



Published in final edited form as:

*Soc Sci Med.* 2015 May ; 133: 233–241. doi:10.1016/j.socscimed.2014.10.037.

## The geography of malaria genetics in the Democratic Republic of Congo: A complex and fragmented landscape

Margaret Carrel<sup>a,\*</sup>, Jaymin Patel<sup>b,1</sup>, Steve M. Taylor<sup>b,c,d</sup>, Mark Janko<sup>e</sup>, Melchior Kashamuka Mwandagaliwa<sup>b</sup>, Antoinette K. Tshefu<sup>f</sup>, Ananias A. Escalante<sup>g</sup>, Andrea McCollum<sup>h</sup>, Md Tauqeer Alam<sup>h</sup>, Venkatachalam Udhayakumar<sup>h</sup>, Steven Meshnick<sup>b</sup>, and Michael Emch<sup>e</sup>

<sup>a</sup>Department of Geographical & Sustainability Sciences, University of Iowa, Iowa City, IA, USA

<sup>b</sup>Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina- Chapel Hill Chapel Hill, NC, USA

<sup>c</sup>Division of Infectious Diseases and International Health, Duke University Medical Center, Durham, NC, USA

<sup>d</sup>Duke Global Health Institute, Durham, NC, USA

<sup>e</sup>Department of Geography, University of North Carolina-Chapel Hill, Chapel Hill, NC, USA

<sup>f</sup>Ecole de Sante Publique, Faculte de Medecine, University of Kinshasa, Kinshasa, Democratic Republic of the Congo

<sup>g</sup>Center for Evolutionary Medicine & Informatics, The Biodesign Institute, Arizona State University, Tempe, AZ, USA

<sup>h</sup>Malaria Branch, Division of Parasitic Diseases and Malaria, Center for Global Health, Centers for Disease Control and Prevention, Atlanta, GA, USA

### Abstract

Understanding how malaria parasites move between populations is important, particularly given the potential for malaria to be reintroduced into areas where it was previously eliminated. We examine the distribution of malaria genetics across seven sites within the Democratic Republic of Congo (DRC) and two nearby countries, Ghana and Kenya, in order to understand how the relatedness of malaria parasites varies across space, and whether there are barriers to the flow of malaria parasites within the DRC or across borders. Parasite DNA was retrieved from dried blood spots from 7 Demographic and Health Survey sample clusters in the DRC. Malaria genetic characteristics of parasites from Ghana and Kenya were also obtained. For each of 9 geographic sites (7 DRC, 1 Ghana and 1 Kenya), a pair-wise  $R_{ST}$  statistic was calculated, indicating the genetic distance between malaria parasites found in those locations. Mapping genetics across the spatial extent of the study area indicates a complex genetic landscape, where relatedness between two proximal sites may be relatively high ( $R_{ST} > 0.64$ ) or low ( $R_{ST} < 0.05$ ), and where distal sites

also exhibit both high and low genetic similarity. Mantel's tests suggest that malaria genetics differ as geographic distances increase. Principal Coordinate Analysis suggests that genetically related samples are not co-located. Barrier analysis reveals no significant barriers to gene flow between locations. Malaria genetics in the DRC have a complex and fragmented landscape. Limited exchange of genes across space is reflected in greater genetic distance between malaria parasites isolated at greater geographic distances. There is, however, evidence for close genetic ties between distally located sample locations, indicating that movement of malaria parasites and flow of genes is being driven by factors other than distance decay. This research demonstrates the contributions that spatial disease ecology and landscape genetics can make to understanding the evolutionary dynamics of infectious diseases.

## Keywords

Democratic Republic of Congo; Disease ecology; Landscape genetics; Malaria

## 1. Introduction

Malaria remains endemic in 104 countries, despite rollback and eradication efforts, and caused an estimated 207 million cases and 627,000 deaths in 2012, primarily in Africa (World Health Organization, 2013). The Democratic Republic of Congo (DRC) in particular suffers one of the highest burdens of malaria in sub-Saharan Africa: health surveys indicate that over a third of adults in the DRC have malaria parasites in their blood (Hay et al., 2010; Taylor et al., 2011). *Plasmodium falciparum*, one species of malaria parasite, causes nearly all infections within the DRC (Taylor et al., 2011). The success of malaria eradication efforts depends not only on local anti-malaria campaigns, but also on preventing reintroduction of malaria parasites from poorly-controlled areas into places newly free of the disease (Le Menach et al., 2011; Moonen et al., 2010). Given the potential for malaria reintroduction to undo progress in control efforts, an understanding of how malaria diffuses across a landscape, and how this diffusion can be limited or stopped, is essential. As yet, however, we have limited insight into how malaria spreads geographically (Lynch and Roper, 2011).

A landscape genetics approach holds promise for filling in this key knowledge gap. Indeed, because of limited movement of mosquitoes (a maximum of 10 km), the most likely means by which malaria spreads is through the movement of people carrying the parasites in their blood (Kaufmann and Briegel, 2004). Landscape genetics allows us to observe parasites in different locations and infer population movement from the distribution of the parasite's genetic markers, while also measuring geographic and landscape factors driving that distribution, either by promoting or by preventing dispersal (Biek and Real, 2010; Holderegger and Wagner, 2006; Manel et al., 2003; Storfer et al., 2010; Storfer et al., 2007). The flow of malaria genes across landscapes has been explored previously and such work has indicate a geographic structure to malaria genetics, that parasite populations can be fragmented depending on the connectivity (i.e. coastal or inland) of locations, that border regions can act as spaces of origin of drug resistance and that population movement threatens success of local eradication programs by re-introduction of malaria (Alam et al.,

2011b; Griffing et al., 2011; Rebaudet et al., 2010; Schultz et al., 2010). Additionally, molecular epidemiology has been used to conclusively determine the source of a malaria outbreak among UN peacekeeping soldiers from Guatemala returning from the DRC, emphasizing the need for molecular surveillance as an integral part of malaria control and elimination program (Patel et al., 2014). As yet, however, few studies have focused on malaria genetics within DRC or connections to other sub-Saharan African countries, which, given the high burden of malaria in DRC and the focus of the global health community on malaria eradication in this region, indicates a need for greater research in this geographic and topical area. Medical geography in particular, with its emphasis on disease ecology and spatial statistics, can make valuable contributions to understanding how infectious diseases of humans are transmitted across spaces and places (Carrel and Emch, 2013).

Using spatial methods to explore the genetic characteristics of malaria parasites found in humans across seven sites in the DRC, and a site each in Ghana and Kenya, we sought to understand whether genetic variation between sites is correlated with geographic space between those sites, how scaling up or down from national to sub-national scale changes the observed relationship between genetics and geography, and whether significant barriers to malaria gene flow exist. Given the complex ecology of malaria, dependent on mosquito vectors from the *Anopheles* genus, and the mobility of infected and susceptible populations across the DRC, we did not anticipate finding a strong correlation between genetic variation and geographic distances, as has been found for other infectious diseases such as H5N1 influenza and rabies (Carrel et al., 2010; Real et al., 2005). Instead, we hypothesized that a more fragmented genetic landscape would be found, reflecting variations in land cover, transportation routes, population movement, and conflict zones that are seen across the DRC.

## 2. Data and methods

In 2007 a Demographic and Health Survey (DHS) was completed on a spatially and statistically representative sample of individuals within the DRC, grouped within household clusters across the country. Survey participants had blood spots taken, from which we were able to retroactively detect malaria parasites via real-time polymerase chain reaction (real-time PCR), as has been previously described (Taylor et al., 2011). This research received institutional review board approval at the Kinshasa School of Public Health, the University of North Carolina–Chapel Hill and the University of Iowa.

To explore how genetic characteristics of malaria parasites vary across the DRC, we selected seven DHS clusters, each with ten or more asymptomatic *P. falciparum*-positive individuals, for analysis. The clusters were selected based on the hypothesis that the Congo River, a principal route of human transportation, would facilitate diffusion of *plasmodium* parasites. Three clusters (81, 88 and 183) were chosen because they were located on or near the Congo River, a principal route of human transportation (Herderschee et al., 2011); two clusters (164 and 211) were chosen because they were not on the river but were approximately the same distance apart from the three river sites as the river sites were from one another; and two final clusters (29 and 203) were chosen because they were far away (> 1000 km on average) from the other clusters (Fig.1). The median distance between clusters

was 709 km (IQR = 883 km). This geographic diversity allowed us to explore whether sites linked via the river, sites the same distance from one another but not river linked and sites geographically distant had differing patterns of genetic relatedness.

These seven clusters contained 82 individuals infected with *P. falciparum*. In addition to the samples from the DRC, 72 and 62 infected individuals previously reported from Ghana and Kenya respectively were included for comparative analyses of parasite microsatellites across political borders (Alam et al., 2011a; McCollum et al., 2012).

Five neutral, meaning not under selective pressure, microsatellite markers on two different chromosomes (C2M33, C2M34, C2M29, C2M27 on chromosome 2; and C3M40 on chromosome 3) were chosen to assess the population structure of parasites. Microsatellites are short repeating sequences of DNA base pairs, typically with high heterozygosity, that can be used to determine genetic relatedness of individuals. All specimens were amplified either through a single-round or hemi-nested Polymerase Chain Reaction as described previously; (Anderson et al., 2000, 1999; Roper et al., 2003). PCR products were then separated by capillary electrophoresis using an ABI 3130xl genetic analyzer. The alleles were scored using GeneMapper software, version 3.7 (Applied Biosystems, Foster City, CA). Alleles were binned, depending on the size of the repeat unit, to the nearest 2 or 3 nucleotides in length.

Since there is a possibility that individuals could be infected with multiple variants of *P. falciparum*, it would be difficult to accurately assign microsatellite allele sizes to particular variants within each individual using capillary electrophoresis. However, because the genetic relatedness index above is calculated at microsatellite marker, the linkage between each marker (i.e. that they were inherited together as an intact unit) is not necessary to ascertain. Therefore, we included all individuals even if they were infected by mixtures of *P. falciparum* variants, as evidenced by mixtures at one or more microsatellite markers. Because the data format for GenAlEx and SPAGeDi requires that each individual have unique microsatellite marker profile, we created “virtual” unique variants for individuals that were infected with multiple variants to ensure we include all possible variants in circulation in our study populations. If two or more alleles were detected on one microsatellite marker, then distinct variants were created that differed from each other only at the marker with multiple alleles. If multiple microsatellite markers contained two alleles, then two distinct variants were created where alleles were randomly assorted for each marker. The same was done if multiple microsatellite markers contained three alleles. If one or more marker had two alleles and one or more marker had three alleles, then two full variants were created using random assortment as above along with a third variant with missing data on marker with two alleles.

## 2.1. Measuring genetic relatedness

We quantified  $\beta$  diversity (or the degree of genetic differentiation between populations) using Slatkin's  $R_{ST}$ . This measure assumes that microsatellites evolve according to the stepwise mutation model (SMM) in which novel alleles are created either by deletion or addition of a single repeated unit of microsatellite with equal probability  $\mu/2$  in both directions (Balloux and Lugon-Moulin, 2002). The  $R_{ST}$  also incorporates the relatedness

between allele sizes into its calculation of relatedness between populations (Anderson et al., 2000; Balloux and Lugon-Moulin, 2002; Holsinger and Weir, 2009).  $R_{ST}$  was calculated in SPAGeDi v1.3 (Hardy and Vekemans, 2002) using the following formula:

$$R_{ST} = \frac{(S - S_w)}{S}$$

where  $S$  is the average squared difference in allele size between all pairs of alleles and  $S_w$  is the average sum of squares of the differences in allele size within each subpopulation. Pairwise  $R_{ST}$  comparisons were calculated using a nested ANOVA approach. Pairwise  $R_{ST}$  genetic distance matrices were created for each of the five microsatellite markers, and one overall  $R_{ST}$  distance matrix was generated, taking into account genetic variation on each marker.

## 2.2. Data analysis

A geographic information systems (GIS) database was built that included DHS cluster locations (offset randomly by the DHS by up to 5 km in order to maintain confidentiality of survey respondents) and was associated with the  $R_{ST}$  distance measures to visualize how genetic relatedness varies across space.

A pair-wise geographic distance matrix was generated for the 7 DRC and Ghana and Kenya sites. First, scatterplots were used to explore how genetic distance between sites varies with the geographic distance between sites. Given the non-independence of the distance matrices, where measures of genetic relatedness or spatial proximity are pair-wise, we then used Mantel tests to quantify the correlation between matrices (Lichstein, 2007; Mantel, 1967). These tests indicate how genetic relatedness increases or decreases as geographic space between locations varies. A Mantel test is a global measure, indicating overall correlation between matrices. Mantel correlograms, in contrast, indicate how this relationship varies across different spatial bands, a local measure of the positive or negative correlation between matrices across various thresholds. Mantel tests and Mantel correlograms were generated for the  $R_{ST}$  and geographic distance matrices measured at 9 sites (7 in DRC, Ghana, Kenya).

As changing the scale of analysis can change the way we understand a phenomenon, we then aggregated the 7 DRC sites into one measure of genetic variability and re-ran the Mantel tests and Mantel correlograms at the national rather than the subnational scale. Doing so allowed us to assess whether the relationships observed between geography and genetics were a function of scale and aggregation, or whether they persist across scales.

Principal Coordinate Analysis (PCoA) was employed to quantify the variation between the different parasite populations using pairwise  $R_{ST}$  comparisons. PCoA plots illustrating the genetic relatedness between the DRC clusters, Ghana and Kenya were generated using GenAlEx v.6.4 (Peakall and Smouse, 2006). By arranging the sites according to genetics rather than geography we were able to understand whether genetically related sites were also those that were geographically proximate.

Identifying the presence of barriers to gene flow between sample locations can indicate if there are breaks in landscape or population mobility that act as stops on the transmission of malaria parasites. Barriers are located in spatial zones of rapid genetic change, and in the case of malaria indicated either low human or mosquito vector movement. Monmonier's algorithm uses maximum-difference methods to determine genetically distinct malaria groups in space where samples on either side of the barrier are more similar to each other than to samples on the other side of the barrier (Manni and Guerard, 2004; Monmonier, 1973; R Development Core Team, 2011). In Monmonier's algorithm, a spatial network is created among the points by means of Voronoi tessellations and Delaunay triangulations. Following creation of the network, the genetic distances between sample locations are associated with the triangulation between those points. Using the method advocated by Monmonier, the exterior network connection with the highest genetic distance is taken as the starting point of the barrier. The barrier is drawn perpendicular to the line connecting the two sample points. From there, the barrier continues moving perpendicular to the next adjacent network connection with the highest value, until either the bounding edge of the network is reached or the barrier has formed a loop by closing on itself. Subsequent barriers start from exterior connections with the highest associated values, and continue until they meet either another barrier or the exterior of the network. Random permutations of the distance matrices and re-drawing of the barriers allows for statistical significance to be assessed.  $R_{ST}$  matrices were randomly permuted 999 times using the *sna* package in R. Barriers were determined using Barrier v2.2 (Manni and Guerard, 2004).

### 3. Results

Parasites from the DRC were genotyped at eight neutral microsatellite markers from 82 parasitemic subjects tested via PCR in seven survey clusters (Fig. 1) either on the Congo River (81, 88, 183), near the Congo River (164, 211) or a substantial distance away (29, 203). Eight subjects were excluded from further analysis, as we were unable to amplify the majority of the microsatellites. More than half of the remaining 74 subjects (52.7%,  $n = 39$ ) were infected by single genetic variants with no mixtures at any locus. 36.5% ( $n = 27$ ) of the subjects had 2 alleles at two or more loci and 10.8% ( $n = 8$ ) of the subjects contained at least one locus with 3 alleles. Among the Kenyan subjects, 33.9% ( $n = 21$ ) were infected by single variants with no mixtures at any locus. 48.4% ( $n = 30$ ) of the subjects had 2 alleles at two or more loci and 17.7% ( $n = 11$ ) of the subjects displayed at least one locus with 3 alleles (McCollum et al., 2012). All Ghanaian parasitemias were selected for a previous study of haplotypes, so only infections containing single variants were genotyped (Alam et al., 2011a).

Pairwise  $R_{ST}$  values measuring genetic relatedness among DRC cluster sites and Ghana and Kenya ranged from below zero (signifying virtual identity) to above zero (indicating increasing divergence). Within the DRC, pairwise  $R_{ST}$  for both proximal and distal sites were high ( $88/183 = 0.1057$ ,  $29/164 = 0.1196$ ) and low ( $88/164 = 0.008$ ,  $29/211 = 0.0381$ ). Values between DRC clusters and Ghanaian and Kenyan parasites exhibited similar ranges of  $R_{ST}$ , some very high ( $164/\text{Ghana} = 0.1774$ ) and some indicating nearly identical genetics ( $164/\text{Kenya} = 0.0055$ ) (Table 1). Connections between sample locations with nearly identical malarial parasite genetics are seen in Fig. 2B.



When the 7 DRC clusters were treated as a single observation, the lowest  $R_{ST}$  was between the DRC and Kenyan parasites ( $R_{ST} = 0.0494$ ), then between DRC and Ghanaian parasites ( $R_{ST} = 0.0864$ ), and the highest degree of genetic difference was observed between Kenyan and Ghanaian parasites ( $R_{ST} = 0.2155$ ) (Fig. 2A). Plotting the relationship between pairwise  $R_{ST}$  and geographic distance appear to indicate a relationship between increased geographic distance and increased genetic distance, although only when the DRC was considered as a single observation (Fig. 3).

While scatter plots suggest an isolation-by-distance phenomenon, with increased genetic change associated with increased geographic space between points, the non-independence of pairwise measures necessitates a more formal statistical analysis. Results of the global Mantel tests for each of the 5 markers and an overall  $R_{ST}$  measure, at both the subnational (7 DRC clusters) and national (1 DRC measure) scales indicate positive but not statistically significant correlation between genetic and geographic distance (Table 2).

Mantel correlograms, representing genetic similarity across geographic distances, indicate that the five microsatellites have varying relationships across geographic space. Points above the zero line indicate similarity (i.e. parasites that are more alike genetically), while points below the zero line indicate dissimilarity. Microsatellites C2M33, C2M34 and the overall measure exhibit a more classically isolation-by-distance pattern, where sites that are geographically proximate have higher similarity, and points fall below the zero line as geographic distance increases (Fig. 4).

Correlograms of 1 DRC measure versus Ghanaian and Kenyan parasites in Fig. 5 exhibit an unusual relationship, with highest genetic similarity at the middle distance (i.e. DRC and Ghana), and a Mantel  $r$  of zero at the low and high geographic distances (i.e. DRC and Kenya, Kenya and Ghana), in contrast to the scatterplot in Fig. 3. Marker C2M29 is an exception to this pattern, however, with high dissimilarity at the middle geographic distance.

The results of the PCoA further reflect the complexity of the relationships between genetics and geography among DRC malarial parasites, in that sites that are proximal geographically, connected by the Congo River (i.e. 183 and 88) are located in opposite spaces in the ordination (Fig. 6). These results do not appear to support the hypothesis that parasite flow is facilitated by the Congo River. Ghanaian parasites are located closest to those in clusters 29 and 203, found in southern and western DRC, while Kenyan parasites in the PCoA are located closest to DRC clusters the north.

Barrier analysis on individual microsatellite matrices and an overall  $R_{ST}$  matrix indicated that no statistically significant barriers to gene flow exist in the dataset. While a large number of potential barriers were returned (Fig. 7), some dividing clusters within the DRC into separate areas and some following political boundaries and separating the DRC from Ghana and Kenya, none were detected in the randomly permuted matrices enough times to achieve significance.

## 4. Discussion

The prevalence of malaria varies geographically, even within a highly malarious country such as the DRC (Taylor et al., 2011). In this study, we show that there is a high level of complexity in the relationships between parasite populations across space. Using the genetic relatedness as a proxy for intermingling of parasite populations, the principle of isolation by distance does appear to hold true for *P. falciparum* at a country level as indicated by the relationships between the DRC, Ghana, and Kenya, even though the Mantel tests did not show a statistically significant relationship. Indeed, the non-linearity of the relationship between genetic and geographic distance observed in the Mantel correlogram, with greater genetic similarity between DRC and Ghanaian samples than between DRC and Kenyan or Kenyan and Ghanaian, suggests that there is potentially greater “friction” to genetic exchange associated with some geographies. Previous research on road connectivity within Sub-Saharan Africa has indicated a relative disconnect between the DRC and cities of East Africa (Gray et al., 2009).

The general isolation by distance relationship observed nationally does not hold true, however, when looking at within-DRC malarial genetic relatedness. Some DRC clusters are closely related to a West African (Ghanaian) site, while others are more closely related to an East African site (Kenya), suggesting that the DRC constitutes a watershed for parasite populations across sub-Saharan Africa. Additionally, when trying to study parasite population structure in a high transmission area, neutral microsatellite markers might not provide the level of resolution needed to draw precise conclusions.

Geographic proximity, or distance, were not the sole determinants of genetic similarity or dissimilarity with the DRC. Additionally, it does not appear that the movement of human hosts along the Congo River increases the interchange of parasite populations. Two clusters which are near each other on the river (88 and 183) were not closely related genetically. Indeed, cluster 183 was genetically anomalous: parasites from this cluster were not genetically related to those from nearby clusters, but were closely related to those from two distant clusters (29 and 203) and to parasites from Kenya. Importantly, cluster 183 was the only cluster in Orientale province. In a separate study of drug resistance haplotypes, parasites from Eastern DRC – including Orientale province – clustered into a population genetically-distinct from those in the balance of the country (Taylor et al., 2013). The reasons for this divergence of parasite populations is unclear, but this province may have been significantly affected by a civil war which lasted until 2006 (Raleigh et al., 2010), particularly, the Tshopo district where cluster 183 is located (Bakari, 2011; United Nations High Commissioner for Human Rights, 2010). This would have possibly limited trade and movement of humans to other parts of the country including the clusters in our study. Additionally, it is possible that individuals in cluster 183, if they were to travel, would go towards Kisangani, a major urban center (Fig. 1), as opposed to traveling along the river towards Kinshasa. Because we did not include data from Kisangani in the analysis, we cannot test this hypothesis in this paper. The DHS also has data on length of residence. In cluster 183, over 97% of survey respondents reported always living in their current village, while the remaining 3% lived in the same village for 12 years. In cluster 88, 87% of respondents reported always living in the same village, while the remaining 13% had lived



there for 9 years or more. As such, this lack of population migration, the potential for short-term movement in different directions (88 towards Kinshasa, 183 towards Kisangani), and potentially less travel along the river due to safety concerns, may explain the genetic unrelatedness between survey clusters.

Other factors that may be involved in gene flow include environmental factors such as land cover change, particularly deforestation, which has been identified as an important mechanism in increased malaria transmission in Latin America and parts of Africa (Guerra et al., 2006). Deforestation rates in DRC have remained low, however, owing to the conflict in the country and lack of international investment, driven by small-scale farming over the long term rather than commercial logging operations. Additionally, long term exposure to deforestation is hypothesized to have little to no effect on transmission, as people begin to adopt protective behavior such as using a bed net (de Castro et al., 2006). Unfortunately, no large-scale study of deforestation in Sub-Saharan Africa has been published, and as such it is difficult to assess the degree to which it has an effect.

This study has several limitations. For the purposes of this study, we selected samples from only seven out of 300 DHS clusters to be characterized on five neutral microsatellites. Selecting more clusters along with additional unlinked microsatellites would have given our results more precision. While statistical significance was not frequently achieved in the Mantel tests or correlograms, potentially due to small sample sizes, the results indicate that there is need for larger studies to examine malaria genetic exchange within and between countries. In particular, incorporating data on human movement between clusters and environmental variation would potentially explain why parasites from geographically close clusters were genetically unrelated. We included only monoinfections from Ghana and this could have underestimated the genetic relationships with other parasite populations. However, we were limited to this as the samples were collected and analyzed as part of a previous study (Alam et al., 2011a).

Malaria genetics in the DRC have a complex and fragmented landscape. Diffusion of malaria parasites is not restricted by political borders, as has been shown for other pathogens such as influenza viruses and as could be hypothesized given that the majority of malaria treatment or eradication programs are self-contained within countries and are not coordinated across border regions (Wallace and Fitch, 2008). This is especially true in the DRC, where malaria control efforts exist in only a few provinces. Additionally, given the unstable socio-political context and conflict prone areas of the DRC and its neighboring countries, movement of people across borders is a frequent event, complicating efforts at malaria control and also shifting patterns of genetic relatedness of parasites.

Malaria control and elimination efforts have made great progress in the past decade (Feachem et al., 2010a; Greenwood, 2009; Mendis et al., 2009). Now, 32 countries have or are planning malaria elimination programs (Feachem et al., 2010b). But previous malaria elimination efforts failed; new programs need to avoid the mistakes that led to these failures (Greenwood, 2009). Our findings about the relatedness of malaria parasites within the DRC and between DRC and Ghana and Kenya indicate one mistake often made in malaria control and elimination: considering countries as isolated entities rather than highly connected units

that need to be addressed simultaneously. The DRC is a large country, which bridges West and East Africa (it is French-speaking in the West and Swahili-speaking in the East). Its geography ranges from rainforest to savannah to mountains, with large urban and rural populations. It borders 9 other African countries. Accordingly, our findings on the spatial epidemiology of malaria in the DRC are likely to be generalizable to many other African countries.

Evidence of close genetic ties between distally located sample locations indicates that the movement of malaria parasites and the flow of genes is being driven by factors other than distance decay. That is, the movement of people across the landscape is helping malaria parasites to skip over geographic space and exchange genetic material with parasites located at great distances. This research demonstrates the contributions that medical geographers and spatial epidemiologists can make to understanding the evolutionary dynamics of infectious diseases. Integrated studies of genetic and spatial relationships may identify modifiable drivers or barriers to malaria parasite diffusion that can help guide malaria control efforts.

## Acknowledgments

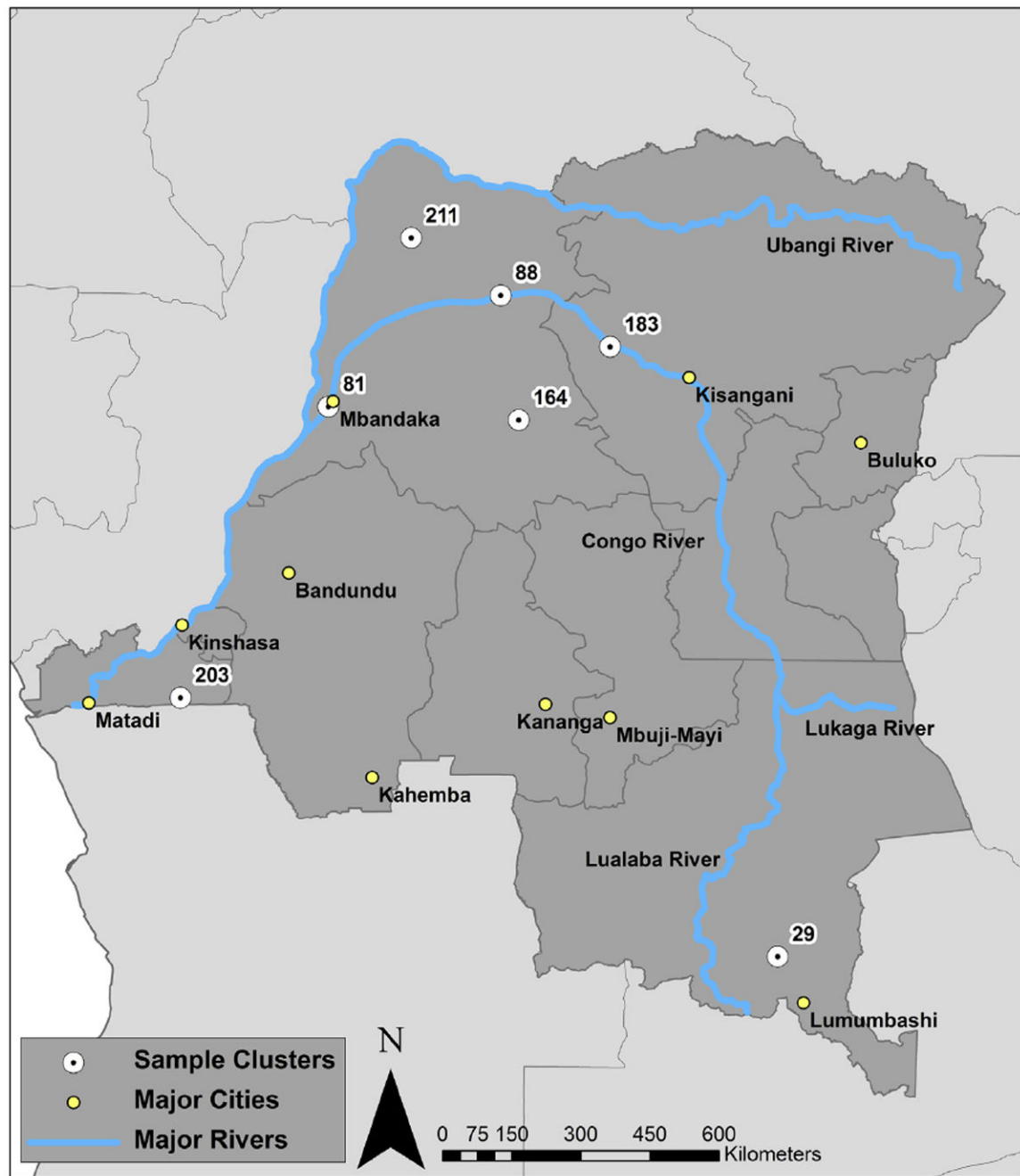
This project was supported by grants from the NSF, the Gillings Innovative Laboratory Fund and the NIH. MC and ME are supported by NSF under award number BCS-1339949. JP and SM are supported by NIH award number 1R56AI097609-01. SMT is supported by the National Institute of Allergy and Infectious Diseases of the NIH under award number K08AI100924. We thank Dr. Augustin Okenge (Programme National de Lutte Contre le SIDA, Kinshasa, DRC), Dr. Jeremie Mwonga (Laboratoire National de Reference SIDA et IST, Kinshasa, DRC), and Ann Way, Mohamed Ayad, and Martin Vaessen (all of Measure DHS, Calverton, MD) for help in obtaining access to the dried blood spots.

## References

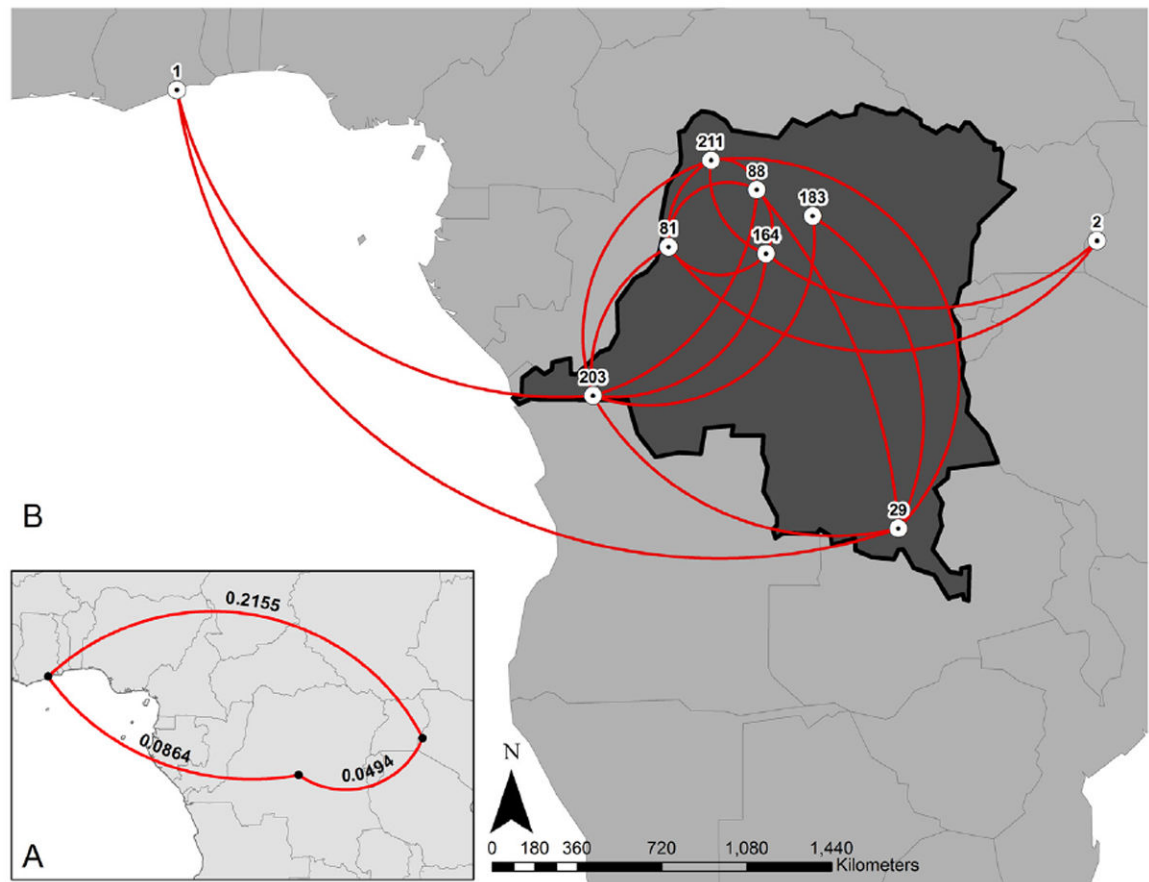
- Alam MT, de Souza DK, Vinayak S, Griffi M, Poe AC, Duah NO, Wilson MD, et al. Selective sweeps and genetic lineages of plasmodium falciparum drug-resistant alleles in Ghana. *J Infect Dis.* 2011a; 203(2):220–227. [PubMed: 21288822]
- Alam MT, Vinayak S, Congpuong K, Wongsrichanalai C, Satimai W, Slutsker L, Udhayakumar V, et al. Tracking origins and spread of sulfadoxine-resistant plasmodium falciparum dhps alleles in Thailand. *Antimicrob Agents Chemother.* 2011b; 55(1):155–164. [PubMed: 20956597]
- Anderson TJ, Haubold B, Williams JT, Estrada-Franco JG, Richardson L, Mollinedo R, French N. Microsatellite markers reveal a spectrum of population structures in the malaria parasite plasmodium falciparum. *Mol Biol Evol.* 2000; 17(10):1467–1482. [PubMed: 11018154]
- Anderson TJ, Su XZ, Bockarie M, Lagog M, Day KP. Twelve microsatellite markers for characterization of Plasmodium falciparum from finger-prick blood samples. *Parasitology.* 1999; 119(2):113–125. [PubMed: 10466118]
- Balloux F, Lugon Moulin N. The estimation of population differentiation with microsatellite markers. *Mol Ecol.* 2002; 11(2):155–165. [PubMed: 11856418]
- Biek R, Real LA. The landscape genetics of infectious disease emergence and spread. *Mol Ecol.* 2010; 19(17):3515. [PubMed: 20618897]
- Carrel M, Emch ME. Genetics: a new landscape for medical geography. *Ann Assoc Am Geogr.* 2013; 103(6):1452–1467. [PubMed: 24558292]
- Carrel MA, Emch M, Jobe RT, Moody A, Wan XF. Spatiotemporal structure of molecular evolution of H5N1 highly pathogenic avian influenza viruses in Vietnam. *PLoS One.* 2010; 5(1):e8631. <http://dx.doi.org/10.1371/journal.pone.0008631>. [PubMed: 20072619]
- de Castro MC, Monte-Mor RL, Sawyer DO, Singer BH. Malaria risk on the Amazon frontier. *Proc Natl Acad Sci U S A.* 2006; 103(7):2452–2457. [PubMed: 16461902]

- Feachem RG, Phillips AA, Hwang J, Cotter C, Wielgosz B, Greenwood BM, Ghebreyesus TA, et al. Shrinking the malaria map: progress and prospects. *Lancet*. 2010a; 376(9752):1566–1578. [PubMed: 21035842]
- Feachem RG, Phillips AA, Targett GA, Snow RW. Call to action: priorities for malaria elimination. *Lancet*. 2010b; 376(9752):1517–1521. [http://dx.doi.org/10.1016/S0140-6736\(10\)61500-0](http://dx.doi.org/10.1016/S0140-6736(10)61500-0). [PubMed: 21035844]
- Gray RR, Tatem AJ, Lamers S, Hou W, Laeyendecker O, Serwadda D, Salemi M, et al. Spatial phylodynamics of HIV-1 epidemic emergence in east Africa. *AIDS Lond Engl*. 2009; 23(14):F9–F17. <http://dx.doi.org/10.1097/QAD.0-b013e32832faf61>.
- Greenwood B. Can malaria be eliminated. *Trans R Soc Trop Med Hyg*. 2009; 103(Suppl. 1):S2–S5. <http://dx.doi.org/10.1016/j.trstmh.2008.10.027>. [PubMed: 19062058]
- Griffing SM, Mixson-Hayden T, Sridaran S, Alam MT, McCollum AM, Cabezas C, Lucas C. South American plasmodium falciparum after the malaria eradication era: clonal population expansion and survival of the fittest hybrids. *PloS One*. 2011; 6(9):e23486. [PubMed: 21949680]
- Guerra CA, Snow RW, Hay SI. A global assessment of closed forests, deforestation and malaria risk. *Ann Trop Med Parasitol*. 2006; 100(3):189–204. <http://dx.doi.org/10.1179/136485906X91512>. [PubMed: 16630376]
- Hardy OJ, Vekemans X. SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol Ecol Notes*. 2002; 2(4):618–620.
- Hay SI, Okiro EA, Gething PW, Patil AP, Tatem AJ, Guerra CA, Snow RW. Estimating the global clinical burden of plasmodium falciparum malaria in 2007. *PLoS Med*. 2010; 7(6):e1000290. [PubMed: 20563310]
- Herderschee J, Kaiser K, Samba DM. Resilience of an African Giant: Boosting Growth and Development in the Democratic Republic of Congo. World Bank-free PDF. 2011
- Holderegger R, Wagner HH. A brief guide to landscape genetics. *Landsc Ecol*. 2006; 21(6):793–796.
- Holsinger KE, Weir BS. Genetics in geographically structured populations: defining, estimating and interpreting FST. *Nat Rev Genet*. 2009; 10(9):639–650. [PubMed: 19687804]
- Kaufmann C, Briegel H. Flight performance of the malaria vectors *Anopheles gambiae* and *Anopheles atroparvus*. *J Vector Ecol*. 2004; 29:140–153. [PubMed: 15266751]
- Le Menach A, Tatem AJ, Cohen JM, Hay SI, Randell H, Patil AP, Smith DL. Travel risk, malaria importation and malaria transmission in Zanzibar. *Sci Reports*. 2011; 1:93.
- Lichstein J. Multiple regression on distance matrices: a multivariate spatial analysis tool. *Plant Ecol*. 2007; 188(2):117–131.
- Lynch C, Roper C. The transit phase of migration: circulation of malaria and its multidrug-resistant forms in Africa. *PLoS Med*. 2011; 8(5):e1001040. [PubMed: 21655316]
- Manel S, Schwartz MK, Luikart G, Taberlet P. Landscape genetics: combining landscape ecology and population genetics. *Trends Ecol Evol*. 2003; 18(4):189–197. [http://dx.doi.org/10.1016/S0169-5347\(03\)00008-9](http://dx.doi.org/10.1016/S0169-5347(03)00008-9).
- Manni F, Guerard E. Barrier v.2.2. 2004
- Mantel N. The detection of disease clustering and a generalized regression approach. *Cancer Res*. 1967; 27(2 Part 1):209–220. [PubMed: 6018555]
- McCollum AM, Schneider KA, Griffing SM, Zhou Z, Kariuki S, Ter-Kuile F, Udhayakumar V, et al. Differences in selective pressure on dhps and dhfr drug resistant mutations in western Kenya. *Malar J*. 2012; 11:77. [PubMed: 22439637]
- Mendis K, Rietveld A, Warsame M, Bosman A, Greenwood B, Wernsdorfer WH. From malaria control to eradication: the WHO perspective. *Trop Med Int Health*. 2009; 14(7):802–809. [PubMed: 19497083]
- Monmonier MS. Maximum-difference barriers: an alternative numerical realization method. *Geogr Anal*. 1973; 5(3):245–261.
- Moonen B, Cohen JM, Snow RW, Slutsker L, Drakeley C, Smith DL, Tanner M, et al. Operational strategies to achieve and maintain malaria elimination. *Lancet*. 2010; 376(9752):1592–1603. [PubMed: 21035841]

- Patel JC, Taylor SM, Juliao PC, Parobek CM, Janko M, Gonzalez LD, Meshnick SR, et al. Genetic evidence of importation of drug-resistant *Plasmodium falciparum* to Guatemala from the Democratic Republic of the Congo. *Emerg Infect Dis*. 2014; 20(6):932–940. <http://dx.doi.org/10.3201/eid2006.131204>. [PubMed: 24856348]
- Peakall R, Smouse PE. GENALEX 6: genetic analysis in excel. population genetic software for teaching and research. *Mol Ecol Notes*. 2006; 6(1):288–295.
- R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; Vienna, Austria: 2011.
- Raleigh C, Linke A, Hegre H, Karlsen J. Introducing acled: an armed conflict location and event dataset special data feature. *J Peace Res*. 2010; 47(5):651–660.
- Real LA, Henderson JC, Biek R, Snaman J, Jack TL, Childs JE, Nadin-Davis S, et al. Unifying the spatial population dynamics and molecular evolution of epidemic rabies virus. *Proc Natl Acad Sci U S A*. 2005; 102(34):12107. [PubMed: 16103358]
- Reboudet S, Bogreau H, Silai R, Lepère J, Bertaux L, Pradines B, Rogier C, et al. Genetic structure of *Plasmodium falciparum* and elimination of malaria, Comoros archipelago. *Emerg Infect Dis*. 2010; 16(11):1686. [PubMed: 21029525]
- Roper C, Pearce R, Breckenkamp B, Gumede J, Drakeley C, Mosha F, Chandramohan D, Sharp B. Antifolate antimalarial resistance in southeast Africa: a population-based analysis. *Lancet*. 2003; 361:1174–1181. [PubMed: 12686039]
- Schultz L, Wapling J, Mueller I, Ntsuke P, Senn N, Nale J, Siba P, et al. Multilocus haplotypes reveal variable levels of diversity and population structure of *Plasmodium falciparum* in Papua New Guinea, a region of intense perennial transmission. *Malar J*. 2010; 9(1):336. [PubMed: 21092231]
- Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP. Landscape genetics: where are we now? *Mol Ecol*. 2010; 19(17):3496–3514. [PubMed: 20723061]
- Storfer A, Murphy MA, Evans JS, Goldberg CS, Robinson S, Spear SF, Waits LP, et al. Putting the “landscape” in landscape genetics. *Heredity*. 2007; 98(3):128–142. <http://dx.doi.org/10.1038/sj.hdy.6800917>. [PubMed: 17080024]
- Taylor SM, Messina JP, Hand CC, Juliano JJ, Muwonga J, Tshefu AK, Meshnick SR, et al. Molecular malaria epidemiology: mapping and burden estimates for the Democratic Republic of the Congo, 2007. *PLoS One*. 2011; 6(1):e16420. [PubMed: 21305011]
- Taylor SM, Antonia AL, Parobek CM, Juliano JJ, Janko M, Emch M, Meshnick SR. *Plasmodium falciparum* sulfadoxine resistance is geographically and genetically clustered within the DR Congo. *Sci Rep*. 2013; 3:1165. [PubMed: 23372922]
- Wallace RG, Fitch WM. Influenza A H5N1 immigration is filtered out at some international borders. *PLoS ONE*. 2008; 3(2):e1697. <http://dx.doi.org/10.1371/journal.pone.0001697>. [PubMed: 18301773]
- United Nations High Commissioner for Human Rights. Report of the Mapping Exercise Documenting the most Serious Violations of Human Rights and International Humanitarian Law Committed within the Territory of the Democratic Republic of the Congo between March 1993 and June 2003. 2010
- World Health Organization. World Malaria Report: 2013. World Health Organization; 2013.



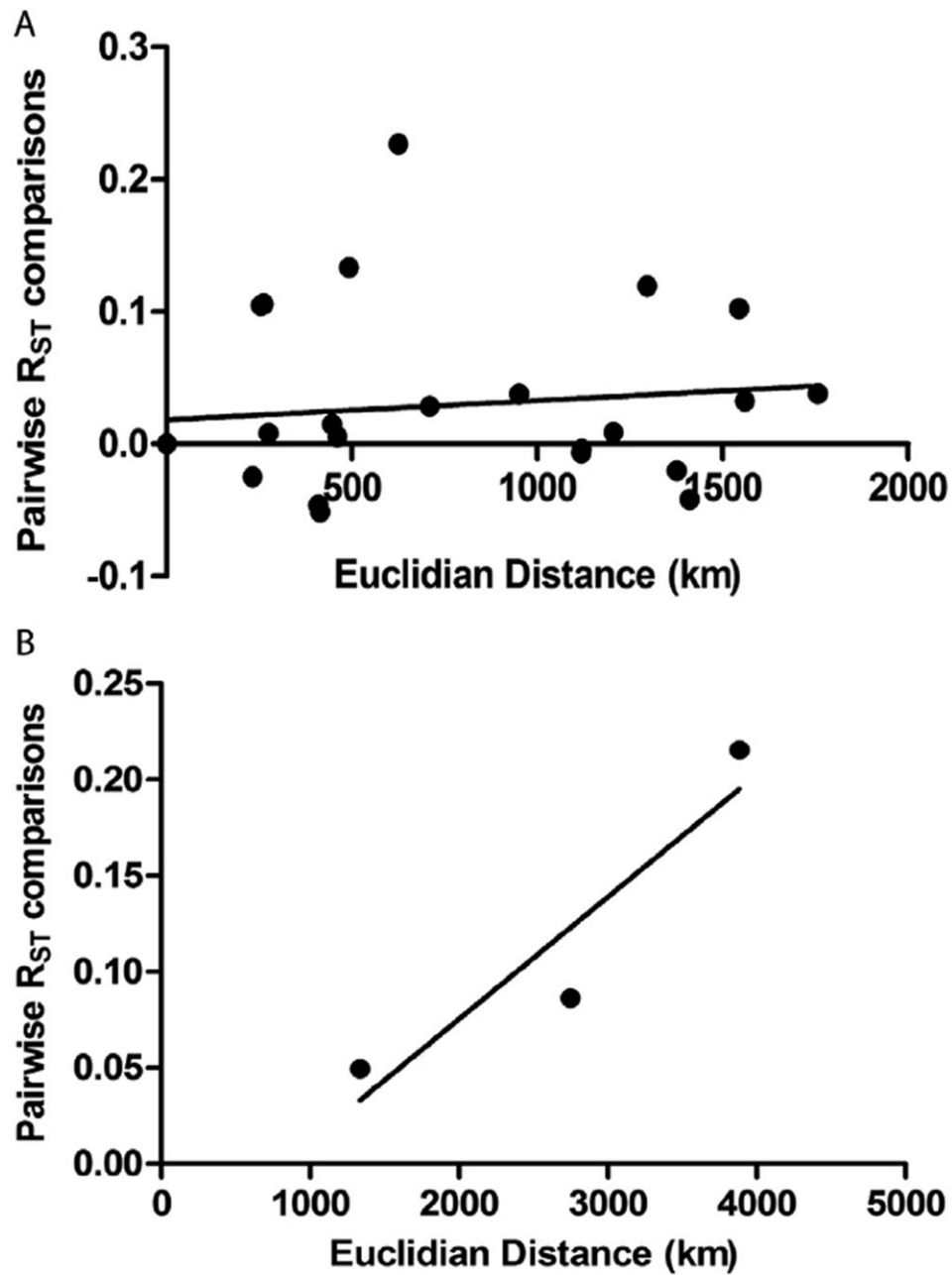
**Fig. 1.**  
Location of the major cities and rivers of the DRC, and the seven DHS clusters (203, 81, 88, 183, 211, 164, and 29) included in the study.



**Fig. 2.**

(A). Pairwise  $R_{ST}$  comparisons between the DRC (with all clusters treated as a single population), Ghana, and Kenya. (B) Genetic relatedness (pairwise  $R_{ST} < 0.05$ ) between the DRC clusters, Ghana (1) and Kenya (2).

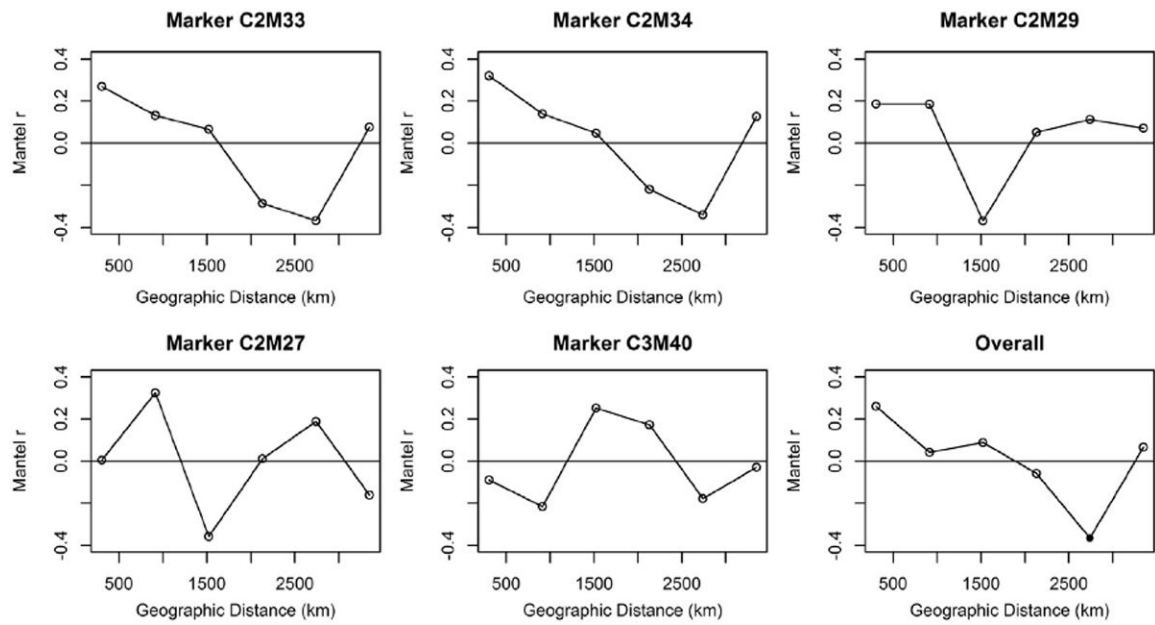




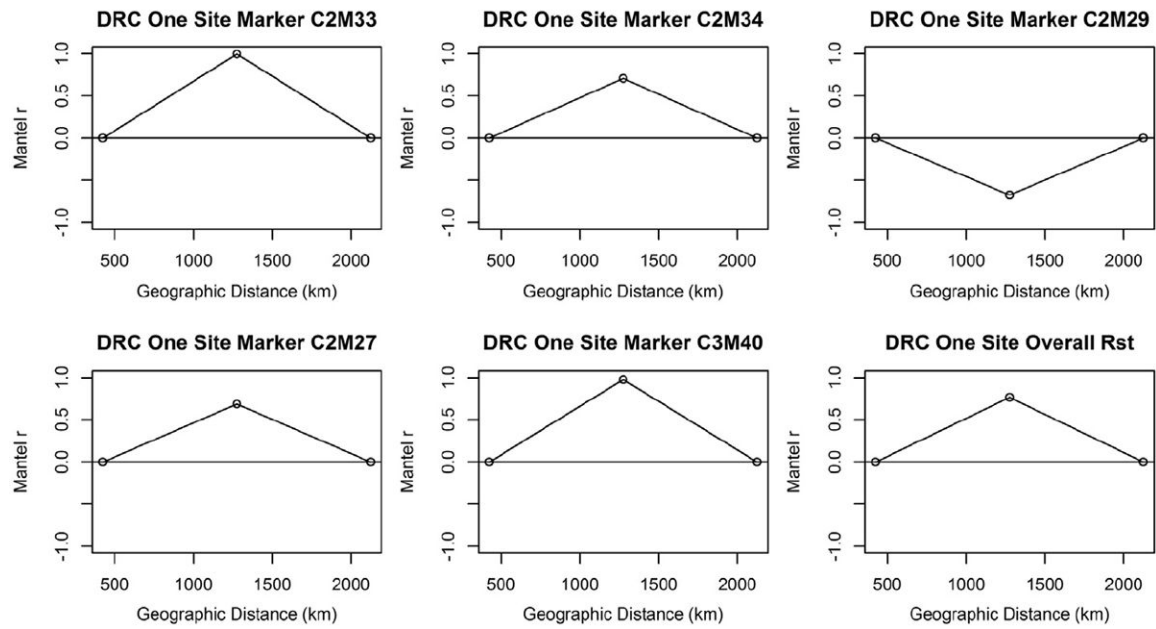
**Fig. 3.**

(A) Pairwise genetic relatedness and Euclidian distance between the seven DRC clusters.

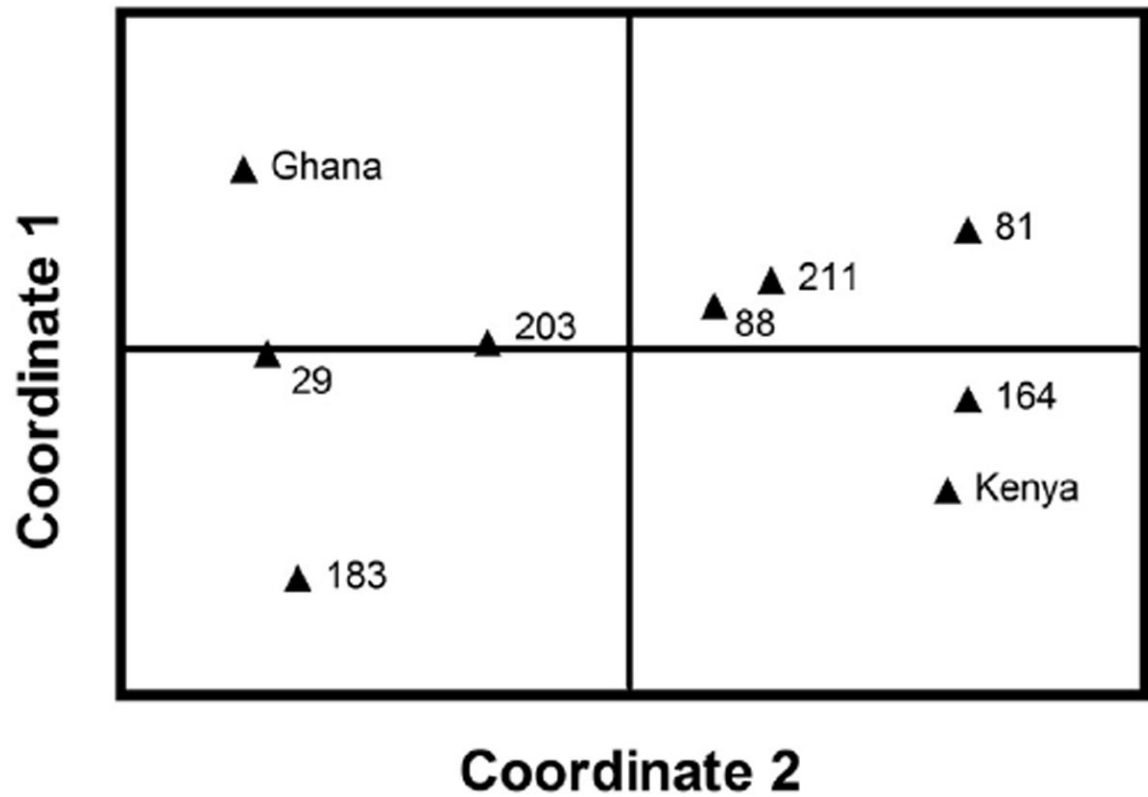
(B). Pairwise genetic relatedness and Euclidian distance between the DRC (with all clusters treated as a single population), Ghana, and Kenya. Black dots represent pairwise comparisons; black line is the linear trend line.



**Fig. 4.** Mantel correlograms indicating genetic similarity (above the zero line) or dissimilarity (below the zero line) across geographic distance thresholds, treating each DRC cluster separately.

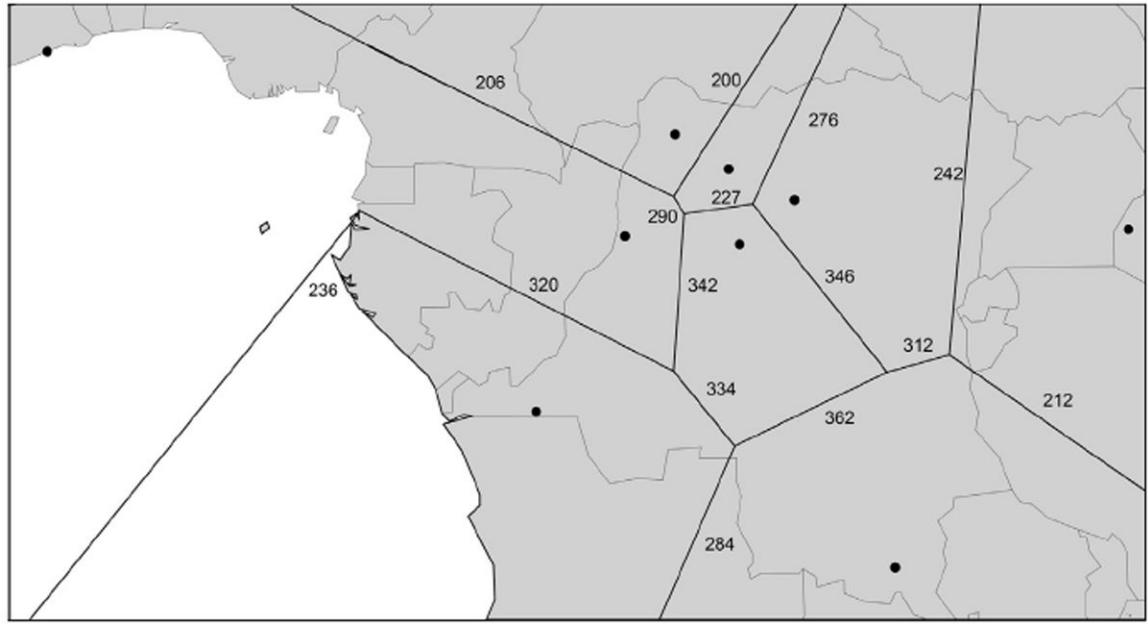


**Fig. 5.** Mantel correlograms indicating genetic similarity (above the zero line) or dissimilarity (below the zero line) across geographic distance thresholds, treating the DRC malarial parasites as one population.



**Fig. 6.**

Principle Coordinates Analysis (PCoA) of genetic relatedness of all seven DRC clusters, Ghana, and Kenya using pairwise  $R_{ST}$  values. Coordinates 1 and 2 account for 54.09 and 36.18% of the variation, respectively (including a 3rd coordinate explains another 9.35% of the variation).



**Fig. 7.**

Locations of barriers found, via Monmonier's algorithm, between malarial parasite populations. The frequency of the barrier being detected by the algorithm is shown next to each barrier. For a barrier to achieve statistical significance it would need to be detected a high percentage of the time (i.e. 950 times out of 1000).

Pairwise overall  $R_{ST}$  values (below diagonal) and geographic distances (in km above diagonal) between the seven DRC clusters, Ghana and Kenya. Small  $R_{ST}$  ( $<0.05$ ) are in bold, and indicate virtual genetic identity between parasite populations.

Table 1

	29	81	88	164	183	203	211	Ghana	Kenya
29 n=12		1543.85	1559.88	1296.18	1375.79	1410.59	1755.65	3518.48	1464.79
81 n=9	0.103		445.87	413.86	624.72	709.51	409.69	2131.6	1802.97
88 n=10	<b>0.032</b>	<b>0.015</b>		274.51	261.96	1118.03	231.19	2439.6	1445
164 n=20	0.120	<b>0.0</b>	<b>0.008</b>		254.42	950.46	460.56	2536.09	1390.97
183 n=10	<b>0.0</b>	0.227	0.106	0.105		1204.87	492.07	2692.69	1196.53
203 n=9	<b>0.0</b>	<b>0.029</b>	<b>0.0</b>	<b>0.038</b>	<b>0.009</b>		1119.16	2126	2221.65
211 n=12	<b>0.038</b>	<b>0.0</b>	<b>0.0</b>	<b>0.006</b>	0.134	<b>0.0</b>		2229.03	1658.28
Ghana n=72	<b>0.032</b>	0.117	0.098	0.177	0.161	<b>0.011</b>	0.104		3884.43
Kenya n=62	0.142	<b>0.033</b>	0.073	<b>0.006</b>	<b>0.085</b>	0.054	<b>0.068</b>	0.215	



**Table 2**

Mantel r (measure of correlation between two matrices) for  $R_{ST}$  and geographic distance between sites, at the subnational and national scales. Statistically significant ( $p < 0.05$ ) values are in bold.

Marker	<u>DRC as 7 clusters</u>		<u>DRC as single site</u>	
	Mantel r	<i>p</i> -value	Mantel r	<i>p</i> -value
C2M33	0.373	0.084	0.919	0.326
C2M34	<b>0.496</b>	<b>0.020</b>	0.948	0.344
C2M29	0.211	0.204	-0.277	0.665
C2M27	0.060	0.350	0.290	0.669
C3M40	-0.113	0.567	0.970	0.322
Overall	0.392	0.057	0.972	0.349