The gut microbiome but not the resistome is associated with urogenital schistosomiasis in preschool-aged children

Derick N.M. Osakunor1✉, Patrick Munk2, Takafira Mduluza3, Thomas N. Petersen2, Christian Brinch2, Alasdair Ivens1, Theresa Chimponda3, Seth A. Amanfo4,5, Janice Murray1,5, Mark E.J. Woolhouse4,5, Frank M. Aarestrup2 & Francisca Mutapi1,5

Helminth parasites have been shown to have systemic effects in the host. Using shotgun metagenomic sequencing, we characterise the gut microbiome and resistome of 113 Zimbabwean preschool-aged children (1–5 years). We test the hypothesis that infection with the human helminth parasite, Schistosoma haematobium, is associated with changes in gut microbial and antimicrobial resistance gene abundance/diversity. Here, we show that bacteria phyla Bacteroidetes, Firmicutes, Proteobacteria, and fungi phyla Ascomycota, Microsporidia, Zoopagomycota dominate the microbiome. The abundance of Proteobacteria, Ascomycota, and Basidiomycota differ between schistosome-infected versus uninfected children. Specifically, infection is associated with increases in Pseudomonas, Stenotrophomonas, Derxia, Thalassospira, Aspergillus, Tricholoma, and Periglandula, with a decrease in Azospirillum. We find 262 AMR genes, from 12 functional drug classes, but no association with individual-specific data. To our knowledge, we describe a novel metagenomic dataset of Zimbabwean preschool-aged children, indicating an association between urogenital schistosome infection and changes in the gut microbiome.
The human gut comprises a diverse ecosystem of microbes, predominantly bacteria, in addition to viruses, fungi and other eukaryotes. Evidence shows that humans rely on the symbiotic relationship with the resident microbial taxa present in humans (microbiota) for extracting essential nutrients from food, as a first line of protection from pathogens, and as a mechanism for shaping the immune system. Shotgun metagenomic sequencing has allowed characterisation of the microbiome (the assembly of genomes of the microbiota) among different human populations, showing considerable heterogeneity. Populations in Africa have been underrepresented in such studies, with a major focus on Western populations. Other studies have included diverse but older populations, not allowing the factors inherent to African childhood to be fully disentangled. Given the potential window of opportunity for influencing health through the microbiome in infants and young children, research focus on this age group is important. Findings from consortiums including the Human Heredity and Health in Africa (H3 Africa) and the HapMap Project will be invaluable for informing nutritionists in Africa.

The composition of the gut microbiome is influenced by age, diet and geography, host genotype, exposure to maternal microbiota, as well as environmental factors including the role of protozoal and helminth parasites. In Africa, children are exposed to several acute and chronic parasitic infections that can impact children's growth and development. In particular, helminth parasites (as shown for schistosome worms) can be contracted by children as young as 6-month-old or less, and these can persist into the second decade of life where they modulate the immune system as well as cause morbidity and pathology. In very young children, the gut microbial population continues to evolve until about age 3–5 years, thus it is important to establish how external factors, especially infections, that young children are exposed to, influence the microbiome.

Schistosomiasis is a disease caused by infection with trematodes of the genus Schistosoma—the predominant human species found in Africa being the urogenital (S. haematobium) and intestinal (S. mansoni) forms. Pathology from the disease is mostly from immunological reactions to trapped eggs attempting to migrate through to the bladder or intestinal lumen, depending on the species involved. The infection causes immunomodulatory effects which help to promote both parasite and host survival, and in preschool-aged children, consequences can extend to malnutrition, poor growth and cognition, reduced vaccine efficacy, and altered prognosis of co-infections. Treatment for schistosomiasis is through administration of the antihelminthic drug, praziquantel, which is effective against all schistosome species.

A number of experimental and human studies, including our own, have examined the association between helminth infections and the structure and composition of the gut microbiome. It has been suggested that the immunomodulatory effects of schistosome infection can extend to the gut microbiota through direct intestinal or systemic interactions. Work in experimental models shows that depletion of the gut bacteria is associated with reduced S. mansoni egg excretion, gut pathology and inflammation. Recently, fluctuations in the composition of the gut microbiota of mice infected with S. mansoni, before and after intestinal damage from egg transmission was shown. This is consistent with a role of the mammalian gut microbiota in the pathogenesis of schistosome infection. However, unlike S. mansoni (intestinal form) that inhabits the same environment as the gut microbiota, S. haematobium predominantly resides in the venous plexus of the bladder (although occasionally in the mesenteric circulation), and thus presents a need to study the indirect systemic impacts of infection on the gut microbiota—a more likely interaction. Correlations between the gut microbiome and systemic diseases such as rheumatoid arthritis suggest the importance of such systemic interactions. Phenotypic and mechanistic studies on systemic interactions between helminths and the microbiome in natural human infections are still in their infancy, and more studies are needed. In a previous study, we found differences in the gut microbiome between S. haematobium infected versus uninfected children, aged up to 13 years old. This was supported by Schneeberger et al., suggesting that genetic and environmental factors may play an additional role. Recently, a study conducted among older children (11–15 years) in Nigeria showed that urogenital schistosomiasis is associated with disruptions in the gut microbiome, suggesting that this may be a further consequence of schistosome infection. However, substantial knowledge gaps on the interaction between the gut microbiome and Schistosoma infection in preschool-aged children still exist. The biggest challenges are demonstrating causation and elucidating mechanistic pathways for any existing interactions.

In addition to the schistosome-microbiota interactions, other interactions relevant to the health of the host have also been reported to occur within the gut ecosystem. One such example is the ability of Salmonella to persist in the body by attaching to intestinal schistosomes, evading repeated antibiotic treatments, increasing the Salmonella population and eventually, potential antibiotic resistance. Furthermore, the microbiome is a reservoir for antimicrobial resistance (AMR) genes (resistome), and provides an ideal environment for AMR gene exchange among the resident and transitory bacterial population. Such interactions are likely to impact the structure and diversity of the resident microbial population, as well as the overall AMR gene composition. Therefore, the structure of AMR genes to bacteria and determined if this was associated with any host-related factors including socio-demography, antibiotic use, current schistosome infection, as well as feeding, growth and nutritional indices. AMR remains one of the largest threats to human health, with numerous calls for antibiotic resistance stewardship worldwide. However, this population of African preschool-aged children is understudied, and almost all AMR gene studies are conducted in industrialised settings. Such settings contrast with low and middle-income countries in terms of access to safe water and sanitation, and access to antibiotics, with or without prescriptions.

Within the framework of a larger paediatric schistosomiasis study in Zimbabwe, the present study focuses on shotgun metagenomic sequencing of stool samples from preschool-aged children, aged 1–5 years old. To add to the repository of information on the gut microbiome and AMR studies in this young population, we characterise the structure and diversity of the gut microbiome (to include the fungi repertoire) and resistome. We apply these data to test the hypothesis that S. haematobium infection is associated with alterations in the gut microbial and AMR gene abundance and diversity. We find that the microbiome but not the resistome is associated with S. haematobium infection, independent of age, sex and village.

**Results**

**Population characteristics.** Of the 113 participants included in the study, the mean age was 3.7 ± 1.1 years, of which 56 were females (49.6%). Sixty-eight (60.2%) and 45 (39.8%) children were from Chihiuri and Mupfure villages, respectively. Antimicrobial use data showed that 58 (51.3%) participants had received antibiotics [amoxicillin (31), co-trimoxazole (27), both (9)], while 18 (15.9%) had not; no information was obtained for the remaining 37 (32.7%) participants. Previous history of praziquantel treatment (for schistosome infection) was reported.
among 29/105 (27.6%) children. *S. haematobium* infection prevalence was 15.9% (18/113), with mean infection intensity of 1.79 eggs/10 ml urine (SEM = 0.76; range = 0–74).

We gathered data on the history of feeding habits and nutritional status of the children. The majority (83.6%) were breastfed, with duration ranging from 2 to 48 months (median = 18 months IQR: 17–20). Children were introduced to solid foods between 1 and 24 months after birth. Diet comprised mainly of traditional maize flour porridges (97%; 96/103), the commercial Cerelac® porridge (1.9%; 2/103), and potatoes (1%; 1/103). Anthropometric measures, adjusted for age, were used to assess nutritional status. Based on the weight-for-height Z scores (WHZ), 3.7% (4/107) of individuals were malnourished and 14.7% (16/109) were stunted, based on the height-for-age Z scores (HAZ).

**Taxonomic composition of the microbiome.** The number of classified read pairs per sample ranged from 3,994,704 to 13,164,482. An average 45.1% of read pairs were mapped to specific reference sequences in the genomic database; this is similar to other studies with the proportion of unmapped reads ranging from 42% to 68%. At any taxonomic level, a putative taxonomic classification could not be assigned to at least 33% of the mapped read pairs and were thus classified as “Unknown”.

In the 113 stool samples, 845 bacteria genera (from 20 unique phyla) and 228 fungi genera (from six unique phyla) were detected. As shown in Fig. 1, the most abundant bacteria phyla in decreasing order were Bacteroidetes (genera: *Prevotella, Bacteroides, Alistipes*), Firmicutes (genera: *Eubacterium, Faecalibacterium, Clostridium, Roseburia*), and Proteobacteria (genus: *Succinatimonas*). The most abundant fungi phyla were Ascomycota (genera: *Protomyces, Aspergillus, Taphrina, Saccharomyces, Candida, Nakaseomyces*), Microsporidia (genus: *Enterocytoszoom*), and Zoopagomycota (genus: *Entomophthora*) [Fig. 2]. These phyla dominated the microbiome and were present in all samples.

**Variation in the microbiome and association with sample metadata.** Principal component analysis (PCA) was used to initially examine variability and patterns in the data set across the first two principal components. At the phylum level, PCA explained 62% and 42.0% of the total variation in fungi and bacteria, respectively. At the genus level, however, PCA explained 34% and 16% of the total variation in fungi and bacteria respectively. The model showed homogeneity in components with no distinct clustering according to metadata and may reflect a high diversity in the cohort. PCA plots and cluster dendrograms for bacteria and fungi content per sample is shown in Supplementary Figs. 1–3.

Permutational multivariate analysis of variance (PERMANOVA) analysis showed a significant effect of age (false discovery rate (FDR) = 0.024) and village (FDR = 0.039) [details shown in Supplementary Fig. 4], schistosome infection status (FDR = 0.039) and schistosome infection intensity (FDR = 0.012) on bacteria genera, across samples. There was also a significant effect of schistosome infection status (FDR = 0.006) and schistosome infection intensity (FDR = 0.006) on fungi genera, across the samples. For both bacteria and fungi genera, no such effects were found for sex, nutritional and growth variables, feeding, previous

---

**Fig. 1 Overview of bacterial microbiota abundance and composition.** From read mapping to the genomic database, abundance was calculated for each microbial taxa across all samples. Stacked bar charts show the most abundant bacteria (a) phyla and (b) genera per sample, proportional to the total microbiota within each sample (n = 113 biologically independent samples). Charts were generated using normalised, zero-corrected abundance matrices. “Unknown” represents abundance data for which a putative taxonomic classification could not be assigned. “-Others” represents abundance data for all other taxa in the abundance data set.
praziquantel treatment, and antibiotic use (FDR > 0.05). Summary output from the analysis is shown in Table 1.

Different genera by schistosome infection status and intensity. We investigated further, via analysis of composition of microbiomes (ANCOM), how specific bacteria and fungi genera were associated with *S. haematobium* infection, while controlling for age, sex and village, followed by evaluation for association with infection intensity. In total, eight genera were identified, five from bacteria (*Pseudomonas*: *W* = 347, *Azospirillum*: *W* = 346, *Stenotrophomonas*: *W* = 292, *Dexia*: *W* = 288, and *Thalassospira*: *W* = 292) and three from fungi (*Aspergillus*: *W* = 75, *Tricholoma*: *W* = 73 and *Periglandula*: *W* = 70). The magnitude of these changes were shown by plotting the abundance of each sample to highlight differences between groups. In schistosome-positive children, the abundance of all but *Azospirillum* was higher (Fig. 3a–e). This observation was consistent with infection intensity [Pseudomonas (*r* = 0.3, *p* = 0.001), Stenotrophomonas (*r* = 0.4, *p* < 0.001), Dexia (*r* = 0.6, *p* < 0.001), Thalassospira (*r* = 0.6, *p* < 0.001) and *Azospirillum* (*r* = -0.4, *p* < 0.001)] as shown in Fig. 3f–j. Likewise, the abundance of *Aspergillus*, *Tricholoma*, and *Periglandula* was higher in schistosome-positive children (Fig. 4a–c) and was consistent with infection intensity as shown in Fig. 4d–f [Aspergillus (*r* = 0.5; *p* < 0.001), *Tricholoma* (*r* = 0.5; *p* < 0.001), and *Periglandula* (*r* = 0.4; *p* < 0.001)].

AMR gene characterisation. An average 0.06% of read pairs were mapped to AMR genes in the ResFinder database. We found evidence of 262 AMR genes, belonging to 12 functional drug class levels. AMR genes belonging to tetracycline was the most common, followed by beta-lactam, macrolide, sulfonamide and nitroimidazole. Of these, the most abundant genes were cfxA6, followed by tet(Q), tet(W), sul2, erm(F) and nimE (Fig. 5).

**Variation in the resistome and association with sample metadata.** PCA was used to initially examine variability and to identify clustering according to individual metadata. The model for the first two components explained 18.0% and 48.0% of the total variability in AMR genes and drug classes respectively. Similarly, there was no clustering according to individual metadata, reflecting high cohort diversity and the role of other factors in influencing the resistome. PCA plots and cluster dendrograms of AMR genes and their drug classes per sample is shown in Supplementary Figs. 5–6.

PERMANOVA analysis did not show any significant association of AMR genes with age, village, sex, feeding, malnutrition, stunting, *S. haematobium* infection, previous praziquantel treatment and antibiotic use on AMR genes. Model summaries of sample metadata and association with AMR genes is shown in Supplementary Data 2.

**Discussion**

Using shotgun metagenomic sequencing, we characterised the structure and composition of the human gut microbiome and resistome in this Zimbabwean preschool population (≤5 years old). Age, dietary and environmental patterns, ethnicity, and geography have a substantial impact on the taxonomic composition of the microbiome. *Prevotella* and *Candida* have been associated with carbohydrate-rich diets, and *Bacteroides* with protein-rich diets. This is a reflection of the dietary lifestyle.
Table 1 Model summaries of sample metadata and association with the gut microbiome.

| Variable                        | n   | Bacteria | Fungi |
|---------------------------------|-----|----------|-------|
|                                 |     | p value  | SSExplain | SSTotal | FDR | p value  | SSExplain | SSTotal | FDR |
| Age (years)                     | 113 | 0.004    | 1344.6   | 82733.4 | 0.024 | 0.082 | 128.6    | 9489.2  | 0.197 |
| Sex                             | 113 | 0.172    | 878.1    | 83200.0 | 0.258 | 0.439 | 82.8     | 9534.9  | 0.671 |
| Village                         | 113 | 0.012    | 1254.0   | 82824.0 | 0.039 | 0.060 | 140.4    | 9477.4  | 0.180 |
| S.h. infection status (pos/neg) | 113 | 0.013    | 1185.1   | 82892.9 | 0.039 | 0.001 | 339.0    | 9278.7  | 0.006 |
| S.h. infection intensity        | 113 | 0.001    | 1514.5   | 82563.6 | 0.012 | 0.001 | 670.1    | 8947.7  | 0.006 |
| Malnourished, yes/no (WHZ)      | 107 | 0.866    | 589.1    | 78498.2 | 0.913 | 0.830 | 59.5     | 9145.6  | 0.830 |
| Stunted, yes/no (HAZ)           | 109 | 0.407    | 754.6    | 79751.0 | 0.542 | 0.611 | 71.5     | 9227.2  | 0.671 |
| Months breastfed                | 90  | 0.082    | 954.9    | 64235.8 | 0.140 | 0.470 | 75.2     | 6985.0  | 0.671 |
| Months solid food introduced    | 102 | 0.913    | 573.1    | 75583.0 | 0.913 | 0.615 | 73.5     | 8792.0  | 0.671 |
| Previous praziquantel treatment | 105 | 0.071    | 991.9    | 77387.4 | 0.140 | 0.233 | 101.6    | 8934.5  | 0.466 |
| Amoxicillin (yes/no)            | 76  | 0.771    | 646.6    | 55603.4 | 0.913 | 0.531 | 78.4     | 6567.3  | 0.671 |
| Co-trimoxazole (yes/no)         | 76  | 0.030    | 1083.6   | 55166.4 | 0.072 | 0.048 | 158.3    | 6487.4  | 0.180 |

The table represents PERMANOVA output for bacteria and fungi genera. Classification of nutritional status was based on a cut off <−2 Z scores52. Schistosome infection intensity was log transformed (log10 [egg count + 1]). S.h. S. haematobium, WHA weight-for-height Z scores, HAZ height-for-age Z scores, pos/neg positive/negative, p value unadjusted p value, FDR adjusted p value (FDR-corrected), SSExplain explained sum of squares, SSTotal total sum of squares.

Fig. 3 Different bacteria genera between schistosome-infected children compared to uninfected children. (a–e) Box plots showing the abundance of specific bacteria genera, grouped by S. haematobium infection status. The horizontal box lines represent the first quartile, the median, and the third quartile. Whiskers denote the range of points within the first quartile −1.5× the interquartile range and the third quartile +1.5× the interquartile range. (f–j) Scatter plots showing linear regression analysis of S. haematobium infection intensity and bacteria genera abundance. The clr-transformed abundance data were used for all plots. Infection status was coded as 0 and 1 for negative (n = 95) and positive (n = 18), respectively. S. haematobium infection intensity was log transformed [log10 (egg count + 1)]. Shaded areas indicate the 95% confidence intervals.

among populations in developing countries5,17, including infants39, and these genera were among the most abundant bacteria and fungi genera found in the current study population. Similar to our previous findings35, we found age but not sex-related associations in bacteria genus diversity. Given that the population in the current study were ≤5 years old, this is consistent with the microbiome being more dynamic in the early years of life before stabilising to a more adult-like state8,12.

Differences observed in the microbiome between developing and developed countries have been attributed to factors inherent to such developing areas5, which may include the role of persistent prevalence of helminth infections, as reviewed by
in all samples, thus make a major contribution to the overall biosis and a link with increased disease risk, progression and finding is uncertain.

Children infected with Ascomycota schistosome-infected versus uninfected children were more abundant in which this association was independent of age, sex and village. This establishment of the gut microbiome in preschool-aged children whose gut microbiome structure is most likely still being established. We also expand on this to include the fungal population or were secondary to changes in the bacterial interaction and systemic interactions. For example, by enhancing the mucosal barrier, tissue repair, production of antimicrobial peptides and reducing dissemination of microbiota to the spleen and liver, the upregulation of IL-22 during helminth infection may favour the abundance of specific microbial taxa.

We identified 262 AMR genes, most of which encoded for resistance to tetracycline, beta-lactam, macrolide and sulfonamide, posing risks to successful treatment of various conditions. Our findings are consistent with observations that schistosome infection is associated with alterations in the diversity and abundance of specific taxonomic groups in the microbiome. In the aforementioned studies, which included preschool and school-aged children, 16S rRNA sequencing showed that Prevotella and Proteobacteria were more abundant in children infected with S. haematobium and S. mansoni. respectively, when compared to uninfected children. To the best of our knowledge, the novelty of the current study is the fact that this population is much younger (<5 years old), an age group whose gut microbiome structure is most likely still being established. Our study thus provides an important insight into helminth infection and its association with changes during the establishment of the gut microbiome in preschool-aged children in endemic areas. We also expand on this to include the fungal component of the microbiome. More relevant to our finding is that this association was independent of age, sex and village.

Phyla that clearly differentiated the microbiome of the schistosome-infected versus uninfected children were Proteobacteria, Ascomycota and Basidiomycota. These were among the top five most abundant bacteria or fungi phyla and were present in all samples, thus make a major contribution to the overall microbiome composition. Proteobacteria has been shown to be present in lower abundances in healthy individuals, and any increases in abundance of members of this phylum confirm dysbiosis and a link with increased disease risk, progression and burden. Attempts have been made to expand the body of knowledge on the fungi repertoire and diversity in the human microbiome, and their association with infection and disease. Studies have suggested that gut fungal populations directly or indirectly help to maintain healthy intestinal homoeostasis and that dysbiosis has immunological consequences. Increases in specific fungi populations such as Aspergillus have been associated with increased eosinophil levels, and an exaggerated Th2 response, also characteristic of schistosome infection, which may explain our observation of the association of schistosome infection with specific fungal populations (Aspergillus, Tricholoma and Periglandula). However, whether our observation was due to primary changes in the fungal population or were secondary to changes in the bacterial population is unclear. To the best of our knowledge, this is the first study to examine such an association and further studies into the role of fungal dysbiosis in schistosome infection are warranted.

Although we cannot infer causation, we are able to determine that for the genera differentiating the microbiome of the schistosome-infected versus uninfected children, there was a positive relationship between microbial abundance and schistosome infection intensity. Hence, it is possible that schistosome infection resulted in alterations in the gut microbiome. However as S. haematobium worms mostly reside in the pelvic venous plexus (although some have occasionally been detected in the intestine in Egyptian autopsies), the effect of infection on the diversity of the microbiota is as suggested for intestinal helminths, but most likely through a more indirect or systemic route than through direct interactions. Mishra et al. have suggested that the immunomodulatory effects of helminths can extend to the gut microbiota through both direct intestinal interactions and systemic interactions. For example, by enhancing the mucosal barrier, tissue repair, production of antimicrobial peptides and reducing dissemination of microbiota to the spleen and liver, the upregulation of IL-22 during helminth infection may favour the abundance of specific microbial taxa.

We identified 262 AMR genes, most of which encoded for resistance to tetracycline, beta-lactam, macrolide and sulfonamide, posing risks to successful treatment of various conditions...
including urinary, enteric and respiratory infections. It is thus not surprising that the data on antimicrobial use for the current study showed predominantly amoxicillin (beta-lactam) and co-trimoxazole (sulfonamide) use in the children. In addition, ceftriaxone and benzylpenicillin (both beta-lactams) and co-trimoxazole are among the most commonly used antibiotics in Zimbabwe.

Increased antimicrobial use impacts the gut microbiota and is selective for AMR in populations. The limited association of the obtainable antimicrobial use data with both the microbiome and resistome in the current study might be surprising. However, a study by Dethlefsen et al. showed that a majority of the bacterial community that was depleted post-ciprofloxacin administration was restored after 4 weeks. Our antimicrobial use data were limited to antibiotic use within the immediate 6 months prior to sampling and was less heterogeneous, thus any marked differences in the microbiota may have been missed. Our findings are consistent with those from recent studies on global sewage samples, which have shown a much stronger association between socio-economic factors related to health, sanitation and education with the resistome, compared to antimicrobial use. This seems more likely to be the case in low- and middle-income countries, where a high contagion—the spread of resistant strains and genes—between individuals may take place, and any antimicrobial use in one individual may have general effects on the population as a whole. Thus, antimicrobial use explains some, but not all variation in AMR genes in this population, and improving such socio-economic factors may improve AMR.

Our study had a few limitations. The cross-sectional study design allowed characterisation of the gut microbiome and its relationship with S. haematobium infection at a single time point. A longitudinal study will inform on the dynamic relationship between the two, as well as the time course or developmental-related trends in the observed profiles. A longitudinal study would also give an indication of the dynamic features of the AMR genes and how they are associated with individual-specific data. Furthermore, relating the presence of AMR genes to measurable phenotypic resistance of bacteria would give a stronger indication of the clinical implications of the AMR genes present.

In conclusion, we characterised, through shotgun metagenomic sequencing, the microbiome (to include the fungi repertoire) and resistome in a preschool population (<5 years old) in Zimbabwe. We identified differences in the gut microbiome between schistosome infected and uninfected children, showing largely an increase in abundance of specific bacteria, and for the first time (to our knowledge), fungi genera in infected children. This association was independent of age, sex and village. Mechanistic studies are required to further explain this relationship. To the best of our knowledge, we also characterised for the first time in this African preschool population, a diversity of AMR genes to bacteria, belonging to various functional drug classes. Our microbiome and resistome data add to publicly available data from different human populations.

**Methods**

**Ethical approval and consent.** The current study is part of a larger paediatric schistosomiasis study, for which ethical and institutional approval was obtained from the Medical Research Council of Zimbabwe (MRCZ/A/1964) and the
University of Edinburgh (Inmutapi-0002), respectively. Permission to conduct the study was obtained from the Mashonaland Central Provincial Medical Director. The study aims and procedures were explained to all participants and their parents/guardians in their local language, Shona. Written informed consent was obtained from the participants’ parents/guardians and recruitment was voluntary with participants free to withdraw from the study at any stage. For the current study, only samples from participants who consented to be part of this sub-study were used.

Study design, population and site. This cross-sectional study forms part of the baseline survey of a larger study on paediatric schistosomiasis conducted in the Shamva district, Northeast Zimbabwe, to compare the health benefits of early treatment of schistosome infections in preschool-aged children, 6 months to 5 years old27. In brief, children within this age group who were lifelong residents, and with no history of tuberculosis, malaria/fever, or recent major illness/surgery were invited to participate in the baseline survey and onward recruitment into the main cohort, if they met further criteria of being negative for S. haematobium by egg count and had no history of antihelminthic treatment. The study followed two groups of schistosoma-negative children for new infections and then compared re-infection rates across two different regimens, following treatment of first schistosome infection. For the current study, a subset of 116 stool samples (from 1–5 year olds) were selected from the baseline survey for microbiome analysis. To be recruited as part of this subset, participants had to consent for their stool samples to be used as part of the current study. A questionnaire was administered at the time of recruitment to gather additional demographic, growth and nutrition, and a clinical history. Clinical records were checked to obtain history of antibiotic use, within the 6 months preceding acquisition of stool samples. Parents/carers were also interviewed to ascertain the health history of the children, and those who had any such history were excluded. Children were tested via parasitological diagnosis for S. mansoni, S. haematobium, and Schistosomiasis (SH). A part of the baseline survey and none who were positive were included in the subset of 116 children in the current study. The subset of children was thus selected based on (a) consent for microbiome analysis, (b) availability of socio-demographic data, (c) availability of parasitology samples (urine and stool samples), (d) availability of test results and clinical history and (e) the current episode of diarrhoea (assessed by questionnaire and visual stool examination).

Samples included in the current study were from two main sites, Mupfure and Chihuri. The sites are located in the Mashonaland Central province, with 125,650 people living in a 99% rural area of 2695 km², according to the 2012 national census76. The inhabitants are primarily subsistence farmers. The area has a high prevalence of S. haematobium (>50%) with low prevalence of S. mansoni and STH (<15%)77, making it ideal for studies on the impacts of urogenital schistosomiasis.

Sample size. The samples used in the current study are from the baseline survey of a larger epidemiological study comparing re-infection rates across two different treatment regimens. As this is a relatively new field in human helminthology, there are limited published studies, with none focusing on the age group in the current study, i.e. 1–5 year olds. Thus there were no published baseline data to inform sample size calculations when we conducted our study. The sample size for the current study was informed by our previous study35 and those of others49,29 in older children with sample sizes ranging from 34–139, from which statistically significant differences were detected in the microbiome of schistosome infected versus uninfected children.

Sample collection, processing and DNA extraction. Urine and stool specimens were collected for parasitological diagnosis of S. haematobium and intestinal helminths, respectively. In summary, about 50 ml of urine sample was collected on three successive days, and a single stool sample was collected on a single day from each participant. Urine bags (Hollister 7511 U-Bag Urine Specimen Collector, Hollister Inc., Chicago, IL, USA) and disposable diapers (for stool samples) were used for sample collection in very young children. Urine samples were examined microscopically for S. haematobium infection following the standard urine filtration method27, and stool samples examined microscopically using the Kato–Katz method26, to exclude S. mansoni and STH (for at least one parasite egg detected). Infection intensity for S. haematobium was defined as the arithmetic mean egg count/10 ml of at least two urine samples collected on three consecutive days. All children who were positive for schistosome infection were treated with a single dose of praziquantel at the standard 40 mg/kg body weight, crushed and administered with squash and sliced bread28 by local nurses.

Aliquots of stool samples in 2 ml cryotubes were stored temporarily at 2–8 °C for a maximum of 24 h prior to processing. For each stool sample, DNA was extracted using the QIAamp DNA Stool Mini Kit (QIAGEN) according to the manufacturer’s instructions. To ensure sample aliquots contained purified DNA, each sample was quantified in-house at the University of Edinburgh using the Qubit fluorometer (ThermoFisher Scientific) prior to shipment for DNA sequencing.

Library preparation and sequencing. DNA samples were shipped on dry ice for library preparation and sequencing at the Beijing Genomics Institute (BGI, Shenzhen, China). At BGI, DNA from the stool samples was quantified using the Qubit fluorometer (ThermoFisher Scientific) and the NanoDrop™ spectrophotometer (ThermoFisher Scientific). As a quality control measure, the integrity of the DNA was assessed using the Agilent 2100 bioanalyzer. DNA was amplified using the QIAquick PCR Purification Kit (Qiagen). Adapter-ligated DNA samples were separated by electrophoresis through a 2% agarose gel to recover the target fragments, and purified using the QIAquick Gel Extraction Kit (Qiagen). Library preparation to enrich the adapter-ligated DNA was done via PCR amplification, size-separated by electrophoresis, and purified using the QIAquick Gel Extraction kit (Qiagen). The final library was quantified using the Agilent 2100 bioanalyzer. The qualifying 116 DNA libraries were amplified using the BOLT system (Illumina), and sequenced on the Illumina HiSeq 4000 platform (Illumina) using paired-end 150 bp sequencing.

Bioinformatics processing: quality control and trimming. Raw FASTQ format sequences from each sample were quality assessed using FASTQC v0.10.0 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The number of read pairs generated per sample ranged from 9,263,538 to 21,350,613 [Supplementary Data 1]. Subsequently, reads were trimmed, to include removing adapters, using BBDuk [BBMap–Bushnell B – https://sourceforge.net/projects/bbmap/] with an output quality Phred threshold score of ≥20 and a minimum read length of 50 bp. K-mer length for finding contaminants was k = 19. We looked for shorter k-mers at read pairs down to k = 11, and reads were trimmed at the right end [Supplementary Data 1].

Bioinformatics processing: mapping of sequence reads. Using a novel reference based mapping and alignment tool, KMA28, the trimmed reads were used as input to align directly against reference sequence databases. The KMA method is designed to improve mapping against redundant databases, and has been shown to outperform existing mapping methods in terms of speed, precision and sensitivity80. In summary, KMA, employs heuristic mapping, which involves directly mapping k-mers between query sequences and selected template databases, including large redundant databases. It then speeds up mapping by using k-mer seeding, and utilises a special version of the Needleman–Wunsch algorithm81 to accurately align regions of matching k-mers. To ensure the best match template for the query reads, multi mapping reads are resolved using a novel sorting scheme, ConClave. The scheme enables assembly of reads which results in a final accurate consensus sequence for the reference sequence, and also rules out bias associated with base calling across different sequencing platforms82.

Bioinformatics processing: microbiome sequence component. To access the microbiome sequence component present in our samples, read pairs and singletons were aligned to a custom reference genomic database (last updated 04.04.2019). Mapped reads were counted as one copy, in cases of read pairs or singletons. Unless otherwise specified below, databases were primarily downloaded via NCBI GenBank: clade specific assembly_summary.txt files (ftp://ncbi.nlm.nih.gov/genomes/genbank). The custom database consisted of the following: bacteria (closed genomes; downloaded 05.02.2019), archaea (downloaded 13.02.2019), MetaHitAssembly (PRIEB674–PRIEB1046; downloaded 01.07.2014), HumanMicrobiome (genomics assemblies downloaded 02.07.2014; downloaded 05.02.2019), plasmid (downloaded 05.02.2019), virus (https://bitbucket.org/genericmicrobiome/kvit_db; downloaded 05.02.2019; https://genom.jgi.doe.gov/pipeline/pages/dynamicOrganismDownload.jsp?organism=IMG_VR; down- loaded 08.01.2019.), fungi (downloaded 13.02.2019), parasites (downloaded 04.04.2019). Sequences selected for the bacteria and bacteria_draft databases from the assembly_summary.txt file were annotated with the tags version_status = "latest" and genome_rep = "Full". Additional entries, assembly_level = "Complete genome" or "Chromosome" in the bacteria database and refseq_category = "representative genome" in the bacteri- a_draft database were also required. The plasmid database was constructed as a subset of the bacteria and bacteria_draft databases; keyword in the FASTA entry header line, "plasmid". The total read count for each microbial community of interest in a sample was calculated as the sum of read counts from each of the databases of interest; bacteria (bacteria, bacteria_draft, MetaHitAssembly, and HumanMicrobiome), fungi, protozoa, and parasites. Sequence mapping statistics are shown in Supplementary Data 1.

The primary (most similar) alignment obtained for mapped sequences was used to assign a putative taxonomy, based on the taxonID obtained. TaxonID’s and associated taxonomy classifications were obtained from downloaded reference microbial genomes from NCBI (ftp://ftp.ncbi.nih.gov/pub/taxonomy/taxdump.tar.gz) and assignment at all taxonomic levels was done. Sequences that had no similarities detected in the nucleotide (nt) database for which we could not assign a putative taxonomic classification were deemed to be unknown sequences, and hence termed “Unknown”. The classification of “unknown” is exquisitely time–sensitive and was appropriate and correct at the time of this analysis. Details of taxonID mapping are shown in Supplementary Data 1.

Bioinformatics processing: AMR gene component. To identify any putative AMR genes present in the samples, the read pairs were aligned to AMR genes (3081
genes) present in the ResFinder database (https://bitbucket.org/genomicsequencing/resfinder_db; downloaded 13.02.2019) with parameters set for a query gene to cover at least 2/5 the length of the reference gene to be selected. Alignments were filtered to retain those exhibiting a selected threshold of identity of 90% (i.e. >90% nucleotide identity between the query and reference gene over at least 90% of the length of reference gene). Sequence mapping statistics are shown in Supplementary Data 1.

Data handling and processing. To account for probable sample-wise sequencing depth differences, as well as a size-dependent probability of observing a reference, mapping counts from the custom genomic database and from the ResFinder database were normalised to the total genome sizes for the genomic database and to the individual gene lengths for the ResFinder database (gene and genome size details in Supplementary Data 1).

The total observed mapping counts are relative, and may account for confounding effects on downstream analyses. This may be due to limitations of an arbitrary total imposed by different sequencing platforms, technical variations in sequencing libraries amounts, or even random variation. To obtain information about the abundances of features in our data set relative to each other, datasets were treated as compositional. Data were transformed using the log-ratio approach as introduced by Aitchison, to make the data symmetric, linear and in a log-ratio coordinate space. However compositional methods such as this do not account for the presence of zeros associated with abundance datasets. To address this, a small pseudocount of half the smallest abundance per feature was added to each respective feature for all the normalised abundance matrices, prior to transformations.

Microbiota abundance data tables with counts, x, and k number of populations (taxa members), were centred log ratio (clr) transformed, defined as:

\[ \text{clr}(x_1, \ldots, x_k) = \log \left( \frac{x_1}{\bar{x}} \right) - \log \left( \frac{x_2}{\bar{x}} \right) - \cdots - \log \left( \frac{x_k}{\bar{x}} \right), \]

where, \( x_i = \left( \frac{1}{k} x_i \right)^{1/3} \) is the geometric mean of the particular composition.

AMR gene abundances were additive log-ratio (alr) transformed, taking the bacterial component of the microbiome (\( x_i \)) as the reference as:

\[ \text{alr}(x_1, \ldots, x_k) = \log \left( \frac{x_1}{\bar{x}} \right) - \log \left( \frac{x_2}{\bar{x}} \right) - \cdots - \log \left( \frac{x_k}{\bar{x}} \right). \]

Unless otherwise stated, clr and alr matrices were used for all downstream analyses. Raw mapping count data and their corresponding clr and alr values for the analysed samples can be found in Supplementary Data 2.

Visualisation. Data visualisation was performed within the R environment. The updated R codes used for analysis of composition of microbiomes (ANCOM) are available on https://github.com/zellerlab/crc_meta/blob/master/src/ANCOM_updated.R OR from the author’s webpage at https://sites.google.com/site/siddharthamandali1988/research.

Received: 1 November 2019; Accepted: 25 February 2020; Published online: 02 April 2020

References
1. Lukes, J., Stensvold, C. R., Jirku-Pomajbikova, K. & Wegener Parfrey, L. Are human intestinal eukaryotes beneficial or commensals? PloS Pathog. 11, e1005039 (2015).
2. Human Microbiome Project Consortium. Structure, function and diversity of the human gut microbiome. Nature 486, 207–214 (2012).
3. Turnbaugh, P. J. et al. The human microbiome project. Nature 449, 804–810 (2007).
4. Brewster, R. et al. Surveying gut microbiome research in Africans: toward improved diversity and representation. Trends Microbiol. 27, 824–835 (2019).
5. De Filippo, C. et al. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. Proc. Natl Acad. Sci. USA 107, 14691–14696 (2010).
6. Rampelli, S. et al. Metagenomics of the Hadza hunter-gatherer gut microbiota. Curr. Biol. 25, 1682–1693 (2015).
7. Schirru, S. L. et al. Gut microbiome of the Hadza hunter-gatherers. Nat. Commun. 5, 3654 (2014).
8. Yatsunenko, T. et al. Human gut microbiome viewed across age and geography. Nature 486, 222–227 (2012).
9. Robertson, R. C., Manges, A. R., Finlay, B. B. & Prendergast, A. J. The human microbiome and child growth—first 1000 days and beyond. Trends Microbiol. 27, 131–147 (2019).
10. International HapMap Project. The International HapMap Project. Nature 426, 789–796 (2003).
11. Mulder, N. et al. H3Africa: current perspectives. Pharrmagenomics Pers. Med. 11, 59–66 (2018).
12. Rodriguez, J. M. et al. The composition of the gut microbiota throughout life, with an emphasis on early life. Micro. Ecol. Health Dis. 26, 26050 (2015).
13. Fan, W., Hua, G., Li, X., Yang, L. & Duan, C. Impact of diet in shaping gut microbiota revealed by a comparative study in infants during the six months of life. J. Microbiol. Biol. Technol. 24, 133–143 (2014).
14. Senghor, B., Sokhna, C., Ruymi, R. & Lagier, J.-C. Gut microbiota diversity according to dietary habits and geographical provenance. *Hum. Microbiome J.* 7, 1-9 (2018).

15. Goodrich, J. K. et al. Human genetics shape the gut microbiome. *Cell* 159, 789–799 (2014).

16. Dominguez-Bello, M. G. et al. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc. Natl Acad. Sci. USA* 107, 11971–11975 (2010).

17. Martinez, I. et al. The gut microbiota of rural papua new guineans: composition, diversity patterns, and ecological processes. *Cell Rep.* 11, 527–538 (2015).

18. Mishra, P., Palma, M., Bleich, D., Loke, P. & Gause, W. Systemic impact of intestinal helminth infections. *Mucosal Immunol.* 7, 753 (2014).

19. Bustinduy, A. L. et al. Impact of polyplasystatic infections on anemia and undernutrition among Kenyan children living in a *Schistosoma haematobium*-endemic area. *Am. J. Trop. Med. Hyg.* 88, 433–440 (2013).

20. Mañana, C. F., Eko, U. F. & Ojo, D. A. Urinary schistosomiasis in preschool children in settlements around Oyan Reservoir in Ogun State, Nigeria: implications for control. *Trop. Med. Int. Health* 8, 78–82 (2003).

21. Bosompem, K. M. et al. Infant schistosomiasis in Ghana: a survey in an irrigation community. *Trop. Med. Int. Health* 9, 917–922 (2004).

22. McManus, D. P. et al. Schistosomiasis. *Nat. Rev. Dis. Prim.* 4, 13 (2018).

23. World Health Organization. *Schistosomiasis*. *WHO/Malaria/2015.15*

24. Pearce, E. J. & MacDonald, A. S. The immunobiology of schistosomiasis. *Am. J. Trop. Med. Hyg.* 78, 538 (2008).

25. McManus, D. P. et al. Schistosomiasis. *Trop. Med. Int. Health* 15, 1197–1205 (2010).

26. Martens, J. K. et al. Changes in the gut microbiota during antibiotic treatment on the fecal microbiota. *PLoS Negl. Trop. Dis.* 10, e0003681 (2015).

27. Jenkins, T. P. et al. Differential human gut microbiome assemblages during soil-transmitted helmint infections in Indonesia and Liberia. *Microbiome* 6, 33 (2018).

28. Kay, G. L. et al. Differences in the faecal microbiota in *Schistosoma mansoni*-infected children vs. uninfected children. *PLoS Negl. Trop. Dis.* 9, e0003861 (2015).

29. Jenkins, T. P. et al. *Schistosoma mansoni* infection is associated with quantitative and qualitative modifications of the mammalian intestinal microbiota. *Sci Rep.* 8, 16 (2018).

30. Cheever, A. W., Kamei, I. A., Elwi, A. M., Mosimann, J. E. & Danner, R. *Schistosoma mansoni* and *S. haematobium* infections in Egypt. II. Quantitative parasitological findings at necropsy. *Am. J. Trop. Med. Hyg.* 26, 702–716 (1977).

31. Scher, J. U. et al. Expansion of intestinal *Prevotella copri* correlates with enhanced susceptibility to arthritis. *Elife* 2, e01202 (2013).

32. Ajibola, O. et al. Urogenital schistosomiasis is associated with signatures of microbiome dysbiosis in Nigerian adolescents. *Sci. Rep.* 9, 10 (2019).

33. Barnhill, A. E., Novozhilova, E., Day, T. A. & Carlson, S. A. Schistosoma-associated *Salmonella* resist antibiotics via specific fimbrial attachments to the flatworm. *Parasit. Vectors* 4, 123 (2011).

34. LoVerde, P. T., Amento, C. & Higashi, G. I. Parasite-parasite interaction of *Schistosoma mansoni* and *Fasciolopsis buski*. *J. Parasitol.* 92, 141, 177–185 (1980).

35. Qin, J. et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464, 59–66 (2010).

36. Hu, Y. et al. Metagene expression analysis of antibiotic resistance genes in a large cohort of human gut microbiota. *Nat. Commun.* 4, 2151 (2013).

37. Fernberg, C., Lofmark, S., Edlund, C. & Jansson, J. K. Long-term impacts of antibiotic exposure on the human intestinal microbiota. *Microbiology* 156, 3216–3233 (2010).

38. De Waele, J. J. et al. Antimicrobial resistance and antibiotic stewardship programs in the ICU: inclusion and persistence in the fight against resistance. *Am. J. Infect. Control* 48, 550–581 (2020).

39. Doron, S. & Davidson, L. E. Antimicrobial stewardship. *Mayo Clin. Proc.* 86, 1113–1123 (2011).

40. Blair, J. M., Webber, M. A., Baylay, A. J., Ogullo, D. O. & Piddock, L. J. Molecular mechanisms of antibiotic resistance. *Nat. Rev. Microbiol.* 13, 42–51 (2015).

41. Holmes, A. H. et al. Understanding the mechanisms and drivers of antimicrobial resistance. *Lancet* 387, 176–187 (2016).

42. Woolhouse, M. E. & Ward, M. J. Microbiology. Sources of antimicrobial resistance. *Science* 341, 1460–1461 (2013).

43. Deschauza, M. et al. Deposition of the gut microbiota in a population with varied ethnic origins but shared geography. *Nat. Med.* 24, 1526–1531 (2018).

44. Gorovitoksai, A., Holmes, S. P. & Huse, S. M. Interpreting *Prevotella* and *Bacteroides* as biomarkers of diet and lifestyle. *Microbiome* 4, 15 (2016).

45. Hoffmann, C. et al. Archaea and fungi of the human gut microbiome: correlations with diet and bacterial resdients. *PLoS ONE* 8, e66019 (2013).

46. Yassour, M. et al. Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability. *Sci. Transl. Med.* 8, 343ra381 (2016).

47. Shin, N. R., Whon, T. W. & Bae, J. W. *Protothecobacteria*: microbial signature of dysbiosis in gut microbiota. *Trends Biotechnol.* 33, 496–503 (2015).

48. Gouba, N., Raoult, D. & Drancourt, M. Eurkaryote culturomics of the gut reveals new species. *PLoS ONE* 9, e1003996 (2014).

49. Gouba, N. & Drancourt, M. Digestive tract mycobiota: a source of infection. *Med. Mal. Infect.* 4, 9–16 (2015).

50. Mukherjee, P. K. et al. Oral mycobiome analysis of HIV-infected patients: identification of *Pichia* as an antagonist of opportunistic fungi. *PLoS Pathog.* 10, e1003996 (2014).

51. Mar Rodriguez, M. et al. Obesity changes the gut human mycobiome. *Sci. Rep.* 5, 14600 (2015).

52. Wheeler, M. L. et al. Immunologic consequences of intestinal fungal dysbiosis. *Cell Host Microbe* 19, 865–873 (2016).

53. Buleksiene, V., Baltruikei, D. & Repčkienė, J. Study of health risk associated with *Aspergillus amstelodami* and its mycotoxic effects. *Ekologija* 3, 42–47 (2006).

54. Jung, L. M. & Loke, P. A role for IL-22 in the relationship between intestinal helmintus, gut microbiota and mucosal immunity. *Int. J. Parasitol.* 43, 253–257 (2013).

55. Frank, T., Gautier, V., Talarmin, A., Bercion, R. & Arlet, G. Characterization of sulphomamide resistance genes and class 1 integron gene cassettes in Enterobacteriaceae, Central African Republic (CAR). *J. Antimicrob. Chemother.* 59, 742–745 (2007).

56. Zimbabwe National Antimicrobial Resistance Core Group. *Situation Analysis of Antimicrobial Use and Resistance in Humans and Animals in Zimbabwe* (Global Antimicrobial Resistance Partnership (GARP), Action on ANtiBiotic Resistance (REACT), Zimbabwe, 2017).

57. Dethlefsen, L., Huse, S., Sogin, M. L. & Relman, D. A. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biol.* 6, e220 (2008).

58. Bartosch, S., Fite, A., Macfarlane, G. T. & McMurdo, M. E. Characterization of bacterial communities in feces from healthy elderly volunteers and hospitalized elderly patients by using real-time PCR and effects of antibiotic treatment on the fecal microbiota. *Appl. Environ. Microbiol.* 70, 3575–3581 (2004).

59. Davies, J. & Davies, D. Origins and evolution of antibiotic resistance. *Microbiol. Mol. Biol. Rev.* 74, 417–433 (2010).
73. Van De Sande-Bruinsma, N. et al. Antimicrobial drug use and resistance in Europe. Emerg. Infect. Dis. 14, 1722 (2008).
74. Collignon, P., Beggs, J. J., Walsh, T. R., Gandra, S. & Laxminarayan, R. Anthropological and socioeconomic factors contributing to global antimicrobial resistance: a univariate and multivariable analysis. Lancet Planet Health 2, e398–e405 (2018).
75. Pehrson, E. C. et al. Interconnected microbiomes and resistomes in low-income human habitats. Nature 533, 212–216 (2016).
76. Zimbabwe National Statistics Agency. Zimbabwe Population Census 2012, http://www.zimstat.co.zw/sites/default/files/img/publications/Population_National_Report.pdf (2012).
77. Midzi, N. et al. Distribution of schistosomiasis and soil transmitted helminthiasis in Zimbabwe: towards a national plan of action for control and elimination. PLoS Negl. Trop. Dis. 8, e3014 (2014).
78. Mott, K. E., Baltes, R., Rambaghia, J. & Baldassini, B. Field studies of a reusable polyamide filter for detection of Schistosoma haematobium eggs by urine filtration. Tropened. Parasitol. 33, 227–228 (1982).
79. Katz, N., Chaves, A. & Pellegrino, J. A simple device for quantitative stool thick-smear technique in Schistosomiasis mansoni. Rev. Inst. Med. Trop. Sao Paulo 14, 397–400 (1972).
80. Clausen, P., Aarestrup, F. M. & Lund, O. Rapid and precise alignment of raw reads against redundant databases with KMA. BMC Bioinformatics 19, 307 (2018).
81. Needleman, S. B. & Wunsch, C. D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. 48, 443–453 (1970).
82. Kaas, R. S., Leekitcharoenphon, P., Aarestrup, F. M. & Lund, O. Solving the problem of comparing whole bacterial genomes across different sequencing platforms. PLoS ONE 9, e104984 (2014).
83. Zankari, E. et al. Identification of acquired antimicrobial resistance genes. J. Antimicrob. Chemother. 67, 2640–2644 (2012).
84. McMurdie, P. J. & Holmes, S. Waste not, want not: why rarefying microbiome datasets is inadmissible. PLoS Comput. Biol. 10, e1003531 (2014).
85. Gloor, G. B., Macklim, J. M., Pawlowsky-ÍGlahn, V. & Egozcue, J. J. Microbiome datasets are compositional: and this is not optional. Front. Microbiol. 8, 2224 (2017).
86. Aitchison, J. The Statistical Analysis of Compositional Data (Chapman and Hall, 1986).
87. Hall, C. M. Statistical analysis of metagenomics data. Genomics Inf. 17, e6 (2019).
88. Arfken, G. in Mathematical Methods for Physicists (ed George Arfken) 229–237 (Academic Press, 1985).
89. Le Cao, K. A. et al. MixMC: a multivariate statistical framework to gain insight into microbial communities. PLoS ONE 11, e0160169 (2016).
90. Oksanen, J. et al. Community Ecology Package, https://CRAN.R-project.org/package=vegan (2016).
91. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. B 57, 289–300 (1995).
92. Mandal, S. et al. Analysis of composition of microbiomes: a novel method for studying microbial composition. Microb. Ecol. Health Dis. 26, 27663 (2015).

Acknowledgements

We thank the local nurses, village health workers and community nurses for their help with the fieldwork. Special thanks to all study participants and their parents/guardians. We also thank members of the Understanding Bilharzia project in Zimbabwe for their technical help, and all the members of the Parasite Immuno-epidemiology Group at the University of Edinburgh for their useful comments in shaping this manuscript. Our research is supported by the Thrasher Research Fund 12440, Wellcome Trust 108061/Z/15/Z, and the Oak Foundation. This research was commissioned in part by the National Institute for Health Research (NIHR) Global Health Research programme (16/136/33) using UK aid from the UK Government. The views expressed in this publication are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. D.N.M.O. is supported by the Darwin Trust of Edinburgh. C.B., T.N.P., P.M., and F.M.A. are funded by the Novo Nordisk Foundation (NNF16OC0021856: Global Surveillance of Antimicrobial Resistance).

Author contributions

D.N.M.O., T.M., F.M.A., M.E.J.W. and F.M. conceptualised and designed the study. D.N.M.O., T.M., T.C., S.A.A., J.M. and F.M. were involved in the fieldwork, sample collection and DNA extractions. D.N.M.O. and T.C., curated the field data. D.N.M.O., J.M. and F.M. organised the sequencing. T.N.P. and A.I. carried out the bioinformatics processing. D.N.M.O., P.M., T.N.P., C.B. and A.I. analysed, produced figures and interpreted the data. D.N.M.O. prepared the draft manuscript and all authors were involved in review, editing and approval of the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s42003-020-0859-7.
Correspondence and requests for materials should be addressed to D.N.M.O.
Reprints and permission information is available at http://www.nature.com/reprints
Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.