Falciparum malaria from coastal Tanzania and Zanzibar remains highly connected despite effective control efforts on the archipelago

Andrew P. Morgan1†, Nicholas F. Brazeau2†, Billy Ngasala3, Lwidiko E. Mhamilawa3,4, Madeline Denton1, Mwinyi Msellem5, Ulrika Morris6, Dayne L. Filer7, Ozkan Aydemir8, Jeffrey A. Bailey8, Jonathan B. Parr1, Andreas Mårtensson4, Anders Bjorkman6 and Jonathan J. Juliano1,2,9*

Abstract

Background: Tanzania’s Zanzibar archipelago has made significant gains in malaria control over the last decade and is a target for malaria elimination. Despite consistent implementation of effective tools since 2002, elimination has not been achieved. Importation of parasites from outside of the archipelago is thought to be an important cause of malaria’s persistence, but this paradigm has not been studied using modern genetic tools.

Methods: Whole-genome sequencing (WGS) was used to investigate the impact of importation, employing population genetic analyses of Plasmodium falciparum isolates from both the archipelago and mainland Tanzania. Ancestry, levels of genetic diversity and differentiation, patterns of relatedness, and patterns of selection between these two populations were assessed by leveraging recent advances in deconvolution of genomes from polyclonal malaria infections.

Results: Significant decreases in the effective population sizes were inferred in both populations that coincide with a period of decreasing malaria transmission in Tanzania. Identity by descent analysis showed that parasites in the two populations shared long segments of their genomes, on the order of 5 cM, suggesting shared ancestry within the last 10 generations. Even with limited sampling, two of isolates between the mainland and Zanzibar were identified that are related at the expected level of half-siblings, consistent with recent importation.

Conclusions: These findings suggest that importation plays an important role for malaria incidence on Zanzibar and demonstrate the value of genomic approaches for identifying corridors of parasite movement to the island.

Keywords: Plasmodium, Malaria, Population genetics

Background

Despite nearly two decades of progress in control, malaria remains a major public health challenge with an estimated 219 million cases and 435,000 deaths in 2017 globally [1]. The mainland of Tanzania has heterogeneous transmission of mainly Plasmodium falciparum, but overall levels of malaria remain high, accounting for approximately 3% of global malaria cases [1]. However,
through a combination of robust vector control and access to efficacious anti-malarial treatment, the archipelago of Zanzibar has been deemed a pre-elimination setting, having only low and mainly seasonal transmission [2]. Despite significant efforts, however, elimination has been difficult to achieve in Zanzibar. The reasons for Zanzibar’s failure to achieve elimination are complex and likely driven by several key factors: (1) as transmission decreases, the distribution of cases changes and residual transmission is more focal and mainly outdoors [3]; (2) a significant number of malaria infections are asymptomatic and thus untreated and remain a source for local transmission [4–7]; and (3) the archipelago has a high level of connectivity with the mainland, thus imported malaria through human travel may play an increasing relative role in transmission.

Genomic epidemiology can supplement traditional epidemiological measures in studies of malaria transmission and biology, thereby helping to direct malaria elimination strategies [8]. Whole-genome sequencing (WGS) can be particularly useful for understanding the history of parasite populations and movement of closely related parasites over geographical distances [9, 10]. Identity by descent (IBD), the sharing of discrete genomic segments inherited from a common genealogical ancestor, has been found to be a particularly good metric for studying the interconnectivity of parasite populations [11–13]. A major obstacle to studying IBD in microorganisms, and in particular malaria, is the presence of multiple clones in a single infection. In order to address this obstacle, recent algorithms have been developed to deconvolve multiple infections into their respective strains from Illumina sequence data [14, 15]. These advances now make it tractable to conduct population genetic analysis of malaria in regions of higher transmission, where infections are often polyclonal.

Decreases in malaria prevalence are hypothesized to be associated with increasing inbreeding in the parasite population, decreased overall parasite genetic diversity and a reduced complexity of infection (COI), defined as a decreased number of infecting clones [8]. This has been shown in pre-elimination settings in Asia as well as in lower transmission regions of Africa [16–18]. It has not been determined if a similar reduction in diversity has occurred in Zanzibar with the significant reduction of malaria in the archipelago. WGS data was used to: (1) characterize the ancestry of parasites in the two regions, (2) determine the levels of genetic diversity and differentiation between archipelago and mainland, (3) determine patterns of relatedness and inbreeding and (4) search for signatures of adaptation and natural selection. Inferred genetic relationships were then examined for evidence of importation of parasites from the higher transmission regions of mainland Tanzania to the lower transmission regions of the Zanzibar archipelago. These findings improve understanding of how importation may affect malaria elimination efforts in Zanzibar.

**Methods**

**Clinical samples**

WGS was attempted on 106 *P. falciparum* isolates collected from subjects with uncomplicated malaria or asymptomatic infection from 2015 to 2017. Forty-three of these were leukodepleted blood collected as part of an in vivo efficacy study of artemether–lumefantrine (AL) in paediatric uncomplicated malaria patients collected from 2015–2017 in Yombo, Bagamoyo District. A remaining 63 isolates were from dried blood spots (DBS) collected in Zanzibar in 2017. These came from cross-sectional surveys of asymptomatic individuals (*n* = 34) and an in vivo efficacy study of artesunate–amodiaquine (ASAQ) with single low dose primaquine (SLDP) in paediatric uncomplicated malaria patients (*n* = 29). These isolates essentially represent a convenience sample. Isolates were not selected for sequencing on the basis of specific clinical or epidemiologic characteristics; however, sequencing was more likely to be successful on isolates from subjects with high parasitaemia. Study participants from Zanzibar were asked to report any overnight travel away from home in the past 4 months. Responses were coded as yes (overnight travel to mainland Tanzania or Kenya) or no (no overnight travel off islands of Zanzibar). Clinical characteristics of the attempted and sequenced samples from each cohort from Zanzibar is provided in Additional file 1: Table S1.

**Generation and sequencing of libraries**

Leukodepleted blood samples and DBS were extracted using QIAamp 96 DNA blood kits per the manufacturer protocol (Qiagen, Hilden, Germany). DNA from leukodepleted blood was acoustically sheared using a Covaris E220 instrument, prepared for sequencing without enrichment using Kappa Hyper library preps, and individually barcoded per manufacturer’s protocol (Kappa Biosystems, Columbus, OH). DNA extracted from DBS was enriched for *P. falciparum* DNA before library prep using two separate selective whole genome amplification (sWGA) reactions. The sWGA approach was adapted from previously published methods and employed two distinct sets of primers designed for *P. falciparum*, including the Probe_10 primer set described previously by Oyola *et al.* and another set of custom primers (JP9) designed using ‘swga’ [19–21]. Phosphorothioate bonds were included between the two most 3’ nucleotides for all primers in both sets to prevent primer degradation. Design and evaluation of these custom primers
and the sWGA approach are described in Additional file 1: Table S2. The two sWGA reactions were carried out under the same conditions. The products of the two sWGA reactions were pooled in equal volumes and acoustically sheared using a Covaris E220 instrument before library preparation using Kappa Hyper library preps. The indexed libraries were pooled and sequenced on a HiSeq 4000 using 2 × 150 chemistry at the University of North Carolina High Throughput Sequencing Facility. Sequencing reads were deposited into the NCBI SRA (Accession numbers: pending).

**Public sequencing data**

Illumina short read WGS data for *P. falciparum* isolates was downloaded from public databases. This included 68 isolates from other regions of Tanzania, collected between 2010 and 2013, as well as 179 isolates from other regions, including Southeast Asia, South Asia, East and West Africa (Additional file 1: Table S3).

**Read alignment and quality control**

Raw paired-end reads were trimmed for adapter sequences with ‘cutadapt’ v1.18 and aligned to the *P. falciparum* 3D7 reference genome (assembly version 3, PlasmoDB version 38: https://plasmodb.org/common/downloads/release-38/Pfalciparum3D7/fasta/data/Plasm oDB-38_Pfalciparum3D7_Genome.fasta) with ‘bwa mem’ v0.7.17-r1188. Duplicates were marked with ‘samtools’ v0.1.24. A position was defined as “callable” if it was covered by ≥ 5 high-quality reads (MQ ≥ 25, BQ ≥ 25), and computed the proportion of callable sites in each isolate was calculated with the Genome Analysis Toolkit (GATK) ‘CallableLoci’ tool v3.8-0. Only isolates with ≥ 70% of the genome callable were used for further analysis.

**Variant discovery and filtering**

Short sequence variants (including SNVs, indels and complex multi-nucleotide variants) were ascertained in parallel in each isolate using GATK ‘HaplotypeCaller’ v.4.0.3.0, then genotyped jointly across the entire cohort with GATK ‘GenotypeGVCFs’ according to GATK best practices. Variant discovery was limited to the core (non-hypervariable) nuclear genome as defined by Miles et al. [22]. Putative SNVs only were filtered using the GATK Variant Quality Score Recalibration (VQSR) method. For training sets, the following datasets were used: QC-passing sites from the *P. falciparum* Genetic Crosses Project release 1.0 (ftp://ngs.sanger.ac.uk/production/malaria/pfcrosses/1.0/); [22]) (true positives, prior score Q30); QC-passing sites from the PGK release v5.1 (ftp://ngs.sanger.ac.uk/production/pf3k/release_5/5.1/) (true positives + false positives, prior score Q15). Site annotations QD, MQ, MQRankSum, ReadPosRankSum, FS, SOR were used and the model was trained with 4 Gaussian components. A VQSLOD threshold — 0.0350 achieved 90% sensitivity for re-discovering known sites in the training sets. All biallelic SNVs with VQSLOD at or above this threshold were retained.

Isolates may contain multiple strains that are haploid resulting in mixed infections with arbitrary effective ploidy. To account for this complexity of infection (COI), prior literature was followed [23] and the following quantities were calculated at each variant site: for each isolate, the within-sample allele frequency (WSAF), the proportion of mapped reads carrying the non-reference allele; the population-level allele frequency (PLAF), the mean of within-sample allele frequencies; and the population-level minor allele frequency (PLMAF), the minimum of PLAF or 1-PLAF. These calculations were performed with ‘vcfdosaf’ (https://github.com/IDEELResearch/vcfdosaf).

**Analyses of mutational spectrum**

Ancestral versus derived alleles at sites polymorphic in *P. falciparum* were assigned by comparison to the outgroup species *Plasmodium reichenowi*. Briefly, an approximation to the genome of the *P. reichenowi*—*P. falciparum* common ancestor (hereafter, “ancestral genome”) was created by aligning the *P. falciparum* 3D7 assembly to the *P. reichenowi* CDC strain assembly (version 3, PlasmoDB version 38: https://plasmodb.org/common/downloads/release-38/PreichenowiCDC/fasta/data/PlasmoDB-38_PreichenowiCDC_Genome.fasta) with ‘nucmer’ v3.1 using parameters “-g 500 -c 500 -l 10” as in [24]. Only segments with one-to-one alignments were retained; ancestral state at sites outside these segments was deemed ambiguous. The one-to-one segments were projected back into the 3D7 coordinate system. Under the assumption of no recurrent mutation, any site polymorphic in *P. falciparum* is not expected to also be mutated on the branch of the phylogeny leading to *P. reichenowi*. Thus, the allele observed in *P. reichenowi* is the ancestral state conditional on the site being polymorphic. Transitions-transversion (Ti:Tv) ratios and mutational spectra were tallied with ‘bcftools stats’ v1.19.

**Analyses of ancestry and population structure**

VQSR-passing sites were filtered more stringently for PCA to reduce artifacts due to rare alleles and missing data. Genotype calls with GQ < 20 or DP < 5 were masked; sites with < 10% missing data and PLMAF > 5% after sample-level filters were retained for PCA, which was performed with ‘akt pca’ v3905c48 [25]. For calculation of f3 statistics, genotype calls with GQ < 10 or DP < 5 were masked; sites with < 10% missing data and PLMAF...
> 1% after sample-level filters were retained. Then f_3 statistics were calculated from WSAsFs rather than nominal diploid genotype calls, using ‘vcfdo f3stat’.

**Estimation of sequence diversity**

Estimates of sequence diversity and differentiation were obtained from the site-frequency spectrum (SFS), which in turn was estimated directly from genotype likelihoods with ‘ANGSD’ 0.921-11-g20b0655 [26] using parameters “-ANGSD” 0.921-11-g20b0655 [26] using parameters “-doCounts 1 -doSaf 1 -GL 2 -minD-

Complexity of infection (COI) and strain deconvolution

Strain deconvolution and inheritance-by-descent analyses

Analyses of natural selection

Estimation of sequence diversity and differentiation were obtained from the site-frequency spectrum (SFS), which in turn was estimated directly from genotype likelihoods with ‘ANGSD’ 0.921-11-g20b0655 [26] using parameters “-doCounts 1 -doSaf 1 -GL 2 -minD-

Demographic inference

Curves of recent historical effective population size were estimated from between-isolate IBD segments with ‘IBDNe’ v07May18-6a4 [29] using length threshold > 3 cM, 20 bootstrap replicates and default parameters otherwise. Local age-adjusted parasite prevalence point estimates (\( \phi PR_{2-10} \)) and credible intervals were obtained from the Malaria Atlas Project [30] via the R package ‘malariaAtlas’ [31].

More remote population-size histories were estimated with ‘smc++’ v1.15.2 [32]. Phased haplotypes from ‘dePloid’ were randomly combined into diploids and parameters estimated separately for mainland Tanzania and Zanzibar populations using fivefold cross-validation via command ‘smc++ cv’, with mutation rate set to \( 10^{-9} \) bp\(^{-1}\) gen\(^{-1}\). Marginal histories from each population were then used to estimate split times using ‘smc++ split’.

Analyses of natural selection

The distribution of fitness effects (DFE) was estimated within mainland Tanzania and Zanzibar populations with ‘polyDFE’ v2.0 using fourfold degenerate sites as putatively-neutral and zerofold degenerate sites as putatively-selected [33]. “Model C” in ‘polyDFE’ parlance—a mixture of a gamma distribution on selection coefficients of deleterious mutations and an exponential distribution for beneficial mutations—was chosen because it does not require a priori definition of discrete bins for selection coefficients, and the gamma distribution can accommodate a broad range of shapes for the DFE of deleterious mutations (expected to represent the bulk of polymorphic sites). Confidence
intervals for model parameters were obtained by non-parametric bootstrap via 20 rounds of resampling over the 100 kb blocks of the input SFS. Because ‘poly-DFE’ fits nuisance parameters for each bin in the SFS, computation time increased and numerical stability decreased for SFS with larger sample sizes. Input SFS were, therefore, smoothed and rescaled to pre-specified sample size of 10 chromosomes each using an empirical-Bayes-like method (https://github.com/CartwrightLab/SoFoS/) re-implemented in ‘sfspy smooth’. Smoothing of input SFS had very modest qualitative effect on the resulting DFE.

The cross-population extended haplotype homozygosity (XP-EHH) statistic was used to identify candidate loci for local adaptation in mainland Tanzania or Zanzibar. Because the statistic requires phased haplotypes and is potentially sensitive to phase-switch errors, only isolates with COI = 1 were used (n = 18 mainland Tanzania, n = 12 Zanzibar). XP-EHH was calculated from haploid genotypes at a subset of 103,982 biallelic SNVs polymorphic among monoclonal isolates with the ‘xpehhbin’ utility of ‘hapbin’ v1.3.0-12-gdb383ad [34]. Raw values were standardized to have zero mean and unit variance; the resulting z-scores are known to have an approximately normal distribution [35] so nominal p-values were assigned from the standard normal distribution. The Benjamini-Hochberg method was used to adjust nominal p-values for multiple testing.

Fig. 1 Ancestry of P. falciparum in Zanzibar and mainland Tanzania. a Location for samples used in this study, coloured by population: orange, mainland Tanzania; purple, Zanzibar; dark grey, published mainland Tanzania isolates from the MalariaGEN P. falciparum Community Project. Other major regional cities show with open circles. b Major axes of genetic differentiation between global P. falciparum populations demonstrated by principal components analysis (PCA) on genotypes at 7122 SNVs with PLMAF > 5%. Each point represents a single isolate (n = 304) projected onto the top two principal components (71% cumulative variance explained); colour-shape combinations indicate country of origin. c-e Population relationships assessed by $f_3$ statistics with focal population indicated at the top of each panel, comparator populations on the vertical axis, and Congolese population as an outgroup. Error bars show 3 times the standard error computed by block-jackknife.
regions and accessible by short-read sequencing [22]—were retained for further analysis. The frequency spectrum was dominated by rare alleles: 151,664 alleles (39.1%) were singletons and 310,951 (80.2%) were present in < 1% of isolates in the dataset. Ancestral and derived states at 361,049 sites (93.1%) were assigned by comparison to the *P. reichenowi* (CDC strain) genome, treating the *reichenowi* allele as ancestral. Similar biases were observed in the mutational spectrum as have been estimated directly from mutation-accumulation experiments [37]: transitions are more common transversions (Ti:Tv = 1.12; previous estimate 1.13), with a large excess of G:C>A:T changes even after normalizing for sequence composition (Additional file 1: Fig. S1). Consistency in the mutational spectrum between independent studies, using different methods for sample preparation and different bioinformatics pipelines, supports the accuracy of genotype calls.

**Ancestry of mainland Tanzania and Zanzibar isolates**

In order to place new isolates in the context of global genetic variation in *P. falciparum*, principal components analysis (PCA) was performed with existing isolates from around the globe (Fig. 1b). A subset of 7122 stringently-filtered sites with PLMAF > 5% (see “Methods”) were retained for PCA to minimize distortion of axes of genetic variation by rare alleles or missing data. Consistent with existing literature, isolates separated into three broad clusters corresponding to southeast Asia, east Africa and west Africa. Mainland Tanzania and Zanzibar isolates fell in the east Africa cluster. This observation was formalized using *f*₃ statistics [38, 39], which measure shared genetic variation in a pair of focal populations *A* and *B* relative to an outgroup population *O*. By calculating *f*₃ across different combinations of comparator populations and holding the outgroup fixed, one can build up an idea of the ancestry of the populations of interest: pairs with relatively larger positive values of *f*₃ are more genetically similar than pairs with relatively smaller *f*₃. The new isolates from Yombo and Zanzibar and published Tanzanian isolates shared mutually greater genetic affinity for each other than for other populations in the panel (Fig. 1c–e); isolates from neighbouring countries Malawi and Kenya were next-closest. Together these analyses support an east African origin for parasites in mainland Tanzania and in Zanzibar.

**Genetic diversity and differentiation**

In order to better understand the population demography and effects of natural selection in the parasite populations, indices of genetic diversity within populations, and the degree to which that diversity is shared across populations, were examined. The genome was partitioned into

---

**Results**

WGS and variant discovery

Genomic data for *P. falciparum* was generated using leukodepleted blood collected from 43 subjects from Yombo, Tanzania (“mainland”) and from DBS collected from 63 subjects from the Zanzibar archipelago (“Zanzibar”; Fig. 1a) using selective whole-genome amplification (sWGA) followed by Illumina sequencing. Thirty-six isolates (84%) from the mainland and 21 isolates (33%) from Zanzibar yielded sufficient data for analysis. These 57 genomes were combined with an additional 68 published genomes from other sites in Tanzania in the MalariaGEN *P. falciparum* Community Project (PICP) and 179 genomes from other sites in Africa and Asia, representing a broad geographic sampling of Africa and Asia [36]. Single-nucleotide variants (SNVs) were ascertained jointly in the global cohort. After stringent quality control on 1.3 million putative variant sites, a total of 387,646 biallelic SNVs in the “core genome”—the 20.7 Mb of the 3D7 reference assembly lying outside hypervariable
four sequence classes—all sites in the core genome; four-fold degenerate ("synonymous") sites; zero-fold degenerate ("nonsynonymous") sites; and coding sites in genes associated with resistance to antimalarial drugs—and several estimators of sequence diversity were calculated in each class (see "Methods"). Levels of sequence diversity at synonymous (putatively neutral) sites were very similar within mainland Tanzania and Zanzibar isolates (theta_pi = 9.0 \times 10^{-4} [95% CI 8.6 \times 10^{-5}–9.4 \times 10^{-4}] vs. 8.4 [95% CI 8.0 \times 10^{-4}–8.7 \times 10^{-4} per site]) and 1.3-fold lower than among previously-published Tanzanian isolates (Fig. 2a). As expected, diversity was lower at non-synonymous sites, which are more likely to be under purifying selection. Tajima's selection. Tajima's non-synonymous sites, which are more likely to be under purifying selection. Tajima's D took negative values in all isolates (Fig. 2a). As expected, diversity was lower among previously-published Tanzanian isolates (median Fst between southeast Asian and African isolates was just 0.0289 [95% bootstrap CI 0.0280–0.0297]; the distribution of Fst between southeast Asian and African isolates is on the order of 0.20 [23]. Thus there exists minimal evidence for genetic differentiation between parasites in mainland Tanzania and Zanzibar.

Patterns of relatedness and inbreeding

Long segments of the genome are shared IBD both within and between isolates. Mean within-isolate segment length was 5.7 cM (95% CI 4.1–7.3 cM, n = 117) on the mainland and 3.7 cM (95% CI 2.8–4.6 cM, n = 80) on Zanzibar in a linear mixed model with individual-level random effects; the full distributions are shown in Fig. 3d. Segments shared between isolates within the mainland population (6.2 cM, 95% CI 5.9–6.6 cM, n = 3279) were longer than segments shared within Zanzibar (4.5 cM, 95% CI 4.1–4.8 cM, n = 592) or between mainland and Zanzibar populations (4.1 cM, 95% CI 3.9–4.3 cM, n = 6506). After accounting for differences in segment length by population, difference in lengths of IBD segments detected between versus within individuals are not significant (mean difference = 0.038 cM, 95% CI −0.10 to 0.23 cM). In a random-mating population the length of a segment shared IBD between a pair of individuals with last common ancestor G generations in the past is exponentially-distributed with mean 100/(2*G) cM. The shared haplotype that observed, with length on the order of 5 cM, are thus consistent with shared ancestry in the past 10 generations—although as many as half of such segments probably date back at least 20 generations [41]. In the presence of inbreeding, IBD sharing persists even longer in time.

Close relationships between isolates from the archipelago and the mainland suggest recent genetic exchange. A threshold of F_{IBD} > 0.25 (half-siblings) was chosen because it implies that two isolates shared at least one common parent in the last outcrossing generation and, therefore, are related as recently as the last 1–2 transmission cycles, depending on background population dynamics. In principle this could result from importation of either insect vectors or human hosts. To investigate the latter possibility, a travel-history questionnaire completed by subjects from Zanzibar was used. Nine subjects reported travel to the mainland in the month before study enrollment; their destinations are shown in Fig. 4a. Ten pairs with F_{IBD} > 0.25 (marked by orange triangles in histogram in Fig. 4b) were identified; all involved a single Zanzibar isolate from a patient who travelled to the coastal town of Mtwara (orange arc in Fig. 4a). It is very likely that this individual represents an imported
Overall, isolates from travellers had slightly higher mean pairwise relatedness to isolates from the mainland (mean $F_{IBD} = 0.0020$, 95% CI 0.0018–0.0021) than did isolates from non-travellers (mean $F_{IBD} = 0.0015$, 95% CI 0.0014–0.0016; Wilcoxon rank-sum test $p = 1.8 \times 10^{-12}$ for difference). But these relationships—spanning 10 or more outcrossing generations—are far too remote to be attributed to the period covered by the travel questionnaire. The pattern likely represents instead the presence of subtle population structure within Zanzibar.

**Demographic history of parasite populations**

The distribution of IBD segment lengths carries information about the trajectory of effective population size in the recent past, up to a few hundred generations before the time of sampling. The site frequency spectrum and patterns of fine-scale linkage disequilibrium carry information about the more remote past. Complementary methods were used to infer recent and remote population demography from phased haplotypes. First, a non-parametric method was applied [29] to infer recent effective population size ($N_e$) from IBD segment lengths separately in mainland Tanzania and Zanzibar populations (Fig. 5a). The method infers a gradual decline of several orders of magnitude in $N_e$ over the past 100 generations to a nadir at $N_e \sim = 5000$ around 15–20 outcrossing generations before the time of sampling. Although the confidence intervals are wide, similar trajectories are inferred in all
three populations (Zanzibar, new mainland Tanzania isolates and published Tanzanian isolates).

Second, more remote population size histories were inferred jointly for mainland Tanzania and Zanzibar and used to estimate the split time between these populations using a sequentially Markovian coalescent method [32]. This family of models has good resolution for relatively remote events, but less precision in the recent past than models based on IBD segments. The result (Fig. 5b) supports a common ancestral population with \( N_e \sim 10^5 \) individuals that underwent a sharp bottleneck followed by rapid growth around 50,000 generations before the present. The time at which the mainland and Zanzibar populations diverged could not be estimated precisely and may have been as recent as 50 or as ancient as 50,000 generations before the present. Trends in \( N_e \) were compared to local trends in parasite prevalence from the Malaria Atlas Project [30] (Fig. 5c). Assuming an interval of approximately 12 months per outcrossing generation [42], the contraction in \( N_e \) may correspond in time to the decrease in prevalence brought about by infection-control measures over the past two decades.

**Natural selection and adaptation**

Finally, several approaches were taken to characterize the effects of natural selection on sequence variation in mainland and Zanzibar populations. The fate of a new mutation—whether it spreads and ultimately becomes fixed, or is lost—is determined by its selection coefficient \( s \), scaled by the effective population size \( N_e \). The distribution of fitness effects (DFE) describes the distribution of \( s \) and can be estimated from the frequency spectrum at putatively-neutral (synonymous) and putatively-selected (non-synonymous) sites (Fig. 6a). Building on previous work in other organisms, the DFE was modelled in each population as a mixture of a gamma distribution (for deleterious mutations, \( N_e s < 0 \)) and an exponential distribution (for beneficial mutations, \( N_e s > 0 \)) [33]. The inference was performed using both the raw SFS and a smoothed representation of the SFS that is more numerically stable and found that results to be similar with both methods. Fitted parameter values are provided in Additional file 1: Table S5, but the discretized representation of the DFE is more amenable to qualitative comparisons (Fig. 6b).

Differences in the DFE between mainland Tanzania and Zanzibar populations are not statistically significant. The great majority of new mutations (mainland: 74%; Zanzibar: 76%) are expected to be very weakly deleterious (\( -0.01 < 4N_e s < 0 \)), and only a small minority are expected to be beneficial (\( 4N_e s > 0 \)) (mainland: 4.5% [95% CI 2.7–29%]; Zanzibar: 2.4% [95% CI 0.56–50%]). The DFE also allows us to estimate that 8.8% (mainland) and 5.2% (Zanzibar) of substitutions since the common ancestor with *P. reichenowi* have been fixed by positive selection; this quantity is known in some contexts as the “rate of adaptive evolution.”
Although the DFE tells us the proportion of polymorphic sites under positive selection, it does not pinpoint which sites those are. To identify signals of recent, population-specific positive selection, the XP-EHH statistic between mainland and Zanzibarian isolates were used [35]. Outliers in the XP-EHH scan, defined as standardized XP-EHH scores above the 99.9th percentile, represent candidates for local adaptation (Additional file 1: Fig. S2). One-hundred four biallelic SNPs in 20 distinct genes passed this threshold (Additional file 1: Table S6). None of these have been associated with resistance to anti-malarial drugs—an important form of local adaptation in this species—but one (PF3D7_0412300) has been identified in a previous selection scan [43]. Prevalences of 54 known drug-resistance alleles are shown in Additional file 1: Table S7 and are similar to previous reports in East Africa [44–46]. None of these loci had $F_{st} > 0.05$ between mainland Tanzania and Zanzibar.

**Discussion**

Zanzibar has been the target of intensive malaria control interventions for nearly two decades following the early implementation of ACT therapies in 2003 [2]. Despite sustained vector control practices and broad access to rapid testing and effective treatment, malaria has not been eliminated from the archipelago [2]. Here WGS of *P. falciparum* isolates from Zanzibar and nearby sites on the mainland was used to investigate ancestry, population structure and transmission in local parasite populations. These data place Tanzanian parasites in a group of east African populations with broadly similar ancestry and level of sequence diversity. There was minimal genomewide signal of differentiation between mainland and Zanzibar isolates.

The most parsimonious explanation for these findings is a source-sink scenario, similar to a previous report in Namibia [47], in which importation of malaria from...
a region of high but heterogeneous transmission (the mainland) is inhibiting malaria elimination in a pre-elimination area (Zanzibar). Using WGS it is shown that the parasite population on the islands remains genetically almost indistinguishable from regions on the mainland of Tanzania. Numerous long haplotypes could be identified that are shared between the populations, on the order of 5 cM, suggesting that genetic exchange between the populations has occurred within the last 10–20 sexual generations. In addition, a Zanzibar isolate is identified that is related at the half-sibling level to a group of mutually-related mainland isolates. This likely represents an imported case and provides direct evidence for recent, and likely ongoing, genetic exchange between the archipelago and the mainland. These observations suggest that parasite movement from the mainland to the archipelago is appreciable and may be a significant hurdle to reaching elimination.

Human migration is critical in the spread of malaria [48], thus the most likely source for importation of parasites into Zanzibar is through human travel to high-risk malaria regions. Multiple studies have been conducted on travel patterns of Zanzibarian residents as it relates to importation of malaria [49–51], one of which estimated that there are 1.6 incoming infections per 1000 inhabitants per year. This is also in accordance with the estimate of about 1.5 imported new infections out of a total of 8 per 1000 inhabitants in a recent epidemiological study [2]. None of these studies have leveraged parasite population genetics to understand importation patterns. Though this study is small, the findings are proof of principle for using genetics to identify specific importation events. These data provides a platform for future genetic surveillance efforts by, for example, design of targeted assays for sequence variants that discriminate mainland from Zanzibari parasites. Such surveillance, including of asymptomatic individuals, would clarify the role of importation versus endemic transmission and potentially identify specific travel corridors to target for interventions. Larger sample sizes would also likely begin to reveal subtle population structure that is not obvious when examining a few dozen isolates.

Malarial infections in Africa are highly polyclonal. This within-host diversity poses technical challenges but also provides information on transmission dynamics. Approximately half of isolates from both the mainland and

![Fig. 6 Characterizing the impact of natural selection on sequence variation.](image-url)

- **a** Site-frequency spectra for putatively neutral (fourfold degenerate) and putatively-selected (zerofold degenerate) sites.
- **b** Inferred distribution of population-scaled selection coefficients ($4N_{es}$) for each population, shown in discrete bins. Dark bars, estimates from raw SFS; light bars, estimates from smoothed SFS. Note logarithmic scale for vertical axis in both panels.
Furthermore, median within-host relatedness (\(F_{\text{IBD}}\)) in the presence of related lineages in the same host. Fisher process [55], but has limited discriminatory power qualitatively consistent with COI estimated by haplotype deconvolution [52], but has limited discriminatory power in the presence of related lineages and the same host. Furthermore, median within-host relatedness (\(F_{\text{IBD}}\)) is 0.25, the expected level for half-siblings, in both mainland and Zanzibar populations. This strongly suggests frequent co-transmission of related parasites in both populations [40]. Estimates of \(F_{\text{IBD}}\) are within the range of estimates from other African populations and add to growing evidence that mixed infections may be predominantly due to co-transmission rather than superinfection even in high-transmission settings [53, 54]. An important caveat of this work is its dependence on statistical haplotype deconvolution. Direct comparison of statistical deconvolution to direct sequencing of single clones has shown that methods like ‘dEploid’ have limited accuracy for phasing the minority haplotype(s) in a mixed infection. Phasing errors tend to limit power to detect IBD between infections, and may cause underestimation of between-host relatedness.

Intensive malaria surveillance over the past several decades provides an opportunity to compare observed epidemiological trends to parasite demographic histories estimated from contemporary genetic data. Estimates of historical effective population size \(N_e\) support an ancestral population of approximately \(10^5\) individuals that grew rapidly around \(10^4\) generations ago, then underwent sharp contraction within the past 100 generations to a nadir around 10–20 generations before the present. Stable estimates of the split time between the mainland and Zanzibar populations could not be obtained, either with a coalescent-based method (Fig. 5b) or with method based on the diffusion approximation to the Wright-Fisher process [55]. This is not surprising given that the shape of joint site frequency spectrum (Additional file 1: Fig. S3), summarized in low \(F_{\text{st}}\) genome-wide, is consistent with near-panmixia. The timing and strength of the recent bottleneck appears similar in mainland Tanzania and Zanzibar isolates and coincides with a decline in the prevalence of parasitemia. However, it should be remembered that the relationship between genetic and census population size—for which prevalence is a proxy—is complex, and other explanations may exist for the observed trends.

Finally, this paper makes the first estimates of the distribution of fitness effects (DFE) in \(P. falciparum\). Although the impact of selection on genetic diversity in this species has long been of interest in the field, previous work has tended to focus on positive selection associated with resistance to disease-control interventions. The DFE is a more fundamental construct that has wide-ranging consequences for the evolutionary trajectory of a population and the genetic architecture of phenotypic variation [56]. Purifying selection is pervasive, but most new alleles (~ 75%) are expected to have sufficiently small selection coefficients that their fate will be governed by drift. The proportion of new mutations expected to be beneficial—the “target size” for adaption—is small, on the order 1–2%. Together these observations imply that even in the presence of ongoing human interventions, patterns of genetic variation in the Tanzanian parasite population are largely the result of drift and purifying selection rather than positive selection. It should be noted that these conclusions are based on the core genome and may not hold for hypervariable loci thought to be under strong selection such as erythrocyte surface antigens. Furthermore, the complex lifecycle of \(P. falciparum\) species also departs in important ways from the assumptions of classical population-genetic models [57]. The qualitative impact of these departures conclusions is hard to determine.

**Conclusion**

The elimination of malaria from Zanzibar has been a goal for many years. This paper presents genomic evidence of continued recent importation of \(P. falciparum\) from mainland Tanzania to the archipelago. Reducing this importation is likely to be an important component of reaching elimination. Investigation of approaches to limit importation, such as screening of travellers or mass drug treatment, is needed. However, the high degree of connectivity between the mainland and the Zanzibar archipelago will make this challenging. It is encouraging that parasite populations in the region appear to be contracting (Fig. 5). These declines are likely due to decreasing transmission but nonetheless need to be interpreted with caution, as they may also be due to other factors that impact effective population size estimates, including violation of model assumptions. The data suggests that larger studies of the relationship between Zanzibarian and mainland parasites will enable further more precise estimates of corridors of importation based on parasite genetics. Genomic epidemiology has the potential to supplement traditional epidemiologic studies in Zanzibar and to aid efforts to achieve malaria elimination on the archipelago.

**Supplementary information**

Supplementary information accompanies this paper at https://doi.org/10.1186/s12936-020-3137-8.

Additional file 1. Additional figures and tables.
Abbreviations
AL: Artemether–lumefantrine; ASAQ: Artesunate–amodiaquine; CMI: Centromer;
COI: Complexity of infection; DBS: Dried blood spots; DFE: Distribution of
fitness effects; FmSS: Median fraction of the genome shared IBD; IBD: Identity
by descent; N* Effective population size; PEA: Principal components analysis;
PFC: PF Community Project; PLAF: Population-level allele frequency; PLMAF:
Population-level minor allele frequency; SFS: Site-frequency spectrum; SLDP:
Single low dose primaquine; SNV: Single nucleotide variant; sWGA: Selective
whole genome amplification; VQSR: Variant Quality Score Recalibration; WGS:
Whole genome sequencing; WSAF: Within-sample allele frequency.

Acknowledgements
We would like to thank the communities and participants who took part in
these studies. We would also like to thank Molly Deutsch-Feldman for helping to
optimize the SWGA protocol.

Authors’ contributions
APM, NFB and JBP designed experiments, conducted analysis and wrote the
manuscript. BN, LEM, MM, and UM collected samples and participated in
manuscript preparation. MD conducted laboratory work and participated in
manuscript preparation. DLF helped develop software and participated in
manuscript preparation. JAB, AM, AB and JJ helped conceive the study, contributed
to the experimental design and wrote the manuscript. All authors read and approved the final manuscript.

Funding
This research was funded by the National Institutes of Health, Grants
R01AI121558, F30AI143172 (NFB), F30MH103925 (APM), and K24AI134990
Funding was also contributed from the Swedish Research Council and Erling-
Persson Family Foundation.

Availability of data and materials
Sequencing reads were deposited into the NCBI SRA (Accession numbers:
090770). This publication uses data generated by the Pf3k project (http://
www.malar iagen .net/proje cts/p-falci parum -commu nity-proje
ct) as described in [36]. Genome sequencing was performed by the Wellcome
Trust Sanger Institute and the Community Projects is coordinated by the
Genomes of cryptic chimpanzee Plasmodium species reveal key evolu-
tionary events leading to human malaria. Nat Commun. 2016;7:11078.

Ethics approval and consent to participate
This analysis was approved by the IRBs at the University of North Carolina
at Chapel Hill, Muhimbili University of Health and Allied Sciences (MUHAS),
Zanzibar Medical Research Ethical Committee and the Regional Ethics Review
Board, Stockholm, Sweden.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
1 Division of Infectious Diseases, Department of Medicine, School of Medi-
cine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA.
2 Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC 27599, USA. 3 Department of Para-
sitology and Medical Entomology, Muhimbili University of Health and Allied
Sciences, Dar es Salaam, Tanzania. 4 Department of Women’s and Children’s
Health, International Maternal and Child Health (IMCH), Uppsala University,
Uppsala, Sweden. 5 Training and Research, Mzazi Mmoja Hospital, Zanzibar,
Tanzania. 6 Department of Microbiology, Tumor and Cell Biology, Karolinska
Institutet, 17177 Stockholm, Sweden. 7 Curriculum in Bioinformatics and Com-
putational Biology, University of North Carolina, Chapel Hill, NC 27599, USA.
8 Department of Laboratory Medicine and Pathology, Brown University,
Providence, RI 02912, USA. 9 Curriculum in Genetics and Molecular Biology,
University of North Carolina, Chapel Hill, NC 27599, USA.

References
1. WHO. World Malaria Report 2018. Geneva: World Health Organization;
2019.
2. Bjorkman A, Shakesy D, Ali AS, Morris U, Mikali H, Abbas AK, et al. From
high to low malaria transmission in Zanzibar-challenges and opportuni-
ties to achieve elimination. BMC Med. 2019;17:14.
3. Bjorkman A, Cook J, Stunock H, Mueller M, Ali A, Xu W, et al. Spatial distri-
bution of falciparum malaria infections in Zanzibar: implications for focal
drug administration strategies targeting asymptomatic parasite carriers.
Clin Infect Dis. 2017;64:1236–43.
4. Bouserna JT, Gouagnc LA, Drakeley Cj, Meuststeged AM, Okech BA, Akim
IN, et al. Plasmodium falciparum gametocyte carriage in asymptomatic
children in western Kenya. Malar J. 2004;3:18.
5. Mavili-Mbumba DF, Nikisema R, Bouyoub-Akotet MK, Bahamontes-Rosa
N, Traoré A, Kombila M. Sub-microscopic gametocyte carriage in febrile
children living in different areas of Gabon. Malar J. 2013;12:375.
6. Okell LC, Bousena T, Griffin JT, Ouédraogo AL, Ghani AC, Drakeley CJ.
Factors determining the occurrence of submicroscopic malaria infections
and their relevance for control. Nat Commun. 2012;3:1237.
7. The mraRA Consultative Group on Diagnoses and Diagnostics. A
research agenda for malaria eradication: diagnoses and diagnostics. PLoS
Med. 2011;8:e1000396.
8. Neafsey DE, Volkman SK. Malaria genomics in the era of eradication. Cold
Spring Harb Perspect Med. 2017;7:a025544.
9. Wesołowski A, Taylor AR, Chang H-H, Ventry R, Tessema S, Bailey JA, et al.
Mapping malaria by combining parasite genomic and epidemiological
data. BMC Med. 2018;16:190.
10. Chang H-H, Park DJ, Galinsky KJ, Schaffner SF, Ndiaye D, Ndir O, et al.
Genomic sequencing of Plasmodium falciparum malaria parasites from
Senegal reveals the demographic history of the population. Mol Biol Evol.
2012;29:3427–39.
11. Taylor AR, Schaffner SF, Cerqueira GC, Nkhoma SC, Anderson TJ, C
Srirawat K, et al. Quantifying connectivity between local Plasmodium falciparum malaria parasite populations using identity by descent. PLoS
Genet. 2017;13:e1007065.
12. Henden L, Lee S, Mueller I, Barry A, Bahlo M. Identity-by-descent analyses
for measuring population dynamics and selection in recombining patho-
gens. PLoS Genet. 2018;14:e1007279.
13. Schaffner SF, Taylor AR, Wong W, Wirth DF, Neafsey DE. hmvmdBD: software
to infer pairwise identity by descent between haploid genotypes. Malar J.
2018;17:196.
14. Zhu SJ, Almagro-Garcia J, McVean G. Deconvolution of multiple infec-
tions in Plasmodium falciparum from high throughput sequencing data.
Bioinformatics. 2018;34:9–15.
15. Zhu SJ, Hendry JA, Almagro-Garcia J, Pearson RD, Amato R, Miles A, et al.
The origins and relatedness structure of mixed infections vary with local prevalence of P. falciparum malaria. bioRxiv. 2019. https://doi.
org/10.1101/387266v4.
16. Auburn S, Benavente ED, Miotto O, Pearson RD, Amato R, Grigg MJ, et al.
Genomic analysis of a pre-elimination Malaysian Plasmodium vivax popu-
lation reveals selective pressures and changing transmission dynamics. Nat
Commun. 2018;9:2585.
17. Daniels RF, Schaffner SF, Wenger EA, Proctor JL, Chang H-H, Wong W, et al.
Modeling malaria genomics reveals transmission decline and rebound in
Senegal. Proc Natl Acad Sci USA. 2015;112:7067–72.
18. Bei AK, Naing M, Deme AB, Daniels RF, Sarr FD, Sokhna C, et al. Dramatic
changes in malaria population genetic complexity in Dielmo and
Ndop, Senegal, revealed using genomic surveillance. J Infect Dis.
2018;217:622–7.
19. Oyola SO, Ariami CV, Hamilton WL, Keke M, Amenga-Etego LN, Ghansah
A, et al. Whole genome sequencing of Plasmodium falciparum from
dried blood spots using selective whole genome amplification. Malar J.
2016;15:597.
20. Sundararaman SA, Plenderleith LJ, Liu W, Loy DE, Leman GH, Lj Y, et al.
Genomes of cryptic chimpanzee Plasmodium species reveal key evolu-
tional events leading to human malaria. Nat Commun. 2016;7:11078.
21. Clarke EL, Sundararaman SA, Seifert SN, Bushman FD, Hahn BH, Brisson D. svga: a primer design toolkit for selective whole genome amplification. Bioinformatics. 2013;30:2071–7.

22. Miles A, Iqbal Z, Vauterin P, Pearson R, Campino S, Theron M, et al. Indels, structural variation, and recombination drive genomic diversity in Plasmodium falciparum. Genome Res. 2016;26:1288–99.

23. Manske M, Miotto O, Campino S, Auburn S, Almagro-Garcia J, Maslen G, et al. Analysis of Plasmodium falciparum diversity in natural infections by deep sequencing. Nature. 2012;487:375–9.

24. Otto TD, Rayner JC, Böhme U, Pain A, Sportsswoude N, Sanders M, et al. Genome sequencing of chimpanzee malaria parasites reveals possible pathways of adaptation to human hosts. Nat Commun. 2014;5:4754.

25. Arthur R, Schulz-Trieglaff O, Cox AJ, O’Connell J. AKT: ancestry and kinship toolkit. Bioinformatics. 2017;33:142–4.

26. Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: analysis of next generation sequencing data. BMC Bioinformatics. 2014;15:356.

27. Browning BL, Browning SR. Improving the malaria genomics in the era of admixture in human history. Genetics. 2012;192:1065–93.

28. Verity R, Aydemir O, Brazeau NF, Watson OJ, Hathaway NJ, Mwanda-Antwi D, et al. High-level Plasmodium falciparum sulfadoxine-pyrimethamine resistance with the concomitant occurrence of septuple haplotype in Tanzania. Malar J. 2015;14:439.

29. Browning SR, Browning BL. Accurate non-parametric estimation of recent effective population size from segments of identity by descent. Am J Hum Genet. 2015;97:404–18.

30. Bhart S, Weiss DJ, Cameron E, Bisanjo D, Mappin B, Dalrymple U, et al. The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. Nature. 2015;526:207–11.

31. Pfeffer DA, Lucas TCD, May D, Harns J, Rozier J, Twohig KA, et al. malarialAtlas: an R interface to global malariometric data hosted by the Malaria Atlas Project. Malar J. 2018;17:352.

32. Terhorst J, Kamm JA, Song YS. Robust and scalable inference of population history from hundreds of unphased whole genomes. Nat Genet. 2017;49:303–9.

33. Tataru P, Mollion M, Glémin S, Bataillon T. Inference of distribution of parasites following meiotic recombination and cotransmission. PLoS Comput Biol. 2018;14:e1005923.

34. Maclean CA, Chue Hong NP, Prendergast JGD. Hapbin: an efficient pathway of adaptation to human hosts. Nat Commun. 2014;5:4754.

35. Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, et al. Natural selection and convergence using population-genomics data. Proc Natl Acad Sci USA. 2013;110:20129–34.

36. Eyre-Walker A, Keightley PD. The distribution of fitness effects of new mutations. Mol Biol Evol. 2015;32:3027–9.

37. Saberi PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, et al. Genome-wide detection and characterization of positive selection in chimpanzee malaria parasites in Thiès, Senegal. Genome Med. 2017;9:5.

38. Hamilton WL, Claessens A, Otto TD, Keke M, Fairhurst RM, Rayner JC, et al. Extreme mutation bias and high AT content in Plasmodium falciparum. Nucleic Acids Res. 2017;45:1889–901.

39. Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, et al. Ancient admixture in human history. Genetics. 2012;192:1065–93.

40. Wong W, Wenger EA, Hartl DL, Wirth DF. Modeling the genetic relatedness of Plasmodium falciparum parasites following meiotic recombination and cotransmission. PLoS Comput Biol. 2018;14:e1005923.

41. Speed D, Balding DJ. Relatedness in the post-genomic era: is it still useful? Nat Rev Genet. 2015;16:33–44.

42. Huber JH, Johnston GL, Greenhouse B, Smith DL, Perkins TA. Quantitative, model-based estimates of variability in the generation and serial intervals of Plasmodium falciparum malaria. Malar J. 2016;15:490.

43. Samad H, Coll F, Preston MD, Ocholla H, Fairhurst RM, Clark TG. Imputation-based population genetics analysis of Plasmodium falciparum malaria parasites. PLoS Genet. 2015;11:e1005131.

44. Kavishe RA, Kaaya RD, Nqagga C, Ofare SG, Kavishe AA, et al. Molecular monitoring of Plasmodium falciparum super-resistance to sulfadoxine–pyrimethamine in Tanzania. Malar J. 2016;15:335.

45. Ngondi JM, Ishengoma DS, Doctor SM, Thwai KL, Keeler C, Mkude S, et al. Surveillance for sulfadoxine–pyrimethamine resistant malaria parasites in the Lake and Southern Zones, Tanzania, using pooling and next-generation sequencing. Malar J. 2017;16:236.

46. Baraka V, Ishengoma DS, Frantis F, Minja DTR, Madebe RA, Ngatunga D, et al. High-level Plasmodium falciparum sulfadoxine-pyrimethamine resistance with the concomitant occurrence of septuple haplotype in Tanzania. Malar J. 2015;14:439.

47. Tatem AJ, Qiuyu Y, Smith DL, Sabot O, Ali AS, Moonen B. The use of mobile phone data for the estimation of the travel patterns and imported Plasmodium falciparum rates among Zanzibar residents. Malar J. 2009;8:287.

48. O’Brien JD, Amenga-Etego L, Li R. Approaches to estimating inbreeding coefficients in clinical isolates of Plasmodium falciparum from genomic sequence data. Malar J. 2016;15:473.

49. Tatem AJ, Qiuyu Y, Smith DL, Sabot O, Ali AS, Moonen B. Travel patterns and imported Plasmodium falciparum rates among Zanzibar residents. In: Travel, health and malaria: the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. 2009;5:e1000695.

50. Tatem AJ, Qiuyu Y, Smith DL, Sabot O, Ali AS, Moonen B. Travel and malaria importation and malaria transmission in Zanzibar. Sci Rep. 2011;1:93.

51. Le Menach A, Tatem AJ, Cohen JM, Hay SI, Randell H, Patil AP, et al. Travel risk, malaria importation and malaria transmission in Zanzibar. Sci Rep. 2012;3:267–70.

52. O’Brien JD, Amenga-Etego L, Li R. Approaches to estimating inbreeding coefficients in clinical isolates of Plasmodium falciparum from genomic sequence data. Malar J. 2016;15:473.

53. Nkhoma Standwell C, Nair S, Cheeseman IH, Rohr-Allegrini C, Singlam S, Nosten F, et al. Close kinship within multiple-genotype malaria parasite infections. Proc Biol Sci. 2012;279:2589–98.

54. Wong W, Griggs AD, Daniels RF, Schaffner SF, Ndiaye D, Bei AK, et al. Genetic relatedness analysis reveals the cotransmission of genetically related Plasmodium falciparum parasites in Thïès, Senegal. Genome Med. 2017;9:5.

55. Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. 2009;5:e1000695.

56. Eyre-Walker A, Keightley PD. The distribution of fitness effects of new mutations. Nat Rev Genet. 2007;8:10–8.

57. Chang H-H, Moss EL, Park DJ, Ndiaye D, Mboup S, Volkman SK, et al. Malaria life cycle intensifies both natural selection and random genetic drift. Proc Natl Acad Sci USA. 2013;110:20129–34.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.