Nanopore-based analysis unravels the genetic landscape and phylogenetic placement of human-infecting *Trichuris* species in Côte d'Ivoire, Tanzania, Uganda, and Laos

Nurudeen Rahman[1,2], **Max A. Bär**[1,2], **Julian Dommann**[1,2], **Eveline Hürlimann**[1,2], **Jean T. Coulibaly**[1,2,3,4], **Said Ali**[5], **Somphou Sayasone**[1,2,6], **Prudence Beinamaryo**[7], **Jennifer Keiser***[1,2], **Pierre H.H. Schneeberger***[1,2]

*These authors contributed equally as senior authors

[1] Department of Medical Parasitology and Infection Biology, Swiss Tropical and Public Health Institute, Allschwil, Switzerland

[2] University of Basel, Basel, Switzerland

[3] Unité de Formation et de Recherche Biosciences, Université Felix Houphouët-Boigny, Abidjan, Côte d'Ivoire

[4] Centre Suisse de Recherches Scientifiques en Côte d'Ivoire, Abidjan, Côte d'Ivoire

[5] Public Health Laboratory Ivo de Carneri, Chake Chake, Pemba, Zanzibar, Tanzania

[6] Lao Tropical and Public Health Institute, Vientiane, Laos

[7] Vector Borne and Neglected Tropical Diseases Division, Ministry of Health, Kampala, Uganda

Abstract

Soil-transmitted helminthiases (STH), including trichuriasis, pose a significant global burden, affecting numerous mammalian hosts. Traditional diagnostic methods have struggled to differentiate species within the *Trichuris* genus, prompting the use of molecular techniques to understand the diversity and zoonotic transmission of *Trichuris* spp. among humans and non-human primates (NHP), dogs, and pigs. In this study, we used nanopore-based full-length ITS2 rDNA sequencing to genetically characterize and confirm the phylogenetic placement of the newly identified human-infecting *Trichuris incognita* and the well-known *Trichuris trichiura*. Using 670 samples from clinical studies in Côte d'Ivoire, Laos, Tanzania, and Uganda, along with publicly available ITS2 sequences, we confirmed two distinct

clusters of human-infecting *Trichuris* species. We found a varying prevalence of *T. incognita* and *T. trichiura* across the four countries and that these were also present in NHP. Our findings demonstrated that ITS2 fragment length accurately differentiates both human-infecting *Trichuris* species and thus makes it a good diagnostic marker for future point-of-care applications in resource-constrained settings.

Introduction

Trichuriasis, an infection caused by the whipworm *Trichuris spp*. is of considerable public health significance and is considered a Neglected Tropical Disease (NTD) (1). This particular NTD has been estimated to disproportionately infect over 465 million people worldwide, mainly prevalent among children living in impoverished communities with limited hygiene or sanitation facilities, leading to a significant burden of 0.64 million disability-adjusted life years (DALYs) lost annually (2). However, several other mammals such as non-human primates (NHP), rodents, ruminants, and marsupials to name a few encompass the more than 20 specific taxonomic groups of hosts associated specifically with the genus *Trichuris,* meaning they are multi-host parasites (3,4).

Previous research approaches using traditional diagnostic methods like Kato-Katz have always regarded *Trichuris sp*. found in humans and NHP as *Trichiura trichiura* except in host-specific cases like the NHP Colobus monkeys (*Trichuris colobae*) (5) and Chacma baboons (*Trichuris ursinus*) (6) where there is a partial understanding of the *Trichuris* multi-host complexity. Morphologically, both eggs and adult worms of *T. trichiura* from humans and NHP are indistinguishable making molecular methods a promising approach (7–10). Recent efforts employed mostly sequences from the internal transcribed spacer (ITS) regions 1 and 2 (ribosomal DNA), mitochondrial markers and β tubulin gene to delineate this cryptic difference, hence resolving the *Trichuris* complex or multi-host specificity (5,9–15). These sequences were used to prove the existence of two separate clusters of *Trichuris* in humans with the consensus being that they both were *T. trichiura* (6,9–15). However, a recent study in Côte d'Ivoire revealed a novel *Trichuris* species, called *T. incognita,* that was

observed in a monophyletic cluster with *T.suis* and *T. colobae* (16). ITS sequences corresponding to this species have been observed in single human patients in Uganda, and Cameroon, as well as several NHP (14). This sequence has also been shown to be a reliable marker to distinguish between *T. suis* isolated from pigs, *T. incognita* isolated from humans, and *T. trichiura* isolated from humans and non-human primates (16,17).

In this study, we sought to investigate the genetic diversity and phylogenetic placement of human-infecting *Trichuris* species in four Low- and middle-income countries (LMIC) across the globe. To do this, we analyzed the full ITS2 region (~500-600bp) from whipworms in stool samples collected from a total of 670 patients from clinical studies in Côte d'Ivoire, Tanzania, Uganda, and Laos. These samples were collected as part of randomized controlled trials (18,19) aiming at testing the efficacy of anthelmintic treatments. We next integrated the dataset generated from these studies with all publicly available and complete ITS2 sequences for the *Trichuris* genus, including sequences derived from humans, non-human primates, pigs, livestock, and rodents from multiple locations. Subsequently, we investigated the genetic makeup and phylogenetic relationships within our study populations and in a broader context. Finally, we identified a potential PCR-based diagnostic marker using the ITS2 fragment length that effectively differentiates between the two known human-infecting species, *Trichuris trichiura* and *Trichuris incognita*.

Results

Description of the study population

The population studied consisted of patients from the Pujini and Dodo shehias on Pemba Island, Tanzania; the Kisoro and Kabale districts in Uganda; the Dabou and Jacqueville regions around the Lagunes district in Côte d'Ivoire; and the Nambak district and Luang Prabang province in Laos (**Fig. 1,** (adapted from Keller et al. (20)). This study included frozen faecal samples from participants aged 6 to 60 years in Côte d'Ivoire, Tanzania and

Laos, as well as ethanol-preserved faecal samples from school-aged children (6-12 years) in Uganda.

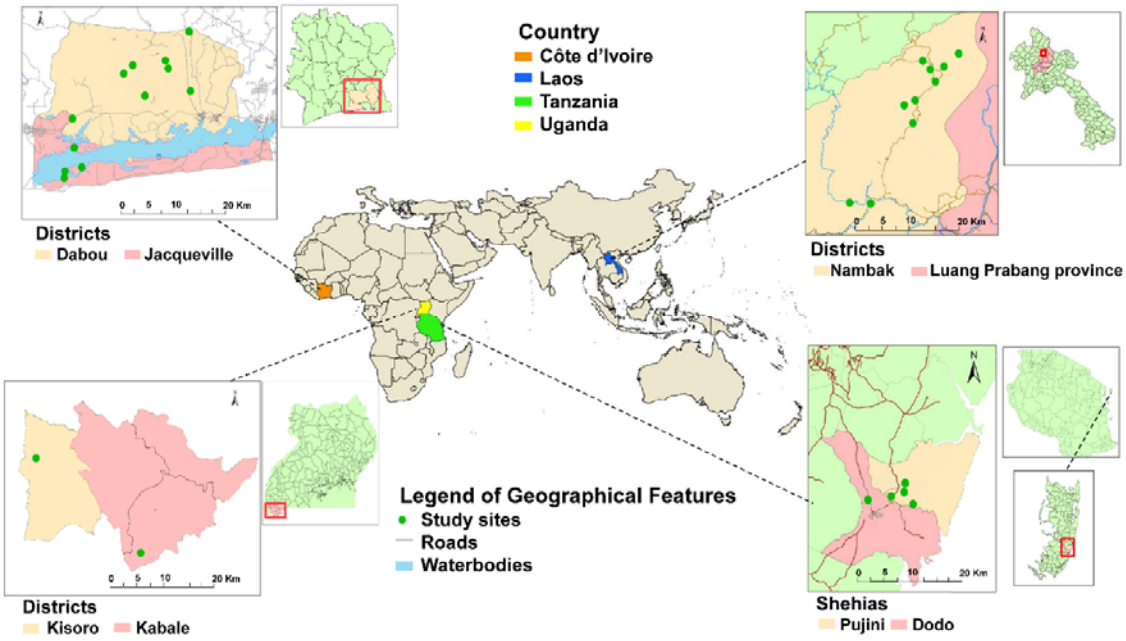


Fig. 1: Sampling locations of the study population. The map of the world highlights the four different study locations: the Lagunes region in Côte d'Ivoire, Pujini and Dodo Shehias on Pemba Island in Tanzania, Nambak districts and Luang Prabang province in Laos PDR (adapted from Keller et al. (20)), and the Kabale and Kisoro districts in Uganda each depicted in varying colours on the map.

Stool samples used in this study were collected in the framework of a double-blind, placebo-controlled randomized trial conducted between September 2018 and June 2020 in Côte d'Ivoire, Tanzania, and Laos (18,21,22) as well as a parallel open-label randomized controlled superiority trial conducted between October to November 2023 in Uganda (registered as NCT06037876; clinicaltrials.gov/). The metadata for the demographics and helminth infection characteristics of the 670 study participants are shown in **Table 1**. The infection intensities as diagnosed with Kato-Katz (KK) and sex distribution of the study participants depict a good balance between treatment groups. Of the 670 participants, 56.0% (375/670) were female and 44.0% (295/670) were male across the four countries. Overall, the Lao participants were older (mean age 23·1 years [SD 15·2]) than the other three cohorts (mean age 16·5 years [13·0] in Côte d'Ivoire, 14·1 years [9.8] in Tanzania, and

15.8 years [12.6] in Uganda). Samples of all 670 participants were included in PCR-based investigations of the internal transcribed spacer 2 (ITS2) region.

Table 1 Description of the study population

	Côte d'Ivoire (n=189)	**Laos (n=156)**	**Tanzania (n=190)**	**Uganda (n=135)**
Age in Years				
6 -12	116 (61.4%)	71 (45.5%)	119 (62.6%)	135 (100%)
13 – 24	32 (16.9%)	14 (9.0%)	49 (25.8%)	0
≥25	41 (21.7%)	71 (45.5%)	22 (11.6%)	0
Sex				
Female	96 (50.8%)	84 (53.8%)	107 (56.4%)	88 (65.2%)
Male	93 (49.2%)	72 (46.2%)	83 (43.6%)	47 (34.8%)
Infection intensity[+]				
Light	144 (76.2%)	136 (87.2%)	144 (75.8%)	118 (87.4%)
Moderate	42 (22.2%)	20 (12.8)	43 (22.6%)	17 (12.6%)
Heavy	3 (1.6%)	0	3 (1.6%)	0

[+] Infection intensity was classified according to mean egg per gram EPG into light (1–999 EPG), moderate (1000–9999 EPG), and heavy (≥10 000 EPG).

Molecular Analysis of ITS2 Nuclear Marker

Following DNA extraction of the faecal samples from the patients partaking in the clinical trials across the four countries, we successfully produced amplicons from 188, 154, 174 and 113 participants in Tanzania, Laos, Côte d'Ivoire, and Uganda, respectively. The resulting ITS2 amplicons were sequenced using a MinION platform to achieve 5 to 9896 unprocessed reads per sample. Using a stringent Divisive Amplicon Denoising Algorithm Version 2, DADA2-based pipeline, we retained a total of 198,792 reads (ranging between 5 to 9420

high-quality reads per sample) that were processed into 474 unique amplicon sequence variants (ASVs) (**Additional file 2: Fig. S1**).

Population Differentiation and Network Analysis

Analyses of haplotypes showed the presence of 233 haplotypes of the *Trichuris* population across the four different countries (**Additional file 2: Table S1**). Ten haplotypes were shared among the ASVs from the different countries, with four haplotypes shared within the four countries, while 223 haplotypes were revealed across all settings (51, 63, 74 and 35 from Côte d'Ivoire, Laos, Tanzania and Uganda, respectively). A few of the haplotypes analyzed showed high frequencies, such as Hap 1 in Laos (19%), Hap 2 in Côte d'Ivoire (31%), Hap 3 in Tanzania (14%) and Hap4 in Laos and Tanzania (18% and 24% respectively), although these haplotypes were also present in varying proportions across the other countries (**Additional file 2: Table S1**). Comparison of these observed haplotypes across the four countries through NCBI BLAST showed that five haplotypes showing 100% identity with already described haplotypes: Hap1, Hap3 and Hap 4 were identical to *T. trichiura* isolated from samples analyzed from Ugandan patients (Accession number KJ588133 and KJ588134 respectively, (9)); Hap2 and Hap9 identified as *T. trichiura* isolated in worms expelled from patients in Cameroon (Accession number GQ301555, (14)). Also, the samples analyzed in Hap2 and Hap9 had a 96% to 100% identity match to ITS2 sequences from the newly described *T. incognita* species (17) following NCBI BLAST to its sequence. The comprehensive list encompassing the BLAST results is available as supplementary material (**Additional file 2: Table S2**).

Table 2: Genetic diversity estimations of *Trichuris* population per region

Country	N	N_H	Gene diversity	π	H_e
Côte d'Ivoire	189	58	0.9889±0.005	0.236±0.113	0.20
Laos	156	70	0.9951±0.003	0.153±0.073	0.13

Tanzania	190	83	0.9951±0.016	0.143±0.069	0.13
Uganda	135	42	0.9929±0.050	0.173±0.083	0.16

N = Number of samples sequenced; N_H = Number of haplotypes; π = Nucleotide diversity ; H_e = Expected heterozygosity

The analysis of genetic diversity indexes revealed that gene diversity within the populations ranged from 0.9889 to 0.9951, with the highest levels in Laos and Tanzania (both 0.9951), and the lowest in Côte d'Ivoire (0.9889). Conversely, nucleotide diversity (π) and expected heterozygosity (H_e) showed a different pattern, with the sample population from Côte d'Ivoire and Uganda exhibiting higher diversity. 58 haplotypes from Côte d'Ivoire had the highest nucleotide diversity (π = 0.236±0.113) and expected heterozygosity (H_e = 0.20), followed by 42 haplotypes from Uganda (π = 0.173±0.083, H_e = 0.16). In contrast, 70 haplotypes from Laos and 83 from Tanzania had lower nucleotide diversity (π = 0.153±0.0730 and 0.143±0.069 respectively) and expected heterozygosity (H_e = 0.13 for both), indicating fewer

Fig. 2: Underlying population genetic structure and network analysis among the four countries A. Haplotype network inferred from ASVs of *Trichuris* species haplotype sequences from Côte d'Ivoire, Laos, Tanzania and Uganda. Each circle represents a unique haplotype and the size of the circle is proportional to the corresponding haplotype frequency. The colours of the pie charts represent the country of origin (red, Côte d'Ivoire; yellow, Laos; purple, Tanzania; blue, Uganda). Hatch marks along the network branches indicate the number of nucleotide differences

differences between sequences in these populations. This suggests that while 153 haplotypes from Laos and Tanzania have extensive allelic variation, the alleles are more similar, whereas 100 haplotypes from Côte d'Ivoire and Uganda have more distinct variations at the nucleotide level (**Table 2**).

Analysis of the haplotype network analyses revealed a reticulate relationship among the ASVs in the four countries scattered in separate clusters (**Fig. 2A**). Specifically, the ASVs present in *Trichuris* populations of Côte d'Ivoire, Laos, Tanzania and Uganda were observed in the two clusters separated by 40 point mutations or nucleotide differences. By comparison, a further internal subdivision among haplotypes with a higher proportion was recorded in Tanzania than in the other three countries while the other sub-cluster revealed a higher proportion of haplotypes from Laos study participants than in the other three countries. The remaining haplotypes were observed in a different cluster, which represented a greater proportion of the *Trichuris* population in Côte d'Ivoire than those found in the Tanzania, Laos and Uganda populations.

Comparison of the pairwise fixation index (FST) values between the sampled populations revealed a range from -0.0007 to 0.1774 (**Fig. 2B**). ASVs of the sampled population from Côte d'Ivoire were highly differentiated from those in Tanzania, Laos and Uganda with FSTs of 0.1774, 0.1692 and 0.14112 respectively ($P < 0.001$ for all FST values). Consequently, ASVs between Laos and Uganda, Laos and Tanzania, as well as Uganda and Tanzania represented an undifferentiated population ($FST < 0.010$, $P > 0.001$).

Phylogenetic and comparative sequence analysis

Phylogenetic analysis of the ITS2 ASVs from the four countries together with publicly available ITS2 sequences of other related species was inferred using neighbour-joining (NJ), and Maximum Likelihood (ML) algorithms and their corresponding tree topologies provided a consistent, robust phylogenetic discrimination (**Fig. 3A**). We identified five clades, each

corresponding to different but taxonomically related host species and broadly matching published whole genome sequences (WGS) of *Trichuris* species. All the pairwise nucleotide identities between sequences generated in this study and publicly available sequences are summarized in **Fig. 3B**. Clade A consisted of published ITS2 sequences of *Trichuris* species of Indian bison, camel, deer and sheep but did not contain any well-characterized WGS of a specific *Trichuris* species. Clade B consisted of published sequences of *Trichuris* species of arvicolids and murid rodents – and included the sequence of an ITS2 region from a previously published WGS of *T. muris* (23,24). Clade C consisted of published sequences of *Trichuris* species of pigs which correctly matched the WGS of *T. suis* (25). Clade D was distinctly differentiated from *T. suis* sequences with nucleotide similarities ranging from 82 to 88% and contains ITS2 sequences from 20.5% (N = 97 samples) included in this study, as well as 14 publicly available sequences of human and NHP species. Sequences from clade D shared approximately 98% - 100% nucleotide sequence identity with the WGS sequence of novel species of human-infecting *Trichuris*, named *T. incognita*, observed in Côte d'Ivoire (17), as well as other published partial sequences from humans in Cameroon (accession number GQ301555) (14), non-human primates (NHP) from Italy and Uganda, and more distantly with Chacma baboons in South Africa (accession number GQ301554) (14) with 90% to 94% nucleotide sequence identity. It is noteworthy that over 52.6% of samples contained in this clade were from Côte d'Ivoire (N = 51) but it also contained samples from other countries, Laos (N = 15 samples), Tanzania (N = 21 samples), and Uganda (N = 10 samples) (**Fig. 3C**). Clade E contained 377 *Trichuris* sequences derived from the human samples in our study as well as 36 ITS2 reference sequences of human and NHP species and 1 WGS sequence corresponding to the most common human-infecting species, *T. trichiura*. Clade E shared 96% - 100% nucleotide sequence identity to published sequences from humans in Ecuador and Uganda (9,23), as well as sequences from captive and non-captive NHP from Uganda, China, South Africa, Italy, Spain, and St. Kitts & Nevis (10,11,14,26,27). It showed a more distant nucleotide sequence identity (79% to 86%) with sequences obtained from Francois Leaf monkeys in China (26). The majority of sequences

clustering in this clade were from Tanzania (N = 148), Laos (N =108), and Uganda (N = 59), but also contained approximately 16.4% (N

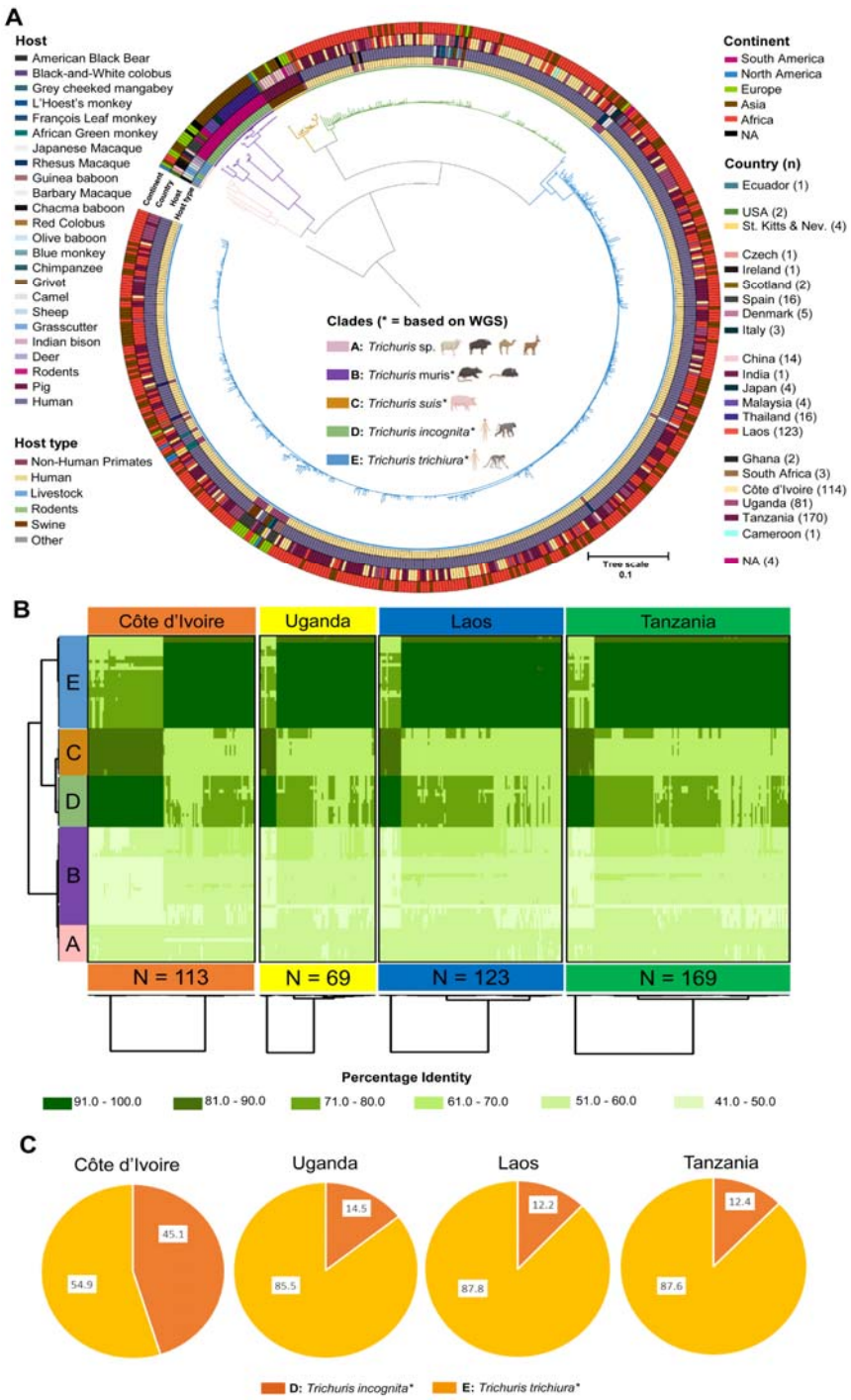


Fig. 3: Phylogenetic and sequence characterization of *Trichuris* population A. Maximum likelihood tree based on the ITS2 rDNA using Tamura-Nei with gamma distribution as the substitution model and *Trichinella spiralis* (GenBank accession: KC006432) as an outgroup. Coloured branches correspond to major clades A to E. External rings (from outer to inner one) represent metadata from the samples (as well as other *Trichuris* reference sequences available from rodents, livestock, pigs, humans, and non-human primates), namely continent, country of origin, host species and host-type respectively. **B.** Pairwise comparison of sequence similarity between human samples (x-axis) from the four countries and reference sequences (y-axis) incorporated in the phylogenetic tree using ward clustering method and Euclidean correlation on OriginPro 2021 graphing software. **C.** Proportion of human-infecting *Trichuris* species observed within Côte d'Ivoire, Uganda Laos, and Tanzania.

= 62) of sequences generated from samples collected in Côte d'Ivoire. Furthermore, clade E was also more distant from clades C and D with nucleotide identity values ranging from 64% to 69%.

Fragment length as a robust diagnostic marker to differentiate human-infecting *Trichuris* species

The fragment lengths - in basepair (bp) - of the ITS2 regions corresponding to the clade structure (clades A to E) are summarized in **Fig. 4**. The length of the ITS2 regions ranged from 424.9±47.6bp for *T. muris*-related sequences, 443.7±7.9bp for *Trichuris* species found in livestock, 579.9±2.4bp for *T. suis*-related sequences, 603.7±13.5bp for *T. incognita*-related sequences, and 539.1±30.4bp for *T. trichiura* related sequences. We observed

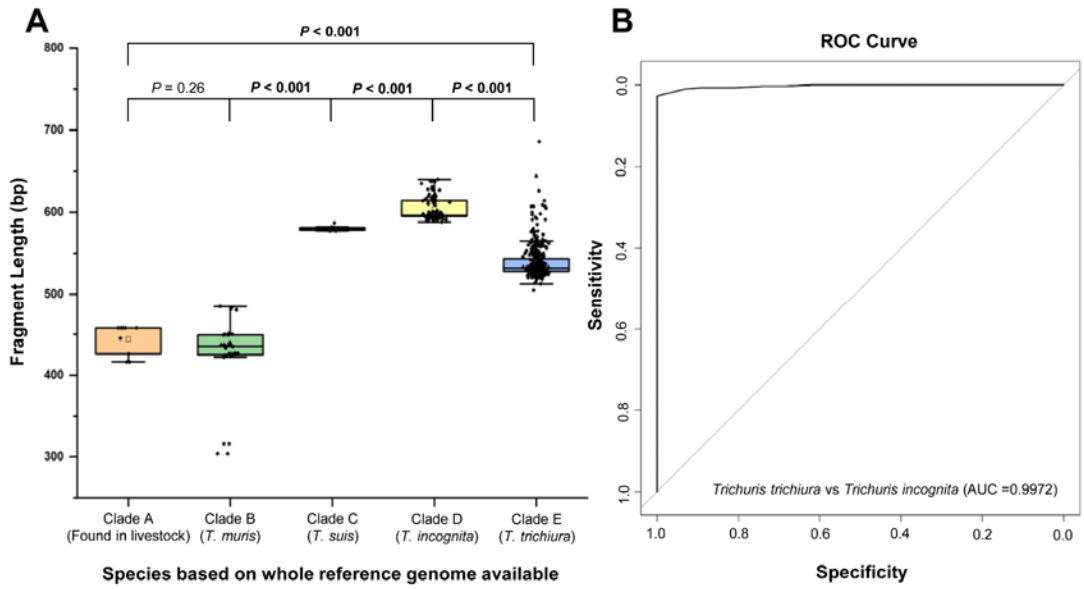


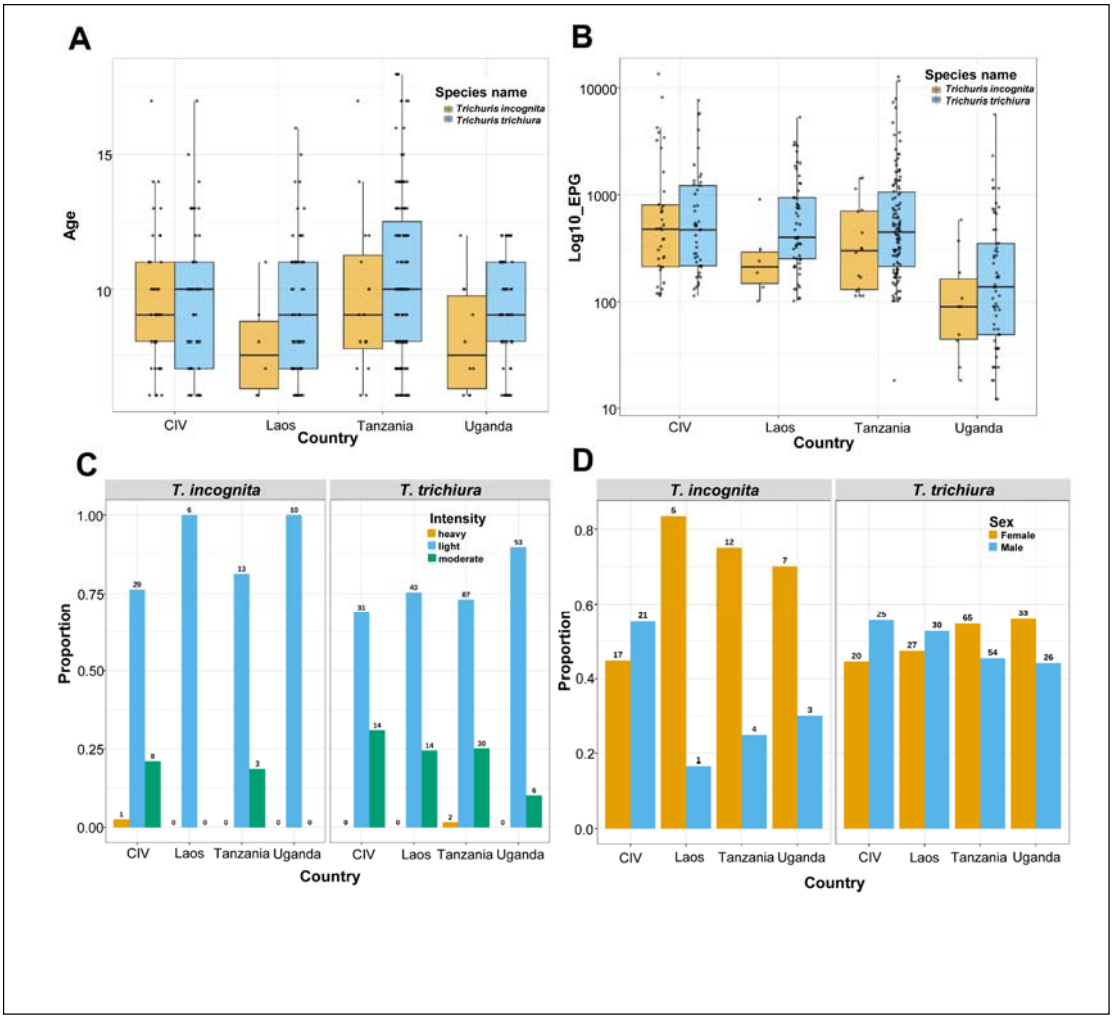
Figure 4: Fragment length could distinguish between different human-infecting *Trichuris* species. **A** Comparison of the fragment length between the five clades. Box plots include the median line, p-value estimated by the one-tailed Mann–Whitney U test. **B** ROC curve showing the performance of random-forest models built by combining the 2 signatures features of fragment length using the lose-1-keep1- m. bp: base pair.

significant differences in fragment length between all clades, except between *T. muris* and the livestock-related clade A.

We then evaluated the predictive power of the ITS2 region length to discriminate between the two human-infecting species, *T. incognita* and *T. trichiura*. For this purpose, we used a

random forest model on our dataset as shown in **Fig. 4B**. The model exhibited comparable classification performance with an area under the receiver operating curve of this model of 99.7% (95% CI: 99.4-100). The fragment length-based random forest model accurately classified 95.4% (Kappa = 85.9%) of samples into their respective *Trichuris* species. Taken together, our results showed that the fragment length is an accurate diagnostic marker to differentiate between the two human-infecting *Trichuris* species.

Human-infecting *Trichuris* species and metadata associations



By focusing on a sub-group of study participants between the ages of 6-18 years, we sought to understand whether there were differences in the metadata associated with both, *T. incognita* and *T. trichiura*. We did not find any association between age and *Trichuris* species in any of the settings, revealing a homogeneous distribution of both species in the

population traditionally most at risk, school-aged children (**Fig. 5A**). Similarly, we did not find any differences in infection intensity (as a categorical variable defined by WHO) between both species (**Fig. 5B/C**). Finally, we found that in 3 out of 4 settings (Laos, Tanzania, and Uganda), females were more likely to be infected with *T. incognita* than males (**Fig. 5D**).

Discussion

The etiological agent of human trichuriasis has traditionally been identified as *T. trichiura* based on morphological characteristics. However, morphological traits alone are often insufficient to distinguish closely related species, especially when adaptive changes or when species divergence has occurred relatively recently. Molecular epidemiological investigations using coding and non-coding regions of nuclear ribosomal DNA have revealed a second *Trichuris* species infecting humans, closely resembling those found in NHP and pigs (14,16,17). Here, we utilized the nuclear ITS2 region to molecularly identify *Trichuris* species infecting humans from samples from randomized controlled trials across diverse geographical regions.

Our phylogenetic analysis of ASVs revealed that *Trichuris* populations from humans were grouped into two distinct *Trichuris* genotypes across these study sites, which we distinguish as Clade D and Clade E (**Fig. 3A**). Haplotype networks, pairwise FST, and pairwise distance analysis (**Fig. 2A-B**; **Fig. 3B**) corroborated these results, indicating well-differentiated and genetically divergent clades across the study sites. These findings align with previous studies using ITS2 rDNA, demonstrating the multi-host affiliation of *Trichuris* species among humans and NHP (9–11,14–16,28,29).

Clade D contained majorly of the ITS2 ASVs from Côte d'Ivoire indicating the establishment of this clade in this setting, which is also where the first human samples came from that led to the specification of this new species (16,17). However, our study not only incorporates the enhanced resolution of full-length ITS2 sequencing but also significantly broadens the geographical scope including samples from another district Côte d'Ivoire and an entirely new

country, Uganda, thereby enriching the diversity and depth of our data. To the best of our knowledge, this is the first time that *T. incognita* has been described in Tanzania, Laos, and Uganda albeit in a smaller proportion suggesting that this species has been infecting humans and bypassing standard microscopy-based diagnostic techniques across at least two continents with endemic trichuriasis for a significant amount of time.

Furthermore, Clade D shares a monophyletic lineage with the *T. suis* clade, with an estimated 82 to 88% nucleotide sequence similarity. Although both share similar physiological features, including overall morphology, egg production, and range of hosts, the evidence derived from the nucleotide sequence similarity between the clades suggests that *T. incognita* and *T. trichiura* evolved in parallel from a distant common ancestor.

In addition, we showed that Clade D contains sequences from samples collected previously from 13 primate species spanning multiple studies and countries, as well as one human patient in Cameroon. With approximately 98-100% nucleotide identity of the ITS2 region among these sequences, it shows that *T. incognita* consists of a multi-host lineage suggesting that it can infect multiple hosts just like *T. trichiura* thus justifying its potential zoonotic origin.

In contrast to Clade D, Clade E contained over 70% of the ASVs generated from the human samples in our study. These sequences clustered closely with those derived from 36 human and NHP species, as well as one whole-genome sequence (WGS). Clade E sequences conformed to the most common human-infecting species, sharing 96% - 100% nucleotide similarity with sequences from humans in Ecuador and Uganda (9,23). Additionally, they matched sequences from both captive and non-captive NHP in Uganda, China, South Africa, Italy, Spain, and St. Kitts & Nevis (10,11,14,26,27). This result is in agreement with a recently published study using ITS2 rDNA (16) and further supports that whipworms in primates comprise a *Trichuris* species complex. Additionally, within this clade, results based on ITS2 revealed an interpopulation similarity of approximately 79% to 86% between *T. trichiura* populations and *Trichuris sp.* from Francois Leaf monkeys, hence, agreeing with Liu

et al. (26) that *Trichuris* populations from Francois Leaf monkeys potentially represent a new *Trichuris* species (**Fig. 3A-B**).

Despite earlier theories suggesting zoonotic transmission between pigs, humans, and NHP from an African origin to Asia and subsequently to South America (30,31), recent evidence suggests limited transmission influenced by natural host dispersion and anthropogenic factors in recent evolutionary history (30,32). Furthermore, comparative analyses of gene duplication, gain, and loss patterns in *T. incognita* compared to *T. suis* and *T. trichiura* indicate a more ancient and divergent lineage (17). All these observations collectively indicate a complex evolutionary history, resulting in a rich diversity of co-evolving species within the *Trichuris* genus. With reports of drug resistance emerging in the veterinary world (33–37), coupled with the recent discovery of *T. incognita* as a result of ivermectin-albendazole unsatisfactory drug efficacy in Côte d'Ivoire compared to other settings (16) as well as expected resistance in human trichuriasis within the next decade (38), it is critical to understand the distribution, transmission of *Trichuris* etiologies and potential factors for reduced efficacy or drug resistance including parasite-related factors (strain analysis). This understanding is essential to sustain the development of effective control strategies and achieve the WHO's 2030 goal of disease elimination as a public health problem (39).

To that end, while next-generation sequencing poses challenges for adopting species-level diagnostic tools in resource-constrained environments, we have identified a potential PCR-based diagnostic marker that can accurately differentiate between the two human-infecting *Trichuris* species, *T. incognita* and *T. trichiura*. This marker simplifies the detection process, eliminating the need for expensive sequencing platforms, and establishes a cost-effective, robust method for distinguishing these morphologically similar species. By using the length of ITS2 fragments, a random forest model validated its diagnostic utility, achieving a 95.4% classification accuracy (Kappa = 85.9%). This demonstrates the reliability of fragment length as a marker in epidemiological studies and control programs, complementing previous findings on ITS2 rDNA's effectiveness in distinguishing closely related nematode species

(11,13,16,28,40–42). The potential use of ITS2 length as a diagnostic marker is particularly suitable for resource-constrained environments and could be combined with PCR or further developed into LAMP assays. While MinION sequencing is relatively affordable and allows high multiplexing with our approach, a length-based diagnostic tool is simpler and more accessible for widespread application (43–46).

Our study has a few drawbacks. First, our observations are based on samples collected from clinical studies in Côte d'Ivoire, Laos, Tanzania, and Uganda, which may not fully capture the global diversity of the newly identified human-infecting *Trichuris* species due to the limited geographical scope. Second, our dependence on nuclear ribosomal DNA for species identification, without comprehensive morphological analysis or whole-genome sequencing, might miss other genetic variations and phenotypic traits. Furthermore, we did not evaluate temporal aspects of infections, such as seasonal variations or long-term changes, and our study did not include an extensive analysis of environmental factors that could influence the distribution and genetic diversity of *Trichuris* species. Importantly, the samples were collected within the context of clinical trials but these represented the status before drug use hence we did not explore how genetic diversity might relate to decreased drug efficacy. While we believe our findings are robust, further studies with broader geographical coverage, and integrated genomic, morphological, and environmental approaches are needed to expand upon these findings.

In conclusion, our study utilized an enhanced resolution of full-length ITS2 rDNA sequencing to identify two distinct human-infecting *Trichuris* species, *T. incognita* and *T. trichiura*, across diverse regions. Our research expands the geographical scope of previous studies, including new regions and countries such as Uganda. We also identified *T. incognita* in Tanzania, and Laos, for the first time, reinvigorating the delimitations of traditional diagnostics in distinguishing between different *Trichuris* species. Additionally, we identified a PCR-based marker using ITS2 fragment length as a powerful diagnostic marker for distinguishing human-infecting *Trichuris* species. This method demonstrates a simplified, cost-effective,

and robust alternative to expensive sequencing approaches and is especially suitable for resource-limited settings. Furthermore, it promises improved epidemiological studies and control programs, hence, aiding the WHO 2030 trichuriasis elimination goal. Future research should refine this diagnostic method and explore other genetic markers in conjunction with ITS2 fragment length to enhance diagnostic precision and provide deeper insights into the evolutionary relationships within the *Trichuris* genus.

Methods

Study design and sample collection

The data and samples used in this study were derived from two studies. The first study was a double-blind, placebo-controlled randomized trial conducted between September 2018 and June 2020 in Côte d'Ivoire, Laos, and Tanzania. The study examined the efficacy and safety of co-administered albendazole and ivermectin versus albendazole monotherapy against whipworm infections in children and adults with the rationale, procedures, and results of the study published previously (18,21,22).

Briefly, a fresh faecal sample was collected from community members aged between 6 and 60 years identified as infected with *T. trichiura* based on a prior stool examination and considered eligible for the trial with an infection intensity of at least 100 eggs per gram of stool. From every sample, about 1g of faeces was transferred into a 2ml screw cap cryotube using a UV sterilized plastic spatula and immediately frozen at -20°C, which were then shipped to Swiss TPH on dry ice at the end of the trial. Samples considered for this analysis represent baseline stool samples provided before any drug administration.

Additionally, data was also derived from a parallel open-label randomized controlled superiority trial conducted between October to November 2023 in two primary schools in Kabale and Kisoro districts, southwestern Uganda among individuals aged 6-12 years to investigate the superiority of co-administered ivermectin plus albendazole to albendazole monotherapy in terms of cure rates against *T. trichiura* infections (registered as

NCT06037876; clinicaltrials.gov/). At baseline screening of the trial, two stool samples were collected at baseline and children found to be infected with *T. trichiura* based on quadruplicate Kato-Katz thick smear readings were enrolled into the study. A portion of the second baseline stool sample (1.5-2 g) from 135 eligible participants was preserved in 95% ethanol, shipped to Swiss TPH in Basel and subjected to amplicon deep sequencing for characterization of *Trichuris* species. Faecal egg count and epidemiological data are available in **(Additional file 2: Table S3)**.

DNA isolation and sample processing

Two separate extraction approaches were used in the isolation of DNA from stool received from the multi-country study and the study from Uganda. For the multi-country study, a total of 658 (226, 191, and 221 respectively from Côte d'Ivoire, Laos, and Tanzania) frozen 100 mg of stool samples from the baseline were extracted using a QiAmp PowerFecal DNA extraction kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol with slight modification. For the faecal samples collected in Southwestern Uganda, 150 - 250mg of samples preserved in Ethanol was used in the extraction of the DNA. Before extraction with the QiAmp PowerFecal DNA extraction kit, a few modification steps were taken to mechanically rupture the *Trichuris* egg shells which include the removal of ethanol from the stool, washing with phosphate buffer saline (PBS), as well as a freeze-thaw-boiling step to enhance egg shell rupture and minimize inhibition (16,47) (see **Additional file 3: Appendix S1** for more details of the modified ethanol-preserved sample extraction protocol).

PCR Amplification, Quantification and Purification

Primers used in the ITS2 rDNA reactions were Trichuris_ITS2_F 5'-ATGTCGACGCTACGCCTGTC-3' and Trichuris_ITS2_R 5'-TAGCCTCGTCTGATCTGAGG-3' (16) with a reverse primer containing an added custom barcodes that were adapted from the Illumina 12bp barcode sequences (48,49) (See **Additional file 2: Table S4** for full primer sequences). For each run, 96-well plates were multiplexed with the Trichuris_ITS2_F

primers and eight custom barcoded reverse primers to amplify the ITS2- regions yielding unique barcodes of 592 bp ITS2 regions for each sample per plate column. LongAmp® Hot Start Taq 2X Master Mix (cat no. M0533S, New England BioLabs, USA) and 2µL of source DNA were used in each amplification. Thermocycling conditions included the following: 94°C for 30 sec initially, 40 cycles of 94°C for 30 sec, 65°C for 15 sec, 65°C for 50 sec, followed by a final extension at 65°C for 10 min carried out on an Eppendorf Mastercycler Nexus Gradient Thermal Cycler (Eppendorf AG, Hamburg, Germany). Barcoded amplicons were quantified directly using a dsDNA HS Assay kit on a Qubit 4 Fluorometer (Invitrogen, USA) according to the manufacturer's protocol to ensure even throughput in the DNA pool using 2 µL of PCR reaction as input. 1% agarose and capillary gel electrophoresis (100V, 20min) were used to verify the amplification of the fragment at the expected band size. Afterwards, sample plates were stored overnight for a short period at 4°C overnight or at -20°C for long-term storage.

Library preparation and sequencing

Nanopore sequencing libraries were prepared using the Native Barcoding Kit 24 v14 (SQK-NBD114.24, Oxford Nanopore Technologies, UK) according to the Native Barcoding Genomic DNA protocol with slight modifications aiming to normalize sample input, allowing for multiplexing as described in more details elsewhere (49). Briefly, samples were pooled column-wise from a 96-well plate before the end-repair step, requiring 12 different outer barcodes for every 8 samples. Each column, representing 8 samples, was processed together with approximately 80 ng input corresponding to 200 fmol of the expected 0.59 kb fragments for the end-repair steps. Following end repair, samples were assessed using the 1X dsDNA HS Assay kit on a Qubit 4 fluorometer. Equimolar masses of the pooled reactions were used for native barcode ligation to ensure equal read numbers per barcode. Adapter ligation and clean-up followed the protocol, utilizing a short-fragment buffer to preserve the ITS2 rDNA fragments. Afterwards, 1 µl of the prepared library was quantified on the Qubit 4.0 (Invitrogen, USA) using the Qubit dsDNA HS Assay Kit (Invitrogen, USA). The protocol

from ONT recommends loading 10-20 fmol of amplicon product onto the Flongle Flow Cell (Flongle Flow Cell, R10 Version, FLO-FLG114, Oxford Nanopore Technologies, UK), so we diluted to a concentration of 2-3 ng/µL in Elution Buffer to load approximately 15 fmol of the library onto the Flongle Flow Cell. The Flongle Flow Cell was primed for loading by flushing the flow cell with 120 µL priming mix (3 µl of Flush Tether in 117 µL of Flow Cell Flush), taking care to avoid the introduction of air bubbles. The library was prepared for loading by mixing 15 µl Sequencing Buffer, 10 µl Loading Beads and 5 µl diluted DNA library, after which the sample was added to a Flongle Flow Cell on a MinION Mk1C (Oxford Nanopore Technologies, UK) and sequenced at depth for 24 h

Read processing and sequencing analysis

All POD5 data resulting from the sequencing of the amplicons were already basecalled and demultiplexed based on the native ONT barcode in real-time on the MinKNow v3.6.5 and the built-in Guppy (v6.5.7) using the high accuracy basecalling model. The base-called POD5 files were then re-run through Dorado (v. 0.5.1; base-caller model dna_r10.4.1_e8.2_400bps_sup@v4.3.0) generating high-quality simplex reads) which was fed through our custom demultiplexing scripts that combine seqkit (v.2.6.1) (50) and custom Python scripts available at https://github.com/STPHxBioinformatics/ITS2_Trichuriasis_nanopore_processing. The script demultiplexes, quality filters, and length filters, as well as utilizes a minimum and maximum, trimmed-read cutoff of 450 bp and 750 bp, respectively with zero mismatches. Consequently, it trims the reads based on the inner PCR barcode (the reverse barcode in this case) ensuring the barcodes are found in the correct orientation. NanoPlot v1.30.1 (51) was then used for descriptive statistics on the runs. Demultiplexed ITS2 reads were then subjected to the R package 'DADA2' version 1.3.2 (52) pipeline to include further filtering to remove reads containing unresolved nucleotides (maxN = 0) as well as reads exceeding the expected error number (maxEE = 15) and size range (425–700 bp). The dataset generated was then used as input to define the error rates and perform the removal of

identical reads (derepFastq) while inferring the composition of the sequenced pool using the dereplicated sequence dataset as input (dadaFs), and removal of chimeric sequences. Following this, Amplicon sequence Variants (ASVs) generated through the DADA2 pipeline with ASVs below 5 reads across all samples were removed.

Genetic Variation and Phylogenetic Analysis

A curated database of reference ITS2 sequences was created by downloading all publicly available full ITS1-5.8S-ITS2 reference sequences of closely related species on NCBI using the search term "Trichuris"[porgn:__txid36086] as well as from the WGS data of *Trichuris muris* (23,24), *Trichuris trichiura* (23), *Trichuris suis* (25) and an adult worm from C□te d'Ivoire (17). Afterwards, seqkit (50) was used to extract only ITS2 sequences within the forward and reverse primers used in this study with the command seqkit -j 4 amplicon –m 2 -p primers.tab -r 1:-1 20240201_Trichuris.txid_36086.fasta --bed > 20240201_Trichuris_txid_36086.bed. The ASVs were then subjected to multiple sequence alignments with the newly curated reference sequences **(Additional file 2: Table S5)**, as well as an outgroup species *Trichinella spiralis* (Accession No: KC006432), using the MAFFT tool (Katoh et al., 2019). Phylogenetic relatedness was inferred using NJ in MEGA v11.0 (53) and the Maximum Likelihood method using RaxML v8.0 (54) employing the jModelTest v0.1.1 (55) to choose the best-to-fit model.

Population Genetic Structure

Alignment from good quality ITS2 ASVs was used for population genetic data and haplotype polymorphic analysis such as numbers of variable sites, number of haplotypes, nucleotide diversity (π), expected heterozygosity, and pairwise FST calculations between the *Trichuris* populations in the four countries were calculated using DnaSP v6.0 (56) and Arlequin v3.5.2 (57) using 1000 permutations. Statistical Parsimonious networks were inferred and visualized using Population Analysis with Reticulate Trees, POPART v1.7 (58) with the connection limit set to 95% and gaps being treated as missing. Furthermore, the number of

nucleotide differences per sequence in each country and other *Trichuris* species incorporated in the phylogenetic analysis was analyzed using the Compute Pairwise Distance prompt based on the number of differences method of MEGA v11 (53) to assess the sequence similarity.

Availability of data and materials

The custom demultiplexing script and DADA2 analysis code used for this study are available at: https://github.com/STPHxBioinformatics/HITS2. The raw ONT sequence data (ITS2 rDNA amplicon sequencing) generated in this study is available in the NCBI Short Read Archive under accession (PRJNA1141773: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA547591).

Acknowledgements

We want to acknowledge the participants and investigators of the IVM-ALB multi-country and FACE-IT efficacy trial study. The authors also thank all the Keiser's lab members for their valuable discussions. Computational analyses were performed using the sciCORE (https://scicore.unibas.ch) scientific computing center at the University of Basel. We are grateful to the European Research Council (No. 101019223) for financial support.

References

1. Bethony J, Brooker S, Albonico M, Geiger SM, Loukas A, Diemert D, et al. Soil-transmitted helminth infections: ascariasis, trichuriasis, and hookworm. The Lancet. 2006 May 6;367(9521):1521–32.

2. Pullan RL, Smith JL, Jasrasaria R, Brooker SJ. Global numbers of infection and disease burden of soil transmitted helminth infections in 2010. Parasites Vectors. 2014 Jan 21;7(1):37.

3. Bundy DAP, Cooper ES. *Trichuris* and Trichuriasis in Humans. In: Baker JR, Muller R, editors. Advances in Parasitology [Internet]. Academic Press; 1989 [cited 2024 Jul 18]. p. 107–73. Available from: https://www.sciencedirect.com/science/article/pii/S0065308X08603322

4. Callejón R, Robles MDR, Panei CJ, Cutillas C. Molecular diversification of Trichuris spp. from Sigmodontinae (Cricetidae) rodents from Argentina based on mitochondrial DNA sequences. Parasitol Res. 2016 Aug 1;115(8):2933–45.

5. Cutillas C, de Rojas M, Zurita A, Oliveros R, Callejón R. Trichuris colobae n. sp. (Nematoda: Trichuridae), a new species of Trichuris from Colobus guereza kikuyensis. Parasitol Res. 2014 Jul 1;113(7):2725–32.

6. Callejón R, Halajian A, Cutillas C. Description of a new species, *Trichuris ursinus* n. sp. (Nematoda: Trichuridae) from *Papio ursinus* Keer, 1792 from South Africa. Infection, Genetics and Evolution. 2017 Jul 1;51:182–93.

7. Oliveros R, Cutillas C, De Rojas M, Arias P. Characterization of four species of Trichuris (Nematoda: Enoplida) by their second internal transcribed spacer ribosomal DNA sequence. Parasitol Res. 2000 Nov 1;86(12):1008–13.

8. Ooi HK, Tenora F, Itoh K, Kamiya M. Comparative Study of Trichuris trichiura from Non-Human Primates and from Man, and Their Difference with T. suis. Journal of Veterinary Medical Science. 1993;55(3):363–6.

9. Ghai RR, Simons ND, Chapman CA, Omeja PA, Davies TJ, Ting N, et al. Hidden Population Structure and Cross-species Transmission of Whipworms (Trichuris sp.) in Humans and Non-human Primates in Uganda. Gasser RB, editor. PLoS Negl Trop Dis. 2014 Oct 23;8(10):e3256.

10. Rivero J, Cutillas C, Callejón R. Trichuris trichiura (Linnaeus, 1771) From Human and Non-human Primates: Morphology, Biometry, Host Specificity, Molecular Characterization, and Phylogeny. Front Vet Sci. 2021 Feb 9;7:626120.

11. Cavallero S, De Liberato C, Friedrich KG, Di Cave D, Masella V, D'Amelio S, et al. Genetic heterogeneity and phylogeny of Trichuris spp. from captive non-human primates based on ribosomal DNA sequence data. Infection, Genetics and Evolution. 2015 Aug;34:450–6.

12. Cavallero S, Montalbano Di Filippo M, Rondón S, Liberato CD, D'Amelio S, Friedrich KG, et al. Nuclear and Mitochondrial Data on Trichuris from Macaca fuscata Support Evidence of Host Specificity. Life. 2020 Dec 31;11(1):18.

13. Nissen S, Al-Jubury A, Hansen TVA, Olsen A, Christensen H, Thamsborg SM, et al. Genetic analysis of Trichuris suis and Trichuris trichiura recovered from humans and pigs in a sympatric setting in Uganda. Veterinary Parasitology. 2012 Aug;188(1–2):68–77.

14. Ravasi DF, O'Riain MJ, Davids F, Illing N. Phylogenetic Evidence That Two Distinct Trichuris Genotypes Infect both Humans and Non-Human Primates. Parkinson J, editor. PLoS ONE. 2012 Aug 28;7(8):e44187.

15. Xie Y, Zhao B, Hoberg EP, Li M, Zhou X, Gu X, et al. Genetic characterisation and phylogenetic status of whipworms (Trichuris spp.) from captive non-human primates in China, determined by nuclear and mitochondrial sequencing. Parasites Vectors. 2018 Dec;11(1):516.

16. Venkatesan A, Chen R, Baer M, Schneeberger PHH, Reimer B, Hürlimann E, et al. Discovery of a novel albendazole-ivermectin refractory Trichuris species, T. incognita, in Côte d'Ivoire patients through fecal DNA metabarcoding [Internet]. bioRxiv; 2024 [cited 2024 Jul 16]. p. 2024.07.04.602123. Available from: https://www.biorxiv.org/content/10.1101/2024.07.04.602123v2

17. Bär MA, Kouamé NA, Touré S, Coulibaly JT, Schneeberger PHH, Keiser J. Genetic makeup of *Trichuris incognita* , a novel *Trichuris* species naturally infecting humans and displaying signs of resistance to drug treatment [Internet]. 2024 [cited 2024 Jul 30]. Available from: http://biorxiv.org/lookup/doi/10.1101/2024.06.11.598441

18. Hürlimann E, Keller L, Patel C, Welsche S, Hattendorf J, Ali SM, et al. Efficacy and safety of co-administered ivermectin and albendazole in school-aged children and adults infected with Trichuris trichiura in Côte d'Ivoire, Laos, and Pemba Island, Tanzania: a double-blind, parallel-group, phase 3, randomised controlled trial. The Lancet Infectious Diseases. 2022 Jan;22(1):123–35.

19. Hürlimann E, Palmeirim, Marta. Efficacy of Ivermectin-albendazole vs Albendazole Alone in School-aged Children Infected With Trichuris Trichiura in Uganda. 2024.

20. Keller L, Hürlimann E, Patel C, Welsche S, Sayasone S, Ali SM, et al. Insights gained from conducting a randomised controlled trial on Ivermectin-Albendazole against *Trichuris trichiura* in Côte d'Ivoire, Lao PDR and Pemba Island. In: Rollinson D, Stothard JR, editors. Advances in Parasitology [Internet]. Academic Press; 2021 [cited 2024 Jul 25]. p. 253–76. Available from: https://www.sciencedirect.com/science/article/pii/S0065308X2030110X

21. Keller L, Welsche S, Patel C, Sayasone S, Ali SM, Ame SM, et al. Long-term outcomes of ivermectin-albendazole versus albendazole alone against soil-transmitted helminths: Results from randomized controlled trials in Lao PDR and Pemba Island, Tanzania. Fairfax KC, editor. PLoS Negl Trop Dis. 2021 Jun 30;15(6):e0009561.

22. Patel C, Hürlimann E, Keller L, Hattendorf J, Sayasone S, Ali SM, et al. Efficacy and safety of ivermectin and albendazole co-administration in school-aged children and adults infected with Trichuris trichiura: study protocol for a multi-country randomized controlled double-blind trial. BMC Infect Dis. 2019 Dec;19(1):262.

23. Foth BJ, Tsai IJ, Reid AJ, Bancroft AJ, Nichol S, Tracey A, et al. Whipworm genome and dual-species transcriptome analyses provide molecular insights into an intimate host-parasite interaction. Nat Genet. 2014 Jul;46(7):693–700.

24. Howe KL, Bolt BJ, Shafie M, Kersey P, Berriman M. WormBase ParaSite − a comprehensive resource for helminth genomics. Molecular and Biochemical Parasitology. 2017 Jul 1;215:2–10.

25. Jex AR, Nejsum P, Schwarz EM, Hu L, Young ND, Hall RS, et al. Genome and transcriptome of the porcine whipworm Trichuris suis. Nat Genet. 2014 Jul;46(7):701–6.

26. Liu GH, Gasser RB, Nejsum P, Wang Y, Chen Q, Song HQ, et al. Mitochondrial and Nuclear Ribosomal DNA Evidence Supports the Existence of a New Trichuris Species in the Endangered François' Leaf-Monkey. Janke A, editor. PLoS ONE. 2013 Jun 20;8(6):e66249.

27. Yao C, Walkush J, Shim D, Cruz K, Ketzis J. Molecular species identification of Trichuris trichiura in African green monkey on St. Kitts, West Indies. Vet Parasitol Reg Stud Reports. 2018 Jan;11:22–6.

28. Liu GH, Zhou W, Nisbet AJ, Xu MJ, Zhou DH, Zhao GH, et al. Characterization of *Trichuris trichiura* from humans and *T. suis* from pigs in China using internal transcribed spacers of nuclear ribosomal DNA. J Helminthol. 2014 Mar;88(1):64–8.

29. Sirima C, Bizet C, Hamou H, Červená B, Lemarcis T, Esteban A, et al. Soil-transmitted helminth infections in free-ranging non-human primates from Cameroon and Gabon. Parasites Vectors. 2021 Dec;14(1):354.

30. Hawash MBF, Betson M, Al-Jubury A, Ketzis J, LeeWillingham A, Bertelsen MF, et al. Whipworms in humans and pigs: origins and demography. Parasites Vectors. 2016 Dec;9(1):37.

31. Nejsum P, Betson M, Bendall RP, Thamsborg SM, Stothard JR. Assessing the zoonotic potential of Ascaris suum and Trichuris suis: looking to the future from an analysis of the past. Journal of Helminthology. 2012 Jun;86(2):148–55.

32. Hawash MBF, Andersen LO, Gasser RB, Stensvold CR, Nejsum P. Mitochondrial Genome Analyses Suggest Multiple Trichuris Species in Humans, Baboons, and Pigs from Different Geographical Regions. Lok JB, editor. PLoS Negl Trop Dis. 2015 Sep 14;9(9):e0004059.

33. Beleckė A, Kupčinskas T, Stadalienė I, Höglund J, Thamsborg SM, Stuen S, et al. Anthelmintic resistance in small ruminants in the Nordic-Baltic region. Acta Veterinaria Scandinavica. 2021 Apr 27;63(1):18.

34. Fairweather I, Brennan GP, Hanna REB, Robinson MW, Skuce PJ. Drug resistance in liver flukes. International Journal for Parasitology: Drugs and Drug Resistance. 2020 Apr 1;12:39–59.

35. Kelleher AC, Good B, de Waal T, Keane OM. Anthelmintic resistance among gastrointestinal nematodes of cattle on dairy calf to beef farms in Ireland. Irish Veterinary Journal. 2020 Jul 1;73(1):12.

36. Rose Vineer H. What Modeling Parasites, Transmission, and Resistance Can Teach Us. Veterinary Clinics of North America: Food Animal Practice. 2020 Mar 1;36(1):145–58.

37. Kotze AC, Gilleard JS, Doyle SR, Prichard RK. Challenges and opportunities for the adoption of molecular diagnostics for anthelmintic resistance. International Journal for Parasitology: Drugs and Drug Resistance. 2020 Dec 1;14:264–73.

38. Else KJ, Keiser J, Holland CV, Grencis RK, Sattelle DB, Fujiwara RT, et al. Whipworm and roundworm infections. Nat Rev Dis Primers. 2020 May 28;6(1):44.

39. World Health Organization. Ending the neglect to attain the Sustainable Development Goals: A road map for neglected tropical diseases 2021–2030 [Internet]. Geneva: World Health Organization; 2021 Jan [cited 2024 Jul 22]. Available from: https://www.who.int/publications/i/item/9789240010352

40. Ceballos-Mendiola G, Valero A, Polo-Vico R, Tejada M, Abattouy N, Karl H, et al. Genetic variability of Anisakis simplex s.s. parasitizing European hake (Merluccius merluccius) in the Little Sole Bank area in the Northeast Atlantic. Parasitol Res. 2010 Nov 1;107(6):1399–404.

41. Cutillas C, Callejón R, de Rojas M, Tewes B, Ubeda JM, Ariza C, et al. *Trichuris suis* and *Trichuris trichiura* are different nematode species. Acta Tropica. 2009 Sep 1;111(3):299–307.

42. Wickramasinghe S, Yatawara L, Rajapakse RPVJ, Agatsuma T. Toxocara canis and Toxocara vitulorum: molecular characterization, discrimination, and phylogenetic analysis based on mitochondrial (ATP synthase subunit 6 and 12S) and nuclear ribosomal (ITS-2 and 28S) genes. Parasitol Res. 2009 May 1;104(6):1425–30.

43. George S, Geldhof P, Albonico M, Ame SM, Bethony JM, Engels D, et al. Molecular speciation of soil-transmitted helminths egg isolates collected during six drug efficacy trials in endemic countries. Trans R Soc Trop Med Hyg. 2016 Nov 1;110(11):657–63.

44. Hassan NA, Noor Badi FA, Mohd-Shaharuddin N, Wan Yusoff WS, Lim Y a. L, Chua KH, et al. A conventional multiplex PCR for the detection of four common soil-transmitted nematodes in human feces: development and validation. Trop Biomed. 2022 Mar 1;39(1):135–43.

45. Alomashi GBA, Khudhur HR, Jebur LS. Conventional Multiplex PCR for Identifying Soil Transmitted Helminthes (*Ascaris Lumbricoides*, *Trichuris Trichura* and *Ancylostoma Duodenale*) in Fecal Specimens. Nano Biomedicine and Engineering. 2022 Dec 31;14(4):289–94.

46. Kaisar MMM, Brienen E a. T, Djuardi Y, Sartono E, Yazdanbakhsh M, Verweij JJ, et al. Improved diagnosis of Trichuris trichiura by using a bead-beating procedure on ethanol preserved stool samples prior to DNA isolation and the performance of multiplex real-time PCR for intestinal parasites. Parasitology. 2017 Jun;144(7):965–74.

47. Ayana M, Cools P, Mekonnen Z, Biruksew A, Dana D, Rashwan N, et al. Comparison of four DNA extraction and three preservation protocols for the molecular detection and quantification of soil-transmitted helminths in stool. Cantacessi C, editor. PLoS Negl Trop Dis. 2019 Oct 28;13(10):e0007778.

48. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. The ISME Journal. 2012 Aug 1;6(8):1621–4.

49. Dommann J, Kerbl-Knapp J, Albertos Torres D, Egli A, Keiser J, Schneeberger PHH. A novel barcoded nanopore sequencing workflow of high-quality, full-length bacterial 16S amplicons for taxonomic annotation of bacterial isolates and complex microbial communities. mSystems. 2024 Sep 10;0(0):e00859-24.

50. Shen W, Le S, Li Y, Hu F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. Zou Q, editor. PLoS ONE. 2016 Oct 5;11(10):e0163962.

51. De Coster W, Rademakers R. NanoPack2: population-scale evaluation of long-read sequencing data. Alkan C, editor. Bioinformatics. 2023 May 4;39(5):btad311.

52. Callahan BJ, Wong J, Heiner C, Oh S, Theriot CM, Gulati AS, et al. High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. Nucleic Acids Research. 2019 Oct 10;47(18):e103–e103.

53. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Battistuzzi FU, editor. Molecular Biology and Evolution. 2021 Jun 25;38(7):3022–7.

54. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014 May 1;30(9):1312–3.

55. Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T. ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. Crandall K, editor. Molecular Biology and Evolution. 2020 Jan 1;37(1):291–4.

56. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, et al. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. Molecular Biology and Evolution. 2017 Dec 1;34(12):3299–302.

57. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources. 2010 May;10(3):564–7.

58. Leigh JW, Bryant D. popart: full-feature software for haplotype network construction. Methods in Ecology and Evolution. 2015;6(9):1110–6.