Genomic data reveal a north-south split and introgression history of blood fluke (*Schistosoma haematobium*) populations from across Africa

Roy N Platt II[1], Egie E. Enabulele[1], Ehizogie Adeyemi[2], Marian O Agbugui[3], Oluwaremilekun G Ajakaye[4], Ebube C Amaechi[5], Chika E Ejikeugwu[6], Christopher Igbeneghu[7], Victor S Njom[8], Precious Dlamini[9], Grace A. Arya[1], Robbie Diaz[1], Muriel Rabone[10], Fiona Allan[10], Bonnie Webster[10], Aidan Emery[10], David Rollinson[10, 11], Timothy J.C. Anderson[1]

[1] Texas Biomedical Research Institute, San Antonio TX, United States

[2] Department of Pathology, University of Benin Teaching Hospital, Edo State, Nigeria

[3] Department of Biological Sciences, Edo State University, Uzairue, Nigeria

[4] Department of Animal and Environmental Biology, Adekunle Ajasin University, Nigeria

[5] Department of Zoology, University of Ilorin, Kwara State, Nigeria

[6] Enugu State University of Science and Technology, Nigeria

[7] Department of Medical Laboratory Science, Ladoke Akintola University of Technology, Nigeria

[8] Department of Applied Biology and Biotechnology, Enugu State University of Science and Technology, Nigeria

[9] Central Public Health Offices, Manzini, Swaziland

[10] Science Department, Natural History Museum, London, United Kingdom

[11] Global Schistosomiasis Alliance, London, United Kingdom

Keywords: schistosomes, parasite, *Schistosoma bovis*, hybridization, introgression desert, admixture

Abstract: The human parasitic fluke, *Schistosoma haematobium* hybridizes with the livestock parasite *S. bovis* in the laboratory, but the extent of hybridization in nature is unclear. We analyzed 34.6 million single nucleotide variants in 162 samples from 18 African countries, revealing a sharp

genetic discontinuity between northern and southern *S. haematobium*. We found no evidence for recent hybridization. Instead the data reveal admixture events that occurred 257-879 generations ago in northern *S. haematobium* populations. Fifteen introgressed *S. bovis* genes are approaching fixation in northern *S. haematobium* with four genes potentially driving adaptation. We identified 19 regions that were resistant to introgression; these were enriched on the sex chromosomes. These results (i) demonstrate strong barriers to gene flow between these species, (ii) indicate that hybridization may be less common than currently envisaged, but (iii) reveal profound genomic consequences of interspecific hybridization between schistosomes of medical and veterinary importance.

Funding - This work was funded by NIH NIAID (R01 AI166049).

Introduction

Hybridization and the transfer of alleles via introgression is an important source of genetic variation between species (1). This process allows for allelic variants, which have already been preselected in a donor species, to be introduced into the genome of a recipient species in a single generation. By comparison, it may take multiple generations for random mutation and selection to deliver comparable levels of genetic variation in the absence of introgression (2). As a result, introgressive hybridization, can result in rapid evolution of new genetic traits in hybridizing species. Hybridization between human and animal parasites can lead to the emergence of parasites with novel traits such as increased pathogenicity (3), expanded host range (4), altered transmission dynamics (5) and drug resistance (6). Understanding the frequency and impact of such hybridization events is critical for devising effective disease intervention strategies.

Members of the blood fluke genus *Schistosoma* parasitize a range of mammal species and cause substantial morbidity and economic loss (7). One pair of species, *S. haematobium*, a human parasite, and *S. bovis*, an ungulate parasite common in domestic livestock, are genetically divergent (3-5% and 18% divergent in the nuclear and mitochondrial genomes respectively), but can hybridize and produce viable offspring when given the opportunity (8). Given the close proximity between humans and their livestock and the regular use of the same water bodies, the potential for hybridization between these species is a particular concern and a significant effort has been mounted to identify, monitor and map *S. haematobium* and *S. bovis* hybrids (9). Multiple studies have reported mitochondrial and/or ribosomal DNA from *S. bovis* in *S. haematobium* populations (for examples see 8, 10, 11, 12). The high frequency of individuals with discordant mitochondrial and nuclear markers has been used to argue that hybridization is common and that the zoonotic threat of *S. bovis* should be considered in human schistosomiasis control programs (13).

However, several more recent multi-marker and genomic studies using single nucleotide variants (SNVs), microsatellites markers and whole genome sequence assemblies have suggested that hybridization between *S. haematobium* and *S. bovis* may not occur as frequently as previously postulated. Exome (14), whole genome (15), and microsatellite data (16), and others (17-20) failed to identify evidence of contemporary hybridization in field-collected parasites. Instead, these studies indicate that *S. bovis* and *S. haematobium* are genetically distinct and do not hybridize frequently but evidence for historical hybridization is clearly evident within genomes of *S.*

haematobium. As a result, some *S. bovis* genes have introgressed into the *S. haematobium* population and reached high frequency; evidence of a potential, adaptive introgression event (14, 15, 17, 19).

In this study, we build upon previous work and try to address knowledge gaps by analyzing a comprehensive dataset of 34.6 million genome-wide SNVs from *S. bovis* (n=21) and *S. haematobium* (n=141) samples collected from 18 countries across the African continent. Many of these samples presented discordant mitochondrial and ribosomal DNA profiles and were categorized as *S. haematobium-bovis* hybrids. This expanded dataset, and recent availability of a high-quality *S. haematobium* genome assembly (21), allows for a more detailed examination of the genetic relationships between these two species and the potential consequences of hybridization on their evolution and epidemiology. Our results: (i) reveal a strong discontinuity between northern and southern *S. haematobium* populations; (ii) define similar genomic introgression profiles in *S. haematobium* sampled from locations 3,002 Km apart; (iii) fine-map introgressed genome regions and identify putative genes driving adaptive introgression; (iv) identify two distinct lineages of *S. bovis*-like mitochondrial DNA in northern *S. haematobium*, consistent with rare, ancient introgression and (v) identify "introgression deserts" on the ZW chromosomes consistent with the sex chromosomes maintaining species integrity. These results enhance our understanding of *Schistosoma spp.* epidemiology, with important implications for control efforts.

Results

<u>DNA Sequencing and Genotyping</u> – We examined 219 *Schistosoma* samples from 24 countries (Figure 1A). Just over 80% (n=176) of the samples were collected as part of this study with the remaining 43 samples made available through open access resources. Median genome coverage per sample was 29.7x. After filtering, the final dataset contained 35,817,757 total SNVs, 7,206,957 common (minor allele frequency; MAF>0.05%) SNVs, and 446,162 common unlinked SNVs genotyped from 162 samples (141 *S. haematobium* and 21 *S. bovis*; Figure 1B). NCBI Short Read Archive (SRA) accessions and sample metadata are available in Supplemental Table 1.

<u>Population structure and ancestry</u> - We examined relationships among samples with a PCA of 355,715 unlinked, common, autosomal SNVs (Figure 1C). Each of our samples fell into one of 3 *K*-means clusters along PC1 and PC2. The three clusters corresponded with (a) *S. haematobium*

individuals from northern Africa, (b) southern Africa, and (c) all *S. bovis* samples. The northern population includes samples collected in Cameroon, Cote d' Ivoire, Egypt Gambia, Guinea Bissau, Liberia, Mali, Niger, Nigeria, Senegal and Sudan. The southern population includes samples collected from Angola, Eswatini, Kenya, Madagascar, Namibia, Tanzania, Uganda, Zambia, and Zanzibar. In general, the equator was used to delineate the northern and southern populations. No samples were placed intermediate between the *S. haematobium* and *S. bovis* clusters: we found no evidence for F1 *S. haematobium-bovis* hybrids among these samples. The weighted, Weir and Cockerham F_{ST} (22, 23) between the *S. bovis* and *S. haematobium* samples was high ($F_{ST} \geq 0.74$-0.79). We observed strong subdivision between northern and SE *S. haematobium* populations ($F_{ST} = 0.16$; Figure 1C) with multiple peaks (Figure 2D). There were 275,657 sites showing fixed differences ($F_{ST} = 1$) between *S. bovis* and *S. haematobium* (Supplemental Table 2).

We used Admixture v1.3.0 (24) to quantify ancestry among the samples (Figure 1D, Supplemental Figure 1). We found that when $k = 2$, *S. bovis* and south African *S. haematobium* individuals were exclusively assigned different ancestry components. By contrast *S. haematobium* samples collected from north Africa was a composite of the two population components including 0.5-26.2% (median = 4.2%) of the *S. bovis* population component. Nigeria was an outlier (Kruskal-Wallis H test statistic = 7915, *P*-value = 0.0049; Supplemental Figure 2), showing significantly higher levels of introgression together with extremely high variance in introgression levels (2.8 – 26.2%): further analyses of Nigerian samples will be of considerable interest.

<u>Reference biases</u> - We tested the data for read-mapping reference biases, which could occur if non-*haematobium* species map poorly to the *S. haematobium* reference assembly (GCF_000699445.3). Mapping rates were 77.7%, 82.8% and 76.3% for *S. bovis*, northern *S. haematobium* and southern *S. haematobium* populations (Figure 1A). A *t*-Test failed to identify differential mapping rates between *S. haematobium* and *S. bovis* ($p = 0.397$), suggesting that reference bias does not significantly contribute to the results observed.

<u>Phylogenetics</u> - The species tree generated using SVDquartets and nuclear SNVs revealed a well-resolved topology. We examined 2,500,000 random quartets, representing 8.82% of all possible distinct quartets. Of these, 18.5% (n = 463,571) were incompatible with the final tree (Figure 3). *S. haematobium* and *S. bovis* were resolved into two clades and *S. haematobium* population tips generally fell into clades reflecting geographic relationships.

On average, we were able to assemble 15,558.4 bp of the mitochondrial genome for each sample 4,757 of which were phylogenetically informative sites. The mitochondrial phylogeny (Figure 4) reveals two major mitochondrial haplotypes, one containing *S. haematobium* individuals from across Africa and another clade containing all of the *S. bovis* and 38 north African *S. haematobium*. The presence of *S. haematobium* samples within a larger *S. bovis* clade is consistent with *S. bovis* mitochondrial introgression into *S. haematobium* that has been frequently reported in field samples (ex. 25). Within the *S. bovis* clade, all *S. haematobium* samples with the introgressed *S. bovis* mitochondria fell into two monophyletic groups, clades "A" and "B". mtDNA haplotypes from these two clades were from samples widely distributed in northern Africa. For example, the same clade "A" haplotype was found in samples from Egypt, Niger and Cote d' Ivoire (>3,300 km apart). The clade "B" haplotype was found in Niger, Nigeria and Cote d' Ivoire, a linear distance of 1,171 km. Bootstrap support for each of these major clades was strong (100%). Phylogenetic trees in Newick format are available in the supplementary materials.

<u>Hybridization and Introgression</u> - We used four methods to identify signatures of hybridization and introgression between *S. haematobium* and *S. bovis*. These methods include the f_3, *D*-statistic, local ancestry assignment (RFmix; 26) and phylogenetic discordance (TWISST; 27). First, the (NW: SE, Sb) f_3 =-0.128 (SE= $0.8e^{-3}$, z-score=-156.4) was significantly negative indicating the north African *S. haematobium* population contains *S. bovis* ancestry. Next, we used the *D*-statistic to test for introgression between *S. haematobium* and *S. bovis* while accounting for lineage sorting (Figure 2E). We averaged *D* in 10Kb blocks. *D* was significantly positive (D=0.46, σ_M=0.007, n=30,278) indicating introgression between north African *S. haematobium* and *S. bovis*.

We examined the landscape of introgression across the genome using local ancestry with RFmix. We used 38 southern African *S. haematobium* lacking *S. bovis* introgression and 13 *S. bovis* samples to serve as reference panels for "pure" *S. haematobium* and *S. bovis*. RFmix results showed that ancestry across the genome was not uniformly distributed in the north African population (Figure 2A). Within the north African population, *S. bovis* ancestry blocks ranged in frequency from 0-100% at any particular locus. Each north African *S. haematobium* sample contained 4.1-22.0% *S. bovis* ancestry (median 7.0%). By comparison the median *S. bovis* ancestry was 0.02% and 100% in the southern *S. haematobium* and *S. bovis* control samples, respectively.

We used TWISST v67b9a66 (27) as an independent method for identifying local introgression. TWISST, measures shifting gene tree frequencies across the genome. Trees were generated from 37,200 non-overlapping, 10kb, sliding windows. On average (mean) each window contained 657.3 SNVs. We examined the three possible topologies between northern *S. haematobium*, southern *S. haematobium* and *S. bovis* using *S. margrebowiei* (GCA_944470205.2; 28) as an outgroup (Figure 2B). The expected species tree, with a monophyletic clade of *S. haematobium*, sister to *S. bovis* was the most common with a mean weight of 0.876 across the genome. The discordant topology uniting northern *S. haematobium* and *S. bovis* was the second most abundant topology (weight = 0.085) compared to the topology with southern *S. haematobium* and *S. bovis* (weight = 0.039).

We examined the genome for regions that are devoid of introgressed *S. bovis* in the northern *S. haematobium* population. There were 918 genomic regions lacking *S. bovis* alleles with a median size of 35.8Kb (Figure 2F). With log transformation and robust Z-scores we identified 19 genome regions that were significant outliers in terms of length ranging from 1.13-6 Mb (median 1.67 Mb). Thirteen of the 19 genome regions were on the ZW scaffold and accounted for 32% (28.6 Mb) of its total length.

<u>Introgression profiles in different countries</u> - We examined the pattern of introgression in individual countries of north Africa (Supplemental Figure 3) as determined by RFMix. The overall patterns of introgression across the genome were consistent between north African populations. Pairwise comparisons of introgressed allele frequencies between countries were positively correlated (r = 0.59-0.8; Supplemental Figure 4) despite distances spanning up to 3,000 km.

<u>Impact of introgression on nucleotide diversity and genetic differentiation of *S. haematobium*</u> - We masked introgressed alleles within individual genomes, and recalculated π, F_{ST} and a PCA (Figure 5). Prior to masking, mean nucleotide diversity (π), was 2.3-fold greater in the northern ($\pi = 2.991 \times 10^{-3}$) vs southern ($\pi = 1.278 \times 10^{-3}$) *S. haematobium*, and π was 3.3-fold greater in *S. bovis* ($\pi = 8.329 \times 10^{-3}$) than the entirety of *S. haematobium* ($\pi = 2.523 \times 10^{-3}$). After masking, northern African *S. haematobium* nucleotide diversity was reduced to nearly identical levels seen in the south population: $\pi_{NW} = 2.991 \times 10^{-3}$ to $\pi_{NW} = 1.07 \times 10^{-3}$. By comparison, removing introgressed alleles had no impact on F_{ST} (F_{ST} = 0.154) between northern and southern *S. haematobium.* Additionally, the structure of the PCA was retained, demonstrating that the differentiation between NW and SE Africa is not driven by introgression.

Both PCA (Fig 1 and 5) analyses and Admixture (k=3) plots (Supplemental Figure 1) recovered one *S. bovis* and two *S. haematobium* population components. The distribution of these components, shows a clear distinction between countries in the north and south, with a boundary that extends from Cameroon to Somalia and into Madagascar (Figure 5A). This boundary generally coincides with the distribution of *S. bovis* in north and central Africa (29) with a few exceptions (30, 31). The division of the *S. haematobium* into northern and southern populations was consistent among analyses with one exception. Madagascar was an intermediate population in Admixture analysis (k=3). In the PCA, samples from Madagascar area assigned to the southern cluster, but they form a distinct subgroup that is intermediate between the remaining southern and northern samples.

Dating Introgression - We used the size of introgressed haplotype blocks to estimate the number of generations since hybridization for each *S. haematobium* sample in north Africa (Supplemental Figure 5). This gave estimated hybridization dates of ~257-879 generations ago (Median – 426 generations, 95% confidence intervals = 281.6-764 generations). *S. haematobium* generation time varies from 3-4 (32) months in lab populations, but is estimated to be 6-12 months in wild populations (33). These generation times imply that admixture between *S. haematobium* and *S. bovis* occurred ~106 years ago assuming four generations per year (high transmission) or 426 years assuming one generation per year (low and / or seasonal transmission). Dating estimates varied between countries: median estimates are lowest in Egypt (286.9) and highest in Nigeria (565) despite their relatively close proximity. A one-way ANOVA indicated significant differences in the number of generations since hybridization between countries (p-value = $1.3e^{-10}$; Supplemental Figure 5).

Selection and adaptive introgression - We examined *S. haematobium* and *S. bovis* populations for signatures of selection using normalized, xpEHH (Figure 2C). We found 996 statistically significant xpEHH values after multiple test correction. We combined values within 1Mb to identify 15 genome regions with signatures of positive selection in the northern population and five in the southern population (Supplemental Table 3). The median normalized xpEHH in each of these regions was >|6| and the windows ranged in size from 3 bp to 709,928 bp (mean 139,942 bp).

We combined selection and introgression analyses to identify genome regions showing evidence for adaptive introgression of *S. bovis* alleles into the northern *S. haematobium* population. These

regions contained outlier values for selection (xpEHH), elevated Patterson's D (D ≥ 0) indicative of introgression, high levels of S. *bovis* ancestry (>95%) and significant differentiation from southern *S. haematobium* (F_{ST} ≥ 95TH percentile). Two genome regions met these criteria; chromosome four (NC_067199.1: 28,476,500-28,813,500) and chromosome five (NC_067200.1:9,773,000-10,447,000). These genome regions span 1.01 Mb and 15 genes; eight on chr four and seven on chr five (Table 1). Of the 74,955 SNVs in these regions, 989 are nonsense or missense mutations. We found 37 missense SNVs where the *S. bovis* allele is at or near fixation in the northern population (Supplemental Table 4). All of these variants are on chr four and fall within four genes; leishmanolysin-like peptidase, a Rho GTPase-activating protein 35, Jumonji domain-containing protein six (JMJD6_4), and Jumonji domain-containing protein six (JMJD6_3).

Discussion

Our analysis of >38 million SNVs provides compelling evidence that *S. haematobium* and *S. bovis* are genetically well differentiated. This conclusion is supported by multiple lines of evidence: high F_{ST} values (F_{ST} ≥ 0.74-0.79; Figure 1C; Figure 2D), distinction by PCA (Figure 1C), strong differentiation by ancestry components in Admixture analyses (Figure 1D) and well supported monophyletic clades in the nuclear species tree (Figure 3). The agreement between these approaches suggests that strong barriers to gene flow exist between these two species.

Our analysis revealed that northern African *S. haematobium* are genetically differentiated (F_{ST} = 0.16) from the southern population. The boundary between these populations appears to extend from Cameroon, Gabon, the Central African Republic, South Sudan, and Ethiopia (Figure 5A). When introgressed *S. bovis* alleles were removed from genomic data, F_{ST} between these populations remained unchanged (F_{ST} = 0.154). Hence, genetic differences between the northern and southern populations do not appear to be driven by introgressed *S. bovis* alleles.

Our results indicate barriers to gene flow exist between northern and southern *S. haematobium*. The southern *S. haematobium* clade is monophyletic within a larger clade of northern African *S. haematobium* (Figure 3). This indicates that *S. haematobium* originated in one of the northern African countries and is consistent with previous work that identified the Arabian Peninsula/Asia as a potential ancestral source population (34). It is possible that the two populations are defined by the distribution of their intermediate hosts. Regional differences in parasite compatibility with their intermediate snail hosts can occur within limited geographical areas (35). *S. haematobium*

from North Africa and the Middle East are transmitted by the *Bulinus truncatus*/*tropicus* species complex and parasites from the Afro-tropical region are transmitted by snails of the *Bulinus africanus* group (36). The presence of such a barrier has important implications for our understanding of the ecological and epidemiological factors that shape the distribution and dynamics of these two parasite populations. Further investigation at the population boundaries may provide new insights into biological differences and incompatibilities between northern and southern *S. haematobium* populations.

Four aspects of our results support an ancient introgression hypothesis. First, each of the north African *S. haematobium* samples contain low levels of *S. bovis* ancestry with the exception of the sole Cameroonian sample. Percentages of *S. bovis* ancestry per individual are similar across multiple analyses: introgressed haplotype blocks from RFMix account for 4.1-22% of individual genomes in the northern *S. haematobium* population, while the population component associated with *S. bovis* in Admixture ranges from 5-26.2% at *K*=2.

Second, the landscape of introgressed alleles across the genome is consistent across north African samples (Supplemental Figure 3) and positively correlated (Supplemental Figure 4) despite being separated by ≤3,000 Km. For this profile to be conserved across such a broad distance suggests (A) it occurred in an ancestor of the north African *S. haematobium* or (B) introgressed alleles provided a selective advantage that spread throughout the north African population. Our data support the later with the nuclear phylogeny (Figure 3) showing that the northern population is paraphyletic; the southern population is a monophyletic clade within this group and it lacks introgressed *S. bovis* alleles. We also observe that some introgressed alleles have reached high frequency in the north African population and show signs of selection (Figure 2). Finally, our data suggest that there is a barrier to gene flow/migration between northern and southern *S. haematobium* populations, restricting dispersal of introgressed alleles to the southern population.

Third, mitochondrial DNA provides insights into a minimal number of hybridization events. 58% of northern *S. haematobium* samples contain introgressed S*. bovis* mitochondria (Figure 4). If the introgressed *S. bovis* mitochondria were the result of contemporary hybridization, we would expect sister relationships between *S. bovis* and *S. haematobium* at the terminal branches of the tree. However, we find that introgressed *S. haematobium* individuals with introgressed *S. bovis* mitochondrial genomes form two monophyletic clades. Clade “A” contains samples from Egypt,

Niger, and Cote d' Ivoire, and Clade "B" contains samples from Niger, Nigeria, and Cote d' Ivoire; each clade spanning >1,000 Km. The most parsimonious interpretation of the phylogeny is that the introgressed *S. bovis* mitochondria share two distinct origins and imply at least two admixture events resulting from mating between a *S. bovis* female and *S. haematobium* male that occurred in the distant past. We note that laboratory crosses between *S. haematobium* are often asymmetric, and may only produce offspring when male *S. haematobium* are mated with female *S. bovis* (37) or produce more male offspring (32). As females are the heterogametic sex, F1 females are expected to show reduced fitness (Haldane's rule; 38). This may contribute to the limited number of *S. bovis* mtDNA lineages observed in *S. haematobium populations*.

Fourth, introgressed *S. bovis* nuclear loci are heavily fragmented within the *S. haematobium* genomes indicating multiple generations since introgression. Our estimates of time since introgression span 257-879 generations ago (95% confidence interval). Introgressed loci were measured in tens of kilobases (median = 76.3 Kb) and the largest blocks extended into the megabases (max = 4.05 Mb). This contrasts with early generation hybrids which would have introgressed block lengths spanning, or nearly spanning entire chromosomes (39). One Nigerian sample contained ~25% introgressed DNA, consistent with expectations for a F2 backcross. However, the maximum introgressed fragment size in a Nigerian sample was only 2.83 Mb and median block size in these samples ranged from 47.1-97.6 Kb indicating multiple recombination events. We found that the time since introgression was significantly different between multiple countries (Figure 6). For example, neighboring countries Niger (453 generations) and Nigeria (565 generations) were not significantly different, but introgression in Cote d' Ivoire (385 generations) appears to have occurred more recently. The variation in the estimates of generations since introgression are consistent with several regional introgression events. Alternatively, variation in age estimates between countries could reflect extrinsic factors like seasonality or intervention strategies that could lengthen or reduce generation times within sub-populations. If this were the cause, it is possible that the number of generations that have lapsed since an introgression event may vary between countries.

S. bovis shows 3.3-fold higher diversity than *S. haematobium*, while genetic diversity (π) is 2.3-fold greater in the north *S. haematobium* population than in the south African *S. haematobium* population. When the introgressed *S. bovis* alleles are removed from the analyses, this difference in genetic diversity between the north and south *S. haematobium* populations is reduced to just 1.05-fold and π is not significantly different (Figure 5B). By contrast, F_{ST} values between northern

and southern *S. haematobium* are consistent whether introgressed alleles are considered (F_{ST} = 0.16) or not (F_{ST} = 0.154) and the relationship among samples in the PCAs is nearly identical when introgressed alleles are included or excluded. These results indicate (i) that the elevated π in northern African *S. haematobium* results from *S. bovis* introgression and (ii) that northern and southern *S. haematobium* populations are well differentiated even after removing introgressed *S. bovis* alleles.

Given that introgressed *S. bovis* alleles have persisted in the northern *S. haematobium* population, we examined the data for signals of adaptive introgression. We found two introgressed genome regions with signals of positive selection in the northern population on chr four (NC_067199.1:28,348,440-28,877,530) and chr five (NC_067200.1:9,712,340-10,514,400). Despite the convergence of signals to these regions, we were not able to identify variants driving selection in these regions. We found 37 missense SNVs where the *S. bovis* allele was nearly fixed in the northern population, but none withstood multiple test correction for directional selection. These variants occur in four genes (WormBaseParaSite v18.0; 40), a Rho GTPase-activating protein 35 (n_{SNVs} = 30; MS3_00007803) a Leishmanolysin-like peptidase (n_{SNVs} = 4; MS3_00007802), and two members of the Jumonji domain-containing protein 6 family; JMJD6_4 (n_{SNVs} = 1; MS3_00010935) and JMJD6_3 (n_{SNVs} = 2; MS3_00010934). The same Leishmanolysin-like peptidase (Table 1) has been identified as a candidate for adaptive introgression from *S. bovis* into *S. haematobium* in two previous studies (14, 15). Genes in same invadolysin gene family are known to modulate the snail host immune system in Schistosoma mansoni (41, 42) and this particular gene has been associated with cell migration and invasion in other parasitic taxa (43).

We also observed genomic regions on three chromosomes of the *S. haematobium* samples where *S. bovis* introgression is rare or absent (Figure 2F). These introgression deserts may contain hybrid incompatibility loci that result in reduced fitness of early generation hybrids and present barriers to further introgression. Thirteen of the 19 regions occur on the sex chromosomes. This is consistent with expectations given that a reduced recombination and smaller effective populations size of the sex chromosomes expose deleterious, introgressed alleles to strong selection pressures (44) and could lead to female sterility as predicted by Haldane's rule.

Understanding hybridization and introgression between *S. haematobium* and *S. bovis* is important for disease control. If hybridization between these species is infrequent, then there may be minimal benefit in linking strategies that manage both human (*S. haematobium*) and livestock (*S. bovis*) *Schistosoma* species. Consistent with this, our results from these samples suggest that hybridization between these species is rare, and gene flow is insufficient to break down strong reproductive barriers between these species. However, adaptive introgression has introduced *S. bovis* alleles into *S. haematobium* populations. This is a clear example of alleles being transferred between livestock and human parasites through introgression. Some *S. bovis* alleles have reached high frequency and are likely selectively advantageous. Future work should aim to understand how the introgressed *S. bovis* variants contribute to the fitness of *S. haematobium* individuals. The strong differentiation between northern *S. haematobium* populations, carrying introgressed *S. bovis* alleles and southern *S. haematobium* populations, with no introgression, is of particular interest. Additionally, future work should examine differences between northern and southern *S. haematobium* populations, and test whether they are reproductively isolated.

Several limitations to our study and its conclusions should be noted. First, our results indicate that hybridization between *S. haematobium* and *S. bovis* is rare and ancient when measured on the continental scale. While early generation hybrids between *S. haematobium* and *S. bovis* may be found with further sampling, it is clear that strong barriers to gene flow maintain species integrity. Second, recombination rates have not been quantified in *S. haematobium* so our estimates of age of admixture are based on recombination rates measured in *Schistosoma mansoni* (45). To improve the accuracy of these estimates, direct measures of recombination rates from *S. haematobium* genetic crosses are needed. Third, our results pertain to *S. haematobium* and *S. bovis*. Extant hybridization between other schistosome species (*S. haematobium*/*S. guineensis* and *S. bovis*/*S. curassoni*) have been documented in field collected samples with genomic data (46, 47). Our results suggest that *Schistosoma* species pairs may form a spectrum in hybridization frequency and compatibility. Future work to understand the factors that impact hybridization and present barriers to gene flow between schistosomes species pairs will be of great interest, and can provide a more nuanced understanding of hybridization and potential implications for schistosome control.

Online Methods

Sample collection: description, ethics, and identification – We used samples or data from three sources. i) The first dataset was generated from samples provided by the Schistosomiasis Collection At the Natural History Museum (48) which is housed at the Natural History Museum (London). SCAN samples consisted of individual miracidia and cercariae preserved on Whatman FTA cards (49). We analyzed 114 *S. haematobium* and *S. bovis* samples from 123 individual hosts (snails or humans) and 12 Africa countries. ii) In addition to the SCAN samples, we collected nine adult *Schistosome* worms, presumed to be *S. bovis*, from the intestines of routinely slaughtered cattle from meat vendors at three abattoirs located in Auchi, Benin City, and Enugu in Nigeria. In the laboratory, the mesenteric vessels of each purchased intestines were visually inspected for schistosome parasites. Adult schistosomes were recovered using forceps and washed in saline solution. Adult pairs were separated into males and females before being stored in 96% ethanol for subsequent DNA isolation analyses. iii) Finally, for the third source of data we used whole genome sequence data from NCBI (14, 15, 17, 21, 28, 50).

Samples provided by the SCAN repository were originally collected in accordance with protocols approved by local, state, and national authorities, including the Ministry of Health. The Imperial College Research Ethics Committee (ICREC) at Imperial College London, in conjunction with ongoing Schistosomiasis Control Initiative (SCI) activities, provided additional ethical guidance for samples collected through the CONTRAST program. Ethical clearance and study protocols for Nigerian samples were approved by the National Health Research Ethics Committee of Nigeria (NHREC) (protocol number: NHREC/01/01/2007– 30/10/2020 and approval number: NHREC/01/01/2007– 29/03/2021) and the Institutional Review Board (IRB) of University of Texas Health, San Antonio Texas, United States of America (protocol number: HSC20180612H). Informed consent was obtained from all participants, with processes tailored to ensure understanding and voluntary participation. All data were anonymized to protect participant privacy, and schistosomiasis-positive individuals were treated with a single dose of praziquantel (40 mg/kg). For livestock parasite collection, approval was secured from local veterinarians. No animals were euthanized for research purposes; *Schistosoma* samples were collected during routine activities at abattoirs. Further details on collection methods, ethical approvals, and data availability for public samples can be found in their respective publications documented in Supplemental Table 1.

Provisional species identifications were assigned to cercariae and miracidia based on sampled host. For example, miracidia hatched from eggs collected from human urine samples were

assumed to be *S. haematobium* while miracidia hatched from eggs in cattle feces were assumed to be *S. bovis*. Cercariae collected from snails were identified by Sanger sequencing the mitochondrial cox1 region and the ribosomal internal transcribed spacer (ITS) rDNA region as previously described (20). Downstream genetic analysis with whole genome SNVs was used to confirm and reassign species identifications where necessary.

Library prep and sequencing – DNA from single parasites stored on FTA cards was subjected to whole-genome amplification (WGA) using methods previously described in (49). DNA was extracted from single male adult *S. bovis* worms using the DNeasy® Blood and Tissue kit before subsequent WGA. We quantified amount of schistosome DNA in each WGA sample by real time quantitative PCR (qPCR) reactions using the single copy gene α-tubulin 1 gene markers primers (*S. haematobium*: forward [GGT GGT ACT GGT TCT GGT TT], reverse [AAA GCA CAA TCC GAA TGT TCT AA]; *S. bovis*: forward [ATG GCC TCG TTA TCA ACC AT], reverse [TGG CCT CGT TAT CAA CCA TA] following previously described protocol in (49). DNA sequencing libraries were generated from 500 ng of DNA per sample using the KAPA Hyperplus kit protocol with the following modifications: i) enzymatic fragmentation at 37°C for 10 minutes, ii) adapter ligation at 20°C for an hour, and iii) 4 cycles of library PCR amplification. After qPCR quantification of each library with KAPA Library Quantification Kits, samples with similar concentrations were combined into pools for sequencing at 4nM, while samples with disparate concentrations were equalized in 10 mM Tris-HCl pH 8.5 before pooling. Libraries were sequenced with 150 bp paired-end reads on two Illumina NovaSeq flowcell. All resulting reads were deposited in the NCBI Short Read Archive under BioProject PRJNA636746 and are documented in Supplemental Table 1.

Computing environment - Analyses were conducted on the Texas Biomedical Research Institute's high-performance computing cluster, with worker nodes containing 96 cores and 1 TB of memory. Computational environments were managed using Conda v22.9.0. Environmental recipe files, Jupyter notebooks, and other code can is archived on GitHub (github.com/nealplatt/sch_hae_scan v0.1z) and at DOI:10.5281/zenodo.13124718.

Read filtering and Mapping - Raw reads were quality trimmed with trimmomatic v0.39 (51) using the following parameters: LEADING:10, TRAILING:10, SLIDINGWINDOW:4:15, MINLEN:36, ILLUMINACLIP:2:30:10:1:true. This command removed low quality bases at the beginning and ends of the reads, removed portions of the read where quality dropped below a minimum threshold, trimmed adapter sequences and discarded reads <36 nts. We then mapped the

trimmed reads to the *S. haematobium* reference genome, GCF_000699445.3 (21) with BBMap v38.18 (52). On average the *S. haematobium* and *S. bovis* (GCA_944470425.1) genome assemblies are ~97% similar across their genomes (17) which should minimally affect reference biases when mapping short reads. However, to avoid reference biases we used the 'vslow' and 'minid=0.8' options with BBMap and discarded ambiguously mapping reads ('ambig=toss').

Genotyping, phasing, and filtering - Mapped reads were sorted with SAMtools v1.13 (53) and checked for duplicates with GATK v4.2.0.0's (54) mark_duplicates. Then single nucleotide variants (SNVs) were genotyped with HaplotypeCaller and GenotypeGVCFs. To make the dataset more manageable, we genotyped each chromosome separately using the -L option. Next, we removed all indels and hard filtered SNVs based on QualByDepth ("QD < 2.0"), RMSMappingQuality (MQ < 30.0), FisherStrand (FS > 60.0), StrandOddsRation (SOR > 3.0), MappingQualityRankSumTest (MQRankSum < -12.5), and ReadPosRankSumTest (ReadPosRankSum < -8.0) with GATK's VariantFiltration. We removed multi-allelic sites, and sites with genotype quality (GQ) <20 or read depth (DP) <8 with VCFtools v0.1.16 (55). After these filters were applied we removed genomic sites that were genotyped in ≤50% of individuals and then any individuals that were genotyped at ≤50% of sites.

SNVs on each chromosome were phased using Beagle v 5.2_21Apr21.304 (56) in windows of 20cm and a 10cm overlap. Assuming a uniform recombination rate similar to *S. mansoni* across the genome, these values are comparable to a 6.5 Mb window and a 3.25 Mb step size (45). We used a burn in of 20 iterations and 60 iterations for the phasing run. All phased chromosome VCFs were combined into a single file using vcfcombine from vcflib v1.0.3 (57) before an additional round of post-phase filtering.

In some cases, multiple miracidia were analyzed from a single host potentially adding highly related samples to our dataset and skewing the downstream results. To remove these, we examined kinship coefficients in our samples using the autosomal chromosomes and the "–unrelated" function in king v2.2.7 (58). This parameter identifies second-degree relatives within the dataset that can be removed prior to downstream analyses. Next, we generated a set of SNVs that were common (minor allele frequency; MAF > 0.05) and unlinked. Unlinked loci were filtered with Plink v1.90b6.21 (59) by removing linked SNVs with a pairwise $r^2 > 0.2$. This filter was applied in 25 Kb sliding windows with a 5kb steps. Finally, we used SnpEff v5.1 (60), to identify the impact of these SNVs on the amino acid sequence in coding regions. To do this we imported the *S.*

haematobium reference genome (GCF_000699445.3) along with the associated GenBank annotations to create a custom database.

Principal Component Analyses - We used a series of tools to explore population structure in our data sets. We used common (minor allele frequency; MAF>0.05), unlinked, autosomal SNVs and Plink v1.90b6.21 (59) to generate a principal component analysis (PCA) to examine relationships among the samples. We used a *K*-means clustering algorithm to assign each sample to between 1 and 10 populations with the Kmeans() function in sklearn.cluster v1.2.0 (61). We then used the Elbow method (62) to examine distortion in the model and determine the optimal number of clusters in the data. Once we identified the optimal number of clusters, we assigned each sample within a cluster, and those designations were used to differentiate the *S. haematobium* populations using analyses as below. These cluster assignments were also used validate the assumed species identify of each sample.

Admixture - We examined the ancestry of each sample with Admixture v1.3.0 (24) and the same unlinked, autosomal SNV dataset from the PCA analyses. However, we further thinned the SNV data with VCFtools v0.1.16 (55) ensuring that no two SNVs were within 10kb of each other. This step minimizes any potential effects of linkage on the results. We ran Admixture v1.3.0 (24), allowing for 2 to 20 possible population components, and used the cross-validation error to determine the optimal range (63). Additionally, we randomly selected individuals with ≥99.999% *S. bovis* or *S. haematobium* ancestry in the k=2 analysis to serve as reference samples for each species in downstream analyses.

Nucleotide diversity (π) and Fixation index (F_{ST}) - We used scikit-allel v1.3.5 (64) to calculate nucleotide diversity (π) and the fixation index (F_{ST}) in sliding windows of 10 kb using autosomal, common (MAF >0.05) SNVs allel.windowed_diversity() and allel.windowed_weir_cockerham_fst() functions. The weighted, Weir-Cockerham F_{ST} (22, 23) was measured between species (*S. haematobium* vs. *S. bovis*) and between the *K*-means populations. Next, we used the reference panel, described above, to identify ancestry informative sites between the *S. haematobium* and *S. bovis* samples. We used scikit-allel v1.3.5's (64) allel.weir_cockerham_fst() to calculate F_{ST} at individual sites. Only sites where F_{ST} = 1 where retained.

Biogeography – *S. haematobium* was split into two groups based on the *K*-means clustering analysis of the PCA results. At k=2 Admixture differentiated *S. haematobium* and *S. bovis* samples, but at k=3 Admixture broadly confirmed the presence of two different *S. haematobium* populations. We used the admixture proportion (Q) from k=3, to visualize how the populations were distributed across Africa. The presence of this ancestry component was extrapolated into unsampled geographic regions using the OrdinaryKriging() function implemented in pykrige v1.7.0 with a linear variogram model (65). Geographic distances between samples were calculated with the haversine() v2.8.0 function (https://pypi.org/project/haversine/).

Genome-wide tests for introgression - We used a series of tests to explore the presence of introgression between *S. bovis* and the *S. haematobium* populations. First, we used average_patterson_f3() from scikit-allel v1.3.5 (64) to calculate a normalized f_3 (66) averaged across blocks of 500 SNVs. Next, we tested for gene flow using the *D*-statistic, also known as the ABBA BABA test (67). We used *S. margrebowiei* (GCA_944470205.2; 28) as the outgroup (O), *S. bovis* as the donor population (P3), and the *S. haematobium K*-means populations as the recipients (P1 and P2). We measured *D* across the genome in 500 SNV blocks with moving_patterson_d() in scikit-allel v1.3.5 (64). Introgressed loci were defined when $D>0+2\sigma$.

Local Ancestry Assignment - For local ancestry assignment, we used RFMix v2.03-r0 (26) and TWISST v67b9a66 (27). RFMix v2.03-r0 (26) uses a random forest approach to assign local ancestry to genomic segments by comparing samples to reference panels. For this, we used the reference samples selected from the Admixture analyses. We generated a genetic map using a uniform recombination rate estimated from *S. mansoni* crosses (1 centimorgan = 287,000 bp; 45). The remainder of the parameters were set to the default.

TWISST v67b9a66 (27) uses gene trees sampled from across the genome to identify potentially introgressed loci. It does this by iteratively sampling subtrees from the gene tree and calculating relative support for each of the possible species trees. We generated gene trees from loci containing 500, phased, common (MAF >0.05) SNVs with RAxML-NG v1.1 (68). For each locus we searched for the 10 best trees and then bootstrapped the best tree for 100 replicates using the GTR+ASC_LEWIS substitution model and *S. margrebowiei* as an outgroup. Nodes supported in ≤10 bootstrap replicates were collapsed with Newick Utilities v1.6 (69). The collapsed trees were used as input for TWISST v67b9a66 (27). Samples were assigned to their *K*-means population.

Selection - We compared selection in the *S. haematobium* intra populations using cross-population extended haplotype homozygosity (xpEHH; 70). Unphased xpEHH was measured with selscan v2.0.0 (71). The resulting unphased xpEHH values were normalized with norm v1.3.0 and the '--xpehh flag'. Bonferroni corrected p-values were assigned to each site. Sites with a corrected p-value < 0.01 were considered to be experiencing putative directional selection between the two *S. haematobium* populations.

Identifying putative adaptive introgression –We searched the genome for regions with F_{ST}, Patterson's D, local ancestry, and xpEHH values indicative of adaptive introgression. To do this we examined how these values were distributed across the genome in sliding windows of 337 Kb and 3,370 bp step size; values equivalent to 1% and 0.01% of the autosomal genome. Specifically, we were looking for regions of the genome that are among the most highly differentiated between the two schistosome populations (F_{ST} >= 95th percentile), with statistically significant signals of introgression (Patterson's D > 0) and directional selection (xpEHH p-value < 0.01), and the *S. bovis* alleles are at high frequency in the northern or southern *S. haematobium* populations (>95%). Windows that met these criteria were then merged together if they were within 10Kb of each other to identify loci containing signals of adaptive introgression.

Autosomal Species Tree - To better understand the relationships among the samples, used SVDquartets (72) as implemented in PAUP* v4.0.a.build166 (73) to generate a species tree. We examined 2.5m random quartets along with 100 standard bootstrap replicates. Nodes in the gene trees supported by <10% of bootstrap replicates were collapsed Newick Utilities v1.6 (69).

Dating introgression - Recombination acts to continuously break down introgressed haplotypes. As a result, the size of introgressed haplotype blocks is directly related to the number or generations since hybridization (74). This can be roughly estimated with the formula G=1/LP where G is generations, L is the average length of introgression haplotypes in Morgans, and P is the proportion of the genome from the major parent (75). We identified introgressed blocks and their lengths (L) for each individual with RFMix v2.03-r0 (26) and P was estimated using Admixture (represented as *q*). A one-way ANOVA was used to identify differences in age estimates between populations (countries).

Introgression Deserts – Some regions of the genome may be resistant to introgression. This could present as large regions lacking introgressed alleles. We used the RFMix results to identify regions of the genome where *S. bovis* ancestry was 0% in the north African *S. haematobium* populations. We log-transformed the length of each region and assigned robust Z-scores. Putative introgression deserts were regions with robust Z-scores > 3.

Mitochondrial genome assembly and phylogeny - We used GetOrganelle v1.7.7.0 (76) to *de novo* assemble mitochondrial genomes. Specifically, we used the animal_mt model and 10 rounds of assembly with *k*-mer sizes of 21, 45, 65, 85, and 105. The mitochondrial contigs were then scaffolded with RagTag v2.1.0 (77, 78) and RAxML-NG v1.1 (68) was used to generate a maximum likelihood tree of the mitochondrial genomes. We used a GTR+G substitution model to and 100 starting trees. Nodal support was assessed with 1,000 bootstrap replicates.

Acknowledgements

We thank Sandy Smith and John Heaner for providing computational support at Texas Biomedical Research Institute's High-Performance Computing Cluster. We would like to thank Frederic Chevalier, Winka Le Clec'h, and Kathrin Bailey for providing feedback during the course of this project. This research was funded by the National Institute of Allergy and Infectious Diseases (NIAD R01 AI166049-01). Schistosomiasis Collections at the Natural History Museum (SCAN) was supported by Wellcome Trust grant 1045958/Z/13/Z. The field work referenced in this study relied on various organizations and funding bodies across multiple countries. In Angola, funding was provided by the Calouste Gulbenkian Foundation in collaboration with the Centro de Investigação em Saúde de Angol, the Programa Nacional de Doenças Tropicais Negligenciadas and the Ministério da Saúde de Angola. In Côte d'Ivoire and Niger, the research was funded by SCORE through the Bill & Melinda Gates Foundation, RISEAL-Niger, Université Félix Houphouët-Boigny, and the Centre Suisse de Recherches Scientifiques en Côte d'Ivoire. The Eswatini National Control Program, Environmental Health Unit, and the School Health Program contributed significantly to the work in Eswatini. The CONTRAST project in Kenya, Uganda, and Zambia was funded by the European Commission, the Kenya Medical Research Institute, National Museums of Kenya, Uganda Ministry of Health, and the University of Zambia. In Liberia, the UK Department for International Development provided funding via SCI (now Unlimit Health) and the Ministry of Health of Liberia. Madagascar's efforts were funded through SCAN, Unlimit Health, and the Madagascar Ministry of Public Health. Work in Namibia was funded by the END Fund and the Namibia Ministry of Health and Social Services. In Zanzibar, the ZEST project was funded by

SCORE, the Ministry of Health (Zanzibar), and the Public Health Laboratory—Ivo de Carneri, Pemba.

Individual collectors are listed in the Supplemental Table 1.

Author Contributions

Roy N Platt II: Conceptualization, Formal Analysis, Writing – Original Draft, Writing - Review & Editing. Elisha Enabulele: Conceptualization, Formal Analysis, Investigation, Resources, Writing – Original Draft, Writing - Review & Editing. Ehizogie Adeyemi: Resources. Marian O Agbugui: Resources. Oluwaremilekun G Ajakaye: Resources. Ebube C Amaechi: Resources. Chika E Ejikeugwu: Resources. Christopher Igbeneghu: Resources. Victor S Njom: Resources. Precious Dlamini: Resources. Grace A. Arya: Investigation. Robbie Diaz: Investigation. Muriel Rabone: Resources, Writing - Review & Editing. Fiona Allan: Resources, Writing - Review & Editing. Bonnie Webster: Resources, Writing - Review & Editing. Aidan Emery: Resources, Writing - Review & Editing. David Rollinson: Resources, Writing - Review & Editing. Timothy J.C. Anderson: Conceptualization, Formal Analysis, Writing – Original Draft, Writing - Review & Editing, Supervision.

Competing interests

The authors declare no competing interests.

Corresponding authors

Correspondence to Roy N. Platt II or Timothy J. C. Anderson.

References

1. Taylor SA, Larson EL. Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. Nature ecology & evolution. 2019;3(2):170-7. Epub 2019/01/31. doi: 10.1038/s41559-018-0777-y. PubMed PMID: 30697003.
2. Aguillon SM, Dodge TO, Preising GA, Schumer M. Introgression. Current biology : CB. 2022;32(16):R865-r8. Epub 2022/08/24. doi: 10.1016/j.cub.2022.07.004. PubMed PMID: 35998591; PMCID: PMC10581619.
3. Mixão V, Gabaldón T. Hybridization and emergence of virulence in opportunistic human yeast pathogens. Yeast (Chichester, England). 2018;35(1):5-20. Epub 2017/07/07. doi: 10.1002/yea.3242. PubMed PMID: 28681409; PMCID: PMC5813172.
4. Depotter JR, Seidl MF, Wood TA, Thomma BP. Interspecific hybridization impacts host range and pathogenicity of filamentous microbes. Current opinion in microbiology. 2016;32:7-13. Epub 2016/04/27. doi: 10.1016/j.mib.2016.04.005. PubMed PMID: 27116367.

5. Ravazi A, Oliveira J, Madeira FF, Nunes GM, Reis YVD, Oliveira ABB, Azevedo LMS, Galvão C, Azeredo-Oliveira MTV, Rosa JAD, Alevi KCC. Climate and Environmental Changes and Their Potential Effects on the Dynamics of Chagas Disease: Hybridization in Rhodniini (Hemiptera, Triatominae). Insects. 2023;14(4). Epub 2023/04/27. doi: 10.3390/insects14040378. PubMed PMID: 37103193; PMCID: PMC10143345.
6. King KC, Stelkens RB, Webster JP, Smith DF, Brockhurst MA. Hybridization in Parasites: Consequences for Adaptive Evolution, Pathogenesis, and Public Health in a Changing World. PLoS pathogens. 2015;11(9):e1005098. Epub 2015/09/04. doi: 10.1371/journal.ppat.1005098. PubMed PMID: 26336070; PMCID: PMC4559376.
7. Colley DG, Bustinduy AL, Secor WE, King CH. Human schistosomiasis. Lancet (London, England). 2014;383(9936):2253-64. Epub 2014/04/05. doi: 10.1016/s0140-6736(13)61949-2. PubMed PMID: 24698483; PMCID: PMC4672382.
8. Webster BL, Diaw OT, Seye MM, Webster JP, Rollinson D. Introgressive hybridization of Schistosoma haematobium group species in Senegal: species barrier break down between ruminant and human schistosomes. PLoS Negl Trop Dis. 2013;7(4):e2110. Epub 2013/04/18. doi: 10.1371/journal.pntd.0002110. PubMed PMID: 23593513; PMCID: PMC3617179.
9. Webster JP, Neves MI, Webster BL, Pennance T, Rabone M, Gouvras AN, Allan F, Walker M, Rollinson D. Parasite Population Genetic Contributions to the Schistosomiasis Consortium for Operational Research and Evaluation within Sub-Saharan Africa. The American journal of tropical medicine and hygiene. 2020;103(1_Suppl):80-91. Epub 2020/05/14. doi: 10.4269/ajtmh.19-0827. PubMed PMID: 32400355; PMCID: PMC7351308 role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.
10. Léger E, Borlase A, Fall CB, Diouf ND, Diop SD, Yasenev L, Catalano S, Thiam CT, Ndiaye A, Emery A, Morrell A, Rabone M, Ndao M, Faye B, Rollinson D, Rudge JW, Sène M, Webster JP. Prevalence and distribution of schistosomiasis in human, livestock, and snail populations in northern Senegal: a One Health epidemiological study of a multi-host system. The Lancet Planetary health. 2020;4(8):e330-e42. Epub 2020/08/18. doi: 10.1016/s2542-5196(20)30129-7. PubMed PMID: 32800151; PMCID: PMC7443702.
11. Huyse T, Webster BL, Geldof S, Stothard JR, Diaw OT, Polman K, Rollinson D. Bidirectional introgressive hybridization between a cattle and human schistosome species. PLoS pathogens. 2009;5(9):e1000571. Epub 2009/09/05. doi: 10.1371/journal.ppat.1000571. PubMed PMID: 19730700; PMCID: PMC2731855.
12. Sene-Wade M, Marchand B, Rollinson D, Webster BL. Urogenital schistosomiasis and hybridization between Schistosoma haematobium and Schistosoma bovis in adults living in Richard-Toll, Senegal. Parasitology. 2018;145(13):1723-6. Epub 2018/09/07. doi: 10.1017/s0031182018001415. PubMed PMID: 30185248.
13. Díaz AV, Walker M, Webster JP. Reaching the World Health Organization elimination targets for schistosomiasis: the importance of a One Health perspective. Philosophical transactions of the Royal Society of London Series B, Biological sciences. 2023;378(1887):20220274. Epub 2023/08/21. doi: 10.1098/rstb.2022.0274. PubMed PMID: 37598697; PMCID: PMC10440173.
14. Platt RN, McDew-White M, Le Clec'h W, Chevalier FD, Allan F, Emery AM, Garba A, Hamidou AA, Ame SM, Webster JP, Rollinson D, Webster BL, Anderson TJC. Ancient Hybridization and Adaptive Introgression of an Invadolysin Gene in Schistosome Parasites. Molecular biology and evolution. 2019;36(10):2127-42. Epub 2019/06/30. doi: 10.1093/molbev/msz154. PubMed PMID: 31251352; PMCID: PMC6759076.
15. Rey O, Toulza E, Chaparro C, Allienne JF, Kincaid-Smith J, Mathieu-Begné E, Allan F, Rollinson D, Webster BL, Boissier J. Diverging patterns of introgression from Schistosoma bovis across S.

726 haematobium African lineages. PLoS pathogens. 2021;17(2):e1009313. Epub 2021/02/06. doi:
727 10.1371/journal.ppat.1009313. PubMed PMID: 33544762; PMCID: PMC7891765.
728 16. Djuikwo-Teukeng FF, Kouam Simo A, Allienne JF, Rey O, Njayou Ngapagna A, Tchuem-Tchuente
729 LA, Boissier J. Population genetic structure of Schistosoma bovis in Cameroon. Parasites & vectors.
730 2019;12(1):56. Epub 2019/01/27. doi: 10.1186/s13071-019-3307-0. PubMed PMID: 30678712; PMCID:
731 PMC6346511.
732 17. Oey H, Zakrzewski M, Gravermann K, Young ND, Korhonen PK, Gobert GN, Nawaratna S, Hasan
733 S, Martínez DM, You H, Lavin M, Jones MK, Ragan MA, Stoye J, Oleaga A, Emery AM, Webster BL,
734 Rollinson D, Gasser RB, McManus DP, Krause L. Whole-genome sequence of the bovine blood fluke
735 Schistosoma bovis supports interspecific hybridization with S. haematobium. PLoS pathogens.
736 2019;15(1):e1007513. Epub 2019/01/24. doi: 10.1371/journal.ppat.1007513. PubMed PMID: 30673782;
737 PMCID: PMC6361461.
738 18. Boon NAM, Mbow M, Paredis L, Moris P, Sy I, Maes T, Webster BL, Sacko M, Volckaert FAM,
739 Polman K, Huyse T. No barrier breakdown between human and cattle schistosome species in the
740 Senegal River Basin in the face of hybridisation. International journal for parasitology. 2019;49(13-
741 14):1039-48. Epub 2019/11/18. doi: 10.1016/j.ijpara.2019.08.004. PubMed PMID: 31734338.
742 19. Kincaid-Smith J, Tracey A, de Carvalho Augusto R, Bulla I, Holroyd N, Rognon A, Rey O, Chaparro
743 C, Oleaga A, Mas-Coma S, Allienne JF, Grunau C, Berriman M, Boissier J, Toulza E. Morphological and
744 genomic characterisation of the Schistosoma hybrid infecting humans in Europe reveals admixture
745 between Schistosoma haematobium and Schistosoma bovis. PLoS Negl Trop Dis. 2021;15(12):e0010062.
746 Epub 2021/12/24. doi: 10.1371/journal.pntd.0010062. PubMed PMID: 34941866; PMCID: PMC8741037.
747 20. Pennance T, Allan F, Emery A, Rabone M, Cable J, Garba AD, Hamidou AA, Webster JP, Rollinson
748 D, Webster BL. Interactions between Schistosoma haematobium group species and their Bulinus spp.
749 intermediate hosts along the Niger River Valley. Parasites & vectors. 2020;13(1):268. Epub 2020/05/26.
750 doi: 10.1186/s13071-020-04136-9. PubMed PMID: 32448268; PMCID: PMC7247258.
751 21. Stroehlein AJ, Korhonen PK, Lee VV, Ralph SA, Mentink-Kane M, You H, McManus DP, Tchuenté
752 LT, Stothard JR, Kaur P, Dudchenko O, Aiden EL, Yang B, Yang H, Emery AM, Webster BL, Brindley PJ,
753 Rollinson D, Chang BCH, Gasser RB, Young ND. Chromosome-level genome of Schistosoma
754 haematobium underpins genome-wide explorations of molecular variation. PLoS pathogens.
755 2022;18(2):e1010288. Epub 2022/02/16. doi: 10.1371/journal.ppat.1010288. PubMed PMID: 35167626;
756 PMCID: PMC8846543.
757 22. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure.
758 Evolution; international journal of organic evolution. 1984;38(6):1358-70. Epub 1984/11/01. doi:
759 10.1111/j.1558-5646.1984.tb05657.x. PubMed PMID: 28563791.
760 23. Bhatia G, Patterson N, Sankararaman S, Price AL. Estimating and interpreting FST: the impact of
761 rare variants. Genome Res. 2013;23(9):1514-21. Epub 2013/07/19. doi: 10.1101/gr.154831.113.
762 PubMed PMID: 23861382; PMCID: PMC3759727.
763 24. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated
764 individuals. Genome Res. 2009;19(9):1655-64. Epub 2009/08/04. doi: 10.1101/gr.094052.109. PubMed
765 PMID: 19648217; PMCID: PMC2752134.
766 25. Angora EK, Allienne JF, Rey O, Menan H, Touré AO, Coulibaly JT, Raso G, Yavo W, N'Goran EK,
767 Utzinger J, Balmer O, Boissier J. High prevalence of Schistosoma haematobium × Schistosoma bovis
768 hybrids in schoolchildren in Côte d'Ivoire. Parasitology. 2020;147(3):287-94. Epub 2019/11/16. doi:
769 10.1017/s0031182019001549. PubMed PMID: 31727202; PMCID: PMC10317610.
770 26. Maples BK, Gravel S, Kenny EE, Bustamante CD. RFMix: a discriminative modeling approach for
771 rapid and robust local-ancestry inference. American journal of human genetics. 2013;93(2):278-88. Epub
772 2013/08/06. doi: 10.1016/j.ajhg.2013.06.020. PubMed PMID: 23910464; PMCID: PMC3738819.

773 27. Martin SH, Van Belleghem SM. Exploring Evolutionary Relationships Across the Genome Using
774 Topology Weighting. Genetics. 2017;206(1):429-38. Epub 2017/03/28. doi:
775 10.1534/genetics.116.194720. PubMed PMID: 28341652; PMCID: PMC5419486.
776 28. Comparative genomics of the major parasitic worms. Nature genetics. 2019;51(1):163-74. Epub
777 2018/11/07. doi: 10.1038/s41588-018-0262-1. PubMed PMID: 30397333; PMCID: PMC6349046.
778 29. Moné H, Mouahid G, Morand S. The distribution of Schistosoma bovis Sonsino, 1876 in relation
779 to intermediate host mollusc-parasite relationships. Advances in parasitology. 1999;44:99-138. Epub
780 1999/11/24. doi: 10.1016/s0065-308x(08)60231-6. PubMed PMID: 10563396.
781 30. Pennance T, Ame SM, Amour AK, Suleiman KR, Allan F, Rollinson D, Webster BL. Occurrence of
782 Schistosoma bovis on Pemba Island, Zanzibar: implications for urogenital schistosomiasis transmission
783 monitoring. Parasitology. 2018;145(13):1727-31. Epub 2018/08/09. doi: 10.1017/s0031182018001154.
784 PubMed PMID: 30086805; PMCID: PMC7116046.
785 31. Stothard JR, Lockyer AE, Kabatereine NB, Tukahebwa EM, Kazibwe F, Rollinson D, Fenwick A.
786 Schistosoma bovis in western Uganda. Journal of helminthology. 2004;78(3):281-4. Epub 2004/10/08.
787 doi: 10.1079/joh2004239. PubMed PMID: 15469635.
788 32. Polack B, Mathieu-Bégné E, Vallée I, Rognon A, Fontaine JJ, Toulza E, Thomas M, Boissier J.
789 Experimental Infections Reveal Acquired Zoonotic Capacity of Human Schistosomiasis Trough
790 Hybridization. The Journal of infectious diseases. 2024;229(6):1904-8. Epub 2024/04/26. doi:
791 10.1093/infdis/jiae152. PubMed PMID: 38669235.
792 33. King CH, Muchiri EM, Ouma JH. Evidence against rapid emergence of praziquantel resistance in
793 Schistosoma haematobium, Kenya. Emerging infectious diseases. 2000;6(6):585-94. Epub 2000/11/15.
794 doi: 10.3201/eid0606.000606. PubMed PMID: 11076716; PMCID: PMC2640915.
795 34. Webster BL, Emery AM, Webster JP, Gouvras A, Garba A, Diaw O, Seye MM, Tchuente LA,
796 Simoonga C, Mwanga J, Lange C, Kariuki C, Mohammed KA, Stothard JR, Rollinson D. Genetic diversity
797 within Schistosoma haematobium: DNA barcoding reveals two distinct groups. PLoS Negl Trop Dis.
798 2012;6(10):e1882. Epub 2012/11/13. doi: 10.1371/journal.pntd.0001882. PubMed PMID: 23145200;
799 PMCID: PMC3493392.
800 35. Rollinson D, Stothard JR, Southgate VR. Interactions between intermediate snail hosts of the
801 genus Bulinus and schistosomes of the Schistosoma haematobium group. Parasitology. 2001;123
802 Suppl:S245-60. Epub 2002/01/05. doi: 10.1017/s0031182001008046. PubMed PMID: 11769287.
803 36. Brown DS. Freshwater snails of Africa and their medical importance: CRC press; 1994.
804 37. Taylor MG. Hybridisation experiments on five species of African schistosomes. Journal of
805 helminthology. 1970;44(3):253-314. Epub 1970/01/01. doi: 10.1017/s0022149x00021969. PubMed
806 PMID: 5505355.
807 38. Haldane JBJJog. Sex ratio and unisexual sterility in hybrid animals1922;12:101-9.
808 39. Harris K, Nielsen R. The Genetic Cost of Neanderthal Introgression. Genetics. 2016;203(2):881-
809 91. Epub 2016/04/03. doi: 10.1534/genetics.116.186890. PubMed PMID: 27038113; PMCID:
810 PMC4896200.
811 40. Harris TW, Arnaboldi V, Cain S, Chan J, Chen WJ, Cho J, Davis P, Gao S, Grove CA, Kishore R, Lee
812 RYN, Muller HM, Nakamura C, Nuin P, Paulini M, Raciti D, Rodgers FH, Russell M, Schindelman G, Auken
813 KV, Wang Q, Williams G, Wright AJ, Yook K, Howe KL, Schedl T, Stein L, Sternberg PW. WormBase: a
814 modern Model Organism Information Resource. Nucleic Acids Res. 2020;48(D1):D762-d7. Epub
815 2019/10/24. doi: 10.1093/nar/gkz920. PubMed PMID: 31642470; PMCID: PMC7145598.
816 41. Hambrook JR, Kaboré AL, Pila EA, Hanington PC. A metalloprotease produced by larval
817 Schistosoma mansoni facilitates infection establishment and maintenance in the snail host by interfering
818 with immune cell function. PLoS pathogens. 2018;14(10):e1007393. Epub 2018/10/30. doi:
819 10.1371/journal.ppat.1007393. PubMed PMID: 30372490; PMCID: PMC6224180.

820 42. Hambrook JR, Hanington PC. A cercarial invadolysin interferes with the host immune response
821 and facilitates infection establishment of Schistosoma mansoni. PLoS pathogens. 2023;19(2):e1010884.
822 Epub 2023/02/03. doi: 10.1371/journal.ppat.1010884. PubMed PMID: 36730464; PMCID: PMC9928134.
823 43. McHugh B, Krause SA, Yu B, Deans AM, Heasman S, McLaughlin P, Heck MM. Invadolysin: a
824 novel, conserved metalloprotease links mitotic structural rearrangements with cell migration. The
825 Journal of cell biology. 2004;167(4):673-86. Epub 2004/11/24. doi: 10.1083/jcb.200405155. PubMed
826 PMID: 15557119; PMCID: PMC2172566.
827 44. Fraïsse C, Sachdeva H. The rates of introgression and barriers to genetic exchange between
828 hybridizing species: sex chromosomes vs autosomes. Genetics. 2021;217(2). Epub 2021/03/17. doi:
829 10.1093/genetics/iyaa025. PubMed PMID: 33724409; PMCID: PMC8045713.
830 45. Criscione CD, Valentim CL, Hirai H, LoVerde PT, Anderson TJ. Genomic linkage map of the human
831 blood fluke Schistosoma mansoni. Genome Biol. 2009;10(6):R71. Epub 2009/07/02. doi: 10.1186/gb-
832 2009-10-6-r71. PubMed PMID: 19566921; PMCID: PMC2718505.
833 46. Berger DJ, Léger E, Sankaranarayanan G, Sène M, Diouf ND, Rabone M, Emery A, Allan F, Cotton
834 JA, Berriman M, Webster JP. Genomic evidence of contemporary hybridization between Schistosoma
835 species. PLoS pathogens. 2022;18(8):e1010706. Epub 2022/08/09. doi: 10.1371/journal.ppat.1010706.
836 PubMed PMID: 35939508; PMCID: PMC9387932.
837 47. Landeryou T, Rabone M, Allan F, Maddren R, Rollinson D, Webster BL, Tchuem-Tchuenté LA,
838 Anderson RM, Emery AM. Genome-wide insights into adaptive hybridisation across the Schistosoma
839 haematobium group in West and Central Africa. PLoS Negl Trop Dis. 2022;16(1):e0010088. Epub
840 2022/02/01. doi: 10.1371/journal.pntd.0010088. PubMed PMID: 35100291; PMCID: PMC8803156.
841 48. Emery AM, Allan FE, Rabone ME, Rollinson D. Schistosomiasis collection at NHM (SCAN).
842 Parasites & vectors. 2012;5:185. Epub 2012/09/05. doi: 10.1186/1756-3305-5-185. PubMed PMID:
843 22943137; PMCID: PMC3453491.
844 49. Le Clec'h W, Chevalier FD, McDew-White M, Allan F, Webster BL, Gouvras AN, Kinunghi S,
845 Tchuem Tchuenté LA, Garba A, Mohammed KA, Ame SM, Webster JP, Rollinson D, Emery AM, Anderson
846 TJC. Whole genome amplification and exome sequencing of archived schistosome miracidia.
847 Parasitology. 2018;145(13):1739-47. Epub 2018/05/29. doi: 10.1017/s0031182018000811. PubMed
848 PMID: 29806576; PMCID: PMC6193844.
849 50. Young ND, Jex AR, Li B, Liu S, Yang L, Xiong Z, Li Y, Cantacessi C, Hall RS, Xu X, Chen F, Wu X,
850 Zerlotini A, Oliveira G, Hofmann A, Zhang G, Fang X, Kang Y, Campbell BE, Loukas A, Ranganathan S,
851 Rollinson D, Rinaldi G, Brindley PJ, Yang H, Wang J, Wang J, Gasser RB. Whole-genome sequence of
852 Schistosoma haematobium. Nature genetics. 2012;44(2):221-5. Epub 2012/01/17. doi: 10.1038/ng.1065.
853 PubMed PMID: 22246508.
854 51. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data.
855 Bioinformatics. 2014;30(15):2114-20. Epub 2014/04/04. doi: 10.1093/bioinformatics/btu170. PubMed
856 PMID: 24695404; PMCID: PMC4103590.
857 52. Bushnell B. BBMap: a fast, accurate, splice-aware aligner. Lawrence Berkeley National
858 Lab.(LBNL), Berkeley, CA (United States), 2014.
859 53. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The
860 Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9. Epub 2009/06/10.
861 doi: 10.1093/bioinformatics/btp352. PubMed PMID: 19505943; PMCID: PMC2723002.
862 54. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D,
863 Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing
864 next-generation DNA sequencing data. Genome Res. 2010;20(9):1297-303. Epub 2010/07/21. doi:
865 10.1101/gr.107524.110. PubMed PMID: 20644199; PMCID: PMC2928508.
866 55. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G,
867 Marth GT, Sherry ST, McVean G, Durbin R. The variant call format and VCFtools. Bioinformatics.

2011;27(15):2156-8. Epub 2011/06/10. doi: 10.1093/bioinformatics/btr330. PubMed PMID: 21653522; PMCID: PMC3137218.
56. Browning BL, Tian X, Zhou Y, Browning SR. Fast two-stage phasing of large-scale sequence data. American journal of human genetics. 2021;108(10):1880-90. Epub 2021/09/04. doi: 10.1016/j.ajhg.2021.08.005. PubMed PMID: 34478634; PMCID: PMC8551421.
57. Garrison E, Kronenberg ZN, Dawson ET, Pedersen BS, Prins P. A spectrum of free software tools for processing the VCF variant call format: vcflib, bio-vcf, cyvcf2, hts-nim and slivar. PLoS Comput Biol. 2022;18(5):e1009123. Epub 2022/06/01. doi: 10.1371/journal.pcbi.1009123. PubMed PMID: 35639788; PMCID: PMC9286226.
58. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-MJB. Robust relationship inference in genome-wide association studies2010;26(22):2867-73.
59. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics. 2007;81(3):559-75. Epub 2007/08/19. doi: 10.1086/519795. PubMed PMID: 17701901; PMCID: PMC1950838.
60. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly. 2012;6(2):80-92. Epub 2012/06/26. doi: 10.4161/fly.19695. PubMed PMID: 22728672; PMCID: PMC3679285.
61. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg VJtJomLr. Scikit-learn: Machine learning in Python2011;12:2825-30.
62. Kodinariya TM, Makwana PRJIJ. Review on determining number of Cluster in K-Means Clustering2013;1(6):90-5.
63. Evanno G, Regnaut S, Goudet JJMe. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study2005;14(8):2611-20.
64. Miles A, R M, Ralph P, Kelleher J, Pisupati R, Rae S, Millar T. scikit-allel: v1.3.5. 2022.
65. Murphy BS. PyKrige: Development of a Kriging Toolkit for Python. 2014.
66. Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. Ancient admixture in human history. Genetics. 2012;192(3):1065-93. Epub 2012/09/11. doi: 10.1534/genetics.112.145037. PubMed PMID: 22960212; PMCID: PMC3522152.
67. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH, Hansen NF, Durand EY, Malaspinas AS, Jensen JD, Marques-Bonet T, Alkan C, Prüfer K, Meyer M, Burbano HA, Good JM, Schultz R, Aximu-Petri A, Butthof A, Höber B, Höffner B, Siegemund M, Weihmann A, Nusbaum C, Lander ES, Russ C, Novod N, Affourtit J, Egholm M, Verna C, Rudan P, Brajkovic D, Kucan Ž, Gušic I, Doronichev VB, Golovanova LV, Lalueza-Fox C, de la Rasilla M, Fortea J, Rosas A, Schmitz RW, Johnson PLF, Eichler EE, Falush D, Birney E, Mullikin JC, Slatkin M, Nielsen R, Kelso J, Lachmann M, Reich D, Pääbo S. A draft sequence of the Neandertal genome. Science (New York, NY). 2010;328(5979):710-22. Epub 2010/05/08. doi: 10.1126/science.1188021. PubMed PMID: 20448178; PMCID: PMC5100745.
68. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics. 2019;35(21):4453-5. Epub 2019/05/10. doi: 10.1093/bioinformatics/btz305. PubMed PMID: 31070718; PMCID: PMC6821337.
69. Junier T, Zdobnov EM. The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. Bioinformatics. 2010;26(13):1669-70. Epub 2010/05/18. doi: 10.1093/bioinformatics/btq243. PubMed PMID: 20472542; PMCID: PMC2887050.
70. Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll SA, Gaudet R, Schaffner SF, Lander ES, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng

C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Shen Y, Sun W, Wang H, Wang Y, Wang Y, Xiong X, Xu L, Waye MM, Tsui SK, Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallée C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PK, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PI, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Sham PC, Varilly P, Altshuler D, Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro HM, Qin ZS, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Tsunoda T, Johnson TA, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Zeng C, Zhao H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C, Royal CD, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Clayton EW, Watkin J, Gibbs RA, Belmont JW, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Wheeler DA, Yakub I, Gabriel SB, Onofrio RC, Richter DJ, Ziaugra L, Birren BW, Daly MJ, Altshuler D, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archevêque P, Bellemare G, Saeki K, Wang H, An D, Fu H, Li Q, Wang Z, Wang R, Holden AL, Brooks LD, McEwen JE, Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R, Stewart J. Genome-wide detection and characterization of positive selection in human populations. Nature. 2007;449(7164):913-8. Epub 2007/10/19. doi: 10.1038/nature06250. PubMed PMID: 17943131; PMCID: PMC2687721.

71. Szpiech Z. selscan 2.0: scanning for sweeps in unphased data. bioRxiv; 2021.

72. Chifman J, Kubatko L. Identifiability of the unrooted species tree topology under the coalescent model with time-reversible substitution processes, site-specific rate variation, and invariable sites. Journal of theoretical biology. 2015;374:35-47. Epub 2015/03/21. doi: 10.1016/j.jtbi.2015.03.006. PubMed PMID: 25791286.

73. Swofford DLJhpcfe. PAUP: phylogenetic analysis using parsimony (and other methods), version 4.0 beta2002.

74. Harrison RG, Larson EL. Hybridization, introgression, and the nature of species boundaries. The Journal of heredity. 2014;105 Suppl 1:795-809. Epub 2014/08/26. doi: 10.1093/jhered/esu033. PubMed PMID: 25149255.

75. Schumer M, Cui R, Powell DL, Rosenthal GG, Andolfatto P. Ancient hybridization and genomic stabilization in a swordtail fish. Molecular ecology. 2016;25(11):2661-79. Epub 2016/03/05. doi: 10.1111/mec.13602. PubMed PMID: 26937625.

76. Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, Li DZ. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 2020;21(1):241. Epub 2020/09/12. doi: 10.1186/s13059-020-02154-5. PubMed PMID: 32912315; PMCID: PMC7488116.

77. Alonge M, Lebeigle L, Kirsche M, Jenike K, Ou S, Aganezov S, Wang X, Lippman ZB, Schatz MC, Soyk S. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. Genome Biol. 2022;23(1):258. Epub 2022/12/16. doi: 10.1186/s13059-022-02823-7. PubMed PMID: 36522651; PMCID: PMC9753292.

962 78. Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, Sedlazeck FJ, Lippman ZB, Schatz MC.
963 RaGOO: fast and accurate reference-guided scaffolding of draft genomes. Genome Biol. 2019;20(1):224-.
964 doi: 10.1186/s13059-019-1829-6. PubMed PMID: 31661016.

965

966

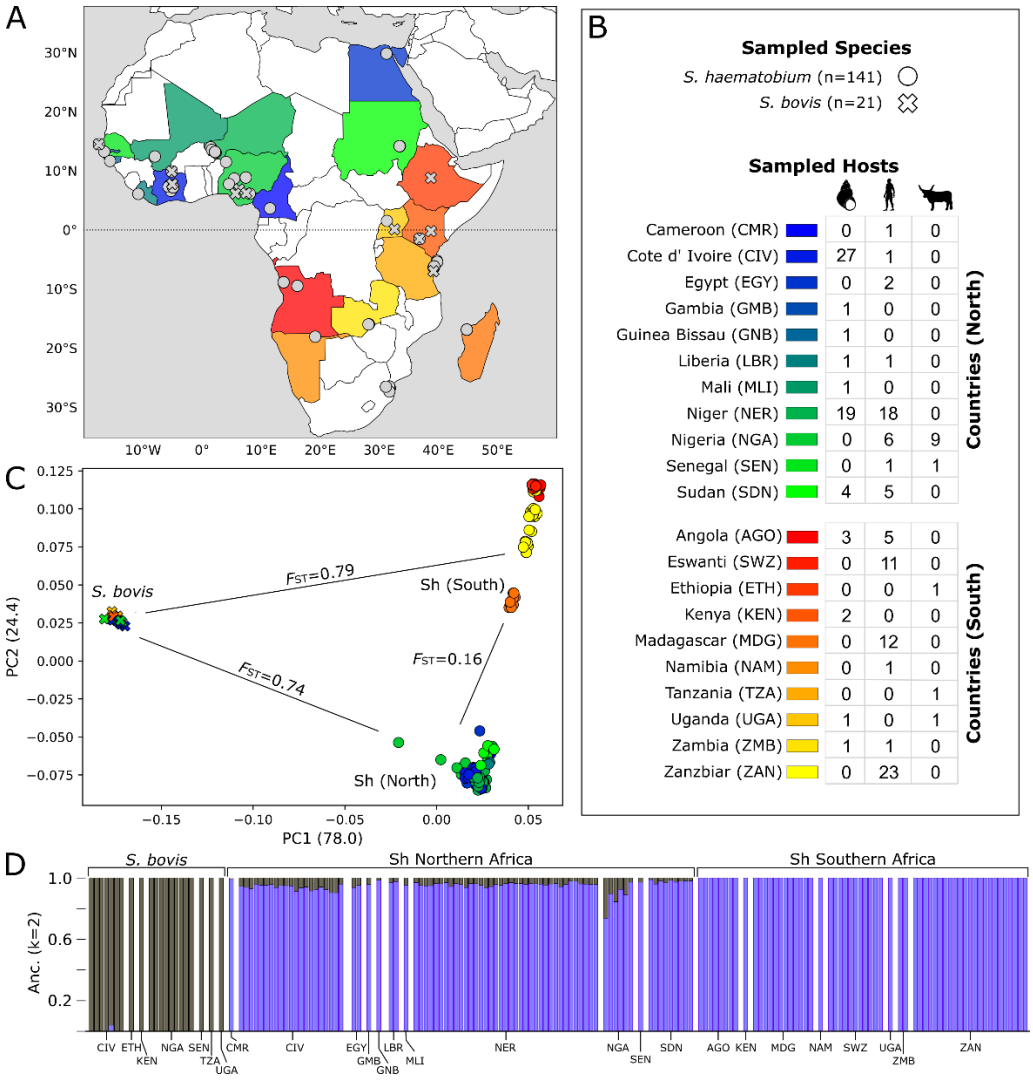


Figure 1. – Sampling localities, sample summary and the population structure of *Schistosoma haematobium* and *S. bovis*. (A) Collection locations for samples used in this study. Where exact coordinates for samples were not readily available we used the country capital as the collection locality. Two populations of *S. haematobium* were identified; northern and southern. The southern population in red-yellow and the northern population in blue to green. (B) A principal component analysis of 355,715 unlinked, common (MAF>0.05), autosomal variants. The three clusters correspond to *S. bovis,* and the northern and southern *S. haematobium* populations. Weighted, Weir-Cockerham F_{ST} values between these populations are shown.

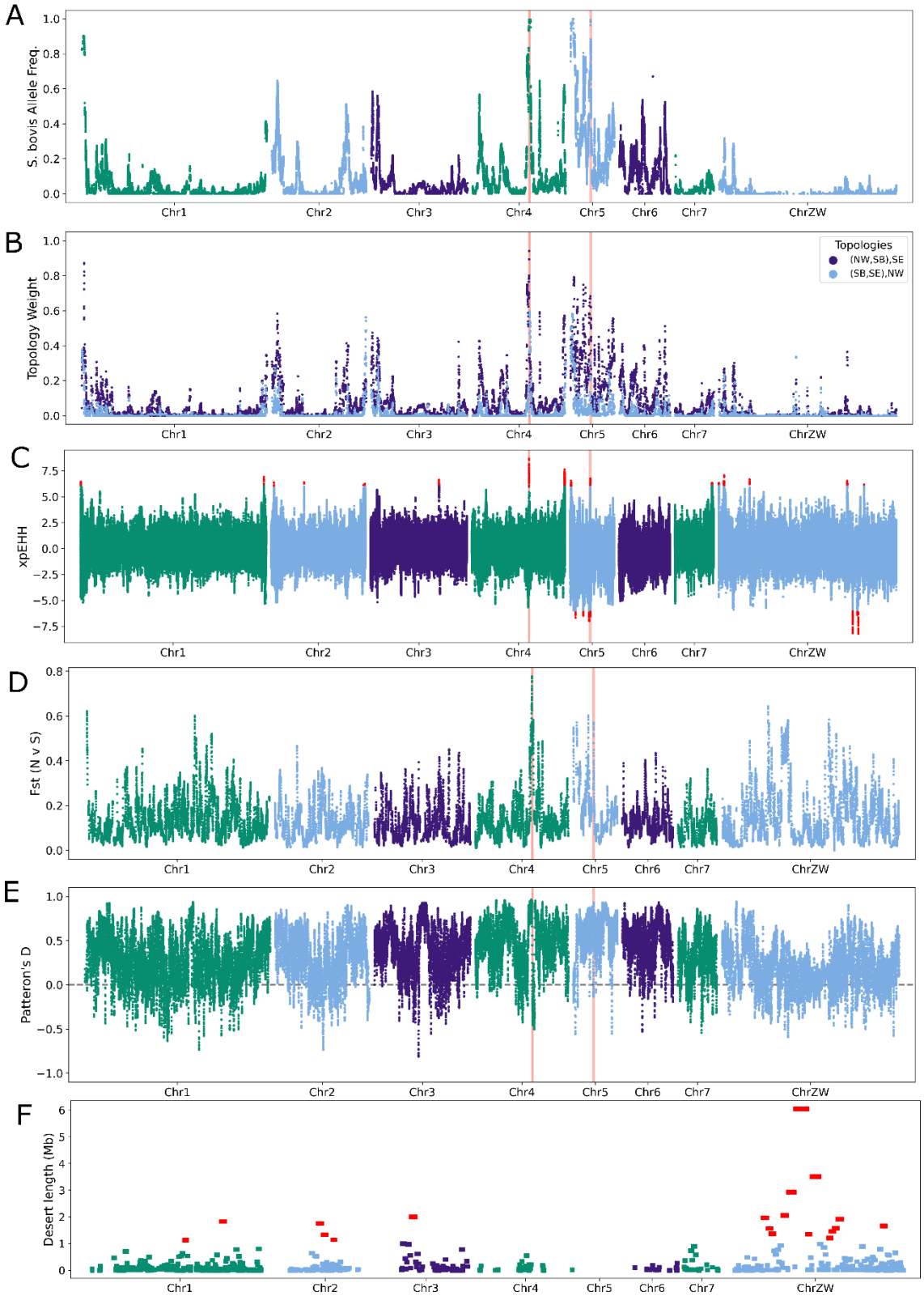


Figure 2. – Local measurements of differentiation, introgression and selection across the genome. (A) The weighted Weir-Cockerham fixation index (*F*ST) between northern and southern Africa, S. haematobium populations was measured across the genome in 10Kb windows. These results indicated multiple, highly differentiated regions between the two populations. (B) Patterson's *D* statistic was

measured to determine if high *F*ST regions were the result of *S. bovis* alleles present in northern populations. *D* measured across the genome was significantly positive indicating the presence of gene flow between S. bovis and north African *S. haematobium* populations. (C) The frequency of *S. bovis* ancestry across the genome in the northern *S. haematobium* population was estimated using RFmix. While the percentage of *S. bovis* alleles in the population are low overall, the *S. bovis* alleles are at or near fixation at loci on Chr4 and Chr5. (D) Gene tree topology weightings across the genome depicting the possible relationships between the northern and south *S. haematobium* populations and *S. bovis* using TWISST. Each locus across the genome is shown as stacked bar plots. While both tools use different methods to depict the relationships between these taxa they recover similar results. (E) Differential selection between *S. haematobium* populations was measured across the genome with extended haplotype homozygosity (xpEHH). Positive values indicate positive selection in the northern population and negative values indicate positive selection in the southern population. Significant xpEHH values ($p<0.05$) after multiple test correction are highlighted in red. (F) Multiple regions of the north African *S. haematobium* genome lacked introgressed *S. bovis* alleles. Regions that are longer than expected by chance are shown ($\text{Z-score}_{\text{Length}} > 3$) in red. These regions, also known as introgression deserts, where not randomly distributed across the genome, with 75% of them occurring on the sex chromosome. Results for FST and Patterson's D are shown after Gaussian smoothing (sigma=3). Pink vertical lines indicate putative regions of adaptive introgression.

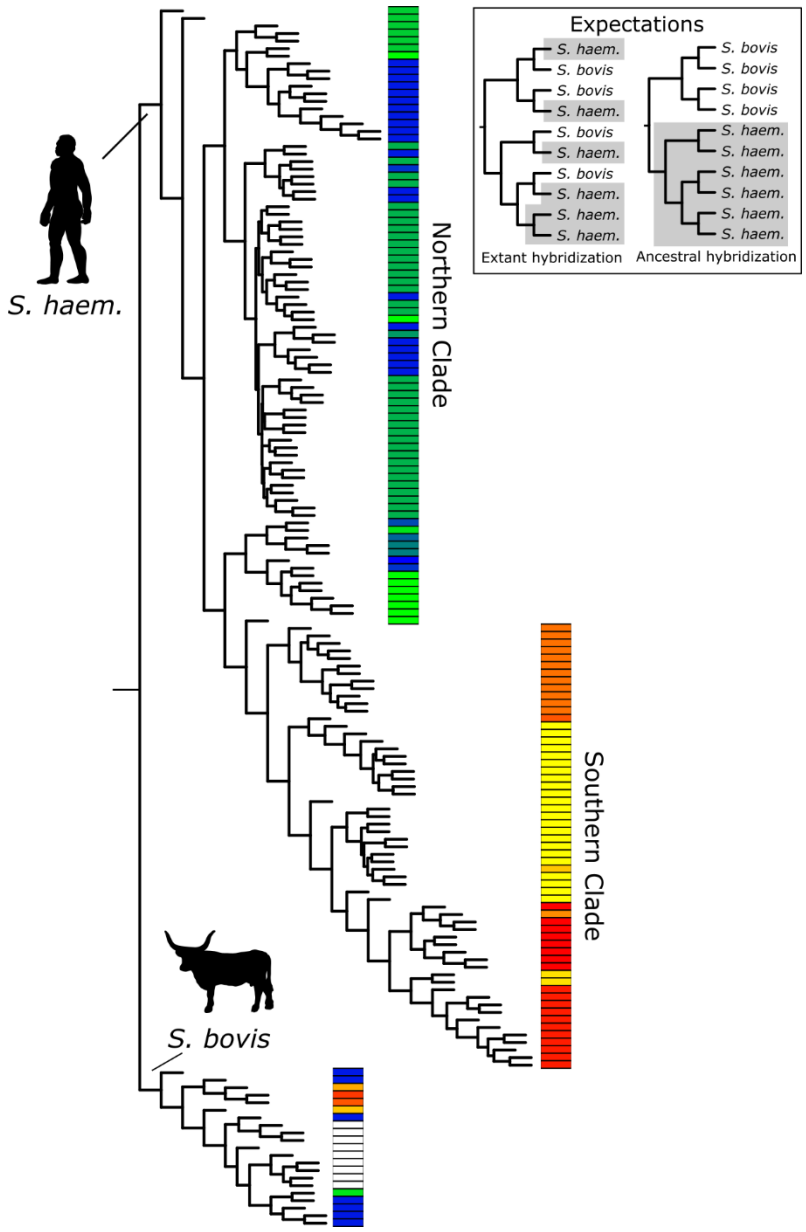


Figure 3 – Species tree of *S. haematobium* and *S. bovis* populations – SVDquartets species tree generated from autosomal SNVs. All nodes were supported by >95% of bootstrap replicates. Phylogenetic relationships between the species can be used to differentiate extant vs ancestral hybridization (inset). The tree shows that both *S. haematobium* and *S. bovis* are monophyletic. Biogeographic partitioning within the tree indicates that *S. haematobium* originated in northern Africa and expanded into southern Africa.

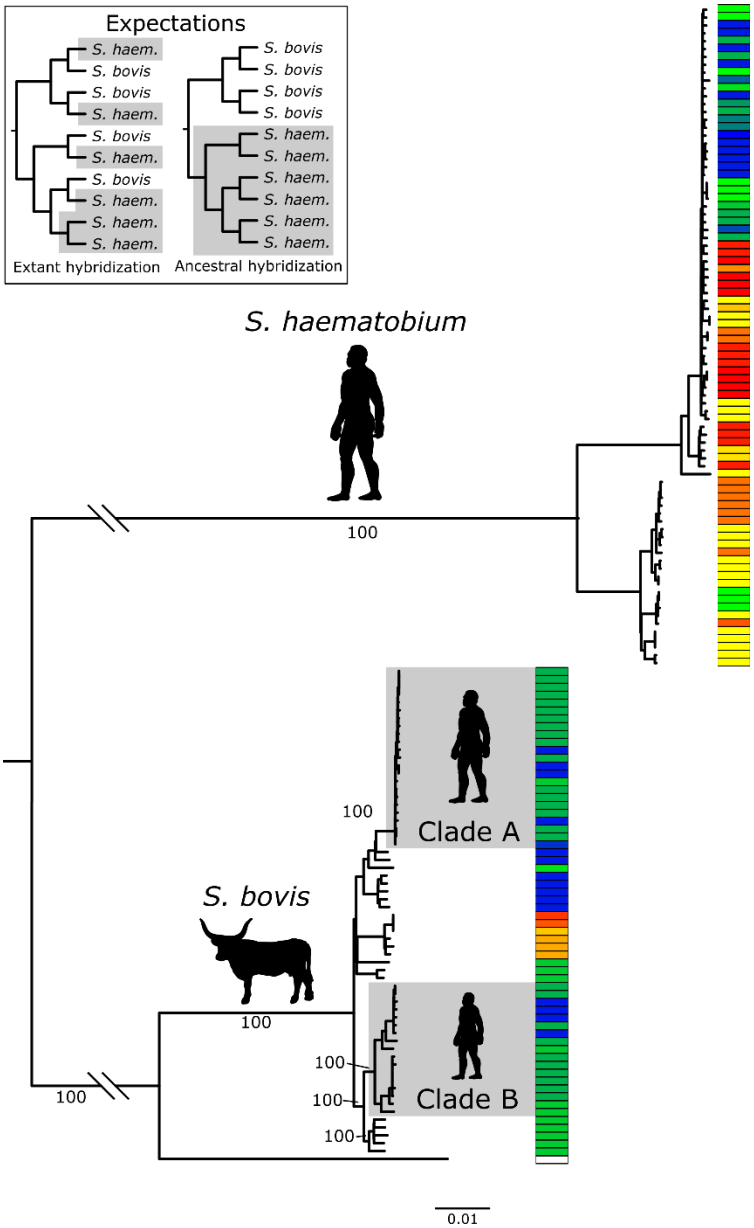


Figure 4. Mitochondrial tree of *S. haematobium* and *S. bovis* - A gene tree was recovered from mitochondrial genome assemblies from each sample. Bootstrap support at select nodes is shown. Phylogenetic relationships between the species can be used to differentiate extant vs ancestral hybridization (inset). Two well supported clades of *S. haematobium* contain an introgressed *S. bovis* mitotype, designated as “A” and “B”. Both the “A” and “B” clades contain samples from north Africa. All south African samples are found within a single clade of the remaining *S. haematobium* samples

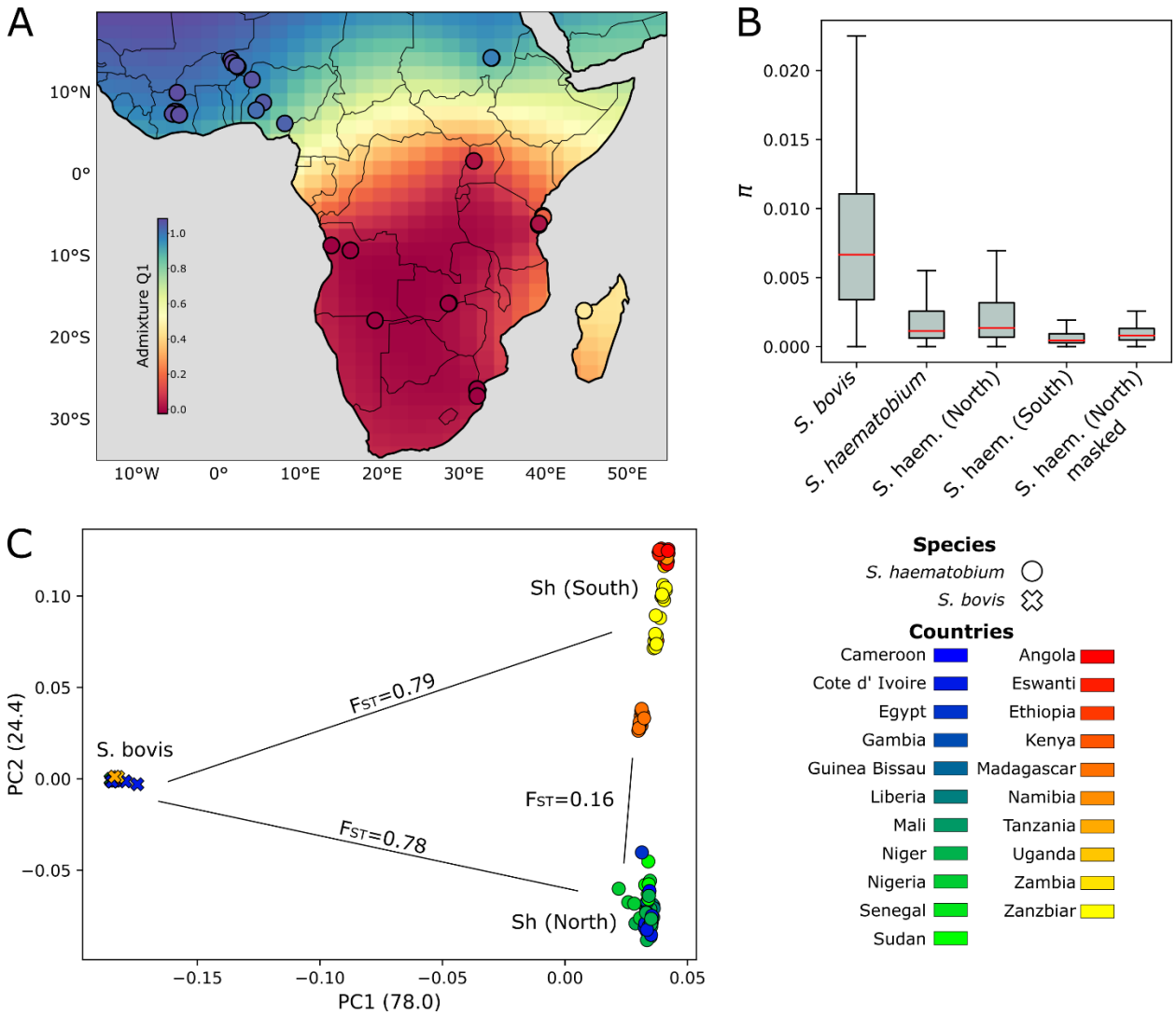
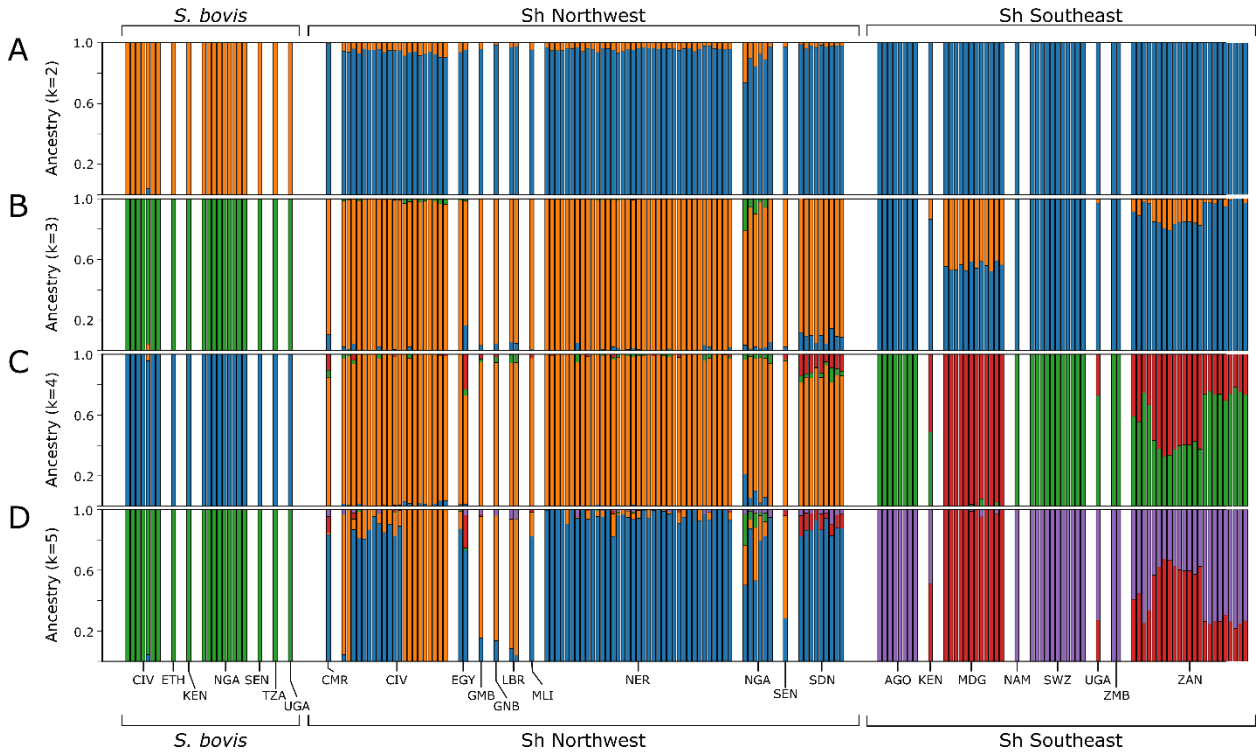
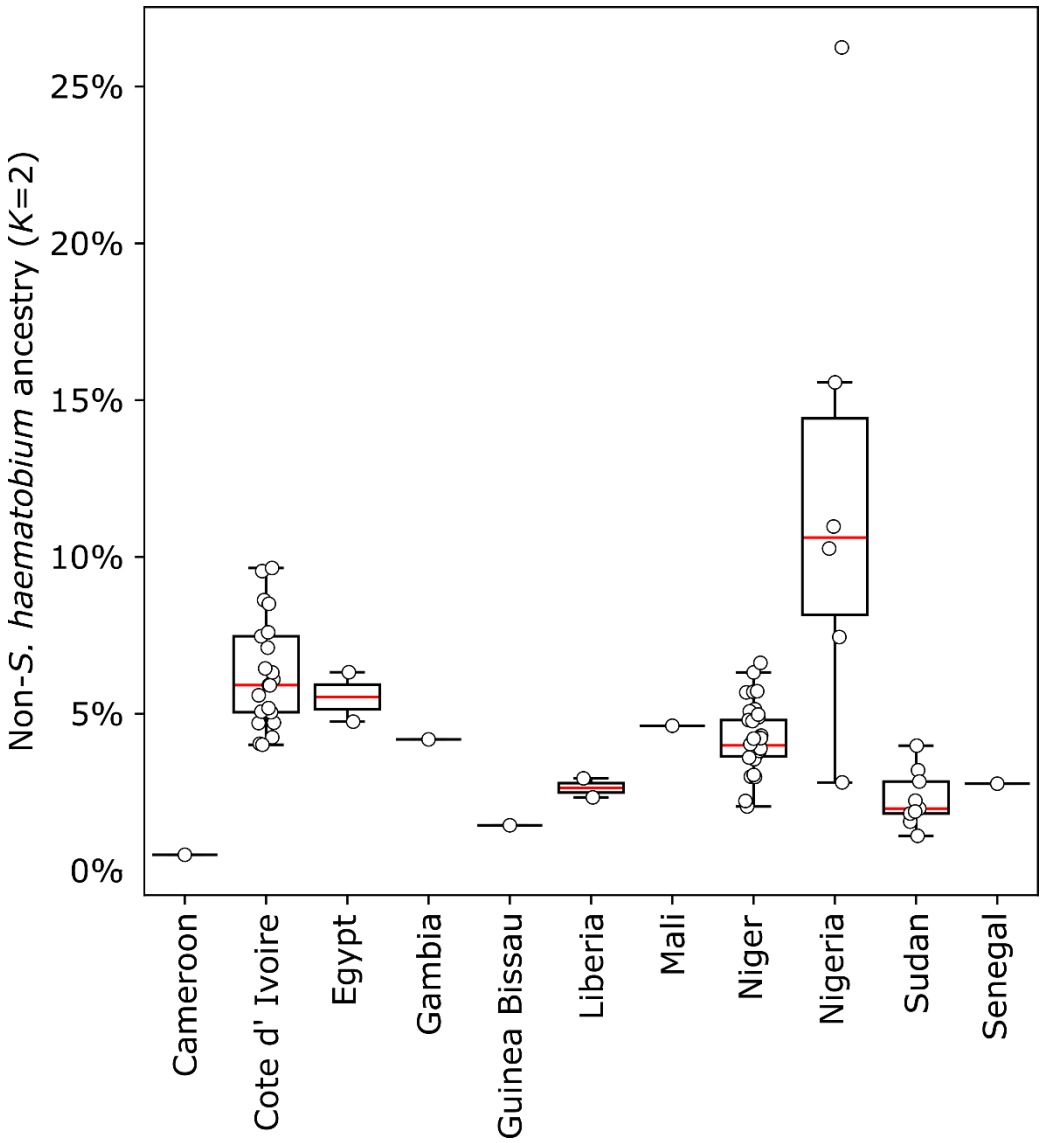


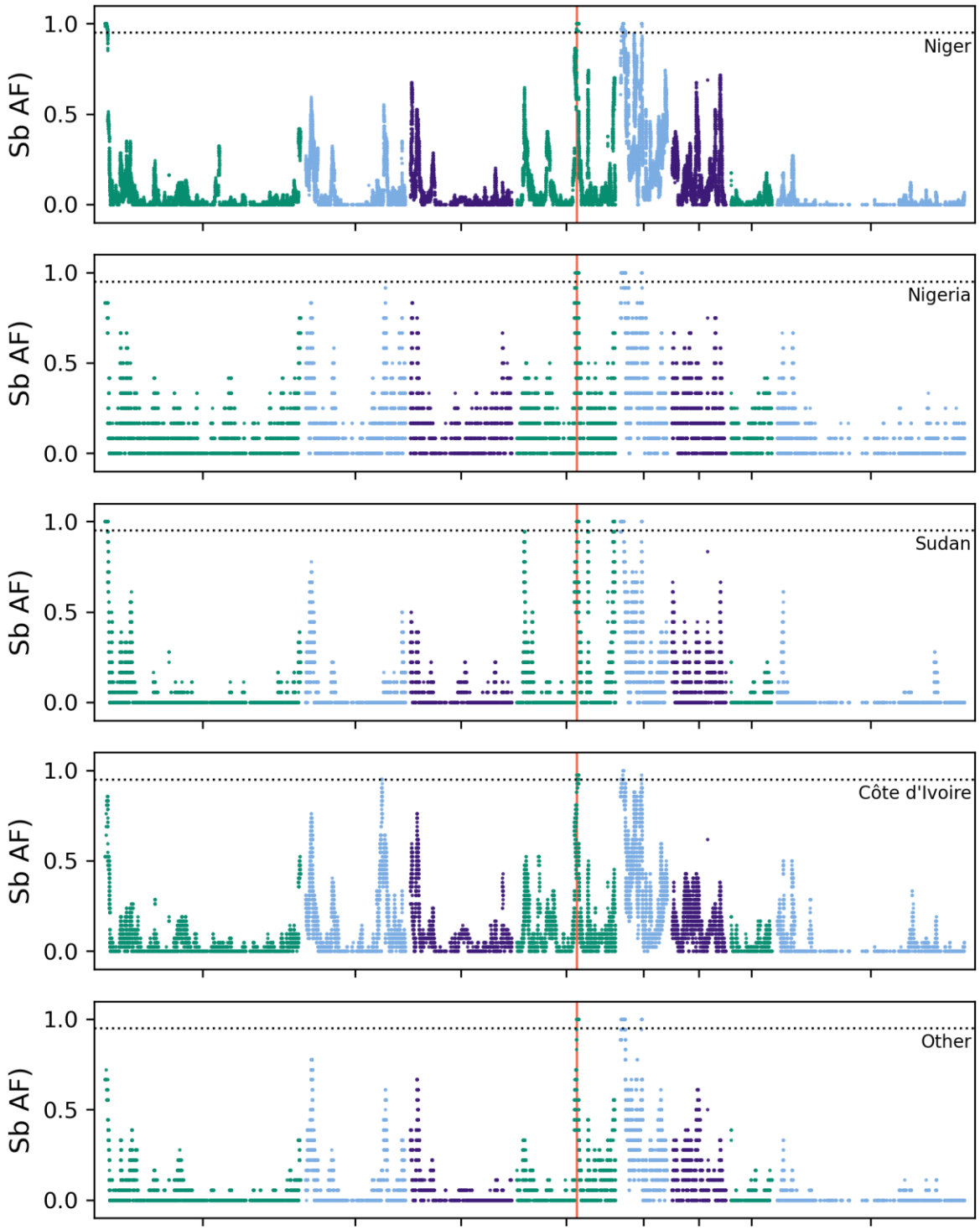
Figure 5 Biogeography of *S. haematobium* is not determined by introgressed *S. bovis* alleles. *S. haematobium* samples were split into two populations by PCA, Admixture and phylogenetic analyses. (A) We used Kriging interpolation to examine the distribution of these populations across Africa using the population component that differentiates the *S. haematobium* populations from one another (B) Nucleotide diversity (π) was calculated in 10kb sliding windows after masking introgressed *S. bovis* alleles present in the northern *S. haematobium population*. π is higher in the northern African *S. haematobium* compared to the southern population. When introgressed *S. bovis* alleles are masked, π is similar for both the southern and northern populations. (C) After masking introgressed *S. bovis* alleles the PCA is similar to Figure 1C. The similarity between the two PCAs show that the genetic differentiation between the northern and southern *S. haematobium* populations is not driven by introgressed *S. bovis* alleles.



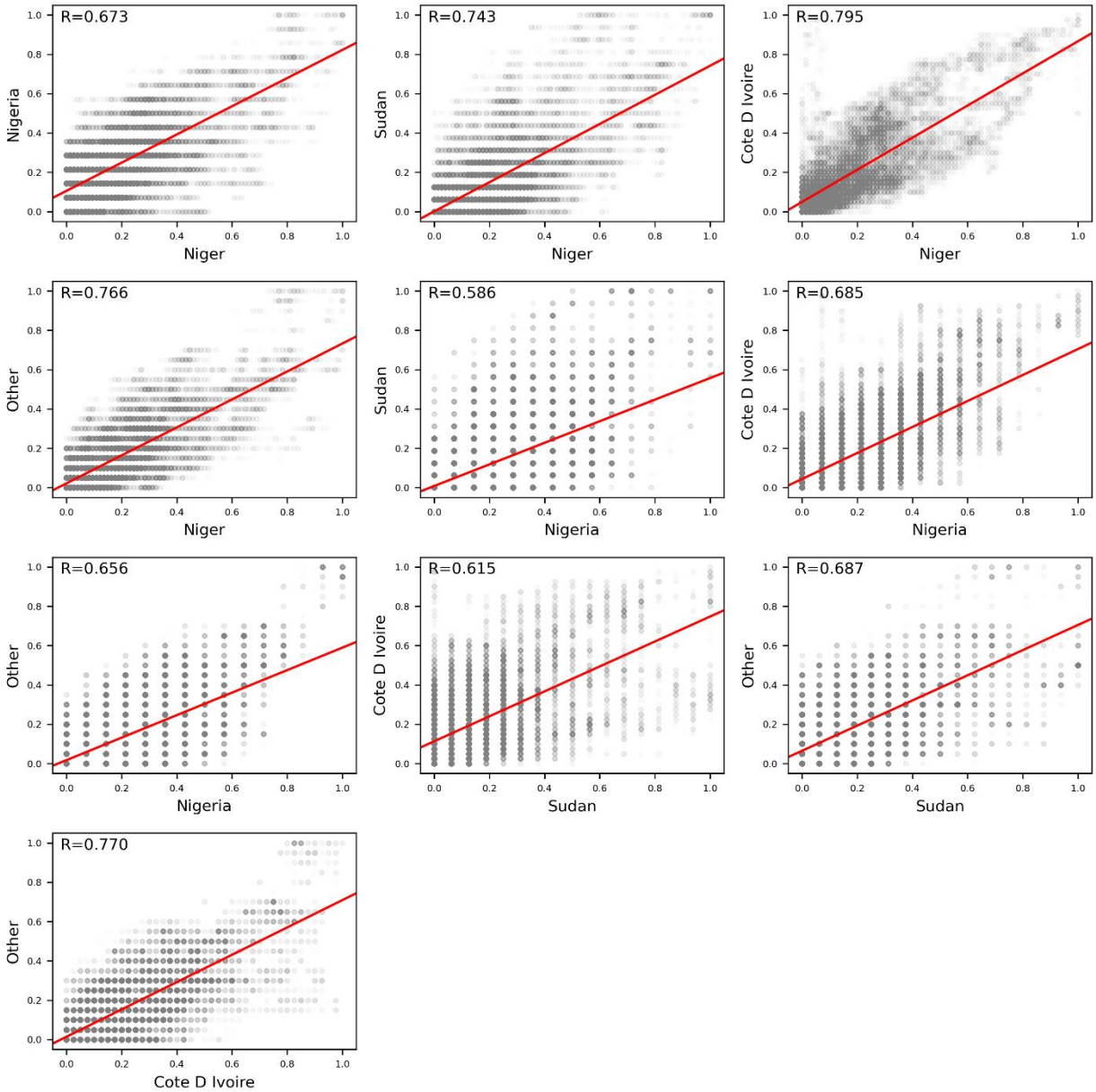
Supplemental Figure 1. Whole genome ancestry assignment with Admixture. We examined multiple different population(k) sizes with Admixture. (A) At k=3, *S. haematobium* and *S. bovis* were separated, and two general populations were identified within the *S. haematobium* samples corresponding to a northern and southern population. (B) An optimal number of populations (k=5) (Evanno *et al*. 2005) shows clear distinctions between the two *S. haematobium* populations. Finally, (C) we generated a reference panel of samples that maximized the *S. haematobium* and *S. bovis* population components from the k=3 results to run a supervised admixture analysis and assign samples as either *S. haematobium* or *S. bovis*. This analysis shows almost all of the northern *S. haematobium* samples contained low levels of *S. bovis* ancestry at k=2, but this percentage various as more population components are added. Country Codes are as follows: "AGO": Angola, "CMR": Cameroon, "CIV": Cote d' Ivoire, "EGY": Egypt, "SWZ": Eswanti, "ETH": Ethiopia, "GMB": Gambia, "GNB": Guinea Bissau, "KEN": Kenya, "LBR": Liberia, "MDG": Madagascar, "MLI": Mali, "NAM": Namibia, "NER": Niger, "NGA": Nigeria, "SEN": Senegal, "SDN": Sudan, "TZA": Tanzania, "UGA": Uganda, "ZMB": Zambia, "ZAN": Zanzibar.



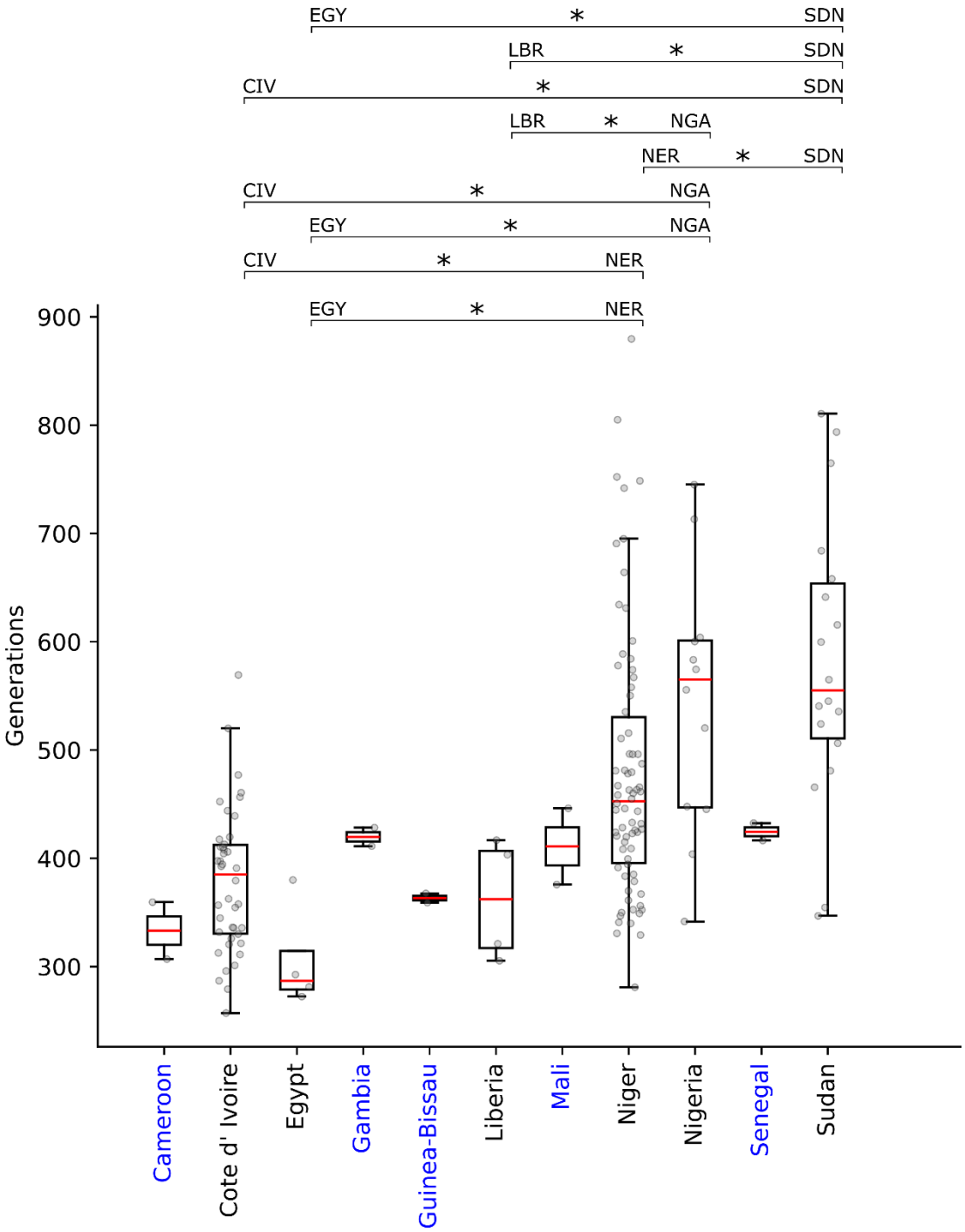
Supplemental Figure 2. Comparison of non-*S. haematobium* ancestry calculated from Admixture (k=2) in each north African sample – Ancestry of each sample was assigned to up to two different population components with Admixture. These two components were maximized in samples from southern Africa or S. bovis samples. By comparison, north African samples were a composite of these two populaitions with low, but varying levels fo the *S. bovis* compponent found in each individual. The population component corresponding to *S. bovis* ancestry was significanly higher in Nigerian that in other north African countries (Kruskal-Wallis H test statistic = 7.915, P-value = 0.0049).



Supplemental Figure 3. *S. bovis* allele frequency across the genome within *S. haematobium* samples from Northern African countries - The frequency of *S. bovis* ancestry across the genome is shown for each of the northwest African countries. In general, the distribution of *S. bovis* alleles is similar for each population. This consistency is an indicator of an ancient introgression event. The dotted line indicates 95% allele frequency.



Supplemental Figure 4. Pairwise comparison of introgressed *S. bovis* allele frequencies within northern *S. haematobium* samples by country – Introgressed *S. bovis allele* frequency is positivley correlated between countries. Pearson's correlation coefficient (*R*) is >0.586 in all comparisons. The correlation of introgressed allele freuqencies between populations up to 3,338 Km apart is expected under an ancient introgression scenario.



Supplemental Figure 5 – Estimated number of generations since admixture with *S. bovis*. We estimated the number of generations since admixture for each sample in the northern *S. haematobium* population by examining the length of introgressed *S. bovis* loci with in the genomes. Individual estimates for each sample for each country as grey points. Results from a one-way ANOVA indicated that age estimates varied significantly between countries. Countries with a single individual, two haplotypes, are shown in blue and were not included in the ANOVA analyses. A "*" indicates p-values < 0.05. Differences in ages may indicate multiple introgression events.

Table1. Genes containing outlier loci.

Chrom	Location	Gene_ID	Gene Name
Chr4	NC_067199.1:28466881-28497752	MS3_00007802	Leishmanolysin-like peptidase
Chr4	NC_067199.1:28497929-28529268	MS3_00007803	MS3_00007803
Chr4	NC_067199.1:28531133-28546611	MS3_00007804	RAD50
Chr4	NC_067199.1:28562068-28562732	MS3_00000457	JMJD6_1
Chr4	NC_067199.1:28571061-28634110	MS3_00010935	JMJD6_4
Chr4	NC_067199.1:28662329-28782419	MS3_00010934	JMJD6_3
Chr4	NC_067199.1:28742276-28747614	MS3_00007805	TY3BI_12
Chr4	NC_067199.1:28785057-28816546	MS3_00010936	JMJD6_5
Chr5	NC_067200.1:9933321-9974793	MS3_00011123	MS3_00011123
Chr5	NC_067200.1:9989479-10043873	MS3_00011124	AK2_3
Chr5	NC_067200.1:10117557-10118821	MS3_00011125	MS3_00011125
Chr5	NC_067200.1:10187316-10209172	MS3_00011126	TSC2
Chr5	NC_067200.1:10219534-10229257	MS3_00009120	MDP1_1
Chr5	NC_067200.1:10245731-10246840	MS3_00000691	MS3_00000691
Chr5	NC_067200.1:10414911-10585361	MS3_00011127	MS3_00011127