Genetic makeup of *Trichuris hominibus*, a novel *Trichuris* species naturally infecting humans and displaying signs of resistance to drug treatment

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, Côte 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, and the emergence of anthelmintic resistance is predicted to occur within the next decade.[1,2] However, data that addresses these discrepancies and provides evidence for resistance does not exist. Here, we present the discovery of a novel *Trichuris* species, *Trichuris hominibus*, and we provide genomic evidence that supports the establishment of drug resistance in humans. Combining long- and short-read sequencing approaches, we assembled a high-quality reference genome of *Trichuris hominibus* to construct the species tree of the *Trichuris* genus and we sequenced a total of 752 individual worms which confirmed the phylogenetic placement and provided evidence supporting benzimidazole resistance. Gene duplication events across the species tree and a genome wide association study point towards β-tubulin variations as a source of albendazole resistance. 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. Efforts should support the development of species-level diagnostic tools, patient-centric treatment approaches in addition to mass-drug administration strategies, the definition of species-specific standard-of-care guidelines, and the incorporation of species diversity in vaccine development.

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*.[1-3] 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.[4] 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.[1,5] To circumvent resistance and improve efficacy against *Trichuris 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.[2]

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] The unmatched ability to evade host immunity with a plethora of immunomodulatory machinery has rendered *Trichuris* a focal point of immunotherapeutic research. 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*. They represent pivotal milestones, offering unprecedented opportunities for the discovery of new helminth-derived biomacromolecules. However, the present genomic data has not been able to explain the low efficacy of benzimidazoles against *T. trichiura*.

Here, we present the discovery of a novel *Trichuris* species, *Trichuris hominibus*, naturally infecting humans in Côte d'Ivoire and displaying 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 genetic makeup of this new species, leveraging whole-genome-sequencing (WGS) data to produce 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 all 1098 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 vs drug-resistant worms. We present genomic data that, for the first time in humans, points towards albendazole resistance through variations on the β-tubulin gene, a result of considerable health relevance.

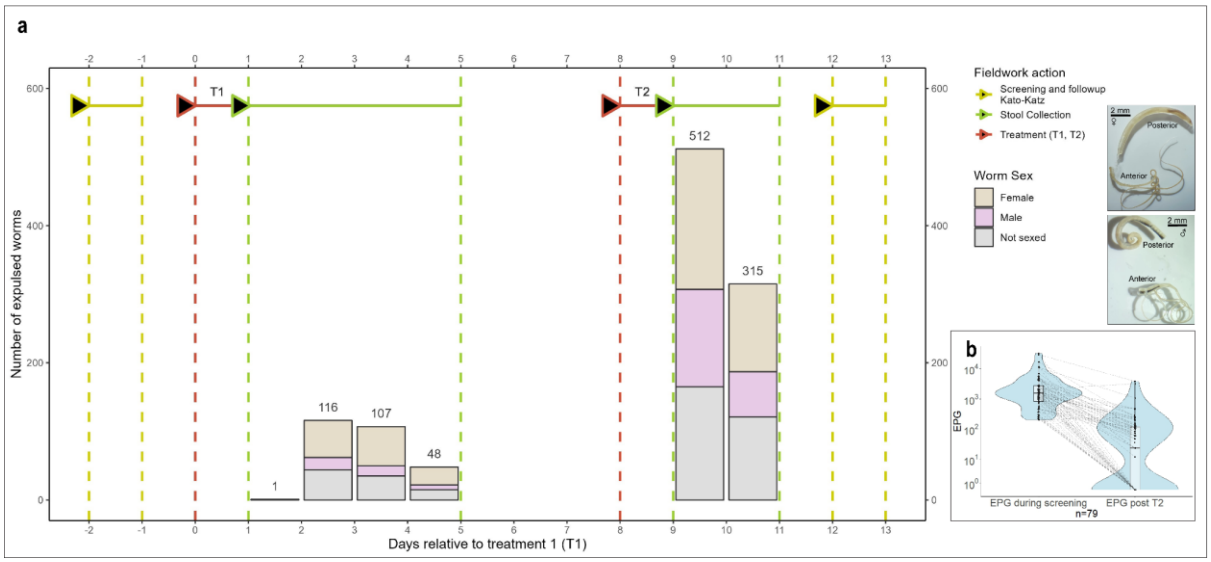
Results

Trichuris expulsion study reveals a new species of Whipworm, resistant 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 (alb-ive) combination treatment (T1) after which stool was collected daily and screened for worms for up to four days. Subsequently, 8 days after the first treatment, the same individuals received a single oral dose oxantel-pamoate treatment (T2) after which stool was collected for 2 consecutive days, as presented in figure 1a. We isolated 271 worms after treatment 1 (T1) with an expulsion peak on days 2 and 3 and 827 worms after treatment 2 (T2) with expulsion peaking on day 1 and 2 after treatment. A total of 1098 *Trichuris* worms were collected. Between one and 57 worms were isolated per host and 708 worms were sexed using bright-field microscopy showing a female to male ratio of 3.42:1 and 1.6:1 during T1 and T2 respectively. A significant correlation (R^2 = 0.34; P = 0.0061) was observed between the number of expulsed worms and the mean egg count per day with a slope of 9300 as shown in SI, figure 1, yielding a fecundity of 14307 eggs per female worm per day. 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. Morphologically, all worms and the observed eggs were indistinguishable and a picture of each worm is provided online.

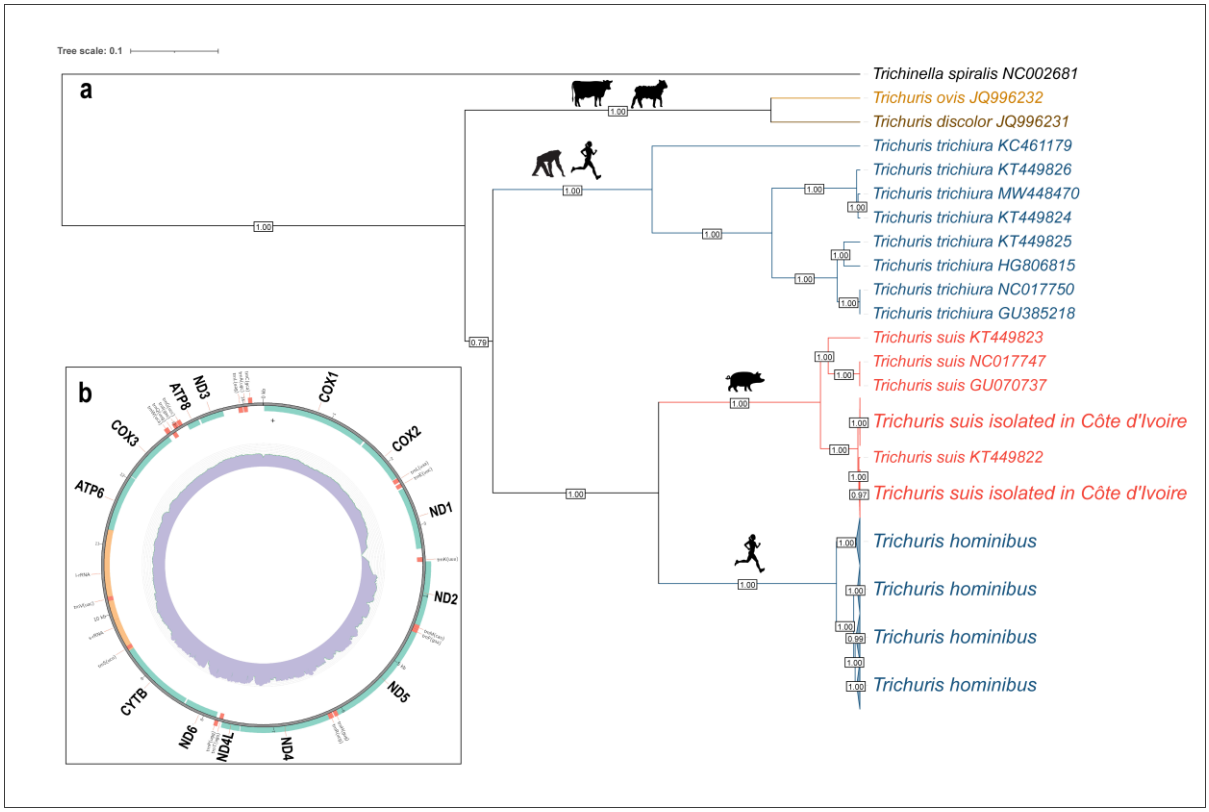
Fig. 1: Timeline of expulsion study including the number of expulsed worms and the female/male distribution of the sequenced worms



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-palmoate treatment, implemented in parts of South America, specifically against *T. trichiura* **b** EPG before treatment 1 compared to after the treatment 2.

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. 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, closer to *T. suis* than to *T. trichiura*, while the phylogenetic placement of the 16 sequenced *T. suis* fall within the expected clade of *T. suis* reference mitochondrial genomes. The phylogenetic tree presented in figure 2a showed the same topology when using nucleic acid sequences and the mt REV substitution model. We tentatively name the species in this newly formed clade *Trichuris hominibus*.

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 hominibus* 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 hominibus*.

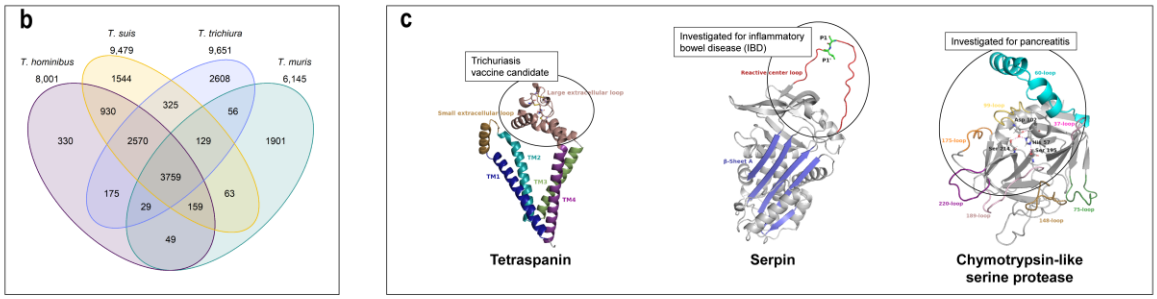
To investigate intra-species phylogeny, we selected the TN96 substitution matrix and added further granularity to the model by partitioning genes and codon positions separately. Using this more flexible model, some uncertainties in the intra-species tree topology could be resolved and 4 subclades were formed with a posterior distribution >0.95. The phylogenetic tree of all assembled *T. hominibus* mitogenomes together with collected metadata on worm sex, host sex, village, coinfection and treatment is presented in the SI, figure 2.

T. hominibus hybrid *de novo* whole genome assembly is mined for potential vaccine candidates and immunotherapeutic agents using comparative genomics

The *de novo* sequenced genome of *T. hominibus* at a coverage of 115X 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, to reduce methodological bias from gene prediction tools. This resulted in the prediction of 13739 genes for the newly discovered species and a total coding sequence length of 21.0 Mb compared to 17.1 Mb in *T. trichiura*. Contigs were clearly assignable to one of three linkage groups based homologous genes from *T. muris*.

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

a	*Trichuris hominibus*	*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*



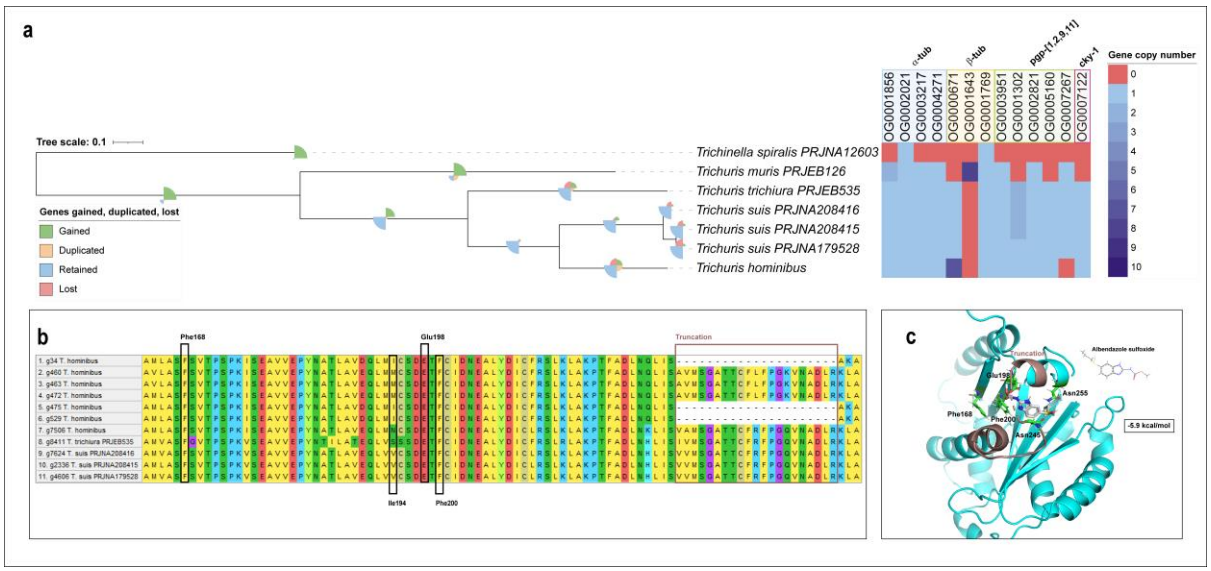
a Whole genome assembly metrics. *bacterial artificial chromosome (BAC) and expressed sequence tag (EST). **b** Venn diagram of orthologous groups from predicted transcripts. **c** Tetraspanin, Serpin and Chymotrypsin-like serine protease identified as genes of interest from analyzing orthologous groups in Venn diagram sections shared by *T. hominibus*, *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. hominibus*, *T. suis*, *T. trichiura* and *T. muris* as shown in figure 2b and present proteins of interest in figure 2c. 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*. Due to the genetic distance of *T. muris*, genes also shared by *T. muris* were included resulting in 204 orthologous groups. 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) and chymotrypsin-like serine proteases (OG0010319) that are shared only by *T. trichiura* and *T. hominibus*. The serpin transcript (encoded by g3902) motives, a reactive center loop which is cleaved at the P1-P1' position and a β-sheet A, which incorporates the cleaved reactive center loop residue upon proteolysis were assigned according to literature reference serpin structures.[17,18] The chymotrypsin-like serine protease (encoded by g3992), reported to degrade trypsin, is the main proteolytic digestive enzyme originating from the pancreas. Loops surrounding the catalytic triad were all present and assigned according to literature naming.[19] The 60-loop, shielding the catalytic pocket, was significantly bigger, compared to reported 60-loops. Finally, a family of tetraspanins (OG0009162) was identified.[20] The tetraspanin identified consists of 4 transmembrane helices (TM1 – TM4), a small extracellular loop (EC1 domain) and a long extracellular loop (EC2 domain) containing 3 disulfid bonds, rendering it a group 2 tetraspanin.[21] A list of all genes and the interproscan annotations are provided in the supporting information.

Gene duplication events across the species tree and a genome wide association study point towards β-tubulin variations as a source of albendazole resistance

We inferred the gene orthology of all genes predicted in the publicly available whole genomes of *Trichuris* species, *T. hominibus* and *Trichinella spiralis* to identify duplication events that might be connected to genes associated with resistance to either albendazole or ivermectin. These resistance-associated genes include β-tubulin 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 albendazole.[22,23] Amongst these genes, one orthologues group (OG0000671) of the β-tubulin gene showed 9 copies of beta tubulin gene in *T. hominibus* and only 1 copy in all *T. suis* and *T. trichiura* reference species as presented in figure 4a. Two genes g456 and g5082 were found to be fragmented. Intriguingly, the α-tubulin gene was only present in one copy. Upon further examination of the 7 complete beta tubulin genes, 3 are predicted to have a truncated phenotype as shown in figure 4b. Figure 4c, shows that the truncation occurs close to the predicted binding site of albendazole sulfoxide. Molecular docking simulations revealed a binding energy of -5.9 kcal/mol for the non-truncated variant and >22 kcal/mol for the truncated variant.

Fig. 4: Duplications across the species tree show a high copy number of β-tubulin genes that exist in a truncated and non-truncated variant.

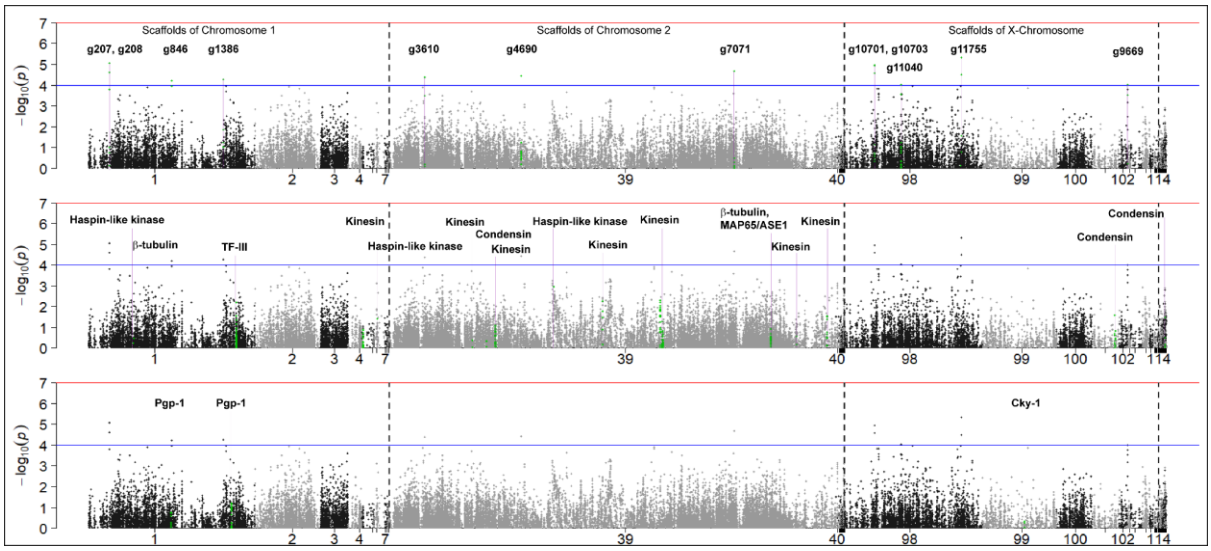


a Species tree indicating gene duplication, gains, losses and retention as pie charts across the tree. Heat map of copy numbers for all genes present in orthologues groups of genes, which have previously been associated to resistance and the α-tubulin gene. **b** Sequence alignment of 7 predicted genes in *T. hominibus* and the orthologous genes in *T. trichiura* and *T. suis*. **c** Autodoc Vina docking simulations of non-truncated β-tubulin variant showing a binding energy of -5.9 kcal/mol.

We then sought to investigate similarly expanded gene families as β-tubulin in *T. hominibus*. GO terms of massively expanded gene families as in ≥ 7 copies per gene in *T. hominibus* 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.

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



To investigate potential SNPs associated with resistance, a genome wide association study was conducted, comparing the populations isolated after the first treatment and second treatment respectively. 11 genes carried suggestive associations, defined as $p < 0.0001$ for the SNP and are listed in the supporting information. One of these genes, g208, stands out as a voltage gated chloride ion channel, carrying an Arg553Lys mutation with a significance value of $p = 2.5 \times 10^{-5}$. The significance of SNPs that were found on numerously expanded gene trees or on genes shown to be associated with either albendazole or ivermectin resistance were all below the suggestive threshold. The β-tubulin gene is of particular interest in this context, although it did not carry known mutations on the Phe168, Glu198 or Phe200 positions, which have been associated to albendazole resistance in veterinary medicine. The strongest association observed was an Asn194Ile mutation with a p-value of $p = 0.12$. As shown in figure 4b, position 194 is varying across species. The *T. trichiura* reference carries a polar Ser sidechain at the 194 position whereas the *T. suis* reference sequences all carry a hydrophobic Val side chain. *T. hominibus* carries hydrophilic (Asn) and hydrophobic (Ile, Met) sidechains on the duplicated beta tubulin genes, where the mutation from the hydrophilic to a non-polar sidechain correlates significantly with loss of drug-sensitivity.

Discussion:

Our investigation and the data provided mark a pioneering milestone in *Trichuris* research, unveiling the discovery of a previously unknown species - *Trichuris hominibus* - naturally infecting humans in Côte d'Ivoire and demonstrating low responsiveness to conventional drug treatments. We present genomic data, that imply β-tubulin variations in resistance mechanisms, raising concerns about the efficacy of current mass deworming strategies. These findings challenge the current understanding of species within the *Trichuris* genus and have profound implications for both parasitology and public health.

The whole-genome assembly forms the basis of our comprehensive genomic analysis and uncovers noteworthy insights into *Trichuris* biology, shedding light on potential mechanisms underlying drug resistance and offering promising avenues for therapeutic intervention. The whole-genome assembly metrics align with established *Trichuris* genomes, underscoring the

robustness of our dataset.[8,9,11] Our exploration of orthologous gene groups shared exclusively by *T. hominibus* 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, the identification of a serpin, that seems to be crucial for human infection of *Trichuris*, may serve as a promising vaccine candidate against human trichuriasis.[14] Finally, this family of proteins has gained particular interest in the context of inflammatory bowels disease (IBD) due to their immunomodulatory potential. Emerging evidence suggests that the serpinome of the gut microbiome plays a key role in IBD physiopathology. Considering the co-evolution of humans with whipworms for millenia, the identification of a serpin uniquely present in human infecting *Trichuris* may be of crucial importance in the pursuit of a treatment against IBD.[13]

A further intriguing gene family, identified only in human infecting *Trichuris* species, belong to the chymotrypsin-like serine proteases. They occur predominantly in the anterior part of the worm and two-thirds are likely to be encoded as secretory proteins. Although trypsin-like serine proteases are implicated in host-immunomodulatory functions, blood coagulation pathways or the degradation of host intestinal mucins, 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. Although tetraspanins are a highly conserved family of four-pass transmembrane proteins, their roles are highly diverse and occur in a variety of essential cellular functions. Knockout experiments have identified there necessity in the immune, hematopoietic, reproductive, genitourinary, epithelial, visual and auditory systems in mice. 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.

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.[27,28] 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.[29,30] 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. The existence of the gene in a truncated and non-truncated version that severely impact the binding energy of albendazole sulfoxide, may be connected to benzimidazole resistance. The occurrence of significant SNPs, as a result of the genome wide association study, further suggests a potential link between β-tubulin variants and prolonged benzimidazole exposure in Côte d'Ivoire. Notably, while β-tubulin exhibited significant duplication, α-tubulin did not, pointing to a targeted adaptive response to albendazole, assuming the same mode of action as in other nematodes.[31,32] Furthermore, the identification of massively expanded gene families associated with microtubule dynamics and mitotic processes provides additional evidence of adaptive evolution under drug pressure,

reinforcing the significance of our observations. 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.[33] MAP65/Ase1, a microtubule-associated protein is reported to promote microtubule flexibility and prevent microtubule severing by cross-linking.[34,35] 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 is connected to microtubule function with regards to the mitotic spindle.[36-38]

Our work builds on preliminary amplicon sequencing data (manuscript under preparation). 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). Results from the GWAS further imply a β-tubulin variation (Asn194Ile) in drug sensitivity. Although the SNP is not pointing towards the *in-silico* predicted active binding site, it is within 2 amino acids of known Glu198Val mutation conferring resistance.[23,39] Asparagine residue Asn255 has shown an increase in albendazole sulfoxide binding affinity *in-silico*. A mutation so close to the active sight from a polar asparagine to a hydrophobic isoleucine sidechain may have a profound impact on the kinetics of the binding reaction of albendazole sulfoxide. Finally, the observation of a mutation on a voltage gated chloride ion channel might be of importance regarding ivermectin drug resistance. A voltage gated ion channel clh-3 is implied in ivermectin adaptation through upregulation in *Haemonchus contortus*.[22]

The expulsion study allowed the collection of a population of *T. hominibus* worms to investigate the potential emergence of drug resistance in Côte d'Ivoire and gain insights into its fecundity and expulsion dynamics. Following the administration of albendazole and ivermectin during the first treatment (T1), 271 out of 1098 worms could be eliminated, which is in-line with previously observed low cure rates in Côte d'Ivoire.[2,40] The expulsion dynamics of T1 exhibited a faster rate compared to reported expulsion dynamics of a single-dose of mebendazole (500 mg) treatment against *T. trichiura*. 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. hominibus* matches literature reported values of approx. 16,700 for *T. trichiura*.[41] Overall, these results suggest that the biology of *T. hominibus* in terms of fecundity is comparable to *T. trichiura*.

Despite the close genetic relationship of *T. hominibus* to *T. suis*, our phylogenetic analysis demonstrates clear separation between *T. hominibus* and *T. suis* isolates from swine hosts in Côte d'Ivoire. While zoonosis events of trichuriasis have been reported, these findings indicate the absence of zoonotic transmission, in recent evolutionary history, between these populations.[42] Moreover, patterns of gene duplication, gain, and loss of function events in *T. hominibus* compared to within *T. suis* suggest a more ancient and divergent evolutionary lineage. The morphological indistinguishability of *T. hominibus*, *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 a novel *Trichuris* species, resistant 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.

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.*[2] 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 g 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 1/2 and 8 onwards, 24h-stool was collected and weighed for 2-3 days. Expelled worms were counted, washed twice with distilled water, once with sterile PBS and then stored either in pure ethanol at 4°C or mashed and stored in RNA-later at -20°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.[43] 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. 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.

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).[44] 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);[45] 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.[46] 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).[47] Coding genes were then extracted and aligned as amino acid and nucleic acid sequences using MAFFT (version 7.490).[48] 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).[49] Aligned concatenated nucleic acid sequences from all assembled and reference mitochondrial genomes were treated as homochromous. 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 were channeled into chopper (version 0.6.0) where reads below 5kb and low quality reads were filtered out and adapters trimmed.[50,51] 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.[52] 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).[53] The outputted contigs were channeled into RagTag (version 2.1.0) for correction and scaffolding using the *Trichuris trichiura* (PRJEB535) reference genome.[54] The corrected scaffolds were then assigned to one of three linkage groups from T. Muris using Liftoff (version 1.6.3),[55] as described in literature.[8] The genome was frozen for quality assessment using BUSCO (version 5.1.2 with Metazoa_obd10 database)[56] and QUAST (version 5.0.2).[57] 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).[58,59] 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.[60] 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. A table of all genes with matching orthologous groups and Gene IDs from worm bank is provided in the supporting information. Species tree with heat map was visualized using iTOL (version 6.8.1).[61] 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).[62,63] Massively expanded gene families were defined as ≥7 copies in *Trichuris hominibus* 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 hominibus*, *T. trichiura* and *T. muris*. A list of these transcripts with their functional annotation is provided in the supporting information. 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).

Molecular docking simulations

Molecular docking simulations were ran similar to literature procedures.[39] Briefly, truncated and non-truncated β-tubulin structures were predicted using alphafold (version 2.2.0). The simulation was ran on AutoDock Vina (version 1.1.2) with albendazole sulfoxide as a ligand, placed between Glu198, Phe200 and Asn255 allowing a 20 Å radius.

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] Plink (version 1.9)

was used to conduct the genome wide association study.[64] A principal component analysis (PCA) was conducted to account for population stratification. The PCA together with information on worm-sex was used in the linear regression option in PLINK. The R package qqman was used to visualize the results.

Data availability

Raw sequencing data is available under the European Nucleotide Archive (ENA) study accession “”. 551 assembled mitochondrial genomes and the de-novo assembled *Trichuris hominibus* 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.

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.

References:

1 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).

2 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).

3 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).

4 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).

5 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).

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 Kuzmin, E., Taylor, J. S. & Boone, C. Retention of duplicated genes in evolution. *Trends Genet* **38**, 59-72, doi:10.1016/j.tig.2021.06.016 (2022).
28 Lallemand, T., Leduc, M., Landes, C., Rizzon, C. & Lerat, E. An Overview of Duplicated Gene Detection Methods: Why the Duplication Mechanism Has to Be Accounted for in Their Choice. *Genes (Basel)* **11**, doi:10.3390/genes11091046 (2020).
29 Kondrashov, F. A. Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc Biol Sci* **279**, 5048-5057, doi:10.1098/rspb.2012.1108 (2012).

30 Zajac, N. *et al.* Gene Duplication and Gain in the Trematode Atriophallophorus winterbourni Contributes to Adaptation to Parasitism. *Genome Biol Evol* **13**, doi:10.1093/gbe/evab010 (2021).
31 Enos, A. & Coles, G. C. Effect of benzimidazole drugs on tubulin in benzimidazole resistant and susceptible strains of Caenorhabditis elegans. *Int J Parasitol* **20**, 161-167, doi:10.1016/0020-7519(90)90096-6 (1990).
32 Dawson, P. J., Gutteridge, W. E. & Gull, K. A comparison of the interaction of anthelmintic benzimidazoles with tubulin isolated from mammalian tissue and the parasitic nematode Ascaridia galli. *Biochem Pharmacol* **33**, 1069-1074, doi:10.1016/0006-2952(84)90515-x (1984).
33 Cushion, T. D., Leca, I. & Keays, D. A. MAPping tubulin mutations. *Front Cell Dev Biol* **11**, 1136699, doi:10.3389/fcell.2023.1136699 (2023).
34 Burkart, G. M. & Dixit, R. Microtubule bundling by MAP65-1 protects against severing by inhibiting the binding of katanin. *Mol Biol Cell* **30**, 1587-1597, doi:10.1091/mbc.E18-12-0776 (2019).
35 Portran, D. *et al.* MAP65/Ase1 promote microtubule flexibility. *Mol Biol Cell* **24**, 1964-1973, doi:10.1091/mbc.E13-03-0141 (2013).
36 Hagstrom, K. A. & Meyer, B. J. Condensin and cohesin: more than chromosome compactor and glue. *Nat Rev Genet* **4**, 520-534, doi:10.1038/nrg1110 (2003).
37 Stutzman, A. V., Liang, A. S., Beilinson, V. & Ikegami, K. Transcription-independent TFIIIC-bound sites cluster near heterochromatin boundaries within lamina-associated domains in C. elegans. *Epigenetics Chromatin* **13**, 1, doi:10.1186/s13072-019-0325-2 (2020).
38 Dai, J., Kateneva, A. V. & Higgins, J. M. Studies of haspin-depleted cells reveal that spindle-pole integrity in mitosis requires chromosome cohesion. *J Cell Sci* **122**, 4168-4176, doi:10.1242/jcs.054122 (2009).
39 Jones, B. P., van Vliet, A. H. M., LaCourse, E. J. & Betson, M. In Silico Docking of Nematode β-Tubulins With Benzimidazoles Points to Gene Expression and Orthologue Variation as Factors in Anthelmintic Resistance. *Frontiers in Tropical Diseases* **3**, doi:10.3389/fitd.2022.898814 (2022).
40 Sprecher, V. P., Coulibaly, J. T., Hurlimann, E., Hattendorf, J. & Keiser, J. Efficacy and Safety of Moxidectin-Albendazole and Ivermectin-Albendazole Combination Therapy Compared to Albendazole Monotherapy in Adolescents and Adults Infected with Trichuris trichiura: A Randomized, Controlled Superiority Trial. *Clin Infect Dis* **77**, 1294-1302, doi:10.1093/cid/ciad387 (2023).
41 Hansen, E. P. *et al.* Faecal egg counts and expulsion dynamics of the whipworm, Trichuris trichiura following self-infection. *J Helminthol* **90**, 298-302, doi:10.1017/S0022149X1500019X (2016).
42 Nejsum, P., Betson, M., Bendall, R. P., Thamsborg, S. M. & Stothard, J. R. Assessing the zoonotic potential of Ascaris suum and Trichuris suis: looking to the future from an analysis of the past. *J Helminthol* **86**, 148-155, doi:10.1017/S0022149X12000193 (2012).
43 Cringoli, G. *et al.* The Mini-FLOTAC technique for the diagnosis of helminth and protozoan infections in humans and animals. *Nature protocols* **12**, 1723-1732, doi:10.1038/nprot.2017.067 (2017).
44 Di Tommaso, P. *et al.* Nextflow enables reproducible computational workflows. *Nat Biotechnol* **35**, 316-319, doi:10.1038/nbt.3820 (2017).
45 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).
46 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).

47 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).
48 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).
49 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).
50 De Coster, W. & Rademakers, R. NanoPack2: population-scale evaluation of long-read sequencing data. *Bioinformatics* **39**, doi:10.1093/bioinformatics/btad311 (2023).
51 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).
52 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).
53 Zimin, A. V. *et al.* The MaSuRCA genome assembler. *Bioinformatics* **29**, 2669-2677, doi:10.1093/bioinformatics/btt476 (2013).
54 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).
55 Shumate, A. & Salzberg, S. L. Liftoff: accurate mapping of gene annotations. *Bioinformatics* **37**, 1639-1643, doi:10.1093/bioinformatics/btaa1016 (2021).
56 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).
57 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).
58 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).
59 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).
60 Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* **20**, 238, doi:10.1186/s13059-019-1832-y (2019).
61 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).
62 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).
63 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).
64 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).