Epidemiology and Global Health

Strong isolation by distance and evidence of population microstructure reflect ongoing *Plasmodium falciparum* transmission in Zanzibar

Reviewed Preprint

Revised by authors after peer review.

About eLife's process

Reviewed preprint version 2
April 26, 2024 (this version)

Reviewed preprint version 1
November 17, 2023

Sent for peer review
August 2, 2023

Posted to preprint server
June 13, 2023

Sean V. Connelly, Nicholas F. Brazeau, Mwinyi Msellem, Billy E. Ngasala, Özkan Aydemir, Varun Goel, Karamoko Niaré, David J. Giesbrecht, Zachary R. Popkin-Hall, Christopher M. Hennelly, Zackary Park, Ann M. Moormann, John Michael Ong'echa, Robert Verity, Safia Mohammed, Shija J. Shija, Lwidiko E. Mhamilawa, Ulrika Morris, Andreas Mårtensson, Jessica T. Lin, Anders Björkman, Jonathan J. Juliano, Jeffrey A. Bailey

MD-PhD Program, University of North Carolina, Chapel Hill, NC 27599 • Research Division, Ministry of Health, Zanzibar, Tanzania • Department of Parasitology and Medical Entomology, Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania • Global Health and Migration Unit, Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden • Department of Medicine, University of Massachusetts Chan Medical School, Worcester, MA • Carolina Population Center, University of North Carolina, Chapel Hill, NC 27599 • Department of Pathology and Laboratory Medicine, Brown University, Providence, RI, 02912 USA • Institute for Global Health and Infectious Diseases, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599 USA • Division of Infectious Diseases, Department of Medicine, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599 USA • Center for Global Health Research, Kenyan Medical Research Institute, Kisumu, Kenya • MRC Centre for Global Infectious Disease Analysis, Imperial College, London • Zanzibar Malaria Elimination Program (ZAMEP), Zanzibar, Tanzania • Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, 17177 Stockholm, Sweden • Department of Global Public Health, Karolinska Institutet, Stockholm, Sweden • Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, 27599 USA • Curriculum in Genetics and Molecular Biology, University of North Carolina, Chapel Hill, NC 27599 USA

Abstract

The Zanzibar archipelago of Tanzania has become a low-transmission area for *Plasmodium falciparum.* Despite being considered an area of pre-elimination for years, achieving elimination has been difficult, likely due to a combination of imported infections from mainland Tanzania, and continued local transmission. To shed light on these sources of transmission, we applied highly multiplexed genotyping utilizing molecular inversion probes to characterize the genetic relatedness of 282 *P. falciparum* isolates collected across Zanzibar and in Bagamoyo District on the coastal mainland from 2016-2018. Overall, parasite populations on the coastal mainland and Zanzibar archipelago remain highly related. However, parasite isolates from Zanzibar exhibit population microstructure due to rapid decay of parasite relatedness over very short distances. This, along with highly related pairs within *shehias*, suggests ongoing low level local transmission. We also identified highly related parasites across *shehias* that reflect human mobility on the main island of Unguja and identified a cluster of highly related parasites, suggestive of an outbreak, in the Micheweni district on Pemba island. Parasites in asymptomatic infections demonstrated higher complexity of infection than those in symptomatic infections, but have similar core genomes.

Our data support importation as a main source of genetic diversity and contribution to the parasite population on Zanzibar, but they also show local outbreak clusters where targeted interventions are essential to block local transmission. These results highlight the need for preventive measures against imported malaria and enhanced control measures in areas that remain receptive for malaria reemergence due to susceptible hosts and competent vectors.

eLife assessment

Connelly and colleagues provide **convincing** genetic evidence that importation from mainland Tanzania is a major source of Plasmodium falciparum lineages currently circulating in Zanzibar. This study also reveals ongoing local malaria transmission and occasional near-clonal outbreaks in Zanzibar. Overall, the manuscript effectively highlights the role of human movements in maintaining residual malaria transmission in an area targeted for intensive control interventions over the past decades and provides clear and **valuable** information for epidemiologists and public health professionals.

Introduction

Malaria cases in Tanzania comprise 3% of globally reported cases, but transmission is heterogeneous, with the coastal mainland witnessing declining but substantial transmission of *Plasmodium falciparum* (Alegana et al., 2021; World Health Organization, 2022). On the other hand, the archipelago of Zanzibar is a pre-elimination setting, with low level seasonal transmission (Björkman et al., 2019). This is largely due to the routine implementation of a combination of effective control measures, including robust vector control and routine access to effective antimalarials (Björkman et al., 2019). Despite these efforts, malaria has been difficult to eliminate from the archipelago. There are several reasons this may be the case: (1) frequent importation of malaria from moderate or high transmission regions of mainland Tanzania and Kenya (Björkman et al., 2019; Le Menach et al., 2011; Lipner et al., 2011; Monroe et al., 2019; Morgan et al., 2020; Tatem et al., 2009); (2) ongoing local transmission due to residual vector capacity despite strong vector control (Björkman et al., 2019); and (3) a reservoir of asymptomatic infections (Björkman & Morris, 2020; Björkman et al., 2019).

Parasite genomics has the potential to help us better understand malaria epidemiology by uncovering population structure and gene flow, providing insight into changes in the parasite population including how parasites move between regions (Neafsey et al., 2021). Genomics has previously been used to study importation and transmission chains in other low transmission settings in Africa and elsewhere (H. H. Chang et al., 2019; Fola et al., 2023; Morgan et al., 2020; Patel et al., 2014; Roh et al., 2019; Sane et al., 2019). Previously, we had investigated the importation of malaria into Zanzibar from the mainland using whole genome sequencing, showing highly similar populations within the mainland and within the archipelago, but also identifying highly related parasite pairs between locations suggesting a role for importation (Morgan et al., 2020). However, this work lacked sufficient samples to assess transmission of parasites within Zanzibar. The larger and spatially rich sample set analyzed in this manuscript offers an opportunity for more refined analyses of transmission across Zanzibar and how parasites are related to those from coastal mainland.

A panel of molecular inversion probes (MIPs), a highly multiplexed genotyping assay, were designed in a previous study to target single nucleotide polymorphisms (SNPs) throughout the *P. falciparum* genome (Aydemir et al., 2018). We leveraged this assay to investigate the genetic

epidemiology of parasites in the coastal mainland and Zanzibar utilizing 391 samples collected from cross-sectional surveys of both asymptomatic infections and symptomatic, uncomplicated malaria cases during 2016-2018. Specifically, we use identity by descent analyses to compare the genetic relatedness of mainland and Zanzibari parasites, and to investigate the geography/spatial relationships of genetically related parasites on the archipelago. We further characterize how the genetic complexity of infections differ by clinical status and describe patterns of antimalarial drug resistance polymorphisms in the parasite populations. In this low transmission setting, these analyses characterize fine-scale local parasite populations that contribute to continued transmission within the region, highlighting a key barrier to malaria elimination in the Zanzibar archipelago.

Methods

Samples from coastal Tanzania (178) and Zanzibar (213) were previously sequenced through multiple studies (**Table 1**, **Supplemental Figure 1**). These samples include 213 dried blood spots (DBS) collected in Zanzibar between February 2016 and September 2017, coming from cross-sectional surveys of asymptomatic individuals (n = 70) and an *in vivo* efficacy study of artesunate-amodiaquine (ASAQ) with single low dose primaquine (SLDP) in pediatric uncomplicated malaria patients in the western and central districts of Unguja island and Micheweni District on Pemba island (n = 143) (Msellem et al., 2020). These samples were geolocalized to *shehias*, the lowest geographic governmental designation of land in Zanzibar, across its two main islands, Unguja and the northern region of Pemba (**Supplemental Figure 1**). Mainland Tanzania samples were collected in rural Bagamoyo District, where malaria transmission persists, and residents frequently travel to Dar es Salaam, the major port from where travelers depart for Zanzibar. Of the mainland Bagamoyo samples, 138 were whole blood collected from 2015-2017 as part of an *in vivo* efficacy study of artemether-lumefantrine (AL) in pediatric uncomplicated malaria patients (Topazian et al., 2022), and the remaining 40 samples were leukodepleted blood collected in 2018 from asymptomatic but RDT-positive children who participated in a study investigating the transmission of *P. falciparum* to colony reared mosquitos. This project leveraged molecular inversion probe (MIP) data from SRA including PRJNA926345, PRJNA454490, PRJNA545345, and PRJNA545347.

In order to place coastal Tanzanian and Zanzibari samples in the context of African *P. falciparum* population structure across multiple regions, MIP data from 147 whole blood samples collected in Ahero District, Kenya from the same parasite clearance study were used (Topazian et al., 2022) in conjunction with a subset of data from 2,537 samples genotyped for a study of the 2013 Demographic Health Survey of the Democratic Republic of the Congo which included samples from DRC, Ghana, Tanzania, Uganda and Zambia (Verity et al., 2020) (see **Supplemental Figure 2**).

Molecular Inversion Probe (MIP) Sequencing

Sequence data for the coastal Tanzanian and Zanzibari samples were generated in a similar fashion across studies. Chelex extracted DNA from DBS and Qiagen Miniprep (Qiagen, Germantown, MD) extracted DNA from leukodepleted blood were used in MIP captures, which were then sequenced as previously described (Aydemir et al., 2018; Verity et al., 2020). Control mixtures of 4 strains of genomic DNA from *P. falciparum* laboratory lines were also sequenced as described previously (Verity et al., 2020). We utilized two MIP panels, one being a genome-wide single nucleotide polymorphism (SNP) MIP panel and the second being a panel with the known drug resistance markers in *Plasmodium falciparum* (Verity et al., 2020). These libraries were sequenced on Illumina Nextseq 500 instrument using 150lJbp paired end sequencing with dual indexing using Nextseq 500/550 Mid-output Kit v2.

Description	Location (District)	Dates	Clinical status†	Sample Size	Age Range (yr)	# in Genome Wide Analysis	# in Drug Resistance Analysis
Community cross-sectional surveys	Zanzibar (Multiple)	2016	A	70	2-70	21	52
in vivo efficacy study of artesunate-amodiaquine (ASAQ) with single low dose primaquine (SLDP) in pediatric uncomplicated malaria patients	Zanzibar (Multiple)	2017	S	143	2-60	117	134
Study of transmission of Plasmodium falciparum to colony reared mosquitos	Mainland (Bagamoyo)	2018	A	40	7-16	34	0
Parasite clearance study of artemether-lumefantrine (AL)	Mainland (Bagamoyo)	2018	S	138	2-11	110	123

†Asymptomatic (A) or Symptomatic (S).

Table 1.

Blood samples from Zanzibar and coastal Tanzania.

MIP variant Calling and Filtering

MIP sequencing data was processed using *MIPTools* (*https://github.com/bailey-lab/MIPTools*), which first merges reads and removes errors and Unique Molecular Identifier (UMI) redundancy with *MIPWrangler* (Aydemir, unpublished). For the genome-wide panel, variant calling was performed using *FreeBayes* within *MIPTools*, for a pooled continuous sample that was filtered for a minimum UMI depth of 10, a within sample allele frequency threshold of 0.01 and a minimum alternate read count of 2 to obtain 5174 variant SNP sites. Utilizing *bcftools* (version 1.15.1), the samples and loci were filtered to only the known targeted SNPs, requiring a minor allele frequency threshold of 0.01, a sample missingness threshold of 10% and loci missingness threshold of 15%. After filtering and subsetting to biallelic sites, 282 samples were left at 1270 loci. The final numbers of samples used for analysis by group are shown in **Table 1**. Sequencing coverage estimates for loci are shown in **Supplemental Figure 3**.

For the drug resistance panel, variant calling was performed as above, with additional *FreeBayes* parameters of a haplotype length of 3 and using the 30 best alleles at a given locus. Three aggregate amino acid summary tables were created with reference amino acid UMI counts, alternate amino acid UMI counts and the coverage depth for each variant, for a total of 309 samples at 2265 SNPs. We focused analysis on the following key known and putative drug resistance molecular marker genes and corresponding mutations: *P. falciparum* (*Pf*) chloroquine resistance transporter (Pfcrt: C72S, M74I, N75E, K76T, T93S, H97Y, F145I, I218F, A220S, Q271E, N326S, M343L, C350R, G353V, I356T, R371I*), Pf* multidrug resistance 1 (Pfmdr1: N86Y, Y184F, S1034C, N1042D, D1246Y), *Pf* dihydrofolate reductase (Pfdhfr: A16V, N51I, C59R, S108N, I164L), *Pf* dihydropteroate synthase (Pfdhps: S436A, S436F, A437G, K540E, A581G, A613T, A613S), *Pf cytochrome b* (Pfcytb: Y268N,Y268S,Y268C) and *Pf* kelch 13 (Pfk13: P441L, F446I, G449A, N458Y, C469F, C469Y, M476I, A481V, Y493H, R515K, P527H, N537I, N537D, G538V, R539T, I543T, P553L, R561H, V568G, P574L, C580Y, R622I, A675V) (World Health Organization, 2020). Prevalence was calculated separately in Zanzibar or mainland Tanzania for each polymorphism by the number of samples with alternative genotype calls for this polymorphism over the total number of samples genotyped and an exact 95% confidence interval using the Pearson-Klopper method was calculated for each prevalence.

Analysis of Population Relatedness and Structure

To investigate genetic relatedness of parasites across regions, identity by descent (IBD) estimates were assessed using the within sample major alleles (coercing samples to monoclonal by calling the dominant allele at each locus) and estimated utilizing a maximum likelihood approach using the *inbreeding_mle* function from the *MIPanalyzer* package (Verity et al., 2020). This approach has previously been validated as a conservative estimate of IBD (Verity et al., 2020). Next, Principal Component Analysis (PCA) was performed to query the comparative genetic variation of the samples through utilizing the genome-wide SNP panel. We pruned 51 samples that had a pairwise IBD of greater than 0.90 to one randomly selected sample as a representative of the clonal population to avoid clonal structure from dominating the analysis. Within-sample allele frequencies were calculated, with an imputation step replacing missing values with the median per each locus, and PCA was performed using the *prcomp* function (Verity et al., 2020) (*R* version 4.2.1).

To include geographic information with querying genetic variation, Discriminant Analysis of Principal Components (DAPC) analysis was used (Jombart, 2008). Pseudohaplotypes were created by pruning the genotype calls at all loci for each sample into a single haplotype, and redundant haplotypes were removed (282 reduced to 272 with unique pseudohaplotypes). DAPC was conducted at the district level, and samples from districts with less than five samples (272 samples to 270 samples in 6 districts) were retained (**Supplemental Figure 4B**). For the main DAPC analysis (**Figure 1B**), highly related isolates were pruned to a single representative

infection (272 reduced to 232) and then included districts with at least 5 samples (232 reduced to 228 samples in 5 districts). The DAPC was performed using the *adegenet* package (Jombart, 2008) with the first 80 PCs based on the cross-validation function *xvalDapc*. To perform K-means clustering, a cluster K of 1 was assigned to the mainland samples while the *kmeans* package was used to find the optimal K to cluster the Zanzibar *shehias* by latitude and longitude. The K-means clustering experiment was used to cluster a continuous space of geographic coordinates in order to compare genetic relatedness in different regions. We selected K = 4 as the inflection point based on the elbow plot (**Supplemental Figure 5**) and based the number to obtain sufficient subsections of Zanzibar to compare genetic relatedness.

To investigate how genetic relatedness varies as distance between pairs increases, isolation by distance was performed across all of Zanzibar and within the islands of Unguja and Pemba. The greater circle (GC) distances between each *shehia* centroid were calculated (within *shehia* distances were equal to 0) to find the distance between each *shehia* in geographic space. Distances were then binned at increments reflecting the max GC distances between regions, which was smallest in Pemba at 12km and much larger for both Unguja (58km) and all of Zanzibar (135km). Within each binned group, mean IBD with 95% Cis are plotted. For graphing IBD connections at the between and within *shehia* level, an IBD threshold of 0.25 (half-siblings) or greater was used (**Figure 4, Supplemental Figure 6, Supplemental Figure 7**). In graphing IBD connections at larger distances between islands or between coastal mainland Tanzania and Zanzibar, a between IBD value of 0.125 (quarter-siblings) or greater was used (**Supplemental Figure 8, Supplemental Figure 9**). These plots were created utilizing *ggraph* in *R* with the nodes being samples and the edges being IBD estimates.

Complexity of Infection (COI), or the number of parasite clones in a given sample, was determined using *THE REAL McCOIL* (v2) categorical method (H.-H. Chang et al., 2017) and the 95% CI was calculated utilizing a nonparametric bootstrap. Fws statistic, which is used to compare the diversity within and between samples in a population, was calculated in *R* version 4.2.1 through the formula, **(1- H_w)/H_p**,, where **H_wg:296:1040** is the within-sample heterozygosity and **H_p** is the heterozygosity across the population, an 95% CIs were calculated utilizing a nonparametric bootstrap. The mean prevalence of antimalarial drug resistance polymorphisms and 95% CIs were calculated using a nonparametric bootstrap method.

Results

Zanzibari falciparum parasites were closely related to coastal mainland parasites but showed higher within than between population IBD and evidence of microstructure on the archipelago

To examine geographic relatedness, we first used principal component analysis (PCA). Zanzibari parasites are highly related to other parasites from East Africa and more distantly related to Central and West African isolates (**Supplemental Figure 2**). PCA analysis of 232 coastal Tanzanian and Zanzibari isolates, after pruning 51 samples with an IBD of greater than 0.9 to one representative sample, demonstrates little population differentiation (**Figure 1A**).

However, after performing K-means clustering of *shehias* in Zanzibar and mainland Tanzania, parasites within each population show more highly related pairs within their respective clusters than between clusters (**Figure 2**). Comparisons of parasite pairs between Zanzibar and coastal Tanzania showed no pairs with an IBD greater than 0.20 (**Figure 2, Supplemental Figure 8**). Similarly, no pairs with an IBD of 0.20 or greater were present in pairwise comparisons between Unguja and Pemba (**Supplemental Figure 9**).

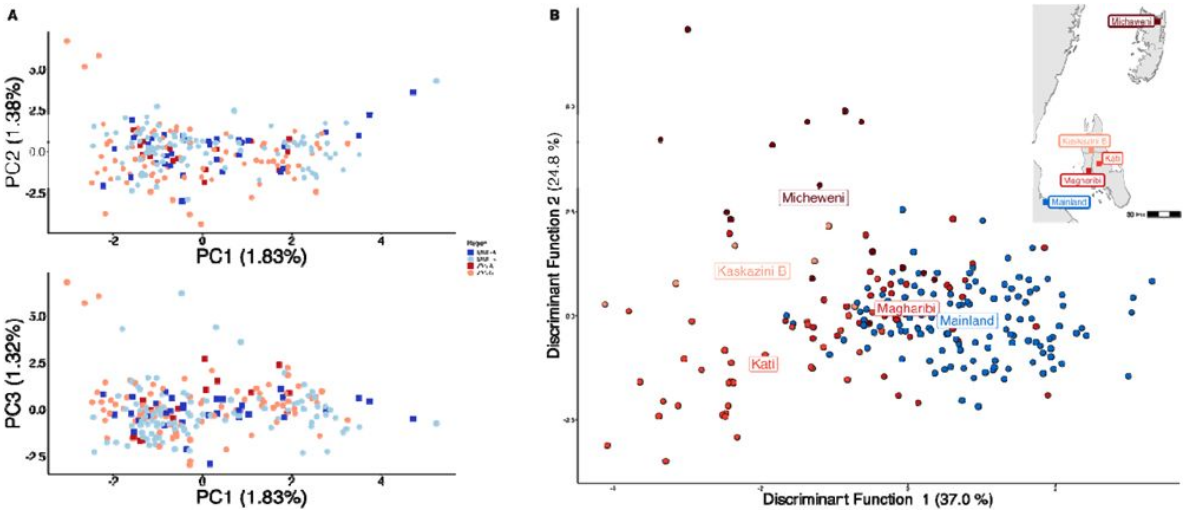

Figure 1.

Parasites between Zanzibar and coastal mainland Tanzania are highly related but microstructure within Zanzibar is apparent.

A) Principal Component Analysis (PCA) comparing parasites from symptomatic vs. asymptomatic patients from coastal Tanzania and Zanzibar. Clusters with an identity by descent (IBD) value of greater than 0.90 were limited to a single representative infection to prevent local-structure of highly related isolates within *shehias* from driving clustering. **B)** A Discriminant Analysis of Principal Components (DAPC) was performed utilizing isolates with unique pseudohaplotypes, pruning highly related isolates to a single representative infection. Districts were included with at least 5 isolates remaining to have sufficient samples for the DAPC. For plotting the inset map, the district coordinates (*e.g.* Mainland, Kati, etc.) are calculated from the averages of the *shehia* centroids within each district.

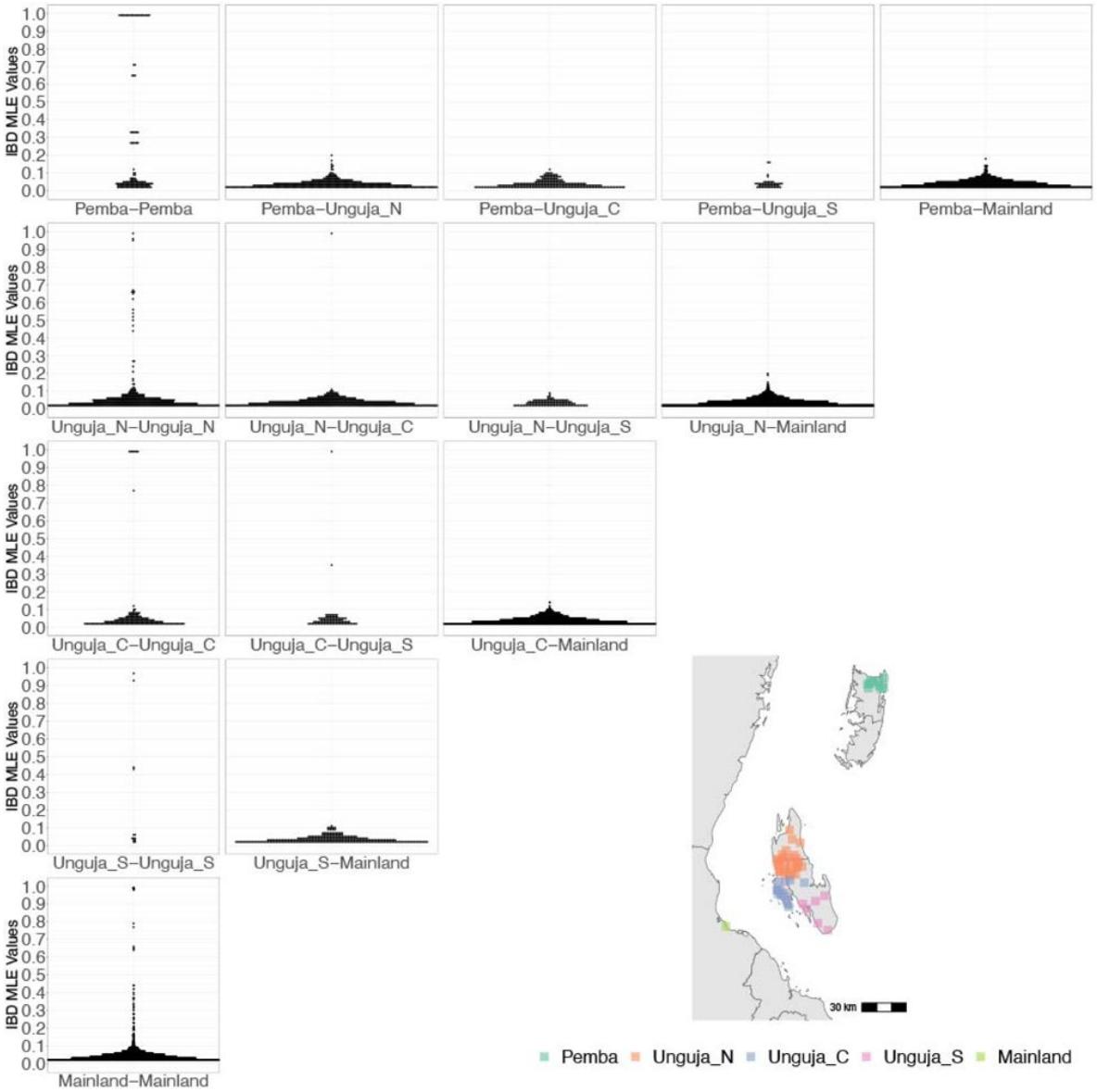

Figure 2.

Coastal Tanzania and Zanzibari parasites have more highly related pairs within their given region than between regions.

K-means clustering of *shehia* coordinates was performed using geographic coordinates all *shehias* present from the sample population to generate 5 clusters (colored boxes). All shehias were included to assay pairwise IBD between differences throughout Zanzibar. K-means cluster assignments were converted into interpretable geographic names Pemba, Unguja North (Unguja_N), Unguja Central (Unguja_C), Unguja South (Unguja_S) and mainland Tanzania (Mainland). Pairwise comparisons of within cluster IBD (column 1 of IBD distribution plots) and between cluster IBD (column 2-5 of IBD distribution plots) was done for all clusters. All IBD values greater than 0 were plotted for each comparison. In general, within cluster IBD had more pairwise comparisons containing high IBD identity.

To further assess the differentiation within the parasite population in Zanzibar, we conducted Discriminant Analysis of Principal Components (DAPC) according to the districts of origin for each isolate. Parasites differentiated geographically, with less variation near the port of Zanzibar town and more differentiation in isolates collected in districts further from the port (**Figure 1B**). This underlying microstructure is also supported by classic isolation by distance analysis (**Figure 3**, **Supplemental Figure 10**). Isolation by distance analysis across all of Zanzibar and within Unguja showed rapid decay of relatedness over very short geographic distances (**Figure 3A** and **3B**). Interestingly in Pemba, mean IBD remained at a similar relatively high level even at longer distances (**Figure 3C**).

Within Zanzibar, parasite clones are shared within and between shehias suggesting local outbreaks

Among the sample pairs in Zanzibar that are highly related (IBD of 0.25 or greater), we see different patterns of genetic relatedness suggesting common local and short distance transmission of clones and occasional long distance transmission (**Figure 4**). In Unguja (**Figure 4A**), we see multiple identical or near identical parasite pairs shared over longer distances, suggesting longer distance gene flow, as well as multiple *shehias* containing highly related pairs. In Northern Pemba, there is one large cluster of highly related parasites shared within and between six *shehias* (**Figure 4B**). Network analysis (**Figure 4C**) for all sample pairs with an IBD of greater than 0.25 from these *shehias* illustrates this, with pairs linked by yellow lines showing the highest IBD. The largest network represents two highly related clusters (groups linked by yellow edges, mean IBD of 0.99) connected by a highly related intermediate (FMH42), suggesting that the clusters are related though parasites that have recombined while on the archipelago. FMH42 links the lower cluster with pairwise IBD of 0.65 and the upper cluster with a pairwise IBD of 0.27. These symptomatic isolates collected from February 2016 to September 2017 in northern Pemba likely derive from sustained transmission from a seeding event.

Network analysis of within *shehia* pairwise IBD sharing in Unguja again shows that there is close relatedness on this small geographic scale (**Supplemental Figure 6**). A cluster of four isolates in the Shakani *shehia* on Unguja island with pairwise IBDs of 0.99 likely reflects ongoing transmission within Shakani, with similar connections in Bambi and Dimani. Meanwhile, a few distant connections likely reflect the extent of human mobility on the island (**Figure 4A**). Similar within district networks on the mainland are shown in **Supplemental Figure** 7.

Compared to symptomatic infections, asymptomatic infections demonstrate greater genetic complexity, especially in coastal Tanzania

Asymptomatic infections were compared to roughly contemporaneously collected isolates from those presenting with acute, uncomplicated malaria. Asymptomatic infections demonstrated greater COI than symptomatic infections, on both the coastal mainland (mean COI 2.5 vs. 1.7, $p < 0.05$, Wilcoxon-Mann-Whitney test) and in Zanzibar (mean COI 2.2 vs. 1.7, $p = 0.05$, Wilcoxon-Mann-Whitney test) (**Figure 5A**). A similar pattern was seen when evaluating Fws, which measures the diversity within a sample compared to the population, with lower Fws in asymptomatic samples consistent with higher within host complexity, with a more pronounced difference on the mainland (**Figure 5B**). Despite these differences, parasites from asymptomatic and symptomatic infections tended to cluster together in PCA analysis, suggesting their core genomes are genetically similar and do not vary based on clinical status (**Figure 1A**).

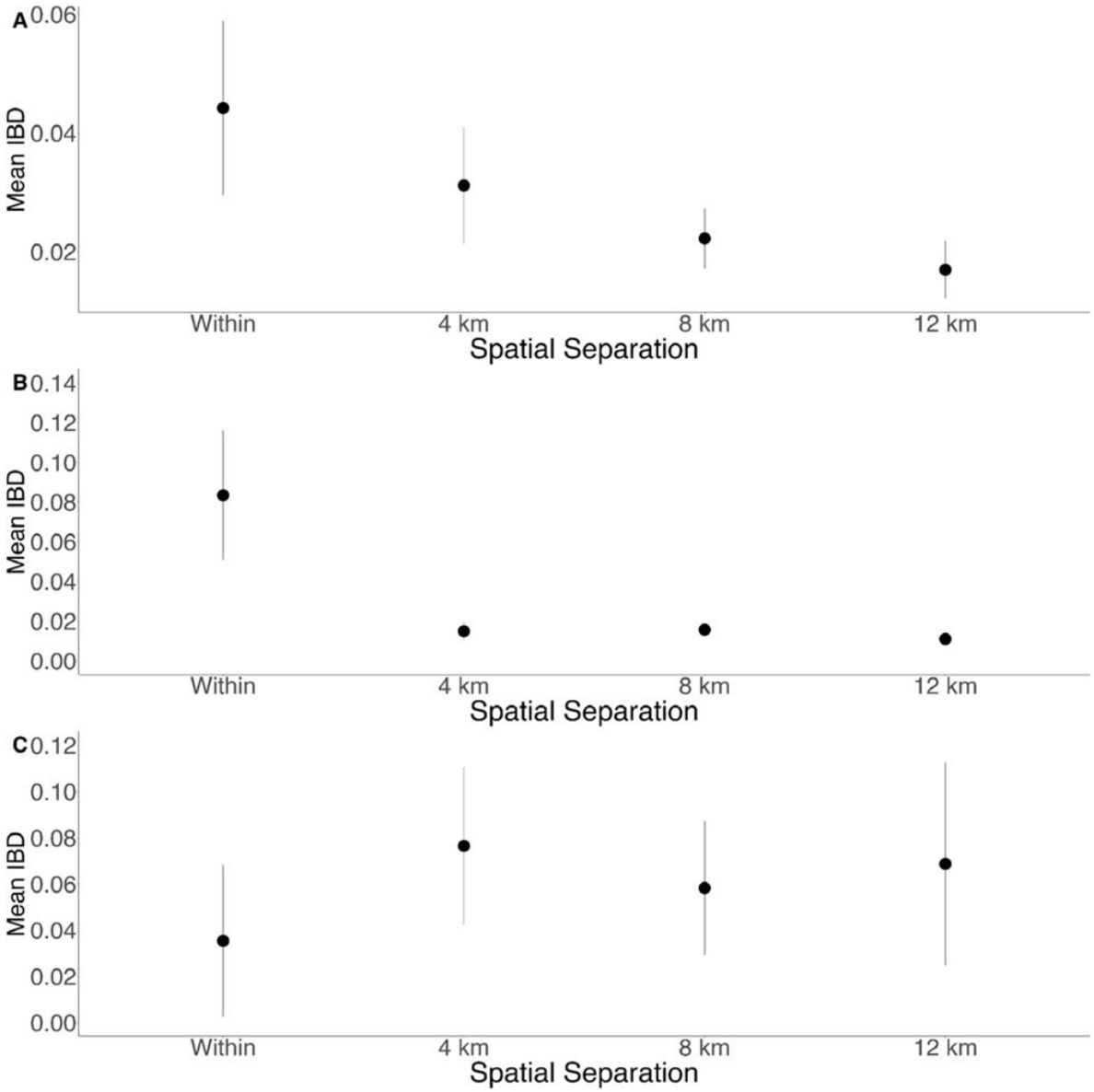

Figure 3.

Isolation by distance is shown between all Zanzibari parasites

(A), only Unguja parasites **(B) and only Pemba parasites (C).** Samples were analyzed based on geographic location, Zanzibar (N=136) (A), Unguja (N=105) (B) or Pemba (N=31) (C) and greater circle (GC) distances between pairs of parasite isolates were calculated based on *shehia* centroid coordinates. These distances were binned at 4km increments out to 12 km. IBD beyond 12km is shown in **Supplemental** Figure 8. The maximum GC distance for all of Zanzibar was 135km, 58km on Unguja and 12km on Pemba. The mean IBD and 95% CI is plotted for each bin.

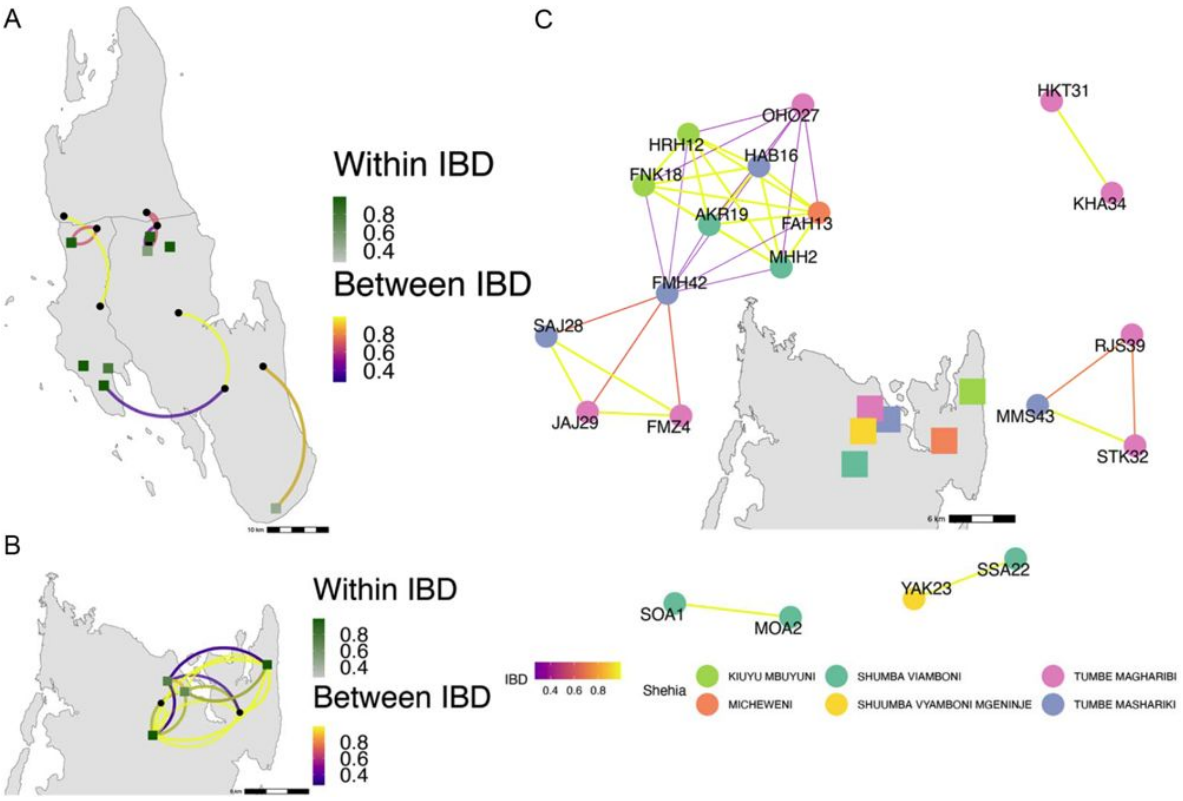

Figure 4.

Highly related pairs span long distances across Zanzibar.

Sample pairs were filtered to have IBD estimates of 0.25 or greater. Within *shehia* pairwise IBD estimates are shown within Unguja **(Panel A)** and Pemba **(Panel B)** as single points, with dark orange representing the greatest degree of IBD. *Shehias* labeled with black dots do not have within IBD estimates of .25 or greater. Between *shehia* IBD reflects pairs of parasites with IBD greater than or equal to 0.25, with the color of the connecting arc representing the degree of IBD and yellow representing maximal connectivity. **Panel C** shows the network of highly related pairs (IBD >0.25) within and between the 6 northern Pemba *shehias* (note: Micheweni is a *shehia* in Micheweni district). Samples (nodes) are colored by *shehia* and IBD estimates (edges) are represented on a continuous scale with increasing width and yellow-shading indicating higher IBD.

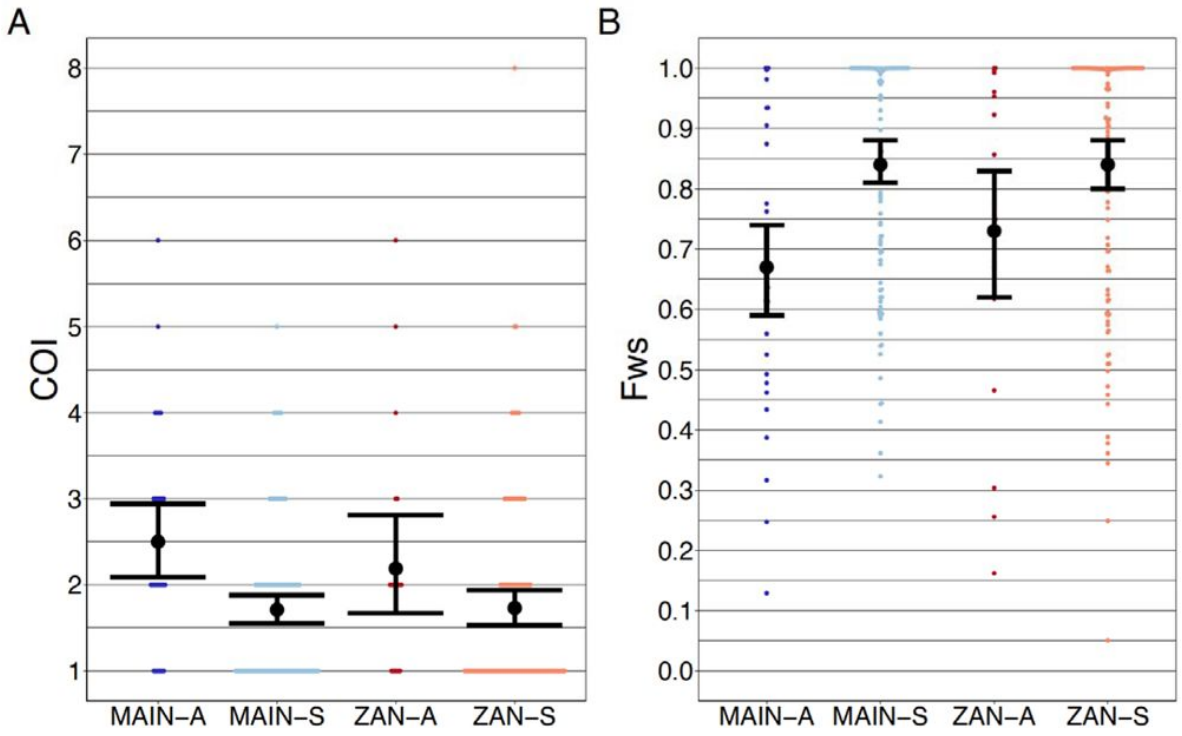

Figure 5.

Complexity of infection (COI) and Fws metric shows a higher COI and lower Fws in asymptomatic than symptomatic infections in both mainland Tanzania and Zanzibar isolates.

COI **(A)** was estimated by the REAL McCOIL's categorical method (H.-H. Chang et al., 2017). Mean COI for asymptomatic was greater than symptomatic infections for all regions (MAIN-A: 2.5 (2.1-2.9), MAIN-S: 1.7 (1.6-1.9), $p < 0.05$, Wilcoxon-Mann-Whitney test and ZAN-A: 2.2 (1.7-2.8), ZAN-S: 1.7 (1.5-1.9), $p = 0.05$, Wilcoxon-Mann-Whitney test). Fws (**B**) was estimated utilizing the formula, $(1 - H_w)/H_p$, where H_w is the within-sample heterozygosity and H_p is the heterozygosity across the population. Mean Fws was less in asymptomatic than symptomatic samples (MAIN-A: 0.67 (0.6-0.7), MAIN-S: 0.85 (0.8-0.9), $p < 0.05$, Wilcoxon-Mann-Whitney test and ZAN-A: 0.73 (0.6-0.8), ZAN-S: 0.84 (0.8-0.9), $p = 0.05$, Wilcoxon-Mann-Whitney test). A nonparametric bootstrap was applied to calculate the mean and 95% confidence interval (CI) from the COI and Fws values.

Drug resistance mutations did not vary between populations

The prevalence of the drug resistance genotypes were quite similar in Zanzibar and coastal Tanzania (**Table 2**). The frequencies of five mutations associated with sulfadoxine/pyrimethamine resistance (Pfdhfr: N51I, C59R, S108N, Pfdhps: A437G, K540E) were quite high with prevalences at or above 0.90. Pfcrt mutations associated with chloroquine and amodiaquine resistance (M74I, N75E, K76T) were all present at approximately 0.05 prevalence (Djimdé et al., 2001; Holmgren et al., 2006). For Pfmdr1, wild type N86 and D1246 were dominant at 0.99 prevalence, which are associated with reduced susceptibility to lumefantrine (Sisowath et al., 2005). No World Health Organization validated or candidate polymorphism in Pfk13 associated with artemisinin resistance were found.

Discussion

In this study, we leverage high-throughput targeted sequencing using molecular inversion probes (MIPs) to characterize the populations and the relationships of *P. falciparum* isolates in Zanzibar and coastal mainland Tanzania. The parasite populations appear to be highly related to each other (**Figure 1A**) when evaluated using SNPs in the core genome. Interestingly, within Zanzibar, structure could be observed, with parasites closer to the main ferry terminal in Zanzibar town clustered more closely with coastal mainland parasites (**Figure 1B**, **Supplemental Figure 1**) compared to parasites that were more geographically distant. This, in combination with the evidence of rapid decline of genetic relatedness with distance on the archipelago (**Figure 3**), is consistent with population microstructure within the island chain. This microstructure within the archipelago is supported by K-means clustering where Zanzibari isolates show higher within cluster than between cluster IBD (**Figure 2**). It is also consistent with isolates with higher IBD (**Figure 4A and B**) in Unguja and Pemba compared to a maximum IBD of 0.20 between Zanzibar and coastal mainland Tanzania (**Supplemental Figure 8**) or between Unguja and Pemba (**Supplemental Figure 9**). Parasite populations within the very low transmission region of Zanzibar may be more isolated than expected, allowing them to differentiate from each other. This may be indicative of very effective local malaria control, yet with continued micro-transmission remaining. Thus, directly targeting local malaria transmission, including the asymptomatic reservoir which contributes to sustained transmission (Barry et al., 2021; Sumner et al., 2021), may be an important focus for ultimately achieving malaria control in the archipelago (Björkman & Morris, 2020). Currently, a reactive case detection program within index case households is being implemented, but local transmission continues and further investigation into how best to control this is warranted (Mkali et al., 2023).

Despite the overall genetic similarity between archipelago populations, we did not find parasite pairs with high levels of IBD between the coastal mainland and Zanzibar, with the highest being 0.20. While this level still represents a significant amount of genetic sharing, similar to a cousin, the lack of higher levels does not allow us to identify specific importation events. This is largely due to the study design, which is based on convenience sampling, the relatively low numbers of samples and lack of sampling from all mainland travel hubs (Bisanzio et al., 2023). Sampling was also denser in Unguja compared to Pemba. On the other hand, we see clear transmission of highly related parasites within each population (IBD > 0.99). In Zanzibar, highly related parasites mainly occur at the range of 20-30km. These results are similar to our previous work using whole genome sequencing of isolates from Zanzibar and mainland Tanzania, showing increased within population IBD compared to between population IBD (Morgan et al., 2020). The network of highly related *P. falciparum* parasites from 6 *shehias* in North Pemba provides an excellent example of likely recent near clonal transmission, consistent with an outbreak (**Figure 4C**). A recent study investigating population structure in Zanzibar also found local population

Mutation	Zanzibar			Mainland		
	Mutant Allele Prevalence‡	CI‡	# Genotyped Samples§	Mutant Allele Prevalence‡	CI‡	# Genotyped Samples§
Pfcrt-M74I	0.054	0.026-0.098	184	0.000	0-0.034	106
Pfcrt-N75E	0.054	0.026-0.098	184	0.000	0-0.034	106
Pfcrt-K76T	0.054	0.026-0.098	184	0.000	0-0.034	106
Pfdhfr-A16V	0.000	0-0.021	173	0.000	0-0.032	112
Pfdhfr-N51I	0.977	0.943-0.994	177	0.964	0.911-0.99	112
Pfdhfr-C59R	0.971	0.934-0.991	174	0.945	0.884-0.98	109
Pfdhfr-S108N	1.000	0.98-1	179	1.000	0.965-1	104
Pfdhfr-S108T	0.000	0-0.02	179	0.000	0-0.035	104
Pfdhfr-I164L	0.000	0-0.02	184	0.000	0-0.037	98
Pfdhps-A437G	1.000	0.98-1	182	1.000	0.968-1	115
Pfdhps-K540E	0.955	0.913-0.98	178	0.964	0.91-0.99	111
Pfdhps-A581G	0.044	0.019-0.085	181	0.107	0.058-0.175	122
Pfk13-K189N	0.023	0.006-0.058	174	0.000	0-0.04	90
Pfk13-K189T	0.078	0.042-0.13	166	0.095	0.042-0.179	84
Pfmdr1-N86Y	0.011	0.001-0.04	180	0.008	0-0.044	124
Pfmdr1-Y184F	0.644	0.57-0.714	180	0.530	0.435-0.624	115
Pfmdr1-D1246Y	0.011	0.001-0.039	184	0.019	0.002-0.067	105
Pfmdr2-I492V	0.430	0.357-0.506	179	0.407	0.302-0.518	86

‡Prevalence was calculated as described in the methods. 95% CI of these polymorphisms were calculated through the Pearson-Klopper method.

§The number of genotyped samples per loci is also shown for each polymorphism.

Table 2.

Drug resistance polymorphism prevalence in Zanzibar and coastal mainland Tanzania.

microstructure in Pemba (Holzschuh et al., 2023). Further, both studies found near-clonal parasites within the same district, Micheweni, and found population microstructure over Zanzibar. Overall, given the findings of microstructure with significant local sharing of highly related strains, these small clusters still potentially drive much of the malaria transmission that occurs within the archipelago through routine human movement or mosquito travel between locales (Huestis et al., 2019). Less frequent longer distance transmission events also occur, likely due to longer range human migration within the islands.

Asymptomatic parasitemia has been shown to be common in falciparum malaria around the globe and has been shown to have increasing importance in Zanzibar (Lindblade et al., 2013; Morris et al., 2015). What underlies the biology and prevalence of asymptomatic parasitemia in very low transmission settings where anti-parasite immunity is not expected to be prevalent remains unclear (Björkman & Morris, 2020). Similar to a few previous studies, we found that asymptomatic infections had a higher COI than symptomatic infections across both the coastal mainland and Zanzibar parasite populations (Collins et al., 2022; Kimenyi et al., 2022; Sarah-Matio et al., 2022). Other studies have found lower COI in severe vs. mild malaria cases (Robert et al., 1996) or no significant difference between COI based on clinical status (Conway et al., 1991; Earland et al., 2019; Kun et al., 1998; Lagnika et al., 2022; Tanabe et al., 2015). In Zambia, one study suggested that infections that cause asymptomatic infection may be genetically different than those that cause symptomatic infection (Searle et al., 2017). However, this study included samples collected over different time periods and relied on a low-density genotyping assay which only investigated the diversity of 24 single nucleotide polymorphisms across the genome. Here, based on SNPs throughout the core genome, we did not see differential clustering of asymptomatic or symptomatic infections in Zanzibar or the mainland (**Figure 1A**), suggesting that these parasite populations remain similar when comparing clinical status. However, this genotyping approach does not address potential variation in the many hypervariable gene families that encode genes known to be associated with pathogenesis (e.g. *var*, *rifin* and *stevor* genes) and does not address differences in expression of genes associated with pathogenesis that may reflect differences in the populations. Investigation with other methods, such as long-read genome sequencing and transcriptional profiling, would be needed to address these differences.

While mutations for partial artemisinin resistance were not observed in K13, other antimalarial resistant mutations of concern were observed. Validated drug resistance mutations linked to sulfadoxine/pyrimethamine resistance (Pfdhfr-N51I, Pfdhfr-C59R, Pfdhfr-S108N, Pfdhps-A437G, Pfdhps-K540E) were found at high prevalence (**Table 2**). Prevalence of polymorphisms associated with amodiaquine resistance (Pfcrt-K76T, Pfmdr1-N86Y, Pfmdr1-Y184F, Pfmdr1-D1246Y) were seen at similar proportions as previous reports (Msellem et al., 2020). The wild-type Pfmdr1-N86 was dominant in both mainland and archipelago populations, concerning for reduced lumefantrine susceptibility. Although polymorphisms associated with artemisinin resistance did not appear in this population, continued surveillance is warranted given emergence of these mutations in East Africa and reports of rare resistance mutations on the coast consistent with spread of emerging Pfk13 mutations (Moser et al., 2021).

Overall, parasites between Zanzibar and coastal mainland Tanzania remain highly related, but population microstructure on the island reflects ongoing low level transmission in Zanzibar, partially driven by asymptomatic infections that potentially constitute a long-term reservoir. This is likely the result of the continued pressure on the population through the implementation of effective control measures. In this study, parasite genomics allows us to parse differences in parasite populations and reveals substructure in an area of low transmission intensity. A recent study identified “hotspot” *shehias*, defined as areas with comparatively higher malaria transmission than other *shehias*, near the port of Zanzibar town and in northern Pemba (Bisanzio et al., 2023). These regions overlapped with *shehias* in this study with high levels of IBD, especially in northern Pemba (**Figure 4**). These areas of substructure represent parasites that differentiated in relative isolation and are thus important locales to target intervention to

interrupt local transmission (Bousema et al., 2012). While a field cluster-randomized control trial in Kenya targeting these hotspots did not confer much reduction of malaria outside of the hotspot (Bousema et al., 2016), if areas are isolated pockets, which genetic differentiation can help determine, targeted interventions in these areas are likely needed, potentially through both mass drug administration and vector control (Morris et al., 2018; Okell et al., 2011). Such strategies and measures preventing imported malaria could accelerate progress towards zero malaria in Zanzibar.

Ethical approvals and consent to participate

This analysis was approved by the IRBs at the University of North Carolina at Chapel Hill (15-1989, 17-0166, 18-1090), Muhimbili University of Health and Allied Sciences (MUHAS), Zanzibar Medical Research Ethical Committee and the Regional Ethics Review Board, Stockholm, Sweden.

Competing interests

The authors have no competing interests to declare.

Data Availability

Parasite sequence data is available through SRA (BioProject PRJNA926345). Code used for analysis is available at: github.com/sconnelly007/TAN_MIP.

Acknowledgements

This research was funded by the National Institutes of Health, grants R01AI121558, R01AI137395, R01AI155730, F30AI143172 and K24AI134990. Funding was also contributed from the Swedish Research Council, Erling-Persson Family Foundation and the Yang Fund. RV acknowledges funding from the MRC Centre for Global Infectious Disease Analysis (reference MR/R015600/1), jointly funded by the UK Medical Research Council (MRC) and the UK Foreign, Commonwealth & Development Office (FCDO), under the MRC/FCDO Concordat agreement and is also part of the EDCTP2 programme supported by the European Union. RV also acknowledges funding by Community Jameel. We would like to thank the communities and participants who took part in these studies. We would like to thank Abebe Fola for his assistance.

Data accessibility statement

Parasite sequence data is available through SRA (BioProject PRJNA926345). Code used for analysis is available at: github.com/sconnelly007/TAN_MIP.

Author contributions

SVC, NB, VG and ZPH conducted analysis and wrote the manuscript. OA, DG, KN, CH and ZP assisted with analysis and participated in manuscript preparation. BN, LEM, SA, SS, MM, UM, AM, and JMO ran the studies in Tanzania and Kenya from which data is derived and participated in manuscript preparation. AB, AM, RV, JTL, JAB and JJJ helped conceive the study, contributed to the experimental design, and wrote the manuscript.

References

Alegana V. A. *et al.* (2021) **Plasmodium falciparum parasite prevalence in East Africa: Updating data for malaria stratification** *PLOS Global Public Health* **1**

Aydemir O. *et al.* (2018) **Drug-Resistance and Population Structure of Plasmodium falciparum Across the Democratic Republic of Congo Using High-Throughput Molecular Inversion Probes** *The Journal of Infectious Diseases* **218**:946–955

Barry A. *et al.* (2021) **Higher gametocyte production and mosquito infectivity in chronic compared to incident Plasmodium falciparum infections** *Nature Communications* **12**

Bisanzio D. *et al.* (2023) **Spatiotemporal dynamics of malaria in Zanzibar, 2015–2020** *BMJ Global Health, 8(1), e009566*

Björkman A., Morris U (2020) **Why Asymptomatic Plasmodium falciparum Infections Are Common in Low-Transmission Settings** *Trends in Parasitology* **36**:898–905

Björkman A. *et al.* (2019) **From high to low malaria transmission in Zanzibar-challenges and opportunities to achieve elimination** *BMC Medicine* **17**

Bousema T., Griffin J. T., Sauerwein R. W., Smith D. L., Churcher T. S., Takken W., Ghani A., Drakeley C., Gosling R (2012) **Hitting hotspots: spatial targeting of malaria for control and elimination** *PLoS Medicine* **9**

Bousema T. *et al.* (2016) **The impact of hotspot-targeted interventions on malaria transmission in Rachuonyo South District in the western Kenyan highlands: A cluster-randomized controlled trial** *PLoS Medicine* **13**

Chang H. H. *et al.* (2019) **Mapping imported malaria in Bangladesh using parasite genetic and human mobility data** *ELife* **8** https://doi.org/10.7554/eLife.43481

Chang H.-H. *et al.* (2017) **THE REAL McCOIL: A method for the concurrent estimation of the complexity of infection and SNP allele frequency for malaria parasites** *PLoS Computational Biology* **13**

Collins K. A. *et al.* (2022) **A Cohort Study on the Duration of Plasmodium falciparum Infections During the Dry Season in The Gambia** *In The Journal of Infectious Diseases* **226**:128–137 https://doi.org/10.1093/infdis/jiac116

Conway D. J., Greenwood B. M., McBride J. S (1991) **The epidemiology of multiple-clone Plasmodium falciparum infections in Gambian patients** *Parasitology, 103 Pt 1* :1–6

Djimdé A. *et al.* (2001) **A molecular marker for chloroquine-resistant falciparum malaria** *The New England Journal of Medicine* **344**:257–263

Earland D., Buchwald A. G., Sixpence A., Chimenya M., Damson M., Seydel K. B., Mathanga D. P., Taylor T. E., Laufer M. K (2019) **Impact of Multiplicity of Plasmodium falciparum Infection on Clinical Disease in Malawi** *The American Journal of Tropical Medicine and Hygiene* **101**:412–415

Fola A. A. *et al.* (2023) **Temporal and spatial analysis of Plasmodium falciparum genomics reveals patterns of parasite connectivity in a low-transmission district in Southern Province, Zambia** *Malaria Journal*

Holmgren G., Gil J. P., Ferreira P. M., Veiga M. I., Obonyo C. O., Björkman A (2006) **Amodiaquine resistant Plasmodium falciparum malaria in vivo is associated with selection of pfcrt 76T and pfmdr1 86Y** *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases* **6**:309–314

Holzschuh A. *et al.* (2023) **Multiplexed ddPCR-amplicon sequencing reveals isolated Plasmodium falciparum populations amenable to local elimination in Zanzibar, Tanzania** *Nature Communications* **14**

Huestis D. L. *et al.* (2019) **Windborne long-distance migration of malaria mosquitoes in the Sahel** *Nature* **574**:404–408

Jombart T (2008) **adegenet: a R package for the multivariate analysis of genetic markers** *Bioinformatics* **24**:1403–1405

Kimenyi K. M., Wamae K., Ngoi J. M., de Laurent Z. R., Ndwiga L., Osoti V., Obiero G., Abdi A. I., Bejon P., Ochola-Oyier L. I. (2022) **Maintenance of high temporal Plasmodium falciparum genetic diversity and complexity of infection in asymptomatic and symptomatic infections in Kilifi Kenya from 2007 to 2018** *Malaria Journal*

Kun J. F., Schmidt-Ott R. J., Lehman L. G., Lell B., Luckner D., Greve B., Matousek P., Kremsner P. G (1998) **Merozoite surface antigen 1 and 2 genotypes and rosetting of Plasmodium falciparum in severe and mild malaria in Lambaréné, Gabon** *Transactions of the Royal Society of Tropical Medicine and Hygiene* **92**:110–114

Lagnika H. O., Moussiliou A., Agonhossou R., Sovegnon P., Djihinto O. Y., Medjigbodo A. A., Djossou L., Amoah L. E., Ogouyemi-Hounto A., Djogbenou L. S (2022) **Plasmodium falciparum msp1 and msp2 genetic diversity in parasites isolated from symptomatic and asymptomatic malaria subjects in the South of Benin** *Parasitology Research* **121**:167–175

Le Menach A., Tatem A. J., Cohen J. M., Hay S. I., Randell H., Patil A. P., Smith D. L. (2011) **Travel risk, malaria importation and malaria transmission in Zanzibar** *Scientific Reports* **1**

Lindblade K. A., Steinhardt L., Samuels A., Kachur S. P., Slutsker L (2013) **The silent threat: asymptomatic parasitemia and malaria transmission** *Expert Review of Anti-Infective Therapy* **11**:623–639

Lipner E., Law M., Barnett E., Keystone J., von Sonnenburg F., Loutan L., Prevots D., Klion A., Nutman T. (2011) **Travel Patterns and Imported Plasmodium falciparum Rates among Zanzibar Residents**

Mkali H. R. *et al.* (2023) **How Real-Time Case-Based Malaria Surveillance Helps Zanzibar Get a Step Closer to Malaria Elimination: Description of Operational Platform and Resources** *Global Health, Science and Practice* **11** https://doi.org/10.9745/GHSP-D-22-00522

Monroe A. *et al.* (2019) **Human behaviour and residual malaria transmission in Zanzibar: findings from in-depth interviews and direct observation of community events** *Malaria Journal* **18**

Morgan A. P. *et al.* (2020) **Falciparum malaria from coastal Tanzania and Zanzibar remains highly connected despite effective control efforts on the archipelago** *Malaria Journal* **19**

Morris U. *et al.* (2018) **A cluster randomised controlled trial of two rounds of mass drug administration in Zanzibar, a malaria pre-elimination setting-high coverage and safety, but no significant impact on transmission** *BMC Medicine* **16**

Morris U. *et al.* (2015) **Characterising temporal trends in asymptomatic Plasmodium infections and transporter polymorphisms during transition from high to low transmission in Zanzibar, 2005-2013** *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases, 33* :110–117

Moser K. A. *et al.* (2021) **Describing the current status of Plasmodium falciparum population structure and drug resistance within mainland Tanzania using molecular inversion probes** *Molecular Ecology* **30**:100–113

Msellem M., Morris U., Soe A., Abbas F. B., Ali A.-W., Barnes R., Frumento P., Ali A. S., Mårtensson A., Björkman A (2020) **Increased sensitivity of Plasmodium falciparum to artesunate/amodiaquine despite 14 years as first-line malaria treatment, Zanzibar** *Emerging Infectious Diseases* **26**:1767–1777

Neafsey D. E., Taylor A. R., MacInnis B. L (2021) **Advances and opportunities in malaria population genomics** *Nature Reviews. Genetics* **22**:502–517

Okell L. C., Griffin J. T., Kleinschmidt I., Hollingsworth T. D., Churcher T. S., White M. J., Bousema T., Drakeley C. J., Ghani A. C (2011) **The potential contribution of mass treatment to the control of Plasmodium falciparum malaria** *PloS One* **6**

Patel J. C. *et al.* (2014) **Genetic Evidence of Importation of Drug-Resistant Plasmodium falciparum to Guatemala from the Democratic Republic of the Congo** *Emerging Infectious Diseases* **20**:932–940

Robert F., Ntoumi F., Angel G., Candito D., Rogier C., Fandeur T., Sarthou J. L., Mercereau-Puijalon O (1996) **Extensive genetic diversity of Plasmodium falciparum isolates collected from patients with severe malaria in Dakar, Senegal** *Transactions of the Royal Society of Tropical Medicine and Hygiene* **90**:704–711

Roh M. E., Tessema S. K., Murphy M., Nhlabathi N., Mkhonta N., Vilakati S., Ntshalintshali N., Saini M., Maphalala G., Chen A. (2019) **High genetic diversity of Plasmodium falciparum in the low-transmission setting of the Kingdom of Eswatini** *The Journal of Infectious Diseases* **220**:1346–1354

Sane R. *et al.* (2019) **Low genetic diversity and complexity of submicroscopic Plasmodium falciparum infections among febrile patients in low transmission areas in Senegal** *PloS One* **14**

Sarah-Matio E. M. *et al.* (2022) **Genetic Diversity of Plasmodium falciparum and Distribution of Antimalarial Drug Resistance Mutations in Symptomatic and Asymptomatic Infections** *Antimicrobial Agents and Chemotherapy* **66**

Searle K. M., Katowa B., Kobayashi T., Siame M. N. S., Mharakurwa S., Carpi G., Norris D. E., Stevenson J. C., Thuma P. E., Moss W. J. (2017) **Distinct parasite populations infect individuals identified through passive and active case detection in a region of declining malaria transmission in southern Zambia** *Malaria Journal*

Sisowath C., Strömberg J., Mårtensson A., Msellem M., Obondo C., Björkman A., Gil J. P (2005) **In vivo selection of Plasmodium falciparum pfmdr1 86N coding alleles by artemether-lumefantrine (Coartem)** *The Journal of Infectious Diseases* **191**:1014–1017

Sumner K. M., Freedman E., Abel L., Obala A., Pence B. W., Wesolowski A., Meshnick S. R., Prudhomme-O'Meara W., Taylor S. M (2021) **Genotyping cognate Plasmodium falciparum in humans and mosquitoes to estimate onward transmission of asymptomatic infections** *Nature Communications* **12**

Tanabe K., Zollner G., Vaughan J. A., Sattabongkot J., Khuntirat B., Honma H., Mita T., Tsuboi T., Coleman R (2015) **Plasmodium falciparum: genetic diversity and complexity of infections in an isolated village in western Thailand** *Parasitology International* **64**:260–266

Tatem A. J., Qiu Y., Smith D. L., Sabot O., Ali A. S., Moonen B (2009) **The use of mobile phone data for the estimation of the travel patterns and imported Plasmodium falciparum rates among Zanzibar residents** *Malaria Journal* **8**

Topazian H. M. *et al.* (2022) **Low complexity of infection is associated with molecular persistence of Plasmodium falciparum in Kenya and Tanzania** *Frontiers in Epidemiology* **2** https://doi.org/10.3389/fepid.2022.852237

Verity R. *et al.* (2020) **The impact of antimalarial resistance on the genetic structure of Plasmodium falciparum in the DRC** *Nature Communications* **11**

World Health Organization (2020) **Report on antimalarial drug efficacy, resistance and response: 10 years of surveillance (2010-2019)** *World Health Organization*

World Health Organization (2022) **World malaria report 2022** *World Health Organization*

Article and author information

Sean V. Connelly
MD-PhD Program, University of North Carolina, Chapel Hill, NC 27599
ORCID iD: 0000-0002-7330-7340

Nicholas F. Brazeau
MD-PhD Program, University of North Carolina, Chapel Hill, NC 27599
ORCID iD: 0000-0003-3976-7965

Mwinyi Msellem
Research Division, Ministry of Health, Zanzibar, Tanzania

Billy E. Ngasala
Department of Parasitology and Medical Entomology, Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania, Global Health and Migration Unit, Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden
ORCID iD: 0000-0001-9300-7719

Özkan Aydemir
Department of Medicine, University of Massachusetts Chan Medical School, Worcester, MA
ORCID iD: 0000-0002-3458-6527

Varun Goel
Carolina Population Center, University of North Carolina, Chapel Hill, NC 27599
ORCID iD: 0000-0002-2933-427X

Karamoko Niaré
Department of Pathology and Laboratory Medicine, Brown University, Providence, RI, 02912 USA
ORCID iD: 0000-0002-7406-9508

David J. Giesbrecht
Department of Pathology and Laboratory Medicine, Brown University, Providence, RI, 02912 USA
ORCID iD: 0000-0003-2498-4705

Zachary R. Popkin-Hall
Institute for Global Health and Infectious Diseases, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599 USA
ORCID iD: 0000-0002-8308-5294

Christopher M. Hennelly
Institute for Global Health and Infectious Diseases, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599 USA
ORCID iD: 0000-0001-5325-6753

Zackary Park
Division of Infectious Diseases, Department of Medicine, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599 USA
ORCID iD: 0000-0002-4743-6624

Ann M. Moormann
Department of Medicine, University of Massachusetts Chan Medical School, Worcester, MA
ORCID iD: 0000-0003-1113-2829

John Michael Ong'echa
Center for Global Health Research, Kenyan Medical Research Institute, Kisumu, Kenya
ORCID iD: 0000-0003-3928-6774

Robert Verity
MRC Centre for Global Infectious Disease Analysis, Imperial College, London
ORCID iD: 0000-0002-3902-8567

Safia Mohammed
Zanzibar Malaria Elimination Program (ZAMEP), Zanzibar, Tanzania

Shija J. Shija
Zanzibar Malaria Elimination Program (ZAMEP), Zanzibar, Tanzania

Lwidiko E. Mhamilawa
Department of Parasitology and Medical Entomology, Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania, Global Health and Migration Unit, Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden

Ulrika Morris
Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, 17177 Stockholm, Sweden
ORCID iD: 0000-0003-4734-5754

Andreas Mårtensson
Global Health and Migration Unit, Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden

Jessica T. Lin
Division of Infectious Diseases, Department of Medicine, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599 USA
ORCID iD: 0000-0002-4516-723X

Anders Björkman
Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, 17177 Stockholm, Sweden, Department of Global Public Health, Karolinska Institutet, Stockholm, Sweden
ORCID iD: 0000-0002-8882-130X

Jonathan J. Juliano
Division of Infectious Diseases, Department of Medicine, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599 USA, Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, 27599 USA, Curriculum in Genetics and Molecular Biology, University of North Carolina, Chapel Hill, NC 27599 USA
ORCID iD: 0000-0002-0591-0850

Jeffrey A. Bailey
Department of Pathology and Laboratory Medicine, Brown University, Providence, RI, 02912 USA
For correspondence: jeffrey_bailey@brown.edu
ORCID iD: 0000-0002-6899-8204

Copyright

Editors

Reviewing Editor
Marcelo Ferreira
University of São Paulo, São Paulo, Brazil

Senior Editor
Dominique Soldati-Favre
University of Geneva, Geneva, Switzerland

Reviewer #1 (Public Review):

Summary:

Zanzibar archipelago is close to achieve malaria elimination, but despite the implementation of effective control measures there is still low level seasonal malaria transmission. This could be due to the frequent importation of malaria from the mainland Tanzania and Kenya, reservoir of asymptomatic infections and competent vectors. To investigate population structure and gene flow of P. falciparum in Zanzibar and mainland Tanzania, they used 178 samples from mainland Tanzania and 213 from Zanzibar that were previously sequenced using molecular inversion probes (MIPs) panels targeting single nucleotide polymorphisms (SNPs). They performed Principal Component Analysis (PCA) and identity by descent (IBD) analysis to assess genetic reladness between isolates. Parasites from coastal mainland Tanzania contribute for the genetic diversity in parasite population in Zanzibar. Despite this, there is a pattern of isolation by distance and microstructure within the achipelago, and evidence of local sharing of highly related strains sustaining malaria transmission in Zanzibar that are important targets for interventions such as mass drug administration and vector control, in addition to measures against imported malaria.

Strengths:

This study presents important samples to understand population structure and gene flow between mainland Tanzania and Zanzibar, especially from rural Bagamoyo District, where malaria transmission persists and there is a major port of entry to Zanzibar. In addition, this study includes a larger set of SNPs, providing more robustness for analyzes such as PCA and IBD. Therefore, the conclusions of this paper are well supported by data.

Comments on revised version:

The authors answered all my questions.

https://doi.org/10.7554/eLife.90173.2.sa1

Reviewer #2 (Public Review):

Summary:

This manuscript describes P. falciparum population structure in Zanzibar and mainland Tanzania. 282 samples were typed using molecular inversion probes. The manuscript is overall well written and shows clear population structure. It follows a similar manuscript published earlier this year, which typed a similar number of samples collected mostly in the same sites around the same time. The current manuscript extends this work by including a large number of samples from coastal Tanzania, and by including clinical samples, allowing for a comparison with asymptomatic samples.

The two studies made overall very similar findings, including strong small-scale population structure, related infections on Zanzibar and the mainland, near-clonal expansion on Pemba, and frequency of markers of drug resistance.

Strengths:

The overall results show a clear pattern of population structure. The finding of highly related infections detected in close proximity shows local transmission and can possibly be leveraged for targeted control.

Comments on revised version:

The authors have addressed my comments.

https://doi.org/10.7554/eLife.90173.2.sa0

Author response:

The following is the authors' response to the original reviews.

> ***eLife assessment***
>
> *Connelly and colleagues provide convincing genetic evidence that importation from mainland Tanzania is a major source of Plasmodium falciparum lineages currently circulating in Zanzibar. This study also reveals ongoing local malaria transmission and occasional near-clonal outbreaks in Zanzibar. Overall, this research highlights the role of human movements in maintaining residual malaria transmission in an area targeted for intensive control interventions over the past decades and provides valuable information for epidemiologists and public health professionals.*
>
> ***Reviewer #1 (Public Review):***
>
> *Zanzibar archipelago is close to achieving malaria elimination, but despite the implementation of effective control measures, there is still a low-level seasonal malaria transmission. This could be due to the frequent importation of malaria from mainland Tanzania and Kenya, reservoirs of asymptomatic infections, and competent vectors. To investigate population structure and gene flow of P. falciparum in Zanzibar and mainland Tanzania, they used 178 samples from mainland Tanzania and 213 from Zanzibar that were previously sequenced using molecular inversion probes (MIPs) panels targeting single nucleotide polymorphisms (SNPs). They performed Principal Component Analysis (PCA) and identity by descent (IBD) analysis to assess genetic relatedness between isolates. Parasites from coastal mainland Tanzania contribute to the genetic diversity in the parasite population in Zanzibar. Despite this, there is a pattern of isolation by distance and microstructure within the archipelago, and evidence of local sharing of highly related strains sustaining malaria transmission in Zanzibar that are important targets for interventions such as mass drug administration and vector control, in addition to measures against imported malaria.*
>
> ***Strengths:***
>
> *This study presents important samples to understand population structure and gene flow between mainland Tanzania and Zanzibar, especially from the rural Bagamoyo District, where malaria transmission persists and there is a major port of entry to Zanzibar. In addition, this study includes a larger set of SNPs, providing more robustness for analyses such as PCA and IBD. Therefore, the conclusions of this paper are well supported by data.*

> ***Weaknesses:***
>
> *Some points need to be clarified:*
>
> *(1) SNPs in linkage disequilibrium (LD) can introduce bias in PCA and IBD analysis. Were SNPs in LD filtered out prior to these analyses?*

Thank you for this point. We did not filter SNPs in LD prior to this analysis. In the PCA analysis in Figure 1, we did restrict to a single isolate among those that were clonal (high IBD values) to prevent bias in the PCA. In general, disequilibrium is minimal only over small distances <5-10kb without selective forces at play. This is much less than the average spacing of the markers in the panel. If there is minimal LD, the conclusions drawn on relative levels and connections at high IBD are unlikely to be confounded by any effects of disequilibrium.

> *(2) Many IBD algorithms do not handle polyclonal infections well, despite an increasing number of algorithms that are able to handle polyclonal infections and multiallelic SNPs. How polyclonal samples were handled for IBD analysis?*

Thank you for this point. We added lines 157-161 to clarify. This section now reads:

“To investigate genetic relatedness of parasites across regions, identity by descent (IBD) estimates were assessed using the within sample major alleles (coercing samples to monoclonal by calling the dominant allele at each locus) and estimated utilizing a maximum likelihood approach using the inbreeding_mle function from the MIPanalyzer package (Verity et al., 2020). This approach has previously been validated as a conservative estimate of IBD (Verity et al., 2020).”

Please see the supplement in (Verity et al., 2020) for an extensive simulation study that validates this approach.

> ***Reviewer #1 (Recommendations For The Authors):***
>
> *(3) I think Supplementary Figures 8 and 9 are more visually informative than Figure 2.*

Thank you for your response. We performed the analysis in Figure 2 to show how IBD varies between different regions and is higher within a region than between.

> ***Reviewer #2 (Public Review):***
>
> *This manuscript describes P. falciparum population structure in Zanzibar and mainland Tanzania. 282 samples were typed using molecular inversion probes. The manuscript is overall well-written and shows a clear population structure. It follows a similar manuscript published earlier this year, which typed a similar number of samples collected mostly in the same sites around the same time. The current manuscript extends this work by including a large number of samples from coastal Tanzania, and by including clinical samples, allowing for a comparison with asymptomatic samples.*
>
> *The two studies made overall very similar findings, including strong small-scale population structure, related infections on Zanzibar and the mainland, near-clonal expansion on Pemba, and frequency of markers of drug resistance. Despite these similarities, the previous study is mentioned a single time in the discussion (in contrast, the previous research from the authors of the current study is more thoroughly discussed). The authors missed an opportunity here to highlight the similar findings of the two studies.*

Thank you for your insights. We appreciated the level of detail of your review and it strengthened our work. We have input additional sentences on lines 292-295, which now reads:

"A recent study investigating population structure in Zanzibar also found local population microstructure in Pemba (Holzschuh et al., 2023). Further, both studies found near-clonal parasites within the same district, Micheweni, and found population microstructure over Zanzibar."

> *Strengths:*
>
> *The overall results show a clear pattern of population structure. The finding of highly related infections detected in close proximity shows local transmission and can possibly be leveraged for targeted control.*
>
> *Weaknesses:*
>
> *A number of points need clarification:*
>
> *(1) It is overall quite challenging to keep track of the number of samples analyzed. I believe the number of samples used to study population structure was 282 (line 141), thus this number should be included in the abstract rather than 391. It is unclear where the number 232 on line 205 comes from, I failed to deduct this number from supplementary table 1.*

Thank you for this point. We have included 282 instead of 391 in the abstract. We added a statement in the results at lines 203-205 to clarify this point, which now reads:

"PCA analysis of 232 coastal Tanzanian and Zanzibari isolates, after pruning 51 samples with an IBD of greater than 0.9 to one representative sample, demonstrates little population differentiation (Figure 1A)."

> *(2) Also, Table 1 and Supplementary Table 1 should be swapped. It is more important for the reader to know the number of samples included in the analysis (as given in Supplementary Table 1) than the number collected. Possibly, the two tables could be combined in a clever way.*

Thank you for this advice. Rather than switch to another table altogether, we appended two columns to the original table to better portray the information (see Table 1).

> *Methods*
>
> *(3) The authors took the somewhat unusual decision to apply K-means clustering to GPS coordinates to determine how to combine their data into a cluster. There is an obvious cluster on Pemba islands and three clusters on Unguja. Based on the map, I assume that one of these three clusters is mostly urban, while the other two are more rural. It would be helpful to have a bit more information about that in the methods. See also comments on maps in Figures 1 and 2 below.*

Cluster 3 is a mix of rural/urban while the clusters 2, 4 and 5 are mostly rural. This analysis was performed to see how IBD changes in relation to local context within different regions in Zanzibar, showing that there is higher IBD within locale than between locale.

> *(4) Following this point, in Supplemental Figure 5 I fail to see an inflection point at K=4. If there is one, it will be so weak that it is hardly informative. I think selecting 4 clusters in Zanzibar is fine, but the justification based on this figure is unclear.*

The K-means clustering experiment was used to cluster a continuous space of geographic coordinates in order to compare genetic relatedness in different regions. We selected this inflection point based on the elbow plot and based the number to obtain sufficient subsections of Zanzibar to compare genetic relatedness. This point is added to the methods at lines 174-178, which now reads:

“The K-means clustering experiment was used to cluster a continuous space of geographic coordinates in order to compare genetic relatedness in different regions. We selected K = 4 as the inflection point based on the elbow plot (Supplemental Figure 5) and based the number to obtain sufficient subsections of Zanzibar to compare genetic relatedness.”

> *(5) For the drug resistance loci, it is stated that "we further removed SNPs with less than 0.005 population frequency." Was the denominator for this analysis the entire population, or were Zanzibar and mainland samples assessed separately? If the latter, as for all markers <200 samples were typed per site, there could not be a meaningful way of applying this threshold. Given data were available for 200-300 samples for each marker, does this simply mean that each SNP needed to be present twice?*

Population frequency is calculated based on the average within sample allele frequency of each individual in the population, which is an unbiased estimator. Within sample allele frequency can range from 0 to 1. Thus, if only one sample has an allele and it is at 0.1 within sample frequency, the population allele frequency would be 0.1/100 = 0.001. This allele is removed even though this would have resulted in a prevalence of 0.01. This filtering is prior to any final summary frequency or prevalence calculations (see MIP variant Calling and Filtering section in the methods). This protects against errors occurring only at low frequency.

> *Discussion:*
>
> *(6) I was a bit surprised to read the following statement, given Zanzibar is one of the few places that has an effective reactive case detection program in place: "Thus, directly targeting local malaria transmission, including the asymptomatic reservoir which contributes to sustained transmission (Barry et al., 2021; Sumner et al., 2021), may be an important focus for ultimately achieving malaria control in the archipelago (Björkman & Morris, 2020)." I think the current RACD program should be mentioned and referenced. A number of studies have investigated this program.*

Thank you for this point. We have added additional context and clarification on lines 275-280, which now reads:

“Thus, directly targeting local malaria transmission, including the asymptomatic reservoir which contributes to sustained transmission (Barry et al., 2021; Sumner et al., 2021), may be an important focus for ultimately achieving malaria control in the archipelago (Björkman & Morris, 2020). Currently, a reactive case detection program within index case households is being implemented, but local transmission continues and further investigation into how best to control this is warranted (Mkali et al. 2023).”

> *(7) The discussion states that "In Zanzibar, we see this both within and between shehias, suggesting that parasite gene flow occurs over both short and long distances." I think the term 'long distances' should be better defined. Figure 4 shows that highly related infections rarely span beyond 20-30 km. In many epidemiological studies, this would still be considered short distances.*

Thank you for this point. We have edited the text at lines 287-288 to indicate that highly related parasites mainly occur at the range of 20-30km, which now reads:

“In Zanzibar, highly related parasites mainly occur at the range of 20-30km.”

> *(8) Lines 330-331: "Polymorphisms associated with artemisinin resistance did not appear in this population." Do you refer to background mutations here? Otherwise, the sentence seems to repeat lines 324. Please clarify.*

We are referring to the list of Pfk13 polymorphisms stated in the Methods from lines 146-148. We added clarifying text on lines 326-329:

“Although polymorphisms associated with artemisinin resistance did not appear in this population, continued surveillance is warranted given emergence of these mutations in East Africa and reports of rare resistance mutations on the coast consistent with spread of emerging Pfk13 mutations (Moser et al., 2021). “

> *(9) Line 344: The opinion paper by Bousema et al. in 2012 was followed by a field trial in Kenya (Bousema et al, 2016) that found that targeting hotspots did NOT have an impact beyond the actual hotspot. This (and other) more recent finding needs to be considered when arguing for hotspot-targeted interventions in Zanzibar.*

We added a clarification on this point on lines 335-345, which now reads:

“A recent study identified “hotspot” shehias, defined as areas with comparatively higher malaria transmission than other shehias, near the port of Zanzibar town and in northern Pemba (Bisanzio et al., 2023). These regions overlapped with shehias in this study with high levels of IBD, especially in northern Pemba (Figure 4). These areas of substructure represent parasites that differentiated in relative isolation and are thus important locales to target intervention to interrupt local transmission (Bousema et al., 2012). While a field cluster-randomized control trial in Kenya targeting these hotspots did not confer much reduction of malaria outside of the hotspot (Bousema et al. 2016), if areas are isolated pockets, which genetic differentiation can help determine, targeted interventions in these areas are likely needed, potentially through both mass drug administration and vector control (Morris et al., 2018; Okell et al., 2011). Such strategies and measures preventing imported malaria could accelerate progress towards zero malaria in Zanzibar.”

> *Figures and Tables:*
>
> *(10) Table 2: Why not enter '0' if a mutation was not detected? 'ND' is somewhat confusing, as the prevalence is indeed 0%.*

Thank you for this point. We have put zero and also given CI to provide better detail.

> *(11) Figure 1: Panel A is very hard to read. I don't think there is a meaningful way to display a 3D-panel in 2D. Two panels showing PC1 vs. PC2 and PC1 vs. PC3 would be better. I also believe the legend 'PC2' is placed in the wrong position (along the Y-axis of panel 2).*
>
> *Supplementary Figure 2B suffers from the same issue.*

Thank you for your comment. A revised Figure 1 and Supplemental Figure 2 are included, where there are separate plots for PC1 vs. PC2 and PC1 vs. PC3.

> *(12) The maps for Figures 1 and 2 don't correspond. Assuming Kati represents cluster 4 in Figure 2, the name is put in the wrong position. If the grouping of shehias is different between the Figures, please add an explanation of why this is.*

Thank you for this point. The districts with at least 5 samples present are plotted in the map in Figure 1B. In Figure 2, a totally separate analysis was performed, where all shehias were clustered into separate groups with k-means and the IBD values were compared between these clusters. These maps are not supposed to match, as they are separate analyses. Figure 1B is at the district level and Figure 2 is clustering shehias throughout Zanzibar.

The figure legend of Figure 1B on lines 410-414 now reads:

"B) A Discriminant Analysis of Principal Components (DAPC) was performed utilizing isolates with unique pseudohaplotypes, pruning highly related isolates to a single representative infection. Districts were included with at least 5 isolates remaining to have sufficient samples for the DAPC. For plotting the inset map, the district coordinates (e.g. Mainland, Kati, etc.) are calculated from the averages of the shehia centroids within each district."

The figure legend of Figure 2 on lines 417-425 now reads:

"Figure 2. Coastal Tanzania and Zanzibari parasites have more highly related pairs within their given region than between regions. K-means clustering of shehia coordinates was performed using geographic coordinates all shehias present from the sample population to generate 5 clusters (colored boxes). All shehias were included to assay pairwise IBD between differences throughout Zanzibar. Pairwise comparisons of within cluster IBD (column 1 of IBD distribution plots) and between cluster IBD (column 2-5 of IBD distribution plots) was done for all clusters. In general, within cluster IBD had more pairwise comparisons containing high IBD identity."

> *(13) Figure 2: In the main panel, please clarify what the lines indicate (median and quartiles?). It is very difficult to see anything except the outliers. I wonder whether another way of displaying these data would be clearer. Maybe a table with medians and confidence intervals would be better (or that data could be added to the plots). The current plots might be misleading as they are dominated by outliers.*

Thank you for this point and it greatly improved this figure. We changed the plotting mechanisms through using a beeswarm plot, which plots all pairwise IBD values within each comparison group.

> *(14) In the insert, the cluster number should not only be given as a color code but also added to the map. The current version will be impossible to read for people with color vision impairment, and it is confusing for any reader as the numbers don't appear to follow any logic (e.g. north to south).*

Thank you very much for these considerations. We changed the color coding to a color blind friendly palette and renamed the clusters to more informative names; Pemba, Unguja North (Unguja_N), Unguja Central (Unguja_C), Unguja South (Unguja_S) and mainland Tanzania (Mainland).

> *(15) The legend for Figure 3 is difficult to follow. I do not understand what the difference in binning was in panels A and B compared to C.*

Thank you for this point. We have edited the legend to reflect these changes. The legend for Figure 3 on lines 427-433 now reads:

"Figure 3. Isolation by distance is shown between all Zanzibari parasites (A), only Unguja parasites (B) and only Pemba parasites (C). Samples were analyzed based on geographic location, Zanzibar (N=136) (A), Unguja (N=105) (B) or Pemba (N=31) (C) and greater circle (GC) distances between pairs of parasite isolates were calculated based on shehia centroid coordinates. These distances were binned at 4km increments out to 12 km. IBD beyond 12km is shown in Supplemental Figure 8. The maximum GC distance for all of Zanzibar was 135km, 58km on Unguja and 12km on Pemba. The mean IBD and 95% CI is plotted for each bin."

> *(16) Font sizes for panel C differ, and it is not aligned with the other panels.*

Thank you for pointing this out. Figure 3 and Supplemental Figure 10 are adjusted with matching formatting for each plot.

> *(17) Why is Kusini included in Supplemental Figure 4, but not in Figure 1?*

In Supplemental Figure 4, all isolates were used in this analysis and isolates with unique pseudohaplotypes were not pruned to a single representative infection. That is why there are additional isolates in Kusini. The legend for Supplemental Figure 4 now reads:

"Supplemental Figure 4. PCA with highly related samples shows population stratification radiating from coastal Mainland to Zanzibar. PCA of 282 total samples was performed using whole sample allele frequency (A) and DAPC was performed after retaining samples with unique pseudohaplotypes in districts that had 5 or more samples present (B). As opposed to Figure 1, all isolates were used in this analysis and isolates with unique pseudohaplotypes were not pruned to a single representative infection."

> *(18) Supplemental Figures 6 and 7: What does the width of the line indicate?*

The sentence below was added to the figure legends of Supplemental Figures 6 and 7 and the legends of each network plot were increased in size:

"The width of each line represents higher magnitudes of IBD between pairs."

> *(19) What was the motivation not to put these lines on the map, as in Figure 4A? This might make it easier to interpret the data.*

Thank you for this comment. For Supplemental Figure 8 and 9, we did not put these lines that represent lower pairwise IBD to draw the reader's attention to the highly related pairs between and within shehias.

> ***Reviewer #2 (Recommendations For The Authors):***
>
> *(1) There is a rather long paragraph (lines 300-323) on COI of asymptomatic infections and their genetic structure. Given that the current study did not investigate most of the hypotheses raised there (e.g. immunity, expression of variant genes), and the overall limited number of asymptomatic samples typed, this part of the discussion feels long and often speculative.*

Thank you for your perspective. The key sections highlighted in this comment, regarding immunity and expression of variant genes, were shortened. This section on lines 300-303 now reads:

"Asymptomatic parasitemia has been shown to be common in falciparum malaria around the globe and has been shown to have increasing importance in Zanzibar (Lindblade et al., 2013; Morris et al., 2015). What underlies the biology and prevalence of asymptomatic parasitemia

in very low transmission settings where anti-parasite immunity is not expected to be prevalent remains unclear (Björkman & Morris, 2020).”

> *(2) As a detail, line 304 mentions "few previous studies" but only one is cited. Are there studies that investigated this and found opposite results?*

Thank you for this comment. We added additional studies that did not find an association between clinical disease and COI. These changes are on lines 303-308, which now reads:

“Similar to a few previous studies, we found that asymptomatic infections had a higher COI than symptomatic infections across both the coastal mainland and Zanzibar parasite populations (Collins et al., 2022; Kimenyi et al., 2022; Sarah-Matio et al., 2022). Other studies have found lower COI in severe vs. mild malaria cases (Robert et al., 1996) or no significant difference between COI based on clinical status (Earland et al. 2019; Lagnika et al. 2022; Conway et al. 1991; Kun et al. 1998; Tanabe et al. 2015)”

> *(3) Table 2: Percentages need to be checked. To take one of several examples, for Pfk13-K189N a frequency of 0.019 for the mutant allele is given among 137 samples. 2/137 equals to 0.015, and 3/137 to 0.022. 0.019 cannot be achieved. The same is true for several other markers. Possibly, it can be explained by the presence of polyclonal infections. If so, it should be clarified what the total of clones sequenced was, and whether the prevalence is calculated with the number of samples or number of clones as the denominator.*

Thank you for this point. We mistakenly reported allele frequency instead of prevalence. An updated Table 2 is now in the manuscript. The method for calculating the prevalence is now at lines 148-151:

“Prevalence was calculated separately in Zanzibar or mainland Tanzania for each polymorphism by the number of samples with alternative genotype calls for this polymorphism over the total number of samples genotyped and an exact 95% confidence interval was calculated using the Pearson-Klopper method for each prevalence.”