



Swansea University
Prifysgol Abertawe



Cronfa - Swansea University Open Access Repository

This is an author produced version of a paper published in:
Wellcome Open Research

Cronfa URL for this paper:
<http://cronfa.swan.ac.uk/Record/cronfa49099>

Paper:

Omedo, I., Mogeni, P., Bousema, T., Rockett, K., Amambua-Ngwa, A., Oyier, I., C. Stevenson, J., Y. Baidjoe, A., de Villiers, E., et. al. (2017). Micro-epidemiological structuring of Plasmodium falciparum parasite populations in regions with varying transmission intensities in Africa. *Wellcome Open Research*, 2, 10
<http://dx.doi.org/10.12688/wellcomeopenres.10784.2>

This is an open access article distributed under the terms of the Creative Commons Attribution Licence.

This item is brought to you by Swansea University. Any person downloading material is agreeing to abide by the terms of the repository licence. Copies of full text items may be used or reproduced in any format or medium, without prior permission for personal research or study, educational or non-commercial purposes only. The copyright for any work remains with the original author unless otherwise specified. The full-text must not be sold in any format or medium without the formal permission of the copyright holder.

Permission for multiple reproductions should be obtained from the original author.

Authors are personally responsible for adhering to copyright and publisher restrictions when uploading content to the repository.

<http://www.swansea.ac.uk/library/researchsupport/ris-support/>



RESEARCH ARTICLE

REVISED **Micro-epidemiological structuring of *Plasmodium falciparum* parasite populations in regions with varying transmission intensities in Africa [version 2; referees: 4 approved]**

Irene Omedo ¹, Polycarp Mogeni ¹, Teun Bousema^{2,3}, Kirk Rockett ⁴, Alfred Amambua-Ngwa⁵, Isabella Oyier¹, Jennifer C. Stevenson^{3,6}, Amrish Y. Baidjoe², Etienne P. de Villiers ^{1,7,8}, Greg Fegan ¹, Amanda Ross ⁹, Christina Hubbart ⁴, Anne Jeffreys⁴, Thomas N. Williams^{1,10}, Dominic Kwiatkowski^{4,11}, Philip Bejon ^{1,12}

¹KEMRI-Wellcome Trust Research Programme, Centre for Geographic Medicine Research-Coast, Kilifi, Kenya

²Department of Medical Microbiology, Radboud University Nijmegen Medical Centre, Radboud Institute for Molecular Life Sciences, Nijmegen, The Netherlands

³London School of Hygiene and Tropical Medicine, London, UK

⁴Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK

⁵Medical Research Council Unit, Fajara, The Gambia

⁶Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA

⁷Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine Research Building, University of Oxford, Oxford, UK

⁸Department of Public Health, Pwani University, Kilifi, Kenya

⁹Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, Basel, Switzerland

¹⁰Department of Medicine, South Kensington Campus, Imperial College London, London, UK

¹¹Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK

¹²Centre for Clinical Vaccinology and Tropical Medicine, University of Oxford, Oxford, UK

v2 First published: 14 Feb 2017, 2:10 (<https://doi.org/10.12688/wellcomeopenres.10784.1>)

Latest published: 08 Sep 2017, 2:10 (<https://doi.org/10.12688/wellcomeopenres.10784.2>)

Abstract

Background: The first models of malaria transmission assumed a completely mixed and homogeneous population of parasites. Recent models include spatial heterogeneity and variably mixed populations. However, there are few empiric estimates of parasite mixing with which to parameterize such models.

Methods: Here we genotype 276 single nucleotide polymorphisms (SNPs) in 5199 *P. falciparum* isolates from two Kenyan sites (Kilifi county and Rachuonyo South district) and one Gambian site (Kombo coastal districts) to determine the spatio-temporal extent of parasite mixing, and use Principal Component Analysis (PCA) and linear regression to examine the relationship between genetic relatedness and distance in space and time for parasite pairs.

Results: Using 107, 177 and 82 SNPs that were successfully genotyped in 133, 1602, and 1034 parasite isolates from The Gambia, Kilifi and Rachuonyo South district, respectively, we show that there are no discrete geographically

Open Peer Review

Referee Status: 

	Invited Referees			
	1	2	3	4
version 2 published 08 Sep 2017			 report	
version 1 published 14 Feb 2017	 report	 report	 report	 report

restricted parasite sub-populations, but instead we see a diffuse spatio-temporal structure to parasite genotypes. Genetic relatedness of sample pairs is predicted by relatedness in space and time.

Conclusions: Our findings suggest that targeted malaria control will benefit the surrounding community, but unfortunately also that emerging drug resistance will spread rapidly through the population.

Keywords

Plasmodium falciparum, malaria, parasite mixing, population structure, micro-epidemiological, targeted control, principal component analysis, genotyping



This article is included in the [KEMRI | Wellcome Trust gateway](#).

1 **Michel Tibayrenc**, Institut de Recherche pour le Développement (IRD), France

2 **Cristian Koepfli** , University of California, Irvine, USA

3 **Christopher Delgado-Ratto** , University of Antwerp, Belgium

4 **Liwang Cui** , Pennsylvania State University, USA

Discuss this article

[Comments \(0\)](#)

Corresponding author: Irene Omedo (iomedo@kemri-wellcome.org)

Author roles: **Omedo I:** Formal Analysis, Software, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Mogeni P:** Formal Analysis, Software, Writing – Review & Editing; **Bousema T:** Conceptualization, Data Curation, Funding Acquisition, Investigation, Methodology, Resources, Writing – Review & Editing; **Rockett K:** Data Curation, Investigation, Writing – Review & Editing; **Amambua-Ngwa A:** Investigation, Resources, Writing – Review & Editing; **Oyier I:** Formal Analysis, Writing – Review & Editing; **Stevenson JC:** Investigation, Writing – Review & Editing; **Baidjoe AY:** Investigation, Writing – Review & Editing; **de Villiers EP:** Formal Analysis, Software, Writing – Review & Editing; **Fegan G:** Formal Analysis, Software, Writing – Review & Editing; **Ross A:** Formal Analysis, Funding Acquisition, Methodology, Software, Writing – Review & Editing; **Hubbart C:** Data Curation, Investigation, Writing – Review & Editing; **Jeffreys A:** Data Curation, Investigation, Writing – Review & Editing; **Williams TN:** Funding Acquisition, Methodology, Resources, Writing – Review & Editing; **Kwiatkowski D** : Conceptualization, Funding Acquisition, Methodology, Resources, Supervision, Writing – Review & Editing; **Bejon P:** Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Methodology, Resources, Software, Supervision, Validation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: Sample collection at the Rachuonyo South site was supported by the Bill and Melinda Gates Foundation, under the Malaria Transmission Consortium, Grant No.45114 and the Grand Challenge Grant No. OPP1024438. Thomas N. Williams is funded by the Wellcome Trust, grant number 091758. Philip Bejon, Polycarp Mogeni and Irene Omedo are funded by the UK Medical Research Council (MRC) and the UK Department for International Development (DFID) under the MRC/DFID Concordat agreement. Polycarp Mogeni is funded by the Gottfried und Julia Bangarter-Rhyner Stiftung and the Novartis Foundation for Medical Biological Research project 13A13. Sample collection in Kilifi was supported by core funding from the Wellcome Trust to the Kenya Programme.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2017 Omedo I *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

How to cite this article: Omedo I, Mogeni P, Bousema T *et al.* **Micro-epidemiological structuring of *Plasmodium falciparum* parasite populations in regions with varying transmission intensities in Africa [version 2; referees: 4 approved]** Wellcome Open Research 2017, 2 :10 (<https://doi.org/10.12688/wellcomeopenres.10784.2>)

First published: 14 Feb 2017, 2:10 (<https://doi.org/10.12688/wellcomeopenres.10784.1>)

REVISED Amendments from Version 1

In response to the reviewer's comments, we have included information on the changes in malaria transmission intensity in each study site during the study period. We have also modified Figure 5 to show the 95% confidence intervals around parasites collected one day apart, and added more information on how this figure was generated. We have additionally provided information on the analysis of SNP subsets, specifically those SNPs typed in EBA175 and AMA1. Figures showing the results of these SNP subset analyses have been included in the [Supplementary materials](#). In the methods section, we have added detailed information on how spatial scan statistics were carried out to identify clusters of parasites with similar genotypes.

Dataset 1 was updated to contain the same parasite sample IDs as those used when reporting parasite pairwise analyses in Dataset 3, Dataset 4 and Dataset 5. The manuscript now also includes data showing the range of malaria positive fraction and parasite prevalence within geographically defined pixels/grids of varying sizes. Finally, we have added more information on the spatial scales over which previous studies have identified *P. falciparum* population structure, and further studies involving the use of gene flow models to show parasite movement between study sites have been proposed.

[See referee reports](#)

Introduction

The earliest models of malaria transmission assumed a completely mixed and homogenous parasite population^{1,2}. However, malaria transmission is highly heterogeneous, and follows the Pareto principle where 80% of infections occur in only about 20% of the population³. Consequently, there is increasing interest in models allowing for spatial heterogeneity and variably mixed populations of parasites⁴⁻⁷. There are now several epidemiological studies describing spatial heterogeneity of malaria on varying geographical scales⁸⁻¹⁹. This heterogeneity is characterized by infection hotspots which usually persist even after transmission has been reduced in surrounding areas^{9,11,20-25}, and thus act as reservoirs of infection^{21,26}. Achieving any meaningful reduction in transmission in regions containing malaria hotspots will require a scale up of control activities, including repeated mass administration of Artemisinin Combination Therapy (ACT) drugs, increased coverage of long lasting insecticide treated nets (LLINs) and intensive indoor residual spraying (IRS). These measures are very costly and may not be realistic for universal coverage in most of the resource-poor endemic countries. Thus, targeted control may be more important, and is likely to be required to eliminate malaria^{3,21,27,28}.

Mathematical models show that targeting hotspots may reduce transmission in surrounding areas^{11,22}. These models, however, assume that hotspots are stable and that mosquito mixing in the community is homogeneous²². Studies have shown that certain species of mosquitoes exhibit some level of site fidelity, where they return to the same homesteads to feed²⁹. If such behaviour is the norm with very little mixing, then this would greatly reduce the community-wide impact of targeted interventions, and interventions would be beneficial only to individuals within the targeted region. If, however, transmission networks operate freely over large geographical areas, then these interventions would likely have an impact beyond the targeted region. Furthermore, parasite evolution

takes place in a micro-epidemiological context and the spread of drug resistance or new antigenic variants through the population will also be critically dependent on the degree of mixing of parasite populations.

Few studies currently provide empiric evidence on the mixing of parasites over space and time, yet this evidence is important as parasite mixing is likely to affect the outcome of targeted control interventions²³. The community-wide impact of targeted control has not been studied extensively, although early controlled trials showed that bed nets were effective at reducing child morbidity and mortality associated with malaria, in villages or communities randomised to the intervention in The Gambia³⁰ and Kilifi³¹. More recent studies have shown that the use of bed nets in a village randomized to intervention in Asembo, western Kenya, also protected individuals just outside the intervention village who were themselves not using bed nets³². A cluster-randomized controlled trial on the impact of targeting integrated control measures to hotspots showed temporally limited effect on reducing transmission in areas surrounding the targeted hotspots²³. In order to inform future targeted control strategies more precise empiric data on parasite mixing is required.

We hypothesized that by genotyping parasites with fine-scale temporal and spatial data we would be able to determine fine-scale structure to the population and infer the degree of parasite mixing over small geographical areas which are likely to be the focus of targeted malaria control programs^{23,27}. We used SNP genotyping of *Plasmodium falciparum* field isolates from three African sites and analysed the genetic relatedness among parasites within individual sites, in order to determine the level of parasite mixing on micro-epidemiological scales in each population. Principal Component Analysis (PCA) was used to detect parasite subpopulations in each site, and tests of spatial autocorrelation including Moran's *I* and spatial scan statistics were used to test for autocorrelation among parasite genotypes. The analyses were carried out at different spatial scales ranging from intensive within-village surveillance through to county-wide surveillance.

Materials and methods

Study sites

P. falciparum infected blood samples were collected from individuals at three sites in two African countries: Kombo coastal districts of The Gambia on the West African coast; Kilifi, Kenya on the East African coast, and Rachuonyo South District in the Western Kenyan highlands. The Gambia has a subtropical climate with a single rainy season between the months of June and October^{33,34}, while Kenya has two rainy seasons, experiencing short rains between October and December and long rains between April and August³⁵. In all three sites, *P. falciparum* is the main causative agent of malaria^{22,33,35} and transmission occurs almost exclusively during and immediately after the rainy seasons^{34,36}. The common vectors in The Gambia are *Anopheles gambiae* s.s., *Anopheles arabiensis* and *Anopheles melas*³⁷, while the common vectors in the Kenyan coast have historically been *A. gambiae* s.s. and *A. funestus*, but a recent shift to *A. arabiensis* and *A. merus* has been detected along the coast³⁸. In Rachuonyo South district, the main vectors transmitting malaria are *A. gambiae* s.l. and *A. funestus*³⁹. Temporal trends show

declining malaria transmission in The Gambia and Coastal Kenya^{17,33,34,40}, although not in Western Kenya⁴¹. Asymptomatic parasite prevalence is lowest in The Gambia at 8.7%⁴², intermediate in Kilifi at 14%⁴³ and slightly higher in Rachuonyo South at 16%⁴⁴. Over the study period, malaria transmission as measured by malaria slide positivity rate fell from 56% in 1998 to 7% in 2009 in Kilifi⁴⁵, and rose slightly in Fajara and Brikama in the Gambia³³.

Ethics statement

Ethical approval for this study was obtained from Kenya Medical Research Institute (KEMRI) Ethical Review Committee (under SSC No. 2239). Written informed consent was obtained from parents/guardians of the study participants. The study methods were carried out in accordance with the approved guidelines.

Sample collection, DNA extraction and Genotyping

5199 *P. falciparum* infected blood samples were collected during hospital admissions and community surveys over a 14-year period from 1998 to 2011. The Gambian samples were collected at Fajara and Brikama health facilities from children aged 8 months to 16 years who were living in the Kombo coastal districts and who were part of a clinical malaria study in 2007–2008³³. The Kilifi samples came from children aged 1 to 6 years who had been recruited into a phase 2b randomized trial looking at the efficacy of the Candidate Malaria Vaccines FP9 ME-TRAP (multiple epitope–thrombospondin-related adhesion protein) and MVA ME-TRAP in 2005⁴⁶, as well as clinical malaria studies looking at antibody responses to Merozoite Surface Protein 2 (MSP2) among individuals 3 weeks to 85 years old⁴⁷; the effect of declining transmission on mortality and morbidity in children up to 14 years old⁴⁰ and definitions of clinical malaria endpoints⁴⁸. The Rachuonyo south samples were collected during a community survey conducted in 2011 as part of a trial looking at the impact of hotspot targeted control interventions on reducing malaria transmission in the wider community²². Prior to genotyping, DNA was extracted from these samples using either ABI prism 6100 Nucleic Acid preparation (Applied Biosystems, Waltham, Massachusetts, USA) or Chelex Extraction.

276 SNPs in 177 genes were typed in the three parasite populations (Dataset 1⁴⁹). The SNPs were selected from a panel of 384 SNPs previously designed for a study on population structure of *P. falciparum* parasites from Africa, Southeast Asia and Oceania⁵⁰ and were chosen based on three criteria:

- polymorphic among three of the most studied and well characterized *P. falciparum* strains (3D7, HB3 and IT).
- uniformly distributed across the parasite genome.
- ease of typing on the sequenom platform.

Genes typed included antigen-encoding, housekeeping and hypothetical genes. 52 and 9 SNPs were typed in the antigen-encoding parasite ligands Erythrocyte Binding Antigen 175 (EBA-175) and Apical Membrane Antigen 1 (AMA-1), respectively. In the Kilifi parasite population, between 158 and 226 SNPs were typed in each sample, while in The Gambia and Rachuonyo south

populations, 131 and 111 SNPs were typed in 143 and 2744 samples, respectively. Genotyping was done on the Sequenom MassARRAY iPLEX platform, which allows multiplexing of up to 40 SNPs in a single reaction well and differentiates alleles based on variations in their mass⁵¹. Locus specific PCR and iPLEX extension primers were designed with the sequenom MassARRAY designer software (Version 3.1) using 3D7 as the reference genome (PlasmoDB release 9.0) (Dataset 2⁵²). A multiplexed PCR reaction was performed by pooling locus-specific primers, and unincorporated dNTPs were dephosphorylated enzymatically using shrimp alkaline phosphatase. Extension primers binding immediately adjacent to the SNP site of interest were then extended by a single nucleotide base, using mass-modified dideoxynucleotides. The extended products were resin cleaned to remove excess salts and the mass of the different alleles determined using MALDI-TOF mass spectrometry.

Sample and SNP cut-off selection criteria

Genotype data was aggregated to determine genotyping success rates for individual samples and SNPs. Samples where >40% of SNP typing failed were excluded from analysis, and among the remaining samples, SNP typing for which >30% of samples failed were further excluded from analysis. The criteria for successful SNP typing were based on the SNP intensity values (r) and allelic intensity ratios (θ). Alleles were called as successful if they were above an intensity cut-off value ranging between 0.5 and 1.0, set depending on the performance of the individual SNP assay, and were classified as failed if they were below this cut-off. For those SNPs that were above the cut-off, allelic intensity ratios ranging between 0 and 1 were used to classify them as homozygous (single parasite genotype infections) or heterozygous (mixed parasite genotype infections). θ values nearing 0 and 1 indicate different homozygous alleles, while intermediate values indicate heterozygous SNPs, representing mixed parasite populations. Where mixed parasite populations were identified, we took the majority SNP calls at each position to indicate the dominant genotype.

Statistical analyses

All statistical analyses of genotype data were conducted in R statistical software (version 3.0.2)⁵³ except for the spatial scan statistics which were computed using SaTScan software (version 9.3)⁵⁴. Analyses were carried out separately for each parasite population, except for the Fixation index (FST) analyses which by definition involve the comparison of populations and so were carried out between samples in the different sites.

In each population, genotype data for all samples was aggregated and analysed collectively. Separate analyses were also carried out for subsets of SNPs typed in EBA 175 (39, 36 and 20 SNPs in The Gambia, Kilifi and Rachuonyo South, respectively) and AMA1 (9 SNPs in The Gambia and 8 SNPs in Kilifi). Only 3 SNPs were genotyped in Rachuonyo South, so this SNP subset was not analysed separately. In the Kilifi population, we ran additional analyses for samples collected from community surveys (asymptomatic infections) and hospital admissions (symptomatic infections).

Calculating pairwise time, distance and SNP differences. Analyses were carried out separately for each of the three sites. Each

parasite was compared to every other parasite in that site (i.e. a pairwise analysis), noting the time, distance and SNP differences between the parasite pair (Dataset 3 – Dataset 5⁵⁵⁻⁵⁷). We took half the lower limit of detection of temporal and spatial differences for parasite pairs collected on the same day and/or at the same location. Parasite pairs collected on the same day were assigned a difference of 0.5 days. For older samples in Kilifi (i.e. collected prior to 2004) where location was known to a 5 km accuracy, pairs collected at the same location were assigned a difference of 2.5km. We had precise geospatial co-ordinates for recent samples in Kilifi (i.e. collected after 2004) as well as all samples from The Gambia and Rachuonyo South, so parasite pairs in these three groups collected from the same location were assigned a difference of 0.02km.

SNP differences were computed by comparing genotype data for parasite pairs within each population and counting the number of SNPs between them. Missing SNP data for each parasite was replaced with the major allele in the respective population, after excluding SNP typing where >30% of assays failed as described above.

Population genetics analyses. Minor allele frequencies were computed for SNPs in each population. Principal components analysis (PCA) was performed using singular value decomposition on a covariance matrix of pairwise SNP differences between parasites in individual populations. To detect inter-population genetic differentiation and within-population genetic diversity, we restricted analysis to 33 SNPs that had been successfully typed in all three populations.

Spatial autocorrelation. Moran's *I* was calculated using geographical coordinates to specify location and scores for the first 3 principal components to specify associated attribute values. Moran's *I* was computed at distance classes of 1 km, 2 km and 5 km, using 100 bootstrap resampling steps to determine statistical significance.

Spatial scan statistics were calculated using SaTScan software and were run separately for each study site. The statistics involved running a purely spatial, retrospective analysis based on a normal probability distribution model using continuous variables (PC scores) and looking for areas with clusters of high PC scores. Latitude and longitude coordinates were used to represent the geographical locations of specific parasites, whereas principal component scores were used to represent individual parasite genotypes. During the analysis, a scanning window that gradually varies in size from including only a single homestead up to 50% of the population moves over the geographical space and at each window size and location, the ratio of parasites with high PCs inside the window versus outside the window is calculated. The window with the highest ratio is noted down as a cluster and its statistical significance is determined after accounting for multiple comparisons using random permutations.

Raster analysis. To identify possible spatial barriers to parasite movement and mixing over short distances, each study area was divided into pixels of varying sizes which were then scored with 1 or 0, based on whether or not a straight line linking any two

parasites crossed their boundaries. These pixels were then used as independent variables in a multivariable linear regression analysis that had the number of SNP differences as the dependent variable. Significance of the coefficient estimates were determined using non-parametric bootstrapping with 100 resampling steps.

To test for correlations between transmission intensity and population genetics at fine scale, each pixel was assigned the mean of the PC scores and either Malaria Positive Fraction (for Kilifi data) or asymptomatic parasite prevalence by PCR (for Rachuonyo) for all samples found within that pixel. The correlation between PC score and MPF or between PC score and parasite prevalence was tested by Spearman's rank ordered correlation coefficient.

Dataset 1. Information on the 276 SNPs genotyped in 177 genes in *P. falciparum* parasite populations from The Gambia, Kilifi and Rachuonyo South

<https://dx.doi.org/10.6084/m9.figshare.5383969>

The columns contain the following information: study_location, site of sample collection; sample_id, unique sample identifier; gene_symbol, gene name (if available); chr_valid, chromosome; coord_valid= base position of SNP on chromosome; sequence_code, SNP name; assay_code, name of assay; rsnumber, unique SNP identifier in dbSNP; reference_allele, 3D7 reference allele, alternative_allele, alternative allele; single_letter_code, IUPAC code for SNPs; result, genotype call after processing; allele1, IUPAC code for allele 1; allele2, IUPAC code for allele 2; allele_ratio1, proportion of allele 1; allele_ratio2, proportion of allele 2; pass_fail, coding of SNP based on availability of valid genotype (pass) or lack of a valid genotype (fail). Geospatial data for homestead location is considered sensitive data and therefore cannot be made open access. However, it can be accessed through a request to our data governance committee, using the email address mmunene@uat/newsite.

Dataset 2. Sequenom assay design information

<http://dx.doi.org/10.6084/m9.figshare.4640719>

Data includes the locus and IPLEX specific primers used in the sequenom reaction to amplify and type the SNPs of interest. Gene product, gene product name; Gene_symbol, gene name; Chromosome, chromosome location of gene; SNP position on chromosome, SNP site; reference_allele, 3D7 reference allele; alternative_allele, alternative allele; sequence, 3D7 reference sequence spanning the SNP site; first_pcrp, first PCR primer sequence; second_pcrp, second PCR primer sequence; extension_primer, IPLEX extension primer sequence; extension1_call, IPLEX primer with extended SNP; extension1_mass, Mass of the extended IPLEX primer; extension1_sequence, sequence of extended IPLEX primer; extension2_call= IPLEX primer with alternative extended allele; extension2_mass, Mass of the extended IPLEX primer with alternative allele; extension2_sequence, sequence of extended IPLEX primer with alternative allele.

Dataset 3. SNP, distance and time differences between *P. falciparum* parasite pairs in The Gambia population

<http://dx.doi.org/10.6084/m9.figshare.4640722>

Differences were computed for all parasite pairwise comparisons. Sample_id and sample_id_x are unique sample identifiers; snps represent the number of snp differences between parasite pairs; km_distance represents geographical distance, in kilometres, between parasite pairs; time_diff represents the temporal distance, in days, between parasite pairs.

Dataset 4. SNP, distance and time differences between *P. falciparum* parasite pairs in the Kilifi population

<http://dx.doi.org/10.6084/m9.figshare.4640725>

Differences were computed for all parasite pairwise comparisons. Sample_id and sample_id_x are unique sample identifiers; snps represent the number of snp differences between parasite pairs; km_distance represents geographical distance, in kilometres, between parasite pairs; time_diff represents the temporal distance, in days, between parasite pairs.

Dataset 5. SNP and distance differences between *P. falciparum* parasite pairs in the Rachuonyo South population

<http://dx.doi.org/10.6084/m9.figshare.4640728>

Differences were computed for all parasite pairwise comparisons. Sample_id and sample_id_x are unique sample identifiers; snps represent the number of snp differences between parasite pairs; km_distance represents geographical distance, in kilometres, between parasite pairs.

Results**Study populations**

5199 *P. falciparum* parasite isolates were collected from the Kombo coastal districts in The Gambia, and Kilifi County and Rachuonyo South district in Kenya (Figure 1) between 1998 and 2011. 107, 177 and 82 SNPs were successfully genotyped in 133, 1602, and 1034 parasite isolates from The Gambia, Kilifi and Rachuonyo South district, respectively (Table 1). 26, 57 and 49 SNPs were present at frequencies of 5% and above in The Gambia, Kilifi and Rachuonyo, respectively. In each of the populations, there was a positive correlation between SNP assay performance and parasite density.

In all study sites, separate analyses of EBA175 and AMA1 did not reveal qualitatively different results from the pooled analyses (Supplementary figure 1 – Supplementary figure 3) and only the results of the pooled analyses are presented here. In the Kilifi population, results were similar between the community surveys and hospital admissions. Here we present the results of the combined analyses of these data subsets.

Parasite genetic diversity and population differentiation

Weir and Cockerham's fixation index (F_{ST}) estimates showed that the level of differentiation amongst the three populations was 0.046, comparable with results of other studies of African populations^{58,59}. Pairwise population analysis gave F_{ST} values of 0.041 between Kilifi and Rachuonyo South, 0.078 between The Gambia and Kilifi and 0.108 between The Gambia and Rachuonyo South, showing the greatest genetic differentiation between The Gambia and Rachuonyo South parasite populations.

Analysis of within-population genetic diversity (π), based on a set of 33 SNPs that had been typed in samples from all three populations, showed that parasites in Rachuonyo South had the highest genetic diversity with an average of 3.384 (95% CI: 3.380 – 3.388) SNP differences per parasite pair. Those in The Gambia had the lowest SNP differences per parasite pair at an average of 2.867 (95% CI: 2.836 – 2.898) SNPs, while Kilifi had intermediate genetic diversity at 3.229 (95% CI: 3.226 – 3.231) SNP differences per parasite pair.

Principal Component Analysis (PCA) was carried out separately for each population using the 107, 177 and 82 SNPs that were successfully typed in The Gambian, Kilifi and Rachuonyo South parasite populations. Cumulatively, the first three principal components accounted for 36.1% (PC1=18.4%, PC2=10.4%, PC3=7.3%) of the variability seen in The Gambia, 13.2% (PC1=5.1%, PC2=4.4%, PC3=3.7%) of the variability seen in Kilifi and 12.7% (PC1=4.4%, PC2=4.3%, PC3=4%) of the variability seen in Rachuonyo South. We were unable to resolve parasite populations into distinct sub-populations using principal component analysis (Figure 2 and Figure 3, Supplementary Figure 4).

Global and local spatial autocorrelation analysis

Having not seen sub-populations by PCA alone, we then included spatial analyses to test for spatial structure to the principal component values. Moran's *I* analysis for spatial autocorrelation showed slight positive correlations for parasites that were statistically significant for at least one principal component at 2 km and below in The Gambia, 5 km and below in Kilifi, and 1 km and below in Rachuonyo South (Figure 4).

Spatial scan statistics using SaTScan identified statistically significant ($p \leq 0.01$) clusters of distinct parasite sub-populations of different sizes in Kilifi and Rachuonyo South. In Kilifi, one cluster with a radius of 1.54 km ($p=0.01$) containing 15 parasites was detected, while in Rachuonyo South, a smaller cluster of genetically distinct parasites was detected with a radius of 0.5 km ($p=0.001$) containing 14 parasites. No clusters were detected in The Gambian population, indicating that parasites did not group into distinct sub-populations in this study site.

Spatio-temporal variations in genetic differences between parasite isolates

We examined the effect of distance and time separating parasite pairs on genetic relatedness to determine the spatial extent and rate of parasite mixing. We used linear regression models where the number of SNP differences between parasite pairs was an outcome predicted by the distance between parasite pairs and the time between parasite pairs. Time was not included for the Rachuonyo South population as the samples were collected in a single cross-sectional survey taken over a few days. Across all three datasets, distance was independently associated with increasing variation in genotype, i.e. the further apart in space any two parasites were, the greater the number of SNP differences between them. In The Gambia and Kilifi populations, time was also shown to be associated with increasing variation in genotype, with parasite pairs collected further apart in time having greater number of genetic differences. Additionally, in The Gambia and Kilifi populations, time interacted antagonistically with distance to attenuate the effect of distance on genotype relatedness (Figure 5). This means that the genetic differences between any two parasites increased with distance, but at a decreasing rate when time between these samples increased. We observed that in The Gambian population, parasites acquired SNP differences over distance at a slower rate than in the Kilifi and Rachuonyo populations.

Bootstrapping the analyses (to take into account the linked nature of pairwise observations) gave statistically significant effects of distance, time and the interaction between distance and time (Table 2).

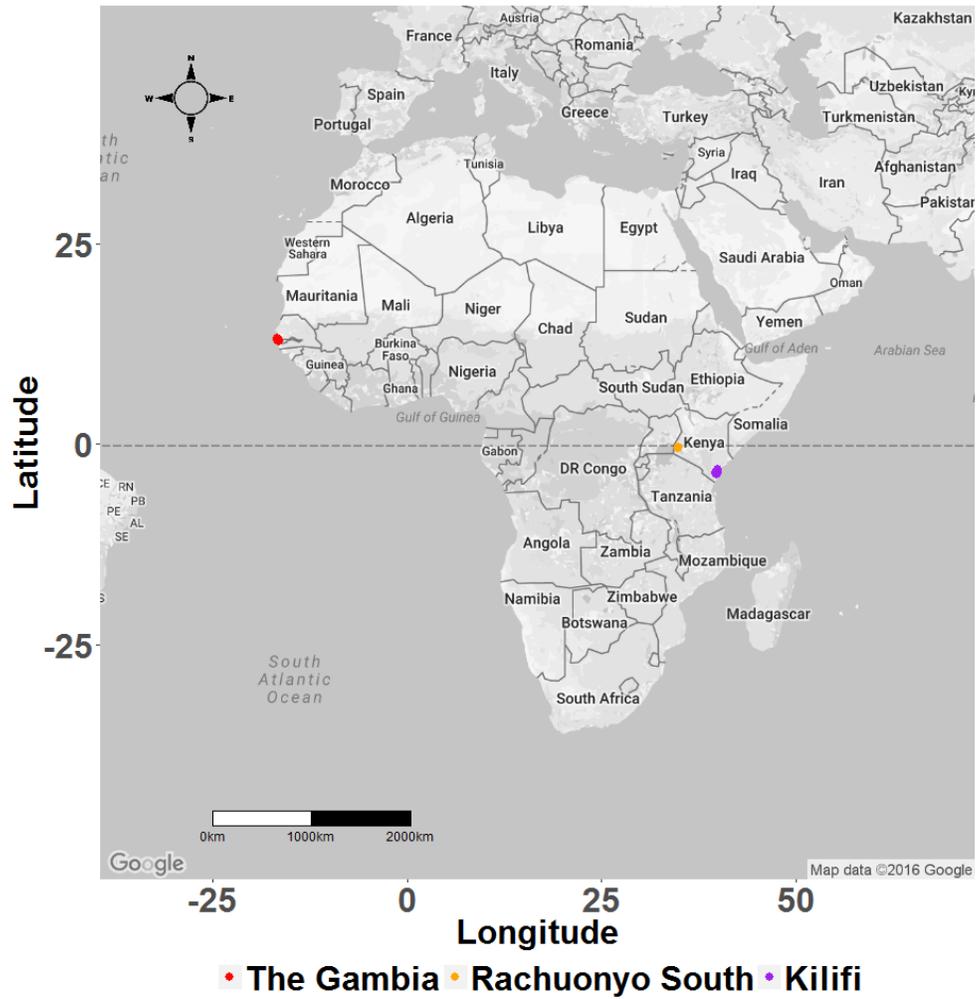


Figure 1. Map of Africa showing the three study sites. The study was conducted on *P. falciparum* samples collected in The Gambia, West Africa and Rachuonyo South District and Kilifi County in Kenya, East Africa.

Table 1. Summary of information on *P. falciparum* infected blood samples collected from The Gambia, Kilifi and Rachuonyo South study sites.

Study site	Contributing study	Study period	Average parasite density	Samples genotyped	Samples analysed	SNPs genotyped	SNPs analysed
The Gambia (Kombo Coastal Districts)	Clinical malaria study	Sep '07 – Dec '08	406,093	143	133	131	107
Kilifi	Community surveys	Feb – Oct '05	4562	748	195	240	177
Kilifi	Clinical malaria surveys	Jul '98 – Apr '10	352,428	1564	1407	240	177
Rachuonyo South	Community surveys	2011	NA	2744	1034	111	82

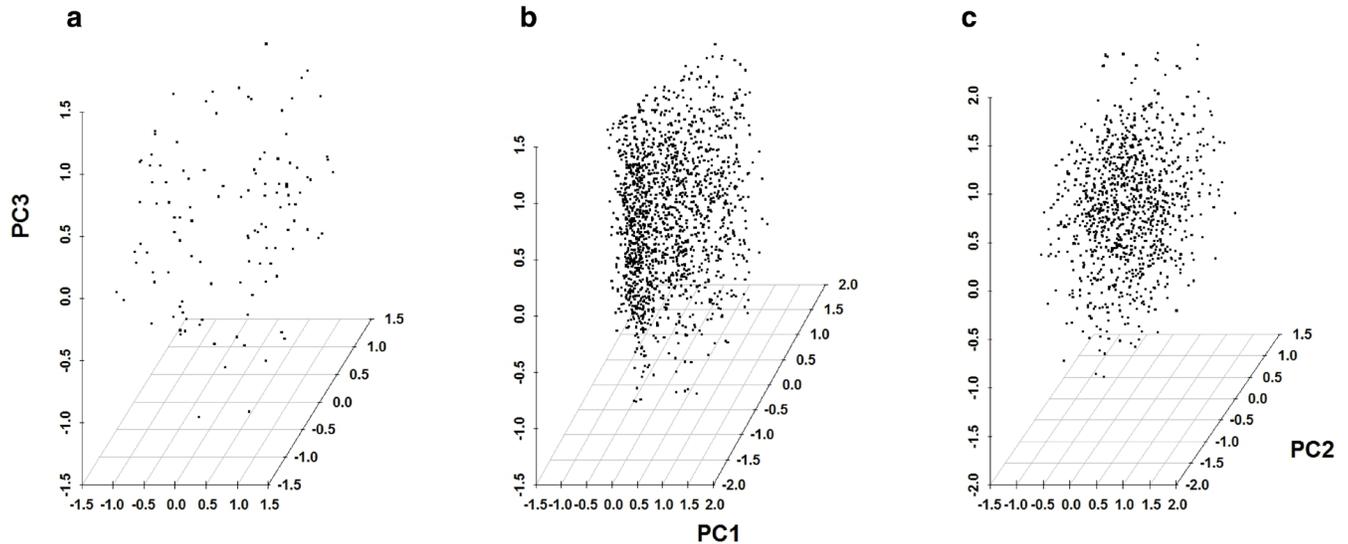


Figure 2. Plots of Principal Component Analysis scores for *P. falciparum* parasite populations in the study sites. Each point represents one of 133 parasites in The Gambia (a), 1602 parasites in Kilifi (b) and 1034 parasites in Rachuonyo South (c). Genetic structuring was not observed for any of the parasite populations based on these three principal components. Cumulatively, the first three principle components accounted for 36.1% (PC1=18.4%, PC2=10.4%, PC3=7.3%), 13.2% (PC1=5.1%, PC2=4.4%, PC3=3.7%) and 12.7% (PC1=4.4%, PC2=4.3%, PC3=4%) of the variability seen in The Gambia, Kilifi and Rachuonyo South populations, respectively.

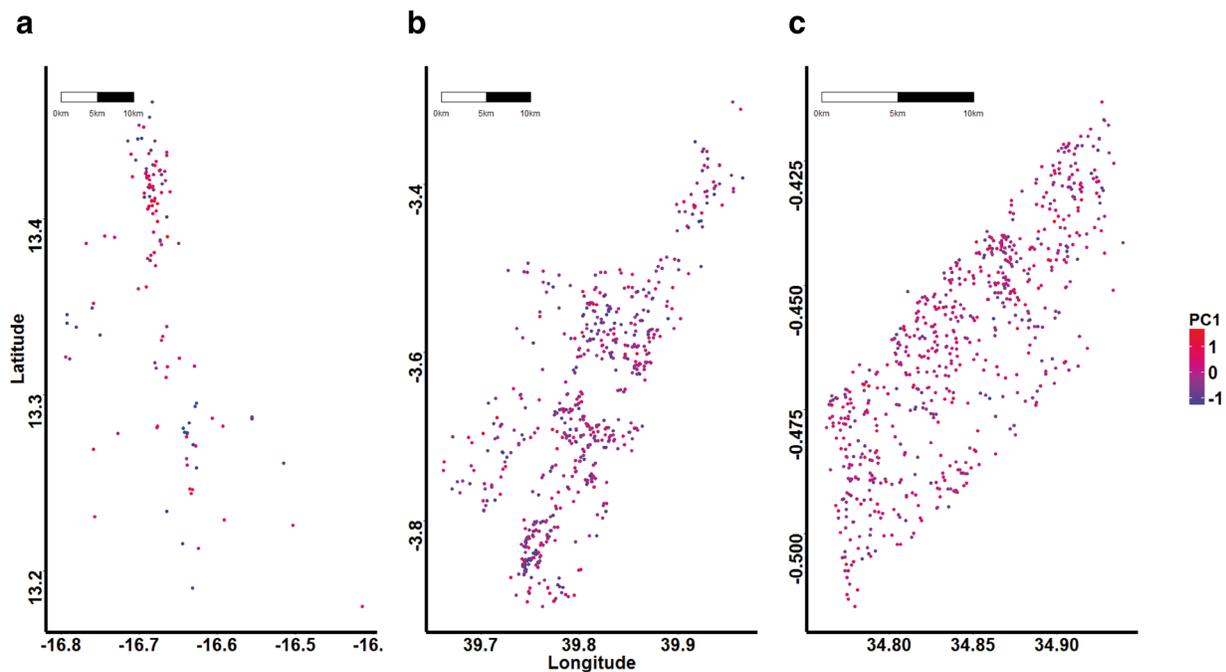


Figure 3. Geographic distribution of *P. falciparum* parasite genotypes based on scores for the first principal component. Each point represents the location of an individual parasite isolate and the colour shading represents distinct genotypes for parasites in (a) The Gambia, (b) Kilifi and (c) Rachuonyo South study sites.

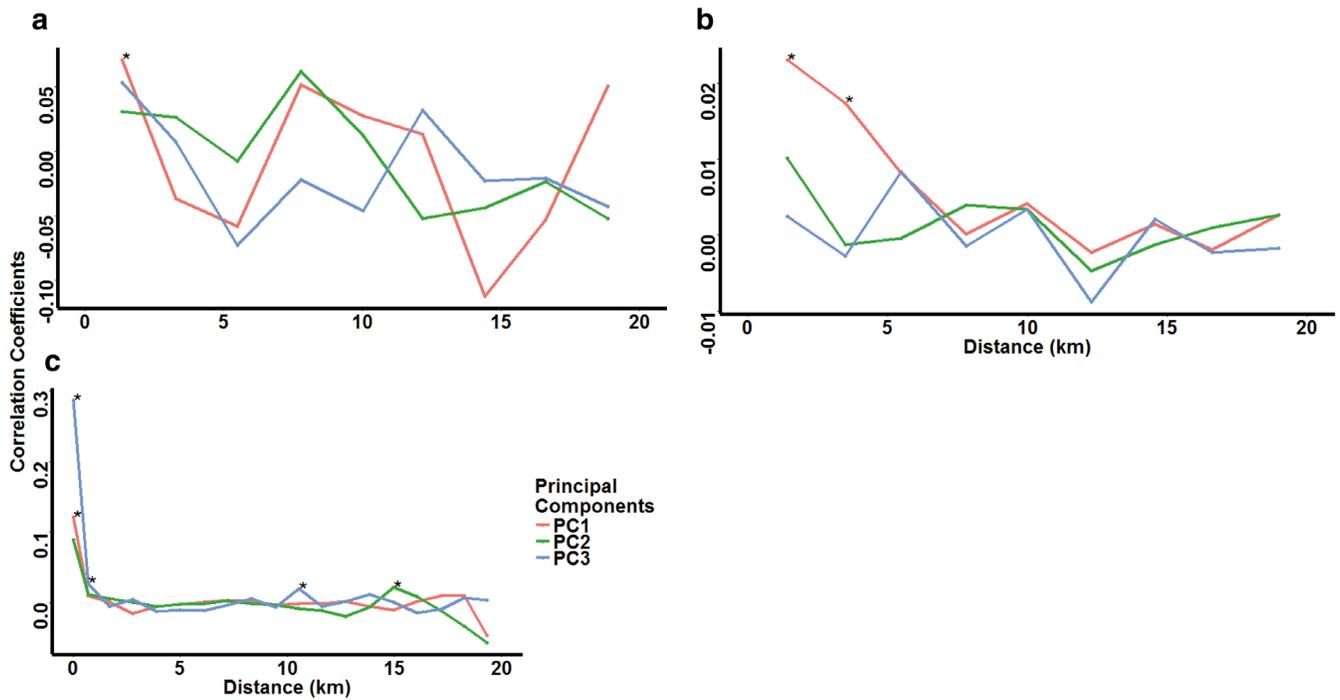


Figure 4. Moran's I spatial autocorrelation analysis for the first three principal components. Coefficients were computed at distance classes of 2 km for (a) The Gambia and (b) Kilifi, and 1 km for (c) Rachuonyo South parasite populations. Asterisks indicate distances at which parasites have significant ($p < 0.01$) autocorrelations. In The Gambia and Kilifi populations, only a few samples were collected from the same location, so Moran's I was not computed at this distance (0 km).

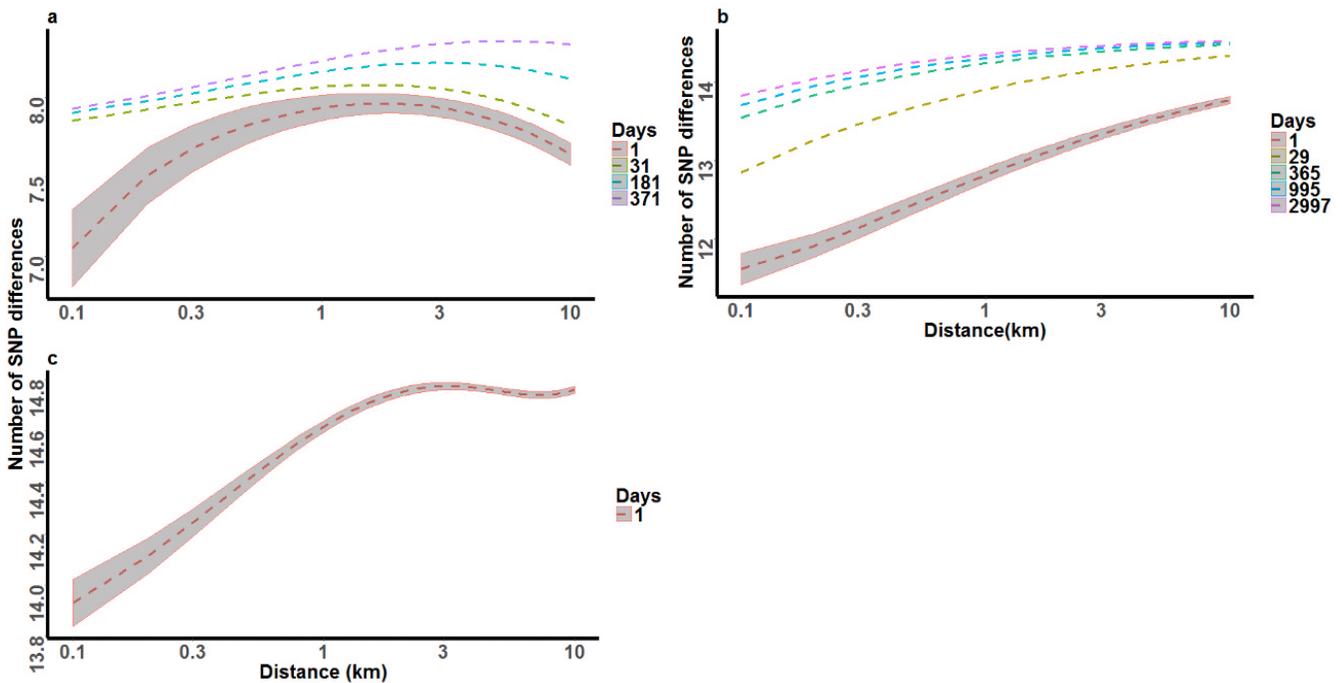


Figure 5. Effects of time-distance interaction on the number of SNP differences between parasite pairs. Dashed lines represent time intervals separating parasite pairs in (a) The Gambia, (b) Kilifi and (c) Rachuonyo South study sites. 95% confidence intervals are included around the 1-day curves in each study site. 100 pairwise analyses were used to generate the curves at each time point. 107, 177 and 82 SNPs were analysed in The Gambia, Kilifi and Rachuonyo South parasite populations, respectively. Dummy data used to generate the graphs contained 8 SNPs in the Gambia, 14 SNPs in Kilifi and 10 SNPs in Rachuonyo south.

Table 2. 95% bootstrap confidence intervals for the linear effects of time, distance and the interaction of time and distance on changes in SNP differences between parasite pairs.

	Time (days)	Distance (km)	Time-Distance interaction
The Gambia	-0.005 - -0.001 (p=0.004)	0.086 - 0.723 (p<0.001)	-0.0003 - -0.002 (p=0.003)
Kilifi	0.190 - 0.647 (p<0.001)	0.297 - 1.363 (p=0.001)	-0.453 - -0.072 (p=0.003)
Rachuonyo South	-	0.0104 - 0.275 (p=0.018)	-

Values represent the change in the number of SNP differences between parasite pairs per day (time), per kilometre (distance) and per day/kilometre (time-distance interaction). Time, distance and the log product of time and distance (time-distance interaction) were log transformed prior to running the regression analyses.

Identification of geographical barriers to parasite movement

We conducted raster analysis by pixels to examine a) the spatial relationship between distinct parasite genotypes as represented by the principal component analysis and either malaria positive fraction (MPF) data (in Kilifi) or PCR positive data (in Rachuonyo South) and b) the presence of possible spatial barriers to parasite movement that would act as factors. “The range of MPF and parasite prevalence per pixel varied depending on the size of the pixels analysed. In the Kilifi population, MPF ranged from 0 – 100% (interquartile range (IQR) = 20%) for the 0.5km pixels; 0 – 100% (IQR = 14%) for the 1.0km pixels; 20 – 83% (IQR = 7%) for the 2km pixels; and 33 – 63% (IQR = 4.7%) for the 4km pixels. In the Rachuonyo South population, PCR positive prevalence varied from 0 – 75% (IQR = 19.4%) for the 0.5km pixels; 0 – 47% (IQR = 17.4%) for the 1.0km pixels; 3.5 – 35.8% (IQR = 14.3%) for the 2km pixels; and 6.2 – 33.4% (IQR = 7.8%) for the 4km pixels.”

The analysis of principal components did not show any consistent or statistically strong associations with markers of transmission intensity (i.e. malaria positive fraction and prevalence of asymptomatic parasitaemia by PCR) (Supplementary Figure 5).

Bootstrapping the multivariable linear regression analysis of pairwise comparisons of samples for SNP differences using 189, 703 and 340 pixels for The Gambia, Kilifi and Rachuonyo South, respectively, showed that the majority of pixels were not significant influences on SNP differences (Supplementary Figure 6). The few pixels that were significant (p<0.05) were non-significant after applying Bonferroni correction to account for multiple testing. Furthermore the distribution of p values was uniform for each dataset (mean p value ~0.5 in each population).

Discussion

As malaria transmission declines, targeted control at the micro-epidemiological scale is likely to be important in eliminating malaria in any remaining transmission foci. The effectiveness of such targeted measures will depend on the extent of parasite mixing in and around these foci²³. In the current analysis, we did not identify any population structure by simple inspection of the Principal components derived from SNP genotyping in The Gambia, Kilifi and Rachuonyo South (Figure 2 and Figure 3), indicative of a parasite population that is well mixed. However we did not conclude that there was no structure to the population, only that we could not identify it in the absence of spatial data.

We therefore went on to analyse the genotype data using spatio-temporal data, and identified spatial autocorrelation using Moran's *I* in all three populations, with statistical significance (p<0.01) for the first principal component in The Gambia and Kilifi and the third principal component in Rachuonyo South (Figure 4). Overall, the consistent pattern observed in the Moran's *I* analyses was that of spatial auto-correlation at close proximity (i.e. at a range of a few km), and little or no auto-correlation at larger distances. The auto-correlation was modest in effect size but statistically significant with p values ranging from 0.01 to 0.001 at < 1 km. However, using scan statistics we identified only two specific clusters of distinct parasite sub-populations based on PC scores, one in Kilifi and another in Rachuonyo South. The limited evidence of specific local clusters of parasite sub-populations in the face of evidence of spatial auto-correlation over the whole study site implies that there is a high degree of mixing among parasites within the study sites, leading to limited clustering of parasites into genetically distinct sub-populations.

We further looked at the effect of time, distance and time-distance interaction on the variation in SNP differences between parasite pairs within individual study sites. Since the number of days differed for almost all parasite pairs, dummy data were included in the regression analysis to enable the generation of time-distance interaction graphs. For each study site, a distance range of 1 – 10km (with an interval of 0.1km between adjacent distances) was used. Temporal distance with 14 and 10 day intervals were assigned to parasite pairs in Kilifi and the Gambia, respectively, whereas time was not considered for the Rachuonyo South population. Constant SNP differences of 14, 10 and 8 were used for parasite pairs in Kilifi, Rachuonyo South and the Gambia, respectively. We found that time and distance were independently associated with increasing variation between parasite genotypes (i.e. the further apart in time or space two parasites were, the greater the genetic differences observed between them). However, in the case of The Gambia and Kilifi populations where we had longitudinal data, time was shown to interact antagonistically with distance, with an increase in time reducing the variations in genetic differences between parasites as distance between the parasites increased (Figure 5). This implies that distance between samples was no longer predictive of genetic variation when there were longer time periods between samples, indicating that, given enough time, even parasites that are separated by large distances would get a chance to interact and recombine, especially if they are not geographically isolated. The number of

SNP differences were seen to plateau at approximately 1km in the Gambia, 3km in Kilifi and 10km in Rachuonyo South. This may be attributed to the characteristics of the local parasite population, which in turn may be explained by the distribution of human settlement in the areas sampled, for example in the Gambia, homesteads tend to be clustered together in distinct, autonomous villages whereas in Rachuonyo South there is a denser and more uniform pattern of human settlement over the study area, enabling the interaction of parasites over a much larger distance.

Lack of genetic structuring of parasite populations observed in this study is indicative of a population that is well mixed. This observation of a highly mixing parasite population is in agreement with results of similar studies of parasite genetic diversity and population differentiation using microsatellites^{58,60,61}, immune selected genes^{62,63} and SNPs⁶⁴. These studies were carried out in parasite populations from different geographical regions representing a diverse range of transmission intensities, from the highest in Africa and Oceania, intermediate in Southeast Asia, to the lowest in South America. However, other studies have shown population structure when looking at the same population^{50,65-67}, although these analyses were carried out on larger geographical scales than those analysed here and mostly involved analyses at provincial, country or continental levels. Population structure was most evident in regions with low transmission intensities such as South America or Southeast Asia⁵⁸, and less evident in Africa where transmission intensity is much higher⁶¹.

On an international level, for example, some studies have been able to distinguish between Senegalese and Thai parasite isolates using a 24-SNP barcode⁶⁸, and another study using 4 SNPs out of a set of 384 SNPs was able to resolve East and West African parasites⁵⁰, showing that parasite populations can be resolved on a large geographical scale. A study in Senegal was also able to identify population structure among parasites using a 24 SNP barcode, despite a high level of similarity among the parasites analysed⁶⁹. It is possible that more detailed genotyping using a larger number of markers, for instance by whole genome sequencing, would start to identify mutations that are private to particular sub-populations at a finer geographical scale, although the degree of mixing observed here suggests that discrete populations are unlikely.

We identified spatial autocorrelation among parasites in the different study areas. However, most of these correlations were found over short distances, pointing to the existence of parasite sub-populations over small spatial scales. This indicates the presence of clusters of genetically distinct parasites at micro-epidemiological scales within the study sites. Previous studies have identified parasite sub-populations based on clustering of serological responses to the important antigen *Plasmodium falciparum* Erythrocyte Membrane Protein 1 (*PfEMP1*) in children in Kilifi⁷⁰, supporting our observations of parasite sub-populations at this site. In Papua New Guinea, sub-populations of parasites have also been identified at a micro-epidemiological scale using *PfEMP1*⁷¹, indicating that this may be a good marker for population differentiation at the micro-epidemiological level.

Studies on hotspots of symptomatic malaria infection have identified hotspots or clusters of infections down to the level of individual

homesteads in Kilifi⁹. The lack of consistent correlations between parasite genotypes and infection prevalence shown through raster analysis of pixels in this study ([Supplementary Figure 5](#)) indicate that infections within higher incidence areas are likely not caused by distinct parasite sub-populations. Instead, such infections are likely caused by parasites that are well mixed within the general population. Our inability to detect barriers to parasite movement over short distances indicates that parasites move freely within the study areas, and the spatial extent of such parasites may be limited only by the ecology and dispersal range of mosquito vectors. Furthermore, recent examination of the epidemiology of hotspots shows that they occur at the full range of spatial scales, with a pattern of spatial auto-correlation that does not show a discontinuity at any scale (i.e. a smooth semi-variogram)⁹. This further argues against the existence of discrete “units” of transmission with sub-populations of parasites.

This has implications for public health interventions that may target transmission hotspots. If hotspots consist of distinct parasite populations that do not mix with parasite populations in the wider parasite community, the impact of hotspot-targeted interventions beyond the hotspot boundaries can be expected to be limited. If parasites mix freely, as suggested by our data, the impact of hotspot-targeted interventions may affect community-wide malaria transmission. This assumes that hotspots can be detected, are stable in time²⁰ and the spread of parasite populations indeed primarily occurs from hotspots to the surrounding community²³.

This study had some limitations. First, the number of SNPs typed was relatively small, and this would have limited our power to detect genetic structuring among the highly similar parasite populations, especially in The Gambia. Detecting structuring in highly similar parasite populations may require either a much larger panel of SNPs or the use of more informative SNPs, as shown in the study by Campino *et al*, 2011⁵⁰. Advances in sequencing technologies have increased the use of whole genome sequence data in the analysis of *P. falciparum* parasite population genetics, and this has led to the identification of hundreds of thousands of SNPs, most of which are present at very low frequencies especially in African parasite populations⁷². Additional analyses will require the use of whole genome sequence data to identify rare variants and distinguish between closely related parasites, thus allowing parasite population structure to be analysed at fine spatial scales. However, despite the small SNP panel used in this study, we were still able to detect population structuring on a micro-epidemiological scale. Our analysis suggests that this structure was a uniform spatial and temporal auto-correlation rather than driven by discrete clusters of parasites at specific locations. Despite the limitations of our SNP typing and sample size we can therefore conclude that any specific clustering is less prominent as a feature than the auto-correlations in space and time that we can detect.

A second limitation is that we conducted our study in only two sites in Kenya, and one site in the Gambia. It may be premature to generalize our results more widely and an analysis of more sites will be required to make confident generalizations. On the other hand the three sites selected do demonstrate differing transmission intensities typical of many endemic Sub Saharan African countries, and this was reflected in the level of genetic diversity observed in

the populations. Furthermore, our findings are consistent across all three sites. Nevertheless, patterns of parasite mixing may differ between populations based on distinctive features such as geographic isolation and patterns of human movement. Further data are required to make more general conclusions. Furthermore, as transmission continues to decline and malaria programmes gradually shift their focus from control to elimination, the analysis of parasite gene flow between different transmission foci, e.g. Kilifi and Rachuonyo South, will become increasingly important in informing the mitigation measures needed to prevent importation of parasites as a result of human movement and migration. These analyses were not carried out in the current study since the numbers of common SNPs between the two Kenyan sites was low, and we only had parasites from one timepoint in Rachuonyo South, hence we were unable to conduct an informative analysis of gene flow between sites.

Finally, we used genetic data to show that there is high parasite movement and mixing within individual study sites. Additional analyses using gene flow models, e.g. as implemented in Migrate-N software, can be used to further validate our hypothesis of rapid gene flow and to confirm whether the parasites are part of a panmictic population or whether there exists underlying population structure, as well as to determine directionality of parasite movement between different populations, assuming that such populations can be identified within the region.

In conclusion, we have shown that *Plasmodium falciparum* parasite populations mix evenly within The Gambia, Kilifi and Rachuonyo South and there appear to be no detectable geographical barriers to parasite movement over short distances within these sites. That said, autocorrelations of genotype were detected at the micro-epidemiological level. We would conclude that control strategies that efficiently target hotspots will likely benefit the wider community outside the hotspots at the District/County level (we are however unable to comment on larger geographical scales), although this is likely to be affected by factors such as the underlying transmission level, heterogeneity of transmission, and patterns of human movement²³. On the other hand, based on the high level of parasite mixing observed at each study site, we would predict that ineffective application of control interventions such as mass drug administration that result in residual foci of transmission would lead to rapid re-infection of the wider community, and also that parasites acquiring mutations conferring drug resistance or immunological escape would spread rapidly at the micro-epidemiological level. This underscores the need for effective and sustained control until malaria elimination is achieved.

Data availability

Figshare: Dataset 1: Information on the 276 SNPs genotyped in 177 genes in *P. falciparum* parasite populations from The Gambia,

Kilifi and Rachuonyo South, doi: <https://dx.doi.org/10.6084/m9.figshare.5383969>⁴⁹

Figshare: Dataset 2: Sequenom assay design information, doi: <http://dx.doi.org/10.6084/m9.figshare.4640719>⁵²

Figshare: Dataset 3: SNP, distance and time differences between *P. falciparum* parasite pairs in The Gambia population, doi: <http://dx.doi.org/10.6084/m9.figshare.4640725>⁵⁵

Figshare: Dataset 4: SNP, distance and time differences between *P. falciparum* parasite pairs in the Kilifi population, doi: <http://dx.doi.org/10.6084/m9.figshare.4640725>⁵⁶

Figshare: Dataset 5: SNP and distance differences between *P. falciparum* parasite pairs in the Rachuonyo South population, doi: <http://dx.doi.org/10.6084/m9.figshare.4640728>⁵⁷

Author contributions

P.B, D.K and T.B conceived and designed the study. T.B, A.A, J.C.S, A.Y.B, T.N.W, K.R, D.K. were involved in sample and/or data collection. C.H, A.J, K.R genotyped the SNPs. I.A.O, P.B. P.M., A.R, E.V, I.O and G.F analysed the data. I.A.O wrote the first draft manuscript. All authors reviewed, edited and approved the final manuscript.

Competing interests

No competing interests were disclosed.

Grant information

Sample collection at the Rachuonyo South site was supported by the Bill and Melinda Gates Foundation, under the Malaria Transmission Consortium, Grant No.45114 and the Grand Challenge Grant No. OPP1024438. Thomas N. Williams is funded by the Wellcome Trust, grant number 091758. Philip Bejon, Polycarp Mogeni and Irene Omedo are funded by the UK Medical Research Council (MRC) and the UK Department for International Development (DFID) under the MRC/DFID Concordat agreement. Polycarp Mogeni is funded by the Gottfried und Julia Bangerter-Rhyner Stiftung and the Novartis Foundation for Medical Biological Research project 13A13. Sample collection in Kilifi was supported by core funding from the Wellcome Trust to the Kenya Programme.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

The paper is published with the permission of the Director of KEMRI.

Supplementary material

Supplementary figure 1. Time-distance interaction curves showing the effect of distance on changes in genetic variation between Gambian *P. falciparum* parasite pairs over time. The analyses were carried out for **a)** pooled SNPs, **b)** EBA-175, **c)** AMA1 and **d)** ‘other’ SNPs (all SNPs excluding EBA175 and AMA1). Dashed lines represent time intervals separating parasite pairs at 1 day (red), 1 month (green), 6 months (blue) and 1 year (purple). The interaction term was log-transformed prior to running the analysis.

[Click here to access the data.](#)

Supplementary figure 2. Time-distance interaction curves showing the effect of distance on changes in genetic variation between Kilifi *P. falciparum* parasite pairs over time. The analyses were carried out for **a)** pooled SNPs, **b)** EBA-175, **c)** AMA1 and **d)** ‘other’ SNPs (all SNPs excluding EBA175 and AMA1). Dashed lines represent time intervals separating parasite pairs at 1 day (red), 1 month (green), 6 months (blue) and 1 year (purple). The interaction term was log-transformed prior to running the analysis.

[Click here to access the data.](#)

Supplementary figure 3. Time-distance interaction curves showing the effect of distance on changes in genetic variation between Rachuonyo South *P. falciparum* parasite pairs over time. The analyses were carried out for **a)** pooled SNPs and **b)** EBA-175. Parasites were all collected during a single cross sectional study thus time was not considered.

[Click here to access the data.](#)

Supplementary Figure 4. Geographical distribution of *P. falciparum* parasite genotypes based on scores for the second (PC2) and third (PC3) principal components. Each point represents an individual parasite isolate and the colour shading represents distinct genotypes for parasites in The Gambia (**d** and **g**), Kilifi (**e** and **h**), and Rachuonyo South (**f** and **i**) study sites.

[Click here to access the data.](#)

Supplementary Figure 5. Raster analysis by pixels. This was carried out to determine the spatial relationship between distinct parasite genotypes as represented by principal component analysis and either malaria positive fraction (MPF) or PCR positive fraction (PPF) data. **(a)** and **(b)** show the distribution of scores for the first principal component (PC1) and MPF over a 1 km × 1 km grid area of Kilifi. **(d)** and **(e)** show the distribution of scores for the first principal component and PPF over a 1 km × 1 km grid area of Rachuonyo South. Spearman’s correlation coefficients computed to show the relationship between parasite genotypes and either MPF **(c)** or PPF **(f)** showed no strong associations between genotypes and the two markers of transmission.

[Click here to access the data.](#)

Supplementary Figure 6. Raster analysis by pixels to examine the presence of spatial barriers to parasite movement. The pixel plots represent p values of bootstrapped linear regression correlation coefficients and show the significance of different geographical locations in acting as barriers to parasite mixing. Individual grid sizes were of approximately 1 km × 1 km in **(a)** Kilifi and **(c)** The Gambia and 0.5 km × 0.5 km in **(b)** Rachuonyo South. The colour key in each case indicates the range of p values from 0.0001 to 1. Significant p values shown on the plot were non-significant after applying Bonferroni correction to account for multiple testing.

[Click here to access the data.](#)

References

- Smith DL, Battle KE, Hay SI, *et al.*: **Ross, macdonald, and a theory for the dynamics and control of mosquito-transmitted pathogens.** *PLoS pathog.* 2012; **8**(4): e1002588.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Reiner RC Jr, Perkins TA, Barker CM, *et al.*: **A systematic review of mathematical models of mosquito-borne pathogen transmission: 1970–2010.** *J R Soc Interface.* 2013; **10**(81): 20120921.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Woolhouse ME, Dye C, Etard JF, *et al.*: **Heterogeneities in the transmission of infectious agents: implications for the design of control programs.** *Proc Natl Acad Sci U S A.* 1997; **94**(1): 338–342.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Perkins TA, Scott TW, Le Menach A, *et al.*: **Heterogeneity, mixing, and the spatial scales of mosquito-borne pathogen transmission.** *PLoS Comput Biol.* 2013; **9**(12): e1003327.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Prosper O, Ruktanonchai N, Martcheva M: **Assessing the role of spatial heterogeneity and human movement in malaria dynamics and control.** *J Theor Biol.* 2012; **303**: 1–14.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Acevedo MA, Prosper O, Lopiano K, *et al.*: **Spatial heterogeneity, host movement and mosquito-borne disease transmission.** *PLoS One.* 2015; **10**(6): e0127552.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Rumisha SF, Smith T, Abdulla S, *et al.*: **Modelling heterogeneity in malaria transmission using large sparse spatio-temporal entomological data.** *Glob Health Action.* 2014; **7**: 22682.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Baidjoe AY, Stevenson J, Knight P, *et al.*: **Factors associated with high**

- heterogeneity of malaria at fine spatial scale in the Western Kenyan highlands. *Malar J*. 2016; 15: 307.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
9. Bejon P, Williams TN, Nyundo C, *et al.*: A micro-epidemiological analysis of febrile malaria in Coastal Kenya showing hotspots within hotspots. *eLife*. 2014; 3: e02130.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 10. Bhatt S, Weiss DJ, Cameron E, *et al.*: The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. *Nature*. 2015; 526(7572): 207–211.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 11. Bousema T, Drakeley C, Gesase S, *et al.*: Identification of hot spots of malaria transmission for targeted malaria control. *J Infect Dis*. 2010; 201(11): 1764–1774.
[PubMed Abstract](#) | [Publisher Full Text](#)
 12. Cook J, Reid H, Iavro J, *et al.*: Using serological measures to monitor changes in malaria transmission in Vanuatu. *Malar J*. 2010; 9: 169.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 13. Cook J, Speybroeck N, Sochantara T, *et al.*: Sero-epidemiological evaluation of changes in *Plasmodium falciparum* and *Plasmodium vivax* transmission patterns over the rainy season in Cambodia. *Malar J*. 2012; 11: 86.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 14. Gething PW, Casey DC, Weiss DJ, *et al.*: Mapping *Plasmodium falciparum* Mortality in Africa between 1990 and 2015. *N Engl J Med*. 2016; 375(25): 2435–2445.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 15. Gething PW, Patil AP, Smith DL, *et al.*: A new world malaria map: *Plasmodium falciparum* endemicity in 2010. *Malar J*. 2011; 10: 378.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 16. Machault V, Vignolles C, Pagès F, *et al.*: Spatial heterogeneity and temporal evolution of malaria transmission risk in Dakar, Senegal, according to remotely sensed environmental data. *Malar J*. 2010; 9: 252.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 17. Noor AM, Kinyoki DK, Mundia CW, *et al.*: The changing risk of *Plasmodium falciparum* malaria infection in Africa: 2000–10: a spatial and temporal analysis of transmission intensity. *Lancet*. 2014; 383(9930): 1739–1747.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 18. Oduro AR, Conway DJ, Schellenberg D, *et al.*: Seroepidemiological and parasitological evaluation of the heterogeneity of malaria infection in the Gambia. *Malar J*. 2013; 12: 222.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 19. Okebe J, Affara M, Correa S, *et al.*: School-based countrywide seroprevalence survey reveals spatial heterogeneity in malaria transmission in the Gambia. *PLoS One*. 2014; 9(10): e110926.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 20. Bejon P, Williams TN, Lijander A, *et al.*: Stable and unstable malaria hotspots in longitudinal cohort studies in Kenya. *PLoS Med*. 2010; 7(7): e1000304.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 21. Bousema T, Griffin JT, Sauerwein RW, *et al.*: Hitting hotspots: spatial targeting of malaria for control and elimination. *PLoS Med*. 2012; 9(1): e1001165.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 22. Bousema T, Stevenson J, Baidjoe A, *et al.*: The impact of hotspot-targeted interventions on malaria transmission: study protocol for a cluster-randomized controlled trial. *Trials*. 2013; 14: 36.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 23. Bousema T, Stresman G, Baidjoe AY, *et al.*: The Impact of Hotspot-Targeted Interventions on Malaria Transmission in Rachuonyo South District in the Western Kenyan Highlands: A Cluster-Randomized Controlled Trial. *PLoS Med*. 2016; 13(4): e1001993.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 24. Dolgin E: Targeting hotspots of transmission promises to reduce malaria. *Nat Med*. 2010; 16(10): 1055.
[PubMed Abstract](#) | [Publisher Full Text](#)
 25. Kangoye DT, Noor A, Midega J, *et al.*: Malaria hotspots defined by clinical malaria, asymptomatic carriage, PCR and vector numbers in a low transmission area on the Kenyan Coast. *Malar J*. 2016; 15: 213.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 26. Smith DL, McKenzie FE, Snow RW, *et al.*: Revisiting the basic reproductive number for malaria and its implications for malaria control. *PLoS Biol*. 2007; 5(3): e42.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 27. Carter R, Mendis KN, Roberts D: Spatial targeting of interventions against malaria. *Bull World Health Organ*. 2000; 78(12): 1401–1411.
[PubMed Abstract](#) | [Free Full Text](#)
 28. Ruktanonchai NW, DeLeenheer P, Tatem AJ, *et al.*: Identifying Malaria Transmission Foci for Elimination Using Human Mobility Data. *PLoS Comput Biol*. 2016; 12(4): e1004846.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 29. McCall PJ, Moshia FW, Njunwa KJ, *et al.*: Evidence for memorized site-fidelity in *Anopheles arabiensis*. *Trans R Soc Trop Med Hyg*. 2001; 95(6): 587–590.
[PubMed Abstract](#) | [Publisher Full Text](#)
 30. Alonso PL, Lindsay SW, Armstrong JR, *et al.*: The effect of insecticide-treated bed nets on mortality of Gambian children. *Lancet*. 1991; 337(8756): 1499–1502.
[PubMed Abstract](#) | [Publisher Full Text](#)
 31. Nevill CG, Some ES, Mung'ala VO, *et al.*: Insecticide-treated bednets reduce mortality and severe morbidity from malaria among children on the Kenyan coast. *Trop Med Int Health*. 1996; 1(2): 139–146.
[PubMed Abstract](#) | [Publisher Full Text](#)
 32. Hawley WA, Phillips-Howard PA, ter Kuile FO, *et al.*: Community-wide effects of permethrin-treated bed nets on child mortality and malaria morbidity in western Kenya. *Am J Trop Med Hyg*. 2003; 68(4 Suppl): 121–127.
[PubMed Abstract](#)
 33. Ceasay SJ, Casals-Pascual C, Nwakanma DC, *et al.*: Continued decline of malaria in The Gambia with implications for elimination. *PLoS One*. 2010; 5(8): e12242.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 34. Ceasay SJ, Casals-Pascual C, Erskine J, *et al.*: Changes in malaria indices between 1999 and 2007 in The Gambia: a retrospective analysis. *Lancet*. 2008; 372(9649): 1545–1554.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 35. Scott JA, Bauni E, Moisi JC, *et al.*: Profile: The Kilifi Health and Demographic Surveillance System (KHDSS). *Int J Epidemiol*. 2012; 41(3): 650–657.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 36. Mwesigwa J, Okebe J, Affara M, *et al.*: On-going malaria transmission in The Gambia despite high coverage of control interventions: a nationwide cross-sectional survey. *Malar J*. 2015; 14: 314.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 37. Caputo B, Nwakanma D, Jawara M, *et al.*: *Anopheles gambiae* complex along The Gambia river, with particular reference to the molecular forms of *An. gambiae* s.s. *Malar J*. 2008; 7: 182.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 38. Mwangangi JM, Mbogo CM, Orindi BO, *et al.*: Shifts in malaria vector species composition and transmission dynamics along the Kenyan coast over the past 20 years. *Malar J*. 2013; 12: 13.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 39. Stevenson J, St Laurent B, Lobo NF, *et al.*: Novel vectors of malaria parasites in the western highlands of Kenya. *Emerg Infect Dis*. 2012; 18(9): 1547–1549.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 40. O'Meara WP, Bejon P, Mwangi TW, *et al.*: Effect of a fall in malaria transmission on morbidity and mortality in Kilifi, Kenya. *Lancet*. 2008; 372(9649): 1555–1562.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 41. Okiro EA, Alegana VA, Noor AM, *et al.*: Malaria paediatric hospitalization between 1999 and 2008 across Kenya. *BMC Med*. 2009; 7: 75.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 42. Sonko ST, Jaiteh M, Jafali J, *et al.*: Does socio-economic status explain the differentials in malaria parasite prevalence? Evidence from The Gambia. *Malar J*. 2014; 13: 449.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 43. Midega JT, Smith DL, Olotu A, *et al.*: Wind direction and proximity to larval sites determines malaria risk in Kilifi District in Kenya. *Nat Commun*. 2012; 3: 674.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 44. Stevenson JC, Stresman GH, Gitonga CW, *et al.*: Reliability of school surveys in estimating geographic variation in malaria transmission in the western Kenyan highlands. *PLoS One*. 2013; 8(10): e77641.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 45. Mogeni P, Williams TN, Fegan G, *et al.*: Age, Spatial, and Temporal Variations in Hospital Admissions with Malaria in Kilifi County, Kenya: A 25-Year Longitudinal Observational Study. *PLoS Med*. 2016; 13(6): e1002047.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 46. Bejon P, Mwacharo J, Kai O, *et al.*: A phase 2b randomised trial of the candidate malaria vaccines FP9 ME-TRAP and MVA ME-TRAP among children in Kenya. *PLoS Clin Trials*. 2006; 1(6): e29.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 47. Polley SD, Conway DJ, Cavanagh DR, *et al.*: High levels of serum antibodies to merozoite surface protein 2 of *Plasmodium falciparum* are associated with reduced risk of clinical malaria in coastal Kenya. *Vaccine*. 2006; 24(19): 4233–4246.
[PubMed Abstract](#) | [Publisher Full Text](#)
 48. Olotu A, Fegan G, Williams TN, *et al.*: Defining clinical malaria: the specificity and incidence of endpoints from active and passive surveillance of children in rural Kenya. *PLoS One*. 2010; 5(12): e15569.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 49. Omedo I, Mogeni P, Bousema T, *et al.*: Dataset 1. Information on the 276 SNPs genotyped in 177 genes in *P. falciparum* parasite populations from The Gambia, Kilifi and Rachuonyo South. *Figshare*. 2017.
[Data Source](#)
 50. Campino S, Auburn S, Kivinen K, *et al.*: Population genetic analysis of *Plasmodium falciparum* parasites using a customized Illumina GoldenGate genotyping assay. *PLoS One*. 2011; 6(6): e20251.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 51. Gabriel S, Ziaugra L, Tabbaa D: SNP genotyping using the sequenom MassARRAY iPLEX platform. *Curr Protoc Hum Genet*. 2009; Chapter 2: Unit 2.12.
[PubMed Abstract](#) | [Publisher Full Text](#)
 52. Omedo I, Mogeni P, Bousema T, *et al.*: Dataset 2: Sequenom assay design

- information. *Figshare*. 2017.
[Data Source](#)
53. RCoreTeam: **R: A language and environment for statistical computing.** R Foundation for Statistical Computing. Vienna, Austria. 2013.
[Reference Source](#)
 54. Kulldorf M: **SaTScan v9.3: Software for the spatila and space-time scan statistics.** 2014.
[Reference Source](#)
 55. Omedo I, Mogeni P, Bousema T, *et al.*: **Dataset 3: SNP, distance and time differences between *P. falciparum* parasite pairs in The Gambia population.** *Figshare*. 2017.
[Data Source](#)
 56. Omedo I, Mogeni P, Bousema T, *et al.*: **Dataset 4: SNP, distance and time differences between *P. falciparum* parasite pairs in the Kilifi population.** *Figshare*. 2017.
[Data Source](#)
 57. Omedo I, Mogeni P, Bousema T, *et al.*: **Dataset 5: SNP and distance differences between *P. falciparum* parasite pairs in the Rachuonyo South population.** *Figshare*. 2017.
[Data Source](#)
 58. Anderson TJ, Haubold B, Williams JT, *et al.*: **Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*.** *Mol Biol Evol.* 2000; **17**(10): 1467–1482.
[PubMed Abstract](#) | [Publisher Full Text](#)
 59. Manske M, Miotto O, Campino S, *et al.*: **Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing.** *Nature.* 2012; **487**(7407): 375–379.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 60. Bakhiet AM, Abdel-Muhsin AM, Elzaki SE, *et al.*: ***Plasmodium falciparum* population structure in Sudan post artemisinin-based combination therapy.** *Acta Trop.* 2015; **148**: 97–104.
[PubMed Abstract](#) | [Publisher Full Text](#)
 61. Oyebola MK, Idowu ET, Nyang H, *et al.*: **Microsatellite markers reveal low levels of population sub-structuring of *Plasmodium falciparum* in southwestern Nigeria.** *Malar J.* 2014; **13**: 493.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 62. Bordbar B, Tuikue Ndam N, Renard E, *et al.*: **Genetic diversity of VAR2CSA ID1-DBL2Xb in worldwide *Plasmodium falciparum* populations: impact on vaccine design for placental malaria.** *Infect Genet Evol.* 2014; **25**: 81–92.
[PubMed Abstract](#) | [Publisher Full Text](#)
 63. Duan J, Mu J, Thera MA, *et al.*: **Population structure of the genes encoding the polymorphic *Plasmodium falciparum* apical membrane antigen 1: implications for vaccine design.** *Proc Natl Acad Sci U S A.* 2008; **105**(22): 7857–7862.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 64. Mobegi VA, Duffy CW, Amambua-Ngwa A, *et al.*: **Genome-wide analysis of selection on the malaria parasite *Plasmodium falciparum* in West African populations of differing infection endemicity.** *Mol Biol Evol.* 2014; **31**(6): 1490–1499.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 65. Bogreau H, Renaud F, Bouchiba H, *et al.*: **Genetic diversity and structure of African *Plasmodium falciparum* populations in urban and rural areas.** *Am J Trop Med Hyg.* 2006; **74**(6): 953–959.
[PubMed Abstract](#)
 66. Pumpaibool T, Amathau C, Durand P, *et al.*: **Genetic diversity and population structure of *Plasmodium falciparum* in Thailand, a low transmission country.** *Malar J.* 2009; **8**: 155.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 67. Anderson TJ, Nair S, Sudimack D, *et al.*: **Geographical distribution of selected and putatively neutral SNPs in Southeast Asian malaria parasites.** *Mol Biol Evol.* 2005; **22**(12): 2362–2374.
[PubMed Abstract](#) | [Publisher Full Text](#)
 68. Daniels R, Volkman SK, Milner DA, *et al.*: **A general SNP-based molecular barcode for *Plasmodium falciparum* identification and tracking.** *Malar J.* 2008; **7**: 223.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 69. Daniels RF, Schaffner SF, Wenger EA, *et al.*: **Modeling malaria genomics reveals transmission decline and rebound in Senegal.** *Proc Natl Acad Sci U S A.* 2015; **112**(22): 7067–7072.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 70. Bejon P, Turner L, Lavstsen T, *et al.*: **Serological evidence of discrete spatial clusters of *Plasmodium falciparum* parasites.** *PLoS One.* 2011; **6**(6): e21711.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 71. Tessema SK, Monk SL, Schultz MB, *et al.*: **Phylogeography of *var* gene repertoires reveals fine-scale geospatial clustering of *Plasmodium falciparum* populations in a highly endemic area.** *Mol Ecol.* 2015; **24**(2): 484–497.
[PubMed Abstract](#) | [Publisher Full Text](#)
 72. MalariaGEN *Plasmodium falciparum* Community Project: **Genomic epidemiology of artemisinin resistant malaria.** *eLife.* 2016; **5**: pii: e08714.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Referee Status:



Version 2

Referee Report 08 September 2017

<https://doi.org/10.21956/wellcomeopenres.13603.r25837>



Christopher Delgado-Ratto 

Global Health Institute, University of Antwerp, Antwerp, Belgium

I acknowledge the answers to my comments by the authors. I have no further comments on this new version.

Competing Interests: No competing interests were disclosed.

Referee Expertise: Molecular epidemiology, population genetics

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Referee Report 19 June 2017

<https://doi.org/10.21956/wellcomeopenres.11628.r23565>



Liwang Cui 

Department of Entomology, Pennsylvania State University, State College, PA, USA

This study analyzed large sample sets of malaria parasites taken from the western and eastern coasts of Africa (The Gambia and Kenya) and genotyped at 276 SNPs. For two of the sample sets, parasites were collected at different time points, allowing identification of population changes over time and space. Overall, the analysis was sound and results were well explained. The authors also notified the limitations of the study. For example, inclusion of additional parasite samples between these western and eastern sites, and use of more SNP markers would validate whether the conclusions drawn here represent the whole African continent.

Comments:

1. The assumption for comparing the temporally collected samples is that malaria case numbers have been reduced, which might lead to genetic isolation and structuring of parasite populations. It would be great if malaria epidemiology at the beginning and end of sample collection in the sites where samples were collected is clearly stated. It is possible that despite the overall reduction in

malaria cases, some of the sites may represent hotspots where malaria epidemiology remained more or less unchanged over the time. As a result, this would make the parasite populations and genetics relatively stable over the time.

2. The inclusion of numerous SNPs for this type of analysis is a nice practice. However, the authors may want to separate those that are clearly under selection (such as EBA175 and AMA1), since these mutations are subject to strong immune selection and will have different evolutionary trajectories as compared to more neutral SNPs.
3. More detailed comparison of the two Kenyan sites might be interesting to see whether gene flow between these sites exists, given that these sites are relatively closely located, yet separated by potential gene flow barriers (such as the rift valley).

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

I cannot comment. A qualified statistician is required.

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Referee Expertise: Molecular epidemiology

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 29 Aug 2017

Irene Omedo, KEMRI- Wellcome Trust Research Programme, Kenya

We are grateful for this review and the helpful comments and suggestions that have been made. We have included a point-by-point response (in bold) to the issues raised.

The assumption for comparing the temporally collected samples is that malaria case numbers have been reduced, which might lead to genetic isolation and structuring of parasite populations. It would be great if malaria epidemiology at the beginning and end of sample collection in the sites where samples were collected is clearly stated. It is possible that despite the overall reduction in malaria cases, some of the sites may represent hotspots where malaria epidemiology remained

more or less unchanged over the time. As a result, this would make the parasite populations and genetics relatively stable over the time.

The following statement has been added to show the changing epidemiology of malaria during the study period:

“Over the study period, malaria transmission, as measured by malaria slide positivity rate, fell from 56% in 1998 to 7% in 2009 in Kilifi¹, and rose slightly in Fajara and Brikama in the Gambia².”

There was no data showing temporal variation in malaria transmission in Rachuonyo South because the samples were collected in a single cross-sectional survey.

The inclusion of numerous SNPs for this type of analysis is a nice practice. However, the authors may want to separate those that are clearly under selection (such as EBA175 and AMA1), since these mutations are subject to strong immune selection and will have different evolutionary trajectories as compared to more neutral SNPs.

We have included the following additional text to show the number of SNPs typed in EBA175 and AMA1:

“Separate analyses were also carried out for subsets of SNPs typed in EBA 175 (39, 36 and 20 SNPs in The Gambia, Kilifi and Rachuonyo South, respectively) and AMA1 (9 SNPs in The Gambia and 8 SNPs in Kilifi). Only 3 SNPs were genotyped in Rachuonyo South, so this SNP subset was not analysed separately.”

And expounded on the section describing the outcome of these analyses:

“In all study sites, separate analyses of EBA175 and AMA1 did not reveal qualitatively different results from the pooled analyses (supplementary figures 1- 3) and only the results of the pooled analyses are presented here.”

More detailed comparison of the two Kenyan sites might be interesting to see whether gene flow between these sites exists, given that these sites are relatively closely located, yet separated by potential gene flow barriers (such as the rift valley).

The current study focused on parasite movement and mixing within small, geographically-defined areas, and hence concentrated on analysing parasite genetics within individual sites. Furthermore, we did not have sufficient number of SNPs typed in these two populations to carry out a meaningful comparison. That said, your comment is important, and has been noted as a recommendation for future work, in the statement below:

“Furthermore, as transmission continues to decline and malaria programmes gradually shift their focus from control to elimination, the analysis of parasite gene flow between different transmission foci, e.g. Kilifi and Rachuonyo South will become increasingly important in informing the mitigation measures needed to prevent importation of parasites as a result of human movement and migration. These analyses were not carried

out in the current study since the numbers of common SNPs between the two Kenyan sites was low, and we only had parasites from one timepoint in Rachuonyo South, hence we were unable to conduct an informative analysis of gene flow between sites”

Competing Interests: No competing interests were disclosed.

Referee Report 18 April 2017

<https://doi.org/10.21956/wellcomeopenres.11628.r21343>



Christopher Delgado-Ratto 

Global Health Institute, University of Antwerp, Antwerp, Belgium

This is a study that used SNP genotyping data finely analysed to describe the geographic structuring of *Plasmodium falciparum* parasites at micro-epidemiological level in three regions from Gambia and Kenya.

The authors were not able to compare the parasite populations among the study sites due to the samples were originally obtained for studies with different study designs (differences in sampling time, study population and design). The genetic diversity and clustering may not only be affected by geographic location and time but also by different ways of sampling the data. Say so, I appreciated that the authors focused in the population dynamics within the study sites.

Regarding the hypothesis that exists gene flow within the study sites, gene flow models could be also useful to prove such genetic exchange of parasites. There are various software that may help on this matter, i.e. Migrate-n.

Specific remarks:

Conclusions section:

- This paragraph is not fully justified on basis of the results: “following mass-treatment campaigns we would predict that if residual foci of transmission are retained this will rapidly lead to reinfection of the wider community, and that parasites acquiring mutations conferring drug resistance or immunological escape will be rapidly spread at a micro-epidemiological level.”

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Referee Expertise: Molecular epidemiology

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 29 Aug 2017

Irene Omedo, KEMRI- Wellcome Trust Research Programme, Kenya

We would like to thank the reviewer for his comments, which helped to improve the manuscript. we have included our responses below (in bold):

Regarding the hypothesis that exists gene flow within the study sites, gene flow models could be also useful to prove such genetic exchange of parasites. There are various software that may help on this matter, i.e. Migrate-n.

Migrate-n is a useful software for measuring migration rates between populations and is definitely useful for measuring gene flow within and between the different sites. It however, seems to need the parasites to be divided into populations and then to test for migration between these populations. Our current analyses aimed to step away from this requirement of specifying the populations a prior, and instead try to identify population structure based on clusters of genetically related parasites. The software also additionally requires the use of allele frequencies, while the current analyses focused on individual level genetic data. However, since such analyses may still be useful for identifying migration rates once distinct parasite populations have been identified, the following statement has been added in the discussion section.

“Finally, we used genetic data to show that there is high parasite movement and mixing within individual study sites. Additional analyses using gene flow models, e.g. as implemented in Migrate-N software, can be used to further validate our hypothesis of rapid gene flow and to confirm whether the parasites are part of a panmictic population or whether there exists underlying population structure, as well as to determine directionality of parasite movement between different populations, assuming that such populations can be identified within the region.”

This paragraph is not fully justified on basis of the results: “following mass-treatment campaigns we would predict that if residual foci of transmission are retained this will rapidly lead to reinfection of the wider community, and that parasites acquiring mutations conferring drug resistance or immunological escape will be rapidly spread at a micro-epidemiological level.”

The statement has been re-written as follows to clarify the message:

“On the other hand, based on the high level of parasite mixing observed at each study site, we would predict that ineffective application of control interventions such as mass drug administration that result in residual foci of transmission would lead to rapid re-infection of the wider community, and also that parasites acquiring mutations conferring drug resistance or immunological escape would spread rapidly at the micro-epidemiological level. This underscores the need for effective and sustained control until malaria elimination is achieved.”

Competing Interests: No competing interests were disclosed.

Referee Report 11 April 2017

<https://doi.org/10.21956/wellcomeopenres.11628.r21345>



Cristian Koepfli 

University of California, Irvine, Irvine, CA, USA

This is a relevant study, assessing the ability to identify small-scale foci of transmission and parasite gene flow to surrounding areas based in SNP-typing. While it is overall clearly presented and well written, more detail, in particular in the results section, would help to better understand the data, and to assess its power.

Specific comments:

Abstract:

1. Please state how many samples and SNP markers were included in the final analysis.
2. I wonder whether “relatedness in space and time” is the correct term, or “distance in space and time” would be more appropriate.

Results:

1. In the part on “Identification of geographical barriers to parasite movement” it would be useful to include the range of prevalence or MPF per pixel analyzed.
The second paragraph of this part is difficult to follow, as the term ‘cluster’ is used consistently, without further indication on what the clusters represent. It would help to include a sentence describing that spatial clusters were analyzed based on the PCA values of all isolates found within the cluster. Thus, clusters of isolates differing from all other isolates were identified. The same is the case in the discussion. What sizes were the clusters identified, and how many haplotypes were included per cluster?
2. Figure 5: Given that for almost every pair of samples the number of days differs, how were the days for the different curves calculated? I assume each color represents a range, yet only a point estimate is given.

Also, please indicate in brackets for each curve the number of samples included. For example, how many samples were available for the 1-day and 31-days analysis in The Gambia? Could the apparent reduction in SNP difference at 10 km be a chance finding due to limited sample size?

Including the number of SNPs analyzed in each population would further help to interpret the data. E.g. it is interesting that in Rachuonyo South the proportion of different SNPs is approx. twice as high as in the other sites, yet this is only evident when Figure 5 is compared to Table 1.

Would it be possible to include confidence intervals for the 1-day curves in the figure? This would help to understand the power of the data. For example, the statement in the abstract "Genetic relatedness of sample pairs is predicted by relatedness in space and time" suggests that genetic relatedness can be inferred, once the distance by space and time is known. This is however difficult to assess without more detail on the variance of the data.

3. In Table 2, what is the unit of the results showed? I assume it is SNP-difference/day (or SNP-difference/km), with days and distance log-transformed. Please state if/how data was transformed.

Discussion:

1. Paragraph 3 of the discussion could be expanded. At what spatial scales was population structure found in previous studies (as compared to the approx. 50 km range of the present study)? Have any of these studies included relatedness? This information would help to assess the feasibility to identify foci of higher transmission, and to estimate the level of gene flow to surrounding areas in different transmission settings.
2. The number of SNP differences plateaus at approx. 1 km in The Gambia, 3 km in Kilifi, and increases up to 10 km in Rachuonyo South. Are there possible explanations for these differences due to the characteristics of the local parasite populations?

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 29 Aug 2017

Irene Omedo, KEMRI- Wellcome Trust Research Programme, Kenya

We would like to sincerely thank the reviewer for taking time to review this work. We appreciate the comments and issues raised, and have addressed them on a point-by-point basis as indicated (in bold).

Abstract:

Please state how many samples and SNP markers were included in the final analysis.

The following statement has been inserted in the abstract to show the number of samples and SNPs used in the final analysis.

"Using 107, 177 and 82 SNPs that were successfully genotyped in 133, 1602, and 1034 parasite isolates from The Gambia, Kilifi and Rachuonyo South district, respectively, we show that there are no discrete geographically restricted parasite sub-populations."

I wonder whether “relatedness in space and time” is the correct term, or “distance in space and time” would be more appropriate.

“Relatedness in space and time” has been replaced with “distance in space and time”.

Results:

In the part on “Identification of geographical barriers to parasite movement” it would be useful to include the range of prevalence or MPF per pixel analyzed.

The second paragraph of this part is difficult to follow, as the term ‘cluster’ is used consistently, without further indication on what the clusters represent. It would help to include a sentence describing that spatial clusters were analyzed based on the PCA values of all isolates found within the cluster. Thus, clusters of isolates differing from all other isolates were identified. The same is the case in the discussion. What sizes were the clusters identified, and how many haplotypes were included per cluster?

The following statement has been added to show the MPF and parasite prevalence range analysed per pixel. The interquartile range of both measures at each pixel size has also been added.

“The range of MPF and parasite prevalence per pixel varied depending on the size of the pixels analysed. In the Kilifi population, MPF ranged from 0 – 100% (interquartile range (IQR) = 20%) for the 0.5km pixels; 0 – 100% (IQR = 14%) for the 1.0km pixels; 20 – 83% (IQR = 7%) for the 2km pixels; and 33 – 63% (IQR = 4.7%) for the 4km pixels. In the Rachuonyo South population, PCR positive prevalence varied from 0 – 75% (IQR = 19.4%) for the 0.5km pixels; 0 – 47% (IQR = 17.4%) for the 1.0km pixels; 3.5 – 35.8% (IQR = 14.3%) for the 2km pixels; and 6.2 – 33.4% (IQR = 7.8%) for the 4km pixels.”

The second paragraph under ‘spatial autocorrelation’ heading in the methods section has been expanded to indicate what clusters represent and how they are identified, and now reads as follows:

“Spatial scan statistics were calculated using SaTScan software and were run separately for each study site. The statistics involved running a purely spatial, retrospective analysis based on a normal probability distribution model using continuous variables (PC scores) and looking for areas with clusters of high PC scores. Latitude and longitude coordinates were used to represent the geographical locations of specific parasites, whereas principal component scores were used to represent individual parasite genotypes. During the analysis, a scanning window that gradually varies in size from including only a single homestead up to 50% of the population moves over the geographical space and at each window size and location, the ratio of parasites with high PCs inside the window versus outside the window is calculated. The window with the highest ratio is noted down as a cluster and its statistical significance is determined after accounting for multiple comparisons using random permutations.”

The number of samples contained in each significant cluster has been added in the results section and now reads as follows:

“In Kilifi, one cluster with a radius of 1.54 km ($p=0.01$) containing 15 parasites was detected, while in Rachuonyo South, a smaller cluster of genetically distinct parasites was detected with a radius of 0.5 km ($p=0.001$) containing 14 parasites”.

And the following section has been added in the discussion section to make it clear that the clusters were based on PC scores:

“However, using scan statistics we identified only two specific clusters of distinct parasite sub-populations based on PC scores, one in Kilifi and another in Rachuonyo South.”

Figure 5: Given that for almost every pair of samples the number of days differs, how were the days for the different curves calculated? I assume each color represents a range, yet only a point estimate is given.

The following statement has been added in the results section to show how the graphs were generated:

“Since the number of days differed for almost all parasite pairs, dummy data were included in the regression analysis to enable the generation of time-distance interaction graphs. For each study site, a distance range of 1 – 10km (with an interval of 0.1km between adjacent distances) was used. Temporal distance with 14 and 10 day intervals were assigned to parasite pairs in Kilifi and the Gambia, respectively, whereas time was not considered for the Rachuonyo South population. Constant SNP differences of 14, 10 and 8 were used for parasite pairs in Kilifi, Rachuonyo South and the Gambia, respectively.”

Also, please indicate in brackets for each curve the number of samples included. For example, how many samples were available for the 1-day and 31-days analysis in The Gambia? Could the apparent reduction in SNP difference at 10 km be a chance finding due to limited sample size?

The number of samples used to draw each curve has been noted in the Figure legend. We agree with the reviewer that the decrease past 10km is likely due to limited sample size and is there probably not significant.

Including the number of SNPs analyzed in each population would further help to interpret the data. E.g. it is interesting that in Rachuonyo South the proportion of different SNPs is approx. twice as high as in the other sites, yet this is only evident when Figure 5 is compared to Table 1.

The number of SNPs analysed in each population has been added to the figure legend.

Would it be possible to include confidence intervals for the 1-day curves in the figure? This would help to understand the power of the data. For example, the statement in the abstract “Genetic relatedness of sample pairs is predicted by relatedness in space and time” suggests that genetic relatedness can be inferred, once the distance by space and time is known. This is however

difficult to assess without more detail on the variance of the data.

Figure 5 has been regenerated to show the confidence intervals for the 1-day curves.

In Table 2, what is the unit of the results showed? I assume it is SNP-difference/day (or SNP-difference/km), with days and distance log-transformed. Please state if/how data was transformed.

The following statement has been added to Table 2 to show the units of measurement of the results.

“Values represent the change in the number of SNP differences between parasite pairs per day (time), per kilometre (distance) and per day/kilometre (time-distance interaction). Time, distance and the product of time and distance (time-distance interaction) were log transformed prior to running the regression analyses.”

Discussion:

Paragraph 3 of the discussion could be expanded. At what spatial scales was population structure found in previous studies (as compared to the approx. 50 km range of the present study)? Have any of these studies included relatedness? This information would help to assess the feasibility to identify foci of higher transmission, and to estimate the level of gene flow to surrounding areas in different transmission settings.

The paragraph has been expanded and now reads as follows:

“Lack of genetic structuring of parasite populations observed in this study is indicative of a population that is well mixed. This observation of a highly mixing parasite population is in agreement with results of similar studies of parasite genetic diversity and population differentiation using microsatellites ^{52, 54, 55}, immune selected genes ^{56, 57} and SNPs ⁵⁸. These studies were carried out in parasite populations from different geographical regions representing a diverse range of transmission intensities from highest in Africa and oceania, intermediate in Southeast Asia, and lowest in south America. However, other studies have shown population structure when looking at the same population ^{48, 59–61}, although these analyses were carried out on larger geographical scales than those analysed here and mostly involved analyses at provincial, country or continental levels. Population structure was most evident in regions with low transmission intensities such as south America or southeast Asia, and less evident in Africa where transmission intensity is much higher.”

The number of SNP differences plateaus at approx. 1 km in The Gambia, 3 km in Kilifi, and increases up to 10 km in Rachuonyo South. Are there possible explanations for these differences due to the characteristics of the local parasite populations?

Although the exact reason for the plateau observed is currently unknown, the following statement has been added to postulate possible reasons for the observation:

“The number of SNP differences were seen to plateau at approximately 1km in the

Gambia, 3km in Kilifi and 10km in Rachuonyo South. This may be attributed to the characteristics of the local parasite population, which in turn may be explained by the distribution of human settlement in the areas sampled, for example in the Gambia, homesteads tend to be clustered together in distinct, autonomous villages whereas in Rachuonyo South there is a denser and more uniform pattern of human settlement over the study area, enabling the interaction of parasites over a much larger distance.”

Competing Interests: No competing interests were disclosed.

Referee Report 29 March 2017

<https://doi.org/10.21956/wellcomeopenres.11628.r21178>



Michel Tibayrenc

Maladies Infectieuses et Vecteurs Ecologie, Génétique, Evolution et Contrôle (MIVEGEC), Institut de Recherche pour le Développement (IRD), Montpellier, France

This is a fine population genetic analysis of 3 samples taken in Gambia and Kenya, relying on the typing of 5199 samples by 276 SNPs. I have little to say about this work, which uses sound approaches and yields clear conclusions. A few remarks:

1. How can heterozygous genotypes be detected in haploid populations of the parasite?
2. As noted by the authors themselves, using 276 SNPs is rather limited. Genetic studies dealing with human populations at nowadays routinely rely on 500000 SNPs or more. One main feature of such studies is that microgeographical structures are detected mostly from low frequency variants and rare variants, which of course are undetectable when using a limited set of markers. Moreover, these low frequency and rare variants are supposed to be highly relevant for phenotypic expression, in particular disease susceptibility and are largely responsible for recent and localized evolution in human populations. (see for example Leslie et al. (2015)¹). It is most probable that these patterns exist in parasite populations too. The authors should discuss this point more, since it is probably one of the main avenues of future researches in microbiology.

References

1. Leslie S, Winney B, Hellenthal G, Davison D, Boumertit A, Day T, Hutnik K, Royrvik EC, Cunliffe B, Wellcome Trust Case Control Consortium 2, International Multiple Sclerosis Genetics Consortium, Lawson DJ, Falush D, Freeman C, Pirinen M, Myers S, Robinson M, Donnelly P, Bodmer W: The fine-scale genetic structure of the British population. *Nature*. 2015; **519** (7543): 309-14 [PubMed Abstract](#) | [Publisher Full Text](#)

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 29 Aug 2017

Irene Omedo, KEMRI- Wellcome Trust Research Programme, Kenya

We would like to sincerely thank the reviewer for taking time to review our work. We appreciate your comments, and have included a point-by-point response to them as follows (our responses are in bold).

How can heterozygous genotypes be detected in haploid populations of the parasite?

The use of the words “homozygous” and “heterozygous” in the context of haploid organisms has been clarified to mean single parasite genotype infections and mixed parasite genotype infections, respectively.

As noted by the authors themselves, using 276 SNPs is rather limited. Genetic studies dealing with human populations at nowadays routinely rely on 500000 SNPs or more. One main feature of such studies is that microgeographical structures are detected mostly from low frequency variants and rare variants, which of course are undetectable when using a limited set of markers. Moreover, these low frequency and rare variants are supposed to be highly relevant for phenotypic expression, in particular disease susceptibility and are largely responsible for recent and localized evolution in human populations. (see for example Leslie et al. (2015)¹). It is most probable that these patterns exist in parasite populations too. The authors should discuss this point more, since it is probably one of the main avenues of future researches in microbiology.

The following additional information has been added in the discussion section to address this point:

“Advances in sequencing technologies have increased the use of whole genome sequence data in the analysis of *P. falciparum* parasite population genetics, and this has led to the identification of hundreds of thousands of SNPs, most of which are present at very low frequencies especially in African parasite populations³. Additional analyses will require the use of whole genome sequence data to identify rare variants and distinguish between closely related parasites, thus allowing parasite population structure to be analysed at fine spatial scales.”

Competing Interests: No competing interests were disclosed.