Genomic and morphological characterization of *Trichuris incognita*, a Human-Infecting *Trichuris* species

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. For decades, it was thought to be caused by one species of whipworm, *Trichuris trichiura*. Significant differences in response rates to the best available anthelmintic treatment — a combination of albendazole and ivermectin — for trichuriasis exist across clinical trial reports yet the underlying reasons are unclear. Here, we present and characterize a human-infecting *Trichuris* species that we named *Trichuris incognita*. Combining long- and short- read sequencing approaches, we assembled a high-quality reference genome of *T. incognita* to construct the species tree of the *Trichuris* genus using 12434 orthologous groups. We sequenced a total of 747 individual worms which were used to confirm 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 of *T. incognita*, compared to the canonical *T. trichiura* species. We conducted a genome wide association study (GWAS) comparing 160 albendazole-ivermectin sensitive to 492 drug non-sensitive worms, which did not point towards an adaptation to drug pressure within the same species. Our results demonstrate that trichuriasis can be caused by multiple species of whipworm. This discovery, coupled with high tolerability of *T. incognita* to albendazole-ivermectin marks a significant shift in how we understand and approach whipworm infections.

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 (PC).[2]

Mebendazole was first mentioned in 1976 by the WHO as a suitable anthelmintic drug for PC, shortly followed by other benzimidazoles such as albendazole. First reports of mass drug administration using benzimidazoles date back to 1984, while ivermectin (Mectizan®) is employed since 1987 mainly for onchocerciasis.[3-5] With more than 3.3 billion tablets of albendazole distributed up until August 2020 and 5 billion treatments of Mectizan® until

January 2025, concern regarding anthelmintic resistance extending to humans, is rising. Reports of decreased efficacy and documented resistance to the same drugs within other helminths in livestock support projections suggesting the emergence of resistance within the next decade.[6,7] While modelling predicts an adaptation to drug treatment may evolve within a decade in helminths, conclusive evidence has not yet been provided.[7]

To circumvent resistance and improve efficacy against *T. trichiura* the use of a combination therapy of albendazole and ivermectin has been taken forward. Recent clinical trials with albendazole-ivermectin, revealed alarmingly low cure rates against what was presumed to be *T. trichiura* in Côte d'Ivoire, in contrast to higher efficacy observed in Laos and Pemba, Tanzania.[8] In collaboration with A. Venkatesan *et al.* amplicon sequencing data pointed towards underlying genomic differences and potentially a new species of human infecting *Trichuris* in Côte d'Ivoire.[9] Furthermore, N. Rahman *et al* show that this species is well established across multiple continents.[10] This species has co-evolved with humans, potentially for millennia, yet has remain evasive to our attention.

T. trichiura, a longstanding companion of humans, has intrigued scientists as both symbiotic ally and evasive parasite, sparking debate over the transition between these roles.[11-13] The relationship with its human host is considered intimate as the head is buried into the mucosa of the large intestine.[14] Helminth derived molecules are indicated to be potent regulators or stimulators of the host immune system and it is thought that these effects may be harnessed for immunotherapeutic agents for infectious, inflammatory, and autoimmune diseases affecting both high- and low-income countries such as type 1 diabetes and multiple sclerosis.[12,15-18] Notably, only the genomes of three *Trichuris* species have been published, *T. suis*,[17] *T. muris*,[15] and the human-invasive *T. trichiura*.[13,15]

We present a new species of *Trichuris*, *Trichuris incognita*, naturally infecting humans in Côte d'Ivoire. *T. incognita* displays a high tolerance to conventional drug treatments and is characterized by a high copy number of β-tubulin (TUBB) genes. Phylogenetic analysis based on the species tree of 12434 orthologous groups of genes, places the isolated worms within a new, monophyletic clade which is genetically closer to *T. suis* than to *T. trichiura*. We elucidate the genome of this new species, leveraging whole-genome-sequencing (WGS) data to produce, a high quality reference genome of this species. Finally, we provide the first ever dataset of human infecting helminths comparing pre- and post-treatment WGS data. 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 GWAS.

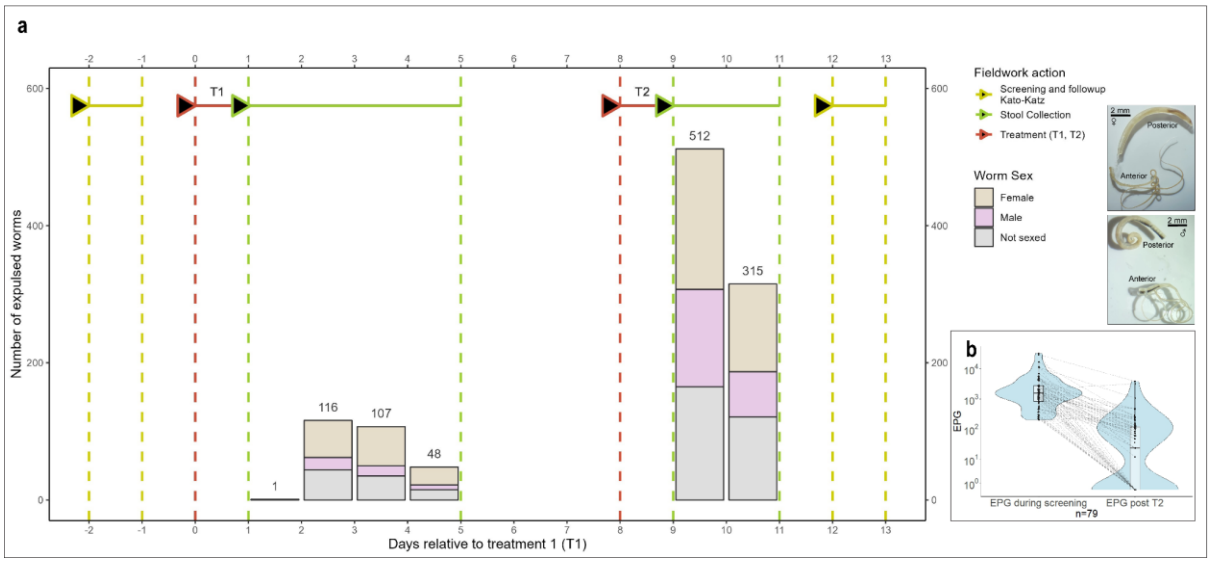
Results

Trichuris expulsion study confirms a new species of whipworm infecting humans, responding poorly to conventional albendazole-ivermectin treatment

We screened 670 children in 7 villages, namely 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 provided enough DNA for genomic analyses. The female to male ratio of the worm sex 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 *T. suis*, 25 individual worms from 2 swine hosts were collected directly from the caecum. A picture of each worm which was used to identify the worm sex is provided on github.

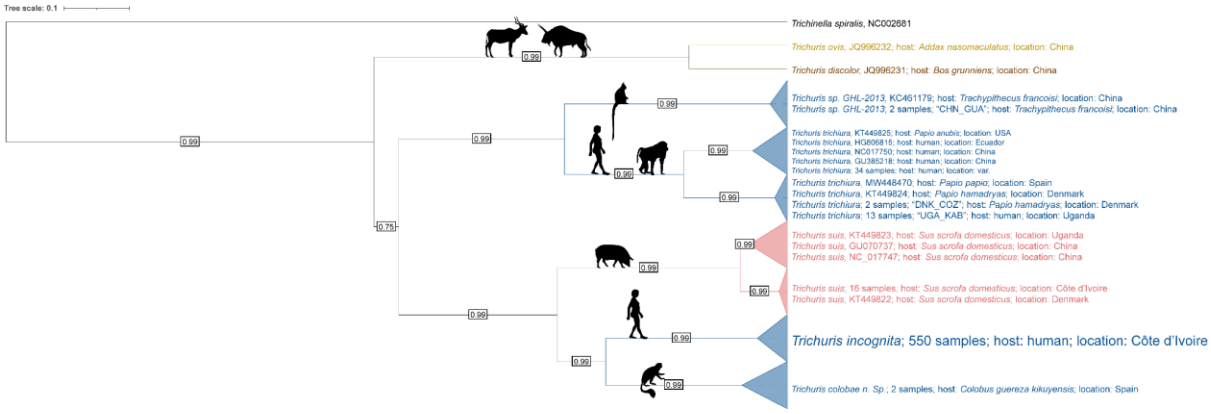
Figure 1: Timeline of expulsion study including the number of expulsed worms and the female/male distribution of the sequenced worms



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. Further morphological data is provided in Figure 3. **b** EPG before T1 and after T2.

To place these new isolates of whole worms from humans, a phylogenetic analysis on the mitogenome was conducted to place this species into context of the *Trichuris* genus. The phylogenetic analysis was done using the coding regions of the mitochondrial genomes of 535 individual worms that yielded a complete mitogenome and is presented in SI Figure 2. The reconstructed mitogenomes contained 13 protein coding genes (*cox*1-3, *nad*1-6, *nad*4L, *atp*6, *atp*8 and *co*B). 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. Previously reported sequences of a *Trichuris* species, isolated from a colobus monkey in Spain show a close genomic relationship with the species from Côte d'Ivoire. Finally, the phylogenetic tree of only *T. incognita* samples was plotted in SI Figure 2 to see if any of the branching within the species correlates to treatment, worm sex, host sex, host, coinfection with other worms, however no correlation could be identified. This species was named *Trichuris incognita* and has now been identified across the African and Asian continent.[9,10]

Figure 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.

Morphological description of *T. incognita*

This parasite has a threadlike structure, filiform. Both female and male can be divided into a thin hair-like anterior part and broader handle-like posterior section. The posterior section of the female worms is slightly ventrally incurved compared to the strong ventral incurvature of the male, adopting a Fibonacci spiral-like structure. The overall length of the worms (displayed in Figure 1a) is 45 mm and 49 mm and the ratio of anterior to posterior is 2.5 and 3.1 for female and male worms respectively, which is comparable to observed *T. trichiura* lengths.[19] The main organs of the male posterior end are the intestine, testis and ejaculatory duct, which run in parallel along the long axis of the body (Figure 3a). At the height of the ending of the testis, the ejaculatory duct and intestine join, forming the cloaca, which opens at the posterior end of the male body. The distal cloacal tube contains the spicule. The spicule is surrounded by a shiny spicule sheath (Figure 3b, 3c) with a granular surface structure. The female posterior starts at the esophagus-intestinal junction, where the vulva is located (Figure 3d). The vagina shows thick walls, which connect back into the oviduct (Figure 3e). The cuticle of the anterior structure shows two patterns, one, which is striated with transverse grooves on one side and the other, a tuberculate band (Figure 3g). Within the anterior part is the stichosome, comprised of a row of stichocytes surrounding the esophagus which starts at the mouth (Figure 3i), and continues to the posterior part of the worm merging into the intestine (Figure 3d) and marking the division between anterior and posterior section. The ova present in the oviduct contained two polar plugs, were 68 μM long and 29 μM wide (Figure 3f) placing them closer to a larger observed egg population of two distinct ova groups of *T. trichiura* in literature. [20]

Figure 3: Morphology of *T. incognita*

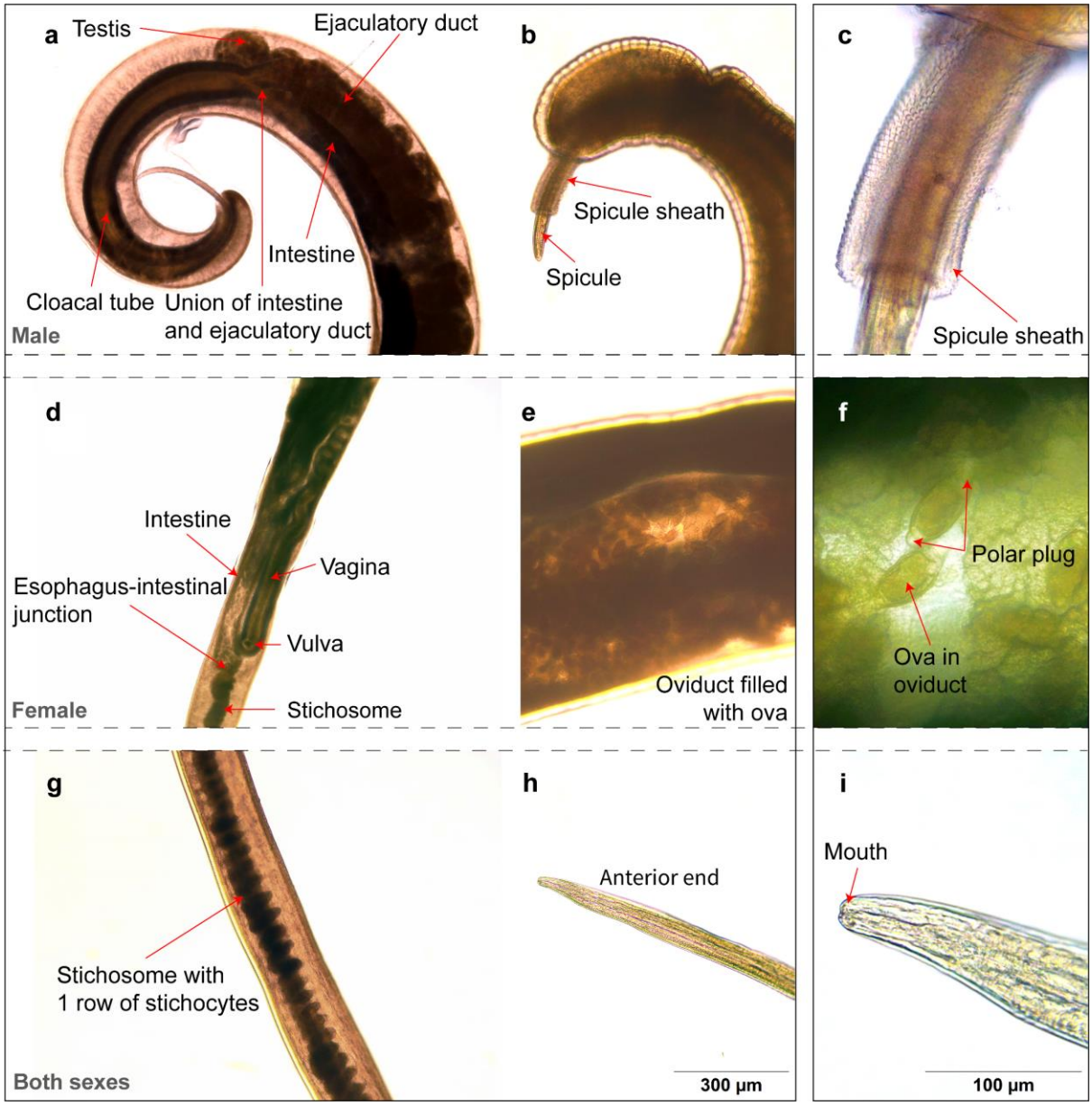


Figure 3: Morphology of *T. incognita*. **a** Posterior end showing male organs including testis, ejaculatory duct, intestine, union of intestine and ejaculatory duct, cloacal tube and spicule with a ruptured sheath. **b** Male spicule with an intact sheath partially covering the spicule. **c** 40X magnification of the shiny spicular sheath showing a surface structure. **d** Morphology showing female reproductive organs at the intersection between anterior and posterior part of the worm including stichosome, esophagus-intestinal junction, intestine, vagina and vulva. **e** Enlarged oviduct filled with *T. incognita* ova. **f** 40X magnification on ova showing polar plugs. The ova are 68 µM long and 29 µM wide in this image. **g** Anterior showing one row of stichocytes and 2 types of cuticle patterns. The left side showing transverse grooves and the right side the tuberculate band. **h** Anterior ending leading up to the mouth. **i** 40X magnification of the mouth.

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

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, compared to 80.57 Mb of *T. trichiura*, 63.84 Mb, 73.46 Mb and 70.45 Mb of *T. suis* and 111.83 Mb of *T. muris*. The *T. incognita* genome consisting of 387 contigs with an N50 of 11.78 Mb, and is presented in Table 1. 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*.

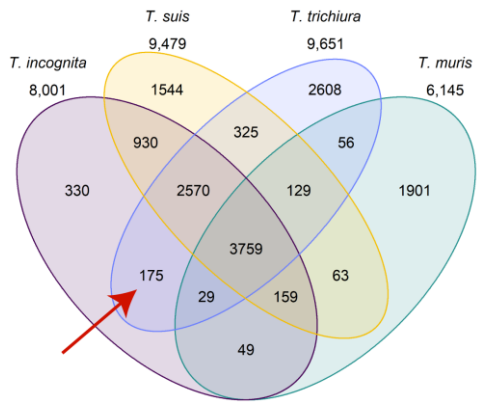
Table 1: Whole genome assembly metrics

	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 single 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	10	9.7	9.9	10.3	7
Missing BUSCOs [%]	24.6	23.9	24	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*

*bacterial artificial chromosome (BAC)

Table 2: Venn diagram of orthologous groups from predicted transcripts, where 175 were shared between *T. incognita* and *T. trichiura* and their GO enrichment analysis

Venn diagram



Category	GO term	p-value	Description
Cellular component	GO:0005615	3.10E-21	extracellular space
	GO:0000786	4.02E-17	nucleosome
Molecular function	GO:0003676	1.29E-52	nucleic acid binding
	GO:0004830	1.21E-33	tryptophan-tRNA ligase activity
	GO:0030527	2.78E-18	structural constituent of chromatin
	GO:0003723	6.72E-18	RNA binding
Biological process	GO:0006436	1.21E-33	tryptophanyl-tRNA aminoacylation
	GO:0044571	3.17E-19	[2Fe-2S] cluster assembly
	GO:0051259	3.17E-19	protein complex oligomerization

Availability of 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 Table 2. 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* or *T. muris*. A GO term enrichment analysis was performed on this subset of genes using the predicted genes in *T. incognita* as a background gene set and the full list of significant terms is provided in SI Table 3. A list of GO enriched terms provided in Table 2 indicate extracellular space to be the most significantly

enriched term in the category of cellular components along with the nucleosom. Nucleic acid binding, tryptophan-tRNA ligase activity, structural constituent of chromatin and RNA binding were the most significant in molecular functions and finally tryptophanyl-tRNA aminoacylation, [2Fe-2S] cluster assembly and protein complex oligomerization in biological process. Finally, considering that the anterior section is in a most intimate relationship with the host, we searched for genes within this gene-set which Jex *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.[17] We identified one family of each serpins (OG0000305) and chymotrypsin-like serine proteases (OG0010319) that are shared only by *T. trichiura* and *T. incognita* (SI Figure 4). Additionally, we identified, a family of tetraspanins (OG0009162) which are investigated as vaccine candidates in schistosomiasis.[21,22]

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

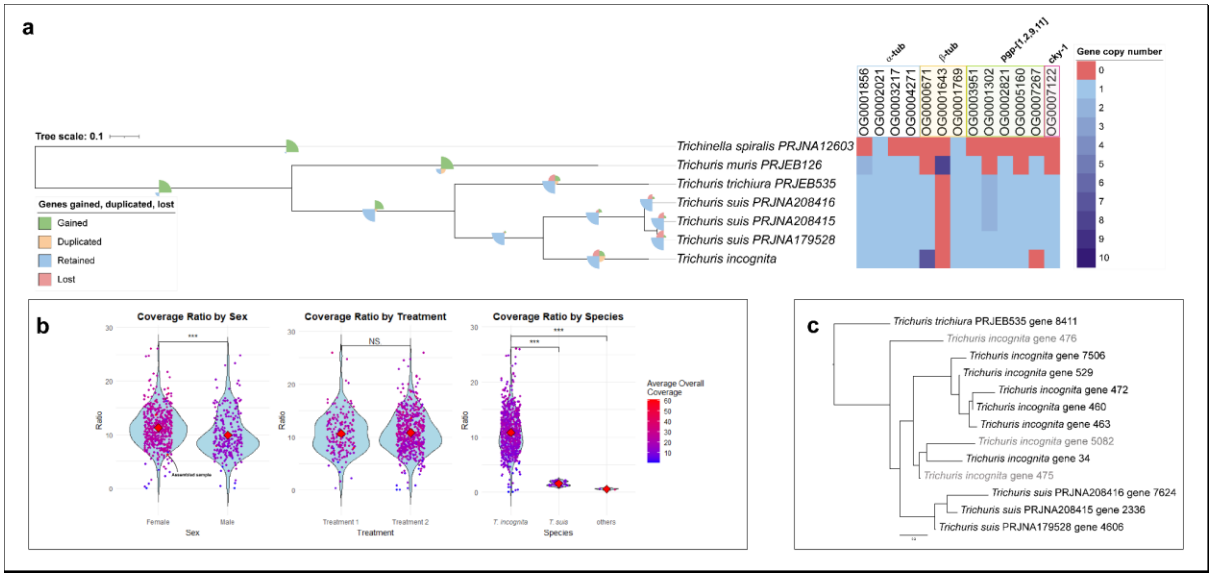
The species tree of the *Trichuris* genus was inferred using species tree inference from all genes (STAG)[23,24] with 12434 orthologous groups of the braker2 predicted genes in the publicly available whole genomes of *Trichuris* species, *T. incognita* and *Trichinella spiralis*. The species tree shows 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.[25,26] Amongst these genes, one orthologuous group (OG0000671) of the TUBB 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 5 shows the 7 predicted TUBB 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 TUBB genes as shown in SI Figure 5. To further investigate the existence of these predicted variants especially if the variants of the 194 position were reproducible in an independent Nanopore sequencing experiment, a region was amplified that was present in all genes and contained the predicted variants (primer sequences are provided in SI Figure 6). As shown in SI Figure 6, the truncation was only present in the *in silico* prediction and further transcriptomic investigations are required as our sequencing experiments are of genomic data and not transcripts. 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, indicating that out all copies of the TUBB gene, one copy is present in the Asn form. 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 site 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.

Figure 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* (n=747) to *T. suis* (n=26) from Côte d'Ivoire and 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 from a two tailed t-test. **c** TUBB gene tree of OG0000671, timescale indicates numbers of substitutions per site.

We then sought to conduct an unbiased investigation of gene families which are as expanded or more expanded compared to the β-tubulin gene in *T. incognita* and identified gene families associated with microtubule function and mitotic processes. 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 (SI Table 4). 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 TUBB 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 and MAP65/Ase1, which promotes microtubule flexibility and prevents microtubule severing by cross-linking.[27,28] Finally, Condensin, Transcription factor (TF) IIIC and a haspin like kinases are involved in chromatid condensation, organization and cohesion respectively.

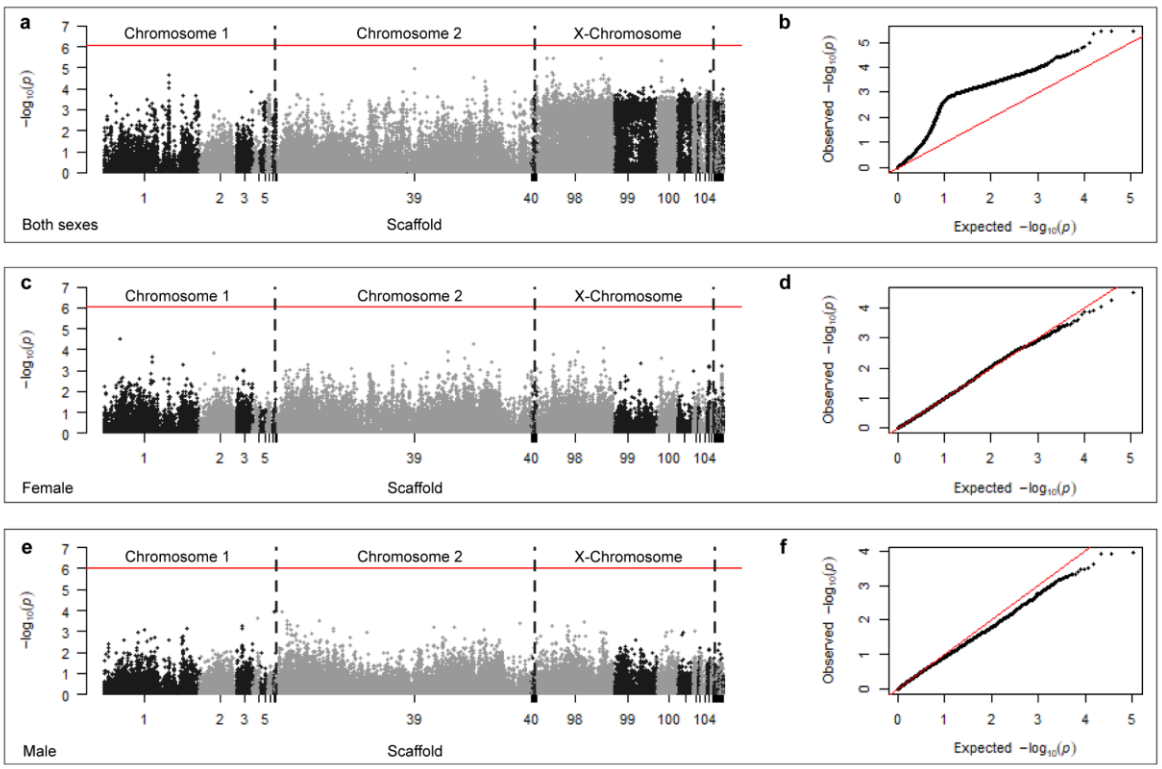
To extrapolate the duplication events from the assembled genome of one single worm to the population, we leveraged the ratio of the average coverage of illumina reads of the amplified TUBB region compared to the longest autosomal contig present in the genome of all 747 worms as shown in Figure 4b. Intriguingly, 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 of 26 *T. suis* worms isolated in Côte d'Ivoire was used as a comparison, where, intriguingly, only 1-2 copies of the TUBB was present. Thus, one unique feature about *T. incognita* is the high copy number of the TUBB gene. The coverage ratios were further compared to coverage ratios from publically available raw sequencing data that 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 indicating 1 copy of the TUBB gene.

Genome wide association study does not provide substantial evidence of adaptation to drug treatment

To investigate potential missense SNPs located on predicted genes and associated with resistance, a genome wide association study (GWAS) was conducted with sequencing data of 747 individual worms, comparing the populations isolated after T1 and T2 respectively of which 492 drug non-sensitive and 160 drug sensitive isolated passed the filters as specified in the methodology section. Both a logistical and linear model was used to investigate 56273 missense mutations and a genome wide significance level of 8.89E-7 was established using the Bonferroni correction at α=0.05 and 56273 mutations. In a second run, population stratification was taken into account by incorporating the first principle component in the analysis. QQ-plots showed consistent artificial inflation when using combined data of male and female worms as seen in Figure 5b. However, when running only male or female isolates, the artificial inflation was mitigated (Figure 5d, 5f). The QQ-plots were most consistent in the logistic model without population stratification as shown in Figure 5, however, no SNPs with genome wide significance were identified. Using a linear model without population stratification (SI Figure 7), a total of 5 genes carried SNPs with genome wide significance. Significant genes included g598, where no pfam annotation was found, g1428, which contains a Cadherin-like domain, g1478, which contains a fragile site-associated protein C-terminus domain and a Kiaa1109 N-terminal region, g2562, which contains a pao retrotransposon peptidase domain and finally g11724, which contains a RNA recognition motif (a.k.a. RRM, RBD, or RNP domain). Gene sequences can be found in the Trichuris_Incognita_Proteins.fasta file provided. No further missense SNPs were identified with a genome wide significance, challenging previously reported adaptation to drug treatment at the Phe168, Glu198 or Phe200 positions of TUBB in *T. trichiura*. The strongest association of a missense 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 SNPs were found on expanded gene families.

Figure 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 missense SNPs associated to treatment outcome (T1 vs T2) using a logistic model in PLINK. Variants were filtered at a minor allele frequency of 0.05, a missing genotype frequency of 0.05, and a deviation from the Hardy-Weinberg Equilibrium at a significance level of p=0.001. The red line indicates the Bonferroni corrected significance of 8.89E-7 at α=0.05 and 56273 mutations. The scaffolds are separated by black/grey colors. **a** and **b** Data of both sexes showing an artificial inflation of p-values in the QQ-plot. 190 males, 462 females, 16 ambiguous, 492 cases as in drug non-sensitive and 160 controls. **c** and **d** Only female SNPs. 0 males, 462 females, 328 cases as in drug non-sensitive and 134 controls. **e** and **f** Only male SNPs. 190 males, 0 females, 164 cases as in drug non-sensitive and 26 controls. No mutations were associated with the grouping into T1 and T2 at genome wide significance using the logistic model.

Discussion:

Our investigation and the presented data represent a milestone in *Trichuris* research. We present for the first time the discovery and description of a newly occurring species of *Trichuris* - *Trichuris incognita* - in humans. This finding challenges the current understanding of species within the *Trichuris* genus, which assumes that only one species of *Trichuris* infects humans, and therefore carries profound implications for both parasitology and public health.

The existence of a species of *Trichuris*, phylogenetically distinct from the canonical human-infective species, has been proposed in recent studies involving amplicon sequencing by Venkatesan *et al*. and Rahman *et al*.[10,29] Adopting the genetic species concept, 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.[13,15,17] 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.[15] The prediction of genes in the publicly available *T. trichiura*, *T. suis*, *T. muris* and *Trichinella spiralis* genomes, and the newly assemble *T. incognita* genome, using the same reference database, allowed the exploration of orthologous gene groups and duplication events across the species tree. The

identification of a family of serpins and chemotryphsin-like serine proteases uniquely present in human infecting *Trichuris* may be of particular interest, as these proteins, which were identified to be upregulated in the stichosome of *T. suis*, have gained particular interest in the field of immunomodulatory diseases in general and vaccine development for *T. spiralis*.[17,30,31]

Our investigation of duplication events across the species tree revealed a unique characteristic of *T. incognita*: A high copy number of TUBB genes. Duplication events are encountered across the tree of life and constitute an important mechanism to provide new genetic material. In eukaryotes, including *C. elegans*, gene duplication rates are estimated to be as frequent as single-nucleotide polymorphisms. However, they have never been investigated within the *Trichuris* genus.[32,33] 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.[34,35] 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. This is in contrast to the low copy number of TUBB in *T. trichiura*. It is of note that *T. trichiura* also does not respond well to single benzimidazole treatment. Leveraging coverage ratios, duplication events could be confirmed in the whole population for each individual worm. Notably, while β-tubulin exhibited significant duplication, α-tubulin did not. Furthermore, massively expanded gene families are associated with microtubule dynamics and mitotic processes. Kinesin, 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.[36] MAP65/Ase1, a microtubule-associated protein is reported to promote microtubule flexibility and prevent microtubule severing by cross-linking.[27,28] These findings suggest that the duplications observed are associated to repairing and protecting microtubule functionality which is intriguing considering the mode of action of albendazole, inhibiting microtubule formation.[37,38] 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.[39-41] *T. muris* is the only other organism with a high TUBB copy number and it has been shown that albendazole or mebendazole have no effect on *T. muris* with an $IC_{50} > 200$ μg/mL.[42] However, comparing the TUBB copy number of the population obtained after T1 to the population obtained after T2 showed no significant difference indicating that duplications are not the result of a recent adaptation. 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 back several centuries.

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. 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 from *T. suis* to *T. incognita* have been reported, these findings suggest that there was a zoonotic event sometime in the past between the human and swine host that was successful, but could not be observed currently.[43] However, the occurrence of closely related *Trichuris* isolates from the colobus monkeys (*Colobus guereza kikuyensis*) raises the question, if these might belong to the same species of *Trichuris* and have zoonotic potential similar to *T. trichiura* species, which are found both within primates and humans.[13] Furthermore, J. Rivero *et al.* identified further *Trichuris* species, which were closely related to the one found in the colobus monkey and may be the same species as *T. incognita* based on

ITS sequences or *cox*1 and *co*b genes.[19] If *T. incognita* were able to infect non-human hosts, which is the case for *T. trichiura,* this would provide a refugia from drug pressure, which might prevent the establishment of anthelmintic resistance.[44]

With the unique experimental setup of the expulsion study, we were able to categorize helminths into drug sensitive and non-sensitive and thus enabling a genome wide association study (GWAS) to investigate the potential establishment of resistance to albendazole-ivermectin. MDA has been ongoing in Côte d'Ivoire for multiple decades,[4] which is well within the timeframe of a possible resistance establishment in STH through PC.[7] The most surprising finding of the GWAS may be, that there is no conclusive evidence to show a clear adaptation to drug treatment in this new species. Depending on the regression model chosen for the GWAS, logistic vs linear, and the inclusion of principal components to account for population stratification or not, SNPs of genome wide significance with acceptable QQ-plots were only observed using a linear model in the male population without considering population stratification. One interesting feature of an identified gene, g1428, is the presence of a cadherin-like domain, as cadherin expression has been reported to be affected by ivermectin in cancer cells.[45] However, while these hits will be subject of future investigation, there is no clear indication of resistance establishment in the analysis conducted. 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[26] and transmembrane proteins such as the chloride ion channel avr-15,[46] all of which did not show an association with genome wide significance in this study.

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 were expulsed, which is in-line with previously observed low cure rates in Côte d'Ivoire.[8,47] The female to male ratio is comparable to recent reports for *T. trichiura*.[48] 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*.[48] Overall, these results suggest that the biology of *T. incognita* in terms of fecundity is comparable to *T. trichiura*. Also morphologically, *T. incognita* and *T. trichiura* were comparable. The presented morphological characteristics of *T. incognita* do not enable a morphological distinction to *T. trichiura* indicating the limits of the morphological species concept *Trichuris* genus.[19] 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.

In summary, our study presents an in-depth characterization of *T. incognita* found to be less susceptible to albendazole-ivermectin treatment. This species reveals a high TUBB copy number, yet the GWAS does not provide conclusive evidence of adaptation to drug pressure. The whole genome presented in this study and the sequencing dataset on drug sensitive and non-sensitive individual worms, isolated from humans, is the first of its kind and provides a unique data resource to investigate resistance establishment in *Trichuris*, in addition to drug targets and vaccine candidates in the field of trichuriasis.

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 (2018). https://doi.org/10.1016/s0140-6736(18)32279-7

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 (2020). https://doi.org/10.1371/journal.pntd.0008505

3 World Health, O. & Onchocerciasis Control Programme in the Volta River Basin, A. Drug development. (Onchocerciasis Control Programme in the Volta River Basin Area, Ouagadougou, 1976).

4 Bland, J., Viedma, C., Davies, P. & Ozorio, P. This wormy world [full issue]. *World health*, 2-31 (1984).

5 Ogoussan, K. T. & Hopkins, A. Mectizan((R)) procurement and delivery for onchocerciasis mass drug administration programmes. *Acta Trop* **120 Suppl 1**, S173-176 (2011). https://doi.org/10.1016/j.actatropica.2010.10.004

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

7 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 (2024). https://doi.org/10.1038/s41467-024-45027-2

8 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 (2022). https://doi.org/10.1016/s1473-3099(21)00421-7

9 Venkatesan, A. *et al.* Trichuriasis in Human Patients from Cote d'Ivoire Caused by Novel Trichuris incognita Species with Low Sensitivity to Albendazole/Ivermectin Combination Treatment. *Emerg Infect Dis* **31**, 104-114 (2025). https://doi.org/10.3201/eid3101.240995

10 Rahman, N. *et al.* Nanopore-based analysis unravels the genetic landscape and phylogenetic placement of human-infecting Trichuris species in Côte d'Ivoire, Tanzania, Uganda, and Laos. *bioRxiv*, 2024.2007.2031.605962 (2024). https://doi.org/10.1101/2024.07.31.605962

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

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

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

14 Else, K. J. *et al.* Whipworm and roundworm infections. *Nature Reviews Disease Primers* **6** (2020). https://doi.org/10.1038/s41572-020-0171-3

15 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 (2014). https://doi.org/10.1038/ng.3010

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

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

18 Chakraborty, P., Aravindhan, V. & Mukherjee, S. Helminth-derived biomacromolecules as therapeutic agents for treating inflammatory and infectious diseases: What lessons do we get

483 from recent findings? *Int J Biol Macromol* **241**, 124649 (2023).
484 https://doi.org/10.1016/j.ijbiomac.2023.124649
485 19 Rivero, J., Cutillas, C. & Callejon, R. Trichuris trichiura (Linnaeus, 1771) From Human and Non-
486 human Primates: Morphology, Biometry, Host Specificity, Molecular Characterization, and
487 Phylogeny. *Front Vet Sci* **7**, 626120 (2020). https://doi.org/10.3389/fvets.2020.626120
488 20 Yoshikawa, H., Yamada, M., Matsumoto, Y. & Yoshida, Y. Variations in egg size ofTrichuris
489 trichiura. *Parasitology Research* **75**, 649-654 (1989). https://doi.org/10.1007/BF00930964
490 21 Tran, M. H. *et al.* Tetraspanins on the surface of Schistosoma mansoni are protective antigens
491 against schistosomiasis. *Nat Med* **12**, 835-840 (2006). https://doi.org/10.1038/nm1430
492 22 Susa, K. J., Kruse, A. C. & Blacklow, S. C. Tetraspanins: structure, dynamics, and principles of
493 partner-protein recognition. *Trends Cell Biol* (2023).
494 https://doi.org/10.1016/j.tcb.2023.09.003
495 23 Emms, D. M. & Kelly, S. (2018).
496 24 Emms, D. M. & Kelly, S. STRIDE: Species Tree Root Inference from Gene Duplication Events.
497 *Molecular Biology and Evolution* **34**, 3267-3278 (2017).
498 https://doi.org/10.1093/molbev/msx259
499 25 Laing, R. *et al.* Transcriptomic analyses implicate neuronal plasticity and chloride homeostasis
500 in ivermectin resistance and response to treatment in a parasitic nematode. *PLoS Pathog* **18**,
501 e1010545 (2022). https://doi.org/10.1371/journal.ppat.1010545
502 26 Doyle, S. R. *et al.* Genomic landscape of drug response reveals novel mediators of anthelmintic
503 resistance. *Preprint* (2022). https://doi.org/10.1101/2021.11.12.465712
504 27 Burkart, G. M. & Dixit, R. Microtubule bundling by MAP65-1 protects against severing by
505 inhibiting the binding of katanin. *Mol Biol Cell* **30**, 1587-1597 (2019).
506 https://doi.org/10.1091/mbc.E18-12-0776
507 28 Portran, D. *et al.* MAP65/Ase1 promote microtubule flexibility. *Mol Biol Cell* **24**, 1964-1973
508 (2013). https://doi.org/10.1091/mbc.E13-03-0141
509 29 Venkatesan, A. *et al.* (2024). https://doi.org/10.1101/2024.07.04.602123
510 30 Mkaouar, H. *et al.* Gut Serpinome: Emerging Evidence in IBD. *Int J Mol Sci* **22** (2021).
511 https://doi.org/10.3390/ijms22116088
512 31 Xu, J. *et al.* Influence of adjuvant formulation on inducing immune response in mice
513 immunized with a recombinant serpin from Trichinella spiralis. *Parasite Immunol* **39** (2017).
514 https://doi.org/10.1111/pim.12437
515 32 Lynch, M. & Conery, J. S. The Evolutionary Fate and Consequences of Duplicate Genes. *Science*
516 **290**, 1151-1155 (2000). https://doi.org/10.1126/science.290.5494.1151
517 33 Kuzmin, E., Taylor, J. S. & Boone, C. Retention of duplicated genes in evolution. *Trends Genet*
518 **38**, 59-72 (2022). https://doi.org/10.1016/j.tig.2021.06.016
519 34 Kondrashov, F. A. Gene duplication as a mechanism of genomic adaptation to a changing
520 environment. *Proc Biol Sci* **279**, 5048-5057 (2012). https://doi.org/10.1098/rspb.2012.1108
521 35 Zajac, N. *et al.* Gene Duplication and Gain in the Trematode Atriophallophorus winterbourni
522 Contributes to Adaptation to Parasitism. *Genome Biol Evol* **13** (2021).
523 https://doi.org/10.1093/gbe/evab010
524 36 Cushion, T. D., Leca, I. & Keays, D. A. MAPping tubulin mutations. *Front Cell Dev Biol* **11**,
525 1136699 (2023). https://doi.org/10.3389/fcell.2023.1136699
526 37 Enos, A. & Coles, G. C. Effect of benzimidazole drugs on tubulin in benzimidazole resistant and
527 susceptible strains of Caenorhabditis elegans. *Int J Parasitol* **20**, 161-167 (1990).
528 https://doi.org/10.1016/0020-7519(90)90096-6
529 38 Dawson, P. J., Gutteridge, W. E. & Gull, K. A comparison of the interaction of anthelmintic
530 benzimidazoles with tubulin isolated from mammalian tissue and the parasitic nematode
531 Ascaridia galli. *Biochem Pharmacol* **33**, 1069-1074 (1984). https://doi.org/10.1016/0006-
532 2952(84)90515-x

533 39 Hagstrom, K. A. & Meyer, B. J. Condensin and cohesin: more than chromosome compactor
534 and glue. *Nat Rev Genet* **4**, 520-534 (2003). https://doi.org/10.1038/nrg1110
535 40 Stutzman, A. V., Liang, A. S., Beilinson, V. & Ikegami, K. Transcription-independent TFIIIC-
536 bound sites cluster near heterochromatin boundaries within lamina-associated domains in C.
537 elegans. *Epigenetics Chromatin* **13**, 1 (2020). https://doi.org/10.1186/s13072-019-0325-2
538 41 Dai, J., Kateneva, A. V. & Higgins, J. M. Studies of haspin-depleted cells reveal that spindle-
539 pole integrity in mitosis requires chromosome cohesion. *J Cell Sci* **122**, 4168-4176 (2009).
540 https://doi.org/10.1242/jcs.054122
541 42 Keiser, J., Tritten, L., Adelfio, R. & Vargas, M. Effect of combinations of marketed human
542 anthelmintic drugs against Trichuris muris in vitro and in vivo. *Parasites & Vectors* **5**, 292
543 (2012). https://doi.org/10.1186/1756-3305-5-292
544 43 Nejsum, P., Betson, M., Bendall, R. P., Thamsborg, S. M. & Stothard, J. R. Assessing the
545 zoonotic potential of Ascaris suum and Trichuris suis: looking to the future from an analysis of
546 the past. *J Helminthol* **86**, 148-155 (2012). https://doi.org/10.1017/S0022149X12000193
547 44 Vercruysse, J. *et al.* Is anthelmintic resistance a concern for the control of human soil-
548 transmitted helminths? *Int J Parasitol Drugs Drug Resist* **1**, 14-27 (2011).
549 https://doi.org/10.1016/j.ijpddr.2011.09.002
550 45 Tang, M. *et al.* Ivermectin, a potential anticancer drug derived from an antiparasitic drug.
551 *Pharmacol Res* **163**, 105207 (2021). https://doi.org/10.1016/j.phrs.2020.105207
552 46 Dent, J. A., Smith, M. M., Vassilatis, D. K. & Avery, L. The genetics of ivermectin resistance in
553 Caenorhabditis elegans. *Proceedings of the National Academy of Sciences* **97**, 2674-2679
554 (2000). https://doi.org/10.1073/pnas.97.6.2674
555 47 Sprecher, V. P., Coulibaly, J. T., Hurlimann, E., Hattendorf, J. & Keiser, J. Efficacy and Safety of
556 Moxidectin-Albendazole and Ivermectin-Albendazole Combination Therapy Compared to
557 Albendazole Monotherapy in Adolescents and Adults Infected with Trichuris trichiura: A
558 Randomized, Controlled Superiority Trial. *Clin Infect Dis* **77**, 1294-1302 (2023).
559 https://doi.org/10.1093/cid/ciad387
560 48 Hansen, E. P. *et al.* Faecal egg counts and expulsion dynamics of the whipworm, Trichuris
561 trichiura following self-infection. *J Helminthol* **90**, 298-302 (2016).
562 https://doi.org/10.1017/S0022149X1500019X
563 49 Cringoli, G. *et al.* The Mini-FLOTAC technique for the diagnosis of helminth and protozoan
564 infections in humans and animals. *Nat Protoc* **12**, 1723-1732 (2017).
565 https://doi.org/10.1038/nprot.2017.067
566 50 Di Tommaso, P. *et al.* Nextflow enables reproducible computational workflows. *Nat*
567 *Biotechnol* **35**, 316-319 (2017). https://doi.org/10.1038/nbt.3820
568 51 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence
569 data. *Bioinformatics* **30**, 2114-2120 (2014). https://doi.org/10.1093/bioinformatics/btu170
570 52 Jin, J. J. *et al.* GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of
571 organelle genomes. *Genome Biol* **21**, 241 (2020). https://doi.org/10.1186/s13059-020-02154-
572 5
573 53 Meng, G., Li, Y., Yang, C. & Liu, S. MitoZ: a toolkit for animal mitochondrial genome assembly,
574 annotation and visualization. *Nucleic Acids Res* **47**, e63 (2019).
575 https://doi.org/10.1093/nar/gkz173
576 54 Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7:
577 improvements in performance and usability. *Mol Biol Evol* **30**, 772-780 (2013).
578 https://doi.org/10.1093/molbev/mst010
579 55 Bouckaert, R. *et al.* BEAST 2.5: An advanced software platform for Bayesian evolutionary
580 analysis. *PLoS Comput Biol* **15**, e1006650 (2019).
581 https://doi.org/10.1371/journal.pcbi.1006650
582 56 De Coster, W. & Rademakers, R. NanoPack2: population-scale evaluation of long-read
583 sequencing data. *Bioinformatics* **39** (2023). https://doi.org/10.1093/bioinformatics/btad311

584 57 Sun, J., Li, R., Chen, C., Sigwart, J. D. & Kocot, K. M. Benchmarking Oxford Nanopore read
585 assemblers for high-quality molluscan genomes. *Philos Trans R Soc Lond B Biol Sci* **376**,
586 20200160 (2021). https://doi.org/10.1098/rstb.2020.0160
587 58 Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome*
588 *Biol* **20**, 257 (2019). https://doi.org/10.1186/s13059-019-1891-0
589 59 Zimin, A. V. *et al.* The MaSuRCA genome assembler. *Bioinformatics* **29**, 2669-2677 (2013).
590 https://doi.org/10.1093/bioinformatics/btt476
591 60 Alonge, M. *et al.* Automated assembly scaffolding using RagTag elevates a new tomato system
592 for high-throughput genome editing. *Genome Biol* **23**, 258 (2022).
593 https://doi.org/10.1186/s13059-022-02823-7
594 61 Shumate, A. & Salzberg, S. L. Liftoff: accurate mapping of gene annotations. *Bioinformatics* **37**,
595 1639-1643 (2021). https://doi.org/10.1093/bioinformatics/btaa1016
596 62 Manni, M., Berkeley, M. R., Seppey, M., Simao, F. A. & Zdobnov, E. M. BUSCO Update: Novel
597 and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring
598 of Eukaryotic, Prokaryotic, and Viral Genomes. *Mol Biol Evol* **38**, 4647-4654 (2021).
599 https://doi.org/10.1093/molbev/msab199
600 63 Mikheenko, A., Prjibelski, A., Saveliev, V., Antipov, D. & Gurevich, A. Versatile genome
601 assembly evaluation with QUAST-LG. *Bioinformatics* **34**, i142-i150 (2018).
602 https://doi.org/10.1093/bioinformatics/bty266
603 64 Kuznetsov, D. *et al.* OrthoDB v11: annotation of orthologs in the widest sampling of organismal
604 diversity. *Nucleic Acids Res* **51**, D445-D451 (2023). https://doi.org/10.1093/nar/gkac998
605 65 Bruna, T., Hoff, K. J., Lomsadze, A., Stanke, M. & Borodovsky, M. BRAKER2: automatic
606 eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein
607 database. *NAR Genom Bioinform* **3**, lqaa108 (2021). https://doi.org/10.1093/nargab/lqaa108
608 66 Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative
609 genomics. *Genome Biol* **20**, 238 (2019). https://doi.org/10.1186/s13059-019-1832-y
610 67 Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree
611 display and annotation. *Nucleic Acids Res* **49**, W293-W296 (2021).
612 https://doi.org/10.1093/nar/gkab301
613 68 Altenhoff, A. M. *et al.* OMA standalone: orthology inference among public and custom
614 genomes and transcriptomes. *Genome Res* **29**, 1152-1163 (2019).
615 https://doi.org/10.1101/gr.243212.118
616 69 Train, C. M., Pignatelli, M., Altenhoff, A. & Dessimoz, C. iHam and pyHam: visualizing and
617 processing hierarchical orthologous groups. *Bioinformatics* **35**, 2504-2506 (2019).
618 https://doi.org/10.1093/bioinformatics/bty994
619 70 Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform.
620 *Bioinformatics* **26**, 589-595 (2010). https://doi.org/10.1093/bioinformatics/btp698
621 71 Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage
622 analyses. *Am J Hum Genet* **81**, 559-575 (2007). https://doi.org/10.1086/519795

623

624

625

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*.[8] 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.[49] 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).[50] 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);[51] 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.[52] 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).[53] Coding genes were then extracted and aligned as amino acid and nucleic acid sequences using MAFFT (version 7.490).[54] 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).[55] 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. 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.[56,57] 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.[58] 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) which allowed for paired end illumina reads and nanopore reads.[59] A total of 1564 contigs were assembled and the raw assembly is provided on github. The BUSCO showed a completeness of 64.3% of complete BUSCOs, 11.2% fragmented and 24.5% missing and the N50 was 0.26 Mb. The outputted contigs were channeled into RagTag (version 2.1.0) for correction and scaffolding using the *Trichuris trichiura* (PRJEB535) reference genome.[60] The corrected scaffolds were then assigned to one of three linkage groups from T. Muris using Liftoff (version 1.6.3),[61] as described in literature.[13] The genome was frozen for quality assessment using BUSCO (version 5.1.2 with Metazoa_obd10 database)[62] and QUAST (version 5.0.2).[63] 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).[64,65] 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, orthologous group functional annotation and GO term enrichment analysis

Orthofinder (version 2.5.5) was used to infer orthology and the species tree of the braker2 predicted genes.[66] 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 WormBase ParaSite is provided. Species tree with heat map was visualized using iTOL (version 6.8.1).[67] 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).[68,69] 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). GO terms were extracted from the InterProScan annotated genes file of *T. incognita* (background GO terms) and of *T. incognita* genes shared with *T. trichiura* but not with *T. suis* or *T. muris* (GO terms of interest). A fishers exact test was performed on the GO term counts of each GO term in the intersection compared to the background GO terms. All gene lists of T. incognita and the InterProScan annotations are provided on github.

Pipeline for variant calling

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[13] full code of which are provided online (https://github.com/max-baer). Raw reads were first processed using Trimmomatic (version 0.39).[51] BWA (version 0.7.17) and SAMtools (version 1.14) were used to index the assembled reference genome, and align, sort and filter raw reads.[70] Flagstat and Kraken2 (version 2.1.1) were used as QC tools. The coverage ratio was used to classify the gender of each worm by analyzing the coverage ratio of the largest X-chromosome scaffold to the largest scaffold on chromosome 2 and is provided on github. Next, the standard GATK (version 4.2.6.1) pipeline was followed for variant calling, base quality score recalibration (BQSR) and genotyping. First, duplicates were marked using the MarkDuplicatesSpark option in GATK. Basecalling was done in a first round to obtain a subset of very likely SNPs for base quality score recalibration. Plots were generated for QUAL, DP, QD, FS, MQ, MQRankSum, SQR, ReadPosRankSum. Heterozygosity was set to 0.015 and indel heterozygosity to 0.01.[13] Initial hard filtering of the variants for BQSR was done in a way to exclude 5% of SNVs with the following lowest quality metrics: QUAL < 194 || QD < 10.3 || MQ < 43.61 || FS > 9.9 || SOR > 2.30 || MQRankSum < -1.269 || ReadPosRankSum < -1.174 for SNPs and QUAL < 101 || QD < 5.3 || FS > 11.4 || ReadPosRankSum < -1.336 for indels. Next, BQSR was conducted before generating a merged gVCF which was used for genotyping and then filtered for SNVs using the following hard filter expression: QUAL < 35 || QD < 1.3 || MQ < 28.76 || FS > 50.5 || SOR > 2.30 || MQRankSum < -4.261 || ReadPosRankSum < -1.421 for SNPs and QUAL < 32 || QD < 0.5 || FS > 30.1 || ReadPosRankSum < -1.836 for indels. Idels were filtered out and the variants were further filtered using the genotype filter expression DP < 3 and a minor allele frequency of 0.02 which resulted in a total of 6’282’017 SNPs. Next, genes in the VCF file were annotated using BCFtools (version 1.15) and BEDTools (version 2.30.0) and the braker2 produced gff file containing all genes of *T. incognita*. Using SnpEff (version 5.2c) a total of 111994 missense SNPs were identified on the genes of *T. incognita* leading to a different amino acid. This set of SNPs was used as an input in the GWAS.

Pipeline for genome wide association study

Plink (version 1.9) was used to conduct the genome wide association study.[71] First the phenotype data on treatment one (T1) or treatment 2 (T2) was added using the “--pheno” option. Next the variants were filtered using a minor allele frequency of 0.05, a missing genotype frequency of 0.05, and a deviation from the Hardy-Weinberg Equilibrium at a significance level of p=0.001 resulting in 56273 variants which were used for association testing. The worm sex data generated from coverage ratios in the previous pipeline was used to add the worm sex to each sample with the “--update-sex” option. First a simple GWAS was run using linear and logistic model with the “--assoc" or “--logistic” option respectively, with both sexes, only females and only males. Next, a principal component analysis was conducted to account for population stratification. The option “--pca" was used on the file containing all filtered variants. Analysis of the principle components revealed an eigenvalue of 52.7 for the

first principle component and 3.1 for the second, indicating that the consideration of only the first principle component will account for most population stratification. Next the logistic and linear model was run under consideration of the first principle component by using the options "--covar" and "--covar-number 1" and the "--linear" or "--logistic" option for the linear or logistic model respectively for both sexes, only females and only males respectively. To visualize the results and generate the QQ-plots the R package qqman was used.

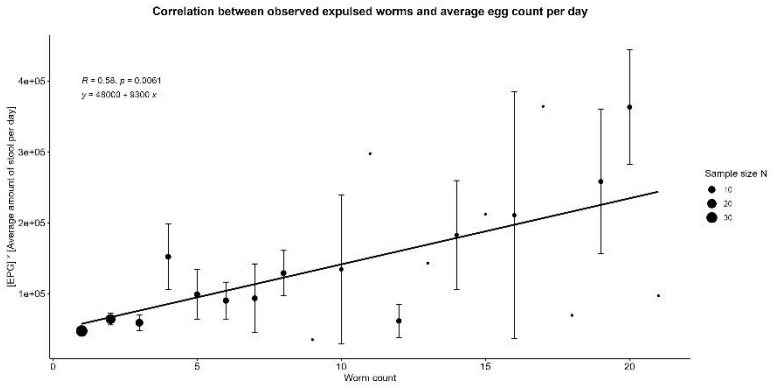
Data availability

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.

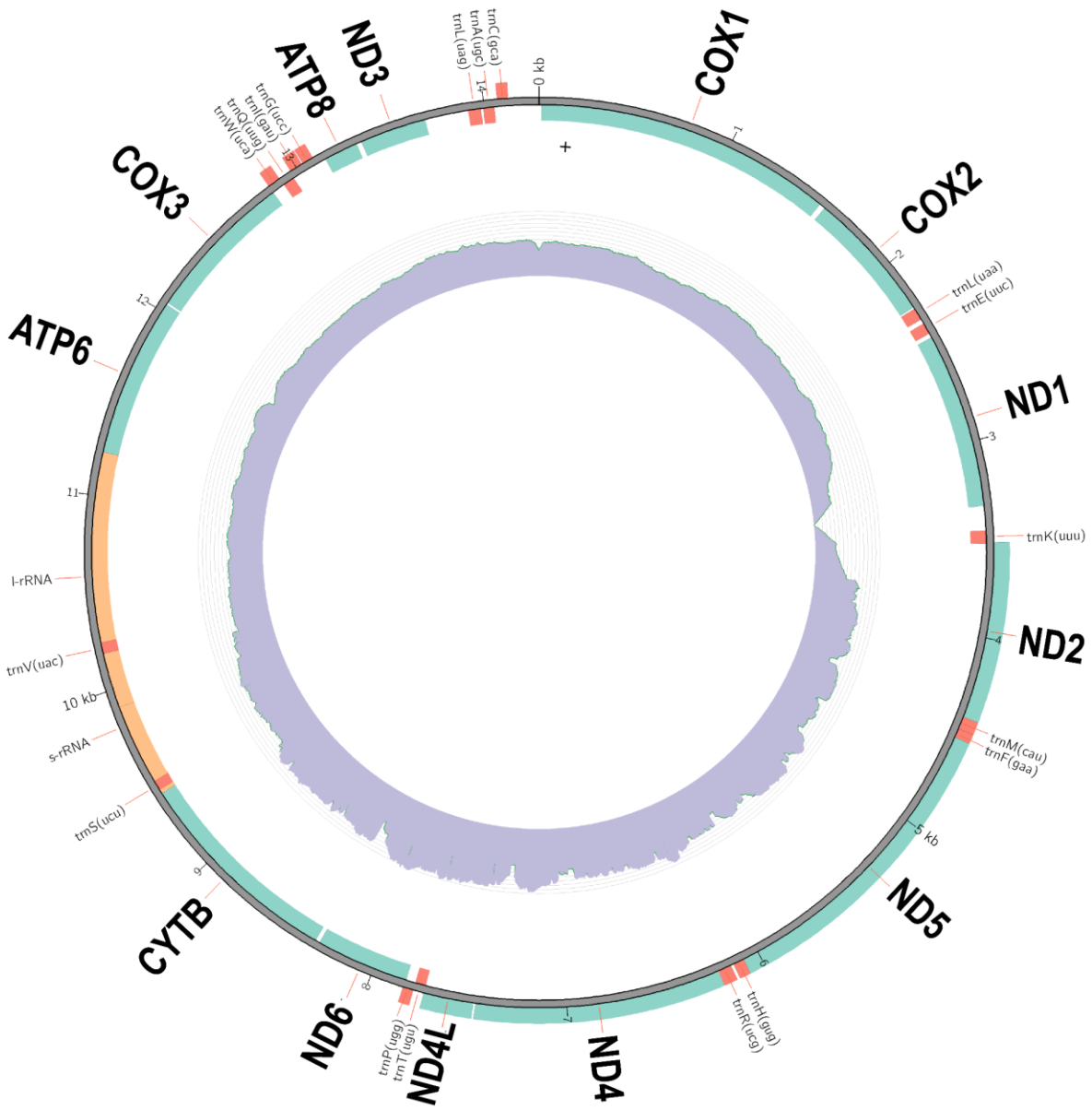
Code availability

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.

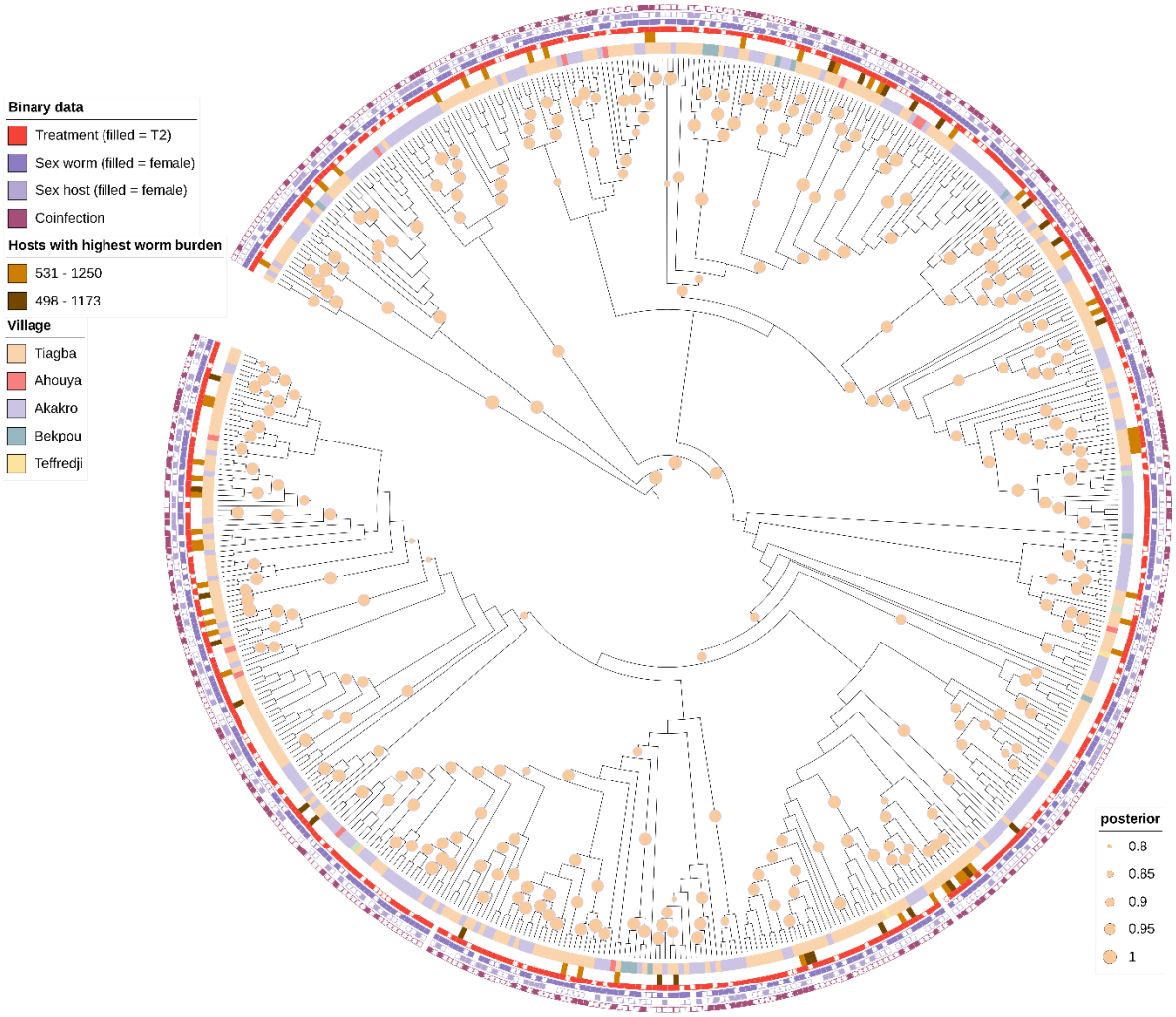
Figures



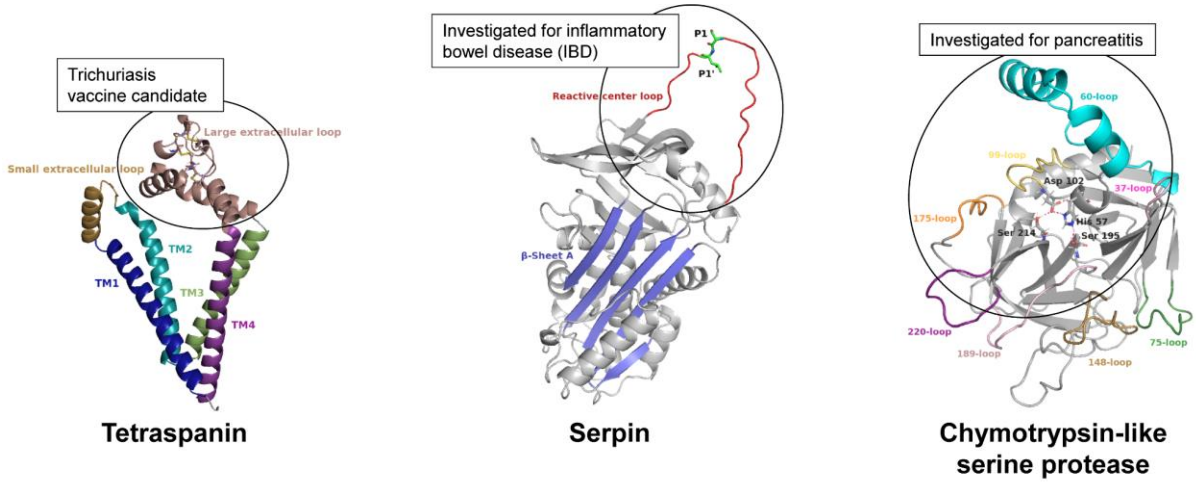
SI Figure 1: Correlation between observed expulsed worms and average egg count per day. Error bars indicate standard error with no error bar present if there was only one value.



SI Figure 2: Representative image of a reconstructed mitochondrial genome containing 13 protein coding genes: *cox*1-3, *nad*1-6, *nad*4L, *atp*6, *atp*8 and *co*B. Turquois represents the coding sequence of genes, the inner ring represents the normalized coverage, light orang represents s- and l-rRNA



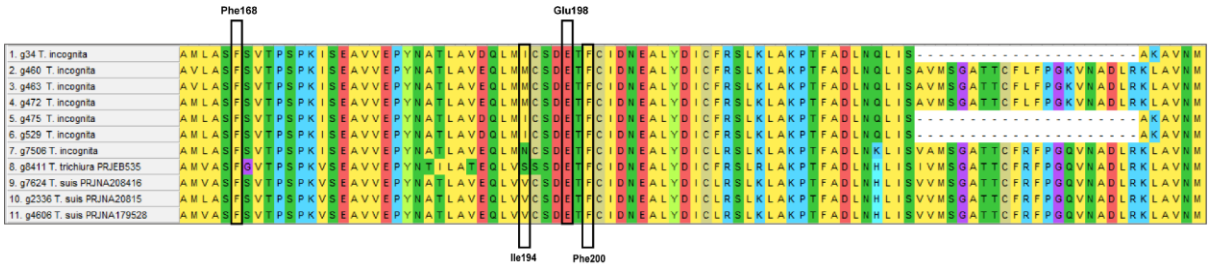
SI Figure 3: Phylogenetic tree based on mitochondrial sequences from 535 isolated *T. incognita* worms from human hosts. The circle on branches indicates the posterior distribution. The outmost 4 rings represent binary metadata. A filled red tile indicates that the worm was isolated after T2 where the empty tile indicates isolation after T1. The dark violet tile indicates the worm sex and the light violet tile the host sex while purple indicates if a coinfection with other worms was observed. The middle ring indicates two hosts: A host with the ID number "531 – 1250" in light brown and a host with the ID number "498 – 1173" in dark brown, showing that the worms isolated from the two different hosts are scattered throughout the phylogenetic tree. The inner circle indicates the village from which the worm was isolated.



853

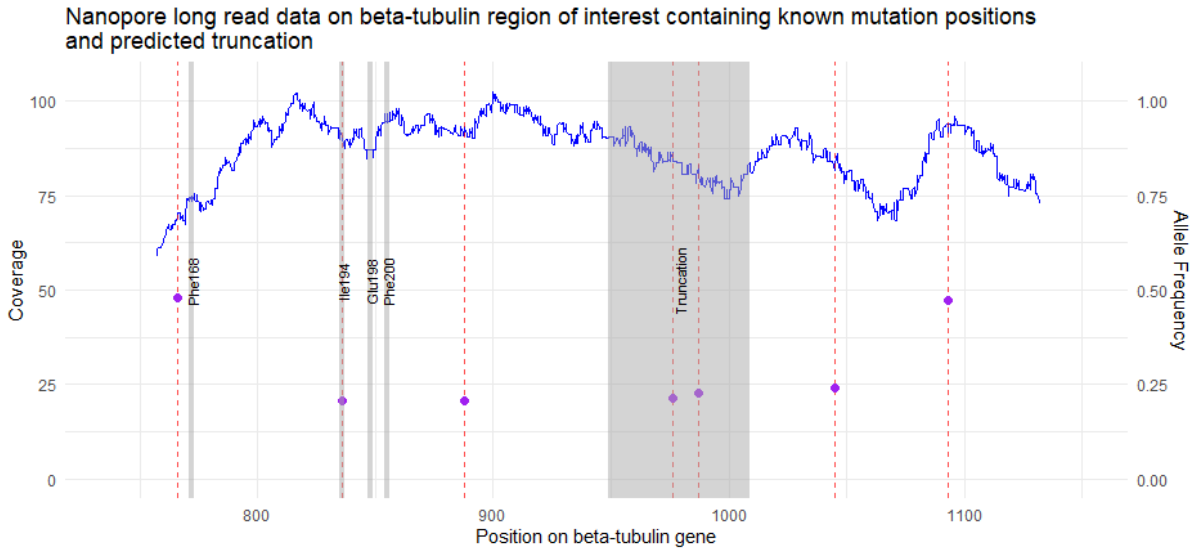
854 SI Figure 4: Tetraspanin, Serpin and Chymotrypsin-like serine protease from analyzing orthologous
855 groups in Venn diagram sections shared by *T. incognita*, *T. trichiura* (Serpin and Chymotrypsin-like
856 serine protease) and *T. muris* (Tetraspanin).

857



858

859 SI Figure 5: Beta tubulin genes found in *T. incognita* compared to *T. suis* and *T. trichiura*,
860 predicted from the braker2 pipeline. Black indicates positions carrying known mutations
861 (Phe168, Glu198, Phe200), observed variable position (Ile194). A truncation was predicted *in*
862 *silico* on 3 of the 7 genes. Gene IDs from top to bottom: g34, g460, g463, g472, g475, g529
863 and g7506.



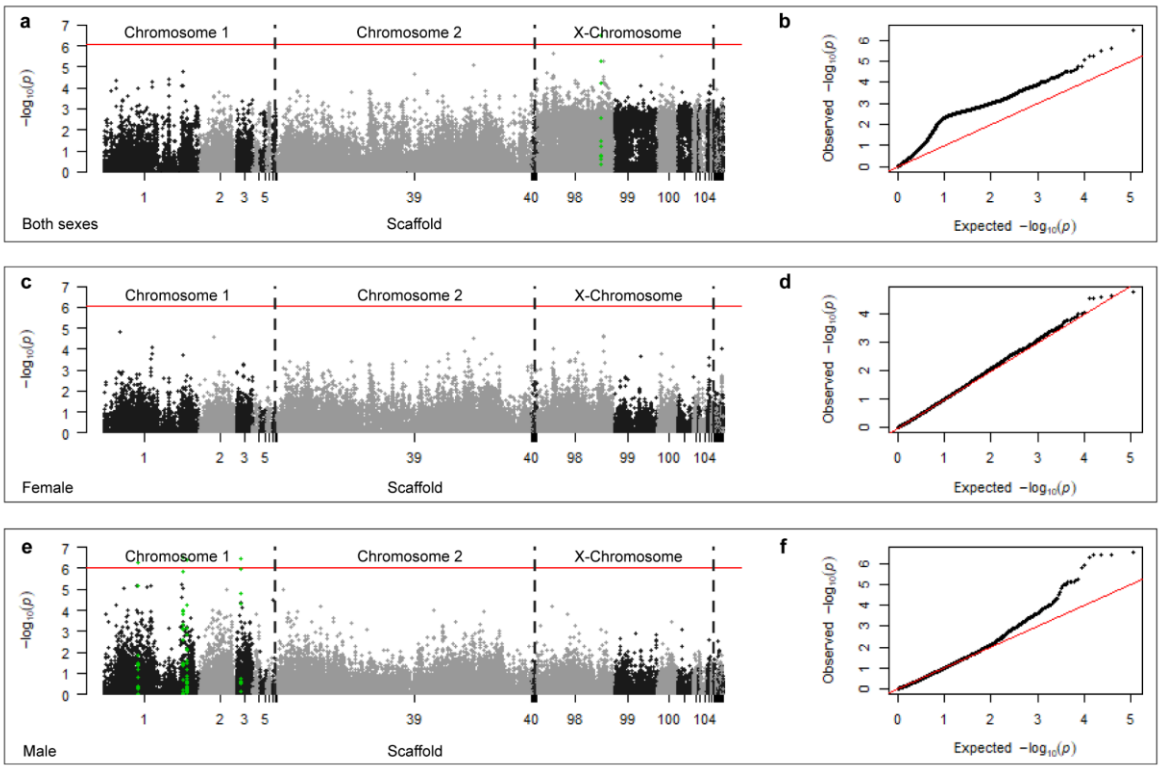
864

865 SI Figure 6: Coverage and variants from an independent PCR and Nanopore genome
866 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 (Ile194) and predicted truncation indicating that the truncation was not observed from the amplified genomic sequence.

Forward primer sequence: AAAGAGACCGGACATTTCGC.

Reversed primer sequence: TGAATTGCCTGGTTTCTAGAATG.



SI Figure 7: GWAS displaying p-values of missense SNPs associated to treatment outcome (T1 vs T1) using a linear model in PLINK. Variants were filtered at a minor allele frequency of 0.05, a missing genotype frequency of 0.05, and a deviation from the Hardy-Weinberg Equilibrium at a significance level of p=0.001. The red line indicates the Bonferroni corrected significance of 8.89E-7 at α=0.05 and 56273 mutations. The scaffolds are separated by black/grey colors. The green dots indicate all SNPs found on the specific gene of interest carrying a SNP with genome wide significance. **a** and **b** Data of both sexes. 190 males, 462 females, 16 ambiguous, 492 cases as in drug non-sensitive and 160 controls. **c** and **d** Only female SNPs. 0 males, 462 females, 328 cases as in drug non-sensitive and 134 controls. **e** and **f** Only male SNPs. 190 males, 0 females, 164 cases as in drug non-sensitive and 26 controls. Significant genes included g598, where no pfam annotation was found, g1428, which contains a Cadherin-like domain, g1478, which contains a fragile site-associated protein C-terminus domain and a Kiaa1109 N-terminal region, g2562, which contains a Pao retrotransposon peptidase domain and finally g11724, which contains a RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain). Gene sequences can be found in the Trichuris_Proteins.fasta file provided on github.

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. GO term enrichment analysis results

"GO-Term_enrichment_analysis.xlsx"

SI Table 4. Highly duplicated genes.

“Main_Phenotype_File_sex.txt”

SI Table 5. Worm ID’s with sex for PLINK input and phenotype of drug sensitive or non sensitive: Sex code ('1' = male, '2' = female, '0' = unknown), Phenotype value ('1' = control, '2' = case, '-9'/'0'/non-numeric = missing data if case/control)