Phylogenetic placement, whole genome assembly, and genome-wide association study of *Trichuris incognita*: a novel human-infecting species causing trichuriasis, with low sensitivity to albendazole-ivermectin

Max A. Bär [1,2], Nadège A. Kouamé [3,4], Sadikou Touré [4], Jean T. Coulibaly [1,2,3,4], Pierre H. H. Schneeberger [†] [1,2], Jennifer Keiser [†] [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, Cote d'Ivoire

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

Abstract

Trichuriasis is a neglected tropical disease that affects as many as 500 million individuals and can cause significant morbidity, even death. For decades, it is thought to be caused by one species of whipworm, *Trichuris trichiura*. Discrepancies in response rates to benzimidazole treatment for trichuriasis exist across clinical trial reports yet the underlying reasons are unclear, and the emergence of anthelmintic resistance is predicted to be established within the next decade. Here, we present the discovery of a novel *Trichuris* species, *Trichuris incognita*, causing trichuriasis, and we investigate variants in context of drug resistance. Combining long- and short- read sequencing approaches, we assembled a high-quality reference genome of *Trichuris incognita* to construct the species tree of the *Trichuris* genus using 12434 orthologous groups and we sequenced a total of 747 individual worms which confirmed the phylogenetic placement. Comparative genomic analysis of genes implied in the resistance of either albendazole or ivermectin in helminths revealed a high copy number of the beta-tubulin (TUBB) gene, present in the whole population, compared to the canonical *T. trichiura* species. A genome wide association study comparing 181 albendazole-ivermectin sensitive to 544 drug non-sensitive worms implied a HERC4 gene with genome wide significance, which has been reported to be downregulated by ivermectin. Our results demonstrate that trichuriasis can be caused by multiple species of whipworm, challenging the long-held belief that it is exclusively due to *Trichuris trichiura*. This discovery, coupled with decreased sensitivity to the best available anthelmintic treatment — a combination of albendazole and ivermectin — marks a significant shift in how we understand and approach whipworm infections. We anticipate our findings to have a profound impact on how we understand whipworm infections with implications on all present and future de-worming strategies across the globe.

Introduction

Soil-transmitted helminthiasis is the most prevalent neglected tropical disease (NTD) worldwide. Over one billion people, particularly in impoverished nations, may develop debilitating chronic infections due to one of the four most common soil-transmitted helminths (STHs): *Ascaris lumbricoides*, *Trichuris trichiura*, and the two hookworm species, *Necator americanus* and *Ancylostoma duodenale*.[1] Despite their genetic diversity, these parasites are classified under the umbrella of STHs by the World Health Organization (WHO) and are treated with benzimidazoles, the cornerstone of large-scale deworming efforts through preventative chemotherapy.[2] With more than 3.3 billion tablets distributed up until August

2020, concern regarding benzimidazole resistance extending to humans, is rising. Reports of decreased efficacy and documented resistance to the same drugs in livestock support projections suggesting the emergence of resistance within the next decade.[3,4] To circumvent resistance and improve efficacy against *T. trichiura* the use of a combination therapy of albendazole and ivermectin has been taken forward. However, recent clinical trials with albendazole-ivermectin, revealed alarmingly low cure rates against *T. trichiura* in Côte d'Ivoire, compared to other regions like Laos and Pemba, Tanzania, hinting to a potential emergence of resistance also of the currently best available treatment for STH infections.[5]

Trichuris trichiura, a longstanding companion of humans, that has co-evolved over millennia, has intrigued scientists as both symbiotic ally and evasive parasite, sparking debate over the transition between these roles.[6-8] Whole-genome sequencing data have yielded invaluable insights into drug discovery and vaccine candidates for trichuriasis, as well as into the broader landscape of infectious, inflammatory, and autoimmune diseases affecting both high- and low-income countries.[7,9-12] Notably, three genomes of *Trichuris* species have been published, *T. suis*, *T. muris*, and the human-invasive *T. trichiura*.

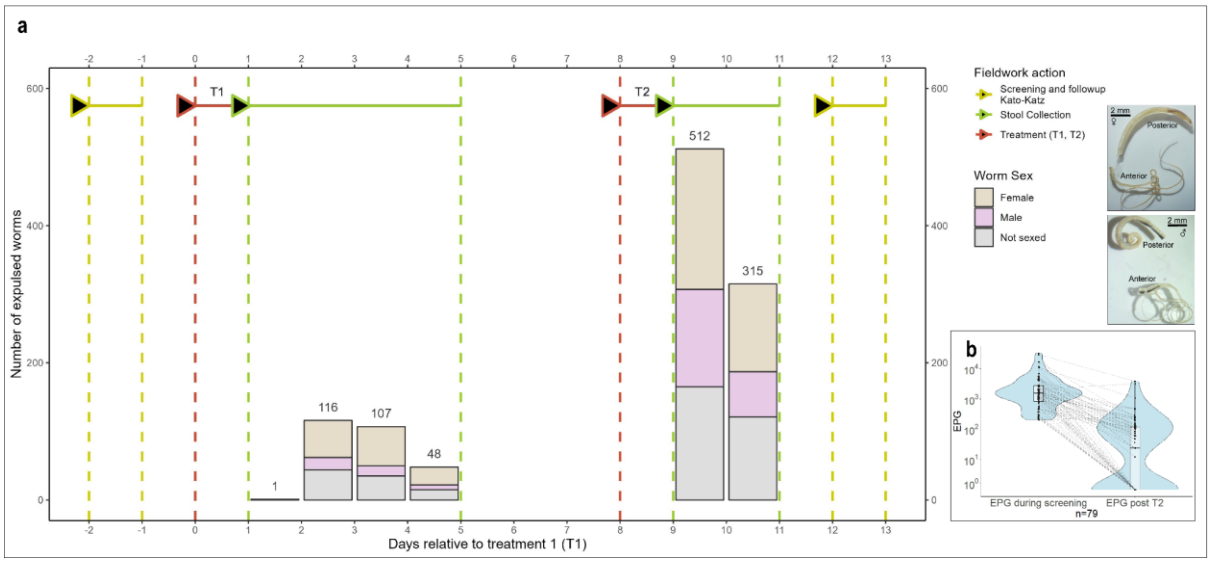
Here, we present genomic evidence supporting the existence of a *Trichuris* species, *Trichuris incognita*, naturally infecting humans in Côte d'Ivoire and displaying signs of resistance to conventional drug treatments. Phylogenetic analysis based on the species tree of 12434 orthologous groups of genes places the isolated worms within a new clade, more closely related to *T. suis* than to *T. trichiura*, limiting the use of available reference genomes. We elucidate the genome of this new species, leveraging whole-genome-sequencing (WGS) data to produce for the first time a high quality reference genome. Comparative genomic analyses offer a path to potential vaccine candidate for trichuriasis and immunotherapeutic agents relevant to Crohn's disease and pancreatitis. Finally, we investigate the whole genome sequencing data of 747 isolates in context of resistance by looking at duplication events across the species tree and conducting a genome wide association study to identify potential single nucleotide polymorphisms comparing drug-sensitive to drug-resistant worms. The GWAS introduces a HERC4 gene, with genome wide significance associated to drug treatment, as a new gene, potentially involved in ivermectin resistance in helminths.

Results

Trichuris expulsion study confirms a new species of whipworm, not responding to conventional albendazole-ivermectin treatment

We screened 670 children in 7 villages Akakro, Ahouya, N'Doumikro, Tiagba, Bekpou and Teffredji in southern Côte d'Ivoire for trichuriasis. 243 participants were enrolled in the study and received a single dose albendazole-ivermectin combination treatment (T1) after which stool was collected daily and screened for worms for up to four days. Subsequently, 8 days after T1, the same individuals received a single oral dose of the trichuricidal drug oxantel-pamoate (T2) after which stool was collected for 2 consecutive days, as presented in Figure 1a. We isolated 271 worms after T1 with an expulsion peak on days 2 and 3 and 827 worms after T2 with expulsion peaking on day 1 and 2 after treatment. A total of 1098 *Trichuris* worms were collected, of which 747 were suitable for genomic analyses. The female to male ratio was 3.42:1 and 1.6:1 during T1 and T2 respectively. A significant correlation ($R^2 = 0.34$; $P = 0.0061$, SI Figure 1) was observed between the number of expulsed worms and the mean egg count per day, yielding a fecundity of 14307 eggs per female worm per day, as presented in SI Figure 1. Figure 1b shows near complete expulsion of all worms as the median EPG dropped from 1524 to 24 at the follow up time point on day 12. To assess relationship of human worms with their local domestic animal pendant, 25 individual worms from 2 swine hosts were collected directly from the caecum. A picture of each worm is provided online.

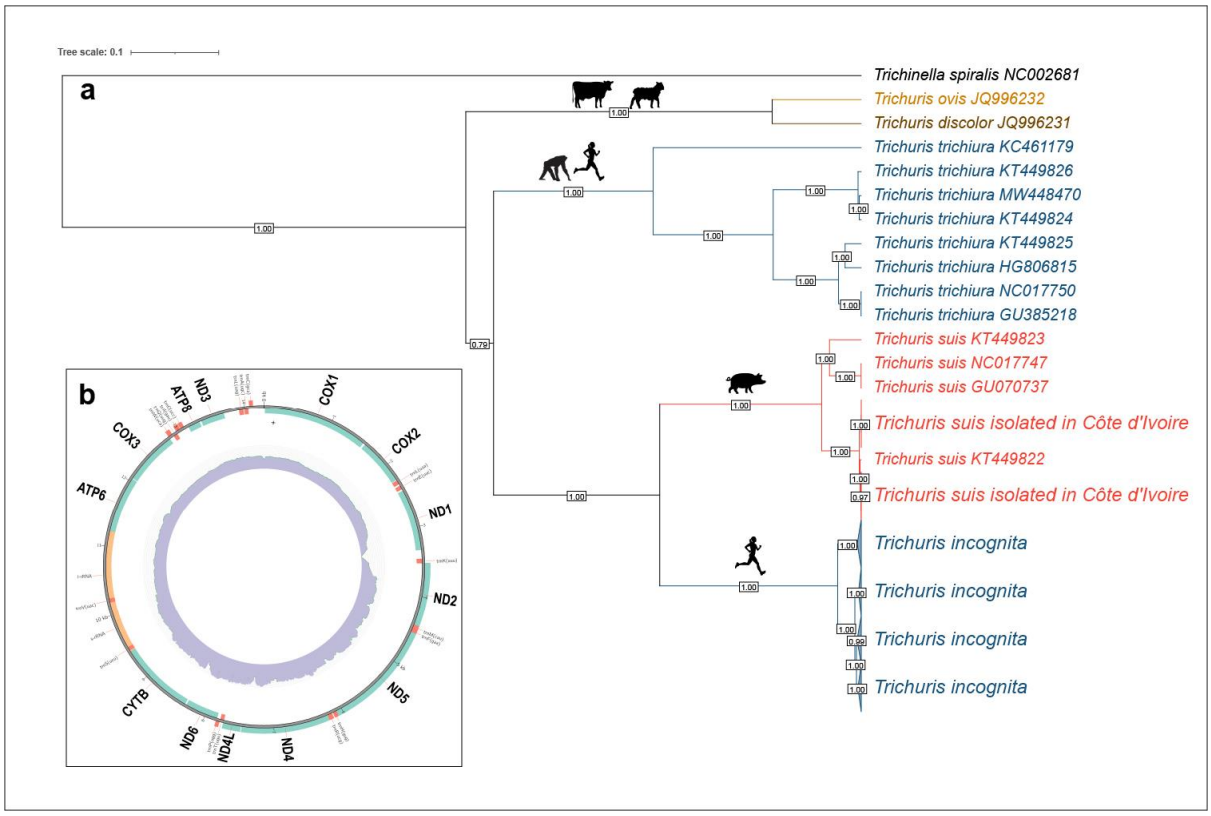
Fig. 1: Timeline of expulsion study including the number of expulsed worms and the female/male distribution of the sequenced worms



a Timeline of expulsion study including the number of expulsed worms and the female/male distribution. T1: albendazole-ivermectin combination treatment, standard-of-care. T2: Oxantel-pamoate treatment. Representative pictures of an expulsed female and male worm are provided on the top right and bottom right respectively. **b** EPG before T1 and after T2.

Next, a phylogenetic analysis of the *Trichuris* worms isolated in Côte d'Ivoire was done using the coding regions of the mitochondrial genomes of 535 individual worms that yielded a complete mitogenome and is presented in Figure 2a. The reconstructed mitogenomes contained 13 protein coding genes (*cox*1-3, *nad*1-6, *nad*4L, *atp*6, *atp*8 and *co*B) as shown in Figure 2b. Concatenated nucleic acid sequences from all assembled and reference mitochondrial genomes showed that the *Trichuris* species encountered in Côte d'Ivoire forms a separate clade from the canonical human infective *T. trichiura*, closer to *T. suis*, while the phylogenetic placement of the 16 sequenced *T. suis* fall within the expected clade of *T. suis* reference mitochondrial genomes. Intraspecies branching was further investigated and presented in SI Figure 2. However, no correlation could be established between treatment, worm sex, host sex, host, coinfection with other worms or village and sub-clade formation within the non-canonical *Trichuris* species. The species is further referred to as *Trichuris incognita*.

Fig. 2: Phylogenetic placement of expulsed *Trichuris* worms from Côte d'Ivoire



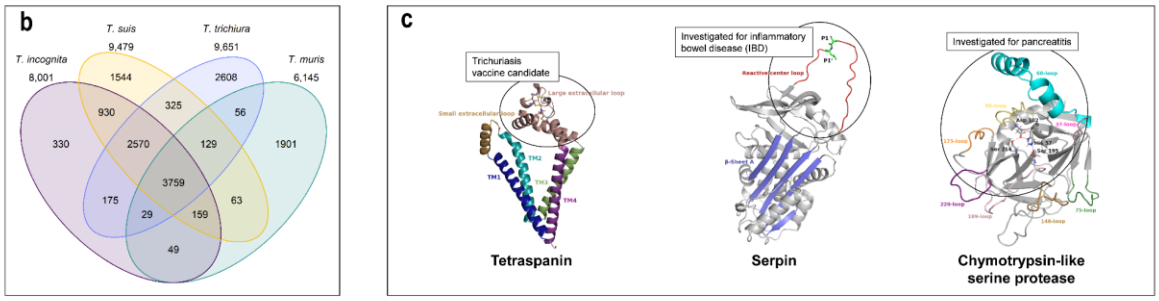
a Bayesian inferred phylogenetic tree based on the amino acid sequences of genes in the mitochondrial genomes of 535 *Trichuris incognita* worms and available reference mitochondrial sequences, using *Trichinella spiralis* as an outgroup. Timescale indicates numbers of substitutions per site. **b** Reconstructed mitochondrial genome of *Trichuris incognita*. Green indicates the genes used for the mitogenome, yellow the ribosomal subunits, red the t-rna's and the coverage displayed in the middle.

T. incognita hybrid *de novo* whole genome assembly and comparative genomics

The *de novo* sequenced genome of *T. incognita* at a coverage of 50X Nanopore and 65X Illumina reads resulted in a length of 84.26 Mb, consisting of 387 contigs with an N50 of 11.78 Mb, and is presented in Figure 3a. The genome completeness amounted to 75.4% of whole, duplicated, and fragmented single copy orthologues as determined by BUSCO. GC content was 43%. Genes were predicted *de novo* for all available reference whole genomes using the braker2 pipeline and the metazoa database, resulting in the prediction of 13739 genes for *T. incognita*.

Fig. 3: Whole genome assembly metrics and comparative genomic analysis of available whole genomes of *Trichuris* species.

a	*Trichuris incognita*	*Trichuris trichiura*	*Trichuris suis*	*Trichuris suis*	*Trichuris suis*	*Trichuris muris*	*Trichinella spiralis*
Accession	NA	PRJEB535	PRJNA208416	PRJNA208415	PRJNA179528	PRJEB126	PRJNA12603
Genome Size [Mb]	84.26	80.57	70.45	73.46	63.84	111.83	63.51
Number of contigs	387	113	1131	1470	306	803	6819
N50 [Mb]	11.78	11.3	0.45	0.5	1.32	28.94	6.37
GC percentage [%]	42.8	42.3	43.5	43.6	43.31	44.6	33.9
Complete BUSCOs [%]	61.3	64.3	65.6	65.4	65.2	61.5	64.8
Duplicated BUSCOs [%]	2.7	1.8	0.4	0.7	0.4	3.9	3.8
Fragmented BUSCOs [%]	11.4	10.0	10.0	9.7	9.9	10.3	7.0
Missing BUSCOs [%]	24.6	23.9	24.0	24.2	24.5	24.3	24.4
Number of genes	13739	11809	11224	11884	9755	20537	13254
Total exon number	82285	71304	69897	71543	66000	93618	103944
Mean exon length [bp]	255	240	235	239	222	281	193
Median exon length [bp]	152	150	150	151	148	164	136
Total intron number	67462	58645	58015	58973	55584	71798	88078
Mean intron length [bp]	267	291	274	278	277	315	187
Median intron length [bp]	57	57	56	57	56	57	74
Total coding sequence [Mb]	20.98	17.14	16.45	17.07	14.63	26.23	20.03
Sequencing technique	Illumina + Nanopore	Pacbio	Illumina	Illumina	Pacbio	Pacbio + Illumina	Whole genome shotgun*



a Whole genome assembly metrics. *bacterial artificial chromosome (BAC) and expressed sequence tag (EST). **b** Venn diagram of orthologous groups from predicted transcripts, where 175 were shared between *T. incognita* and *T. trichiura* and 29 were shared also with *T. muris* and considered. **c** Tetraspanin, Serpin and Chymotrypsin-like serine protease identified as genes of interest from analyzing orthologous groups in Venn diagram sections shared by *T. incognita*, *T. trichiura* and *T. muris*.

Having a *Trichuris* genome, closely related to *T. suis* but from a species naturally infecting humans, poses a unique opportunity to explore genes that might be crucial for the adaptation to a human host. We looked into the occurrence of orthologous groups throughout the genomes of *T. incognita*, *T. suis*, *T. trichiura* and *T. muris* as shown in Figure 3b and present proteins of interest in Figure 3c. Leveraging the close genetic relationship with *T. suis*, predicted transcripts were investigated that are shared with *T. trichiura* but do not occur in *T. suis*. We searched for genes which Gasser et. al identified to be upregulated in the stichosome and potentially have immunomodulatory effects like galactins, serpins, venom allergen–like proteins, apyrase or calreticulin and chymotrypsin-like serine proteases.[13-16] We identified one family of each serpins (OG0000305)[17,18] and chymotrypsin-like serine proteases (OG0010319)[19] that are shared only by *T. trichiura* and *T. incognita*. Additionally, we identified, a family of tetraspanins (OG0009162) which are investigated as vaccine candidates in Schistosomiasis.[20,21] A list of all genes and the interproscan annotations are provided in the supporting information.

The species tree confirms the existence of a new species of *Trichuris*, and reveals highly duplicated genes associated to microtubule dynamics and mitotic processes

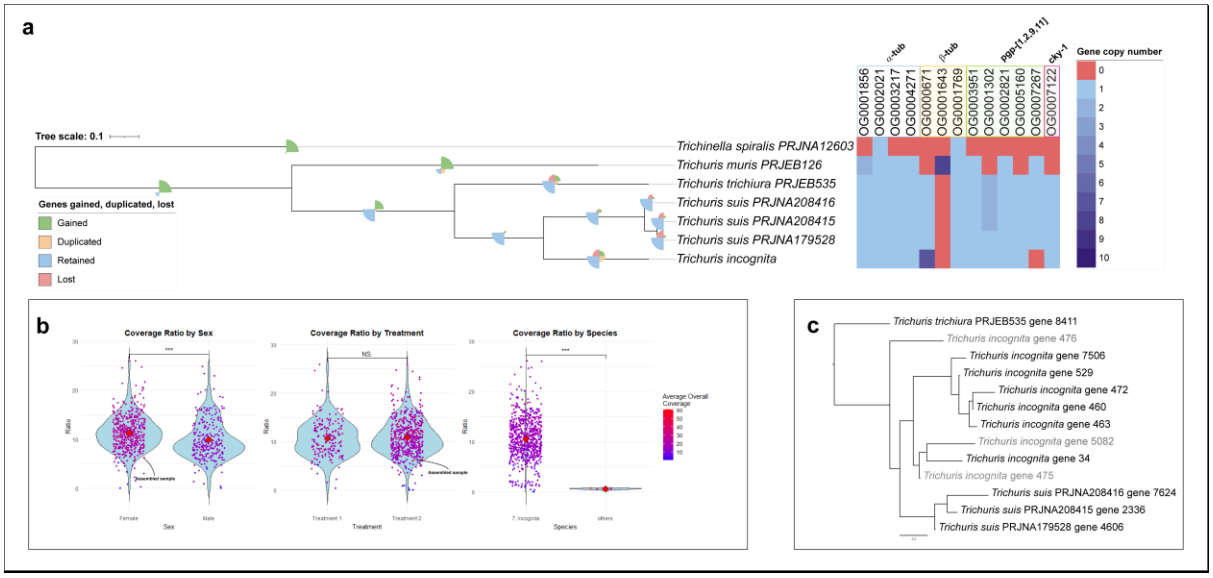
The species tree of the *Trichuris* genus was inferred using 12434 orthologous groups of the braker2 predicted genes in the publicly available whole genomes of *Trichuris* species, *T. incognita* and *Trichinella spiralis*. The species tree shows the same topology as for the mitogenome inferred phylogeny.

We further inferred the gene orthology of all genes predicted in the publicly available whole genomes of *Trichuris* species, *T. incognita* and *Trichinella spiralis* to identify duplication events

of genes that have been associated with resistance to either albendazole or ivermectin in helminths. These resistance-associated genes include β-tubulin (TUBB) for albendazole or glutamate-gated chloride channel subunits (*glc*-1, *glc*-2, *glc*-3, *avr*-14, *avr*-15), voltage gated chloride ion channels (*clh*-3), sodium:potassium:chloride symporter (*nkcc*-1), bestrophin chloride channel (*best*-19), p-glycoprotein ABC transporters (*pgp*-1, *pgp*-2, *pgp*-9, *pgp*-11), and transcription factor *cky*-1, for ivermectin respectively.[22,23] Amongst these genes, one orthologuous group (OG0000671) of the β-tubulin gene showed 9 copies of beta tubulin gene in *T. incognita* and only 1 copy in all *T. suis* and *T. trichiura* reference species as presented in Figure 4a. One gene was fragmented into 2 sections, g475 and g476, and g5082 was also found to be fragmented. Intriguingly, the α-tubulin gene was only present in one copy. SI Figure 3 shows the 7 predicted β-tubulin genes from the braker 2 pipeline. None of the genes show previously encountered Phe168, Glu198 or Phe200 mutations. Position 194, showed the existence of 3 different variants carrying either Ile, Met or a polar Asn. Reference sequences of *T. trichiura*, and *T. suis* also show different variants at this position. *In silico*, a truncation was predicted in 3 of the 7 β-tubulin genes as shown in SI Figure 3. To further investigate the existence of these predicted variants, the region was amplified and sequenced using Nanopore long-read sequencing. As shown in SI Figure 4, the truncation was only present in the *in silico* prediction and further transcriptomic investigations are required. However, the different variants at position 194 were confirmed by the independent sequencing experiment and the coverage ratio indicated that 8% of all reads carried an Asn, 37% carried an Ile, and 55% a Met. Notably, these variants are very close to the predicted binding site of albendazole, and are at variable sites in reference sequences.

The gene tree of the orthologous group OG0000671 shown in Figure 4c gives insight into the history of these duplication events. The gene tree shows a similar distance in number of substitutions per sight from *T. suis* gene 4606 to *T. incognita* gene 34, compared to the distance from *T. incognita* gene 472 to *T. incognita* gene 34. This indicates that interspecies genetic variation is similar to intraspecies genetic variation for at least 2 clades of TUBB genes. Clade one including gene 7506, 529, 472, 460, 463 and clade two including gene 34. Genes 475, 476 and 5082 were not considered, as they were found to be fragmented. Assuming an average mutation rate of 2×10^{-8} point mutations per gene per generation in *C. elegans* as an estimate, the duplication events would be dated to a time before mass drug administration started. However, no data on molecular clocks exists for *T. trichiura* and no evidence to support the suitability of this molecular clock model for this species.

Fig. 4: Duplications across the species tree show a high copy number of β-tubulin genes.



a Species tree indicating gene duplication, gains, losses and retention as pie charts across the tree. Heat map of copy numbers for all genes present in orthologues groups of genes, which have previously been associated to resistance and the α-tubulin gene. **b** Coverage ratio compared between worm sex (n=747), treatment (n=747) and from *T. incognita* to other species. Three *T. suis* genomes and one *T. trichiura* genome were used to calculate the average coverage in the species containing only one gene in OG0000671. *** signifies a p-value < 0.001. **c** TUBB gene tree of OG0000671, timescale indicates numbers of substitutions per site.

To extrapolate the duplication events from the assembled genome of one single worm to the population we leveraged the ratio of the average coverage of the amplified TUBB region compared to the longest autosomal contig present in the genome in the illumina shotgun sequencing data of all 747 worms as shown in Figure 4b. Intriguingly, the gene copy number was reflected in the average overall coverage showing a significant difference between worm sex (p-value < 0.001) but not showing any significance with regard to treatment. On average 11 copies were present in female worms and 8 copies in male worms. To compare the coverage ratios against species, where only one gene copy was identified, the raw sequencing data was downloaded for *T. suis* (PRJNA20815, PRJNA20816, PRJNA179528) and *T. trichiura* (PRJNA304165), the same region of TUBB was used and the coverage ratio was calculated again from the amplified TUBB region to the autosomal region. All reference coverage ratios were similar and far below the coverage ratio of *T. incognita*.

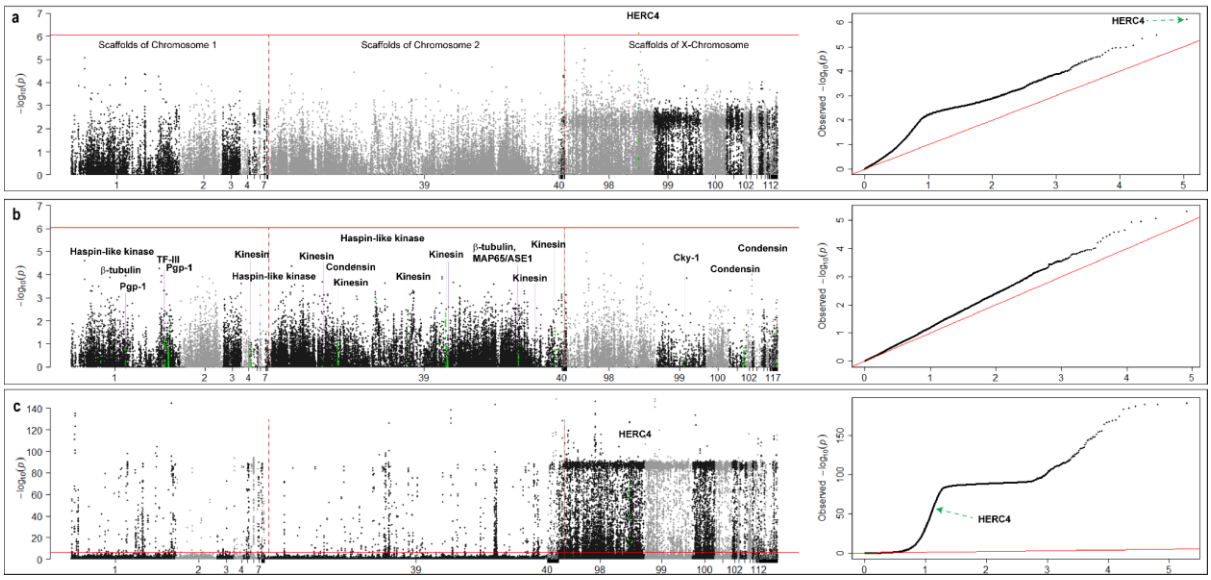
We then sought to investigate gene families which are as expanded or more expanded compared to the β-tubulin gene in *T. incognita*. GO terms of massively expanded gene families as in ≥7 copies per gene in *T. incognita* and ≤4 in all other species were investigated and a list of gene families and their annotation is provided in the supporting information. From 81 gene families, 68 could be associated to DNA transposable elements and retrotransposons, 6 remained unknown and without blast hits and 7 could be assigned a functionality. Apart from the β-tubulin gene, other genes related to microtubule function and flexibility were found within these 7 assignable orthologous groups. Amongst these highly duplicated genes is the motor protein kinesin, MAP65/Ase1, Condensin, Transcription factor (TF) IIIC and a haspin like kinase.

Genome wide association study shows a single putative hit, HERC4, associated with drug treatment, indicating a potential adaptation to drug exposure

To investigate potential SNPs associated with resistance, a genome wide association study was conducted with sequencing data of 747 individual worms, comparing the populations

isolated after the T1 and T2 respectively. Population stratification and sex bias was taken into consideration and a genome wide significance level of 8.89e-7 x 10^{-7} was established using the Bonferroni correction. A single hit was identified on the sex-linked chromosome, on gene 11724, which was blasted against the NCBI database and assigned to a HERC4 gene as shown in Figure 5a. No further variants were identified with a genome wide significance, challenging previously reported adaptation to drug treatment at the Phe168, Glu198 or Phe200 positions of TUBB. The strongest association of an SNP associated to drug treatment was the Asn194Ile mutation with a p-value of p = 0.12 which is far below the genome wide significance level. Furthermore, no genome wide significant SNP's were found on expanded gene families as shown in Figure 5b.

Fig. 5: Genome wide association study of worms expulsed after albendazole-ivermectin treatment (T1) versus worms expulsed after oxantel-pamoate treatment (T2).



GWAS displaying p-values of each SNP associated to treatment and corresponding to a non-silent mutation. The red line indicates the Bonferroni corrected significance of 8.89e-7 x 10^{-7} at α=0.05 and 56270 mutations. **a** Single SNP showing genome wide significance, located on a HERC4 gene. **b** Mutations on highly expanded gene families, including β-tubulin, which has been associated with albendazole resistance and genes associated with ivermectin resistance. **c** Mutations associated with worm sex.

The qq-plots are presented in Figure 5 and show an increased density of signals around -$\log_{10}$(p) = 2.5 for the Manhattan plot presented in Figure 5a. This shoulder has been attributed to the difference in worm sex distributions after T1 and T2. Although sex is taken into account when running the calculation in PLINK, the differences are still visible after correction. Figure 5c shows the GWAS when dividing the subpopulations into worm sex, achieving -$\log_{10}$(p) values of over 90, located primarily on the sex chromosome. The HERC4 gene, highlighted in green, shows a strong association also with sex, and the observed value is approximately 1.5 log units higher than the expected value, indicating that this association with sex might be artificially inflating the -$\log_{10}$(p) value in Figure 5a and that the mutation might not be significant on a genome wide level. When removing all sex attributed SNP's with a genome wide significance, the qq-plot shows comparably little difference between expected and observed values, however the HERC4 gene is then filtered out.

Discussion:

Our investigation and the data provided mark a pioneering milestone in *Trichuris* research, providing substantial evidence for the discovery of a previously unknown species - *Trichuris incognita* - naturally infecting humans in Côte d'Ivoire and demonstrating low responsiveness to standard drug treatments. We present the new species tree of the *Trichuris* genus, provide a high-quality whole genome, show genes essential for human parasitism through comparative genomic analyses and highly duplicated gene families such as the TUBB gene. A genome wide association study identified a single putative SNP, linked to resistance on the HERC4 gene and showed that reported mutations of the TUBB gene, postulated to confer with resistance, are not linked to drug resistance in *T. incognita*. These findings challenge the current understanding of species within the *Trichuris* genus and have profound implications for both parasitology and public health.

The existence of a species of *Trichuris*, phylogenetically distinct species from the canonical human-infective species, has been proposed in recent preprints by amplicon sequencing from Trichuris eggs by Venkatesan et al. and Rahman et al.{Venkatesan, 2024 #253} Our data confirm the existence of this species at the whole genome level, using the species tree constructed from 12434 orthologous groups, and 535 mitogenomes of 747 individually isolated and sequenced worms, all falling into the same clade. The whole-genome assembly metrics align with established *Trichuris* genomes, underscoring the robustness of our dataset.[8,9,11] However, structural variations might be missed, as the *T. muris* genome was used as a reference for scaffolding, similar to what was done in literature.[9] Our exploration of orthologous gene groups shared exclusively by *T. incognita* and *T. trichiura* – but not *T. suis* - unveils three gene families of interest: a serpin, a chymotrypsin-like serine protease and a tetraspanin gene family, that may all play a crucial role in human parasitism.

Serpins constitute the largest and most widely distributed superfamily of protease inhibitors and play a crucial role in escaping from host immunity.[18] They are excreted by *Trichuris* species and were reported to be upregulated in the stichosome of *T. suis*.[11] While serpins are promising antigens for *Trichinella spiralis* vaccination, this family of proteins has gained particular interest in the context of inflammatory bowels disease (IBD) due to their immunomodulatory potential.[13,14] A further intriguing gene family, identified only in human infecting *Trichuris* species, belong to the chymotrypsin-like serine proteases. Chymotrypsin is reported to degrade trypsin, the main digestive protease originating from the pancreas. The ability to degrade pancreatic proteases may be crucial for the survival of this enteric parasite and a starting point for drug development not only against trichuriasis but also in the field of pancreatitis.[11,19] Finally, we also identified a tetraspanin gene family. Tetraspanins are a highly conserved family of four-pass transmembrane proteins and their roles are highly diverse and occur in a variety of essential cellular functions. They have been identified as suitable antigens in schistosomiasis vaccination. For example, the large extracellular loop of *S. mansoni* tetraspanin-2 (Sm-TSP-2) has been employed successfully as a vaccine against murine schistosomiasis.[20,24-26] Notably, no effective human vaccine currently exists for trichuriasis, making the identification of crucial tetraspanins involved in human infection particularly promising for vaccine development.

With the unique experimental setup of the expulsion study, we were able to categorize the expulsed worms into drug sensitive and non-sensitive enabling a genome wide association study (GWAS). HERC4 was identified as the gene carrying the single SNP with genome wide significance. It is a member of the HECT-type E3 ligase family containing a HECT domain that catalyzes the formation of a thioester with ubiquitin, ultimately leading to ubiquitination of a substrate. HERC4 has found to be overexpressed in various cancer cells and involved in cell

proliferation modulation and highly expressed in mouse testis, specifically during spermiogenesis.[27-30] Intriguingly, ivermectin is reported to downregulate expression of HERC4 in ovarian cancer cells.[31,32] This is in contrast to reported quantitative trait loci (QTL) associated with ivermectin resistance in *H. contortus* and *C. elegans* found on the cky-1 gene and transmembrane proteins such as the chloride ion channel avr-15, all of which did not show an association with genome wide significance. We hypothesize that the resistance to drug treatment may be a combination of recent adaptation to ivermectin and general tolerance towards albendazole given the observation of duplication events.

Duplication events are encountered across the tree of life and constitute an important mechanism to provide new genetic material. The rates of gene duplication are estimated to be as frequent as the rates of single-nucleotide polymorphisms and the *C. elegans* genome is made up of 33% duplicated genes.[33,34] Drawing upon previous studies linking gene duplication to drug adaptation across diverse parasitic organisms, and deducing modes of action of drugs from observations of duplication events, we sought to investigate gene duplication events in connection to drug insensitivity.[35,36] In total, 7 variants of the β-tubulin gene were identified in the whole genome, providing substantial redundancy that may allow for adaptation through sub-functionalization or dosage compensation. Leveraging coverage ratios, duplication events could be confirmed in the whole population. Notably, while β-tubulin exhibited significant duplication, α-tubulin did not. Furthermore, the identification of massively expanded gene families associated with microtubule dynamics and mitotic processes may be connected with the tolerance this species has towards conventional treatments, assuming the same mode of action as in other nematodes.[37,38] Kinesin, for example, acts as intracellular shuttle and can de-polymerize microtubules. It has been shown to bind to negatively charged residues E410, D417 and E421 on the β-tubulin subunit.[39] MAP65/Ase1, a microtubule-associated protein is reported to promote microtubule flexibility and prevent microtubule severing by cross-linking.[40,41] These findings suggest that the duplications observed are associated to repairing and protecting microtubule functionality. Further, highly duplicated genes including Condensin, Transcription factor (TF) IIIC and a haspin like kinase were associated to chromatid condensation, organization and cohesion respectively. While these genes have not been associated to benzimidazole resistance, they are linked to mitotic processes which in turn are connected to microtubule function with regards to the mitotic spindle.[42-44] However, if partial duplication is a result of the drug pressure from mass drug administration, warrants further investigation.

Despite the close genetic relationship of *T. incognita* to *T. suis*, our phylogenetic analysis demonstrates clear separation between *T. incognita* and *T. suis* isolates from swine hosts in Côte d'Ivoire. While zoonosis events of trichuriasis have been reported, these findings suggest that there was a zoonotic event sometime in the past that was successful, but could not be observed currently.[45] Moreover, patterns of gene duplication, gain, and loss of function events in *T. incognita* compared to within *T. suis* suggest a more ancient and divergent evolutionary lineage. The morphological indistinguishability of *T. incognita*, *T. trichiura* and *T. suis* worms and eggs emphasizes the need of diagnostic tools able to differentiate between these worms to gain further insight into the transmission dynamics and prevalence of this new species, as it responds differently to drug treatment.

The expulsion study allowed also to gain insights into its fecundity and expulsion dynamics. Following the administration of albendazole and ivermectin during T1, only 271 out of 1098 worms could be eliminated, which is in-line with previously observed low cure rates in Côte d'Ivoire.[5,46] The female to male ratio is comparable to recent reports for *T. trichiura*. The fecundity of around 14,300 eggs per female per day of *T. incognita* matches literature reported

values of approx. 16,700 for *T. trichiura*.[47] Overall, these results suggest that the biology of *T. incognita* in terms of fecundity is comparable to *T. trichiura*.

In summary, our study presents a novel *Trichuris* species, less susceptible to conventional albendazole-ivermectin treatment and offers evidence of adaptive genomic changes associated with drug exposure. The whole genome presented in this study provides a unique data resource, not only for drug targets and vaccine candidates in the field of trichuriasis research, but also in the field of infectious, inflammatory, and autoimmune diseases.

Acknowledgements

We would like to acknowledge Dr. John Gilleard and Dr. Abhinaya Venkatesan from the Faculty of Veterinary Medicine, Host-Parasite Interactions Program, University of Calgary, AB T2N1N4, Canada for the fruitful discussions on *Trichuris* phylogeny. We are grateful to the European Research Council for financial support (Nr. 101019223).

References:

1 James, S. L. *et al.* Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet* **392**, 1789-1858, doi:10.1016/s0140-6736(18)32279-7 (2018).

2 Montresor, A. *et al.* The global progress of soil-transmitted helminthiases control in 2020 and World Health Organization targets for 2030. *PLoS Negl Trop Dis* **14**, e0008505, doi:10.1371/journal.pntd.0008505 (2020).

3 Rose Vineer, H. *et al.* Increasing importance of anthelmintic resistance in European livestock: creation and meta-analysis of an open database. *Parasite* **27**, 69, doi:10.1051/parasite/2020062 (2020).

4 Coffeng, L. E., Stolk, W. A. & de Vlas, S. J. Predicting the risk and speed of drug resistance emerging in soil-transmitted helminths during preventive chemotherapy. *Nat Commun* **15**, 1099, doi:10.1038/s41467-024-45027-2 (2024).

5 Hürlimann, E. *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* **22**, 123-135, doi:10.1016/s1473-3099(21)00421-7 (2022).

6 Rook, G. A. W. The old friends hypothesis: evolution, immunoregulation and essential microbial inputs. *Frontiers in Allergy* **4**, doi:10.3389/falgy.2023.1220481 (2023).

7 Ghedin, E. Panning for molecular gold in whipworm genomes. *Nat Genet* **46**, 661-663, doi:10.1038/ng.3018 (2014).

8 Doyle, S. R. *et al.* Population genomics of ancient and modern Trichuris trichiura. *Preprint*, doi:10.1101/2021.10.21.464505 (2021).

9 Foth, B. J. *et al.* Whipworm genome and dual-species transcriptome analyses provide molecular insights into an intimate host-parasite interaction. *Nat Genet* **46**, 693-700, doi:10.1038/ng.3010 (2014).

10 McSorley, H. J., Hewitson, J. P. & Maizels, R. M. Immunomodulation by helminth parasites: defining mechanisms and mediators. *Int J Parasitol* **43**, 301-310, doi:10.1016/j.ijpara.2012.11.011 (2013).

11 Jex, A. R. *et al.* Genome and transcriptome of the porcine whipworm Trichuris suis. *Nat Genet* **46**, 701-706, doi:10.1038/ng.3012 (2014).

12 Chakraborty, P., Aravindhan, V. & Mukherjee, S. Helminth-derived biomacromolecules as therapeutic agents for treating inflammatory and infectious diseases: What lessons do we get from recent findings? *Int J Biol Macromol* **241**, 124649, doi:10.1016/j.ijbiomac.2023.124649 (2023).

13 Mkaouar, H. *et al.* Gut Serpinome: Emerging Evidence in IBD. *Int J Mol Sci* **22**, doi:10.3390/ijms22116088 (2021).
14 Xu, J. *et al.* Influence of adjuvant formulation on inducing immune response in mice immunized with a recombinant serpin from Trichinella spiralis. *Parasite Immunol* **39**, doi:10.1111/pim.12437 (2017).
15 Jancso, Z., Hegyi, E. & Sahin-Toth, M. Chymotrypsin Reduces the Severity of Secretagogue-Induced Pancreatitis in Mice. *Gastroenterology* **155**, 1017-1021, doi:10.1053/j.gastro.2018.06.041 (2018).
16 Mosztbacher, D., Jancso, Z. & Sahin-Toth, M. Loss of chymotrypsin-like protease (CTRL) alters intrapancreatic protease activation but not pancreatitis severity in mice. *Sci Rep* **10**, 11731, doi:10.1038/s41598-020-68616-9 (2020).
17 Sanrattana, W., Maas, C. & de Maat, S. SERPINs-From Trap to Treatment. *Front Med (Lausanne)* **6**, 25, doi:10.3389/fmed.2019.00025 (2019).
18 Bao, J. *et al.* Serpin functions in host-pathogen interactions. *PeerJ* **6**, e4557, doi:10.7717/peerj.4557 (2018).
19 Goettig, P., Brandstetter, H. & Magdolen, V. Surface loops of trypsin-like serine proteases as determinants of function. *Biochimie* **166**, 52-76, doi:10.1016/j.biochi.2019.09.004 (2019).
20 Tran, M. H. *et al.* Tetraspanins on the surface of Schistosoma mansoni are protective antigens against schistosomiasis. *Nat Med* **12**, 835-840, doi:10.1038/nm1430 (2006).
21 Susa, K. J., Kruse, A. C. & Blacklow, S. C. Tetraspanins: structure, dynamics, and principles of partner-protein recognition. *Trends Cell Biol*, doi:10.1016/j.tcb.2023.09.003 (2023).
22 Laing, R. *et al.* Transcriptomic analyses implicate neuronal plasticity and chloride homeostasis in ivermectin resistance and response to treatment in a parasitic nematode. *PLoS Pathog* **18**, e1010545, doi:10.1371/journal.ppat.1010545 (2022).
23 Doyle, S. R. *et al.* Genomic landscape of drug response reveals novel mediators of anthelmintic resistance. *Preprint*, doi:10.1101/2021.11.12.465712 (2022).
24 Chen, L. *et al.* Protection and immunological study on two tetraspanin-derived vaccine candidates against schistosomiasis japonicum. *Parasite Immunol* **38**, 589-598, doi:10.1111/pim.12338 (2016).
25 Mekonnen, G. G. *et al.* Characterisation of tetraspanins from Schistosoma haematobium and evaluation of their potential as novel diagnostic markers. *PLoS Negl Trop Dis* **16**, e0010151, doi:10.1371/journal.pntd.0010151 (2022).
26 Hayon, J., Weatherhead, J., Hotez, P. J., Bottazzi, M. E. & Zhan, B. Advances in vaccine development for human trichuriasis. *Parasitology*, 1-12, doi:10.1017/S0031182021000500 (2021).
27 Zhu, Q., Yang, X. & Lv, Y. HERC4 modulates ovarian cancer cell proliferation by regulating SMO-elicited hedgehog signaling. *Biochim Biophys Acta Gen Subj* **1868**, 130557, doi:10.1016/j.bbagen.2023.130557 (2024).
28 Zheng, Y. *et al.* HERC4 Is Overexpressed in Hepatocellular Carcinoma and Contributes to the Proliferation and Migration of Hepatocellular Carcinoma Cells. *DNA Cell Biol* **36**, 490-500, doi:10.1089/dna.2016.3626 (2017).
29 Zeng, W. L. *et al.* Expression of HERC4 in lung cancer and its correlation with clinicopathological parameters. *Asian Pac J Cancer Prev* **16**, 513-517, doi:10.7314/apjcp.2015.16.2.513 (2015).
30 Rodriguez, C. I. & Stewart, C. L. Disruption of the ubiquitin ligase HERC4 causes defects in spermatozoon maturation and impaired fertility. *Dev Biol* **312**, 501-508, doi:10.1016/j.ydbio.2007.09.053 (2007).
31 Mutlu, M. *et al.* Small molecule induced STING degradation facilitated by the HECT ligase HERC4. *Nat Commun* **15**, 4584, doi:10.1038/s41467-024-48922-w (2024).

456 32 Li, N., Zhao, L. & Zhan, X. Quantitative proteomics reveals a broad-spectrum antiviral property
457 of ivermectin, benefiting for COVID-19 treatment. *J Cell Physiol* **236**, 2959-2975,
458 doi:10.1002/jcp.30055 (2021).
459 33 Kuzmin, E., Taylor, J. S. & Boone, C. Retention of duplicated genes in evolution. *Trends Genet*
460 **38**, 59-72, doi:10.1016/j.tig.2021.06.016 (2022).
461 34 Lallemand, T., Leduc, M., Landes, C., Rizzon, C. & Lerat, E. An Overview of Duplicated Gene
462 Detection Methods: Why the Duplication Mechanism Has to Be Accounted for in Their Choice.
463 *Genes (Basel)* **11**, doi:10.3390/genes11091046 (2020).
464 35 Kondrashov, F. A. Gene duplication as a mechanism of genomic adaptation to a changing
465 environment. *Proc Biol Sci* **279**, 5048-5057, doi:10.1098/rspb.2012.1108 (2012).
466 36 Zajac, N. *et al.* Gene Duplication and Gain in the Trematode Atriophallophorus winterbourni
467 Contributes to Adaptation to Parasitism. *Genome Biol Evol* **13**, doi:10.1093/gbe/evab010
468 (2021).
469 37 Enos, A. & Coles, G. C. Effect of benzimidazole drugs on tubulin in benzimidazole resistant and
470 susceptible strains of Caenorhabditis elegans. *Int J Parasitol* **20**, 161-167, doi:10.1016/0020-
471 7519(90)90096-6 (1990).
472 38 Dawson, P. J., Gutteridge, W. E. & Gull, K. A comparison of the interaction of anthelmintic
473 benzimidazoles with tubulin isolated from mammalian tissue and the parasitic nematode
474 Ascaridia galli. *Biochem Pharmacol* **33**, 1069-1074, doi:10.1016/0006-2952(84)90515-x
475 (1984).
476 39 Cushion, T. D., Leca, I. & Keays, D. A. MAPping tubulin mutations. *Front Cell Dev Biol* **11**,
477 1136699, doi:10.3389/fcell.2023.1136699 (2023).
478 40 Burkart, G. M. & Dixit, R. Microtubule bundling by MAP65-1 protects against severing by
479 inhibiting the binding of katanin. *Mol Biol Cell* **30**, 1587-1597, doi:10.1091/mbc.E18-12-0776
480 (2019).
481 41 Portran, D. *et al.* MAP65/Ase1 promote microtubule flexibility. *Mol Biol Cell* **24**, 1964-1973,
482 doi:10.1091/mbc.E13-03-0141 (2013).
483 42 Hagstrom, K. A. & Meyer, B. J. Condensin and cohesin: more than chromosome compactor
484 and glue. *Nat Rev Genet* **4**, 520-534, doi:10.1038/nrg1110 (2003).
485 43 Stutzman, A. V., Liang, A. S., Beilinson, V. & Ikegami, K. Transcription-independent TFIIIC-
486 bound sites cluster near heterochromatin boundaries within lamina-associated domains in C.
487 elegans. *Epigenetics Chromatin* **13**, 1, doi:10.1186/s13072-019-0325-2 (2020).
488 44 Dai, J., Kateneva, A. V. & Higgins, J. M. Studies of haspin-depleted cells reveal that spindle-
489 pole integrity in mitosis requires chromosome cohesion. *J Cell Sci* **122**, 4168-4176,
490 doi:10.1242/jcs.054122 (2009).
491 45 Nejsum, P., Betson, M., Bendall, R. P., Thamsborg, S. M. & Stothard, J. R. Assessing the
492 zoonotic potential of Ascaris suum and Trichuris suis: looking to the future from an analysis of
493 the past. *J Helminthol* **86**, 148-155, doi:10.1017/S0022149X12000193 (2012).
494 46 Sprecher, V. P., Coulibaly, J. T., Hurlimann, E., Hattendorf, J. & Keiser, J. Efficacy and Safety of
495 Moxidectin-Albendazole and Ivermectin-Albendazole Combination Therapy Compared to
496 Albendazole Monotherapy in Adolescents and Adults Infected with Trichuris trichiura: A
497 Randomized, Controlled Superiority Trial. *Clin Infect Dis* **77**, 1294-1302,
498 doi:10.1093/cid/ciad387 (2023).
499 47 Hansen, E. P. *et al.* Faecal egg counts and expulsion dynamics of the whipworm, Trichuris
500 trichiura following self-infection. *J Helminthol* **90**, 298-302, doi:10.1017/S0022149X1500019X
501 (2016).
502 48 Cringoli, G. *et al.* The Mini-FLOTAC technique for the diagnosis of helminth and protozoan
503 infections in humans and animals. *Nat Protoc* **12**, 1723-1732, doi:10.1038/nprot.2017.067
504 (2017).
505 49 Di Tommaso, P. *et al.* Nextflow enables reproducible computational workflows. *Nat*
506 *Biotechnol* **35**, 316-319, doi:10.1038/nbt.3820 (2017).

50 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120, doi:10.1093/bioinformatics/btu170 (2014).
51 Jin, J. J. *et al.* GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol* **21**, 241, doi:10.1186/s13059-020-02154-5 (2020).
52 Meng, G., Li, Y., Yang, C. & Liu, S. MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization. *Nucleic Acids Res* **47**, e63, doi:10.1093/nar/gkz173 (2019).
53 Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772-780, doi:10.1093/molbev/mst010 (2013).
54 Bouckaert, R. *et al.* BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* **15**, e1006650, doi:10.1371/journal.pcbi.1006650 (2019).
55 De Coster, W. & Rademakers, R. NanoPack2: population-scale evaluation of long-read sequencing data. *Bioinformatics* **39**, doi:10.1093/bioinformatics/btad311 (2023).
56 Sun, J., Li, R., Chen, C., Sigwart, J. D. & Kocot, K. M. Benchmarking Oxford Nanopore read assemblers for high-quality molluscan genomes. *Philos Trans R Soc Lond B Biol Sci* **376**, 20200160, doi:10.1098/rstb.2020.0160 (2021).
57 Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol* **20**, 257, doi:10.1186/s13059-019-1891-0 (2019).
58 Zimin, A. V. *et al.* The MaSuRCA genome assembler. *Bioinformatics* **29**, 2669-2677, doi:10.1093/bioinformatics/btt476 (2013).
59 Alonge, M. *et al.* Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol* **23**, 258, doi:10.1186/s13059-022-02823-7 (2022).
60 Shumate, A. & Salzberg, S. L. Liftoff: accurate mapping of gene annotations. *Bioinformatics* **37**, 1639-1643, doi:10.1093/bioinformatics/btaa1016 (2021).
61 Manni, M., Berkeley, M. R., Seppey, M., Simao, F. A. & Zdobnov, E. M. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Mol Biol Evol* **38**, 4647-4654, doi:10.1093/molbev/msab199 (2021).
62 Mikheenko, A., Prjibelski, A., Saveliev, V., Antipov, D. & Gurevich, A. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* **34**, i142-i150, doi:10.1093/bioinformatics/bty266 (2018).
63 Kuznetsov, D. *et al.* OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Res* **51**, D445-D451, doi:10.1093/nar/gkac998 (2023).
64 Bruna, T., Hoff, K. J., Lomsadze, A., Stanke, M. & Borodovsky, M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom Bioinform* **3**, lqaa108, doi:10.1093/nargab/lqaa108 (2021).
65 Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* **20**, 238, doi:10.1186/s13059-019-1832-y (2019).
66 Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* **49**, W293-W296, doi:10.1093/nar/gkab301 (2021).
67 Altenhoff, A. M. *et al.* OMA standalone: orthology inference among public and custom genomes and transcriptomes. *Genome Res* **29**, 1152-1163, doi:10.1101/gr.243212.118 (2019).
68 Train, C. M., Pignatelli, M., Altenhoff, A. & Dessimoz, C. iHam and pyHam: visualizing and processing hierarchical orthologous groups. *Bioinformatics* **35**, 2504-2506, doi:10.1093/bioinformatics/bty994 (2019).
69 Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-575, doi:10.1086/519795 (2007).

Supporting Information

Methods:

Expulsion study design fieldwork

This trial was conducted in the Dabou and Jacqueville districts of Côte d'Ivoire. This area was selected, because communities with high endemicity were identified as part of a previous trial on *T. trichiura*.[5] All community members aged 6-12 who were positive for *T. trichiura* in duplicate Kato-Katz smears, with an infection intensity of at least 200 eggs per gram of stool (EPG), were eligible for trial inclusion. Excluded from the trial were individuals with reported major systemic illnesses or chronic disease, known allergy to study medications and having received anthelmintic treatment with ivermectin in the past two weeks.

This trial was conducted in accordance with the protocol, International Conference on Harmonization Good Clinical Practice E6 (R2) (ICH-GCP) and the current version of the Helsinki Declaration. Parents or guardians of participating children provided written informed consent. Children provided written assent. All authors take responsibility for the accuracy and completeness of the data and the fidelity of the trial to the protocol, which is available together with the statistical analysis plan in appendix 1. The protocol was approved by independent ethics committees in Côte d'Ivoire (reference numbers 156-22/MSHPCMU/CNESVS-kp and ECCI00918) and Switzerland (reference number AO_2022_00028).

Treatment and sample collection from humans

Guardians/caregivers were invited to participate in an information session. The research team explained the purpose and procedures of the study, as well as potential benefits and risks of participation. The timeline of the procedures is summarized in Figure 1a. For initial diagnosis, participants were asked to provide one stool sample from which duplicate Kato-Katz thick smears using 41·7 mg of stool were prepared and assessed under a light microscope for the identification of *T. trichiura*, *A. lumbricoides*, and hookworm ova by laboratory technicians. 10% of all Kato-Katz slides were randomly chosen for subsequent quality control by picking every tenth slide of all slides read by each laboratory technician on the respective day. All eligible participants were treated with albendazole-ivermectin at day 0. Albendazole, the product of Glaxo Smith Kline (Zentel®), was a single tablet of 400 mg. Ivermectin, the product of Merck (Stromectol®), was given at a dose of of 200 µg/kg. On day 8 a single oral dose of 20 mg/kg of oxantal pamoate was administered. Oxantel pamoate tablets (500 mg) were provided by the University of Basel.

From day 1 or 2 and 8 onwards, 24h-stool was collected and weighed for 2-3 days. Expulsed worms were counted, washed twice with distilled water, once with sterile PBS and then stored in pure ethanol at 4°C. A picture of each worm in ethanol was taken and is provided on online (https://github.com/max-baer).

Sample collection from pigs

To obtain *T. suis* worms, a list of all pig keepers was drawn by the help of local village chief. Pig keepers were gathered in formal meetings where the purpose of the study was explained. Pig keepers were asked to give written informed consent to participate in the study. Fecal samples were then collected directly from the rectum by use of gloves lubricated with glycerin by a veterinarian. The samples were analyzed using the Mini Flotac technique.[48] 2 pigs that tested positive for *T. suis* roaming freely in Tiagba were sacrificed and worms were collected directly from the cecum and stored in either pure ethanol at 4 °C or mashed and stored in RNA-later at -20 °C.

DNA extraction and high-throughput sequencing

Worms stored in absolute ethanol were photographed and used for sequencing experiments. The posterior end of the worms was severed and DNA was extracted using a DNeasy Blood & Tissue Kit (Qiagen, Cat: 69504) according to the manufacturer's protocol from the anterior part of the worm. Sequencing libraries for illumina were prepared using the NEBNext® Ultra™ II FS DNA Library Prep Kit (New England Biolabs, Cat: E7805L) according to the manufacturer's protocol after which DNA concentration and library size was determined (HS NGS Fragment Analyzer). Libraries were sequenced using 150 bp paired-end chemistry at an average 20X coverage on an Illumina NovaSeq6000 platform with an S4 flow cell at the Genomics Facility Basel as part of the Department of Biosystems Science and Engineering. Sequencing libraries for Nanopore were prepared using the Native Barcoding Kit 24 V14 (Oxford Nanopore Technologies, Cat: SQK-NBD114.2) from a single whole female worm and sequenced on a PromethION flow cell and PromethION 2 device.

PCR and Nanopore sequencing experiment of the TUBB gene

Two primers were selected surrounding the regions containing known mutations (Phe168, Glu198, Phe200), observed variable position (Ser194) and predicted truncation site. The primer sequences were AAAGAGACCGGACATTTCGC for the forward primer and TGAATTGCCTGGTTTCTAGAATG for the reversed primer and the amplified region was 402 base pairs. Genomic DNA from one sequenced worm was used as input ant the OneTaq® One-Step RT-PCR Kit was used to amplify the region during 40 cycles at an annealing temperature of 56°C for 30 seconds and an extension temperature of 68°C for 1 minute, although the RT function of the Kit was redundant. Sequencing libraries for Nanopore were prepared using the native barcoding kit 24 V14 (Oxford Nanopore Technologies, Cat: SQK-NBD114.2), and sequenced on a Flongle flow cell and the MinION device. The dorado basecaller was used for duplex basecalling. Reads were filtered by alignment score (AS>600) and length (>390 BP) and aligned using minimap2 to the reference TUBB sequence. Variants were called using bcftools and coverage plots were created in R.

Pipeline for raw read processing and mitochondrial genome assembly

For all sequencing, assembly and annotation pipelines Nextflow (version 23.04.1 build 5866) was used with Java (version 11.0.3).[49] All outputs of confirmed Nextflow runs and code is provided online (https://github.com/max-baer). Raw reads were first processed using Trimmomatic (version 0.39);[50] for PE reads, adapters were removed and N-bases trimmed. Next reads were channeled to getorganelle (version 1.7.7.0) to reconstruct the mitochondrial genome.[51] Runs which resulted in a complete mitochondrial genome were further processed by flipping in cases where the reverse complement was generated and adjustment to COX1 as starting gene. Reference and newly constructed mitochondrial genomes were annotated and visualized using MitoZ (version 3.6).[52] Coding genes were then extracted and aligned as amino acid and nucleic acid sequences using MAFFT (version 7.490).[53] All reconstructed mitochondrial genomes and the aligned MAFFT files used as input for phylogenetic inference are provided online (https://github.com/max-baer).

Phylogenetic analysis

BEAST2 (version 2.7.5) was used for all phylogenetic inference with a pure birth process (Yule process).[54] Aligned concatenated nucleic acid sequences from all assembled and reference mitochondrial genomes were treated as homochronous. The JC69 substitution model with gamma-distributed rate heterogeneity was used (JC69 + Γ_4) and a strict molecular clock was assumed. Trees and clock models were linked and all model parameters were estimated jointly. The prior distributions of the model parameters are listed in Supplementary Table1. A

Markov Chain Monte Carlo was run for each analysis. Tracer (version 1.7.2) was used to assess convergence and effective sample size. The percentage of samples discarded as burn-in was at least 10%. For the amino acid inferred species tree, 2 assembled sequences of each subclade clade with a posterior distribution over 0.95 were selected at random. The mito REV substitution model was used and all genes were partitioned separately, allowing for different rates for each gene. Effective sample size (ESS) was at least 900 for each of the inferred parameters. The intraspecies phylogeny was inferred using a TN96 + Γ_4 model. Each gene was partitioned separately and codon positions 1 and 2 were partitioned separately to 3. The mean mutation rates of the genes and there codon positions are provided in the supporting information and show a faster mutation rate for codon position 3 in all cases.

Pipeline for Hybrid-genome assembly and gene prediction

Nextflow (version 23.04.1 build 5866) was used with Java (version 11.0.3). Nanopore raw reads from a single worm were channeled into chopper (version 0.6.0) where reads below 5kb and low quality reads were filtered out and adapters trimmed.[55,56] Trimmed reads were channeled into Kraken2 (version 2.1.1) and reads with a contamination ratio >0.5 were filtered out with a custom R (version 4.3.0) script, resulting in a 50X coverage.[57] Illumina reads of the same worm at a 65X coverage were channeled into Kraken2 where reads of impurities were identified and filtered out. Illumina and Nanopore channels were combined to run a hybrid de-novo genome assembly using MaSuRCA (version 4.0.9).[58] The outputted contigs were channeled into RagTag (version 2.1.0) for correction and scaffolding using the *Trichuris trichiura* (PRJEB535) reference genome.[59] The corrected scaffolds were then assigned to one of three linkage groups from T. Muris using Liftoff (version 1.6.3),[60] as described in literature.[8] The genome was frozen for quality assessment using BUSCO (version 5.1.2 with Metazoa_obd10 database)[61] and QUAST (version 5.0.2).[62] At this point all available reference whole genomes of *Trichuris* species and *Trichinella spiralis* were channeled into the pipeline to run a de-novo gene prediction on all reference genomes and reduce methodological bias from gene prediction software. Repeat regions were masked using RepeatMasker (version 4.1.4) before channeling the genome to BRAKER (version 3.0.6, Braker2 pipeline using the Metazoa protein database from OrthoDB 11).[63,64] Gene prediction statistics including exon and intron count, mean length and median length were gathered using a custom bash script.

Phylogenetic orthology inference, species tree and orthologous group functional annotation

Orthofinder (version 2.5.5) was used to infer orthology and the species tree of the braker2 predicted genes.[65] 81425 out of 92202 (88.3%) genes were classified into 12434 orthologous groups with a mean group size of 6.5 and median group size of 5. Gene ontology of orthologous groups of interest was inferred using InterProScan (version 5.63-95.0). A list of resistance associated genes was downloaded from WormBase and blasted against all orthologous groups using an E-value threshold of $1e^{-50}$ to identify orthologous groups of resistance associated genes. In SI Table 2 a list of all genes with matching orthologous groups and Gene IDs from worm bank is provided. Species tree with heat map was visualized using iTOL (version 6.8.1).[66] Duplicated, gained, lost and retained genes across the species tree were inferred using OMA standalone (version 2.6.0) and pyHam (version 1.1.11).[67,68] Massively expanded gene families were defined as ≥7 copies in *Trichuris incognita* and ≤4 in all other species and a list is provided in the supporting information together with the functional annotation. The orthology of predicted transcripts was inferred analogously. A custom R-script was used to visualize the Venn diagram and identify transcripts shared by *Trichuris incognita*, *T. trichiura* and *T. muris*. SI table 1, a list of these transcripts with their functional annotation

is provided. Protein structures were predicted using Alphafold (version 2.2.0, template date cutoff was set to 2021-09-01) and visualized with PyMOL (version 2.5).

Pipeline for genome wide association study

Nextflow (version 23.04.1 build 5866) was used with Java (version 11.0.3). Variants were called by transferring a published protocol by Doyle et. al. into Nextflow[8] full details of which are provided online (https://github.com/max-baer). Plink (version 1.9) was used to conduct the genome wide association study.[69] In brief, a principal component analysis (PCA) was conducted to account for population stratification. The PCA together with information on worm-sex was accounted for and a linear regression model was chosen in PLINK. The R package qqman was used to visualize the results and generate the QQ-plots.

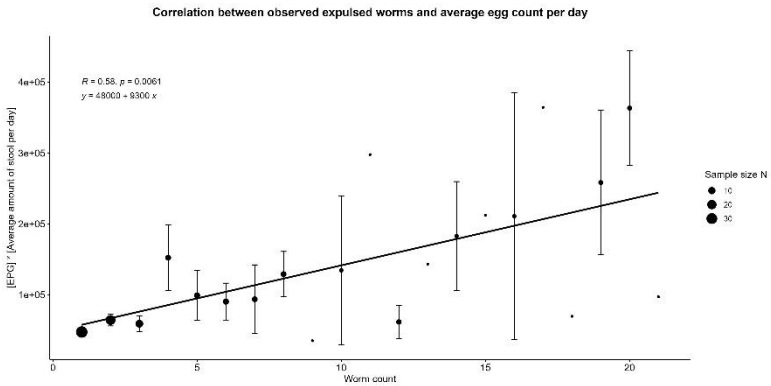
Data availability

Raw sequencing data is available under the European Nucleotide Archive (ENA) study accession “”. 551 assembled mitochondrial genomes including *T. incognita* and *T. suis* and the de-novo assembled *Trichuris incognita* reference genome are available on from https://github.com/max-baer. The publicly available mitochondrial genomes used for comparative analysis (Accession numbers: “KT449826”, “GU385218“, “HG806815“, “KT449822“, “KT449823“, “GU070737“, “JQ996231“, “JQ996232“, “KC461179“, “NC_017747”, “NC_018597”, “NC_018596”, “NC_028621”, “NC_002681”) are available from ENA and NCBI. There are no restrictions on data availability.

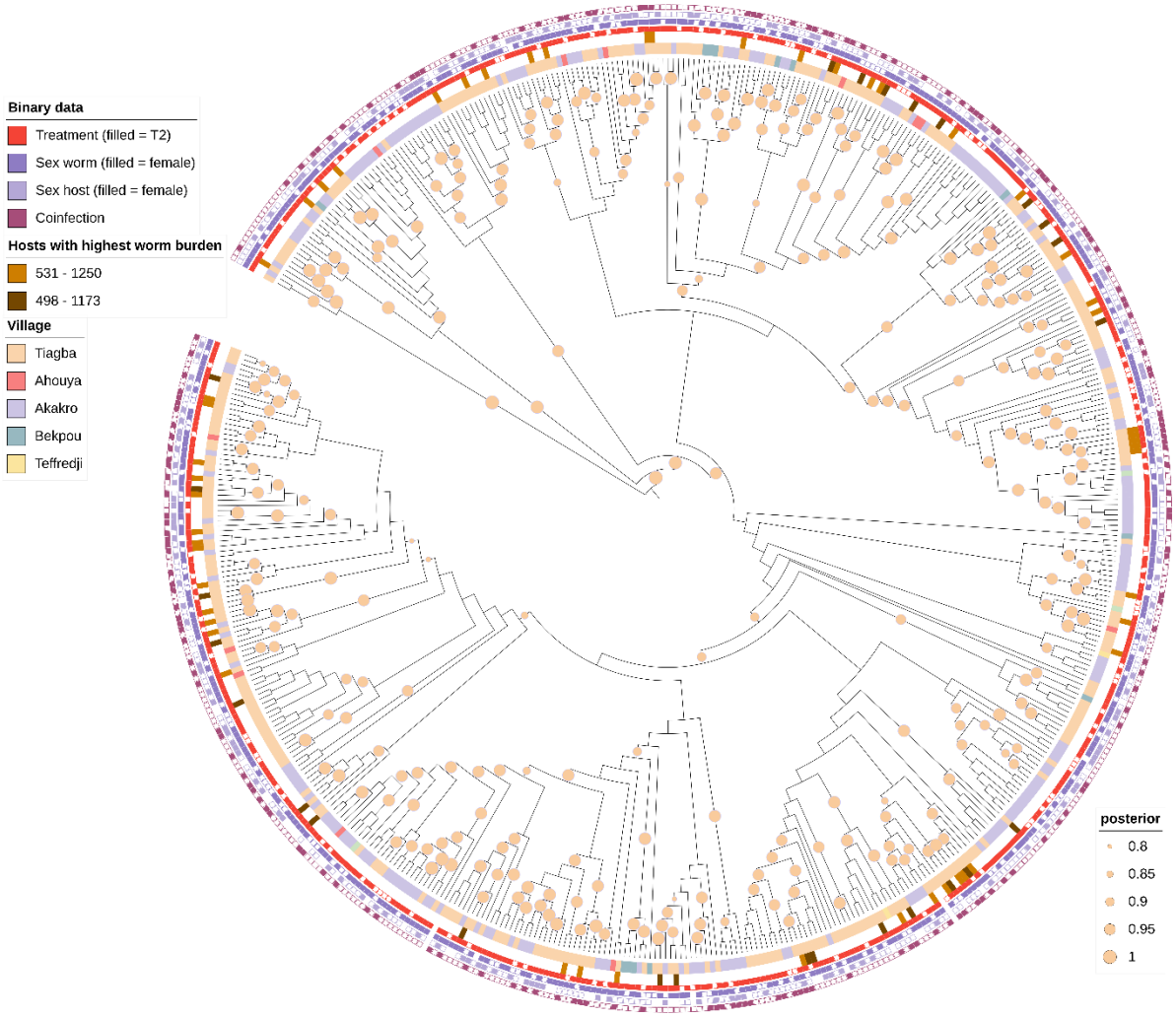
Code availability

Custom code to analyze data and reproduce the figures presented is available at https://github.com/max-baer. Where available all input scripts and log files are provided including nextflow log files documenting a complete run, BEAST2 XML input and log files.

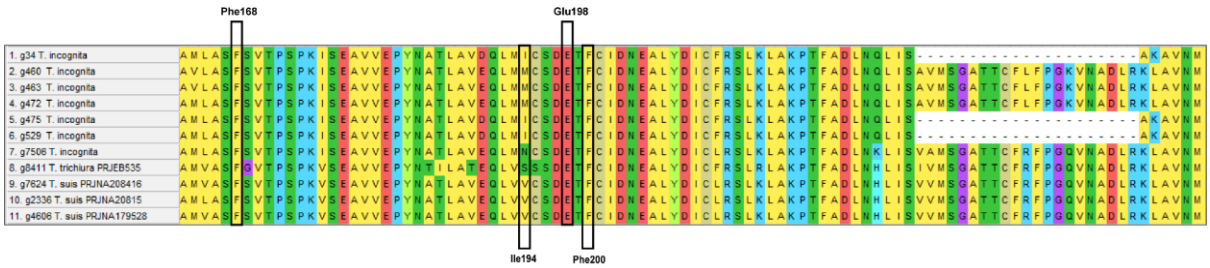
Figures



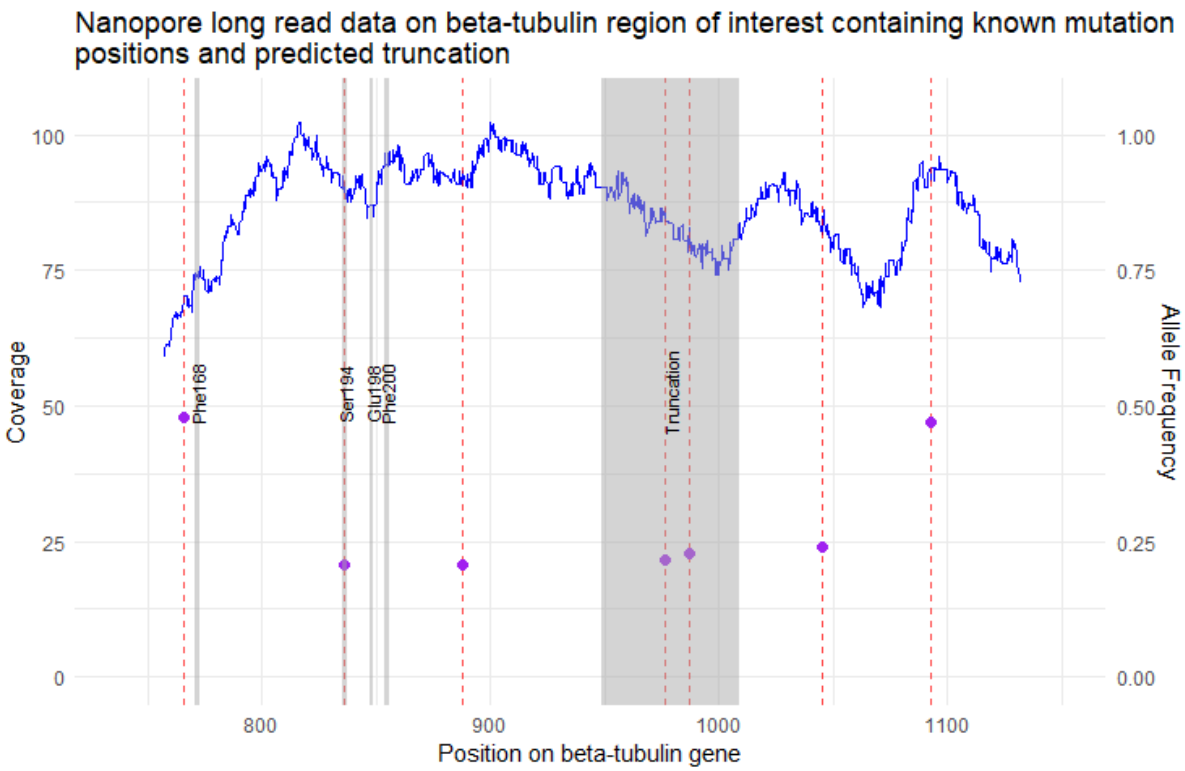
SI Figure 1: Correlation between observed expulsed worms and average egg count per day.



SI Figure 2: Phylogenetic tree based on mitochondrial sequences from 535 isolated *T. incognita* worms from human hosts.



SI Figure 3: Beta tubulin genes found in *T. incognita* compared to *T. suis* and *T. trichiura*, predicted from the braker2 pipeline. Black indicates positions carrying known mutations (Phe168, Glu198, Phe200), observed variable position (Ser194). A truncation was predicted *in silico* on 3 of the 7 genes.



SI Figure 4: Coverage and variants from an independent PCR and Nanopore genome sequencing experiment of the regions containing known mutations of TUBB and the *in silico* predicted truncation. Blue line indicates coverage, red lines the called variants using bcftools with the purple dot indicating allele frequency. Gray indicating positions carrying known mutations (Phe168, Glu198, Phe200), observed variable position (Ser194) and predicted truncation indicating that the truncation was not observed from the amplified genomic sequence.

Forward primer sequence: AAAGAGACCGGACATTTCGC.

Reversed primer sequence: TGAATTGCCTGGTTTCTAGAATG.

Tables

"Gene_list_venn_intersection.xlsx"

SI Table 1. Gene list of cross section in Venn diagram.

"Gene_list_resistance_associated_literature.xlsx"

SI Table 2. Gene list of resistance associated genes towards albendazole or ivermectin in helminths.

"Orthologous_group_list_duplications.xlsx"

SI Table 3. Highly duplicated genes.