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

Sean V. Connelly[1], Nicholas F. Brazeau[1], Mwinyi Msellem[2], Billy E. Ngasala[3,14], Ozkan Aydemir[8], Varun Goel[5], Karamoko Niaré[4], David J. Giesbrecht[4], Zachary R. Popkin-Hall[7], Christopher M. Hennelly[7], Zackary Park[6], Ann M. Moormann[8], John Michael Ong'echa[9], Robert Verity[10], Safia Mohammed[11], Shija J. Shija[11], Lwidiko E. Mhamilawa[3], Ulrika Morris[12], Andreas Mårtensson[15], Jessica T. Lin[6], Anders Björkman[12,13], Jonathan J. Juliano[6,16,17*], Jeffrey A. Bailey[4*]

1: MD-PhD Program, University of North Carolina, Chapel Hill, NC 27599
2: Research Division, Ministry of Health, Zanzibar, Tanzania
3: Department of Parasitology and Medical Entomology, Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania
4: Department of Pathology and Laboratory Medicine, Brown University, Providence, RI, 02912 USA
5: Carolina Population Center, University of North Carolina, Chapel Hill, NC 27599
6: Division of Infectious Diseases, Department of Medicine, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599 USA
7: Institute for Global Health and Infectious Diseases, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599 USA
8: Department of Medicine, University of Massachusetts Chan Medical School, Worcester, MA
9: Center for Global Health Research, Kenyan Medical Research Institute, Kisumu, Kenya
10: MRC Centre for Global Infectious Disease Analysis, Imperial College, London
11: Zanzibar Malaria Elimination Program (ZAMEP), Zanzibar, Tanzania
12: Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, 17177 Stockholm, Sweden
13: Department of Global Public Health, Karolinska Institutet, Stockholm, Sweden.
14: Department of Women's and Children's Health, International Maternal and Child Health (IMCH), Uppsala University, Uppsala, Sweden
15: Department of Women's and Children's Health, Global Health and Migration, Uppsala University, Uppsala, Sweden
16: Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, 27599 USA
17: Curriculum in Genetics and Molecular Biology, University of North Carolina, Chapel Hill, NC 27599 USA

*: co-senior authors

Keywords: *Plasmodium falciparum,* malaria, population microstructure, Tanzania, Zanzibar, molecular inversion probes

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 391 *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, in Zanzibar the parasite population exhibits 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 that importation remains 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.

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 (A. 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 (A. 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 (A. 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 (A. Björkman et al., 2019); and (3) a reservoir of asymptomatic infections (A. Björkman et al., 2019; Anders Björkman & Morris, 2020).

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, Taylor, & MacInnis, 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; Morgan et al., 2020; Moser et al., 2021; 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 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.

Molecular inversion probes (MIPs) are a highly multiplexed genotyping assay used 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.

METHODS

Samples from coastal Tanzania (178) and Zanzibar (213) were previously sequenced through multiple studies (**Table 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* structure, 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).

Molecular Inversion Probe (MIP) Sequencing: Sequence data for these 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 150 bp paired end sequencing with dual indexing using Nextseq 500/550 Mid-output Kit v2.

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 **Supplemental Table 1**. Sequencing coverage estimates for loci are shown in **Supplemental Figure 2**.

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. After filtering loci with greater than or equal to 5 UMIs, we were left with 309 samples at 2265 SNPs. SNPs with less than 0.05% population frequency were removed. We focused analysis on the following 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).

Analysis of Population Relatedness and Structure

Identity by descent (IBD) estimates were assessed using the within sample major alleles and estimated utilizing a maximum likelihood approach using the *inbreeding_mle* function from the *MIPanalyzer* package (Verity et al., 2020). For the Principal Component Analysis (PCA) 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). For the Discriminant Analysis of Principal Components (DAPC) analysis, 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 3B**). 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 (Beugin & Michaud, 2020) 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 (**Supplemental Figure 4**). For the isolation by distance analysis across all of Zanzibar, as well as within the islands of Unguja and Pemba, the greater circle distances between each *shehia* centroid was calculated (within *shehia* distances were equal to 0) and the 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 9, Supplemental Figure 10**). 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 6, Supplemental Figure 7**). These plots were created utilizing *ggraph* in *R* with the nodes being samples and the edges being IBD estimates. Complexity of Infection (COI) 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 was calculated in *R* version 4.2.1 through the formula, $(1 - H_w)/H_p$, where H_w is the within-sample heterozygosity and H_p is the heterozygosity across the population, and 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 5**). PCA analysis of 232 coastal Tanzanian and Zanzibari isolates 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 6**). Similarly, no pairs with an IBD of 0.20 or greater were present in pairwise comparisons between Unguja and Pemba (**Supplemental Figure 7**).

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 8**). 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 9**). 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 10**.

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**).

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 0.02 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**) 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 6**) or between Unguja and Pemba (**Supplemental Figure 7**). 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 (Anders Björkman & Morris, 2020).

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, we see this both within and between *shehias*, suggesting that parasite gene flow occurs over both short and long distances. 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 microstructure in Pemba (Holzschuh et al., 2023). 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, Steinhardt, Samuels, Kachur, & Slutsker, 2013; U. 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 (Anders 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 (Sarah-Matio et al., 2022). Potential reasons for this phenomenon include low level immune response resulting in chronic infection with multiple clones over time in asymptomatic individuals (Felger et al., 2012), and dominance of more virulent clones in shorter symptomatic infections resulting in lower COI (Collins et al., 2022). 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. Finally, any collection of "asymptomatic" isolates is likely to be heterogeneous, including those that are pre-symptomatic, those who may have been previously treated, and detection of multiple strains is also reliant on the ability of the genotyping method or sampling strategy to sensitively capture minor strains (minority genomes).

Through sequencing validated drug resistance polymorphisms, 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. Polymorphisms associated with artemisinin resistance did not appear in this population. However, 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., 2020).

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” *shehia*s, 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). A potential strategy is targeting these focal regions through both mass drug administration and vector control to eliminate these sources of transmission (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.

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.

ORCID

Sean V. Connelly: 0000-0002-7330-7340

Nicholas F. Brazeau: 0000-0003-3976-7965

Billy E. Ngasala: 0000-0001-9300-7719

Ozkan Aydemir: 0000-0002-3458-6527

Varun Goel: 0000-0002-2933-427X

Karamoko Niaré: 0000-0002-7406-9508

David J. Giesbrecht: 0000-0003-2498-4705

360 Zachary R. Popkin-Hall: 0000-0002-8308-5294

361 Christopher M. Hennelly: 0000-0001-5325-6753

362 Zackary Park: 0000-0002-4743-6624

363 Ann M. Moormann: 0000-0003-1113-2829

364 John Michael Ong'echa: 0000-0003-3928-6774

365 Robert Verity: 0000-0002-3902-8567

366 Ulrika Morris: 0000-0003-4734-5754

367 Jessica T. Lin: 0000-0002-4516-723X

368 Anders Björkman: 0000-0002-8882-130X

369 Jonathan J. Juliano: 0000-0002-0591-0850

370 Jeffrey A. Bailey: 0000-0002-6899-8204

REFERENCES

Alegana, V. A., Macharia, P. M., Muchiri, S., Mumo, E., Oyugi, E., Kamau, A., … Snow, R. W. (2021). Plasmodium falciparum parasite prevalence in East Africa: Updating data for malaria stratification. *PLOS Global Public Health*, *1*(12), e0000014.

Aydemir, O., Janko, M., Hathaway, N. J., Verity, R., Mwandagalirwa, M. K., Tshefu, A. K., … Bailey, J. A. (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*(6), 946–955.

Barry, A., Bradley, J., Stone, W., Guelbeogo, M. W., Lanke, K., Ouedraogo, A., … Bousema, T. (2021). Higher gametocyte production and mosquito infectivity in chronic compared to incident Plasmodium falciparum infections. *Nature Communications*, *12*(1), 2443.

Beugin, M.-P., & Michaud, F. (2020). *adegenet: Exploratory Analysis of Genetic and Genomic Data*. Retrieved from http://datacompass.lshtm.ac.uk/id/eprint/1765/

Bisanzio, D., Lalji, S., Abbas, F. B., Ali, M. H., Hassan, W., Mkali, H. R., … Reithinger, R. (2023). Spatiotemporal dynamics of malaria in Zanzibar, 2015–2020. *BMJ Global Health*, *8*(1), e009566.

Björkman, A., Shakely, D., Ali, A. S., Morris, U., Mkali, H., Abbas, A. K., … Mårtensson, A. (2019). From high to low malaria transmission in Zanzibar—challenges and opportunities to achieve elimination. *BMC Medicine*, Vol. 17. doi: 10.1186/s12916-018-1243-z

Björkman, Anders, & Morris, U. (2020). Why Asymptomatic Plasmodium falciparum Infections Are Common in Low-Transmission Settings. *Trends in Parasitology*, *36*(11), 898–905.

Bousema, T., Griffin, J. T., Sauerwein, R. W., Smith, D. L., Churcher, T. S., Takken, W., … Gosling, R. (2012). Hitting hotspots: spatial targeting of malaria for control and elimination. *PLoS Medicine*, *9*(1), e1001165.

Chang, H. H., Wesolowski, A., Sinha, I., Jacob, C. G., Mahmud, A., Uddin, D., … Buckee, C. (2019). Mapping imported malaria in Bangladesh using parasite genetic and human mobility data. *ELife*, *8*. doi: 10.7554/eLife.43481

Chang, H.-H., Worby, C. J., Yeka, A., Nankabirwa, J., Kamya, M. R., Staedke, S. G., … Greenhouse, B. (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*(1), e1005348.

Collins, K. A., Ceesay, S., Drammeh, S., Jaiteh, F. K., Guery, M. A., Lanke, K., … Claessens, A. (2022). A Cohort Study on the Duration of *Plasmodium falciparum* Infections During the Dry Season in The Gambia. *The Journal of Infectious Diseases*, Vol. 226, pp. 128–137. doi: 10.1093/infdis/jiac116

Djimdé, A., Doumbo, O. K., Cortese, J. F., Kayentao, K., Doumbo, S., Diourté, Y., … Plowe, C. V. (2001). A molecular marker for chloroquine-resistant falciparum malaria. *The New England Journal of Medicine*, *344*(4), 257–263.

Felger, I., Maire, M., Bretscher, M. T., Falk, N., Tiaden, A., Sama, W., … Smith, T. A. (2012). The dynamics of natural Plasmodium falciparum infections. *PloS One*, *7*(9), e45542.

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*(4), 309–314.

Holzschuh, A., Lerch, A., Gerlovina, I., Fakih, B. S., Al-mafazy, A.-W. H., Reaves, E. J., … Koepfli, C. (2023). Highly multiplexed ddPCR-amplicon sequencing reveals strong Plasmodium falciparum population structure and isolated populations amenable to local elimination efforts in Zanzibar (p. 2023.01.05.522832). doi: 10.1101/2023.01.05.522832

Huestis, D. L., Dao, A., Diallo, M., Sanogo, Z. L., Samake, D., Yaro, A. S., … Lehmann, T. (2019). Windborne long-distance migration of malaria mosquitoes in the Sahel. *Nature*, *574*(7778), 404–408.

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*, 93.

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*(6), 623–639.

Lipner, E., Law, M., Barnett, E., Keystone, J., von Sonnenburg, F., Loutan, L., … Nutman, T. (2011). Travel Patterns and Imported Plasmodium falciparum Rates among Zanzibar Residents. In *Hospitality and Health* (pp. 78–92). Apple Academic Press.

Monroe, A., Mihayo, K., Okumu, F., Finda, M., Moore, S., Koenker, H., … Harvey, S. (2019). Human behaviour and residual malaria transmission in Zanzibar: findings from in-depth interviews and direct observation of community events. *Malaria Journal*, *18*(1), 220.

Morgan, A. P., Brazeau, N. F., Ngasala, B., Mhamilawa, L. E., Denton, M., Msellem, M., … Juliano, J. J. (2020). Falciparum malaria from coastal Tanzania and Zanzibar remains highly connected despite effective control efforts on the archipelago. *Malaria Journal*, Vol. 19. doi: 10.1186/s12936-020-3137-8

Morris, U., Msellem, M. I., Mkali, H., Islam, A., Aydin-Schmidt, B., Jovel, I., … Björkman, A. (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*(1), 215.

Morris, U., Xu, W., Msellem, M. I., Schwartz, A., Abass, A., Shakely, D., … Martensson, A. (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., Aydemir, O., Hennelly, C., Kobayashi, T., Shields, T., Hamapumbu, H., … on behalf of the Southern and Central African International Center for Excellence. (2021). Temporal and spatial analysis of Plasmodium falciparum genomics reveals patterns of connectivity in a low-transmission district in Southern Province, Zambia (p. 2021.10.14.21264576). doi: 10.1101/2021.10.14.21264576

Moser, K. A., Madebe, R. A., Aydemir, O., Chiduo, M. G., Mandara, C. I., Rumisha, S. F., … Bailey, J. A. (2020). Describing the current status of Plasmodium falciparum population structure and drug resistance within mainland Tanzania using molecular inversion probes. *Molecular Ecology*. doi: 10.1111/mec.15706

Msellem, M., Morris, U., Soe, A., Abbas, F. B., Ali, A.-W., Barnes, R., … Björkman, A. (2020). Increased sensitivity of Plasmodium falciparum to artesunate/amodiaquine despite 14 years as first-line malaria treatment, Zanzibar. *Emerging Infectious Diseases*, *26*(8), 1767–1777.

Neafsey, D. E., Taylor, A. R., & MacInnis, B. L. (2021). Advances and opportunities in malaria population genomics. *Nature Reviews. Genetics*, *22*(8), 502–517.

Okell, L. C., Griffin, J. T., Kleinschmidt, I., Hollingsworth, T. D., Churcher, T. S., White, M. J., … Ghani, A. C. (2011). The potential contribution of mass treatment to the control of Plasmodium falciparum malaria. *PloS One*, *6*(5), e20179.

Patel, J. C., Taylor, S. M., Juliao, P. C., Parobek, C. M., Janko, M., Gonzalez, L. D., … Meshnick, S. R. (2014). Genetic Evidence of Importation of Drug-Resistant Plasmodium falciparum to Guatemala from the Democratic Republic of the Congo. *Emerging Infectious Diseases*, *20*(6), 932–940.

Roh, M. E., Tessema, S. K., Murphy, M., Nhlabathi, N., Mkhonta, N., Vilakati, S., … Others. (2019). High genetic diversity of Plasmodium falciparum in the low-transmission setting of the Kingdom of Eswatini. *The Journal of Infectious Diseases*, *220*(8), 1346–1354.

Sane, R., Talla, C., Diouf, B., Sarr, F. D., Diagne, N., Faye, J., … Niang, M. (2019). Low genetic diversity and complexity of submicroscopic Plasmodium falciparum infections among febrile patients in low transmission areas in Senegal. *PloS One*, *14*(4), e0215755.

Sarah-Matio, E. M., Guillochon, E., Nsango, S. E., Abate, L., Ngou, C. M., Bouopda, G. A., … Morlais, I. (2022). Genetic Diversity of Plasmodium falciparum and Distribution of Antimalarial Drug Resistance Mutations in Symptomatic and Asymptomatic Infections. *Antimicrobial Agents and Chemotherapy*, *66*(8), e0018822.

Searle, K. M., for the Southern Africa International Centers of Excellence for Malaria Research, Katowa, B., Kobayashi, T., Siame, M. N. S., Mharakurwa, S., … 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*, Vol. 16. doi: 10.1186/s12936-

017-1810-3

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*(6), 1014–1017.

Sumner, K. M., Freedman, E., Abel, L., Obala, A., Pence, B. W., Wesolowski, A., … Taylor, S. M. (2021). Genotyping cognate Plasmodium falciparum in humans and mosquitoes to estimate onward transmission of asymptomatic infections. *Nature Communications*, *12*(1), 909.

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*, 287.

Topazian, H. M., Moser, K. A., Ngasala, B., Oluoch, P. O., Forconi, C. S., Mhamilawa, L. E., … Juliano, J. J. (2022). Low complexity of infection is associated with molecular persistence of Plasmodium falciparum in Kenya and Tanzania. *Frontiers in Epidemiology*, *2*. doi: 10.3389/fepid.2022.852237

Verity, R., Aydemir, O., Brazeau, N. F., Watson, O. J., Hathaway, N. J., Mwandagalirwa, M. K., … Juliano, J. J. (2020). The Impact of Antimalarial Resistance on the Genetic Structure of Plasmodium falciparum in the DRC. *Nature Communications*, *11*(1), 2107.

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.

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.

507 **TABLES**
508
509 **Table 1. Blood samples from Zanzibar and coastal Tanzania†.**

Description	Location (District)	Dates	Symptomatic (S) or Asymptomatic (A)	Number of Samples	Age range (yr)
Community cross-sectional surveys	Zanzibar (Multiple)	2016	A	70	2-70
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
Parasite clearance study of artemether-lumefantrine (AL)	Mainland (Bagamoyo)	2018	S	138	2-11
Study of transmission of *Plasmodium falciparum* to colony reared mosquitos	Mainland (Bagamoyo)	2018	A	40	7-16

510 †see **Supplemental Figure 1** for full sample details.

511

512 **Table 2. Drug resistance polymorphism prevalence in Zanzibar and coastal mainland Tanzania.**

Mutation	Zanzibar			Mainland		
	Mutant Allele Prevalence‡	CI‡	# Genotyped Samples§	Mutant Allele Prevalence ‡	CI‡	# Genotyped Samples§
Pfcrt-M74I	0.02	0.003-0.043	140	ND†	-	70
Pfcrt-N75E	0.019	0.003-0.042	140	ND†	-	70
Pfcrt-K76T	0.02	0.002-0.043	140	ND†	-	70
Pfdhfr-A16V	ND†	-	142	ND†	-	72
Pfdhfr-N51I	0.969	0.94-0.993	146	0.913	0.853-0.965	76
Pfdhfr-C59R	0.922	0.884-0.954	146	0.895	0.827-0.955	75
Pfdhfr-S108N	1	1-1	141	1	1-1	71
Pfdhfr-S108T	ND†	-	141	ND†	-	71
Pfdhfr-I164L	ND†	-	149	ND†	-	66
Pfdhps-A437G	0.915	0.872-0.952	144	0.902	0.845-0.954	100
Pfdhps-K540E	0.917	0.876-0.955	145	0.92	0.86-0.968	84
Pfdhps-A581G	0.03	0.008-0.055	147	0.077	0.035-0.126	108
Pfk13-K189N	0.019	0.002-0.044	137	ND†	-	57
Pfk13-K189T	0.074	0.036-0.118	135	0.116	0.042-0.205	54
Pfmdr1-N86Y	0.013	0-0.032	143	0.007	0-0.021	113
Pfmdr1-Y184F	0.513	0.445-0.585	153	0.44	0.343-0.526	106
Pfmdr1-D1246Y	0.01	0-0.026	153	0.012	0-0.038	79

513 †ND = Not Detected

514 ‡Mean prevalence and 95% CI of polymorphisms calculated through a nonparametric bootstrap are
515 shown below based on region, either in Zanzibar or mainland Tanzania.

516 §The number of genotyped samples per loci is also shown for each polymorphism.
517

FIGURES

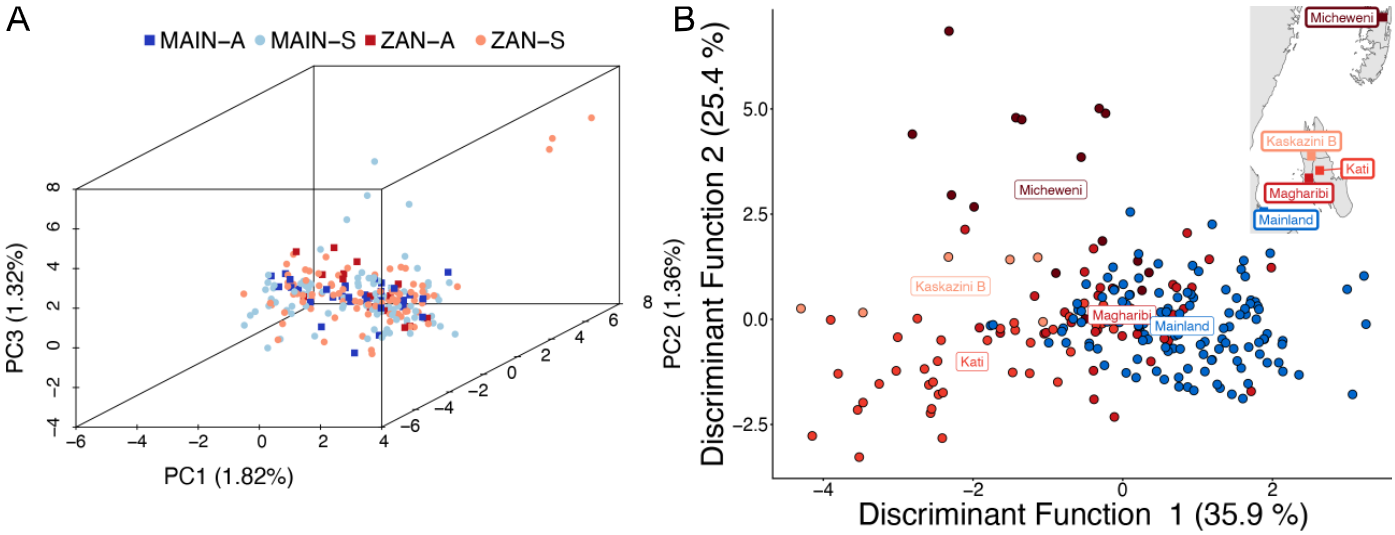


Figure 1. Parasites between Zanzibar and coastal mainland Tanzania are highly related but microstructure within Zanzibar is apparent. A) Principal Component Analysis (PCA) comparing parasites from symptomatic vs. asymptomatic patients from coastal Tanzania and Zanzibar. Clusters with an identity by descent (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 and including districts with at least 5 samples present. 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 *shehias* was performed using geographic coordinates of samples to generate 5 clusters (colored boxes). 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.

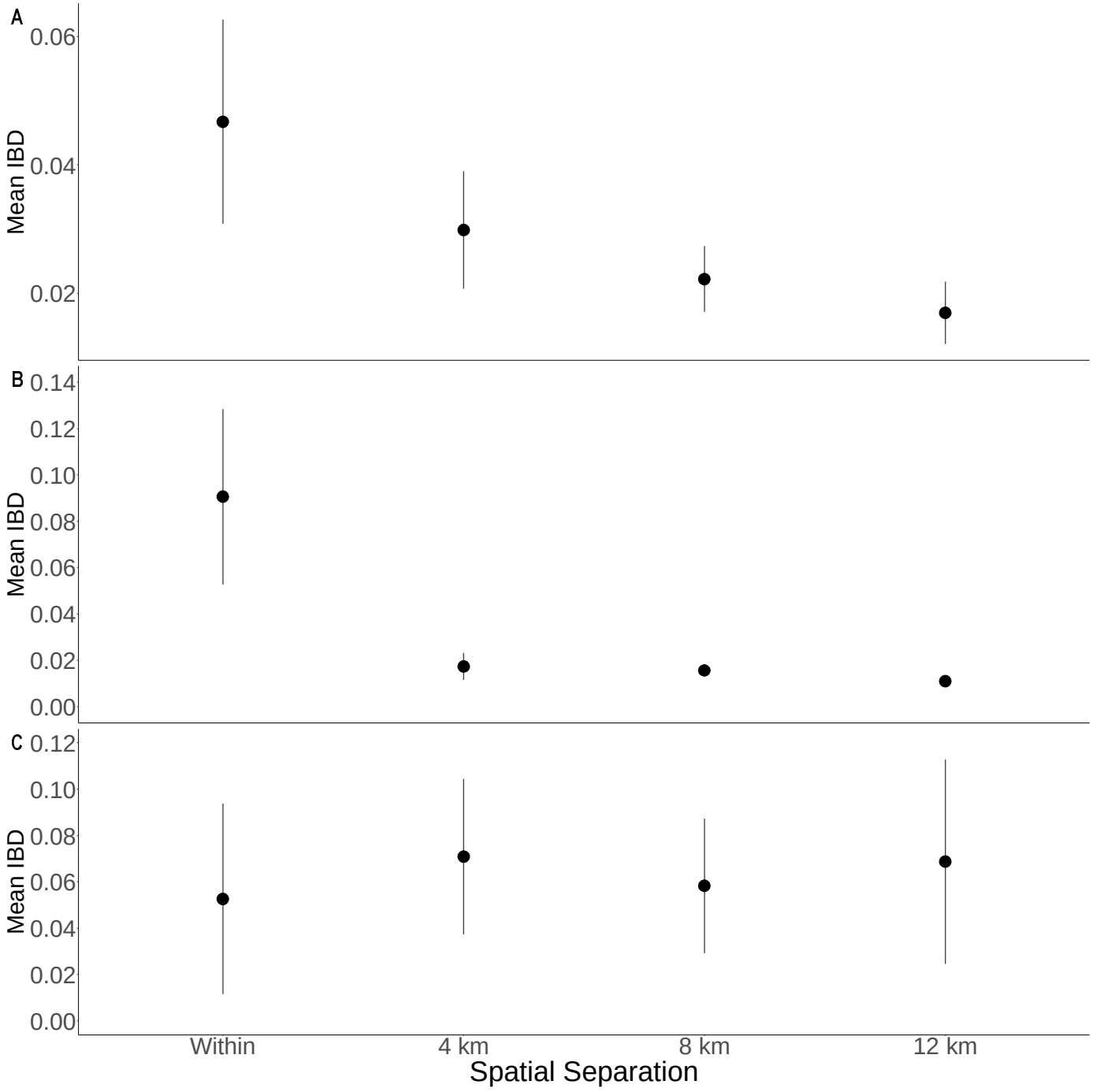
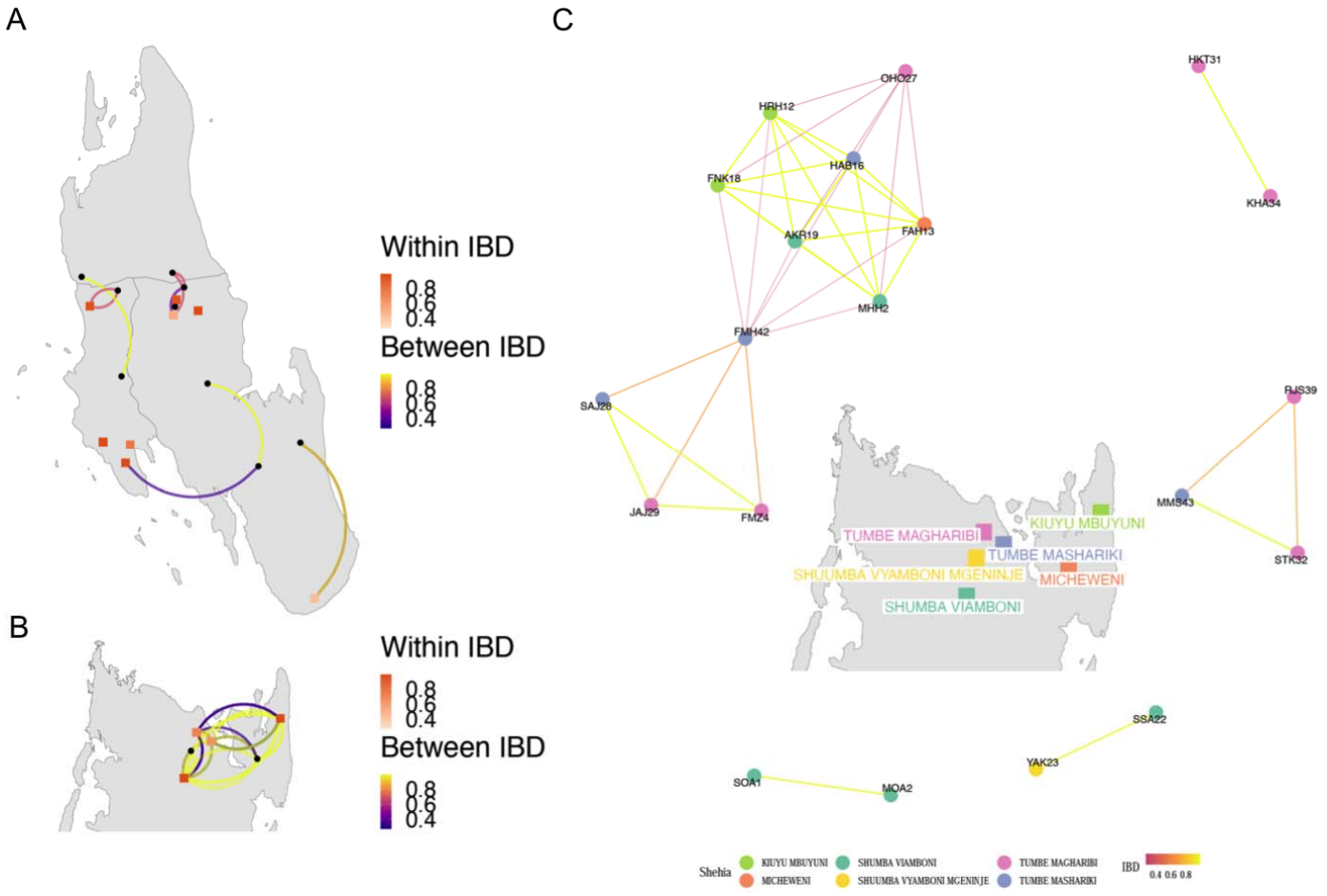


Figure 3. Isolation by distance is shown between all Zanzibari parasites (A), only Unguja parasites (B) and only Pemba parasites (C). Samples were analyzed based on geographic location, Zanzibar (N=136) (A), Unguja (N=105) (B) or Pemba (N=31) (C) and greater circle (GC) distances between pairs of parasite isolates were calculated based on *shehia* centroid coordinates. For (A) and (B), 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 and 58km on Unguja. For (C), these distances were binned at 4km increments due to smaller amounts of pairwise comparisons due to fewer samples being available within Pemba compared to Panel A and B. The maximum GC distance for Pemba was 12km. The mean IBD and 95% CI is plotted for each bin.



548

549 **Figure 4. Highly related pairs span long distances across Zanzibar.** Sample pairs were filtered to
550 have IBD estimates of 0.25 or greater. Within *shehia* pairwise IBD estimates are shown within Unguja
551 **(Panel A)** and Pemba **(Panel B)** as single points, with dark orange representing the greatest degree of
552 IBD. Between *shehia* IBD reflects pairs of parasites with IBD greater than or equal to 0.25, with the color
553 of the connecting arc representing the degree of IBD and yellow representing maximal connectivity. **Panel**
554 **C** shows the network of highly related pairs (IBD >0.25) within and between the 6 northern Pemba *shehias*
555 (note: Micheweni is a *shehia* in Micheweni district). Samples (nodes) are colored by *shehia* and IBD
556 estimates (edges) are represented on a continuous scale with increasing width and yellow-shading
557 indicating higher IBD.

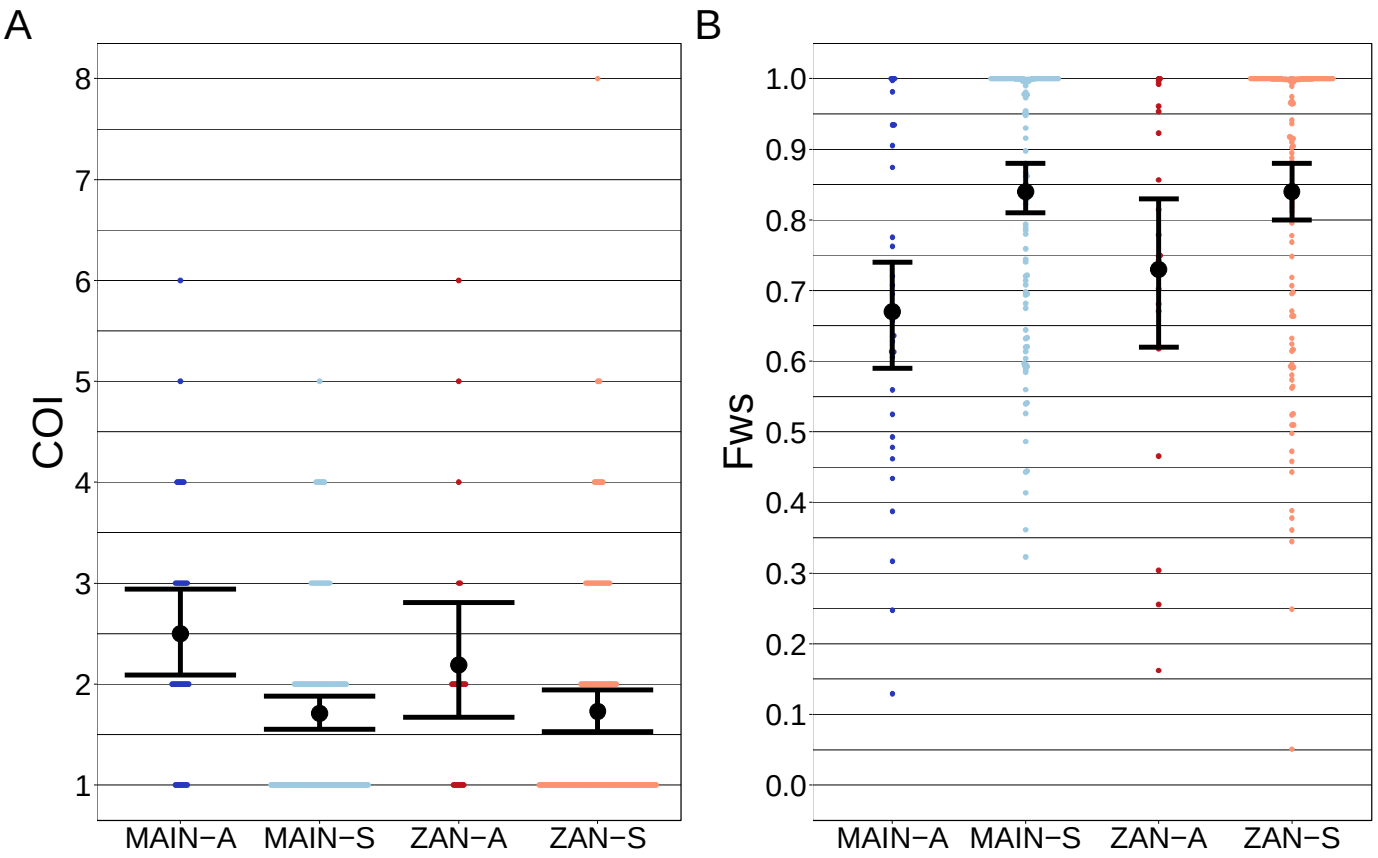


Figure 5. Complexity of infection (COI) and Fws metric shows a higher COI and lower Fws in asymptomatic than symptomatic infections in both mainland Tanzania and Zanzibar isolates. COI (**A**) was estimated by the REAL McCOIL's categorical method (H.-H. Chang et al., 2017). Mean COI for asymptomatic 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.