Why do late- stage parasitic disease control measures fail: leveraging whole-genome sequencing to understand factors that contributed to schistosomiasis reemergence in China

Hannah D. Guss, Yannick Z. Francioli, Elise Grover, Wei Zou, Andrew Hill, Kristen J. Wade, Hamish Pike, Yang Liu, Bo Zhong, & Todd Castoe

Abstract:

Schistosomiasis is a neglected tropical disease caused by schistosome blood flukes that affects over 200 million worldwide(Weerakoon et al. 2015) . Global efforts to eradicate the disease have had substantial success in some endemic regions, while other regions lag in progress. Accordingly, regions of the world that have made progress towards elimination may provide key insight into the paths to success, and distinct challenges of eradicating the disease. China has previously achieved near elimination levels through intensive control regimes including mass administration of the anthelminthic drug Praziquantel yet continued to experience re-emergence of schistosomiasis as recently as the early 2000s(Wang et al. 2009). The factors that have enabled persistence and re-emergence despite otherwise low prevalence remain poorly understood. To address these questions, we analyze whole genome sequences of 270 archival schistosome miracidial genomes sampled from 63 human hosts across a network of villages in the mountainous region of China in 2007 immediately following reemergence. We use these data to estimate parasite population demography, genomic diversity, patterns of parasite relatedness, and worm burden per host to understand factors that may contribute to regional persistence and continued transmission.

Approach:

- To understand the factors driving the re-emergence and expansion of schistosomiasis in Sichuan Province despite strong control measures and historically low prevalence, we conducted whole-genome sequencing of hundreds of miracidia.
- Leveraging our high-resolution genomic data, we reconstructed parasite population structure, effective population sizes, and genomic diversity to assess the impacts of intensive control measures.
- We used these data to address several key questions:
    - How many distinct parasite lineages contributed to re-emergence, and what does this imply about the number of independent introduction events?
    - How do parasite population genomic patterns align with expectations for a low-prevalence, controlled region versus scenarios of alternative mammalian host reservoirs sustaining transmission?
    - How have control measures and associated bottlenecks influenced parasite inbreeding, genomic variation, and selection?
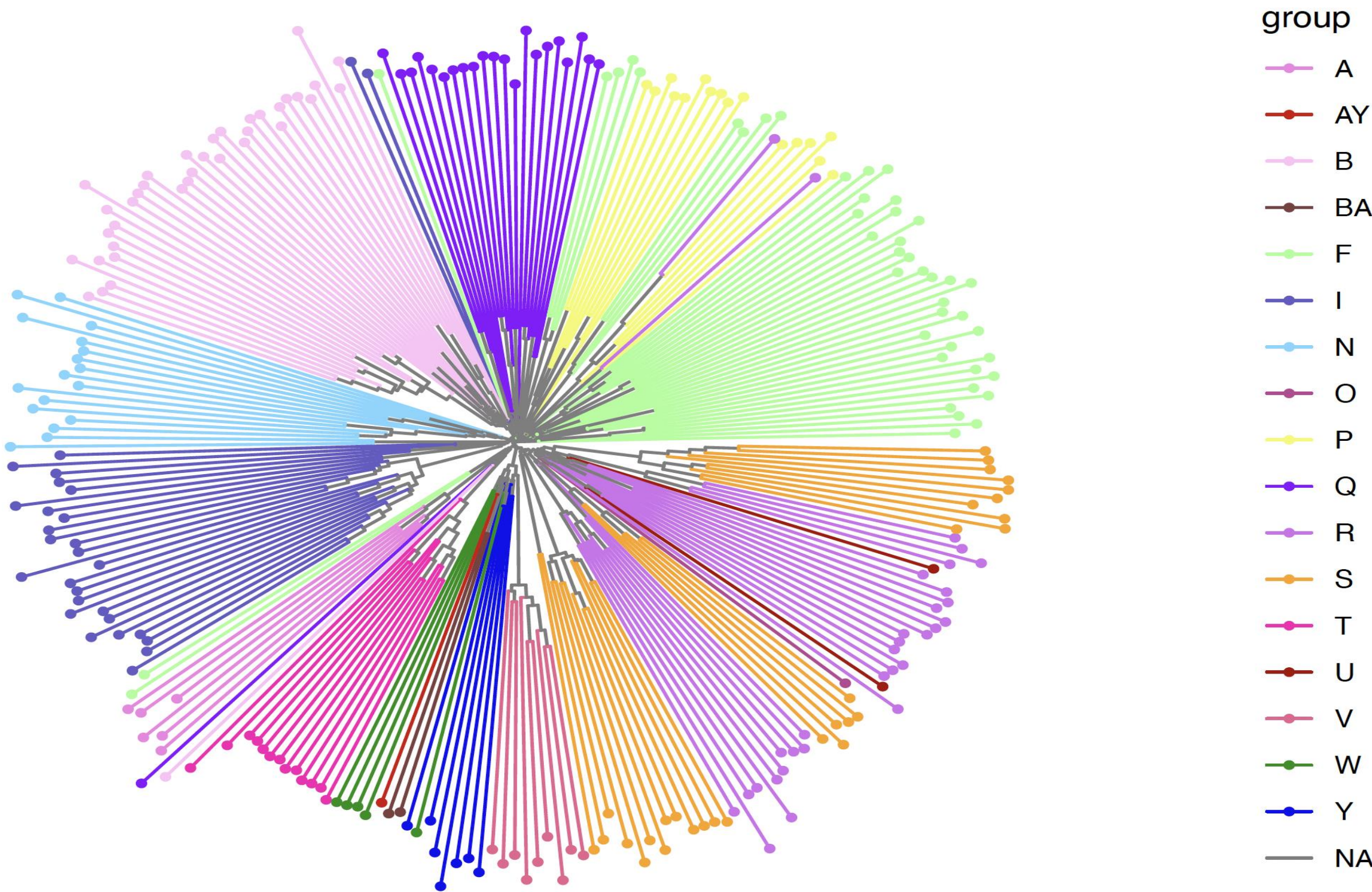
Neighbor Joining Tree: Figure1: A Neighbor joining tree of sampled miracidia colored by village.

Methods:

- In total, 270 whole-genome-amplified miracidia samples from 2007 were successfully sequenced at >20x genome coverage and included in further analyses.

- These 270 samples represented a total of 52 human hosts from 17 villages in Sichuan Province, China. Across these samples, the average paired-end sequencing reads per sample was 75 million, the mean depth of coverage was 37x, and the percent of reads that mapped to the reference *S. japonicum* genome (Luo et al. 2022) was 95%.

Relatedness Network Colored by Village and Edge Distance
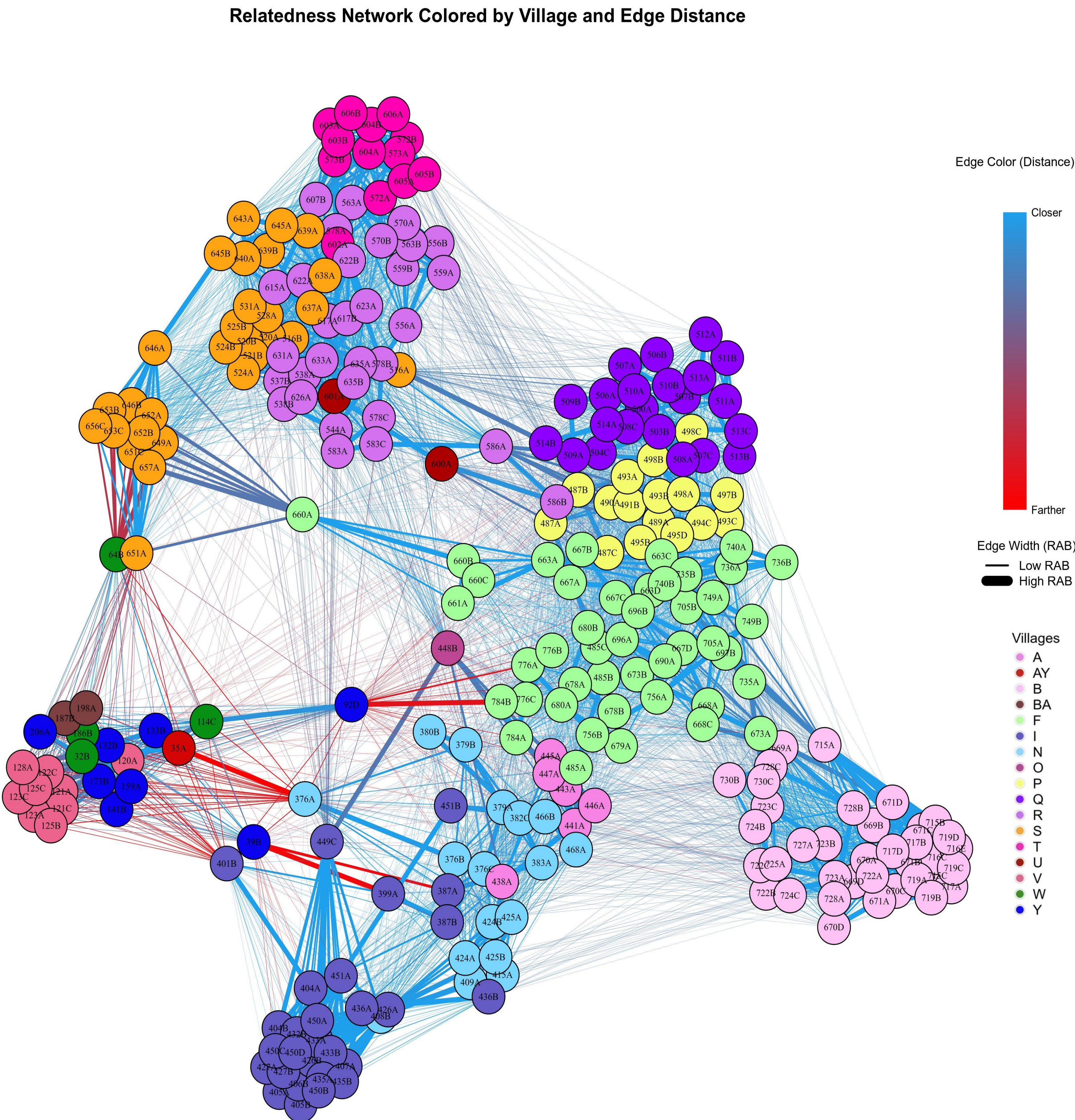
Networks of relatedness: Figure2: A network built on pairwise relatedness estimates(RAB). Each circle is a miracid, which is colored by their corresponding village. The width of the connective lines is representative of the RAB value, with the RAB cutoff between high and low Rab being 0.02, and the color of the line represents the distance between the village the sample belongs too.

Results:

- NGSrelate-based relatedness analysis uncovers patterns of genetic connectivity, showing that village V has the highest mean relatedness (0.589), suggesting a more isolated lineage, while village F, with the lowest relatedness (0.039), may represent a more genetically mixed cluster. Heatmap visualizations further reveal that closely situated villages tend to cluster together, while long-range relatedness links indicate possible historical or recent gene flow between distant populations(Fig2).

- Parasite populations are highly structured within villages, but some inter-village transmission is evident. High relatedness between parasites from different villages and co-infected hosts with multiple lineages suggest transmission hubs enabling parasite spread.

Results:

- Phylogenetic analysis demonstrates village-based clustering, where samples from the same village tend to group together, and geographically close villages also show clustering patterns. However, some villages exhibit a more dispersed distribution, suggesting a more complex genetic relationship(Fig.1).

- Relatedness estimates show that parasites from the same village tend to be highly genetically similar, indicating that transmission is primarily occurring at a micro-scale within villages. This highlights the importance of household- and village-level interventions in breaking transmission cycles.
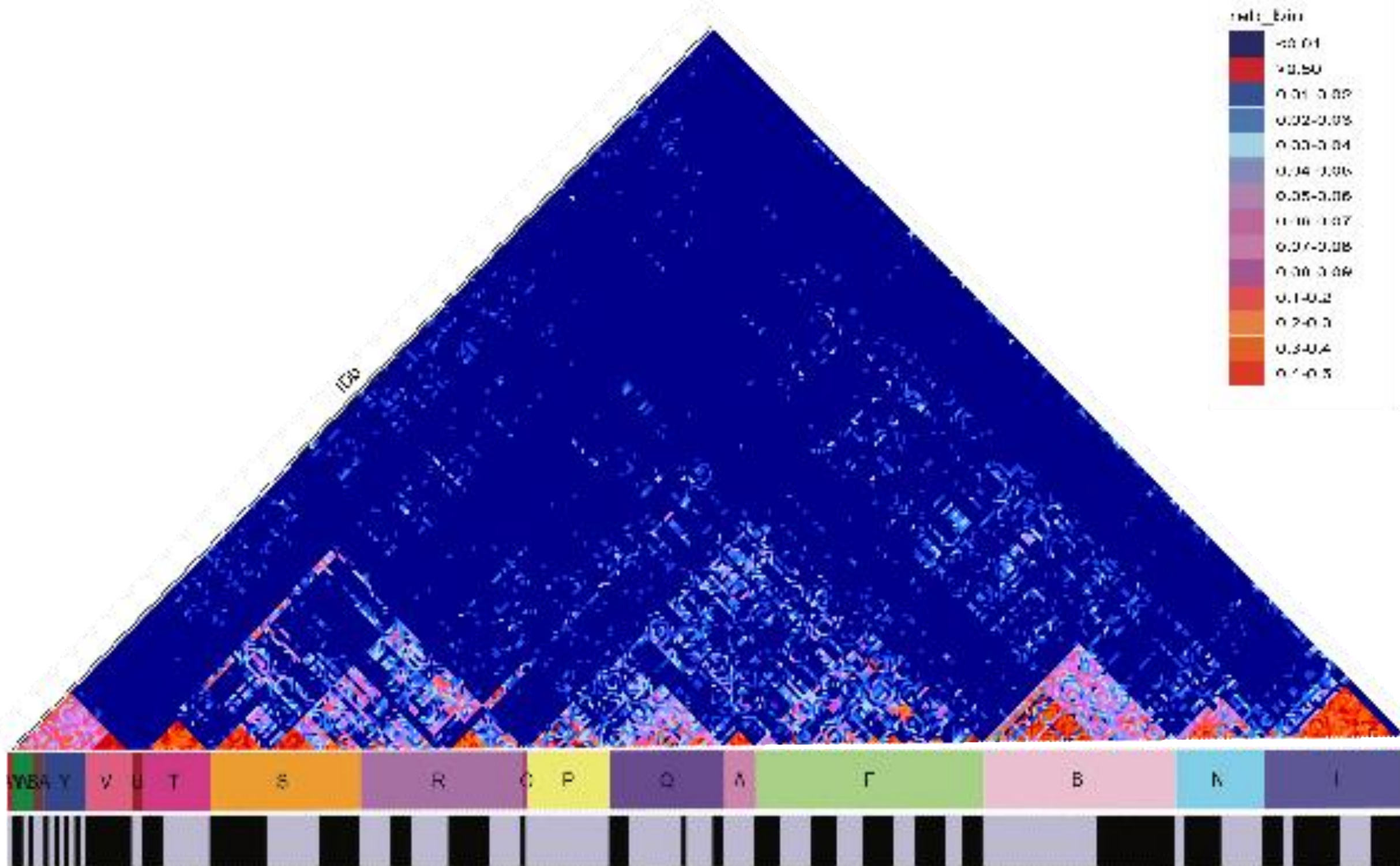
Relatedness Heatmap: Figure3: Triangle Heatmap of pairwise allele-sharing(RAB) between *Schistosoma japonicum* miracidia samples. RAB values are colored and binned into specific thresholds. Villages are identified by color and letter; hosts are demarcated by bars below village.

Conclusions:

- Our findings indicate that highly localized patterns of genetic relatedness suggest human hosts within villages remain key drivers of transmission, allowing parasite populations to persist despite mass drug administration (MDA) efforts.

Citations

1. Luo, F. *et al.* A chromosome-level genome of the human blood fluke *Schistosoma japonicum* identifies the genomic basis of host-switching. *Cell Rep.* **39**, 110638 (2022).
2. Weerakoon, K. G. *et al.* Advances in the diagnosis of human schistosomiasis. *Clin. Microbiol. Rev.* **28**, 939–967 (). 2015
3. Wang, L.-D. *et al.* China's new strategy to block *Schistosoma japonicum* transmission: experiences and impact beyond schistosomiasis. *Trop. Med. Int. Health* **14**, 1475–1483 (2009).