Supplementary Information

Global diversity of soil-transmitted helminths reveals population-biased genetic variation that impacts diagnostic targets

[Origin of samples and data](#_mezl7bx7l060) 2

[ENA accessions and Bioprojects for datasets used in this study](#_rzwqaxueeyzw) 8

[References](#_tf8lv1fs8axs) 10

**Origin of samples and data**

We aimed to source worm isolates and egg/faecal data with strong infections for any soil-transmitted helminth (STH) present. Below is a breakdown of the number and types of samples per country, any relevant ethical approvals, and accession numbers relating to publicly available data. The 150 samples sequenced in this study are specified below (‘this study’) and in the main text.

1. Argentina (n = 1).
   1. n = 1; whole genome sequencing data of DNA extracted from concentrated pools of eggs of all STH species; the original study protocol was approved by the bioethics committee of Colegio de Médicos de la Provincia de Salta and the IRBs of BCM (protocol number H-34926); citation: this study. Key collaborators: Nicolas R Caro, Ruben O Cimino, Alejandro J Krolewiecki, Rojelio Mejia.

1. Bangladesh (n = 10).
   1. n = 10; whole genome sequencing data of DNA extracted from faecal samples; the original study protocol was approved by the Ethical Review Committee at icddr,b (PR-14105), the Committee for the Protection of Human Subjects at the University of California, Berkeley (2014-08-6658), and the institutional review board at Stanford University (27864); citation: this study. Key collaborators: John M Colford Jr, Jade Benjamin-Chung, Steven A Williams.

1. Benin (n = 25).
   1. n = 25; whole genome sequencing data of DNA extracted from faecal samples, from the DeWorm3 project. One sample did not yield sequencing reads. The original study protocol was approved by the Institut de Recherche Clinique au Bénin (IRCB) through the National Ethics Committee for Health Research (002-2017/CNERS-MS) from the Ministry of Health, the Human Subjects Division at the University of Washington (STUDY00000180) and the Data Safety and Monitoring Committee (DSMC); citation: this study. Key collaborators: Moudachirou Ibikounlé, Adrian JF Luty, Judd L Walson.

1. Cameroon (n = 6).
   1. n = 1; whole genome sequencing data of DNA extracted from a *S. mansoni* worm; ENA study accession: [PRJEB2679](https://www.ebi.ac.uk/ena/browser/view/PRJEB2679); Sample accession code: ERR103050; citation:[(Crellen et al., 2016)](https://paperpile.com/c/zev5C5/SWWJB).
   2. n = 5, whole genome sequencing data of DNA extracted from concentrated eggs from faecal samples; ENA study accession: PRJEB44010; sample accession codes: ERR9805789-93; citation:[(Doyle et al., 2022)](https://paperpile.com/c/zev5C5/CBlZ2).

1. China (n = 8)
   1. n = 1, whole genome sequencing data of DNA extracted from an *N. americanus* worm*;* ENA study accession: [PRJNA304165](https://www.ebi.ac.uk/ena/browser/view/PRJNA304165); sample accession code: SRR2968128; citation:[(Pilotte et al., 2016)](https://paperpile.com/c/zev5C5/jPZQQ).
   2. n = 7, whole genome sequencing data of DNA extracted from *T. trichiura* worms; ENA study accession: PRJEB44010; sample accession codes: ERR9805779-785; citation:[(Doyle et al., 2022)](https://paperpile.com/c/zev5C5/CBlZ2).

1. Democratic Republic of Congo (n = 2).
   1. n = 1, whole genome sequencing data of DNA extracted from samples; citation:[(Papaiakovou et al., 2023)](https://paperpile.com/c/zev5C5/aCmla).
   2. n = 1, whole genome sequencing data of DNA extracted from a faecal extract, part of the same protocol approval as above. The original study protocol was approved by the Democratic Republic of Congo / University Hospital, Ghent University, Belgium (M104; Catholic University of Bukavu, Democratic Republic of the Congo [Ref: UCB/CIE/NC/016/2016], the Ministry of Public Health, Democratic Republic of the Congo [Ref: 062/CD/DPS/SK/2017]); citation: this study.

1. Ecuador (n = 1).
   1. n = 1, whole genome sequencing data of DNA extracted from an *A. lumbricoides* worm, ENA study accession: [PRJNA304165](https://www.ebi.ac.uk/ena/browser/view/PRJNA304165); Sample accession code: SRR2968217; citation:[(Pilotte et al., 2016)](https://paperpile.com/c/zev5C5/jPZQQ).

1. Ethiopia (n = 5).
   1. n = 2, whole genome sequencing data of DNA extracted from faecal samples; ETH\_ET018, ETH\_ET103 in[(Papaiakovou et al., 2023)](https://paperpile.com/c/zev5C5/aCmla); ENA study accession: [PRJNA847183](https://www.ebi.ac.uk/ena/browser/view/PRJNA847183); sample accession codes: SAMN28922051-SAMN28922052; citation:[(Papaiakovou et al., 2023)](https://paperpile.com/c/zev5C5/aCmla).
   2. n = 2, whole genome sequencing data of DNA extracted from two *A. lumbricoides* worms. The original study protocol was approved by Ethical Review Committee, Faculty of Medicine and Health Sciences / University Hospital, Ghent University, Belgium [Ref: B670201627755 and PA2014/003], Jimma University, Ethiopia (Ref: RPGC/547/2016); citation: this study. Key collaborators: Piet Cools, Bruno Levecke, Zeleke Mekonnen.
   3. n = 1; whole genome sequencing data of DNA extracted from a faecal sample. The original study protocol was approved by the Ethical Review Committee, Faculty of Medicine and Health Sciences / University Hospital, Ghent University, Belgium (Ref: B670201627755 and PA2014/003), and by the Jimma University, Ethiopia (Ref: RPGC/547/2016); citation: this study. Key collaborators: Piet Cools, Bruno Levecke, Zeleke Mekonnen.

1. Fiji (n = 2).
   1. n = 2, whole genome sequencing data of two aliquots of DNA extracted from a single *A. lumbricoides* worm; the worm was provided by the Natural History Museum, London, UK under registration number: 2012.11.19.1, *Ascaris lumbricoides* Linnaeus, 1758 -- Ascaridinae; Ascarididae; Ascaridoidea; Spirurina; Rhabditida; Chromadorea, 1, spirit material.

1. Guadeloupe (n = 4).
   1. n = 4, whole genome sequencing data of DNA extracted from *S. mansoni* worms; ENA study accession: [PRJEB3054](https://www.ebi.ac.uk/ena/browser/view/PRJEB3054); Sample accession codes: ERR539842-45; citation:[(Crellen et al., 2016)](https://paperpile.com/c/zev5C5/SWWJB).

1. Honduras (n = 8).
   1. n = 8, whole genome sequencing data of DNA extracted from *T. trichiura* worms; ENA study accession: PRJEB44010; sample accession codes: ERR9805798-805; citation:[(Doyle et al., 2022)](https://paperpile.com/c/zev5C5/CBlZ2).

1. India (n = 25).
   1. n = 25, whole genome sequencing data of DNA extracted from faecal samples, part of the DeWorm3 project. The original study protocol was approved by the Christian Medical College Institutional Review Board in Vellore, India (10392). The study was also approved by The Human Subjects Division at the University of Washington (STUDY00000180) and the genome skimming project was revised by the Data Safety and Monitoring Committee (DSMC); citation: this study. Key collaborators: Sitara SR Ajjampur, Malathi Manuel, Judd L Walson.

1. Italy (n = 2).
   1. n = 2; whole genome sequencing data of DNA extracted from faecal samples; CAM1 & CAM2 in Papaiakovou et al (2023); ENA study accession: PRJNA847183; sample accession code: SAMN28922036-37; citation:[(Jenkins et al., 2018; Papaiakovou et al., 2023)](https://paperpile.com/c/zev5C5/aCmla+qK7Uv).

1. Kenya (n = 76)
   1. n = 68, whole genome sequencing data of DNA extracted from individual *A. lumbricoides* worms; ENA study accession: PRJNA511012; sample accession codes: SRX5228374-SRX5228441; citation:[(Easton et al., 2020)](https://paperpile.com/c/zev5C5/1ypHy).
   2. n = 7; whole genome sequencing data of DNA extracted from faecal samples; the original study protocol was approved by the Scientific and Ethics Review Committees (ERC) of the Kenya Medical Research Institute (KEMRI, SSC #1820); citation:[(Onkanga et al., 2016; Secor et al., 2020)](https://paperpile.com/c/zev5C5/CPcaG+agg0l). Key collaborators: Maurice Odiere, Pauline Mwinzi.
   3. n = 1; whole genome sequencing data of DNA extracted from a *S. mansoni* worm; ENA study accession: PRJEB2679; citation:[(Crellen et al., 2016)](https://paperpile.com/c/zev5C5/SWWJB).

1. Malawi (n = 25).
   1. n = 25; whole genome sequencing data of DNA extracted from faecal samples, part of the DeWorm3 project. The original study protocol was approved by The London School of Hygiene and Tropical Medicine (12013), The College of Medicine Research Ethics Committee (P.04/17/2161) in Malawi. The study was also approved by The Human Subjects Division at the University of Washington (STUDY00000180) and the genome skimming project was revised by the Data Safety and Monitoring Committee (DSMC); citation: this study. Key collaborators: Robin Bailey, David Chaima, Khumbo Kalua, Judd L Walson, Stefan Witek-McManus.

1. Malaysia (n = 650).
   1. n = 650; whole genome sequencing data of DNA extracted from faecal samples; ENA study accession: [PRJNA797994](https://www.ebi.ac.uk/ena/browser/view/PRJNA797994); sample accession codes: SAMN25042866-25043515; citation:[(Tee et al., 2022)](https://paperpile.com/c/zev5C5/8mYT3).

1. Mozambique (n = 20).
   1. n = 20; whole genome sequencing data of DNA extracted from concentrated eggs from faecal samples; the original study protocol (WASH-IT) was approved by the National Bioethics Committee for Health in Mozambique. ENA study accession: PRJEB53235. Key collaborators: Maria Cambra-Pellejà, Anélsio Cossa, Javier Gandasegui, Berta Grau-Pujol, Inácio Mandomando, Maria Martínez-Valladares, Augusto Messa Jr., Osvaldo Muchisse, Jose Muñoz, Valdemiro Novela, Charfudin Sacoor.

1. Myanmar (n = 38)
   1. n = 6; whole genome sequencing data of DNA extracted from faecal samples: SAMN28922044 (MMR\_TKU23), SAMN28922045 (MMR\_TKU25), SAMN28922049 (MMR\_TKU102), SAMN28922046 (MMR\_NDK63), SAMN28922047 (MMR\_NDK92), SAMN28922048 (MMR\_NDK113) in Papaiakovou et al (2023); Study accession number: [PRJNA847183](https://www.ncbi.nlm.nih.gov/bioproject/847183); citation:[(Papaiakovou et al., 2023)](https://paperpile.com/c/zev5C5/aCmla).
   2. n = 32; whole genome sequencing data of DNA extracted from faecal samples; the original study protocol was approved by Imperial College London, UK (Ethical Review Ref: 17IC4249 and 17IC4249 NoA1); citation: this study. Key collaborators: Roy M Anderson, Julia Dunn.

1. Nigeria (n = 11).
   1. n = 11; whole genome sequencing data of DNA extracted from faecal samples; the original study protocol was approved by the Health Research Ethics Committee of the Kebbi State Ministry of Health, Nigeria (reference number:105:23/2021); citation: this study. Key collaborator: Olumide Ajibola.

1. Puerto Rico (n = 1).
   1. n = 1; whole genome sequencing data froma *S. mansoni* worm; ENA study accession: PRJEB2679; sample accession code: ERR046038; citation:[(Crellen et al., 2016)](https://paperpile.com/c/zev5C5/SWWJB).

1. Senegal (n = 1).
   1. n = 1; whole genome sequencing data of DNA extracted from an *S. mansoni* worm; ENA study accession: PRJEB2679; sample accession code: ERR103049; citation:[(Crellen et al., 2016)](https://paperpile.com/c/zev5C5/SWWJB).

1. South Africa (n = 7)
   1. n = 7; whole genome sequencing data of DNA extracted from faecal samples; the original study protocol was approved by the Biomedical Research Ethics Administration, University of KwaZulu-Natal KZN (Ref BF029/07); citation: this study. Key collaborator: Eyrun F Kjetland.

1. Sri Lanka (n = 7).
   1. n = 6; whole genome sequencing data of DNA extracted from faecal samples; ENA study accession: PRJNA847183; sample accession codes: SAMN28922038-43; citation:[(Papaiakovou et al., 2023)](https://paperpile.com/c/zev5C5/aCmla).
   2. n = 1; whole genome sequencing data of DNA extracted from faecal samples. The original study protocol was approved by the Ethical Review Committee, Faculty of Medicine, University of Peradeniya, Sri Lanka [Ref: 2015/EC/58]; citation: this study. Key collaborators: Cinzia Cantacessi, Timothy P Jenkins.

1. Tanzania (n = 5).
   1. n = 5; whole genome sequencing data of DNA extracted from concentrated eggs from faecal samples; ENA study accession: PRJEB44010; sample accession codes: ERR9805806-810; citation:[(Doyle et al., 2022)](https://paperpile.com/c/zev5C5/CBlZ2).

1. Thailand (n = 15).
   1. n = 15; whole genome sequencing data of DNA extracted fromindividual *S. stercoralis* worms; ENA study accession: PRJNA602131; accession codes: SRR10915458-5472; Citation:[(Aupalee et al., 2020)](https://paperpile.com/c/zev5C5/HMxKp).

1. Uganda (n = 47)
   1. n = 32; whole genome sequencing data of DNA extracted from faecal samples (4 Gb, for 31 samples, 12 Gb for BLANK sample); the original study protocol was approved by the UVRI Research Ethics Committee, as well as the Uganda National Council for Science and Technology and the University of Manchester Research Ethics Committee[(Houlder et al., 2023)](https://paperpile.com/c/zev5C5/F0jVp). Key collaborators: Emma Houlder, Andrew S MacDonald, Harriet Mpairwe.
   2. n = 1; whole genome sequencing data of DNA extracted from a *T. trichiura* worm; ENA study accession: PRJNA304165; sample accession code: SRR2968131; citation:[(Pilotte et al., 2016)](https://paperpile.com/c/zev5C5/jPZQQ).
   3. n = 12; whole genome sequencing data of DNA extracted from *T. trichiura* worms; ENA study accession: PRJEB44010; sample accession codes: ERR9805811-22; citation:[(Doyle et al., 2022)](https://paperpile.com/c/zev5C5/CBlZ2).
   4. n = 2; whole genome sequencing data of DNA extracted from *S. mansoni* worms; ENA study accession: PRJEB2679; sample accession code: ERR119615; citation:[(Crellen et al., 2016)](https://paperpile.com/c/zev5C5/SWWJB).

1. USA (n = 1).
   1. n = 1; whole genome sequencing data from *S. stercoralis*; ENA study accession: [PRJEB2679](https://www.ebi.ac.uk/ena/browser/view/PRJEB2679); sample accession code: ERR066168.

### 

### 

### **ENA accessions and Bioprojects for datasets used in this study**

1. *Schistosoma mansoni* ENA data ([ERR103050](https://www.ebi.ac.uk/ena/browser/view/ERR103050), [ERR539842](https://www.ebi.ac.uk/ena/browser/view/ERR539842),[ERR539843](https://www.ebi.ac.uk/ena/browser/view/ERR539843), [ERR53984](https://www.ebi.ac.uk/ena/browser/view/ERR539843)4, [ERR53984](https://www.ebi.ac.uk/ena/browser/view/ERR539843)5,[ERR119614](https://www.ebi.ac.uk/ena/browser/view/ERR119614), [ERR11961](https://www.ebi.ac.uk/ena/browser/view/ERR119614)5, [ERR997461](https://www.ebi.ac.uk/ena/browser/view/ERR997461)):
   1. BioProject Number:
      1. PRJEB3054:<https://www.ebi.ac.uk/ena/browser/view/PRJEB3054>
      2. PRJEB2679:<https://www.ebi.ac.uk/ena/browser/view/PRJEB2679>
   2. Link to study: https://doi.org/10.1038/srep20954
2. *Trichuris trichiura* worm and egg data from Honduras, China, Uganda under the European Nucleotide Archive (ENA)
   1. BioProject Number: [ERP128004](http://www.ebi.ac.uk/ena/browser/view/ERP128004)

<https://www.ebi.ac.uk/ena/browser/view/PRJEB44010> (Doyle)

* 1. Link to the study:https://doi.org/10.1038/s41467-022-31487-x

1. *Necator*, *Ascaris*, *Trichuris* worm data from Accession :
   1. BioProject Number: PRJNA304165:<https://www.ebi.ac.uk/ena/browser/view/PRJNA304165>
   2. Link to the study: https://doi.org/10.1371/journal.pntd.0004578
2. Faecal metagenomes from multiple countries, under the project accession:
   1. BioProject Number PRJNA847183 (IJP Papaiakovou) =<https://www.ebi.ac.uk/ena/browser/view/PRJNA847183>
   2. Link to the study: https://doi.org/10.1016/j.ijpara.2022.12.002
3. Easton *Ascaris* worms from Kenya under:
   1. BioProject Number PRJNA511012 =<https://www.ebi.ac.uk/ena/browser/view/PRJNA511012>;
   2. Link to the study = https://doi.org/10.7554/eLife.61562
4. Faecal metagenomes from Malaysia
   1. BioProject Number: PRJNA797994,<https://www.ebi.ac.uk/ena/browser/view/PRJNA797994>
   2. Link to the study: https://doi.org/10.1186/s40168-022-01385-x
5. Concentrated egg samples from Mozambique can be found under
   1. BioProject Number PRJEB53235 =<https://www.ebi.ac.uk/ena/browser/view/PRJEB53235>
   2. Link to the study: NA
6. Schistosoma mansoni data from Puerto Rico and Senegal:
   1. BioProject Number: [PRJEB31375](https://www.ebi.ac.uk/ena/browser/view/PRJEB31375)<https://www.ebi.ac.uk/ena/browser/view/PRJEB31375?show=reads>;
   2. Link to the study = https://doi.org/10.1038/s41467-021-24958-0
7. Strongyloides stercoralis worms from Thailand
   1. BioProject Number: PRJNA602131 =<https://www.ebi.ac.uk/ena/browser/view/PRJNA602131>
   2. Link to the study: https://doi.org/10.1186/s13071-020-04115-0
8. Faecal extracts from Uganda were generated as part of this study:
   1. Link to the study:<https://doi.org/10.1038/s41467-023-37502-z> - data are available from the corresponding author of that study
9. Strongyloides dataset used:
   1. BioProject Number: PRJEB2679,<https://www.ebi.ac.uk/ena/browser/view/ERR066168>
   2. Link to the study: NA

### 

### **References**

[Aupalee, K., Wijit, A., Singphai, K., Rödelsperger, C., Zhou, S., Saeung, A., & Streit, A. (2020). Genomic studies on Strongyloides stercoralis in northern and western Thailand. *Parasites & Vectors*, *13*(1), 250.](http://paperpile.com/b/zev5C5/HMxKp)

[Crellen, T., Allan, F., David, S., Durrant, C., Huckvale, T., Holroyd, N., Emery, A. M., Rollinson, D., Aanensen, D. M., Berriman, M., Webster, J. P., & Cotton, J. A. (2016). Whole genome resequencing of the human parasite Schistosoma mansoni reveals population history and effects of selection. *Scientific Reports*, *6*, 20954.](http://paperpile.com/b/zev5C5/SWWJB)

[Doyle, S. R., Søe, M. J., Nejsum, P., Betson, M., Cooper, P. J., Peng, L., Zhu, X.-Q., Sanchez, A., Matamoros, G., Sandoval, G. A. F., Cutillas, C., Tchuenté, L.-A. T., Mekonnen, Z., Ame, S. M., Namwanje, H., Levecke, B., Berriman, M., Fredensborg, B. L., & Kapel, C. M. O. (2022). Population genomics of ancient and modern Trichuris trichiura. *Nature Communications*, *13*(1), 1–12.](http://paperpile.com/b/zev5C5/CBlZ2)

[Easton, A., Gao, S., Lawton, S. P., Bennuru, S., Khan, A., Dahlstrom, E., Oliveira, R. G., Kepha, S., Porcella, S. F., Webster, J., Anderson, R., Grigg, M. E., Davis, R. E., Wang, J., & Nutman, T. B. (2020). Molecular evidence of hybridization between pig and human Ascaris indicates an interbred species complex infecting humans. *eLife*, *9*. https://doi.org/](http://paperpile.com/b/zev5C5/1ypHy)[10.7554/eLife.61562](http://dx.doi.org/10.7554/eLife.61562)

[Houlder, E. L., Costain, A. H., Nambuya, I., Brown, S. L., Koopman, J. P. R., Langenberg, M. C. C., Janse, J. J., Hoogerwerf, M. A., Ridley, A. J. L., Forde-Thomas, J. E., Colombo, S. A. P., Winkel, B. M. F., Galdon, A. A., Hoffmann, K. F., Cook, P. C., Roestenberg, M., Mpairwe, H., & MacDonald, A. S. (2023). Pulmonary inflammation promoted by type-2 dendritic cells is a feature of human and murine schistosomiasis. *Nature Communications*, *14*(1), 1–12.](http://paperpile.com/b/zev5C5/F0jVp)

[Jenkins, T. P., Formenti, F., Castro, C., Piubelli, C., Perandin, F., Buonfrate, D., Otranto, D., Griffin, J. L., Krause, L., Bisoffi, Z., & Cantacessi, C. (2018). A comprehensive analysis of the faecal microbiome and metabolome of Strongyloides stercoralis infected volunteers from a non-endemic area. *Scientific Reports*, *8*(1), 15651.](http://paperpile.com/b/zev5C5/qK7Uv)

[Onkanga, I. O., Mwinzi, P. N. M., Muchiri, G., Andiego, K., Omedo, M., Karanja, D. M. S., Wiegand, R. E., Secor, W. E., & Montgomery, S. P. (2016). Impact of two rounds of praziquantel mass drug administration on Schistosoma mansoni infection prevalence and intensity: a comparison between community wide treatment and school based treatment in western Kenya. *International Journal for Parasitology*, *46*(7), 439–445.](http://paperpile.com/b/zev5C5/CPcaG)

[Papaiakovou, M., Fraija-Fernández, N., James, K., Briscoe, A. G., Hall, A., Jenkins, T. P., Dunn, J., Levecke, B., Mekonnen, Z., Cools, P., Doyle, S. R., Cantacessi, C., & Littlewood, D. T. J. (2023). Evaluation of genome skimming to detect and characterise human and livestock helminths. *International Journal for Parasitology*, *53*(2), 69–79.](http://paperpile.com/b/zev5C5/aCmla)

[Pilotte, N., Papaiakovou, M., Grant, J. R., Bierwert, L. A., Llewellyn, S., McCarthy, J. S., & Williams, S. A. (2016). Improved PCR-Based Detection of Soil Transmitted Helminth Infections Using a Next-Generation Sequencing Approach to Assay Design. *PLoS Neglected Tropical Diseases*, *10*(3), e0004578.](http://paperpile.com/b/zev5C5/jPZQQ)

[Secor, W. E., Wiegand, R. E., Montgomery, S. P., Karanja, D. M. S., & Odiere, M. R. (2020). Comparison of school-based and community-wide mass drug administration for schistosomiasis control in an area of western Kenya with high initial Schistosoma mansoni infection prevalence: A cluster randomized trial. *The American Journal of Tropical Medicine and Hygiene*, *102*(2), 318–327.](http://paperpile.com/b/zev5C5/agg0l)

[Tee, M. Z., Er, Y. X., Easton, A. V., Yap, N. J., Lee, I. L., Devlin, J., Chen, Z., Ng, K. S., Subramanian, P., Angelova, A., Oyesola, O., Sargsian, S., Ngui, R., Beiting, D. P., Boey, C. C. M., Chua, K. H., Cadwell, K., Lim, Y. A. L., Loke, P. ’ng, & Lee, S. C. (2022). Gut microbiome of helminth-infected indigenous Malaysians is context dependent. *Microbiome*, *10*(1), 214.](http://paperpile.com/b/zev5C5/8mYT3)