* with hybrid 18S rDNA haplotypes in Guangxi, Southern China. *PLoS Neglected Tropical Diseases* 13, e0007396.