Population genomics of modern and ancient *Trichuris trichiura*

Stephen R. Doyle\*+, Martin Jensen Søe\*, Peter Nejsum, Martin Sikora,Martha Betson, Phillip J. Cooper, Lifei Peng, Xing-Quan Zhu, Ana Sanchez, Gustavo Adolfo Fontecha Sandoval, Cristina Cutillas, Bruno Levecke, Matthew Berriman, Brian Lund Fredensborg, Eske Willerslev, Kurt H. Kjær, Christian Moliin Outzen Kapel

\*Co-first authors

+ Corresponding author

**Wellcome Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, United Kingdom**

Stephen R. Doyle: [stephen.doyle@sanger.ac.uk](mailto:stephen.doyle@sanger.ac.uk)

Matthew Berriman: [mb4@sanger.ac.uk](mailto:mb4@sanger.ac.uk)

**Department of Plant and Environmental Sciences, University of Copenhagen, 1871 Frederiksberg, Denmark**

Martin Jensen Søe: [martin.j.soe@gmail.com](mailto:martin.j.soe@gmail.com)

Christian Moliin Outzen Kapel: [chk@plen.ku.dk](mailto:chk@plen.ku.dk)

Brian Lund Fredensborg: [blf@plen.ku.dk](mailto:blf@plen.ku.dk)

**Department of Veterinary Disease Biology, University of Copenhagen, 1871 Frederiksberg, Denmark**

Peter Nejsum: [pn@clin.au.dk](mailto:pn@clin.au.dk)

**Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, 1350 Copenhagen K, Denmark**

Martin Sikora: [martin.sikora@sund.ku.dk](mailto:martin.sikora@sund.ku.dk)

Kurt H. Kjær: [kurtk@sund.ku.dk](mailto:kurtk@sund.ku.dk)

Eske Willerslev: [ewillerslev@sund.ku.dk](mailto:ewillerslev@sund.ku.dk)

**School of Veterinary Medicine, Faculty of Health and Medical Sciences, University of Surrey, Daphne Jackson Road, Guildford, GU2 7AL, United Kingdom**

Martha Betson: [m.betson@surrey.ac.uk](mailto:m.betson@surrey.ac.uk)

**Institute of Infection and Immunity, St George’s University of London, London, United Kingdom**

Phillip J. Cooper: [pcooper@sgul.ac.uk](mailto:pcooper@sgul.ac.uk)

**Department of Parasitology, School of Basic Medical Sciences, Guangdong Medical**

**University, Zhanjiang, Guangdong Province 524005, People’s Republic of China**

Lifei Peng: [lifpeng@hotmail.com](mailto:lifpeng@hotmail.com)

**College of Veterinary Medicine, Shanxi Agricultural University, Taigu, Shanxi**

**Province 030801, People’s Republic of China**

Xing-Quan Zhu: [xingquanzhu1@hotmail.com](mailto:xingquanzhu1@hotmail.com)

**Health Sciences, Brock University, St. Catharines, Ontario, L2S 3A1, Canada**

Ana Sanchez: [ana.sanchez@brocku.ca](mailto:ana.sanchez@brocku.ca)

**Microbiology Research Institute, Ciudad Universitaria, Universidad Nacional Autónoma de Honduras, Tegucigalpa, Honduras**

Gustavo Adolfo Fontecha Sandoval: [gustavo.fontecha@unah.edu.hn](mailto:gustavo.fontecha@unah.edu.hn)

**Departamento de Microbiología y Parasitología, Facultad de Farmacia, Universidad de Sevilla, Sevilla, Spain**

Cristina Cutillas: [cutillas@us.es](mailto:cutillas@us.es)

**Department of Virology, Parasitology and Immunology, Ghent University, Ghent, Belgium**

Bruno Levecke: [Bruno.Levecke@UGent.be](mailto:Bruno.Levecke@UGent.be)

# 

# Abstract

The neglected tropical disease Trichuriasis is caused by *Trichuris trichiura*, a soil-transmitted gastrointestinal nematode that infects as many as 700 million people worldwide. Using whole-genome sequencing of geographically distributed modern isolates collected from human and primate hosts, together with ancient samples preserved in archaeologically defined latrine and deposits representing the oldest eukaryotic pathogens sequenced to date, we present the first population genomics study of *T. trichiura*. We describe continent-scale genetic structure and high genetic relatedness between human and baboon infective parasites relative to other primate infective parasites. Admixture and population demographic analyses support a stepwise distribution of genetic variation which is highest in Uganda, consistent with an African origin and subsequent translocation with human migration. Finally, genome-wide scans of adaptation within human isolates and between human and animal isolates reveals local regions of genetic differentiation between geographically distinct populations. These genome-wide data provide insight into zoonotic reservoirs of human-infective *T. trichiura* and will support future efforts toward the implementation of genomic epidemiology of this globally important soil-transmitted helminth.

# 

# Introduction

The human-infective whipworm, *Trichuris trichiura*, is a soil-transmitted helminth responsible for Trichuriasis, a neglected tropical disease (NTD) estimated to affect as many as 700 million people worldwide. As a gastrointestinal parasite, an infection begins by the ingestion of embryonated eggs in contaminated food or soil; eggs migrate to the large intestine and hatch, after which emerging larvae burrow and establish an intracellular niche within intestinal epithelia where they develop to adult stages that can remain *in situ* for years. Although low infection burdens can remain asymptomatic, chronic infections cause a range of debilitating gastrointestinal symptoms and can lead to nutritional deficiencies and delays in physical and cognitive development, especially in children. As such, whipworm infections are primarily treated with benzimidazole (albendazole or mebendazole) anthelmintics; in endemic regions, treatment is targeted toward pre-school and school-aged children, women of reproductive age and high-risk working adults via annual or biannual mass drug administration (MDA) campaigns. In 2018, more than 676 million school-aged children, representing 53% treatment coverage of at-risk children, were treated with anthelmintics. It is almost certain that this number will only increase in the near future, as soil-transmitted helminths are targeted for elimination as a public health problem by 2030 in line with the WHO Sustainable Development Goals for NTDs [(World Health Organization, 2020)](https://paperpile.com/c/RvxIQI/guPt).

Humans have been parasitised by whipworms for millennia. Although now generally restricted to tropical and subtropical regions [(Pullan et al., 2014)](https://paperpile.com/c/RvxIQI/iY9i), *T. trichiura* was once a truly globally distributed worm; parasite eggs have been found in human coprolites (fossilised faeces) from archaeological sites dated up to 7,100 BC [(Araujo et al., 2008; Ledger et al., 2020, 2019)](https://paperpile.com/c/RvxIQI/2kEx+CPNz+yMOs), including sites in Europe and North America where infections are no longer prevalent [(Gonçalves et al. 2003; Søe et al. 2018; Graff et al. 2020; Ledger et al. 2021)](https://paperpile.com/c/RvxIQI/D7DU+Pu5P+K1LM+5aKL). However, whipworms are known to infect a broad range of mammals; over 70 species have been described within the genus *Trichuris*, and while generally host-specific, cross-host species transmission of individual parasite species have been reported, including between humans and non-human primates [(Betson et al., 2015; Cavallero et al., 2015; Ghai et al., 2014; Ravasi et al., 2012)](https://paperpile.com/c/RvxIQI/t7FC+TZrk+upiQ+n2Nc). Parasites that can infect human and non-human hosts represent a credible challenge to control campaigns, as non-humans hosts may act as a reservoir where parasites can evade treatment and can subsequently act as a source to reinfect treated populations. An understanding of the historical and modern dispersal of *T. trichiura* in human and non-human hosts is, therefore, a necessary component required to achieve elimination. This aim, to eliminate soil-transmitted helminths by MDA, is threatened by the emergence of resistance to the benzimidazole compounds used to control them, as has been observed in other parasitic nematodes (and in particular, veterinary helminths) frequently exposed to the drug [(Rose Vineer et al., 2020)](https://paperpile.com/c/RvxIQI/ogvt). Although evidence to suggest resistance is emerging in *T. trichiura* is limited, understanding population connectivity within and between endemic regions will inform the likelihood and rate by which resistance alleles might spread when they arise.

Here we describe the first population genomic analysis of *T. trichuris*. Using whole-genome sequencing data of modern isolates collected from human and primate hosts, together with ancient samples preserved in archaeological defined latrine and deposits [(Søe et al., 2018)](https://paperpile.com/c/RvxIQI/Pu5P), we describe broad and fine-scale genetic structure, admixture, and population demographics of geographically and genetically distinct populations, and explore genome-wide evidence of local adaptation within human isolates and between human and animal isolates. These genome-wide variant data will support future efforts toward the implementation of genomic epidemiology of this globally important soil-transmitted helminth.

# 

# Results

## Sequencing of modern and ancient *Trichuris trichiura*

We have generated whole-genome sequencing data from 56 modern and 17 ancient isolates (**Supplementary Table 1**), resulting in an average coverage of 9.45× and 0.66× of the nuclear genomes and 504× and 95× of the mitochondrial genomes, respectively (**Supplementary Table 2**). The samples analysed are derived from a broad geographic distribution, including populations in Africa, the Americas, China, and Europe (**Fig. 1a**). The modern isolates were composed of 49 parasites obtained from human hosts, as well as 7 samples obtained from captive animal hosts, including 2 parasites from baboons (*Papio hamadryas*), 2 from colobus monkeys (*Colobus guereza kikuyensis*) and 3 from leaf monkeys (*Trachypithecus francoisi*).

The ancient isolates were obtained from archaeological latrines and dig sites, primarily from Denmark as well as from the Netherlands and Lithuania, which have been dated to span the last 1000 years (**Supplementary Fig. 1**). These samples thus represent the oldest helminth samples, and likely the oldest eukaryotic pathogens [(Spyrou et al., 2019)](https://paperpile.com/c/RvxIQI/C4Pn), from which whole-genome sequencing data has been derived to date. Reads derived from ancient samples displayed increased sample damage due to deamination relative to the modern samples (**Supplementary Fig. 2**), particularly in the 5’ 2 bp of the reads, which was removed prior to downstream analyses.

Joint genotyping followed by stringent filtering identified a total of 1,647 mtDNA and 6,571,976 nuclear variants in the sample cohort. Considering the variation in depth of coverage and degree of variant missingness between samples, we subset the variant data into analyses containing either mitochondrial or nuclear datasets from samples with a minimum depth of coverage of 3× at 80% of the sites. Depth of coverage analyses comparing autosomal and sex-linked scaffolds revealed 25 male (XY; expected 0.5× depth of sex-linked scaffolds relative to autosomes) and 15 female (XX; expected 1:1 ratio of sex-to-autosomal depth) worms in the individual worm sequencing data; the datasets derived from pooled eggs (ancient, Cameroon, Ethiopia & Tanzania) revealed intermediate converge due to the presence of mixed-sex within the pools (**Supplementary Fig. 3**). Depending on the analysis, the colobus and leaf monkey samples were also excluded.

To investigate the broad-scale genetic diversity within the global cohort, we first performed a principal component analysis of mitochondrial variation (**Fig. 1b**). Three genetically defined clusters were identified, clearly separating China and America (Ecuador and Honduras) from a third cluster of mixed origin. Closer inspection of this third cluster identifies closely related but genetically distinguishable ancient, modern Ugandan and baboon isolates (**Fig. 1c).** To provide further granularity, we analysed nuclear genetic variation with a subset of higher quality samples, providing clearer differentiation between geographically distinct human isolates, as well as the ancient and baboon isolates (**Fig. 1d**). We also sampled additional populations from Africa (Cameroon, Tanzania, & Ethiopia), however, there was limited genetic information due to the very low sequencing coverage obtained; only unembryonated eggs rather than adult worms were sampled, and thus, DNA concentration was very low and suboptimal for sequencing [(Doyle et al., 2019)](https://paperpile.com/c/RvxIQI/LDEE). To address this, we reassessed genetic relatedness using identity-by-state (IBS) and IBS covariance analyses using ANGSD, which is more tolerant to low-coverage and sparse genetic data [(Korneliussen et al., 2014)](https://paperpile.com/c/RvxIQI/LGx7); here, some degree of genetic signal based on within-Africa clustering of these samples was identified (**Supplementary Fig. 4 & 5**). However, further sampling is clearly needed to precisely place isolates from these regions relative to the larger and better genetically defined cohort.

Whipworm isolates obtained from leaf and colobus monkeys were found to be highly genetically distinct from the human/baboon group and to each other (**Supplementary Fig. 6a**). To better understand their phylogenetic placement in the *Trichuris* genus, we assembled mitochondrial genomes for all modern worm isolates using an iterative baiting and mapping approach and compared them to publicly available whole mitochondrial genomes. Whipworms from leaf monkeys form a sister group to the human and non-human primate clades, whereas colobus-infective whipworms group more closely with pig whipworm *T. suis* isolates in a separate cluster **(Supplementary Fig. 6b)**, providing further support that these animal-derived *Trichuris spp*. are genetically distinct species from the human- and baboon-infective *T. trichiura*.

### 

### Figure 1. Global sampling distribution and broad-scale genetic relatedness of modern and ancient *Trichuris trichiura*

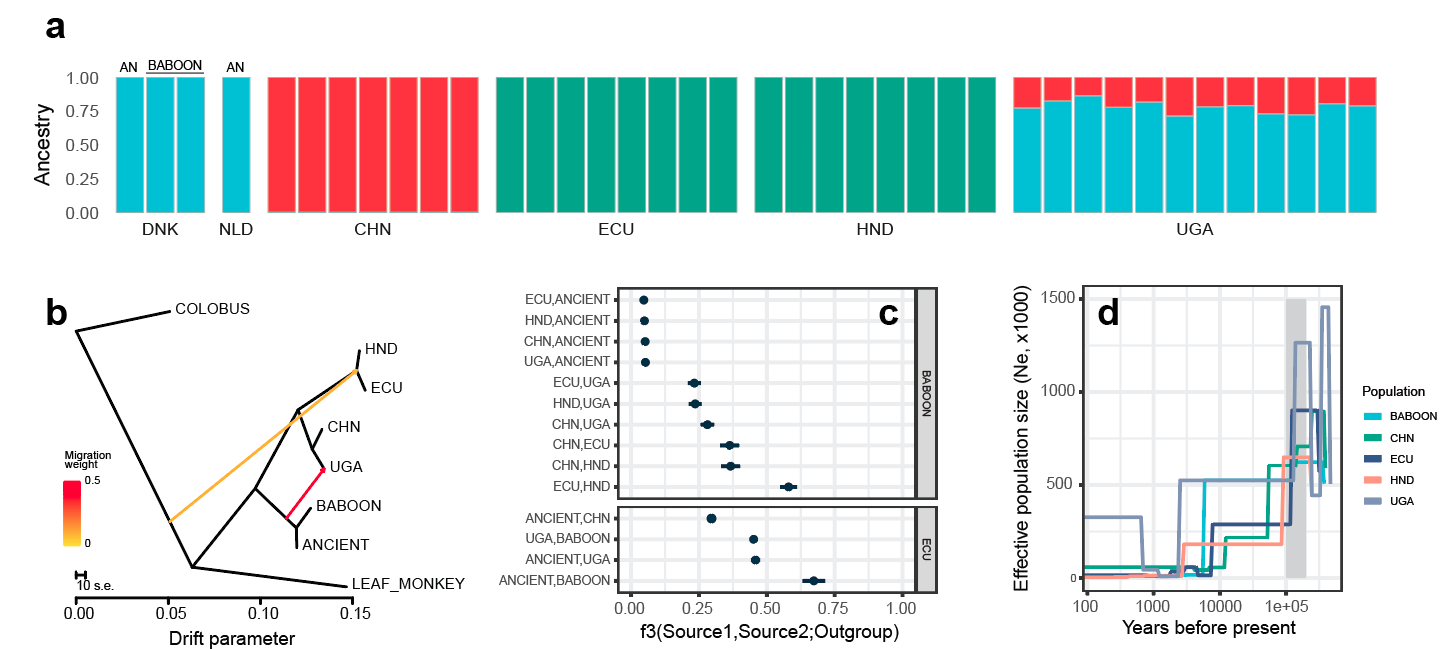
**a.** World map showing the approximate sampling locations of isolates used in the study, highlighting geographic regions of modern and ancient sampling, as well as host species. **b.** Principal component analysis of mitochondrial diversity (56 samples, 802 variants). **c.** Zoom in on a cluster of samples indicated by the dashed box in panel b, containing ancient, Ugandan, and baboon-derived samples. **d.** Principal component analysis of nuclear diversity using a subset of higher quality samples (39 samples, 2,544,110 variants).

## Patterns of admixture

To more formally describe the genetic relationships between populations, we first used NGSadmix to visualise the ancestral composition of samples. We estimated three ancestry components (K = 3) (**Fig. 2a**), largely consistent with the global population structure in the mitochondrial and nuclear PCAs (**Fig. 1b,c,d**); ancient and baboon clustered, which was distinct from each of China and American (Ecuador and Honduras) populations. Ugandan samples showed mixed ancestry of China and ancient/baboon samples. To explore this further, we determined admixture proportions across a range of K values (K = 2-10) (**Supplementary Fig. 7**). Subtle evidence of shared ancestry between China and the Americas was observed. Uganda samples displayed that most diverse ancestry profile, with complex mixtures of private components not found elsewhere. Similarly, Ecuador and Honduras also increased in diverse components throughout the range, however, these were largely shared. Interestingly, the ancient and baboon samples maintained a single fixed ancestry profile throughout the conditions tested.

To begin to quantify admixture, we used Treemix to estimate migration between nodes under a maximum likelihood framework (**Fig. 2b**). Two migration edges were supported (see **Supplementary Fig. 8** for trees and associated residual heatmaps foredges = 0-5), emphasising the strong genetic relationship between Ugandan, ancient and baboon isolates. Interestingly, there was evidence implicating low level admixture between the base of the tree leading to colobus isolates and Ecuador/Honduras isolates. While this doesn’t directly implicate hybridisation between colobus-infective and human-infective isolates specifically, hybridisation between the phylogenetically related but pig-infective *T. suis* and *T. trichiura* has been described previously in Ecuador [(Meekums et al., 2015)](https://paperpile.com/c/RvxIQI/zILC). We further quantified these admixture results using *f*-statistics on the nuclear variants (**Fig. 2c**). Using baboon samples as an outgroup, we demonstrate the close genetic relationship between geographically localised Ecuador and Honduras populations. However, we further identify the relationship between China and the American populations, and that this relationship is closer than between Ugandan and American populations. Investigations of mitochondrial markers of rrnL and nad1 from humans in Uganda, China and Ecuador have identified close similarity between samples from China and Ecuador, which are distinctly different to those from Uganda [(Hawash et al., 2016)](https://paperpile.com/c/RvxIQI/hTAi). Using baboon as an outgroup reduced our ability to measure more subtle differences between the closely related baboon, ancient, and Ugandan populations; to address this, we used Ecuador as an outgroup, which provided greater resolution to show the closer relationship between baboon and ancient isolates, relative to comparison with Ugandan isolates.

Collectively, these findings allow us to hypothesise about the potential dispersal of *T. trichiura*. First, population demographic analyses reveal that all populations show a substantial decline in effective population size between 100-200 kya (**Fig. 2d**); although we have made specific assumptions about important, but unmeasurable parameters of this model (specifically, mutation rate and generation time), these data are consistent with population bottlenecks as a result of early human migration out of Africa. The intermediate *f*3 statistics between Uganda and China suggests a stepwise migration pattern, first from Africa to Asia and then from Asia to the Americas, the latter of which would be consistent with human migration and crossing of the Beringia land bridge between the two continents. This migration was distinct to that which established in Europe, indicative by the lower degree of shared variation between ancient and Chinese isolates, but almost no sharing between ancient and the Americas. It has been argued that some intestinal parasites may not have survived the Beringia crossing due to unfavourable environmental conditions to sustain their lifecycle, and thus, genetic connectivity in parasite populations reflect a separate peopling of the Americas through trans-pacific or coastal migrations [(Araujo et al., 2008)](https://paperpile.com/c/RvxIQI/2kEx). To explore this here, we compared the proportion of variation shared between Ugandan, Chinese and American isolates. While a large proportion of variation was private (in total, 34.8% of variants are present only in one of the three populations) or shared by all three populations (25.8%), 5.25% of variants are shared between Ugandan and American populations that are absent from Chinese isolates (**Supplementary Fig. 9**); while tenuous due to the low sample numbers and sparse geographic sampling, such variation would support a low level but independent introduction of parasites to the Americas from Africa.



### 

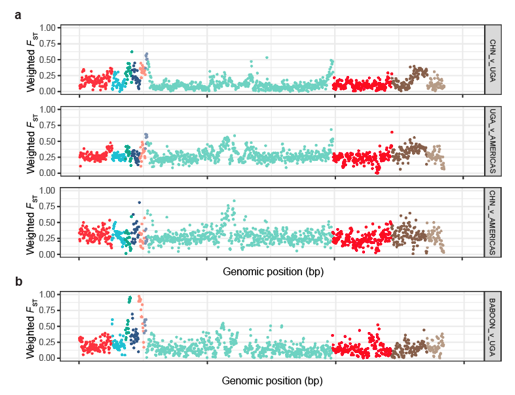
### Figure 2. Fine-scale genetic relationships and admixture between global populations

**a**. Admixture plot depicting population ancestry proportions determined using NGSadmix for K = 3, variants = 1,788,398. See Supplementary Fig. 7 for full analysis of K from 2 to 10. Ancient (AN) and baboon samples are highlighted. **b**. Treemix maximum likelihood tree of ancient and modern samples, including Colobus and Leaf monkey samples as outgroups, showing two migration edges. See Supplementary Fig. 8 for full analysis of migration edges from 0 to 5. **c.** Outgroup *f*3 statistics, comparing allele frequency correlations between two source populations (indicated on the left side of the panel), relative to an outgroup population (right side of panel). Positive *f*3 values imply shared or correlated genetic drift between Source 1 and Source 2 populations. **d.** Population demographic history of each population, determined using smc++, comparing effective population size (*N*e) over recent evolutionary history. The grey box highlights the time period between 100-200 kya, coinciding with human migration out of Africa.

## Genome-wide patterns of selection and adaptation

Distinct geographic clustering of Ugandan, Chinese, and the American populations prompted us to examine evidence of diversifying selection indicative of local adaptation within the human infective parasites. Mean nucleotide diversity varied between populations, with the highest diversity observed in Uganda, followed by China and baboon, and then the Americas, consistent with a global dispersal of *T. trichiura* originating in Africa (**Supplementary Fig. 10**). Tajima’s *D* was consistent, on average, between modern human populations (**Supplementary Fig. 11a**), however, local regions of positive and negative values were observed genome-wide **(Supplementary Fig. 11b)**. The baboon population presented with a bimodal distribution, as a result of high levels of positive Tajima’s *D* on the sex-linked scaffolds; this observation is likely explained by the fact that both worm samples were male (**Supplementary Fig. 3**; MN\_DNK\_COZ\_PH\_001 & MN\_DNK\_COZ\_PH\_002), and thus, lacked variation as a result of being hemizygous. A genome-wide analysis of divergence, calculated using a sliding window of *F*ST between pairs of populations, identified local regions of differentiation that may be indicative of local adaptation (**Fig. 3a**). For each pairwise comparison - UGA vs CHINA (Weir and Cockerham weighted *F*ST = 0.130), UGA vs Americas (*F*ST = 0.270), and China vs Americas (*F*ST = 0.275) - we identified 55, 63 and 57 genes in region of high divergence (defined as the top 1%), respectively. Analysis of gene ontology (GO) terms for each of the genesets (**Supplementary Tables 3, 4, and 5** for each pairwise comparison, respectively) did not, however, identify any enriched terms that would have suggested selection of multiple genes of common function. We extended this analysis to compare Ugandan and baboon isolates, given their close genetic relationship despite being isolated from different host species. Genome-wide mean *F*ST was intermediate (*F*ST = 0.177) to that of the human-specific analyses, re-emphasising that despite divergence, they are within the species-level of diversity expected of *T. trichiura*; however, genome-wide, novel regions of differentiation were observed, particularly on the sex-linked scaffolds (**Fig. 3b**). Regions of high divergence (top 1%) contained 62 genes (**Supplementary Table 6**), however, as in the previous comparison, no GO term enrichment was observed.

Finally, we explored variation in and around the β-tubulin gene. Large scale treatment of parasites using benzimidazole anthelmintics (as is used by MDA programmes against soil-transmitted helminths including *T. trichiura*) is known to select for genetic variation in β-tubulin conferring resistance to treatment. This is particularly evident in veterinary parasites [(Kwa et al., 1993; Sallé et al., 2019)](https://paperpile.com/c/RvxIQI/cBJJ+3SdM), where selection for benzimidazole resistance has rendered benzimidazole derivatives almost ineffective in some regions of the world [(Rose Vineer et al., 2020)](https://paperpile.com/c/RvxIQI/ogvt), and emerging evidence suggests that MDA may be exerting similar selection pressure on soil-transmitted parasites [(Diawara et al., 2009, 2013)](https://paperpile.com/c/RvxIQI/PMvA+0YBN). We identify 9 polymorphic sites in the sex-linked β-tubulin gene (Gene id: TTRE\_0000877201; Location: Trichuris\_trichiura\_1\_001: 10684531..10686350) segregating within modern human populations, however, no variation within the codon positions P167, P198, or P200 that are typically associated with benzimidazole resistance was observed (**Supplementary Fig. 12a**), consistent with previous studies [(Bennett et al., 2002; Hansen et al., 2013a, 2013b)](https://paperpile.com/c/RvxIQI/P6WJ+uLTq+UI0X). Further, little evidence of broader-scale genetic change on standing genetic variation in the region surrounding the β-tubulin gene that might be associated with positive selection on a gene within that region (**Supplementary Fig. 12b,c**). We do not have any data on their phenotypic response or exposure history to anthelmintic treatment, and thus, we cannot determine whether the absence of known benzimidazole-resistance associated variants in our data means that (i) all populations are phenotypically susceptible to anthelmintic treatment, or (ii) populations may or may not be phenotypically susceptible to benzimidazole treatment, but that selection on variation elsewhere in the genome, and not on variation in the β-tubulin gene, is responsible for drug response in *T. trichiura*. Future studies exploiting genome-wide analyses on parasite populations that are phenotypically well-defined for drug response [(Doyle and Cotton, 2019)](https://paperpile.com/c/RvxIQI/6upB) is needed to define the genetic architecture of drug response in *T. trichiura*. It will be particularly important to disentangle known variation in the tolerance to benzimidazole drugs by *T. trichiura* [*(Keiser and Utzinger, 2008; Patel et al., 2020)*](https://paperpile.com/c/RvxIQI/DP8X+mXf7) from the emergence of loss of efficacy due to high drug exposure [(Levecke et al., 2014)](https://paperpile.com/c/RvxIQI/GMqa), especially if the promise of the use of molecular diagnostics to monitor the emergence of anthelmintic resistance is to be realised [(Gandasegui et al., 2020; Kotze et al., 2020; Rashwan et al., 2017; Vlaminck et al., 2020)](https://paperpile.com/c/RvxIQI/c5H7+sF20+buSy+itre).



### Figure 3. Genome-wide comparison of genetic variation

**a.** Comparison human-infective *T. trichiura* from genetically and geographically defined populations, as per Fig. 1b & d. Pairwise *F*ST was measured in 50 kbp windows between China and Uganda (top), Uganda and Americas (Ecuador and Honduras)(middle), and China and Americas (bottom).

**b.** Comparison of closely related human-infective Ugandan and baboon-infective *T. trichiura*.

# Discussion

The whipworm *Trichuris trichiura* is one of the most highly prevalent and globally distributed helminth pathogens of humans. Here, we have undertaken a broad survey of genome-wide genetic diversity between modern and ancient isolates of *T. trichiura*. For the first time, we describe the genomic diversity of ancient parasites and their genetic connectivity to modern parasite populations, identify genetically distinct geographically defined populations, and evidence of differential proportions of admixture between them. Population demographic analyses support the hypothesis that global populations suffered severe population bottlenecks coinciding with human migration out of Africa, reinforcing the important role that human migration has played on the spread of modern helminths throughout the world [(Araujo et al., 2008; Choi et al., 2016; Hawash et al., 2016; Sallé et al., 2019; Small et al., 2019)](https://paperpile.com/c/RvxIQI/cBJJ+UIoP+2kEx+FD1M+hTAi).

Comparative analysis of human and animal infective isolates demonstrate that parasites isolated from baboons likely represent a zoonotic reservoir of human-infective parasites. The degree by which human and animal occupy similar ecological niches likely influence the likelihood of cross species transmission; baboon are large ground dwelling primates that can be found close to human activity, and therefore, likely share a common source of infection of soil-transmitted helminths such as *T. trichiura*. This is in contrast to smaller, tree dwelling primates including leaf and colobus monkeys that would have less interaction with humans. Genome-wide analyses between human and baboon parasites do, however, reveal discrete regions of the genome that differ between hosts, potentially indicative of some degree of host adaptation. Further sampling of human and a broader range of animal-infective parasites, particularly where cohabitation exists, combined with genome-wide analyses of selection should provide further insight into the mechanisms by which host-specificity is maintained and/or cross-species transmission is tolerated.

Our data form the genetic framework for genomic epidemiology of *T. trichiura*, a highly important soil-transmitted helminth that infects over 600 million people worldwide. It is almost certain that genomic surveillance will become an important tool for informing helminth control campaigns in the future; such data could be used to interpret variation between and changes within parasite populations, for example, population decline due to effective control or distinguishing between transmission, recrudescence, and/or loss of efficacy of the drugs used to control them. These require a comprehensive understanding of the underlying genetic diversity of the parasite throughout its range; our data provides a significant first step towards this goal, and will be enriched by further sampling throughout endemic regions to characterise finer-scale genetic connectivity and effective parasite transmission zones.

## Acknowledgements

We thank: the University of Copenhagen for funding to M.J.S. and all laboratory expenses through the KU2016 initiative, ‘The Genomic History of Denmark’; Mads XXX from Copenhagen Zoo and Jens F. Dahlerup from ... for ….

SRD is supported by a UKRI Future Leaders Fellowship (MR/T020733/1).

# 

## Author contributions

CMOK, PN, MJS, and MS designed the study. MJS extracted DNA from ancient isolates and prepared NGS libraries for sequencing for all isolates. PN extracted DNA from modern worm isolates. MBet, PJC, LP, XQZ, AS, GAFS, CC, BL oversaw collection of modern worm isolates. BLF, EW, KHK, and MBer provided expertise and advice throughout the study. CMOK, PN, MJS, and SRD analysed and interpreted results. SRD led and performed the bioinformatics analyses, drafted the manuscript and figures. All authors have read and approved the final version of the manuscript.

# 

# 

# Methods

## Sampling of ancient and modern *Trichuris spp.* from humans and animals

Sampling details specific to each group or population of parasites are outlined below. As the focus of the study is on the genetics of whipworms, no patient-specific data or identifiers were included in any analyses performed. In all cases where parasites were collected from human patients, informed written consent was obtained from all participants or their guardians (in cases where the participant was a child) after being informed about the study in both English and/or the local language. Samples were collected in line with ethical consideration and clearances as follows:

*Ancient: Trichuris* sp.eggs were isolated from ancient latrine samples in which a metagenomic study has previously identified DNA from *T. trichiura* [(Søe et al., 2018)](https://paperpile.com/c/RvxIQI/Pu5P). The eggs were isolated from ancient human latrine or deposit sites from Denmark (1000-1700 AD), the Netherlands (1350-1850 AD), Lithuania (1550-1580 AD) (**Supplementary Fig. 1**).

*China*: Worm samples, previously defined as genetically distinct from human-infective *T. trichiura* by mitochondrial genome and nuclear genetic markers [(Liu et al., 2013)](https://paperpile.com/c/RvxIQI/LIEv), were collected directly from the caecum of a human patient during a surgery in Zhanjiang People’s Hospital in Zhanjiang, Guangdong Province, China (CHN\_GUA). This study was approved by the Animal Ethics Committee of the Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences. Worms were also collected during necropsy from the caecum of a captive *Trachypithecus francoisi* (François Leaf-Monkey), which was humanely euthanized due to acute gastric dilation (People’s Republic of China) (CHN\_GUA [zoo]).

*Colobus*: Worms, previously defined as *Trichuris colobae n. Sp*. [(Cutillas et al., 2014)](https://paperpile.com/c/RvxIQI/rBUv), were collected from the caecum during necropsy of *Colobus guereza kikuyensis* (Mantled Guereza or Eastern Black-and-White Colobus) (ESP\_MAL [zoo]), which died of natural causes at the Fuengirola Zoo (Málaga, Spain).

*Ecuador*: Worms were collected as a joint effort by Phillip J. Cooper and Martha Betson (ECU\_QUI and ECU\_TEL). Approval for these studies was obtained from The Ethics and Welfare Committee at The Royal Veterinary College, University of London, UK (M2013 0005).

*Honduras*: Samples were collected as a joint effort by Ana Sanchez (Brock University, Canada) and Gustavo A. F. Sandoval (Universidad Nacional Autónoma de Honduras (UNAH), Honduras) (HND\_OLA and HND\_SAL). Samples collection protocols were reviewed and approved by the The Brock University Bioscience Research Ethics Board and Comité de Ética de Investigación – Maestría en Enfermedades Infecciosas Y Zoonóticas – Facultad de Ciencias – UNAH.

*Leaf monkey:* In the Zoo of XXXX, China a François Leaf monkey (*Trachypithecus francoisi*) was euthanized due to acute gastric dilation and Trichuris were collected from the caecum during post-mortem examination*.*

*Uganda*: Worms were recovered after anthelmintic treatment from the faeces of children in Uganda (UGA\_KAB) as previously described [(Olsen et al., 2009)](https://paperpile.com/c/RvxIQI/7wQJ). Permission was obtained from the Ministry of Health and the National Council of Science and Technology in Uganda. A subset of worms were recovered from a Danish patient infected with *T. trichiura* from Uganda (UGA\_DNK) as previously described [(Hansen et al., 2016)](https://paperpile.com/c/RvxIQI/nNRU). Approval for these studies was obtained from The Danish Central Medical Ethics Committee.

*Cameroon*: Stool samples were collected with focus on monitoring the efficacy of mebendazole for the treatment of soil-transmitted helminths [any previous publications?]. Permission was obtained from the National Ethics Committee in Yaoundé, Cameroon (Sep 2011).

*Tanzania:* Stool samples were collected[any previous publications?]. Permission was obtained from the Ministry of Health, Zanzibar Revolutionary Government, Tanzania (July 2012).

*Ethiopia:* Stool samples were collected [any previous publications?]. Permission was obtained from the Ethical Review Board at Jimma University (Sep 2011).

Individual worms were washed extensively in tab water or saline solution and preserved in 70% ethanol. When possible, the sex of individual worms were determined through microscopic examination prior to DNA extraction.

## DNA extraction and high-throughput sequencing

Ancient samples were processed for egg extraction in a dedicated paleoparasitological laboratory at the Department of Plant and Environmental Sciences (PLEN), University of Copenhagen. Helminth eggs were concentrated by selecting particles based on flotation in high density liquid sedimentation and size filtration. The egg isolate samples were then processed for DNA extraction in dedicated aDNA laboratories at the Centre for Geogenetics (CGG), University of Copenhagen, in accordance with strict aDNA-specific requirements. DNA extraction was performed using the PowerLyzer PowerSoil DNA isolation kit (MO BIO Laboratories, Carlsbad, California) with minor modifications. The complete protocol has been described previously [(Søe et al., 2018)](https://paperpile.com/c/RvxIQI/Pu5P).

DNA from modern worm isolates was prepared as follows: (i) for samples collected from stool following treatment with anthelmintics, directly from a human patient via colonoscopy, and from baboon, DNA was extracted from whole adult worms using the MasterPureTM DNA Purification kit (Epicentre Biotechnologies) at the Department of Veterinary Disease Biology (VDB), UCPH following the manufacturer’s protocol with the following exception: worms were first homogenized in 300 µL lysis solution using a matching disposable plastic pestle and then incubated at 56°C for three hours; (ii) for the colobus samples, DNA extraction from whole adult worms was performed at the Departmento de Microbiología y Parasitología, Universidad de Sevilla, Sevilla, Spain, using a DNeasy Blood & Tissue Kit (Qiagen) according to the manufacturer’s protocol; (iii) for the leaf monkey samples, DNA extraction was performed at the Department of Parasitology, Lanzhou Veterinary Research Institute, Lanzhou, China as previously described [(Liu et al., 2013)](https://paperpile.com/c/RvxIQI/LIEv). DNA (400 ng) was fragmented to approximately 300-700 bp using a BioRuptor (Diagenode) using 2 or 4 cycles of 15 sec ON / 90 sec OFF at instrument setting HIGH. Fragmented DNA was concentrated using a MinElute kit (Qiagen), after which DNA concentration was determined (Qubit 2.0 dsDNA HS kit) and library size confirmed (Agilent Bioanalyzer High Sensitivity DNA kit).

DNA sequencing libraries were prepared for all sample types, modern and ancient DNA extracts, using the NEBNext DNA Sample Prep Master Mix Set for 454 (E6070) kit with a modified protocol previously described for ancient samples [(Søe et al., 2018)](https://paperpile.com/c/RvxIQI/Pu5P). DNA libraries were amplified in a single PCR reaction of 50 µL using: half of the prepared DNA library (12.5 µL), 5 U Taq Gold (Life Technologies), 1x buffer Gold, 2 mM MgCl2, 0.25 mM dXTP, 0.2 mM of primers (PE1.0 and Illumina multiplex primer), for 8-16 cycles depending on library strength. Libraries were sequenced using 100 bp single-end (ancient isolates) or 100 bp paired-end (modern isolates) chemistry on a HiSeq 2000/2500 platform at The Danish National High-Throughput DNA Sequencing Centre (now Centre for GeoGenetics Sequencing Core).

## Raw read processing and mapping to the reference genome

Raw reads were first processed using AdapterRemoval2 [(Schubert et al., 2016)](https://paperpile.com/c/RvxIQI/xASA); for both SE and PE reads, adapters were removed and N bases trimmed, and for PE reads, R1 and R2 reads were collapsed where possible. Where multiple lanes of data were generated, trimmed read sets were merged prior to mapping.

Mapping was performed using BWA mem [(Li, 2013)](https://paperpile.com/c/RvxIQI/tavP) to an unpublished but significantly improved reference genome of *T. trichiura* (available here: ftp://[ngs.sanger.ac.uk/production/pathogens/parasites/Trichuris/trichuria/](http://ngs.sanger.ac.uk/production/pathogens/parasites/Trichuris/trichuria/)). Originally described by Foth et al. [(2014)](https://paperpile.com/c/RvxIQI/oI8m/?noauthor=1), the new assembly used here is larger (80.57 Mb vs 75.49 Mb), more contiguous (N50 = 11.3 Mb [vs 0.07 Mb] & N50n = 2 [vs 265]), and in fewer scaffolds overall (n = 113 vs 4156) relative to the published version. For the ancient samples, trimmed SE reads were mapped, whereas for modern samples, trimmed PE and SE reads (merged PE and trimmed SE) were mapped independently and subsequently merged. For all samples, duplicate reads were identified using picard MarkDuplicates (<http://broadinstitute.github.io/picard/>).

Putative deamination damage (represented by excessive C-to-T and G-to-A substitutions in the ends of reads) was assessed using Pmdtools (<https://github.com/pontussk/PMDtools>) [(Skoglund et al., 2014)](https://paperpile.com/c/RvxIQI/4uzR). This analysis revealed bias in the terminal 2 bp, particularly in the ancient samples; to account for this, the mapped reads were trimmed by 2 bp for all samples (ancient and modern) using bamUtils trimBam [(Jun et al., 2015)](https://paperpile.com/c/RvxIQI/eOru). Genome coverage was determined using samtools bedcov per scaffold and in 100 kbp windows. To explore putative causes of mapping variation, we also ran Kraken2 [(Wood et al., 2019)](https://paperpile.com/c/RvxIQI/6Vt4) to estimate the degree of contamination in the raw sequencing reads. This analysis showed that while each sample contained a small degree of “contamination” evidenced by hits to the kraken database (minikraken2\_v1\_8GB), it did not explain overall variation in mapping. Quality control and quantitative comparison between samples was undertaken at each stage of the pipeline and visualised using Multiqc [(Ewels et al., 2016)](https://paperpile.com/c/RvxIQI/mlKY).

## 

## 

## Variant calling

Variant calling was performed with GATK Haplotype Caller, first by generating per sample GVCF files (--min-base-quality-score 20 --minimum-mapping-quality 30 --standard-min-confidence-threshold-for-calling 30), followed by joint genotyping of the sample cohort. Heterozygosity was adjusted from default settings based on estimates from raw reads using GenomeScope [(Vurture et al., 2017)](https://paperpile.com/c/RvxIQI/DY8F). To improve efficiency, each task was split by sample and by scaffold and run in parallel, before merging to produce the final call sets.

The raw output was split into mitochondrial and nuclear variants, and then split again into single nucleotide variants (SNV) and indels, and filtered independently. A “hard filtering” approach was employed based on removing relevant tails of variant distributions (typically the upper and/or lower 1%) on the following quality metrics: QUAL, DP, MQ, SOR, FS, QD, MQRankSum, and ReadPosRankSum. Variants were further filtered to ensure they met the following criteria: minimum and maximum alleles = 2; minor allele > 0.02; Hardy Weinberg Equilibrium = >1E-6 (nuclear variants); per sample missingness > 0.8; per genotype depth > 3. Finally, variants were removed if they were found in regions of the genome where the reference genome guided mapping of kmers (k = 35) was poor or not unique determined using SNPable (<http://lh3lh3.users.sourceforge.net/snpable.shtml>), retaining regions of the genome where overlapping 35-mers mapped uniquely and without 1-mismatch (75.02% of the genome kept). In summary, 6,571,976 variants passed these filters.

## 

## Population structure analyses

Broad-scale genetic relatedness between samples and populations was explored by principal component analysis using the R package SNPrelate [(Zheng et al., 2012)](https://paperpile.com/c/RvxIQI/LjPY). This was performed on both mitochondrial and nuclear variants separately, and with and without the colobus and leaf monkey samples, which were found as outliers in all analyses.

To further examine the relationships between samples we generated identify-by-state (IBS) and IBS covariance matrices using ANGSD [(Korneliussen et al., 2014)](https://paperpile.com/c/RvxIQI/LGx7) (parameters: -minMapQ 30 -minQ 20 -GL 2 -doMajorMinor 1 -doMaf 1 -SNP\_pval 2e-6 -doIBS 1 -doCounts 1 -doCov 1 -makeMatrix 1 -minMaf 0.05). This approach took mapped reads in a bam file directly as input (as opposed to variants in a VCF), and calculated genotype likelihoods per variant site which are more sensitive to poorer quality and/or low coverage data.

## 

## Phylogenetics of reconstructed and publicly available mitochondrial genomes

Preprocessed sequence data from modern isolates were mapped to published mitochondrial genomes of *T. trichiura* (accession numbers: KT449826, GU385218 and HG806815 [an extract from the T. trichiura whole-genome shotgun sequence]), *T. suis* (KT449822, KT449823, GU070737), *T. discolor* (JQ996231), *T. ovis* (JQ996232) and that of the Leaf-Monkey, Trichuris sp. (KC461179) reporting only the best hit. The published genome with the most hits was then used as a reference in guided assemblies using the mitochondrial baiting and iterative mapping approach, MITObim (v1.8) [(Hahn et al., 2013)](https://paperpile.com/c/RvxIQI/4rW6). Subsequent manual curation of the sequence was performed to remove redundant, overlapping sequence at the ends of the mitochondrial contigs, and the starting position adjusted to the first codon of the COXI gene. For the ancient samples, consensus mitochondrial genome sequences have previously been called for the 10 samples (of 13 in total) that provided 100% bases with coverage to the KT449826 *T. trichiura* reference sequence [(Søe et al., 2018)](https://paperpile.com/c/RvxIQI/Pu5P). Consensus sequences were annotated using the MITOS Webserver [(Bernt et al., 2013)](https://paperpile.com/c/RvxIQI/Xq0J). Mitochondrial genomes were aligned with all *Trichuris sp*. reference genomes publicly available at NCBI using mafft v7 [(Katoh and Standley, 2013)](https://paperpile.com/c/RvxIQI/dbPD), with the global-pair setting at 1,000 iterations. A Neighbor Joining tree was built using the Jukes-Cantor nucleotide distance model and 1,000 bootstraps in CLC Sequence Viewer 7 (Qiagen), which was also used for visualisation.

## 

## Admixture and population demographic analyses

Admixture was determined using NGSadmix [(Skotte et al., 2013)](https://paperpile.com/c/RvxIQI/aW2k). Briefly, genotype likelihoods were extracted from using vcftools, after which NGSadmix (-P 4 -minMaf 0.05 -misTol 0.9) was run over a range of K values from 2 to 10. The optimal value of K was determined by iteratively running NGSadmix as above 5 times, changing the seed value on each run. The log likelihood of each run (all iterations of K = 2-10, s = 1-5) were used to determine the optimal value of K using Evanno’s method [(Evanno et al., 2005)](https://paperpile.com/c/RvxIQI/aC09) on the Clumpak webserver [(Kopelman et al., 2015)](https://paperpile.com/c/RvxIQI/T2Vk).

Treemix [(Pickrell and Pritchard, 2012)](https://paperpile.com/c/RvxIQI/desq) was performed on the nuclear dataset, including leaf and colobus monkey samples as outliers. Variants were first filtered using vcftools (--max-missing 1), followed by further pruning to minimise variants in linkage disequilibrium. Customs scripts (ldPruning.sh & vcf2treemix.sh) were modified from <https://github.com/speciationgenomics/scripts>. The optimal number of migration edges was estimated using the R package OptM (<https://cran.r-project.org/web/packages/OptM/> ).

*f*-statistics were calculated using ADMIXTOOLS [(Patterson et al., 2012)](https://paperpile.com/c/RvxIQI/MgTI) using the qp3Pop tool to calculate *f*3 data. All combinations of source 1 and source 2 populations were determined, using either baboon or Ecuador samples as the outgroup. Customs scripts (convertVCFtoEigenstrat.sh) were modified from [https://github.com/](https://github.com/speciationgenomics/scripts)[joanam](https://github.com/joanam)/[scripts](https://github.com/joanam/scripts).

Population demographics were determined using smc++ [(Terhorst et al., 2017)](https://paperpile.com/c/RvxIQI/fcPO). For each population, variant sites present in all individuals were extracted using vcftools (--max-missing 1), after which smc++ vcf2smc was run per scaffold. Estimated population sizes were fit to the data for all scaffolds using smc++ estimate, which used the *Caenorhabditis elegans* mutation rate (2.9e-9 mutations per site per generation)[(Denver et al., 2009)](https://paperpile.com/c/RvxIQI/w1hZ) as a proxy for *T. trichuris* mutation rate, which is currently unknown. Finally, effective population size per generation was scaled based on an estimate of 3 generations per year. As the ancient samples were a pooled population of eggs rather than individual parasites, they violated assumptions of the model and were, therefore, removed from the analysis.

## 

## Genome-wide genetic diversity analyses

Genome-wide nucleotide diversity, Tajima’s *D*, and pairwise *F*ST was determined using vcftools in 50 kbp non-overlapping sliding windows. To improve the visualisation of the genome-wide comparisons, analyses were restricted to scaffolds in chromosomal linkage groups, which represented 89.83% of the total genome assembly. For analyses in which “Americas” were defined as a single population, samples from Ecuador and Honduras were combined.

The annotation of the unpublished genome assembly used in this analysis has not yet been completed. To enable the identification of genes within outlier regions of the genome-wide analyses, we performed a liftover of existing gene models and gene identifiers from the published version of the *T. trichiura* assembly currently available in WormBase Parasite [(Howe et al., 2017)](https://paperpile.com/c/RvxIQI/NEPa) (<https://parasite.wormbase.org/Trichuris_trichiura_prjeb535/Info/Index/>, Version: [WBPS15](https://parasite.wormbase.org/info/about/release-log.html)) using liftoff [(Shumate and Salzberg, 2020)](https://paperpile.com/c/RvxIQI/N6gn). By retaining gene identifiers from the original assembly, cross validation of gene hits within WormBase Parasite could be performed. In total, 8451 of 9650 gene features (~88%) were transferred; of the genes that did not transfer, ~80% were classified as contamination based on hits to the uniprot\_reference\_proteome database using DIAMOND [(Buchfink et al., 2015)](https://paperpile.com/c/RvxIQI/PYmh). We performed gene ontology term analysis using gProfiler [(Raudvere et al., 2019; Shumate and Salzberg, 2020)](https://paperpile.com/c/RvxIQI/N6gn+BdOS), applying a g:SCS multiple testing correction method and a significance threshold of 0.05, restricting the complete geneset to only genes that had successfully been lifted from the original annotation.

Analysis of variation in β-tubulin was performed using vcftools (--site-pi --maf 0.01), using a bed file of exon coordinates.

## 

## Data availability

Raw sequencing data is available XXX. The

*Trichura trichiura* genome assembly is available at ftp://[ngs.sanger.ac.uk/production/pathogens/parasites/Trichuris/trichuria/](http://ngs.sanger.ac.uk/production/pathogens/parasites/Trichuris/trichuria/) .

## 

## Code availability

Custom code to analyse data and reproduce the figures presented is available at <https://stephenrdoyle.github.io/ancient_trichuris/> and Zenodo.

# 

# References

[Araujo, A., Reinhard, K.J., Ferreira, L.F., and Gardner, S.L. (2008). Parasites as probes for prehistoric human migrations? Trends Parasitol. *24*, 112–115.](http://paperpile.com/b/RvxIQI/2kEx)

[Bennett, A.B., Anderson, T.J.C., Barker, G.C., Michael, E., and Bundy, D.A.P. (2002). Sequence variation in the Trichuris trichiura beta-tubulin locus: implications for the development of benzimidazole resistance. Int. J. Parasitol. *32*, 1519–1528.](http://paperpile.com/b/RvxIQI/UI0X)

[Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M., and Stadler, P.F. (2013). MITOS: improved de novo metazoan mitochondrial genome annotation. Mol. Phylogenet. Evol. *69*, 313–319.](http://paperpile.com/b/RvxIQI/Xq0J)

[Betson, M., Søe, M.J., and Nejsum, P. (2015). Human Trichuriasis: Whipworm Genetics, Phylogeny, Transmission and Future Research Directions. Current Tropical Medicine Reports *2*, 209–217.](http://paperpile.com/b/RvxIQI/n2Nc)

[Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. Nat. Methods *12*, 59–60.](http://paperpile.com/b/RvxIQI/PYmh)

[Cavallero, S., De Liberato, C., Friedrich, K.G., Di Cave, D., Masella, V., D’Amelio, S., and Berrilli, F. (2015). Genetic heterogeneity and phylogeny of Trichuris spp. from captive non-human primates based on ribosomal DNA sequence data. Infect. Genet. Evol. *34*, 450–456.](http://paperpile.com/b/RvxIQI/t7FC)

[Choi, Y.-J., Tyagi, R., McNulty, S.N., Rosa, B.A., Ozersky, P., Martin, J., Hallsworth-Pepin, K., Unnasch, T.R., Norice, C.T., Nutman, T.B., et al. (2016). Genomic diversity in Onchocerca volvulus and its Wolbachia endosymbiont. Nat Microbiol *2*, 16207.](http://paperpile.com/b/RvxIQI/UIoP)

[Cutillas, C., de Rojas, M., Zurita, A., Oliveros, R., and Callejón, R. (2014). Trichuris colobae n. sp. (Nematoda: Trichuridae), a new species of Trichuris from Colobus guereza kikuyensis. Parasitol. Res. *113*, 2725–2732.](http://paperpile.com/b/RvxIQI/rBUv)

[Denver, D.R., Dolan, P.C., Wilhelm, L.J., Sung, W., Lucas-Lledó, J.I., Howe, D.K., Lewis, S.C., Okamoto, K., Thomas, W.K., Lynch, M., et al. (2009). A genome-wide view of Caenorhabditis elegans base-substitution mutation processes. Proc. Natl. Acad. Sci. U. S. A. *106*, 16310–16314.](http://paperpile.com/b/RvxIQI/w1hZ)

[Diawara, A., Drake, L.J., Suswillo, R.R., Kihara, J., Bundy, D.A.P., Scott, M.E., Halpenny, C., Russell Stothard, J., and Prichard, R.K. (2009). Assays to Detect β-Tubulin Codon 200 Polymorphism in Trichuris trichiura and Ascaris lumbricoides. PLoS Neglected Tropical Diseases *3*, e397.](http://paperpile.com/b/RvxIQI/0YBN)

[Diawara, A., Halpenny, C.M., Churcher, T.S., Mwandawiro, C., Kihara, J., Kaplan, R.M., Streit, T.G., Idaghdour, Y., Scott, M.E., Basáñez, M.-G., et al. (2013). Association between response to albendazole treatment and β-tubulin genotype frequencies in soil-transmitted helminths. PLoS Negl. Trop. Dis. *7*, e2247.](http://paperpile.com/b/RvxIQI/PMvA)

[Doyle, S.R., and Cotton, J.A. (2019). Genome-wide Approaches to Investigate Anthelmintic Resistance. Trends Parasitol. *35*, 289–301.](http://paperpile.com/b/RvxIQI/6upB)

[Doyle, S.R., Sankaranarayanan, G., Allan, F., Berger, D., Jimenez Castro, P.D., Collins, J.B., Crellen, T., Duque-Correa, M.A., Ellis, P., Jaleta, T.G., et al. (2019). Evaluation of DNA Extraction Methods on Individual Helminth Egg and Larval Stages for Whole-Genome Sequencing. Front. Genet. *10*, 826.](http://paperpile.com/b/RvxIQI/LDEE)

[Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. *14*, 2611–2620.](http://paperpile.com/b/RvxIQI/aC09)

[Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics *32*, 3047–3048.](http://paperpile.com/b/RvxIQI/mlKY)

[Foth, B.J., Tsai, I.J., Reid, A.J., Bancroft, A.J., Nichol, S., Tracey, A., Holroyd, N., Cotton, J.A., Stanley, E.J., Zarowiecki, M., et al. (2014). Whipworm genome and dual-species transcriptome analyses provide molecular insights into an intimate host-parasite interaction. Nat. Genet. *46*, 693–700.](http://paperpile.com/b/RvxIQI/oI8m)

[Gandasegui, J., Martínez-Valladares, M., Grau-Pujol, B., Krolewiecki, A.J., Balaña-Fouce, R., Gelaye, W., van Lieshout, L., Kepha, S., Mandomando, I., Muñoz, J., et al. (2020). Role of DNA-detection–based tools for monitoring the soil-transmitted helminth treatment response in drug-efficacy trials. PLOS Neglected Tropical Diseases *14*, e0007931.](http://paperpile.com/b/RvxIQI/sF20)

[Ghai, R.R., Simons, N.D., Chapman, C.A., Omeja, P.A., Davies, T.J., Ting, N., and Goldberg, T.L. (2014). Hidden population structure and cross-species transmission of whipworms (Trichuris sp.) in humans and non-human primates in Uganda. PLoS Negl. Trop. Dis. *8*, e3256.](http://paperpile.com/b/RvxIQI/TZrk)

[Gonçalves, M.L.C., Araújo, A., and Ferreira, L.F. (2003). Human intestinal parasites in the past: new findings and a review. Mem. Inst. Oswaldo Cruz *98 Suppl 1*, 103–118.](http://paperpile.com/b/RvxIQI/D7DU)

[Graff, A., Bennion-Pedley, E., Jones, A.K., Ledger, M.L., Deforce, K., Degraeve, A., Byl, S., and Mitchell, P.D. (2020). A comparative study of parasites in three latrines from Medieval and Renaissance Brussels, Belgium (14th-17th centuries). Parasitology *147*, 1443–1451.](http://paperpile.com/b/RvxIQI/K1LM)

[Hahn, C., Bachmann, L., and Chevreux, B. (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads--a baiting and iterative mapping approach. Nucleic Acids Res. *41*, e129.](http://paperpile.com/b/RvxIQI/4rW6)

[Hansen, E.P., Tejedor, A.M., Thamsborg, S.M., Alstrup Hansen, T.V., Dahlerup, J.F., and Nejsum, P. (2016). Faecal egg counts and expulsion dynamics of the whipworm, Trichuris trichiura following self-infection. J. Helminthol. *90*, 298–302.](http://paperpile.com/b/RvxIQI/nNRU)

[Hansen, T.V.A., Nejsum, P., Olsen, A., and Thamsborg, S.M. (2013a). Genetic variation in codons 167, 198 and 200 of the beta-tubulin gene in whipworms (Trichuris spp.) from a range of domestic animals and wildlife. Vet. Parasitol. *193*, 141–149.](http://paperpile.com/b/RvxIQI/P6WJ)

[Hansen, T.V.A., Thamsborg, S.M., Olsen, A., Prichard, R.K., and Nejsum, P. (2013b). Genetic variations in the beta-tubulin gene and the internal transcribed spacer 2 region of Trichuris species from man and baboons. Parasit. Vectors *6*, 236.](http://paperpile.com/b/RvxIQI/uLTq)

[Hawash, M.B.F., Betson, M., Al-Jubury, A., Ketzis, J., LeeWillingham, A., Bertelsen, M.F., Cooper, P.J., Littlewood, D.T.J., Zhu, X.-Q., and Nejsum, P. (2016). Whipworms in humans and pigs: origins and demography. Parasit. Vectors *9*, 37.](http://paperpile.com/b/RvxIQI/hTAi)

[Howe, K.L., Bolt, B.J., Shafie, M., Kersey, P., and Berriman, M. (2017). WormBase ParaSite − a comprehensive resource for helminth genomics. Molecular and Biochemical Parasitology *215*, 2–10.](http://paperpile.com/b/RvxIQI/NEPa)

[Jun, G., Wing, M.K., Abecasis, G.R., and Kang, H.M. (2015). An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. Genome Res. *25*, 918–925.](http://paperpile.com/b/RvxIQI/eOru)

[Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.](http://paperpile.com/b/RvxIQI/dbPD)

[Keiser, J., and Utzinger, J. (2008). Efficacy of current drugs against soil-transmitted helminth infections: systematic review and meta-analysis. JAMA *299*, 1937–1948.](http://paperpile.com/b/RvxIQI/mXf7)

[Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A., and Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Mol. Ecol. Resour. *15*, 1179–1191.](http://paperpile.com/b/RvxIQI/T2Vk)

[Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics *15*, 356.](http://paperpile.com/b/RvxIQI/LGx7)

[Kotze, A.C., Gilleard, J.S., Doyle, S.R., and Prichard, R.K. (2020). Challenges and opportunities for the adoption of molecular diagnostics for anthelmintic resistance. Int. J. Parasitol. Drugs Drug Resist. *14*, 264–273.](http://paperpile.com/b/RvxIQI/c5H7)

[Kwa, M.S., Kooyman, F.N., Boersema, J.H., and Roos, M.H. (1993). Effect of selection for benzimidazole resistance in Haemonchus contortus on beta-tubulin isotype 1 and isotype 2 genes. Biochem. Biophys. Res. Commun. *191*, 413–419.](http://paperpile.com/b/RvxIQI/3SdM)

[Ledger, Ledger, Rowan, Marques, Sigmier, Šarkić, Redžić, Cahill, and Mitchell (2020). Intestinal Parasitic Infection in the Eastern Roman Empire During the Imperial Period and Late Antiquity. American Journal of Archaeology *124*, 631.](http://paperpile.com/b/RvxIQI/CPNz)

[Ledger, M.L., Anastasiou, E., Shillito, L.-M., Mackay, H., Bull, I.D., Haddow, S.D., Knüsel, C.J., and Mitchell, P.D. (2019). Parasite infection at the early farming community of Çatalhöyük. Antiquity *93*, 573–587.](http://paperpile.com/b/RvxIQI/yMOs)

[Levecke, B., Montresor, A., Albonico, M., Ame, S.M., Behnke, J.M., Bethony, J.M., Noumedem, C.D., Engels, D., Guillard, B., Kotze, A.C., et al. (2014). Assessment of anthelmintic efficacy of mebendazole in school children in six countries where soil-transmitted helminths are endemic. PLoS Negl. Trop. Dis. *8*, e3204.](http://paperpile.com/b/RvxIQI/GMqa)

[Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.](http://paperpile.com/b/RvxIQI/tavP)

[Liu, G.-H., Gasser, R.B., Nejsum, P., Wang, Y., Chen, Q., Song, H.-Q., and Zhu, X.-Q. (2013). Mitochondrial and nuclear ribosomal DNA evidence supports the existence of a new Trichuris species in the endangered françois’ leaf-monkey. PLoS One *8*, e66249.](http://paperpile.com/b/RvxIQI/LIEv)

[Meekums, H., Hawash, M.B.F., Sparks, A.M., Oviedo, Y., Sandoval, C., Chico, M.E., Stothard, J.R., Cooper, P.J., Nejsum, P., and Betson, M. (2015). A genetic analysis of Trichuris trichiura and Trichuris suis from Ecuador. Parasit. Vectors *8*, 168.](http://paperpile.com/b/RvxIQI/zILC)

[Olsen, A., Namwanje, H., Nejsum, P., Roepstorff, A., and Thamsborg, S.M. (2009). Albendazole and mebendazole have low efficacy against Trichuristrichiura in school-age children in Kabale District, Uganda. Trans. R. Soc. Trop. Med. Hyg. *103*, 443–446.](http://paperpile.com/b/RvxIQI/7wQJ)

[Patel, C., Coulibaly, J.T., Schulz, J.D., N’Gbesso, Y., Hattendorf, J., and Keiser, J. (2020). Efficacy and safety of ascending dosages of albendazole against in preschool-aged children, school-aged children and adults: A multi-cohort randomized controlled trial. EClinicalMedicine *22*, 100335.](http://paperpile.com/b/RvxIQI/DP8X)

[Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient Admixture in Human History. Genetics *192*, 1065–1093.](http://paperpile.com/b/RvxIQI/MgTI)

[Pickrell, J.K., and Pritchard, J.K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. *8*, e1002967.](http://paperpile.com/b/RvxIQI/desq)

[Pullan, R.L., Smith, J.L., Jasrasaria, R., and Brooker, S.J. (2014). Global numbers of infection and disease burden of soil transmitted helminth infections in 2010. Parasit. Vectors *7*, 37.](http://paperpile.com/b/RvxIQI/iY9i)

[Rashwan, N., Scott, M., and Prichard, R. (2017). Rapid Genotyping of β-tubulin Polymorphisms in Trichuris trichiura and Ascaris lumbricoides. PLoS Negl. Trop. Dis. *11*, e0005205.](http://paperpile.com/b/RvxIQI/itre)

[Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., and Vilo, J. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res. *47*, W191–W198.](http://paperpile.com/b/RvxIQI/BdOS)

[Ravasi, D.F., O’Riain, M.J., Davids, F., and Illing, N. (2012). Phylogenetic evidence that two distinct Trichuris genotypes infect both humans and non-human primates. PLoS One *7*, e44187.](http://paperpile.com/b/RvxIQI/upiQ)

[Rose Vineer, H., Morgan, E.R., Hertzberg, H., Bartley, D.J., Bosco, A., Charlier, J., Chartier, C., Claerebout, E., de Waal, T., Hendrickx, G., et al. (2020). Increasing importance of anthelmintic resistance in European livestock: creation and meta-analysis of an open database. Parasite *27*, 69.](http://paperpile.com/b/RvxIQI/ogvt)

[Sallé, G., Doyle, S.R., Cortet, J., Cabaret, J., Berriman, M., Holroyd, N., and Cotton, J.A. (2019). The global diversity of Haemonchus contortus is shaped by human intervention and climate. Nat. Commun. *10*, 4811.](http://paperpile.com/b/RvxIQI/cBJJ)

[Schubert, M., Lindgreen, S., and Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res. Notes *9*, 88.](http://paperpile.com/b/RvxIQI/xASA)

[Shumate, A., and Salzberg, S.L. (2020). Liftoff: accurate mapping of gene annotations. Bioinformatics.](http://paperpile.com/b/RvxIQI/N6gn)

[Skoglund, P., Northoff, B.H., Shunkov, M.V., Derevianko, A.P., Pääbo, S., Krause, J., and Jakobsson, M. (2014). Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. Proceedings of the National Academy of Sciences *111*, 2229–2234.](http://paperpile.com/b/RvxIQI/4uzR)

[Skotte, L., Korneliussen, T.S., and Albrechtsen, A. (2013). Estimating individual admixture proportions from next generation sequencing data. Genetics *195*, 693–702.](http://paperpile.com/b/RvxIQI/aW2k)

[Small, S.T., Labbé, F., Coulibaly, Y.I., Nutman, T.B., King, C.L., Serre, D., and Zimmerman, P.A. (2019). Human Migration and the Spread of the Nematode Parasite Wuchereria bancrofti. Mol. Biol. Evol. *36*, 1931–1941.](http://paperpile.com/b/RvxIQI/FD1M)

[Søe, M.J., Nejsum, P., Seersholm, F.V., Fredensborg, B.L., Habraken, R., Haase, K., Hald, M.M., Simonsen, R., Højlund, F., Blanke, L., et al. (2018). Ancient DNA from latrines in Northern Europe and the Middle East (500 BC–1700 AD) reveals past parasites and diet. PLOS ONE *13*, e0195481.](http://paperpile.com/b/RvxIQI/Pu5P)

[Spyrou, M.A., Bos, K.I., Herbig, A., and Krause, J. (2019). Ancient pathogen genomics as an emerging tool for infectious disease research. Nat. Rev. Genet. *20*, 323–340.](http://paperpile.com/b/RvxIQI/C4Pn)

[Terhorst, J., Kamm, J.A., and Song, Y.S. (2017). Robust and scalable inference of population history from hundreds of unphased whole genomes. Nat. Genet. *49*, 303–309.](http://paperpile.com/b/RvxIQI/fcPO)

[Vlaminck, J., Cools, P., Albonico, M., Ame, S., Chanthapaseuth, T., Viengxay, V., Do Trung, D., Osei-Atweneboana, M.Y., Asuming-Brempong, E., Jahirul Karim, M., et al. (2020). Piloting a surveillance system to monitor the global patterns of drug efficacy and the emergence of anthelmintic resistance in soil-transmitted helminth control programs: a Starworms study protocol. Gates Open Res *4*, 28.](http://paperpile.com/b/RvxIQI/buSy)

[Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H., Gurtowski, J., and Schatz, M.C. (2017). GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics *33*, 2202–2204.](http://paperpile.com/b/RvxIQI/DY8F)

[Wood, D.E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. Genome Biol. *20*, 257.](http://paperpile.com/b/RvxIQI/6Vt4)

[World Health Organization (2020). Ending the neglect to attain the Sustainable Development Goals: A road map for neglected tropical diseases 2021–2030 (World Health Organization, Geneva).](http://paperpile.com/b/RvxIQI/guPt)

[Zheng, X., Levine, D., Shen, J., Gogarten, S.M., Laurie, C., and Weir, B.S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics *28*, 3326–3328.](http://paperpile.com/b/RvxIQI/LjPY)