UNIVERSITY OF CALIFORNIA,
IRVINE

Landscape Genetics of African Malaria Parasite and Its Vectors

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Ecology and Evolutionary Biology

by

Elizabeth Hemming-Schroeder

Dissertation Committee:
Associate Professor Guiyun Yan, Chair
Associate Professor Jose Ranz
Associate Professor Andrew Noymer

2018

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to express the deepest appreciation to my committee chair, Guiyun Yan, who has provided continual support and outstanding research opportunities. Without his guidance and persistent help this dissertation would not have been possible.

I would like to thank my committee members, Professor Jose Ranz and Professor Andrew Noymer, for their support and invaluable advice throughout my graduate career. In addition, I would like to thank my past and current lab mates for their mentorship and input along the way, especially Dr. Eugenia Lo. Finally, thank you to the staff and technicians from the the Kenya Medical Research Institute (KEMRI) at Kisumu for sample collection, as well as their hospitality during my stay.

Financial support was provided by National Institutes of Health grants R01 AI050243, D43 TW01505, and U19 AI129326, the University of California, Irvine Chancellor's Club Fellowship, and a Department of Education GAANN Fellowship.

CURRICULUM VITAE

Elizabeth Hemming-Schroeder

2012	E. David Cronon Stewardship Fellow of the Lakeshore Nature Reserve, Madison, WI
2013	Assistant to Vector-borne Disease Epidemiologist, Wisconsin Department of Health Services, Madison, WI
2013	Undergraduate Research Assistant, Medical Entomology Laboratory, University of Wisconsin, Madison
2013	B.Sc. in Biology, Certificate in Global Health, University of Wisconsin, Madison
2016	M.Sc. in Ecology and Evolutionary Biology, University of California, Irvine
2018	Ph.D. in Ecology and Evolutionary Biology, University of California, Irvine

FIELD OF STUDY

Molecular Epidemiology of Malaria

PUBLICATIONS

Zhong D, Lo E, Wang X, Yewhalaw D, Zhou G, Atieli HE, Githeko A, Hemming-Schroeder E, Lee MC, Afrane Y, Yan G. Multiplicity and molecular epidemiology of Plasmodium vivax and Plasmodium falciparum infections in East Africa. Malar J. 2018;17(1):185.

Hemming-Schroeder E, Lo E, Salazar C, Puente S, Yan G. Landscape Genetics: A Toolbox for Studying Vector-Borne Diseases. Frontiers Ecol Evol. 2018; 8(6):21.

Hemming-Schroeder E, Umukoro E, Lo E, Fung B, Tomás-Domingo P, Zhou G, Zhong D, Dixit A, Atieli H, Githeko A, Vardo-Zalik A. Impacts of antimalarial drugs on *Plasmodium falciparum* drug resistance markers, Western Kenya, 2003–2015. Am J Trop Med Hyg. 2018; 98(3):692-699.

Hemming-Schroeder E, Strahl S, Yang E, Nguyen A, Lo E, Zhong D, Atieli H, Githeko A, Yan G. Emerging Pyrethroid Resistance among Anopheles arabiensis in Kenya. Am J Trop Med Hyg. 2018; 98(3):704-709.

Lo E, Hemming-Schroeder E, Yewhalaw D, Nguyen J, Kebede E, Zemene E, Getachew S, Tushune K, Zhong D, Zhou G, Petros B, Yan G. Transmission dynamics of co-endemic Plasmodium vivax and P. falciparum in Ethiopia and prevalence of antimalarial resistant genotypes. PLoS Negl Trop Dis. 2017; 11(7):e0005806.

Lo E, Nguyen K, Nguyen J, Hemming-Schroeder E, Xu J, Etemesi H, Githeko A, Yan G. Plasmodium malariae Prevalence and csp Gene Diversity, Kenya, 2014 and 2015. Emerg Infect Dis. 2017;23(4):601.

Lo E, Lam N, Hemming-Schroeder E, Nguyen J, Zhou G, Lee MC, Yang Z, Cui L, Yan G. Frequent Spread of Plasmodium vivax Malaria Maintains High Genetic Diversity at the Myanmar-China Border without Distance and Landscape Barriers. J Infect Dis. 2017; 216(10):1254-63.

Hemming-Schroeder E, Lo E. Evidence of a hard selective sweep for artemisinin resistant Plasmodium falciparum. Lancet Infect Dis. 2017; 17(5):462-3.

Lo E, Nguyen J, Oo W, Hemming-Schroeder E, Zhou G, Yang Z, Cui L, Yan G. Examining Plasmodium falciparum and P. vivax clearance subsequent to antimalarial drug treatment in the Myanmar-China border area based on quantitative real-time polymerase chain reaction. BMC Infect Dis. 2016; 16 (1):154.

ABSTRACT OF THE DISSERTATION

Landscape Genetics of African Malaria Parasite and Its Vectors

By

Elizabeth Hemming-Schroeder

Doctor of Philosophy in Ecology and Evolutionary Biology

University of California, Irvine, 2018

Professor Guiyun Yan, Chair

During a time of intensive antimalarial campaigns, it is crucial to understand the effects these campaigns have on the population genetics of malaria parasites and mosquitoes, particularly with respect to genes associated with drug and insecticide resistance. In addition, as countries approach malaria elimination, it will be critically important to understand the underlying factors that cause malaria epidemics or help to sustain malaria transmission in order to effectively control malaria and achieve elimination. Therefore, my dissertation research aims to A) evaluate the impact that public health interventions have on the population genetics of malaria parasites and mosquitoes; and B) assess the relative impact that key ecological factors have on the dispersal, measured through gene flow, of malaria parasites and vectors. To address these aims, I collected malaria parasite and mosquito samples in Kenya and genotyped them for molecular markers associated with drug and insecticide resistance, as well as neutral markers to infer gene flow. I tested the link between key ecological factors (temperature, precipitation, vegetation index, topographic wetness index, human population density, and distance to roads) and spatial genetic structure between populations using landscape genetic analytic methods. I found a recent increase in drug resistance markers associated with the antimalarial drug used to prevent malaria

in pregnancy, as well as an increase in polymorphisms associated with increased tolerance to the partner drug of the first-line treatment for malaria. I found a key mutation to be associated with insecticide resistance in *Anopheles arabiensis* in Kenya, as well as that this mutation is common throughout Western Kenya. Finally, I found that high human population density promotes dispersal of *An. gambiae s.s.*, high temperatures and low vegetation indices promote dispersal of *An. arabiensis,* and that physical barriers to human travel, such as lakes, may prevent dispersal of *P. falciparum* in Kenya. These findings allow us to identify areas susceptible to the introduction of malaria parasites and malaria vectors, as well as drug and insecticide resistance.

INTRODUCTION

Optimistic from the promising potential of DDT as an effective insecticide for malaria vectors, the World Health Organization (WHO) initiated the Global Malaria Eradication Program (GMEP) in 1955. GMEP interventions were largely focused on indoor residual spraying (IRS) of DDT with little flexibility to adapt eradication programs to local malaria epidemiology situations.[1] Within just fourteen years, after multiple resurgences of malaria, the notion of worldwide malaria eradication was abandoned.[1] The failure of the GMEP has since largely been attributed to a lack of scientific research to support antimalarial interventions.[1] In addition, decreased susceptibility to DDT and waning monetary interest in malaria eradication further contributed to the disbandment of the GMEP.[1]

Today, approximately 1.2 billion people worldwide remain at high risk of malaria.[2] Recently, there has been resurgence in monetary support to eradicate malaria, especially from the Bill and Melinda Gates Foundation.[3] As countries approach malaria elimination and pre-elimination phases, it will be critically important to understand the underlying factors which promote and prevent migration of malaria parasites and vectors in order to effectively sustain local malaria control and elimination. In addition to understanding potential susceptibilities to malaria parasite reintroduction, knowledge of how malaria parasites and vectors migrate is significant to understanding how insecticide and drug resistance spreads. Moreover, during a time of intensive antimalarial campaigns, it is crucial to understand the effects these campaigns have on the genetic structure of malaria parasites and mosquitoes, particularly with respect to genes associated with drug and insecticide resistance. Therefore, through this dissertation, I aim to A) assess the impact that public health interventions have on malaria parasites and mosquitoes; and B) evaluate the relative impact that key ecological factors have on the dispersal

of malaria parasites and vectors. These findings will help to inform antimalarial interventions in Kenya. In addition, these methods and findings may be broadly applicable to informing interventions in other countries, especially as malaria eradication is being considered.

Landscape genetics can be an extremely useful approach to improve our understanding of how vector-borne diseases spread.[4] Broadly, landscape genetics aims to quantify the effects of landscape heterogeneity on microevolutionary processes, such as gene flow, genetic drift, and/or natural selection.[5] Landscape genetics has primarily and traditionally been used by conservation biologists, such as to identify vulnerable populations and identify areas where corridors are needed to promote gene flow.[6] More recently, landscape genetics has been used to study various infectious diseases, such as chronic wasting disease,[7,8] rabies in domestic dogs,[9] raccoon rabies,[10-13] hantavirus,[14,15] H5N1 avian influenza,[16] and malaria.[17-19] Landscape genetic analysis begins by developing hypotheses of how landscape factors resist gene flow.[20] To test isolation-by-resistance hypotheses, landscape genetics integrates analytical tools across multiple disciplines, including landscape ecology, population genetics, and spatial statistics. It requires the use of molecular markers to measure genetic variation and to infer gene flow. Other tools needed are geographic information systems, remote sensing, population genetics, and statistical and mathematical modeling techniques.[21] Moreover, landscape genetics analysis of vector-borne diseases is particularly complex, as several factors must be considered when designing a study, such as those related to the environment, vector mobility, and human and pathogen mobility. However, for this same reason, this method can be useful to parse out numerous factors and identify potential hotspot areas of disease movement for targeted public health interventions and containment of disease and drug resistance.[22]

Kenya provides an ideal setting for my proposed research because of its heterogeneous landscape, from the malaria endemic lowlands spanning the coastline of the Indian ocean, that rise inland to an arid interior. Further west leads to the malaria epidemic highlands which are bisected by the Rift Valley, and finally to the west are the malaria endemic lowlands surrounding Lake Victoria. Kenya has a warm and humid climate along the eastern coastline, with wildlife-rich savannah grasslands inland towards the central highland. Toward the west, there is a warm and humid climate around Lake Victoria, hilly areas and temperate forested, Kakamega Forest, which is relic of an East African rainforest. The land cover varies from the dense tropical forests of the reserve and the mangrove forests along the coastline of Indian Ocean, to the shrubbery of the arid desert lands of the North to the thick mountainous forests and alpine vegetation along the slopes of the snow-capped mountains.

Kenya is in the control phase of malaria with 70% of the total population of 44.9 million people living in areas of high malaria transmission.[23] Although significant progress has been made in decreasing malaria cases and malaria deaths, malaria still remains a major public health problem in Kenya accounting for 18% of outpatient visits an 6% of hospital admission.[24] The most common method of malaria protection in Kenya is insecticide treated bed nets (ITN). In 2014, 34% of households had at least one ITN for every two people.[24] Artemether-lumefantrine is the first-line treatment for *P. falciparum* malaria. The major malaria species is *Plasmodium falciparum* and the major vector species are *An. gambiae* s.s., *An. arabiensis*, and *An. funestus*.

The overall goal of my dissertation is to gain knowledge of how ecological factors and public health interventions impact population genetic structure of malaria parasites and mosquitoes in Kenya. My dissertation is therefore divided into four chapters: chapters one and two consider public health interventions; and chapters three and four focus on ecological factors

influencing population structure. In the first chapter, I investigate how antimalarial drug policies and use have impacted *P. falciparum* molecular markers associated with increased drug tolerance over a thirteen year period. In the second chapter, I examine insecticide resistance in *An. arabiensis* populations across Kenya through the geographic distribution of knockdown resistance mutations and bioassays. In chapter three, I investigate the landscape, climatic, and social factors driving the population structure of *Anopheles gambiae s.s.* and *An. arabiensis.* Finally, in the fourth chapter, I consider the factors influencing the population structure of *P. falciparum.*

Through this research, I found a recent increase in drug resistance markers associated with the antimalarial drug used to prevent malaria in pregnancy, as well as an increase in polymorphisms associated with increased tolerance to the partner drug of the first-line treatment for malaria. I found a key mutation to be associated with insecticide resistance in *Anopheles arabiensis* in Kenya, as well as that this mutation is common throughout Western Kenya. I identified high human population density as the factor primarily promoting dispersal of *An. gambiae s.s.,* while high temperatures and low vegetation indices drive the dispersal patterns of *An. arabiensis.* Finally, I found that physical barriers to human travel, such as lakes, may prevent dispersal of *P. falciparum* in Kenya. These findings related to prevalent genetic markers associated with antimalarial drug and insecticide resistance have important implications for maintaining effective public health policies. In addition, by identifying factors important for promoting the dispersal of malaria vectors and parasites, we improve our understanding of the areas which would be susceptible to the re-introduction of malaria parasites and malaria vectors following antimalarial interventions, as well as the areas which are most susceptible to the invasion of drug resistant parasites and insecticide resistant mosquitoes.

References

1. Nájera JA, González-Silva M, Alonso PL, 2011. Some lessons for the future from the Global Malaria Eradication Programme (1955–1969). *PLoS Med 8*(1): e1000412.
2. World Health Organization. World Malaria Report 2014. Geneva: WHO; 2014 http://www.who.int/malaria/publications/world_malaria_report_2013/en/
3. Tanner M, Savigny DD, 2008. Malaria eradication back on the table. *Bull of World Health Organ 86*(2): 82.
4. Biek R, Real LA, 2010. The landscape genetics of infectious disease emergence and spread. *Mol Ecol 19*: 3515–3531.
5. Balkenhol N, Waits L P, Dezzani RJ, 2009. Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. *Ecography 32*: 818–830.
6. Storfer A, Murphy MA, Spear SF, Holderegger, R, Waits LP, 2010. Landscape genetics: where are we now? *Mol Ecol 19*: 3496–3514.
7. Blanchong JA, Samuel MD, Scribner KT, Weckworth BV, Langenberg JA, Filcek KB, 2008. Landscape genetics and the spatial distribution of chronic wasting disease. *Biol Lett 4*: 130–133.
8. Robinson SJ, Samuel MD, Rolley RE, and Shelton P, 2013. Using landscape epidemiological models to understand the distribution of chronic wasting disease in the Midwestern, U S A. *Landscape Ecol 28*: 1923–1935.
9. Brunker K, Hampson K, Horton DL, Biek R, Randolph SE, 2012. Integrating the landscape epidemiology and genetics of RNA viruses: rabies in domestic dogs as a model *Parasitology 139*: 1899–1913.

10. Rees EE, et al, 2008. Assessing a landscape barrier using genetic simulation modelling: implications for raccoon rabies management. *Prev Vet Med 86*: 107–123.
11. Rees EE, et al, 2009. Landscape modelling spatial bottlenecks: implications for raccoon rabies disease spread. *Biol Lett 5*: 387–390.
12. Cullingham CI, Kyle CJ, Pond BA, Rees EE, White BN, 2009. Differential permeability of rivers to raccoon gene flow corresponds to rabies incidence in Ontario, Canada. *Mol Ecol 18*: 43–53.
13. Côté H, Garant D, Robert K, Mainguy J, Pelletier F, 2012. Genetic structure and rabies spread potential in raccoons: the role of landscape barriers and sex-biased dispersal. *Evol Appl 5*: 393–404.
14. Guivier, E, et al, 2011. Landscape genetics highlights the role of bank vole metapopulation dynamics in the epidemiology of Puumala hantavirus. *Mol Ecol* 20: 3569–3583.
15. Dubois A, et al, 2017. Microevolution of bank voles (*Myodes glareolus*) at neutral and immune-related genes during multiannual dynamic cycles: consequences for Puumala hantavirus epidemiology. *Infect Genet Evol 49*: 318–329.
16. Carrel M, Wan XF, Nguyen T, Emch M, 2011. Genetic Variation of highly pathogenic H5N1 avian influenza viruses in vietnam shows both species-specific and spatiotemporal associations. *Avian Dis 55*: 659–666.
17. Carrel, et al, 2015. The geography of malaria genetics in the Democratic Republic of Congo: a complex and fragmented landscape. *Soc Sci Med 133*: 233–24.

18. Lo, E, et al, 2017. Transmission dynamics of co-endemic Plasmodium vivax and P falciparum in Ethiopia and prevalence of antimalarial resistant genotypes. *PLoS Negl Trop Dis* 11: e0005806.
19. Lo E, et al, 2017. Frequent spread of Plasmodium vivax malaria maintains high genetic diversity at the Myanmar-China Border, without distance and landscape barriers. *J Infect Dis 216*: 1254–1263.
20. Shirk AJ, Landguth EL, Cushman SA, 2017. A comparison of individual-based genetic distance metrics for landscape genetics. *Mol Ecol Resour 17*: 1308–1317.
21. Manel S, Holderegger R, 2013. Ten years of landscape genetics. *Trends Ecol Evol 28*, 614–621.
22. Hemming-Schroeder E, Lo E, Salazar C, Puente S, Yan G, 2018. Landscape Genetics: A Toolbox for Studying Vector-Borne Diseases. *Frontiers in Ecology and Evolution 6*: 21
23. World Health Organization. World Malaria Report 2015. Geneva: WHO; 2015 http://www.who.int/malaria/publications/country-profiles/profile_ken_en.pdf
24. President's Malaria Initiative. Malaria Operational Plan FY 2016. 2016 http://www.pmi.gov/docs/default-source/default-document-library/malaria-operational-plans/fy16/fy-2016-kenya-malaria-operational-plan.pdf?sfvrsn=5

CHAPTER 1

Impacts of antimalarial drugs on *Plasmodium falciparum* drug resistance markers, Western Kenya, 2003-2015

Originally published in the American Journal of Tropical Medicine and Hygiene (2018). 98(3): 692 – 699. doi: 10.4269/ajtmh.17-0763

Abstract

Antimalarial drug resistance has threatened global malaria control since chloroquine resistant *Plasmodium falciparum* emerged in Asia in the 1950s. Understanding the impacts of changing antimalarial drug policy on resistance is critical for resistance management. *Plasmodium falciparum* isolates were collected from 2003-2015 in Western Kenya and analyzed for genetic markers associated with resistance to chloroquine (*Pfcrt),* sulfadoxine-pyrimethamine (*Pfdhfr/Pfdhps*), and artemether-lumefantrine (*PfKelch13/Pfmdr1*) antimalarials. In addition, household antimalarial drug use surveys were administered. *Pfcrt* 76T prevalence decreased from 76% to 6% from 2003 to 2015. *Pfdhfr/Pfdhps* quintuple mutants decreased from 70% in 2003 to 14% in 2008, but increased to near fixation by 2015. SP 'super resistant' alleles *Pfdhps* 581G and 613S/T were not detected in the 2015 samples that were assessed. The *Pfmdr1* N86-184F-D1246 haplotype associated with decreased lumefantrine susceptibility increased significantly from 4% in 2005 to 51% 2015. No *PfKelch13* mutations that have been previously associated with artemisinin resistance were detected in the study populations. The increase in *Pfdhfr/Pfdhps* quintuple mutants that associates with sulfadoxine-pyrimethamine resistance may be resulted from the increased usage of sulfadoxine-pyrimethamine for intermittent preventative therapy in pregnancy (IPTp) and for malaria treatment in the community. Prevalent *Pfdhfr/Pfdhps* mutations call for careful monitoring of sulfadoxine-pyrimethamine resistance and effectiveness of the current IPTp program in Kenya. Additionally, the commonly occurring *Pfmdr1* N86-184F-D1246 haplotype associated with increased lumefantrine tolerance calls for surveillance of AL efficacy in Kenya, as well as consideration for a rotating ACT regimen.

Introduction

Antimalarial drug resistance has significantly hindered malaria control efforts and played a key role in shaping global drug policies since the first reports of chloroquine (CQ) resistance arose from Southeast Asia in 1957.[1] Since then, due to widespread drug resistance, global recommendations for the first-line treatment of malaria have changed from CQ to sulfadoxine-pyrimethamine (SP), and again, most recently, from SP to artemisinin-combination therapy (ACT).[1] Since both CQ and SP drug resistance arose in Southeast Asia before spreading to Africa,[2] the emergence of ACT resistance in several Southeast Asian countries and recent report on the emergence of indigenous artemisinin-resistant *P. falciparum* in Africa[3] triggers major concern on the efficacy of malaria control programs in Africa where most of the global malaria burden falls.[4]

Though delayed clearance of the parasite following artemisinin treatment has been reported in African countries, such as Kenya,[5] Nigeria,[6] and Angola,[7] the association of African *PfKelch13*mutations with clinical resistance is not clear and mutations associated with artemisinin resistance in Southeast Asia have yet to be commonly observed in Africa.[8, 9, 10, 11, 12] Close monitoring and resistance validation of *PfKelch13* mutations, as well monitoring for mutations associated with ACT partner drug resistance in East Africa will be critical to detecting the spread of ACT resistance from Southeast Asia to Africa or indigenous emergence. Additionally, although ACTs have been implemented as first-line treatment for malaria in Kenya since 2006, other antimalarial drugs, including SP and CQ, continue to be used for treating malaria,[13] further complicating malaria treatment in Kenya. Moreover, intermittent preventative treatment of malaria in pregnancy (IPTp) with SP as prophylaxis for malaria in pregnancy was

adopted as Kenya national policy in 1998,[14] which may lead to continued selection pressure for mutations associated with SP resistance.

Here, we investigated the dynamics of antimalarial drug resistance markers in response to changing antimalarial drug policy in Western Kenya. *Plasmodium falciparum* samples across the years 2003, 2005, 2008, and 2015 were examined, before and after the first mass distribution of artemether-lumefantrine (AL) in Kenya in 2006. Frequencies of amino acid polymorphisms in genes including *Pfcrt* for CQ resistance,[2] *Pfdhfr* and *Pfdhps* for SP resistance,[15] *Pfmdr1* for lumefantrine tolerance,[16] and *PfKelch13* for artemisinin resistance [17] were assessed. We examined whether the observed amino acid changes have been undergoing selection through a longitudinal comparison of mutation frequencies in these drug resistance genes. Understanding the impacts of antimalarial drug policy on molecular markers of drug resistance and monitoring for artemisinin resistance are critical to informing antimalarial drug policy in Kenya.

Methods

Study design and participants

This study was conducted in two sites in Western Kenya: Kakamega (0.282º N, 34.752º E), a low malaria transmission site and Kombewa (0.105º S, 34.520º E), a high malaria transmission site. The differences in malaria transmission intensities is partly attributed to the differences in altitude between sites, where Kakamega is in the highlands (1,430-1,580-m elevation) and Kombewa is a lowland site (1,170-1,300 m). Blood samples were collected from asymptomatic school children between the ages of 6-15 years old in 2003, 2005, 2008, and 2015. Sampling methods were consistent across the years studied. School-aged children were studied because they are among the age groups with the highest risk of malaria infection. A total of 705

P. falciparum isolates were collected between 2003 and 2015 at the two study sites, ranging from 29 to 194 isolates per site per year. Samples with more than one mixed (mutation/wildtype) mutation site were discarded from haplotype analyses, but were included in individual SNP analyses. Blood dots were made on filter paper for genotyping and stored at -20°C until use.

Scientific and ethical clearance was given by the institutional scientific and ethical review boards of the Kenya Medical Research Institute, Kenya and the University of California, Irvine, USA. Written informed consent/assent for study participation was obtained from all consenting heads of households and each individual who was willing to participate in the study.

Procedures

The Saponin/Chelex method was used to extract parasite DNA from dried blood samples.[18] Quantitative polymerase chain reaction (qPCR) of *P. falciparum*-specific 18S rRNA was used to detect *P. falciparum* infections.[19] *P. falciparum* isolates were genotyped at *Pfcrt* for CQ resistance; *Pfdhfr* and *Pfdhps* for SP resistance; *Pfmdr1* for lumefantrine tolerance; and *PfKelch13* for artemisinin resistance (see Supplementary Table 2 for codon positions). For genes *Pfcrt, Pfmdr1, Pfdhfr,* and *Pfdhps*, a restriction enzyme digestion protocol was used to detect specific mutations among samples collected in 2003, 2005, and 2008 [20] and a subset of mutations were confirmed by direct sequencing; for samples in 2015, the mutations of these target genes were assessed by PCR and sequencing.[21, 22] For *PfKelch13,* samples were amplified and sequenced using the published protocol.[17] We used the *Pfcrt*76T mutation as a proxy for CQ resistance and *Pfdhfr*51I-59R-108N/*Pfdhps*437G-540E quintuple mutant for SP resistance because of their strong associations with antimalarial resistance.[2, 15] In addition SP 'super resistant' alleles *Pfdhps* 581G, *Pfdhps* 613S/T, and *Pfdhfr* 164L were examined in 2015

isolates[15].The *Pfmdr1* N86-184F-D1246 haplotype was used as a proxy for reduced AL susceptibility because of evidence that lumefantrine selects for this haplotype.[23, 24, 25] Nonsynonymous mutations in the *PfKelch13* propeller region were assessed for artemisinin resistance, given that single amino acid changes in this region have been associated with *in vivo* and *ex vivo* resistance.[8, 17, 26, 27] Amplified PCR fragments were purified and sequenced from both ends by Sanger sequencing (GENEWIZ, Inc.). All sequences were blasted against NCBI GeneBank database for verification. Sequences were visualized using Chromas v2.5.0, aligned with ClustalX v2.1, and manually edited in Bioedit v7.2.5. Sequences were deposited to Genbank (accession numbers MF344967-MF345825).

Household antimalarial usage surveys

A cross-sectional survey was conducted for a total of 10,519 randomly selected households in Western Kenya in the years 2003, 2007, 2011, and 2016 to assess antimalarial drug usage. For the years 2003 and 2007, surveys were conducted in Kakamega and Kisii counties, and in 2010 and 2016, surveys were conducted in Kakamega and Vihiga counties. No significant differences between sites within years were observed, and so results from multiple sites were pooled for visualization and analysis. Questionnaires were administered to an adult member of each surveyed household. Specifically, in the questionnaires, household heads were asked to name which medicine was used for the family member whom had the most recent malaria episode.

Data Analysis

Two-tailed chi-square tests and Fisher's exact tests were conducted to make pairwise comparisons for mutation frequencies between sites and years for all haplotypes and individual polymorphisms assessed. A Bonferonni correction for 28 tests was applied, placing significance at 0.0018. The 95% confidence intervals (CI) were computed using the binomial distribution. Linkage disequilibrium was tested for all samples with complete genotypes after omitting samples from mixed infections (n=168). Linkage disequilibrium estimates were calculated in Genepop 4.2 for all possible pairs of loci.[28] P-values were calculated using Fisher's tests with a Bonferonni correction for 36 tests across nine loci, placing significance at 0.0014.

Results

Significant changes in frequencies of drug resistance molecular markers were observed with changes in antimalarial drug policy and reported use over the thirteen-year study period in Western Kenya. A decreasing trend in *Pfcrt* 76T mutation, associated with CQ resistance,[2] was observed from 2003 to 2015 at both study sites (Figure 1.1A). However, differences were observed between the sites in 2008 when the *Pfcrt* 76T mutation was observed at a significantly higher frequency in Kakamega at 91.9% than in Kombewa at 61.0%. By 2015, *Pfcrt*76T mutation frequencies declined to 2.7% and 11.8% in Kombewa and Kakamega, respectively.

Individual mutations important for SP resistance[15] revealed varying trends over the years at both study sites (Figure 2). For instance, the frequency of *Pfdhfr* N51I mutation decreased significantly between 2003 and 2008, but rebounded and increased significantly in 2015 (Figure 1.2). Likewise, a significant decrease in *Pfdhfr* C59R mutation was detected at Kombewa from 2003 to 2008, but the mutation frequency bounced back in 2015 to a high level as seen in 2003

(Figure 1.2). While, for *Pfdhfr* S108N, the most important mutation for *in vitro* pyrimethamine resistance,[2] there was relatively little change in mutation frequencies across the years at both study sites (Figure 1.2). *Pfdhps* A437G, and *Pfdhps* K540E showed a similar trend of little change across the years (Figures 1.2). The *Pfdhfr* triple mutant and *Pfdhps* double mutant were also found to be most prevalent haplotypes among the 2015 samples. The 'super resistant' alleles *Pfdhps* 581G and *Pfdhps* 613S/T were not detected in 2015 samples, which was also the result of a 2005 study in Western Kenya.[29] The 'super resistant' allele *Pfdhfr* 164L was detected in one 2015 isolate.

For *Pfdhfr/Pfdhps* haplotype analysis, a total of 300 samples were excluded for having either multiple mixed mutation sites or incomplete haplotypes. While the quintuple *Pfdhfr/Pfdhps* mutant, considered to be fully resistant to SP,[15] decreased between 2005 and 2008 following the policy change to AL in 2004 (though AL was not distributed until 2006) (Figure 1.1B), both study sites experienced a significant increase in *Pfdhfr/Pfdhps* quintuple mutants between 2008 and 2015, exceeding the frequencies seen in 2003 (Figure 1.1A). For instance, at Kombewa, *Pfdhfr/Pfdhps* quintuple mutant frequencies increased by 15-fold from 2008 to 2015; a three-fold increase was also found in Kakamega within the same time period. By 2015, *Pfdhfr/Pfdhps* quintuple mutant frequencies were 96.7% at Kombewa and 77.8% at Kakamega.

Pfmdr1 N86, 184F, and D1246 polymorphisms are associated with decreased lumefantrine susceptibility.[16, 23, 24, 25] Though, N86 may be the most important polymorphism for increased lumefantrine tolerance.[16, 24] A significant decrease in mutation prevalence was observed from 2008 to 2015 for *Pfmdr1* N86Y and *Pfmdr1* D1246Y at both study sites. While, for *Pfmdr1* Y184F, at Kakamega, an increase in mutation frequencies was observed from 2005 to

2015. At Kombewa, there was no significant change in *Pfmdr1* Y184F mutation frequencies across collection years (Figure 1.3).

For *Pfmdr1* haplotype analysis, a total of 154 samples were excluded for having either multiple mixed mutation sites or incomplete haplotypes. A significant increase in the *Pfmdr1* N86-184F-D1246 haplotype frequency, associated with lumefantrine tolerance,[23, 24, 25] was observed at Kombewa and Kakamega between 2005 when frequencies were 4.2% and 4.5%, respectively, and 2015. By 2015, the *Pfmdr1* N86-184F-D1246 frequency at Kombewa was 42.7% and at Kakamega was 66.7%. Estimates of linkage disequilibrium revealed no significant linkage between any of the polymorphisms investigated. Though not statistically significant after applying a Bonferonni correction, locus pairs *Pfdhps* 437/*Pfdhps* 540 and *Pfdhfr* 59/*Pfdphs* 540 were the most closely linked (p=0.006 and p=0.028, respectively).

Eleven unique nonsynonymous mutations were observed in *PfKelch13* among our samples (Figure 1.4). The most common mutations were A578S that was found in four isolates and E612D in three isolates (Figure 1.4). The remaining nine mutations were observed individually in only one isolate: I448M, L457I, C469W, N490S, R513S, S522C, A554S, A569S, and I590F.

According to the self-reported antimalarial drug use for treatment household survey (Figure 1.1B), SP drug use was less than 2% in 2007, one year after the distribution of AL. Nevertheless, the SP drug use increased to 10.8% in 2011 and 12% in 2016, despite the fact that AL was increasingly used as the first-line treatment from 2007 (49.2%) to 2016 (81%; Figure 1.1B). CQ use was at 4.8% in 2003, but was not reported in any of the following years. Antimalarial drug use surveys also revealed that in 2016, 92% of women who were pregnant in the past four months (n=109) took at least one dose of SP as intermittent preventative treatment

of malaria in pregnancy (IPTp) while pregnant (Table 1.1). This level of IPTp-SP coverage is increased from previous years in Western Kenya when coverage was 16% in 2001[30], and 51% in 2008[31].

Discussion

This study examined the impact of past and current antimalarial drug policy and usage on drug resistance genetic markers of *P. falciparum*, the most common and deadly malarial parasite in sub-Saharan Africa.[4] We found that mutations associated with resistance have declined for CQ, but have increased for SP following an initial decline. In addition, the prevalence of polymorphisms associated with lumefantrine tolerance have increased since pre-AL distribution levels. No known mutations associated with artemisinin resistance in Asia were detected.

CQ was retracted as first line antimalarial treatment in 1998 due to increasing and widespread reports of CQ resistance in sub-Saharan Africa.[32] Its resistance is primarily attributed to the mutation *Pfcrt* K76T.[2] As expected, we observed a significant decline in *Pfcrt* K76T prevalence to very low levels by 2015. Our findings corroborated the results from the community surveys that indicated very low CQ usage for antimalarial treatment over the past twelve years. Reduced selection pressure might no longer favor chloroquine resistance mutations, and thus a drastic reduction in mutants was observed from 2003 to 2015. Additionally, AL has been demonstrated to select for chloroquine susceptible parasites.[33] Thus, the observed increase in AL coverage could also favor the wildtype *Pfcrt* polymorphism. The decline of *Pfcrt* K76T mutation in Kenya, which was also reported in previous studies,[32, 34, 35] as well as nearby countries Tanzania[36] and Rwanda;[37] calls for careful study into the possibility for CQ to be reintroduced,

such as in a combination therapy or in limited cases. However, the risk of rapid re-emergence of CQ resistance should be cautiously evaluated prior to a potential reintroduction.

SP replaced CQ as the first-line treatment for malaria in Kenya in 1999. However, by 2003, reports showed that SP effectiveness was also faltering.[38] Quintuple mutants consisting of *Pfdhps* 437/540 and *Pfdhfr* 51/59/108 are considered to be fully resistant to SP [15]. These quintuple mutants were present in 91.3% of the isolates collected in 2015, which was remarkably higher than that observed in 2008 (13.8%). The striking increase in *Pfdhps*/*Pfdhfr* mutants could be partly explained by the observed increase in SP usage from less than 2% in 2007 to 12% in 2016, even though the frequency of SP use in 2016 was still less than that observed in 2003 (45.2%). Another explanation for the drastic rise in quintuple *Pfdhps*/*Pfdhfr* mutants could be the increased use of SP as IPTp, a guideline put forth by the WHO.[39] IPTp with SP as prophylaxis for malaria in pregnancy was adopted as Kenya national policy in 1998.[14] IPTp coverage was low in the first few years, with only 16% of pregnant women in Western Kenya reported taking at least one dose of IPTp-SP in 2001.[30] Since then, the IPT coverage has been increasing over the years with up to 51% of pregnant women in 2008[31] and 92% in 2016 (this study) reported taking at least one dose of IPTp-SP. The substantial increase in coverage of IPTp-SP from 2001 to 2016, coupled with the moderate increase in SP usage for malaria treatment from 2007 to 2016 likely impose selection pressure for SP resistant parasites.

Apart from Kenya, the increasing *Pfdhfr*/*Pfdhps* mutation frequencies have also been reported in several other African countries.[15] The dominance of fully resistant SP mutants is concerning because this could decrease IPTp-SP effectiveness and exacerbate malaria infections.[40] Although WHO recommends the continuation of IPTp-SP in malaria endemic countries across Africa,[39] it is imperative to monitor its effectiveness given the very high levels

of fully resistant mutants observed in this study.[15] The emergence of 'super resistant' alleles, such as *Pfdhfr*164L, *Pfdhps*581G, and *Pfdhps*613S/T sextuple mutant haplotype may further diminish the effectiveness of IPTp-SP.[15, 41, 42] For example, the 581G mutation has been associated with increased parasitemia in pregnant women in Tanzania.[43] Notably, we did not detect such 'super resistant' haplotypes in our study populations despite that they have been previously detected in Western Kenya at varying frequencies.[44, 45] Careful monitoring of SP resistance and emergence of 'super resistant' alleles is critical.

AL was first distributed in Kenya in 2006, following the policy change from SP to AL for first-line antimalarial treatment. Changes in lumefantrine sensitivity have been associated with polymorphisms in the *Pfmdr1* gene.[16, 25] For example, Tanzanian parasites having the *Pfmdr1* N86-184F-D1246 haplotype were able to withstand lumefantrine blood concentrations 15-fold higher than parasites with the 86Y-Y184-1246Y haplotype.[23] Additionally, in Uganda, AL was demonstrated to select for haplotypes with N86 in combination with 184F, D1246, or both.[24] Our findings of a significantly increased prevalence of N86-184F-Y1246 haplotypes since prior to the distribution of AL suggest that this haplotype is being selected for by AL. This finding of a commonly occurring haplotype associated with decreased lumefantrine susceptibility calls for continued surveillance of AL efficacy in Kenya. Additionally, since other ACTs such as artesunate-amodiaquine and dihydroartemisinin-piperaquine pose different selective pressures on *Pfmdr1* haplotypes than AL, rotating ACT regimens may be an effective strategy for delaying ACT partner drug resistance in Kenya.[24, 46]

No mutations associated with artemisinin resistance in Asia have been observed in our study populations. However, other *PfKelch13* mutations were observed at low frequencies. The nonsynonymous A578S *PfKelch13* mutation observed in this and other studies[8, 9, 47, 48, 49] was not

found to be associated with artemisinin resistance when introduced in the Dd2 line.[8] The *PfKelch13* E612D mutation observed in the present study has not been examined in regard to its association with AL resistance, but it has been observed in other parts of Africa.[50] Seven other nonsynonymous mutations were detected in our isolates post-ACT distribution. The fact that none of the Southeast Asian *PfKelch13* mutations were detected in our study populations suggests that there may be a combination of different factors that play a role in artemisinin resistance between the two continents.[47] For example, artemisinin resistance may require additional mutations at secondary loci,[51] such as those candidate SNPs identified by Chebon et al[52] in Kenyan *P. falciparum* on chromosomes 12 and 14. This notion is underscored by the observance of common delayed clearance of the parasite following ACT treatment in Kenya.[5] In addition, a five-year longitudinal study conducted in Uganda found that there was a correlation between the increased usages of ACT in communities with decreased sensitivities of the parasites to the drug.[53] These results suggest the possibility of an independent emergence of artemisinin resistance in Africa, which is not associated with *PfKelch13* mutations. As a result, closer surveillance of widespread ACT usage and deeper analyses of the parasite genome are needed to identify new or potential markers for artemisinin resistance in Africa.

Our study had certain limitations. The present study was limited to two study sites in Western Kenya. It is unclear whether a similar pattern is observed in other parts of Kenya or other countries. We did not examine *Pfmdr1* copy number, which has been shown to be associated with lumefantrine tolerance.[54] In addition, sample sizes for *PfKelch13* were relatively small especially in 2003 due to limited DNA quantity and quality in some of those earlier samples, which limits our ability to detect rare mutations. Lastly, Sanger sequencing of PCR products used in the present study has a lower sensitivity in detecting rare mutations in infections

with multiple clones compared to deep sequencing methods. However, the overall trends reported here would not be affected by these limitations.

The findings from this study have significant implications for malaria control in Kenya. Firstly, the efficacy of IPTp-SP in Kenya could be diminished by the near fixation of fully resistant SP mutants. Secondly, approximately 10% of surveyed patients continue to use SP for malaria treatment despite the policy that AL is the recommended first-line drug. Since fully resistant SP mutants are predominant in this region, it is conceivable that these patients experience high malaria treatment failure rates. Thirdly, we found that artemisinin resistance has not yet spread from Southeast Asia to Western Kenya as evidenced by the absence of *PfKelch13* mutations in Kenya that are known to be associated with drug resistance in Southeast Asia. However, we detected a nonsynonymous mutation *PfKelch13* E612D in multiple isolates that may be a potential candidate for *in vitro* validation for artemisinin resistance. Lastly, we found an increase in *Pfmdr1* haplotypes associated with decreased lumefantrine susceptibility, which calls for continued monitoring of AL effectiveness and potentially implementing multiple first-line ACTs to delay ACT partner drug resistance.

This study sheds light on the long-term dynamics of drug resistance markers in response to antimalarial policy. Our findings suggest that changes in first-line antimalarial treatment and IPTp policies have been followed by dramatic changes in molecular drug resistance markers. In addition, despite policy changes, ineffective drugs continue to be used for extended amounts of time, which may lead to the persistence of drug resistance markers. Understanding the interplay between drug resistance on a molecular level, antimalarial drug usage, and antimalarial drug policy is critical to informing antimalarial drug use policies.

Table 1.1. Intermittent preventative therapy in pregnancy (IPTp) antimalarial drug usage in Western Kenya (n (%)). n is the total number of pregnant women surveyed.

	2001 (n=903)	**2008** (n=444)	**2016** (n=109)
Took at least one dose of IPTp-SP while pregnant	147 (16%)	227 (51%)	100 (92%)
Communities surveyed	Kisii, Bondo	Kisii, Bondo	Kakamega, Vihiga
Data source	Guyatt et al 2004[30]	Gikandi et al 2008[31]	This study

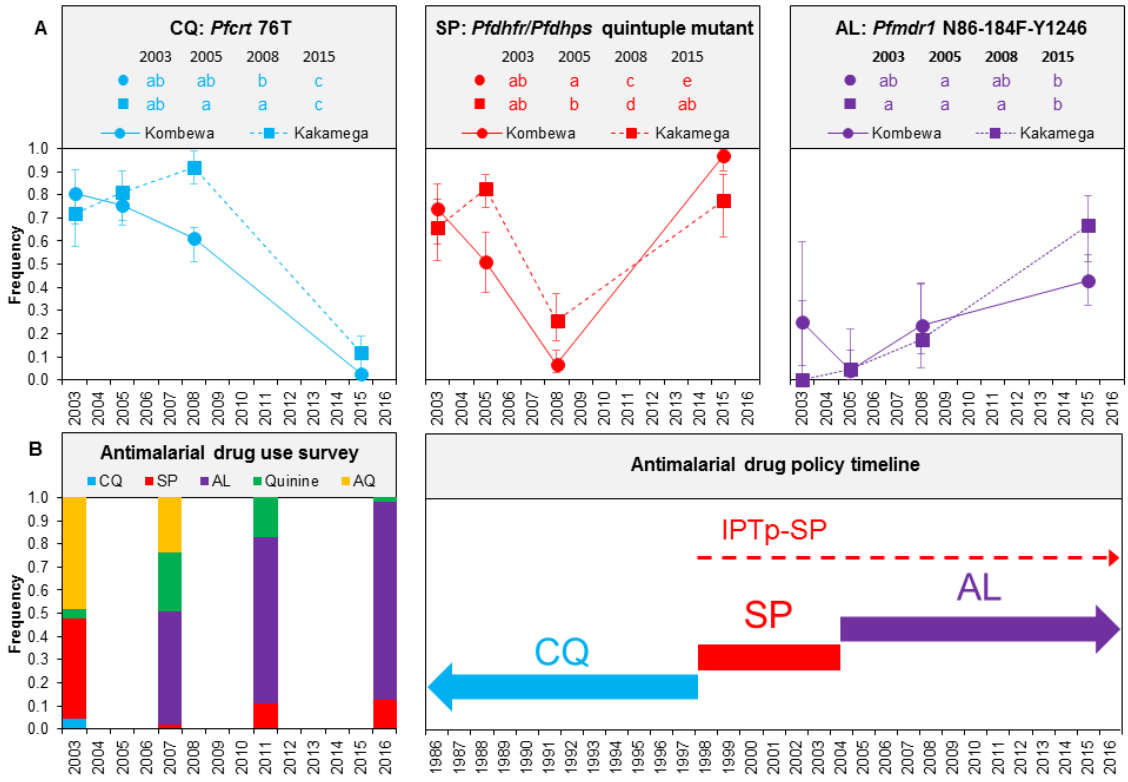


Figure 1.1. A comparison of A) *Plasmodium falciparum* drug resistance molecular markers to B) reported antimalarial drug usage for treatment and antimalarial drug policy timeline in Western Kenya. CQ is chloroquine; SP is sulfadoxine-pyrimethamine; AL is artemether-lumefantrine, AQ is amodiaquine. *Pfhfr/Pfdhps* quintuple mutant is *Pfdhfr*51I-59R-108N/*Pfdhps*437G-540E. Error bars represent 95% confidence intervals. Shared lowercase letters between study sites/years indicate that they are not significantly different from each other. Differing lowercase letters indicate statistically significant differences between study sites/collection years. Statistical significance was determined from the results of Fisher's exact tests with a Bonferroni correction for 28 tests between study sites and years ($p < 0.0018$). The timelines for first-line antimalarial drugs are indicated by bold arrows, whereas the timeline for intermittent preventative therapy in pregnancy (IPTp) is indicated by the dashed arrow.

Figure 1.2. Mutation frequencies for genetic markers associated with sulfadoxine-pyrimethamine resistance in *Plasmodium falciparum* samples collected in 2003-2015 at Kombewa and Kakamega. Differing lowercase letters indicate statistically significant differences between study sites/collection years. Statistical significance was determined from the results of Fisher's exact tests with a Bonferroni correction for 28 tests ($p < 0.0018$).

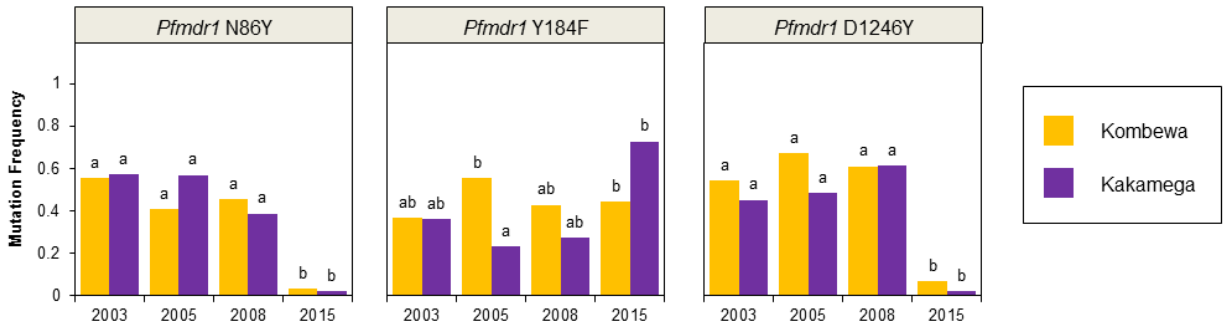


Figure 1.3. Mutation frequencies for genetic markers associated with lumefantrine resistance in *Plasmodium falciparum* samples collected in 2003-2015 at Kombewa and Kakamega. Differing lowercase letters indicate statistically significant differences between study sites/collection years. Statistical significance was determined from the results of Fisher's exact tests with a Bonferroni correction for 28 tests ($p < 0.0018$).

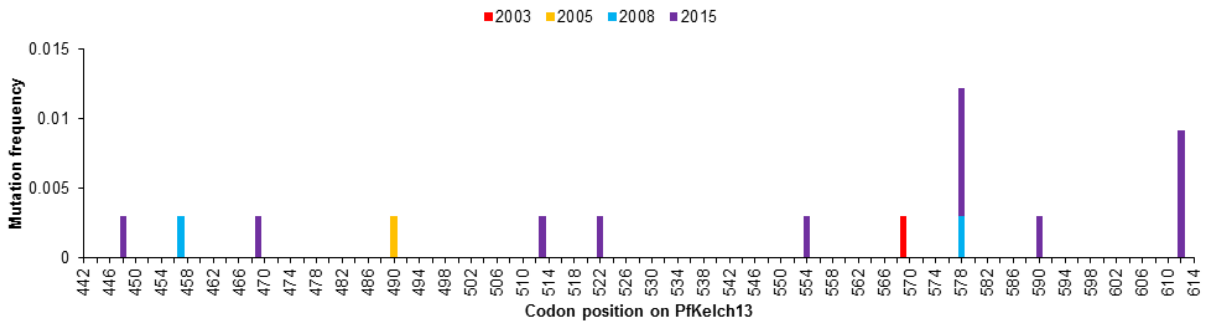


Figure 1.4. Nonsynonymous *PfKelch13* mutations in Western Kenya, 2003-2015.

References

1. World Health Organization. Antimalarial drug combination therapy: report of a WHO technical consultation. Geneva (Switzerland): WHO; 2001. Available at http://apps.who.int/iris/bitstream/10665/66952/1/WHO_CDS_RBM_2001.35.pdf
2. Wongsrichanalai C, Pickard AL, Wernsdorfer WH, Meshnick SR, 2002. Epidemiology of drug-resistant malaria. *Lancet Infect Dis 2*: 209-218.
3. Lu F, et al, 2017. Emergence of indigenous artemisinin-resistant *Plasmodium falciparum* in Africa. *N Engl J Med 376*10: 991-993.
4. World Health Organization. World malaria report 2016. Geneva (Switzerland): WHO; 2016. Available at: http://www.who.int/malaria/publications/world-malaria-report-2016/report/en/
5. Beshir KB, et al, 2013. Residual *Plasmodium falciparum* parasitemia in Kenyan children after artemisinin-combination therapy is associated with increased transmission to mosquitoes and parasite recurrence. *J Infect Dis 208*: 2017-2024.
6. Sowunmi A, Adewoye EO, Gbotsho GO, Happi CT, Sijuade A, Folarin OA, Okuboyejo TM, Michael OS, 2010. Factors contributing to delay in parasite clearance in uncomplicated *falciparum* malaria in children. *Malar J 9*:53.
7. Plucinski MM, et al, 2017. Efficacy of artemether-lumefantrine, artesunate-amodiaquine, and dihydroartemisinin-piperaquine for treatment of uncomplicated *Plasmodium falciparum* malaria in Angola, 2015. *Malar J 16*: 62.
8. Ménard D, Khim N, Beghain J, et al, 2016. A worldwide map of *Plasmodium falciparum K13*-propeller polymorphisms. *N Engl J Med 374*: 2453-2464.

9. Muwanguzi J, Henriques G, Sawa P, Bousema T, Sutherland CJ, Beshir KB, 2016. Lack of *K13* mutations in *Plasmodium falciparum* persisting after artemisinin combination therapy treatment of Kenyan children. *Malar J 15*(1): 1.
10. Cooper RA, Conrad MD, Watson QD, Huezo SJ, Ninsiima H, Tumwebaze P, Nsobya SL, Rosenthal PJ, 2015. Lack of artemisinin resistance in *Plasmodium falciparum* in Uganda based on parasitological and molecular assays. *Antimicrob Agents Chemother 59*: 5061-5064.
11. Boussaroque A, et al, 2016. Emergence of Mutations in the *K13* Propeller Gene of *Plasmodium falciparum* Isolates from Dakar, Senegal, in 2013-2014. *Antimicrob Agents Chemother 60*: 624-627.
12. Leang R, et al, 2015. Evidence of *Plasmodium falciparum* malaria multidrug resistance to artemisinin and piperaquine in western Cambodia: dihydroartemisinin-piperaquine open-label multicenter clinical assessment. *Antimicrob Agents Chemother 59*: 4719-4726.
13. Dixit A, Lee MC, Goettsch B, Afrane Y, Githeko AK, Yan G, 2016. Discovering the cost of care: consumer, provider, and retailer surveys shed light on the determinants of malaria health-seeking behaviours. *Malar J 15*: 179.
14. van Eijk AM, Hill J, Larsen DA, Webster J, Steketee RW, Eisele TP, ter Kuile FO, 2013. Coverage of intermittent preventive treatment and insecticide-treated nets for the control of malaria during pregnancy in sub-Saharan Africa: a synthesis and meta-analysis of national survey data, 2009-11. *Lancet Infect Dis 13*: 1029-1042.
15. Naidoo I, Roper C, 2013. Mapping 'partially resistant', 'fully resistant', and 'super resistant' malaria. *Trends Parasitol 29*: 505-515.

16. Venkatesan M, et al, 2014. Polymorphisms in *Plasmodium falciparum* chloroquine resistance transporter and multidrug resistance 1 genes: parasite risk factors that affect treatment outcomes for *P. falciparum* malaria after artemether-lumefantrine and artesunate-amodiaquine. *Am J Trop Med Hyg. 91*: 833-843.
17. Ariey F, et al, 2014. A molecular marker of artemisinin-resistant *Plasmodium falciparum* malaria. *Nature 505*: 50-55.
18. Bereczky S, Mårtensson A, Gil JP, Färnert A, 2005. Short report: rapid DNA extraction from archive blood spots on filter paper for genotyping of *Plasmodium falciparum*. *Am J Trop Med Hyg 72*: 249-251.
19. Rougemont M, Van Saanen M, Sahli R, Hinrikson HP, Bille J, Jaton K, 2004. Detection of four *Plasmodium* species in blood from humans by 18S rRNA gene subunit-based and species-specific real-time PCR assays. *J Clin Microbiol 42*: 5636-5643.
20. Vardo-Zalik AM, Zhou G, Zhong D, Afrane YA, Githeko AK, Yan G, 2013. Alterations in *Plasmodium falciparum* genetic structure two years after increased malaria control efforts in western Kenya. *Am J Trop Med Hyg 88*: 29-36.
21. Zhang GQ, Guan YY, Sheng HH, Zheng B, Wu S, Xiao HS, Tang LH, 2008. Multiplex PCR and oligonucleotide microarray for detection of single-nucleotide polymorphisms associated with *Plasmodium falciparum* drug resistance. *J Clin Microbiol 46*: 2167-2174.
22. Wang P, Brooks DR, Sims PF, Hyde JE, 1995. A mutation-specific PCR system to detect sequence variation in the dihydropteroatesynthetase gene of *Plasmodium falciparum*. *Mol Biochem Parasitol 71*: 115-125.
23. Malmberg M, Ferreira PE, Tarning J, Ursing J, Ngasala B, Björkman A, Mårtensson A, Gil JP, 2012. *Plasmodium falciparum* drug resistance phenotype as assessed by patient

antimalarial drug levels and its association with *pfmdr1* polymorphisms. *J Infect Dis 207*: 842-847.

24. Taylor AR, Flegg JA, Holmes CC, Guérin PJ, Sibley CH, Conrad MD, Dorsey G, Rosenthal PJ, 2017. Artemether-lumefantrine and dihydroartemisinin-piperaquine exert inverse selective pressure on *Plasmodium falciparum* drug sensitivity-associated haplotypes in Uganda. *Open Forum Infect Dis* 4: 229.
25. Sisowath C, Ferreira PE, Bustamante LY, Dahlström S, Mårtensson A, Björkman A, Krishna S, Gil JP, 2007. The role of *pfmdr1* in *Plasmodium falciparum* tolerance to artemether-lumefantrine in Africa. *Trop Med Int Health 12*: 736-742.
26. Straimer J, et al, 2015. *K13*-propeller mutations confer artemisinin resistance in *Plasmodium falciparum* clinical isolates. *Science 347*: 428-431.
27. Ashley EA, et al, 2014. Spread of artemisinin resistance in *Plasmodium falciparum* malaria. *N Engl J Med 371*: 411-423.
28. Rousset F, 2008. Genepop'007: a complete reimplementation of the genepop software for Windows and Linux. *Mol Ecol Resour 8*: 103-106.
29. Zhong D, Afrane Y, Githeko A, Cui L, Menge DM, Yan G, 2008. Molecular epidemiology of drug-resistant malaria in western Kenya highlands. *BMC Infect Dis 8*: 105.
30. Guyatt HL, Noor AM, Ochola SA, Snow RW, 2004. Use of intermittent presumptive treatment and insecticide treated bed nets by pregnant women in four Kenyan districts. *Trop Med Int Health 9*: 255-261.

31. Gikandi PW, Noor AM, Gitonga CW, Ajanga AA, Snow RW, 2008. Access and barriers to measures targeted to prevent malaria in pregnancy in rural Kenya. *Trop Med Int Health 13*: 208-217.

32. Kiarie WC, Wangai L, Agola E, Kimani FT, Hungu C, 2015. Chloroquine sensitivity: diminished prevalence of chloroquine-resistant gene marker *pfcrt-76* 13 years after cessation of chloroquine use in Msambweni, Kenya. *Malar J 14*: 328.

33. Sisowath C, Petersen I, Veiga MI, Mårtensson A, Premji Z, Björkman A, Fidock DA, Gil JP, 2009. In vivo selection of *Plasmodium falciparum* parasites carrying the chloroquine-susceptible *pfcrt K76* allele after treatment with artemether-lumefantrine in Africa. *J Infect Dis 199*: 750-757.

34. Achieng AO, et al, 2015. Temporal trends in prevalence of Plasmodium falciparum molecular markers selected for by artemether–lumefantrine treatment in pre-ACT and post-ACT parasites in western Kenya. *Int J Parasitol Drugs Drug Resis 5*: 92-99.

35. Lucchi NW, et al, 2015. In vitro and molecular surveillance for antimalarial drug resistance in Plasmodium falciparum parasites in Western Kenya reveals sustained artemisinin sensitivity and increased chloroquine sensitivity. *Antimicrob Agents Chemother 59*: 7540-7547.

36. Mohammed A, et al, 2013. Trends in chloroquine resistance marker, Pfcrt-K76T mutation ten years after chloroquine withdrawal in Tanzania. *Malaria J 12*: 415.

37. Kateera F, et al, 2016. Molecular surveillance of Plasmodium falciparum drug resistance markers reveals partial recovery of chloroquine susceptibility but sustained sulfadoxine-pyrimethamine resistance at two sites of different malaria transmission intensities in Rwanda. *Acta Trop 164*: 329-336.

38. Bousema JT, Gouagna LC, Meutstege AM, Okech BE, Akim NI, Githure JI, Beier JC, Sauerwein RW, 2003. Treatment failure of pyrimethamine-sulphadoxine and induction of *Plasmodium falciparum* gametocytaemia in children in western Kenya. *Trop Med Int Health 8*: 427-430.

39. World Health Organization. WHO policy brief for the implementation of intermittent preventive treatment of malaria in pregnancy using sulfadoxine-pyrimethamine (IPTp-SP). Geneva (Switzerland): WHO; 2013. Available at: http://www.who.int/malaria/publications/atoz/policy_brief_iptp_sp_policy_recommendation/en/

40. McGready R, White NJ, Nosten F, 2011. Parasitological efficacy of antimalarials in the treatment and prevention of *falciparum* malaria in pregnancy 1998 to 2009: a systematic review. *Int J Gynaecol Obstet 118*: 123-135.

41. Braun V, Rempis E, Schnack A, Decker S, Rubaihayo J, Tumwesigye NM, Theuring S, Harms G, Busingye P, Mockenhaupt FP, 2015. Lack of effect of intermittent preventive treatment for malaria in pregnancy and intense drug resistance in western Uganda. *Malar J 14*: 372.

42. Minja DT, et al, 2013. Plasmodium falciparum mutant haplotype infection during pregnancy associated with reduced birthweight, Tanzania. *Emerg Infect Dis 19*: 1446.

43. Harrington WE, Mutabingwa TK, Muehlenbachs A, Sorensen B, Bolla MC, Fried M, Duffy PE, 2009. Competitive facilitation of drug-resistant Plasmodium falciparum malaria parasites in pregnant women who receive preventive treatment. *Proc Natl Acad Sci 106*: 9027-9032.

44. Iriemenam NC, et al, 2012. Temporal trends of sulphadoxine-pyrimethamine (SP) drug-resistance molecular markers in Plasmodium falciparum parasites from pregnant women in western Kenya. *Malar J 11*: 134.

45. Spalding MD, Eyase FL, Akala HM, Bedno SA, Prigge ST, Coldren RL, Moss WJ, Waters NC, 2010. Increased prevalence of the pfdhfr/phdhps quintuple mutant and rapid emergence of pfdhps resistance mutations at codons 581 and 613 in Kisumu, Kenya. *Malar J 9*: 338.

46. Rosenthal PJ, 2013. The interplay between drug resistance and fitness in malaria parasites. *Mol Microbiol 89*: 1025-1038.

47. Kamau E, et al, 2014. *K13*-propeller polymorphisms in *Plasmodium falciparum* parasites from sub-Saharan Africa. *J Infect Dis 211*: 1352-1355.

48. Hawkes M, Conroy AL, Opoka RO, Namasopo S, Zhong K, Liles WC, John CC, Kain KC, 2015. Slow clearance of *Plasmodium falciparum* in severe pediatric malaria, Uganda, 2011-2013. *Emerg Infect Dis 21*: 1237.

49. Mohon AN, Alam MS, Bayih AG, Folefoc A, Shahinas D, Haque R, Pillai DR, 2014. Mutations in *Plasmodium falciparum K13* propeller gene from Bangladesh (2009-2013). *Malar J 13*: 431.

50. MalariaGEN Plasmodium falciparum Community Project, 2016. Genomic epidemiology of artemisinin resistant malaria. *eLife 5*: e08714.

51. Miotto O, et al, 2015. Genetic architecture of artemisinin-resistant *Plasmodium falciparum*. *Nat Genet 47*: 226-234.

52. Chebon LJ, et al, 2016. Genetically determined response to artemisinin treatment in western Kenyan *Plasmodium falciparum* parasites. *PloS one 11*: e0162524.

53. Conrad MD, et al, 2014. Comparative impacts over 5 years of artemisinin-based combination therapies on *Plasmodium falciparum* polymorphisms that modulate drug sensitivity in Ugandan children. *J Infect Dis 210*: 344-353.

54. Sidhu AB, Uhlemann AC, Valderramos SG, Valderramos JC, Krishna S, Fidock DA, 2006. Decreasing *pfmdr1* copy number in *Plasmodium falciparum* malaria heightens susceptibility to mefloquine, lumefantrine, halofantrine, quinine, and artemisinin. *J Infect Dis 194*: 528-535.

CHAPTER 2

Emerging pyrethroid resistance among *Anopheles arabiensis* in Kenya

Originally published in the American Journal of Tropical Medicine and Hygiene

(2018). 98(3): 704 - 709. doi: 10.4269/ajtmh.17-0445

Abstract

Vector control programs, particularly in the form of insecticide treated bed nets, are essential to achieving malaria elimination goals. Recent reports of increasing knockdown resistance (*kdr)* mutation frequencies for *Anopheles arabiensis* in Western Kenya heightens the concern on the future effectiveness of insecticide treated bed nets in Kenya. We examined resistance in *An. arabiensis* populations across Kenya through *kdr* mutations and WHO-recommended bioassays. We detected two *kdr* alleles, L1014F and L1014S. *Kdr* mutations were found in 5 of the 11 study sites, with mutation frequencies ranging from 3% to 63%. In two Western Kenya populations, the *kdr* L1014F allele frequency was as high as 10%. The L1014S frequency was highest at Chulaimbo at 55%. Notably, the *kdr* L1014F mutation was found to be associated with pyrethroid resistance at Port Victoria, but *kdr* mutations were not significantly associated with resistance at Chulaimbo, which had the highest *kdr* mutation frequency among all sites. This study demonstrated the emerging pyrethroid resistance in *An. arabiensis* and that pyrethroid resistance may be related to *kdr* mutations. Resistance monitoring and management are urgently needed for this species in Kenya where resistance is emerging and its abundance is becoming predominant. *Kdr* mutations may serve as a biomarker for pyrethroid resistance in *An. arabiensis*.

Introduction

Despite intensive malaria control efforts, malaria remains a leading cause of morbidity and mortality in Kenya, especially among younger children and pregnant women.[1] Vector control programs, particularly in the form of insecticide treated bed nets (ITNs) are essential to achieving malaria elimination goals[2,3] and have coincided with a decrease in malaria-related morbidity rates in Kenya.[4] However, increasing insecticide resistance threatens the efficacy of antimalarial interventions.[5]

Pyrethroids are the only approved insecticide for use in ITNs.[6] Its low mammalian toxicity and induction of paralysis using nerve stimulation of dysfunctional sodium channels makes it ideal for ITN usage.[5,7] However, a single amino acid change at residue position 1014 in the voltage-gated sodium channel (VGSC) gene of insects has made the insecticide increasingly obsolete. This mutation has been shown to confer knockdown resistance (*kdr*) by decreasing sodium channel affinity for the insecticide binding site.[8] The *kdr* mutations are found as L1014F (*kdr*-west) and L1014S (*kdr*-east) in *Anopheles gambiae*.[9] L1014F refers to a point mutation from leucine to phenylalanine, whereas L1014S represents a mutation from leucine to serine.[9,10] Originally, L1014F was found in Western Africa, hence leading to its name *kdr*-west,[11,12,13,14] whereas, L1014S (*kdr*-east) was found in Eastern Africa.[10,15] However, both mutations are now found throughout Africa and have not been solely concentrated geographically, thus suggesting a shift in *kdr* mutation frequencies in endemic countries.[16,17,18,19,20] Additionally, both *kdr* mutations have been associated with increased susceptibility to *Plasmodium falciparum*, further heightening malaria risk in areas with high insecticide resistance.[21]

Mass distribution of ITNs has been followed by a rapid increase in *kdr* alleles and insecticide resistance in *An. gambiae* s.s.[5] In Kenya, where ITN coverage increased from below

10% in 2004[22] to greater than 80% since 2013,[23] *kdr* mutation frequencies in *An. gambiae* s.s. increased rapidly from 6% in 2001[15] to near fixation at 98% in 2010.[5] In addition to the rise of *kdr* mutation frequencies in *An. gambiae s.s.*, higher ITN usage has led to a species shift from primarily *An. gambiae s.s.* to *An. arabiensis.*[2,24,25,26,27] As such, the contribution of *An. arabiensis* to malaria transmission increases in malaria endemic areas under the current ITN program.

Recently, *kdr* mutation frequencies in *An. arabiensis* from Western Kenya have been found to be increasing and were as high as 13% and 39% at certain localities in 2013.[6,23] Previously, in 2005, *kdr* mutation frequencies were not found to exceed 6% at any locality in Western Kenya[28] and moreover, were not detected in 2009[29]. Although the evasion of ITNs might explain why the frequency of *kdr* mutations and physiological insecticide resistance in *An. arabiensis* has remained relatively low with respect to *An. gambiae* s.s., we expect an increasing in *kdr* mutations for *An. arabiensis* to continue. However, we do not expect *kdr* mutations to increase as rapidly in *An. arabiensis* as they did in *An. gambiae s.s.* due to the reduced selection pressure imposed on *An. arabiensis* which more commonly feed outdoors.

Although ITNs are currently the most cost-effective method of preventing malaria, increased insecticide resistance and outdoor biting reduce their efficacy and present a major threat to malaria control programs.[1] Previous studies have examined the spatial distribution of *kdr* mutations in various *An. arabiensis* populations in Africa,[5, 6, 7, 12] but the association between *kdr* mutations and phenotypic resistance is not well-established. Therefore, this study aimed to examine the link between *kdr* mutations and pyrethroid resistance by comparing genotypes of phenotypically resistant and susceptible mosquitoes.

Materials and Methods

Study design for kdr survey

An. gambiae s.l. larvae were collected from eleven study sites across Kenya between May 2014 and October 2014 (Figure 2.1). No more than five larvae were collected from a given habitat to reduce sampling bias. Sampling bias was tested by comparing mutation frequencies to frequencies when randomly selecting one larvae per habitat and no significant differences were found. Study sites were selected across the diverse geographical regions of Kenya. The major regions were the lowlands surrounding Lake Victoria in Western Kenya (Port Victoria, Homa Bay, Kanyawegi, Chulaimbo, and Miwani), the highlands in Western Kenya (Kamkuywa) the Great Rift Valley in Western Kenya (Kabernet and Marigat), and coastal Kenya (Malindi, Mtwapa, Gazi).

WHO bioassays

To explore the link between *kdr* mutations and pyrethroid resistance, we genotyped phenotypically resistant and susceptible *An. arabiensis*, determined by a standard WHO insecticide susceptibility bioassay.[30] *Anopheles gambiae s.l.* larvae were collected from Port Victoria and Chulaimbo, study sites where *kdr* mutations in *An. arabiensis* had previously been detected,[22] and reared to adults. Adult female mosquitoes 2-3 days old were aspirated into exposure tubes in batches of 15-20 mosquitoes per tube. Tubes were lined with insecticide-impregnated paper, 0.05% deltamethrin. A subset of tubes was only lined with oil paper to serve as controls. In addition, the Kisumu susceptible *An. gambiae s.s.* strain was used as a control. After being held in their respective tubes for 60 minutes, mosquitoes were transferred to a holding tube with 10% sucrose solution and put to standard insectary conditions for 24 hours.

These mosquitoes were screened again. If after 24 hours mosquitoes were knocked down such that they were either dead or unable to fly, they were classified as susceptible.

Procedures

Genomic DNA was extracted from individual mosquitoes using standard ethanol extraction procedures with phenol:chloroform.[31] The final DNA pellet was suspended in 20ul of TE buffer. A Nano Drop 1000 Spectrophotometer was used to quantify DNA concentrations and stock DNA was diluted to an approximate concentration of 1 μg/μl for use in PCR. *Anopheles arabiensis* and *An. gambiae* s.s. were identified within the *An. gambiae* s.l. complex using a ribosomal DNA PCR assay.[32] We genotyped 683 *An. arabiensis* for *kdr* alleles: L1014 (wildtype), L1014F (*kdr*-west), and L1014S (*kdr*-east) using a Taqman probe assay.[33] For detection, the wildtype alleles were labeled with VIC at the 5' and the 1014F and 1014S *kdr* alleles were labeled with 6-FAM.

Statistical analysis

For the WHO bioassay, Fischer's exact tests were performed to make pairwise comparisons for mutation frequencies between resistant and susceptible groups. Odds ratios (OR) were used to quantify the association between *kdr* genotype and insecticide resistance phenotype. Chulaimbo and Port Victoria populations were analyzed separately.

Results

Kdr Survey

A total of 1425 *An. gambiae* s.l specimens were examined (Table 2.1). *Anopheles arabiensis* proportions ranged from 12.8% at Chulaimbo to 100% at Miwani, Bogoria, Gazi,

Mtwapa, and Malindi (Table 1). *Kdr* mutations were detected in five *An. arabiensis* populations: Port Victoria (10.3%), Homa Bay (2.3%), Kamkuywa (2.8%), Kanyawegi (15.8%), and Chulaimbo (63.2%) (Figure 2.1). The 1014F mutation prevalence was highest at Port Victoria (9.2%), Kanyawegi (10.5%), and Chulaimbo (8.5%), but also observed at Kamkuywa (2.9%) and Homa Bay (1.7%). The 1014S mutation was prevalent at Chulaimbo (54.7%) and detected at low frequencies at Port Victoria (1.1%), Homa Bay (0.6%), and Kanyawegi (5.3%). No mutations were observed in populations outside Western Kenya. The population at Chulaimbo was the only population that significantly deviated from Hardy-Weinberg equilibrium with regards to *kdr* alleles (Table 2.1).

WHO Bioassay

The control Kisumu susceptible *An. gambiae* s.s. strain had a mortality rate of 100%. We observed a mortality rate of 82.8% (95% CI [0.792-0.859]) and 73.7% (95% CI [0.610-0.834]) for *An. arabiensis* at Port Victoria and Chulaimbo, respectively. Both mortality rates were lower than the WHO 90% threshold for resistance (Figure 2.2A).

A comparison of *kdr* mutation frequencies between a subset of resistant and susceptible *An. arabiensis* revealed that deltamethrin-resistant mosquitoes had significantly higher frequencies of the L1014F mutation at Port Victoria (OR=3.495, 95% CI [1.809-7.102], $p<0.001$, Fischer's exact test) (Figure 2.2B), supporting the link between the *kdr* mutation and pyrethroid resistance. Although both L1014F and L1014S mutations were detected at Chulaimbo, the highest resistant field population, there was no significant difference in allele frequencies between susceptible and resistant groups ($p=0.078$; Fischer's exact test) (Figure 2.2B). When comparing only the L1014F frequency between groups at Chulaimbo, the difference is

marginally significant (OR=3.957, 95% CI [0.781-21.713], p=0.053; Fischer's exact test) and could be limited by a low sample size in the resistance group (n=14). Whereas there was no significant difference in L1014S frequencies between susceptible and resistant groups (OR=0.525, 95% CI [0.197-1.364], p=0.185, Fischer's exact test).

Discussion

The observed high proportions of *An. arabiensis* in this study demonstrate the ongoing species composition shift from predominantly *An. gambiae* s.s. to *An. arabiensis* in East Africa.[2,24,25,26,27] A decline in *An. gambiae* s.s. relative abundance yet stable population of *An. arabiensis* has been observed in the lowlands of Kenya in conjunction with an increase in ITN coverage.[2,7,23,27] These findings underscore the importance of the role that *An. arabiensis* are playing in maintaining residual malaria transmission, and as such, will present a major barrier to malaria control and elimination. Understanding *An. arabiensis* insecticide resistance mechanisms, as well as monitoring for resistance are essential to achieving malaria elimination goals.

The presence of *kdr* mutations at several sites in Western Kenya indicates the widespread occurrence of *kdr* mutations among *An. arabiensis* populations. In particular, the L1014F mutation, first detected in Kenya in 2012,[6] was observed in four of the five Western Kenya populations in this study. The emergence of L1014F was also found in neighboring malaria endemic countries. L1014F has recently been detected in Tanzania in both *An. gambiae* and *An. arabiensis* populations.[35] Moreover, high frequencies of the L1014F mutation in *An. arabiensis* have been reported from Ethiopia[36,37, 38] and central Sudan.[39] A continual increase in this mutation prevalence in Kenya may cause further concern on the future utility of ITNs.

The rise of the L1014F mutation may be particularly concerning given that this mutation was found to be associated with pyrethroid resistance in *An. arabiensis* in our Port Victoria study population. *Kdr* mutations at Chulaimbo were not significantly associated with pyrethroid resistance. This result could be due to the low frequency of L1014F and presence of the L1014S mutation at this site. The prevalence in L1014F mutations was higher in the resistant group at Chulaimbo, but the difference was not statistically significant. In *An. gambiae s.s.*, the L1014S mutation has been found to be more weakly associated with pyrethroid resistance than the L1014F mutation.[40] Similarly, the L1014F mutation may also have a stronger association with pyrethroid resistance in *An. arabiensis.* In Sudan, there was also a significant association found between the 1014F mutation and DDT and pyrethroid resistance in *An. arabiensis*, but the 1014S mutation was not detected in the populations tested.[39] Further studies are needed to investigate the role of the 1014S and 104F mutations in *An. arabiensis* insecticide resistance. The result also suggests that other mechanisms such as metabolic detoxification or secondary mutations at alternative loci could be involved in pyrethroid resistance in *An. arabiensis* at Chulaimbo, especially given the high levels of resistance at this site. Metabolic resistance using rapid insecticide detoxification due to the overexpression of P450 enzymes has been found to be a common resistance mechanism for *An. arabiensis*.[34,36,41,42]

Interestingly, *kdr* mutations were only observed in *An. arabiensis* specimens from study sites where *An. gambiae* were also common at proportions exceeding 30%. Stump et al.[15] first suggested the possibility that *kdr* alleles could have been introduced into Kenyan *An. arabiensis* populations through introgression. Adaptive introgression of *kdr* alleles has been supported by evidence of consequential contemporary gene flow between *An. arabensis* and *An. gambiae* in East Africa.[43,44] This notion is underscored by findings of identical intron sequences in the

VGSC between the two species in Kenya.[29] Our findings of *kdr* mutations occurring exclusively in *An. arabiensis* populations where *An. gambiae* are common are consistent with the hypothesis that *An. arabiensis* acquire *kdr* mutations through introgression with sympatric *An. gambiae* populations.

Pyrethroid resistance in *An. arabiensis* has been reported in several countries, including Sudan, [39] Ethiopia, [36,45] Malawi,[46] Tanzania,[47] Zanzibar,[48,49] and Kenya.[7] Despite wide-spread resistance in major malaria vectors in sub-Saharan Africa, pyrethroids are the only approved insecticide for use in ITNs.[6] The findings from this study and Abdalla et al.[38] that the L1014F mutation is associated with pyrethroid resistance in *An. arabiensis* provide evidence on the utility of screening *An. arabiensis* populations for *kdr* mutations in informing pyrethroid resistance status and trends. Though, that *kdr* mutations were not associated with resistance at Chulaimbo also highlights the complexity of insecticide resistance and the need for further studies on resistance mechanisms in *An. arabiensis*.

Kdr mutations could potentially increase and spread rapidly in a pattern like that observed for *An. gambiae* from 2001-2010.[5,15] Our results of commonly occurring 1014F mutations associated with pyrethroid resistance in *An. arabiensis* underscores the importance in searching for alternative methods to pyrethroid impregnated bed nets for vector control. High levels of resistance in *An. gambiae* s.s[5], *An. arabiensis* behavioral resistance to ITNs,[2] an increased proportion of *An. arabiensis*, and frequent *kdr* mutations in *An. arabiensis* from Western Kenya could all contribute to compromised efficacy of ITNs. Therefore, complementary interventions targeting outdoor mosquitoes, such as attractive toxic sugar baited traps, habitat reduction, and/or biological larvicides, could be important to improving the overall efficacy of

antimalarial programs, as well as suppressing pyrethroid resistance. These interventions have been effective for vector control in areas such as Mali,[50] Ecuador,[51] Peru,[51] and Kenya.[52]

In summary, we found evidence of widespread *kdr* mutations in Western Kenya and an association between the *kdr* 1014F mutation and pyrethroid resistance in *An. arabiensis.* This result is concerning for the effectiveness of ITNs, especially since *An. arabiensis* is becoming the predominant malaria vector in Kenya and throughout Africa.[2] Monitoring for the spread of insecticide resistance in *An. arabiensis* is critical for resistance management, and consequently, the success of vector control programs.

Table 2.1. Proportion of *Anopheles arabiensis* within the *Anopheles gambiae* s.l. species complex and knockdown resistance (kdr) genotype frequencies with Hardy-Weinberg equilibrium parameters for *An. arabeiensis* collected in Kenya, 2014.

Site	Elevation	*No.*	*An. arabiensis* (%)	Genotype frequencies (%)^					Hardy-Weinberg equilibrium	
				LL	LF	FF	LS	SS	H_E‡	F_{IS}§
Port Victoria	1139	168	56.5	80.4	18.5	0.0	0.0	0.0	0.187	0.013
Homa Bay	1184	133	68.4	95.3	3.5	0.0	1.2	0.0	0.046	-0.019
Kamkuywa	1487	72	52.8	91.9	5.4	0.0	0.0	0.0	0.054	0.000
Kanyawegi	1214	129	47.3	71.1	15.8	2.6	10.5	0.0	0.028	0.050
Chulaimbo	1377	446	12.8	26.9	17.3	0.0	0.0	55.8	0.558	0.690*
Miwani	1161	120	100	100	0.0	0.0	0.0	0.0	0.000	-
Marigat	1004	94	100	100	0.0	0.0	0.0	0.0	0.000	-
Kabernet	1150	101	92.1	100	0.0	0.0	0.0	0.0	0.000	-
Gazi	15	30	100	100	0.0	0.0	0.0	0.0	0.000	-
Mtwapa	66	44	100	100	0.0	0.0	0.0	0.0	0.000	-
Malindi	14	88	100	100	0.0	0.0	0.0	0.0	0.000	-

^L is wildtype at L1014 codon; F is L1014F mutation; S is L1014S mutation
‡H_E expected heterozygosity
§F_{IS} inbreeding coefficient
* Significant deviation from Hardy-Weinberg equilibrium

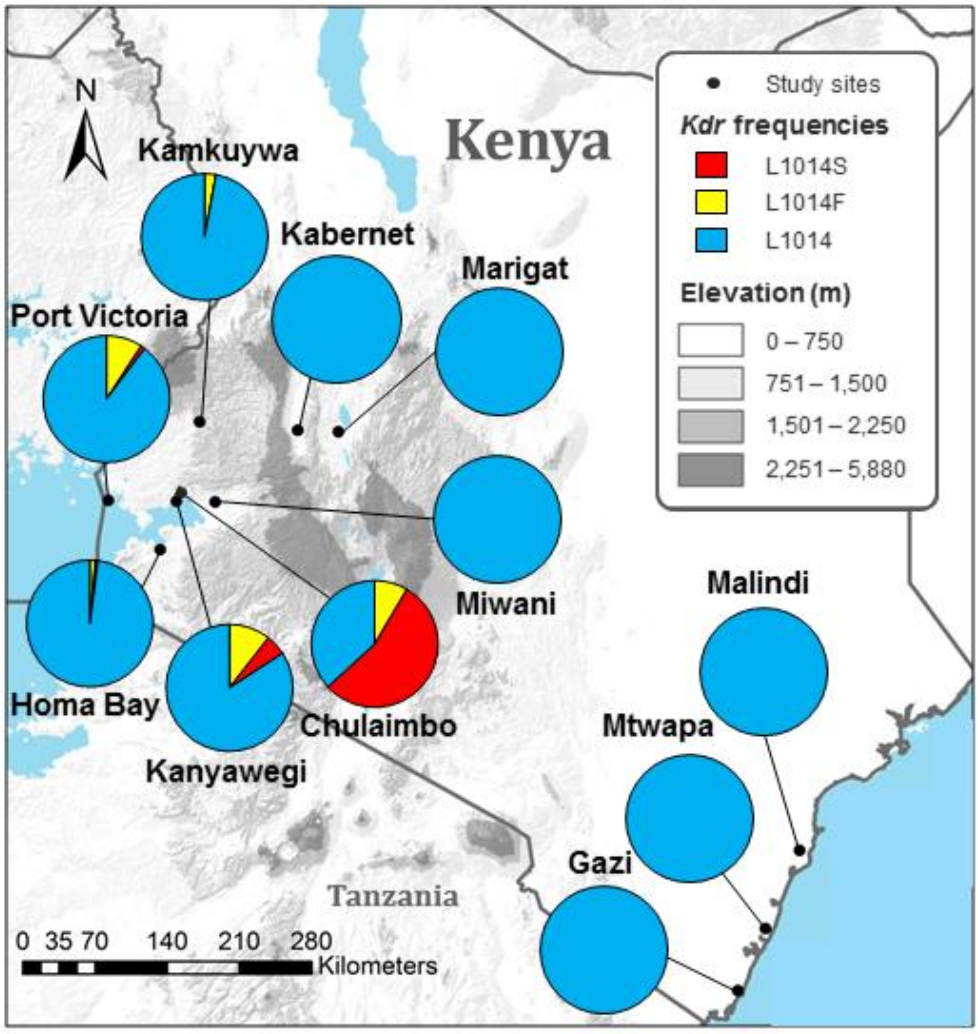


Figure 2.1. Knockdown resistance (*kdr)* allele frequencies in *Anopheles arabiensis* populations across Kenya, 2014. 1014F mutation prevalences: Kanyawegi (10.5%), Port Victoria (9.2%), Chulaimbo (8.5%), Kamkuywa (2.9%), Homa Bay (1.7%), Kabernet (0.0%), Marigat (0.0%), Miwani (0.0%), Gazi (0.0%), Mtwapa (0.0%), Malindi (0.0%). 1014S mutation prevalences: Chulaimbo (54.7%),Port Victoria (1.1%), Homa Bay (0.6%), Kanyawegi (5.3%), Kamkuywa (0.0%), Kabernet (0.0%), Marigat (0.0%), Miwani (0.0%), Gazi (0.0%), Mtwapa (0.0%), Malindi (0.0%).

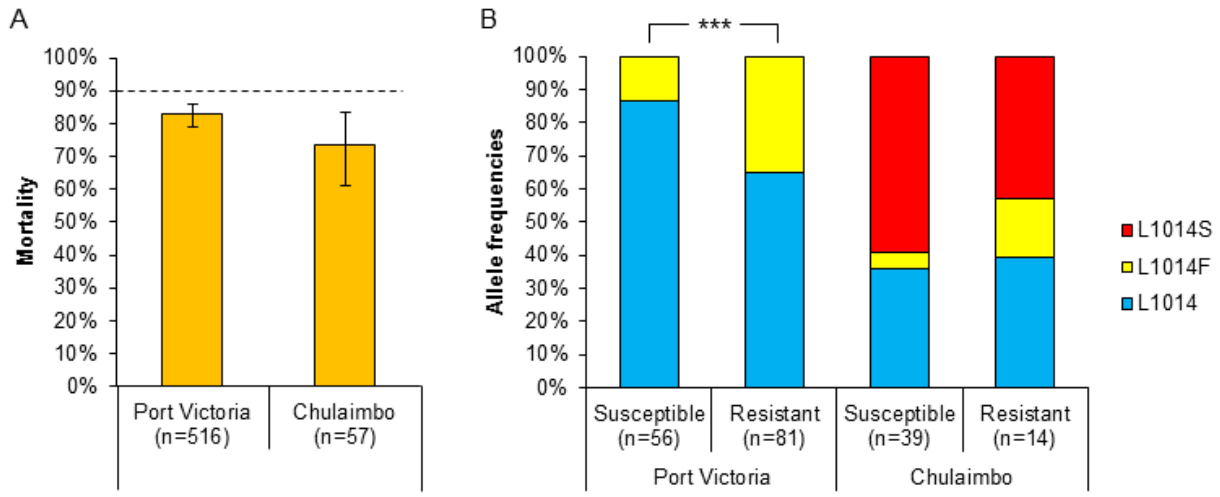


Figure 2.2. Mortality rates (A) and frequencies of knockdown resistance (*kdr)* alleles of susceptible and resistant groups (B) in *Anopheles arabiensis* populations in Kenya. The dotted line indicates WHO threshold for confirmed resistance (90%). *** indicates p<0.001. Error bars indicate 95% confidence interval. Mortality rates at Port Victoria: 82.8% (95% CI [0.792-0.859]); and Chulaimbo: 73.7% (95% CI [0.610-0.834]). 1014F mutation prevalences: Port Victoria Susceptible (13.3%), Port Victoria Resistant (35.2%), Chulaimbo Susceptible (5.1%), Chulaimbo Resistant (17.9%). 1014S mutation prevalences: Port Victoria Susceptible (0.0%), Port Victoria Resistant (0.0%), Chulaimbo Susceptible (59.0%), Chulaimbo Resistant (42.9%)

References

1. World Health Organization, 2014. World malaria report 2014. Geneva: WHO.
2. Bayoh MN, Mathias DK, Odiere MR, Mutuku FM, Kamau L, Gimnig JE, Vulule JM, Hawley WA, Hamel MJ, Walker ED, 2010. Anopheles gambiae: historical population decline associated with regional distribution of insecticide-treated bed nets in western Nyanza Province, Kenya. *Malar J 9*: 62.
3. Ranson H, N'Guessan R, Lines J, Moiroux N, Nkuni Z, Corbel V, 2011. Pyrethroid resistance in African anopheline mosquitoes: what are the implications for malaria control? *Trends Parasitol 27*: 91–98.
4. O'Meara WP, Bejon P, Mwangi TW, Okiro EA, Peshu N, Snow RW, Newton CR, Marsh K, 2008. Effect of a fall in malaria transmission on morbidity and mortality in Kilifi, Kenya. *Lancet 372*: 1555–1562.
5. Mathias DK, Ochomo E, Atieli F, Ombok M, Bayoh MN, Olang G, Muhia D, Kamau L, Vulule JM, Hamel MJ, Hawley WA, 2011. Spatial and temporal variation in the kdr allele L1014S in Anopheles gambiae ss and phenotypic variability in susceptibility to insecticides in Western Kenya. *Malar J 10*: 10.
6. Ochomo E, Subramaniam K, Kemei B, Rippon E, Bayoh NM, Kamau L, Atieli F, Vulule JM, Ouma C, Gimnig J, Donnelly MJ, 2015. Presence of the knockdown resistance mutation, Vgsc-1014F in Anopheles gambiae and An. arabiensis in western Kenya. *Parasit Vectors 8*: 616.
7. Ochomo E, Bayoh MN, Brogdon WG, Gimnig JE, Ouma C, Vulule JM, Walker ED, 2013. Pyrethroid resistance in Anopheles gambiae ss and Anopheles arabiensis in western

Kenya: phenotypic, metabolic and target site characterizations of three populations. Med *Vet Entomol 27*: 156–164.

8. Pauron D, Barhanin J, Amichot M, Pralavorio M, Berge JB, Lazdunski M,1989. Pyrethroid receptor in the insect Na sup+ channel: Alteration of its properties in pyrethroid-resistant flies. *Biochemistry 28*: 1673–1677.
9. Martinez-Torres D, Chandre F, Williamson MS, Darriet F, Bergé JB, Devonshire AL, Guillet P, Pasteur N, Pauron D, 1998. Molecular characterization of pyrethroid knockdown resistance (*kdr*) in the major malaria vector *Anopheles gambiae s.s. Insect Mol Biol 7*: 179–184.
10. Ranson H, Jenson B, Vulule JM, Wang X, Hemingway J, Collins FH, 2000. Identification of a point mutation in the voltage-gated sodium channel gene of Kenyan *Anopheles gambiae* associated with resistance to DDT and pyrethroids. *Insect Mol Biol 9*: 491–497
11. Chandre F, Manguin S, Brengues C, Dossou YJ, Darriet F, Diabate A, Carnevale P, Guillet P, 1999. Current distribution of a pyrethroid resistance gene (kdr) in Anopheles gambiae complex from west Africa and further evidence for reproductive isolation of the Mopti form. *Parassitologia 41*: 319–322.
12. Awolola TS, Brooke BD, Koekemoer LL, Coetzee M, 2003. Absence of the kdr mutation in the molecular 'M'form suggests different pyrethroid resistance mechanisms in the malaria vector mosquito Anopheles gambiae ss. *Trop Med Int Health 8*: 420–422.
13. Fanello C, Petrarca V, Della Torre A, Santolamazza F, Dolo G, Coulibaly M, Alloueche A, Curtis CF, Toure YT, Coluzzi M, 2003. The pyrethroid knock-down resistance gene in

the Anopheles gambiae complex in Mali and further indication of incipient speciation within An. gambiae ss. *Insect Mol Biol 12*: 241–245.

14. Yawson AE, McCall PJ, Wilson MD, Donnelly MJ, 2004. Species abundance and insecticide resistance of Anopheles gambiae in selected areas of Ghana and Burkina Faso. *Med Vet Entomol 18*: 372–377.
15. Stump AD, Atieli FK, Vulule JM, Besansky NJ, 2004. Dynamics of the pyrethroid knockdown resistance allele in western Kenyan populations of Anopheles gambiae in response to insecticide-treated bed net trials. *Am J Trop Med Hyg 70*: 591–596.
16. Santolamazza F, Calzetta M, Etang J, Barrese E, Dia I, Caccone A, Donnelly MJ, Petrarca V, Simard F, Pinto J, Della Torre A, 2008. Distribution of knock-down resistance mutations in Anopheles gambiae molecular forms in west and west-central Africa. *Malar J 7*: 74.
17. Sharp BL, Ridl FC, Govender D, Kuklinski J, Kleinschmidt I, 2007. Malaria vector control by indoor residual insecticide spraying on the tropical island of Bioko, Equatorial Guinea. *Malar J 6*: 52.
18. Balkew M, Gebre-Michael T, Hailu A, 2003. Insecticide susceptibility level of Anopheles arabiensis in two agro-development localities in Eastern Ethiopia. *Parassitologia 45*: 1-3.
19. Ndjemaï HN, Patchoké S, Atangana J, Etang J, Simard F, Bilong CF, Reimer L, Cornel A, Lanzaro GC, Fondjo E, 2009. The distribution of insecticide resistance in Anopheles gambiae sl populations from Cameroon: an update. *Trans R Soc Trop Med Hyg 103*: 1127–1138.

20. Himeidan YE, Chen H, Chandre F, Donnelly MJ, Yan G, 2007. Permethrin and DDT resistance in the malaria vector Anopheles arabiensis from Eastern Sudan. *Am J Trop Med Hyg 77*(6): 1066–1068.
21. Alout H, Yameogo B, Djogbénou LS, Chandre F, Dabiré RK, Corbel V, Cohuet A, 2014. Interplay between Plasmodium infection and resistance to insecticides in vector mosquitoes. *J Infect Dis 210*(9): 1464–1470.
22. Noor AM, Amin AA, Akhwale WS, Snow RW, 2007. Increasing coverage and decreasing inequity in insecticide-treated bed net use among rural Kenyan children. *PLoS Med 4*: e255.
23. Wanjala CL, Mbugi JP, Ototo E, Gesuge M, Afrane YA, Atieli HE, Zhou G, Githeko AK, Yan G, 2015. Pyrethroid and DDT resistance and organophosphate susceptibility among Anopheles spp. mosquitoes, western Kenya. *Emerg Infect Dis 21*: 2178.
24. Sougoufara S, Harry M, Doucouré S, Sembène PM, Sokhna C, 2016. Shift in species composition in the Anopheles gambiae complex after implementation of long-lasting insecticidal nets in Dielmo, Senegal. *Med Vet Entomol 30*: 365–368.
25. Mwangangi JM, Muturi EJ, Muriu SM, Nzovu J, Midega JT, Mbogo C, 2013. The role of Anopheles arabiensis and Anopheles coustani in indoor and outdoor malaria transmission in Taveta District, Kenya. *Parasit Vectors 6*: 114.
26. Ototo EN, Mbugi JP, Wanjala CL, Zhou G, Githeko AK, Yan G, 2015. Surveillance of malaria vector population density and biting behaviour in western Kenya. *Malar J 14*: 244.

27. Kitau J, Oxborough RM, Tungu PK, Matowo J, Malima RC, Magesa SM, Bruce J, Mosha FW, Rowland MW, 2012. Species shifts in the Anopheles gambiae complex: do LLINs successfully control Anopheles arabiensis? *PLoS one 7*: e31481.

28. Kamau L, Agai D, Matoke D, Wachira L, Gikandi G, Vulule JM, 2008. Status of insecticide susceptibility in Anopheles gambiae sensu lato and Anopheles funestus mosquitoes from western Kenya. J Insect Sci 8: 11.

29. Kawada H, Dida GO, Ohashi K, Komagata O, Kasai S, Tomita T, Sonye G, Maekawa Y, Mwatele C, Njenga SM, Mwandawiro C, 2011. Multimodal pyrethroid resistance in malaria vectors, Anopheles gambiae ss, Anopheles arabiensis, and Anopheles funestus ss in western Kenya. PloS one 6: e22574.

30. World Health Organization (WHO), 2013. Test procedures for insecticide resistance monitoring in malaria vector mosquitoes. Geneva: WHO.

31. Severson, DW, 1997. RFLP analysis of insect genomes. In The Molecular Biology of Insect Disease Vectors . Netherlands: Springer 1997: 309-320

32. Scott JA, Brogdon WG, & Collins FH, 1993. Identification of single specimens of the Anopheles gambiae complex by the polymerase chain reaction. *Am J Trop Med Hyg 49*: 520–529.

33. Bass C, Nikou D, Donnelly MJ, Williamson MS, Ranson H, Ball A, Vontas J, Field LM, 2007. Detection of knockdown resistance (kdr) mutations in Anopheles gambiae: a comparison of two new high-throughput assays with existing methods. *Malar J 6*: 111.

34. Chen H, Githeko AK, Githure JI, Mutunga J, Zhou G, Yan G, 2008. Monooxygenase levels and knockdown resistance (kdr) allele frequencies in Anopheles gambiae and Anopheles arabiensis in Kenya. *J Med Entomol 45*: 242–250.

35. Kabula B, Kisinza W, Tungu P, Ndege C, Batengana B, Kollo D, Malima R, Kafuko J, Mohamed M, Magesa S, 2014. Co-occurrence and distribution of East (L1014S) and West (L1014F) African knock-down resistance in Anopheles gambiae sensu lato population of Tanzania. *Trop Med Int Health 19*: 331–341.
36. Fettene M, Olana D, Christian RN, Koekemoer LL, Coetzee M, 2013. Insecticide resistance in Anopheles arabiensis from Ethiopia. *Afr Entomol 21*: 89–94.
37. Yewhalaw D, Van Bortel W, Denis L, Coosemans M, Duchateau L, Speybroeck N, 2010. First evidence of high knockdown resistance frequency in Anopheles arabiensis (Diptera: Culicidae) from Ethiopia. *Am J Trop Med Hyg 83*: 122–125.
38. Yewhalaw D, Wassie F, Steurbaut W, Spanoghe P, Van Bortel W, Denis L, Tessema DA, Getachew Y, Coosemans M, Duchateau L, Speybroeck N, 2011. Multiple insecticide resistance: an impediment to insecticide-based malaria vector control program. *PloS one 6*: e16066.
39. Abdalla H, Wilding CS, Nardini L, Pignatelli P, Koekemoer LL, Ranson H, Coetzee M, 2014. Insecticide resistance in Anopheles arabiensis in Sudan: temporal trends and underlying mechanisms. *Parasit Vectors 7*: 213.
40. Lynd A, Weetman D, Barbosa S, Egyir Yawson A, Mitchell S, Pinto J, Hastings I, Donnelly MJ, 2010. Field, genetic, and modeling approaches show strong positive selection acting upon an insecticide resistance mutation in Anopheles gambiae ss. *Mol Biol Evol 27*: 1117–1125
41. Amenya DA, Naguran R, Lo T, Ranson H, Spillings BL, Wood OR, Brooke BD, Coetzee M, Koekemoer LL, 2008. Over expression of a cytochrome P450 (CYP6P9) in a major

African malaria vector, Anopheles funestus, resistant to pyrethroids. *Insect Mol Biol 17*: 19–25.

42. Cuamba N, Morgan JC, Irving H, Steven A, Wondji CS, 2010. High level of pyrethroid resistance in an Anopheles funestus population of the Chokwe District in Mozambique. *PloS one 5*: e11010.
43. Mawejje HD, Wilding CS, Rippon EJ, Hughes A, Weetman D, Donnelly MJ, 2013. Insecticide resistance monitoring of field-collected Anopheles gambiae sl populations from Jinja, eastern Uganda, identifies high levels of pyrethroid resistance. *Med Vet Entomol 27*: 276–283.
44. Weetman D, Steen K, Rippon EJ, Mawejje HD, Donnelly MJ, Wilding CS, 2014. Contemporary gene flow between wild An. gambiae ss and An. arabiensis. *Parasit Vectors 7*: 345.
45. Asale A, Getachew Y, Hailesilassie W, Speybroeck N, Duchateau L, Yewhalaw D, 2014. Evaluation of the efficacy of DDT indoor residual spraying and long-lasting insecticidal nets against insecticide resistant populations of Anopheles arabiensis Patton (Diptera: Culicidae) from Ethiopia using experimental huts. *Parasit Vectors 7*: 131.
46. Mzilahowa T, Chiumia M, Mbewe RB, Uzalili VT, Luka-Banda M, Kutengule A, Mathanga DP, Ali D, Chiphwanya J, Zoya J, Mulenga S, 2016. Increasing insecticide resistance in Anopheles funestus and Anopheles arabiensis in Malawi, 2011–2015. *Malar J 15*: 563.
47. Lwetoijera DW, Harris C, Kiware SS, Dongus S, Devine GJ, McCall PJ, Majambere S, 2014. Increasing role of Anopheles funestus and Anopheles arabiensis in malaria transmission in the Kilombero Valley, Tanzania. *Malar J 13*: 331.

48. Haji KA, Khatib BO, Smith S, Ali AS, Devine GJ, Coetzee M, Majambere S, 2013. Challenges for malaria elimination in Zanzibar: pyrethroid resistance in malaria vectors and poor performance of long-lasting insecticide nets. *Parasit Vectors 6*: 82.

49. Jones CM, Haji KA, Khatib BO, Bagi J, Mcha J, Devine GJ, Daley M, Kabula B, Ali AS, Majambere S, Ranson H, 2013. The dynamics of pyrethroid resistance in Anopheles arabiensis from Zanzibar and an assessment of the underlying genetic basis. *Parasit Vectors 6*: 343.

50. Müller GC, Beier JC, Traore SF, Toure MB, Traore MM, Bah S, Doumbia S, Schlein Y, 2010. Field experiments of Anopheles gambiae attraction to local fruits/seedpods and flowering plants in Mali to optimize strategies for malaria vector control in Africa using attractive toxic sugar bait methods. *Malar J 9*: 262.

51. Kroeger A, Horstick O, Riedl C, Kaiser A, Becker N, 1995. The potential for malaria control with the biological larvicide Bacillus thuringiensis israelensis (Bti) in Peru and Ecuador. *Acta Trop 60*: 47–57.

52. Fillinger U, Lindsay SW, 2006. Suppression of exposure to malaria vectors by an order of magnitude using microbial larvicides in rural Kenya. *Trop Med Int Health 11*: 1629–1642.

CHAPTER 3

Landscape Genetics of *Anopheles gambiae s.s.* and *An. arabiensis* in Kenya

Abstract

Anopheles gambiae s.s. and *An. arabiensis* are major malaria vectors in sub-Saharan Africa. Knowledge of how geographical factors drive the dispersal of malaria vectors can help in insecticide resistance management efforts, as well as planning effective vector control interventions. Dispersal patterns can be inferred from measuring genetic relatedness among populations or inferred gene flow. Here, we collect *Anopheles gambiae s.s.,* and *An. arabiensis* across Kenya and genotype specimens at nine microsatellite loci to measure relatedness. We test associations between pairwise genetic distance and ecological factors hypothesized to influence dispersal using linear mixed effects models. We found that for *An. gambiae s.s.,* high population densities are primarily associated with increased gene flow. High annual precipitation was also associated with increased gene flow. For *An. arabiensis,* we found that high temperatures and low vegetation indices were associated with increased gene flow in Western Kenya. In Eastern Kenya, low vegetation indices were associated with increased gene flow for *An. arabiensis*. By identifying the factors that drive malaria vector dispersal, we improve our understanding of the areas which are susceptible to the invasion of insecticide resistant mosquitoes. In addition, this knowledge can be used to plan effective vector control interventions.

Introduction

Anopheles gambiae s.s. and *An. arabiensis* mosquitoes are major malaria vectors, as well as the most widespread mosquitoes of the *An. gambiae* complex.[1] While the species commonly occupy similar ecological niches, *Anopheles gambiae s.s.* are generally associated with more humid environments, whereas *An. arabiensis* have a higher tolerance for drier environments.[2,3] Another notable difference between the two species is that *An. gambiae s.s.* are highly anthropophagic,[4,5] whereas *An. arabiensis* are more catholic in their feeding behavior.[6] Since mosquitoes primarily disperse to seek blood meals and oviposit,[7] we expect that these differences in habitat and feeding preferences result in ecological variables differentially driving the dispersal patterns of these two malaria vectors. These differences may create a complex system influencing malaria parasite spread. Knowledge of how geographical factors influence the dispersal of malaria vectors can help in efforts to contain insecticide resistance, planning effective vector control interventions, and identifying potential areas susceptible to parasite re-introduction from infected mosquitoes following antimalarial interventions.[8]

Organism dispersal patterns and population connectedness can be inferred from measuring genetic relatedness among populations or gene flow. While studies have reported genetic differentiation between *An. arabiensis* populations, neither physical barriers nor geographic distance has been identified as factors responsible for *An. arabiensis* population structuring.[9-12] Likewise, geographical distance alone does not appear to be a barrier to gene flow among populations of *An. gambiae s.s.*,[12] as Lehmann et al.[13] found high gene flow between populations in Kenya (East Africa) and Senegal (West Africa). However, *An. gambiae s.s.* populations were found to be highly differentiated between Western Kenya and coastal Kenya.[12] The Eastern arm of the Rift Valley, which bisects Kenya, has been speculated to be the cause of

genetic differentiation in *An. gambiae s.s.* populations due to its low temperatures and arid conditions making it inhospitable to agriculture, and as such, lacks human settlements.[12,14] Alternatively, *An. gambiae s.s.* structuring is thought to be largely influenced by environmental heterogeneity.[15] Since the eastern arm of the Rift Valley is characterized by low temperatures, low precipitation, as well as low human population density,[12] we cannot dismiss any of these factors for principally driving population structure of *An. gambiae s.s.*. Thus, here, we disentangle confounding environmental and landscape factors to test the hypothesis that low human population densities primarily restrict gene flow between populations of *An. gambiae s.s.* in Kenya. In addition, we test the hypothesis that alternative factors related to climate and landscape primarily restrict gene flow among *An. arabiensis*, the more zoophagic vector.

Using a landscape genetics approach allows us to rigorously test the impacts of ecological variables on the dispersals of organisms through the inference of population movement from the distribution of genetic markers and the measurement of ecological factors hypothesized to promote or prevent dispersal.[16-20] Thus, here, we use a landscape genetics approach to test two hypotheses related to malaria vector dispersal, measured by gene flow, in Kenya: 1) Low human population densities provide a significant barrier to gene flow for *An. gambiae s.s.*, but not for *An. arabiensis*; 2) Climate and landscape factors primarily restrict gene flow for *An. arabiensis.* By testing relationships between population genetic structure of malaria vectors and ecological factors, we can parse out confounding factors and determine the importance of key variables influencing malaria vector dispersal.[8,16]

Methods

Sample Collection

An. gambiae s.l. larvae were collected between May 2014 and January 2015 from fourteen sites within three distinct geographical areas in Kenya: Western Kenya, Rift Valley, and Coastal Kenya (Table 3.1). Larvae were collected using a standard mosquito dipper. No more than five larvae were collected per habitat to reduce bias. Collected larvae were stored in 100% ethanol until DNA purification.

DNA extraction and species identification

Genomic DNA was extracted using standard ethanol extraction procedures with phenol:chloroform.[21] DNA was eluted into 20 µl of TE buffer. Then, DNA was quantified using a NanoDrop 8000 Spectrophotomer and diluted to a concentration of 1µg/1µl sterile water. We identified *An. arabiensis* and *An. gambiae* s.s. species within the *An. gambiae* s.l. complex using a ribosomal DNA polymerase chain reaction (PCR) assay.[22]

Microsatellite genotyping

Nine microsatellite loci were selected for genotyping *An. gambiae s.s., An. arabiensis, P. falciparum* based on evidence of polymorphism in previous studies, reliable amplification, and having an even distribution across chromosomes (Table 3.2).[24] We used the M13 tailed primer method to fluorescently label our primers.[25] Amplification was conducted in a total volume of 10 µl with 5 µl of 2x DreamTaq Green PCR Master Mix (Thermo Fisher, USA), 0.5 µl of 10µM primer (forward primer with M13 tail), and 1 µl of DNA template. Thermocycling conditions for *An. gambiae s.s.* and *An. arabiensis* were as follows: initial denature of 94°C for 3 min, followed

by 35 amplification cycles of 94°C for 30 sec, annealing temperature (Table 3.2) for 30 sec, and 72°C for 45 sec, and then a final extension of 72°C for 6 min. PCR products were analyzed on an automated 4300 DNA analyzer (Li-Cor, Lincoln, NE), and alleles were quantified with the use of Gene ImagIR 4.33 software (Li-Cor).

Population genetic analysis

We tested for deviation from Hardy-Weinberg equilibrium and allelic richness at each study site in Arlequin.[26] To estimate population structure, we used a model-based approach and an exploratory approach for population clustering. First, we used the R package *adegenet* to do a standard principal component analysis (PCA) to determine genetic relationships among populations.[28] Second, we estimated population structure using STRUCTURE v. 2.3.4, which uses a Bayesian algorithm to group samples into genetically distinct clusters.[29] We tested K=1-7, with six replicates for each K-level, an initial burn-in of 200,000, and then 600,000 Monte Carlo Markov Chain iterations. The program was run using an admixture model. ΔK was used to detect the number of K (clusters).[30] The output data for the best estimate of K were analyzed using CLUMMP to calculate the mean cluster membership coefficients across multiple runs.[31] The CLUMMP output was visualized using DISTRUCT.[32] Pairwise F_{ST} values were calculated in the R package *adegenet* and bootstrapping was done to determine statistical significance at p-value <0.05, with a Bonferonni correction.[28]

Landscape genetic analysis

Since population sampling was not evenly spread across the country, we analyzed Eastern Kenya separately from the rest of Kenya due to the biases uneven sampling can

introduce in analysis.[32] Genetic distance between populations was calculated as the proportion of shared alleles (D_{PS}) measured in the R package PopGenReport.[33,34] When testing the effects of environmental factors on gene flow between populations, between-site characteristics are of the greatest concern.[35] Hence, resistance surfaces were created based on factors hypothesized to prevent or promote gene flow. We created raster files in ArcGIS 10 using climate data from WorldClim (BIO1 and BIO12),[36] vegetation index data from MODIS (MCD12Q1),[37,38] topographic wetness index data from NERC Environmental Information Data Centre,[39] population density from WorldPop (www.worldpop.org.uk), and roads from the Global Roads Open Access Data Set (Table 3.3). All raster files were resampled to a grain size of 1 km. Landscape distance among all pairs of sites was measured using electrical circuit theory in PopGenReport.[34,40] Circuit theory incorporates multiple pathways into the analysis, providing an advantage over the least cost path method. A key shortcoming of landscape genetics has been identified in the methods used to assign resistance values to variables (e.g. forest cover or elevation).[41] For continuous variables, such as precipitation, raw elevation numbers can be used to assign resistance values, but this method assumes a linear response to precipitation. For non-continuous variables, assigning resistance values has most commonly relied upon expert opinion, but this approach introduces biases.[41] To overcome these shortcomings, we used ResistanceGA a package in R to optimize resistance surfaces to our genetic data.[42,43] ResistanceGA uses a genetic algorithm to unbiasedly optimize resistance surfaces. Lastly, linear mixed effects models with the maximum likelihood population effects were used to fit optimized resistance surfaces to genetic data.[44] Akaike information criterion with a penalty for extra parameters (AICc) was used as the measure of model fitness to genetic data.

Results

Population genetic analysis

***An. gambiae s.s.*:** Mean allelic richness (A_R) and expected heterozygosity (H_E) were highest in the four highland sites (A_R: 4.93-6.12; H_E: 0.60-0.68) compared to the three lowland sites (A_R: 4.51-4.80; H_E: 0.42-0.55) (Table 3.4). Mean observed heterozygosity (H_O) was consistently lower than H_E across all populations (Table 3.4).We identified two clusters consistently across six runs in STRUCTURE (Figure 3.1C). PCA analysis of the first and second components (explaining 42.6 and 31.4% of the variance, respectively), revealed a similar clustering pattern as the STRUCTURE analysis (Figure 1B). The four highland populations clustered together along with a lowland population ('Kan'). Whereas two of the westernmost lowland populations clustered together ('Por' and 'Hom') (Figure 3.1A). In addition, pairwise F_{ST} values were generally found to be higher among lowland sites, as well as between lowland and highland sites, than among highland sites (Table 3.5).

***An. arabiensis*:** Mean allelic richness (A_R) ranged from 4.67 at 'Jar' in Eastern Kenya to 8.42 at 'Kan' in Western Kenya. Expected heterozygosity (H_E) was highest at 'Nak' in the Rift Valley (0.73) and lowest at 'Kak' in Eastern Kenya (0.44) (Table 3.4). Mean observed heterozygosity (H_O) was consistently lower than H_E across all populations (Table 3.4). We identified four clusters consistently across six runs in STRUCTURE (Figure 3.2C). PCA analysis of the first and second components (explaining 39.8 and 26.3% of the variance, respectively), also revealed clustering pattern similar to that observed in STRUCTURE (Figure 3.2B). Five of the six Western Kenya populations clustered together, with 'Miw', the easternmost site in Western Kenya, as the exception (Figure 3.2A). Three sites in the Rift Valley ('Nak', 'Gil', and 'Nai') clustered together. Two coastal Kenya populations clustered together, but the other two

clustered with Western Kenya and Rift Valley populations. The majority of pairwise F_{ST} values were significant with no clear trends across regions (Table 3.6).

Landscape genetic anlaysis

Of the six predictor variables tested, human population density, average temperature, and annual precipitation were significant in explaining population genetic structure of *An. gambiae s.s.* populations. For *An. gambiae s.s.*, low human population density (AICc=0), high temperature (AICc=3.31), and low precipitation (AICc=3.54) were associated with an increase in landscape resistance to gene flow (Figure 3.3). For *An. arabiensis*, average temperature (AICc=0) and vegetation index (AICc=1.78) were significant in explaining population genetic structure in Western Kenya, while only vegetation index (AICc=0) was significant in explaining population genetic structure in Eastern Kenya (Table 3.7). Low average temperature and high vegetation index were associated with an increase in landscape resistance to gene flow for *An. gambiae s.s.* (Figure 3.4).

Testing the fit of all combinations of the top single surface models revealed that a combination of all three variables (human population density, annual precipitation, and average temperature) had the highest model weight (0.24), followed by a combination of human population density and precipitation (0.23), and human population density and temperature (0.21) (Table 3.9). In the top model, human population density had the highest percent contribution (71%), followed by precipitation (18%), and temperature (11%) (Figure 3.5). In the second highest ranked model, human population density had the highest contribution (86%), followed by precipitation (14%). Finally in the third highest ranked model, human population density had the highest contribution (96%), followed by temperature (4%). Overall, human

population density and precipitation had the highest contributions in the multi surface models for *An. gambiae s.s.*. Testing the fit of the combinations of the top single surface models for *An. arabiensis* in Western Kenya revealed that combining the temperature and vegetation index surfaces did not improve the landscape resistance model (Table 3.8).

Discussion

Using a landscape genetics framework to test hypotheses related to whether climatic, landscape, or social factors predominantly influence the population structures of *An. gambiae s.s.,* and *An. arabiensis*, we found that contrasting factors influence the population structures of the three species. For *An. gambiae s.s.,* both social (low human population distribution) and climatic (low precipitation and high temperature) factors were the most important factors shaping population structure. While, for *An. arabiensis,* landscape (high vegetation index) and climatic (low temperature) factors were the most important factors for population structuring.

An. gambiae s.s. and *An. arabiensis* are primary vectors of human malaria in sub-Saharan Africa, a disease responsible for 438,000 deaths worldwide annually, with around 90% occurring in Africa.[45] Though the two species commonly co-occur, *An. arabiensis* mosquitoes are more well-adept to disturbed,[46] hot,[47,48] and arid environments.[46] Additionally, *An. arabiensis* are more catholic in their feeding preferences taking blood meals from both human and non-human hosts, especially cattle.[49-51] Therefore, our findings that gene flow between *An. gambiae s.s.* populations is primarily restricted by low human population densities, high temperatures, and low precipitation are consistent with the mosquito's biology. Likewise, that gene flow between *An. arabiensis* populations is primarily restricted by low temperatures and high vegetation index

is also consistent the known ecology of *An. arabienisis*, a species which is known to thrive in arid conditions[46] and feed on non-human hosts.[49-51]

The lack of human settlement on the high plateaus of the eastern arm of the Great Rift Valley, which bisects Kenya, was thought to explain why populations of *An. gambiae s.s.* between Western Kenya and Eastern Kenya were much more distinct than between Western Kenya and Senegal in Western Africa, despite that the two countries are separated by more than 5000 km.[12] Whereas, Western Kenya and coastal Kenya populations are only 700 km apart. [12] This hypothesis was formulated since unlike the eastern arm of the rift, human agricultural activity occurs in a broad band across the area between Senegal and Western Kenya. [12] We tested the long-standing hypothesis that low human population densities between *An. gambiae s.s.* populations provide a barrier to gene flow, and we provide evidence in support of this hypothesis. Moreover, we demonstrate that low human population densities influences *An. gambiae s.s.* structure on a smaller spatial scale than previously suggested, working among populations within Western Kenya. In addition, *An. gambiae s.s.* structuring was thought to be largely influenced by environmental heterogeneity.[15] We provide evidence in support of this hypothesis, as well, though human population density was found to be the primary explanatory variable.

While populations of *An. gambiae s.s.* across the eastern arm of the rift have been found to be highly genetically differentiated, no such differentiation has been found between populations of *An. arabiensis*. [12] This result suggested that areas of low human population densities do not provide a barrier to gene flow for *An. arabiensis*. [12] In addition, no relationship between genetic differentiation and geographical distance was found for *An. arabiensis*, [12] posing the possibility that factors aside from distance and human population densities impact gene flow,

such as climate and geographical features.[1] We provide evidence in support of this conjecture, as forests (high vegetation index) were an important barrier to gene flow in Western and Eastern Kenya and low temperatures were important for population structuring in Western Kenya. While low temperatures were not significant for population structuring in Eastern Kenya, the temperature range was very small in Eastern Kenya (23.5 – 26.7°C), compared to Western Kenya (10.8 – 26.6°C), which may help to explain why low temperatures were only important in driving population structure in Western Kenya.

Overall, we detected significant genetic structuring between populations of *An. gambiae s.s.* and *An. arabiensis* which suggests that populations are more genetically isolated than observed in previous studies in Kenya.[12-14] This trend of increasing fragmentation could be caused by the increase in ITN coverage since 2004 decreasing mosquito abundances.[52] Additionally, we found that genetic structuring for both species did not conform to the isolation-by-distance model, a finding consistent with previous studies on *An. gambiae s.s.* [12] and *An. arabiensis*.[53] In addition to geographical gene flow barriers contributing to a lack of isolation-by-distance, genetic differentiation may also be caused by historical factors.[53] This phenomenon may help to explain the close clustering of certain Eastern Kenya populations with some Western Kenya populations in *An. arabiensis.* Low F_{ST} values between populations across the eastern Rift Valley may be attributed to a range expansion, facilitated by the expansion of human settlement and agriculture, rather than large amounts of contemporary gene flow.[54] Moreover, large amounts of contemporary gene flow across the Rift Valley seems particularly unlikely given that mosquitoes rarely disperse more than a few kilometers throughout a lifetime.[7]

While malaria prevention is primarily mediated through the use of insecticide-impregnated bednets (ITNs), the overuse of these nets has increased resistance in mosquito

vectors.[55] Understanding the population genetic structure of these vectors is useful for tracking the spread of resistant genes and in designing alternative methods of malaria prevention.[53] Additionally, in response to the scale-up in distribution of insecticide treated bednets (ITNs) throughout Kenya and sub-Saharan Africa, *An. arabiensis* have increased in proportional abundance, replacing the more historically dominant and anthropogenic species *Anopheles gambiae* s.s. and *An. funestus*.[56-59] *Anopheles arabiensis* have been able to persist and thrive despite intensive antimalarial interventions due to their catholic feeding behavior, as well as their tendencies to more commonly feed outdoors where people are not under the protection of ITNs[46] and rest outdoors where they are not affected by indoor residual spraying (IRS).[56,57,60] Since *An. arabiensis* are not as vulnerable to traditional antimalarial vector control interventions, they are likely to play a crucial role in maintaining residual malaria transmission in sub-Saharan Africa even as countries approach pre-elimination and elimination.[56,59] Thus, understanding how ecological features influence *An. arabiensis* dispersal will likely become increasingly important to interrupting residual malaria transmission.

Using a landscape genetics approach to identify important dispersal corridors has great potential for mitigating disease risk.[8,16] In this study, we found that corridors for *An. gambiae s.s.* in Kenya are most likely to be areas of high human population density and high precipitation, whereas corridors for *An. arabiensis* are likely to be areas of low vegetation index and high temperatures. This knowledge is important to improving insecticide resistance management, as well as how parasites can become re-introduced to an area following public health interventions.

Table 3.1. Locality information and sample size by species used for analyses.

					Sample size (n)	
Region	Site	Longitude	Latitude	Elevation	*An. gambiae* s.s.	*An. arabiensis*
Western Kenya Lowlands	Por	34.012	-0.114	1139	58	58
	Ken	34.629	-0.385	1150	0	57
	Hom	34.465	-0.543	1184	42	58
	Kan	34.606	-0.122	1214	60	45
	Miw	34.943	-0.129	1161	0	58
Western Kenya Highlands	Emu	34.618	0.023	1520	51	0
	May	34.578	0.51	0.513	52	0
	Kai	34.899	0.156	1647	29	0
	Kam	34.807	0.571	1487	27	33
Rift Valley	Mar	36.018	0.484	1004	0	58
	Kab	35.663	0.498	1150	0	58
	Nak	35.945	-0.334	2163	0	18
	Gil	35.911	-0.495	2009	0	24
	Nai	36.459	-0.703	2036	0	14
Eastern Kenya	Mus	39.501	-4.394	15	0	36
	Jun	39.742	-3.852	66	0	45
	Jar	39.736	-3.616	16	0	30
	Kak	40.039	-3.170	14	0	57

Table 3.2. Microsatellite markers used for *An. gambiae s.l.* (Zheng et al 1996) genotyping.

An. gambiae s.l.		
Locus	Chromosome	Annealing (°C)
AG2H143	2L	60
AG2H46	2R	55
45C1	3L	60
AG3H577	3L	60
33C1	3R	55
Ag3H249	3R	60
AGXH7	X	57
AGXH99	X	50
1D1	X	50

Table 3.3. Predictor variables used for landscape genetic analysis.

Category	Variable	Source
Climate	Average Temperature	WorldClim BIO1
	Annual Precipitation	WorldClim BIO12
Social	Human Population Density	Worldpop
	Distance to Roads	Global Roads Open Access Data Set, version 1
Landscape	Vegetation Index	NASA MCD12Q1
	Topographic Wetness Index	NERC Environmental Information Data Centre

Table 3.4. Genetic diversity indices for *Anopheles gambiae s.s.* and *An. arabiensis* in Kenya.

A_R is allelic richness; H_O is observed heterozygosity; H_E is expected heterozygosity; F_{IS} is the fixation index.

	An. gambiae s.s.				*An. arabiensis*			
Site	**A_R**	**H_O**	**H_E**	**F_{IS}**	**A_R**	**H_O**	**H_E**	**F_{IS}**
Por	4.68	0.26	0.42	0.35	7.62	0.29	0.64	0.56
Hom	4.51	0.30	0.48	0.37	6.52	0.22	0.53	0.45
Ken	–	–	–	–	6.44	0.33	0.54	0.46
Kan	4.80	0.42	0.55	0.24	8.42	0.37	0.63	0.61
Kai	4.93	0.30	0.60	0.51	–	–	–	–
Kam	5.76	0.35	0.68	0.49	7.69	0.41	0.62	0.40
Emu	6.47	0.38	0.65	0.38	–	–	–	–
May	6.12	0.40	0.67	0.40	–	–	–	–
Kab	–	–	–	–	6.91	0.30	0.55	0.57
Mar	–	–	–	–	6.56	0.40	0.60	0.48
Nak	–	–	–	–	6.63	0.50	0.73	-0.11
Gil	–	–	–	–	6.15	0.49	0.63	0.54
Nai	–	–	–	–	6.22	0.62	0.68	0.01
Kak	–	–	–	–	5.58	0.25	0.44	0.78
Jar	–	–	–	–	4.67	0.23	0.58	1.00
Jun	–	–	–	–	7.67	0.42	0.71	0.71
Mus	–	–	–	–	6.10	0.27	0.66	0.84

Table 3.5. Pairwise F_{ST} values for *An. gambiae s.s.* populations in Western Kenya. Lower triangle indicates pairwise F_{ST} value and upper triangle indicates statistical significance (P<0.05).

	Por	Hom	Kan	Kai	Kam	Emu	May
Por	0	*	*	*	*	*	*
Hom	0.157	0	*	*	*	*	*
Kan	0.228	0.15	0	*			*
Kai	0.3	0.148	0.104	0			*
Kam	0.192	0.092	0.07	0.041	0		
Emu	0.207	0.067	0.06	0.035	0.031	0	
May	0.19	0.109	0.093	0.085	0.036	0.043	0

Table 3.6. Pairwise F_{ST} values for *An. arabiensis* populations in Kenya. Lower triangle indicates pairwise F_{ST} value and upper triangle indicates statistical significance (P<0.05).

	Por	Hom	Ken	Kan	Miw	Kam	Kab	Mar	Nak	Gil	Nai	Kak	Jar	Jun	Mus
Por	0						*					*	*		*
Hom	0.08	0		*	*	*	*	*	*	*		*		*	*
Ken	0.06	0.14	0			*	*		*		*	*	*	*	*
Kan	0.02	0.09	0.04	0	*			*	*		*	*	*		*
Miw	0.19	0.25	0.17	0.18	0	*	*	*	*	*	*	*	*	*	*
Kam	0.03	0.07	0.06	0.02	0.14	0	*		*	*	*	*	*		*
Kab	0.12	0.18	0.13	0.09	0.25	0.13	0	*	*	*	*	*	*	*	*
Mar	0.25	0.24	0.19	0.18	0.07	0.14	0.25	0			*	*			*
Nak	0.10	0.16	0.10	0.10	0.20	0.09	0.18	0.17	0			*			*
Gil	0.17	0.22	0.16	0.18	0.23	0.16	0.25	0.19	0.04	0		*	*		*
Nai	0.17	0.20	0.17	0.20	0.21	0.16	0.25	0.19	0.03	0.05	0	*	*		*
Kak	0.27	0.11	0.25	0.25	0.11	0.21	0.28	0.18	0.24	0.31	0.33	0	*	*	*
Jar	0.16	0.19	0.20	0.18	0.22	0.21	0.25	0.18	0.13	0.19	0.17	0.27	0	*	*
Jun	0.10	0.18	0.12	0.11	0.18	0.09	0.17	0.15	0.08	0.14	0.11	0.23	0.14	0	*
Mus	0.18	0.20	0.20	0.18	0.20	0.14	0.23	0.20	0.13	0.17	0.16	0.26	0.09	0.14	0

Table 3.7. Top single surface models for *An. gambiae s.l.* in Kenya. Models within 4 AICc points of the top model are displayed. Coefficient (Coeff.) indicates the relationship between the variable and landscape resistance to gene flow.

	Coeff.	LL	AICc	ΔAIcc
***An. gambiae s.s.* (Western Kenya)**				
1) Human Population Density	-	33.91	-60.82	0.00
2) Average Temperature	+	32.25	-57.51	3.31
3) Annual Precipitation	-	32.14	-57.28	3.54
***An. arabiensis* (Western Kenya)**				
1) Average Temperature	-	93.31	-18.129	0.00
2) Vegetation Index	+	92.42	-179.35	1.78
***An. arabiensis* (Eastern Kenya)**				
1) Vegetation Index	+	25.08	-34.16	0

Table 3.8. Multi surface landscape resistance models for *An. gambiae s.l.* in Kenya

Surface	LL	AICc	ΔAICc	weight
***An. gambiae s.s.* (Western Kenya)**				
Population + Precipitation + Temperature	34.24	-61.48	0	0.24
Population + Precipitation	34.21	-61.42	0.06	0.23
Population + Temperature	34.13	-61.25	0.22	0.21
Population	33.91	-60.82	0.66	0.17
Precipitation + Temperature	33.06	-59.12	2.36	0.07
Temperature	32.25	-57.51	3.97	0.03
Precipitation	32.14	-57.28	4.20	0.03
Geographic Distance	31.63	-56.27	5.21	0.02
***An. arabiensis* (Western Kenya)**				
Temperature	93.31	-181.13	0	0.69
Vegetation Index	92.42	-179.35	1.78	0.28
Geographic Distance	89.27	-173.04	8.09	0.01
Vegetation Index + Temperature	89.13	-170.26	8.37	0.01

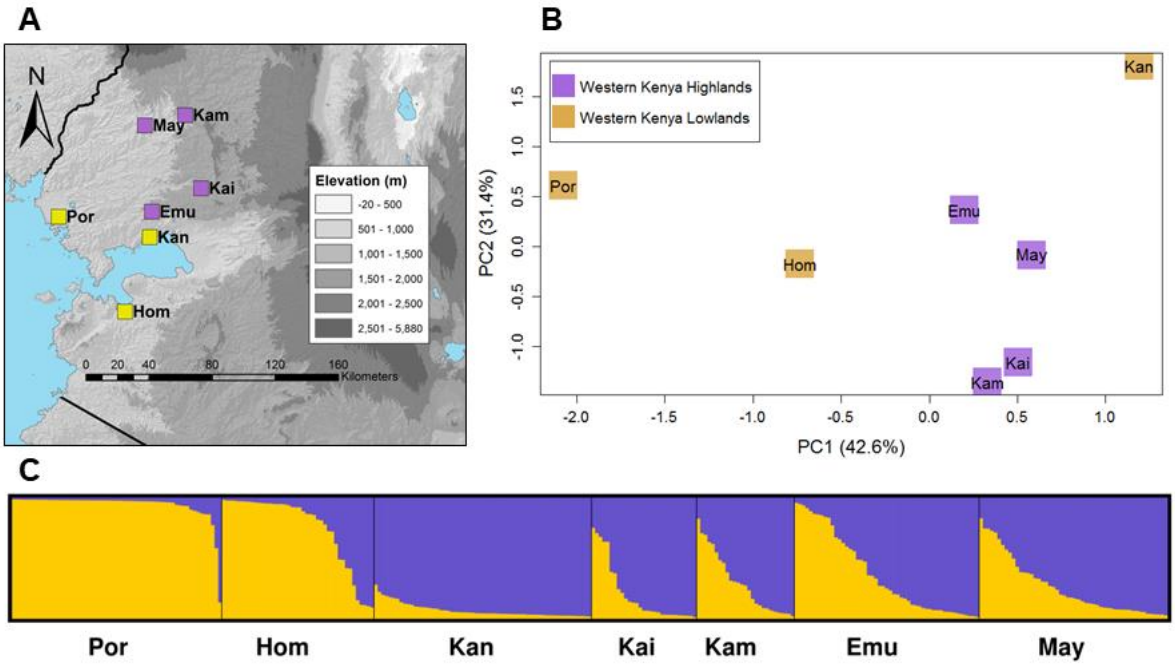


Figure 3.1. Population structure of *An. gambiae s.s.* in Western Kenya A) Map of *An. gambiae s.s.* sites. B) Principal component analysis (PCA) of genetic relatedness. C) Bayesian population structuring analysis from STRUCTURE.

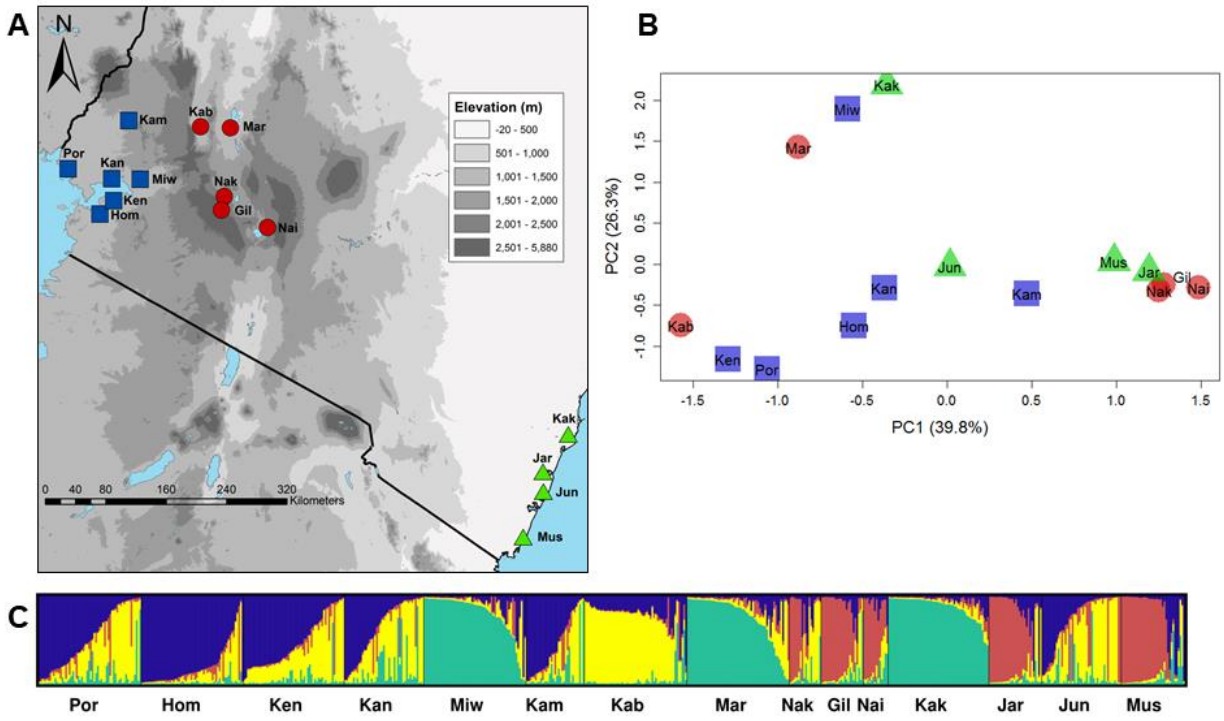


Figure 3.2. Population structure of *An. arabiensis* in Kenya. A) Map of *An. gambiae s.s.* sites. B) Principal component analysis (PCA) of genetic relatedness. C) Bayesian population structuring analysis from STRUCTURE.

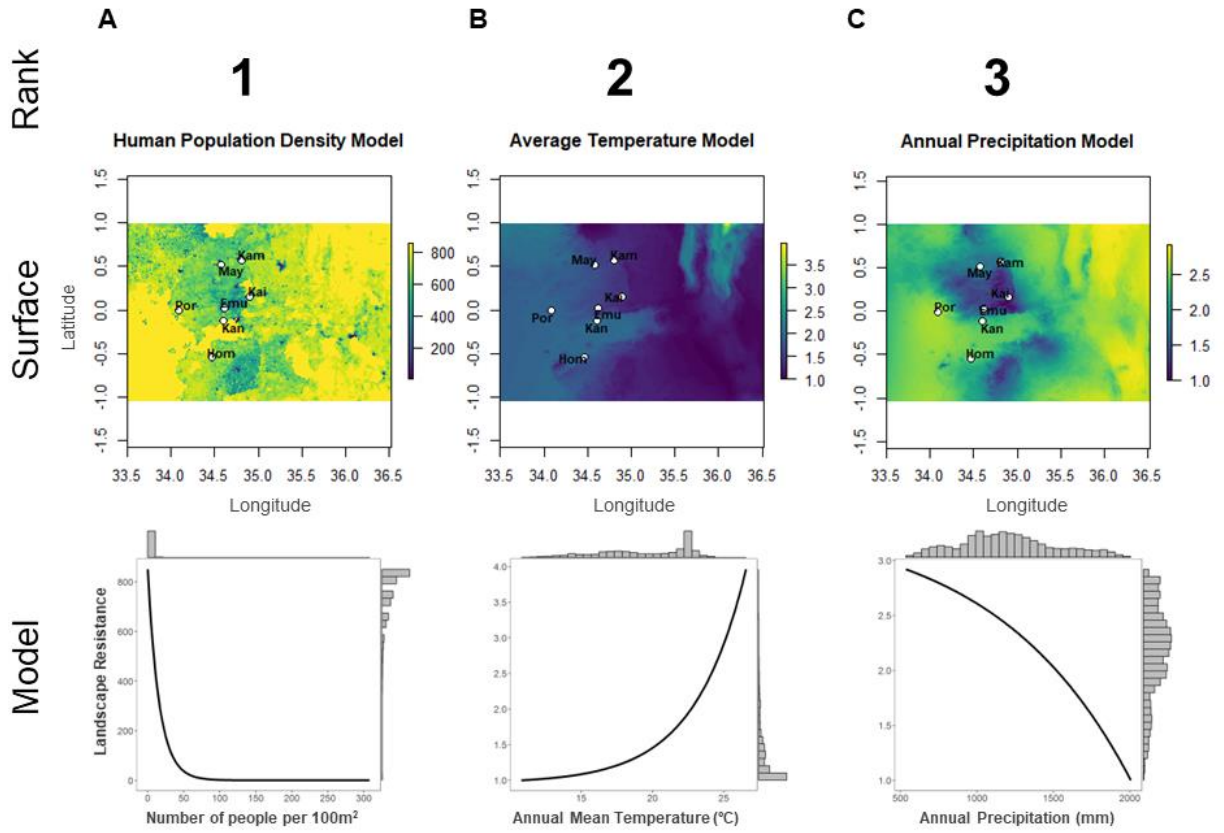


Figure 3.3. Top single surface models for *An. gambiae s.s.* in Western Kenya. A) Highest performing model. B) Second highest performing model. C) Third highest performing model. Surface is the landscape resistance raster map for the top performing surfaces following the best fit transformation. A higher number (yellow) indicates higher landscape resistance to gene flow. Model is the plot showing the transformation of the original, raw data (x-xis) to the optimized landscape resistance to gene flow value (y-axis).

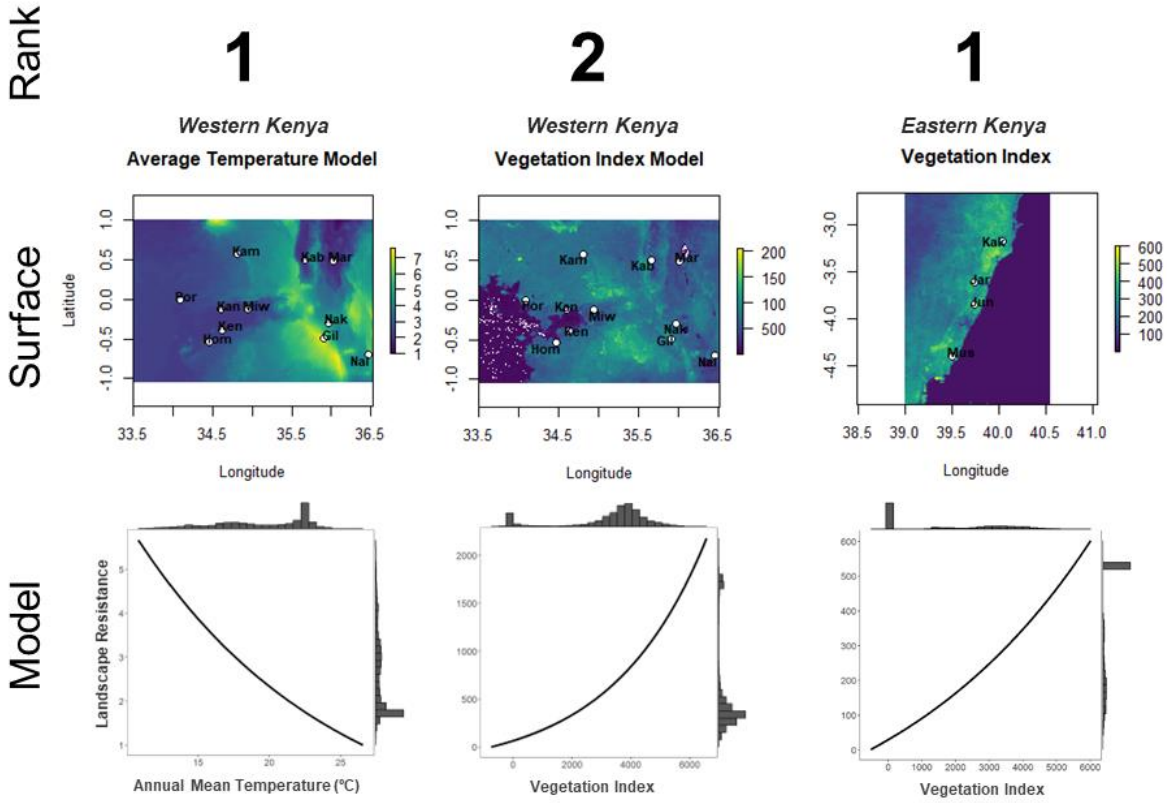
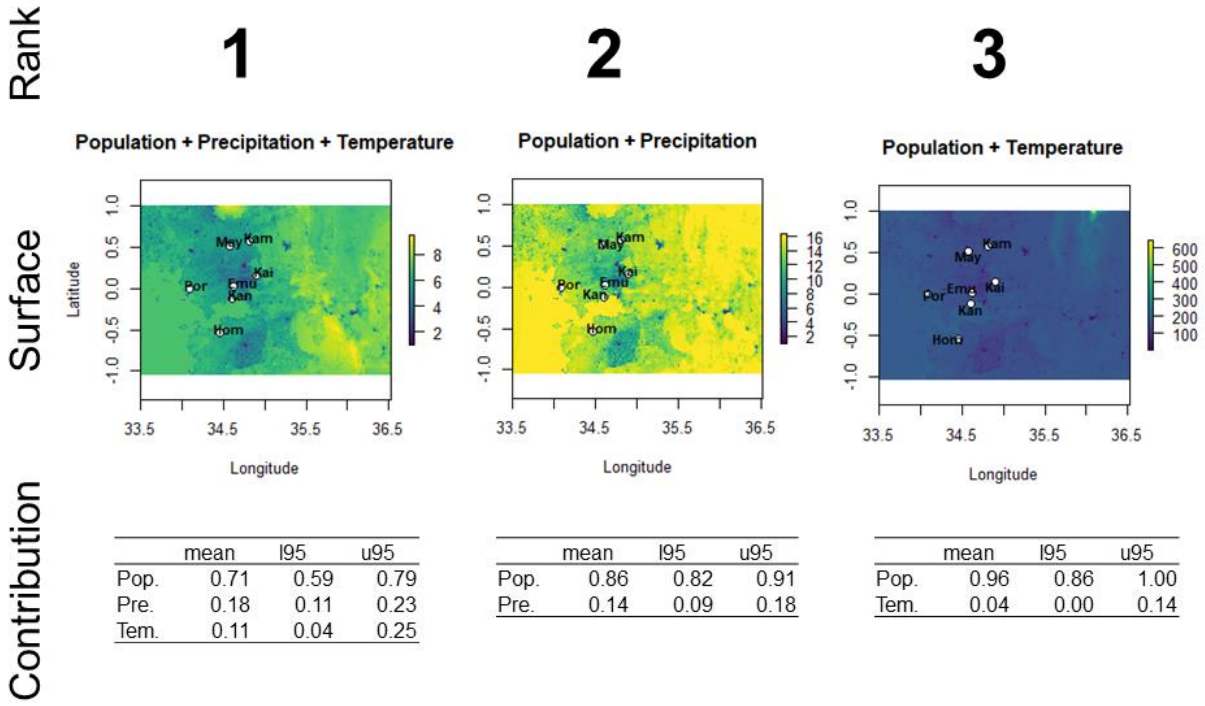


Figure 3.4. Top single surface models for *An. arabiensis* in Western and Eastern Kenya. A) Highest performing model for Western Kenya. B) Second highest performing model for Western Kenya. C) Highest performing model for Eastern Kenya. Surface is the landscape resistance raster map for the top performing surfaces following the best fit transformation. A higher number (yellow) indicates higher landscape resistance to gene flow. Model is the plot showing the transformation of the original, raw data (x-xis) to the optimized landscape resistance to gene flow value (y-axis).



	mean	l95	u95
Pop.	0.71	0.59	0.79
Pre.	0.18	0.11	0.23
Tem.	0.11	0.04	0.25

	mean	l95	u95
Pop.	0.86	0.82	0.91
Pre.	0.14	0.09	0.18

	mean	l95	u95
Pop.	0.96	0.86	1.00
Tem.	0.04	0.00	0.14

Figure 3.5. Top multi surface models for *An. gambiae s.s.* in Western Kenya A) Highest performing multi surface model. B) Second highest performing multi surface model. C) Third highest performing multi surface model. Surfaces are raster map for the top performing surfaces following the best fit transformation. Contribution is the mean proportion contribution of individual surfaces to the multi surface model with lower and upper 95% confidence intervals.

References

1. Coluzzi M, Sabatini A, Petrarca V, Di Deco MA, 1979. Chromosomal differentiation and adaptation to human environments in the *Anopheles gambiae* complex. *Trans R Soc Trop Med Hyg 73*(5):483-97.
2. Wirtz RA, Burkot TR, 1991. Detection of malarial parasites in mosquitoes. In Advances in disease vector research. New York, NY: *Springer,* 77-106.
3. Trape JF, Rogier C, 1996. Combating malaria morbidity and mortality by reducing transmission. *Parasitol Today 12*(6):236-40.
4. Mala AO, et al, 2011. Plasmodium falciparum transmission and aridity: a Kenyan experience from the dry lands of Baringo and its implications for *Anopheles arabiensis* control. *Malar J 10*(1): 121.
5. Macdonald G, 1957. The Epidemiology and Control of Malaria. London, United Kingdom: *Oxford University Press*, 63–99.
6. Gillies M, de Meillon B, 1968. The Anophelini of Africa South of the Sahara (Ethiopian zoogeographical region). Johannesburg, South Africa: *South African Institute of Medical Research*, *2nd Ed*.
7. Service MW. Mosquito (Diptera: Culicidae) dispersal—the long and short of it, 1997. *J Med Entomol 34*(6): 579-88.
8. Hemming-Schroeder E, Lo E, Salazar C, Puente S, Yan G, 2018. Landscape Genetics: A Toolbox for Studying Vector-Borne Diseases. *Front Ecol Evol 6*: 21.

9. Ramsdale CD, Fontaine RE, World Health Organization, 1970. Ecological investigations of *Anopheles gambiae* and *Anopheles funestus*. Geneva: *World Health Organization*.
10. Charlwood JD, Vij R, Billingsley PF, 2000. Dry season refugia of malaria-transmitting mosquitoes in a dry savannah zone of east Africa. *Am J Trop Med Hyg 62*(6):726-32.
11. Aniedu I, 1997. Dynamics of malaria transmission near two permanent breeding sites in Baringo district, Kenya. *Indian J Med Res 105*: 206-11.
12. Kamau L, Mukabana WR, Hawley WA, Lehmann T, Irungu LW, Orago AAS, Collings FH, 1999. Analysis of genetic variability in *Anopheles arabiensis* and *Anopheles gambiae* using microsatellite loci. *Insect Mol Biol 8*(2): 287-97.
13. Lehmann T, Hawley WA, Kamau L, Fontenille D, Simard F, Collins FH, 1996. Genetic differentiation of *Anopheles gambiae* populations from East and West Africa: comparison of microsatellite and allozyme loci. *Heredity 77*(2): 192.
14. Kamau L, Lehmann T, Hawley WA, Orago AS, Collins FH, 1998. Microgeographic genetic differentiation of *Anopheles gambiae* mosquitoes from Asembo Bay, western Kenya: a comparison with Kilifi in coastal Kenya. *Am J Trop Med Hyg 58*(1): 64-9.
15. Touré YT, et al, 1994. Ecological genetic studies in the chromosomal form Mopti of *Anopheles gambiae* s. str. in Mali, West Africa. *Genetica 94*: 213-23.
16. Biek R, Real LA, 2010. The landscape genetics of infectious disease emergence and spread. *Mol Ecol 19*(17): 3515-31.
17. Holderegger R, Wagner HH, 2006. A brief guide to landscape genetics. *Landscape Ecol 21*(6): 793-6.

18. Manel S, Holderegger R, 2013. Ten years of landscape genetics. *Trends Ecol Evol 28*(10): 614-21.

19. Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP, 2010. Landscape genetics: where are we now?. *Mol Ecol 19*(17): 3496-514.

20. Storfer A, et al, 2007. Putting the 'landscape' in landscape genetics. *Heredity 98*(3): 128.

21. Severson, DW, 1997. RFLP analysis of insect genomes. In The Molecular Biology of Insect Disease Vectors . Netherlands: *Springer* 1997: 309-320.

22. Scott JA, Brogdon WG, Collins FH, 1993. Identification of single specimens of the *Anopheles gambiae* complex by the polymerase chain reaction. *Am J Trop Med Hyg 49*(4): 520-9.

23. Bereczky S, MÅrtensson A, Gil JP, FÄrnert A, 2005. Rapid DNA extraction from archive blood spots on filter paper for genotyping of *Plasmodium falciparum*. *Am J Trop Med Hyg 72*(3):249-51.

24. Zheng L, Benedict MQ, Cornel AJ, Collins FH, Kafatos FC, 1996. An integrated genetic map of the African human malaria vector mosquito, *Anopheles gambiae*. *Genetics 143*(2):941-52.

25. Oetting WS, Lee HK, Flanders DJ, Wiesner GL, Sellers TA, King RA, 1995. Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13 tailed primers. *Genomics 30*(3): 450-8.

26. Excoffier L, Lischer HE, 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour 10*(3):564-7.
27. Jombart T, 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics 24*(11): 1403-5.
28. Pritchard JK, Stephens M, Donnelly P, 2000. Inference of population structure using multilocus genotype data. *Genetics 155*(2): 945-59.
29. Evanno G, Regnaut S, Goudet J, 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol 14*(8): 2611-20.
30. Jakobsson M, Rosenberg NA, 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics 23*(14):1801-6.
31. Rosenberg NA, 2004. DISTRUCT: a program for the graphical display of population structure. *Mol Ecol Resour 4*(1): 137-8.
32. Oyler-McCance SJ, Fedy BC, Landguth EL, 2013. Sample design effects in landscape genetics. *Conserv Genet 14*(2): 275-85.
33. Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL, 1994. High resolution of human evolutionary trees with polymorphic microsatellites. *Nature 368*: 455.
34. Adamack AT, Gruber B, 2014. PopGenReport: simplifying basic population genetic analyses in R. *Methods Ecol Evol 5*(4): 384-7.

35. Cushman S, Storfer A, Waits L, 2015. Landscape genetics: concepts, methods, applications. *John Wiley & Sons*.
36. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A, 2005. Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol 25*: 1965-1978.
37. Channan S, Collins K, Emanuel WR, 2014. Global mosaics of the standard MODIS land cover type data. University of Maryland and the Pacific Northwest National Laboratory, College Park, Maryland, USA.
38. Friedl MA, Sulla-Menashe D, Tan B, Schneider A, Ramankutty N, Sibley A, Huang X, 2010. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sens Environ 114*(1): 168-82.
39. Marthews TR, Dadson SJ, Lehner B, Abele S, Gedney N, 2015. High-resolution global topographic index values. NERC Environmental Information Data Centre. https://doi.org/10.5285/6b0c4358-2bf3-4924-aa8f-793d468b92be
40. McRae BH, Dickson BG, Keitt TH, Shah VB, 2008. Using circuit theory to model connectivity in ecology, evolution, and conservation. *Ecology 89*(10): 2712-24.
41. Richardson JL, Brady SP, Wang IJ, Spear SF, 2016. Navigating the pitfalls and promise of landscape genetics. *Mol Ecol 25*(4): 849-63.
42. Peterman WE, Anderson TL, Ousterhout BH, Drake DL, Semlitsch RD, Eggert LS, 2015. Differential dispersal shapes population structure and patterns of genetic differentiation in two sympatric pond breeding salamanders. *Conserv Genet 16*(1):59-69.

43. Peterman WE, 2018. ResistanceGA: An R package for the optimization of resistance surfaces using genetic algorithms. *Methods Ecol Evol 00*: 1-10.

44. Bates D, Maechler M, Bolker B, Walker S, 2014. lme4: Linear mixed-effects models using Eigen and S4. *R package version 1*(7):1-23.

45. World Health Organization, 2016. World malaria report 2015. WHO, Geneva.

46. Sinka ME, et al, 2010. The dominant *Anopheles* vectors of human malaria in Africa, Europe and the Middle East: occurrence data, distribution maps and bionomic précis. *Parasit Vectors 3*(1): 117.

47. Robert V, Awono-Ambene HP, Thioulouse J, 1998. Ecology of larval mosquitoes, with special reference to *Anopheles arabiensis* (Diptera: Culcidae) in market-garden wells in urban Dakar, Senegal. *J Med Entomol 35*(6): 948-55.

48. Kirby MJ, Lindsay SW, 2009. Effect of temperature and inter-specific competition on the development and survival of *Anopheles gambiae* sensu stricto and *An. arabiensis* larvae. *Acta Trop 109*(2) 118-23.

49. Githeko AK, et al, 1996. Some observations on the biting behavior of *Anopheles gambiae ss, Anopheles arabiensis,* and *Anopheles funestus* and their implications for malaria control. *Exp Parasitol 82*(3): 306-15.

50. Massebo F, Balkew M, Gebre-Michael T, Lindtjørn B, 2013. Blood meal origins and insecticide susceptibility of *Anopheles arabiensis* from Chano in South-West Ethiopia. *Parasit Vectors 6*(1): 44.

51. Tirados I, Costantini C, Gibson G, Torr SJ, 2006. Blood-feeding behaviour of the malarial mosquito *Anopheles arabiensis*: implications for vector control. *Medical Vet Entomol 20*(4): 425-37.

52. Wanjala CL, et al, 2015. Pyrethroid and DDT resistance and organophosphate susceptibility among Anopheles spp. mosquitoes, western Kenya. *Emerg Infect Dis 21*: 2178.

53. Nyanjom SR, et al, 2003. Population genetic structure of *Anopheles arabiensis* mosquitoes in Ethiopia and Eritrea. *J Hered 94*(6): 457-63.

54. Donnelly MJ, Cuamba N, Charlwood JD, Collins FH, Townson H, 1999. Population structure in the malaria vector, *Anopheles arabiensis* Patton, in East Africa. *Heredity 83*(4): 408.

55. Chandre F, Darrier F, Manga L, Akogbeto M, Faye O, Mouchet J, Guillet P, 1999. Status of pyrethroid resistance in *Anopheles gambiae* sensu lato. *Bull World Health Organ 77*(3): 230.

56. Kitau J, et al, 2012. Species shifts in the Anopheles gambiae complex: do LLINs successfully control *Anopheles arabiensis*?. *PLoS one 7*(3): e31481.

57. Bayoh MN, et al, 2010. *Anopheles gambiae*: historical population decline associated with regional distribution of insecticide-treated bed nets in western Nyanza Province, Kenya. *Malar J 9*(1): 62.

58. Lwetoijera DW, Harris C, Kiware SS, Dongus S, Devine GJ, McCall PJ, Majambere S, 2014. Increasing role of *Anopheles funestus* and *Anopheles arabiensis* in malaria transmission in the Kilombero Valley, Tanzania. *Malar J 13*(1): 331.

59. Drake JM, Beier JC, 2014. Ecological niche and potential distribution of *Anopheles arabiensis* in Africa in 2050. *Malar J 13*(1): 213.

60. Charlwood JD, et al, 1995. The rise and fall of *Anopheles arabiensis* (Diptera: Culicidae) in a Tanzanian village. *Bull Entomol Res 85*(1):37-44.

CHAPTER 4

Landscape Genetics of *Plasmodium falciparum* in Kenya

Abstract

Malaria spread is complex in that it is affected by the movement of mosquitoes and people. Movement patterns can be inferred from measuring genetic relatedness among populations or inferred gene flow. Despite that an understanding of how malaria spreads across a landscape is critical for preventing re-introduction of malaria parasites, little is known on the factors shaping their genetic structure. We hypothesize that human population connectedness, measured by proximity to roads, primarily drives *Plasmodium falciparum* gene flow in Kenya, as opposed to factors related to mosquito movement. To test this hypothesis, we collected *P. falciparum* malaria parasites across Kenya and genotype specimens at nine microsatellite loci. We tested associations between pairwise genetic distance and ecological factors related to mosquito and human movement using linear mixed effects models. We observed genetic structuring, but did not find any associations between genetic structure and ecological factors tested. We found evidence of a gene flow barrier which coincides with Lake Victoria, suggesting that *P. falciparum* population structure may be largely influenced by human travel. Moreover, we found that populations proximal to large cities were dissimilar to nearby *P. falciparum* populations, which may be indicative of frequent human travel to large cities resulting in parasites being introduced from more distant populations. Knowledge of how malaria parasites spread geographically can be used to identify areas susceptible to malaria parasite re-introductions following a public health intervention, as well as to predict the spread of antimalarial drug resistance.

Introduction

Malaria parasite dispersal patterns are complex, as they are affected by both the movement of mosquitoes and people. Mosquitoes primarily disperse to seek blood meals and oviposit.[1] Thus, factors associated with both larval habitat and blood meal host availability impact dispersal. Mosquito dispersal distance ranges from less than one kilometer to several kilometers throughout a lifetime.[1] Whereas humans can carry malaria parasites over very long distances, though this distance may also depend on environmental or landscape factors, such as factors associated with agricultural capacity and accessibility.[2] Therefore, dispersal patterns of malaria parasites are complex with potentially numerous factors contributing to parasite spread. Despite that an understanding of how malaria spreads across a landscape is critical for preventing malaria re-introduction and sustaining malaria elimination, there is limited knowledge on how malaria spreads geographically.[3,4]

Organism dispersal patterns can be inferred from measuring genetic relatedness among populations or estimated gene flow. Population structure of *P. falciparum* in Kenya is thought to be shaped by human travel patterns which maintain high gene flow across the area.[5,6] While Zhong et al.[7] report small but statistically significant population structure between highland and lowland malaria populations, other studies have reported no genetic structure across the region.[5,6,8,9] Similarly, no significant barriers to gene flow between *P. falciparum* populations were detected in the Democratic Republic of Congo.[4] In Ethiopia, gene flow patterns of *P. falciparum* in Ethiopia were found to reflect seasonal human migration patterns coinciding with seasonal harvest.[10] In addition, environmental heterogeneity and geographical distance were not found to constrain *P. falciparum* gene flow in Ethiopia.[10] Though human movement is thought to largely drive *P. falciparum* gene flow, few geographic barriers to *P. falciparum* gene flow have

been identified. Here, we test if human population connectedness, measured by road accessibility, can explain *P. falciparum* genetic structuring in Kenya.

Landscape genetic analysis allows us to test the impacts of ecological variables on the dispersals of organisms through the measurement of associations between inferred gene flow and ecological factors.[11-15] Thus, here, we use a landscape genetics approach to test the hypothesis that accessibility (distance to roads) primarily drives *P. falciparum* population structure. As countries approach elimination, it is critically important to understand the underlying factors which promote and prevent migration of malaria parasites to sustain malaria control and elimination. Moreover, parasite resistance to antimalarial drugs poses a major threat to malaria elimination, and so understanding the factors which may facilitate resistance spread can help to inform resistance containment strategies.

Methods

Sample Collection

P. falciparum parasites were collected between May 2014 and January 2015 from fourteen sites within three distinct geographical areas in Kenya: Western Kenya, Rift Valley, and Coastal Kenya (Table 4.1). Parasites were collected from symptomatic patients seen at clinics. Blood dots were collected on Whatman filter paper from patients who had a positive malaria rapid diagnostic test (RDT). Filter paper samples were stored at -20°C until use.

DNA extraction and species identification

The Saponin/Chelex method was used to extract the DNA.[16] The final extracted volume was 200 μL. Quantitative polymerase chain reaction (qPCR) was used to detect and quantify the

P. falciparum DNA simultaneously.[17] Amplification was conducted in a 20 µL mixture consisted of 2µL of DNA, 10µL of 2xSYBR Green qPCR Master Mix (Thermo Scientific, USA), and 0.3µL of 10µM primer. Reactions were performed in the CFX96 Touch Real-Time PCR Detection System (Bio-Rad). Cycling conditions were as follows: initial denaturation at 95°C for 3 min, followed by 45 cycles at 94°C for 30 sec, 55°C for 30 sec, 68°C for 1 min, and 95°C for 10, followed by a melting curve step from 65°C to 95°C with 0.5°C increments. Samples with a CQ value less than or equal to 40 were considered positive for *P. falciparum*.

Microsatellite genotyping

Nine microsatellite loci were selected for genotyping *An. gambiae s.s., An. arabiensis, P. falciparum* based on evidence of polymorphism in previous studies, reliable amplification, and having an even distribution across chromosomes.[18] We used the M13 tailed primer method to fluorescently label our primers.[19] Amplification was conducted in a total volume of 10 µl with 5 µl of 2x DreamTaq Green PCR Master Mix (Thermo Fisher, USA), 0.5 µl of 10µM primer (forward primer with M13 tail), and 1 µl of DNA template. Thermocycling conditions were as follows: initial denature of 95°C for 5 min, followed by 45 amplification cycles of 95°C for 30 sec, annealing temperature (Table 4.2) for 30 sec, and 65°C for 45 sec, and then a final extension of 65°C for 7 min. PCR products were analyzed on an automated 4300 DNA analyzer (Li-Cor, Lincoln, NE), and alleles were quantified with the use of Gene ImagIR 4.33 software (Li-Cor).

Population genetic analysis

We calculated expected heterozygosity and Garza-Williamson Index in Arlequin.[20] Allelic richness at each study site was calculated in the *hierfstat* R package.[21] To estimate population structure, we used a model-based approach and an exploratory approach for

population clustering. First, we used the R package *adegenet* to do a standard principal component analysis (PCA) to determine genetic relationships among populations.[22] Second, we estimated population structure using STRUCTURE v. 2.3.4, which uses a Bayesian algorithm to group samples into genetically distinct clusters, K.[23] We tested K=1-7, with six replicates for each K-level, an initial burn-in of 200,000, and then 600,000 Monte Carlo Markov Chain iterations. The program was run using an admixture model. ΔK was used to detect the number of K (clusters).[24] The output data for the best estimate of K were analyzed using CLUMMP to calculate the mean cluster membership coefficients across multiple runs.[25] The CLUMMP output was visualized using DISTRUCT.[26] Pairwise F_{ST} values were calculated in the R package *adegenet* and bootstrapping was done to determine statistical significance at p-value <0.05, with a Bonferonni correction.[22]

We also performed spatial clustering of individual analysis using a Bayesian clustering in a spatial geographic network with TESS under the admixture model.[27-29] Individuals within a population were randomly assigned geographic coordinates over +/- 0.01 degree N-S and E-W, centered on the geographic coordinate of the population. The maximum number of clusters was set to K=5, with 100 replicates for each K-level, an initial burn-in of 10,000, and then 60,000 sweeps. The optimal K was determined based on the Deviance Information Criterion (DIC), a statistical measure of model deviance penalized by the number of K. The runs with the lowest 10% DIC scores for optimal K were further analyzed using CLUMMP and visualized in R using the script written by Jay et al.[28] Finally, for *P. falciparum*, we identified barriers using Monmonier's algorithm, which uses a maximum-difference method to determine genetically distinct groups in space where samples on either side of the barrier.[30,31] Barriers were determined using the R package *adegenet*.[22]

Landscape genetic analysis

Since population sampling was not evenly spread across the country, we analyzed Eastern Kenya separately from the rest of Kenya due to the biases uneven sampling can introduce in analysis.[32] Genetic distance between populations was calculated as the proportion of shared alleles (D_{PS}) measured in the R package *PopGenReport*.[33,34] When testing the effects of environmental factors on gene flow between populations, between-site characteristics are of the greatest concern.[35] Hence, resistance surfaces were created based on factors hypothesized to prevent or promote gene flow. We created raster files in ArcGIS 10 using climate data from WorldClim,[37] vegetation index data from MODIS (MCD12Q1),[38,39] population density from WorldPop (www.worldpop.org.uk), and roads from the Global Roads Open Access Data Set (Table 4.3). All raster files were resampled to a grain size of 1 km. Landscape distance among all pairs of sites was measured using electrical circuit theory in *PopGenReport*.[39,40] Circuit theory incorporates multiple pathways into the analysis, providing an advantage over the least cost path method. A key shortcoming of landscape genetics has been identified in the methods used to assign resistance values to variables (e.g. forest cover or elevation).[40] For continuous variables, such as precipitation, raw elevation numbers can be used to assign resistance values, but this method assumes a linear response to precipitation. For non-continuous variables, assigning resistance values has most commonly relied upon expert opinion, but this approach introduces biases.[40] To overcome these shortcomings, we used *ResistanceGA* a package in R to optimize resistance surfaces to our genetic data.[41,42] ResistanceGA uses a genetic algorithm to unbiasedly optimize resistance surfaces. Lastly, linear mixed effects models with the maximum likelihood population effects were used to fit optimized resistance surfaces to genetic data. Akaike

information criterion with a penalty for extra parameters (AICc) was used as the measure of model fitness to genetic data.

Results

Population genetic analysis

Allelic richness (A_R) was highest at 'Jun' (5.26) and 'Emu' (5.20), while A_R was lowest at 'Por' (3.11) and 'Eld' (3.47) (Table 4.4). Expected heterozygosity (H_E) was highest at 'Jun (0.78) and 'Ken' (0.78) and lowest at 'Emu' (0.43) and 'Kab' (0.45) (Table 4.4). Garza-Williamson indexes (G-W) were low across all populations, indicative of recent bottleneck events (Table 4.4). We identified four clusters consistently across six runs in STRUCTURE (Figure 4.1C). PCA of the first and second components (explaining 28.2 and 16.6% of the variance, respectively) revealed a similar clustering pattern as the STRUCTURE analysis (Figure 3.1B). All sites clustered together with the exceptions of 'Emu' and 'Kab' (Figure 4.1C). In addition, pairwise F_{ST} values were generally found to be non-significant, with the exceptions of pairwise comparisons with 'Emu', 'Kab', 'Gil' and 'Jun' (Table 4.5). In addition, F_{ST} values were significant between populations in Western Kenya: 'Por' and 'Ken' (0.09); 'Hom' and 'Ken' (0.04); 'Hom' and 'May' (0.05); 'Hom' and 'Eld' (0.03).

Landscape genetic analysis

None of the single surface models performed higher than a null model for *P. falciparum* (Table 4.6), so we do not display any of the landscape resistance surfaces. We therefore conducted additional spatial analysis and display those outputs. Through analysis of spatial distribution of ancestry coefficients in Western Kenya, we consistently identified four distinct clusters (Figure 4.2A). Cluster one includes 'Emu'; cluster two includes 'Kab'; cluster three

includes ‘Por’, ‘May’, ‘Emu’, and ‘Mar’; and cluster four includes ‘Hom’, ‘Ken’, ‘Gil’, ‘Nak’ (Figure 4.2A). Barrier analysis revealed a significant barrier between ‘Hom’, ‘Ken’, ‘Emu’, and ‘Por’, which partially coincides with Lake Victoria (Figure 4.2B). In Eastern Kenya, we identified two distinct clusters (Figure 4.3A). Cluster one includes ‘Jun’ and cluster two includes ‘Mus’, ‘Jar’, and ‘Kak’ (Figure 4.3A). Barrier analysis revealed a significant barrier between ‘Jun’ and ‘Jar’ populations (Figure 4.3B).

Discussion

Little to no population genetic differentiation has been found for *P. falciparum* in Kenya, which is thought to be influenced by frequent human travel maintaining extensive gene flow between populations.[5-7] While we detected population differentiation between populations of *P. falciparum* across Kenya, we did not identify any significant factors associated with population structuring. However, we detected a genetic barrier coinciding with Lake Victoria. Lake Victoria provides a natural barrier to human travel, and so this finding provides evidence to support that human travel may largely influence population structure of malaria parasites in Kenya.

Additionally, we identified barriers surrounding ‘Jun’ in Eastern Kenya and ‘Emu’ in Eastern Kenya, as well as that these populations were not closely related to nearby populations. The reason for these divergences at ‘Jun’ and ‘Emu’ may be partly explained by their relatively close proximity to major cities, ‘Jun’ to Mombasa (28 km) and ‘Emu’ to Kisumu (30km). Mombasa is the second most populous city in Kenya, while Kisumu is the third most populous city in Kenya. Following the Gravity Model, where the amount of interaction between two cities is proportional to city size and distance, it is plausible that *P. falciparum* population structure near Mombasa and Kisumu are heavily influenced by frequent human travel from more distant

populations not included in this study.[4,43-45] Thus, close proximity to highly connected city centers may help to explain why parasites from 'Jun' and 'Emu' are not closely related to nearby populations. Rather, the population structure of these populations may be highly influenced by parasite introduction from more distant populations not included in this study, as a result of frequent human travel between larger cities. This notion of human travel characterizing the population structure of these populations is underscored by the finding that genetic diversity was highest at 'Emu' and 'Jun', as measured by allelic richness. The presence of a gene flow barrier across Lake Victoria, as well as the divergence observed in populations proximal to populous cities provide evidence in support of human travel driving the population structure of malaria parasites in Kenya.

Overall, we detected significant genetic structuring between populations of *P. falciparum*, which suggests that populations are more genetically isolated than observed in previous studies in Kenya, which detected little to no structuring.[5,7,46] This trend of increasing fragmentation, particularly in Western Kenya, could be caused by the increase in public health interventions since 2006[47,48] decreasing parasite transmission intensity.[49,50] This notion is underscored by the observed low Garza-Williamson indexes, which are indicative of recent bottleneck events. Malaria admissions in Kenya decreased from more than 500 admissions per 100,000 people in 2009 to less than 100 admissions per 100,000 people in 2014 (the year of this study).[51] Likewise, Anthony et al.[52] reported significant *P. falciparum* genetic structuring following a period of decreasing malaria transmission in the Malaysia. This finding of increasing fragmentation in Kenya has implications for malaria control. For example, *P. falciparum* polymorphisms associated with antimalarial drug resistance may spread more slowly between

populations, as well as local interventions may be more effective at reducing local malaria incidence.

Genetic structuring in Kenya did not conform to the isolation-by-distance model at the national level, a finding consistent with previous studies in Africa.[4,10,53] Further, genetic structuring was not associated with environmental or landscape heterogeneity, a finding which is consistent with *P. falciparum* structuring in the Democratic Republic of Congo[4] and Ethiopia.[10] These findings implicate the importance of human movement not easily captured by road networks in maintaining *P. falciparum* gene flow over large distances. For example, political and territorial conflicts, as well as linguistic and cultural diversity are major drivers of human interaction essential to parasite gene flow.[50,53,54] Moreover, the presence of a gene flow barrier coinciding with Lake Victoria, a natural barrier to human movement further implicates the significance of travel in maintaining parasite movement. Similarly, other natural barriers to travel, such the mountainous terrain across the Malaysian Borneo region[52] and the Andes separating Western and Central South America[55] are thought to restrict parasite gene flow. Thus, it is likely that complex human movement plays a key role in shaping *P. falciparum* structure in Kenya.

This study had certain limitations. First, parasite sampling from symptomatic patients may bias toward rare alleles, as symptomatic infections would be caused most frequently by infection with a parasite of an uncommon genotype.[56,57] This may result in a higher observed genetic diversity than would be in parasites from asymptomatic infections. Second, different landscape or environmental drivers may drive *P. falciparum* dispersal at a smaller or larger scale than tested. Medley et al.[58] found that gene flow of the Asian tiger mosquito was facilitated by highways at a broad spatial scale, but was hindered by forests at a small spatial scale. Therefore,

it is plausible that environmental or landscape factors associated with mosquito movement has a larger impact on *P. falciparum* population structure at a smaller scale, such as the village level. However, since parasite samples were collected at health facilities and not associated with a home address, we cannot test this hypothesis. Third, we are unable to test the hypothesis that the observed divergences in population structure at 'Emu' and 'Jun' is influenced by parasite introductions from more distant populations, since those potential populations were not included in this study.

Using a landscape genetics approach to identify factors facilitating parasite dispersal has great potential for mitigating disease risk.[4,11,59] While we did not identify significant corridors for *P. falciparum* in Kenya, physical barriers to human travel, such as lakes, may impede *P. falciparum* dispersal. In addition, large cities may facilitate the spread of parasites between more distant populations. These findings suggests the possibility that human travel not sufficiently captured by road networks shapes population structure in Kenya. This knowledge is important to understanding how drug resistance spreads, as well as how parasites can become re-introduced to an area following public health interventions. Knowing which areas are most susceptible to parasite re-introduction is essential to planning effective antimalarial interventions and monitoring.

Table 4.1. Locality information and sample size.

Region	Site	Longitude	Latitude	Elevation	Sample size
Western Kenya Lowlands	Por	34.012	-0.114	1139	57
	Ken	34.629	-0.385	1150	58
	Hom	34.465	-0.543	1184	57
Western Kenya Highlands	Emu	34.618	0.023	1520	57
	May	34.578	0.51	0.513	57
	Eld	35.177	0.325	2098	53
Rift Valley	Mar	36.018	0.484	1004	29
	Kab	35.663	0.498	1150	19
	Nak	35.945	-0.334	2163	11
	Gil	35.911	-0.495	2009	19
Eastern Kenya	Mus	39.501	-4.394	15	58
	Jun	39.742	-3.852	66	58
	Jar	39.736	-3.616	16	58
	Kak	40.039	-3.170	14	58

Table 4.2. Microsatellite markers used for *P. falciparum* (Anderson et al. 1999) genotyping.

P. falciparum		
Locus	Chromosome	Annealing (°C)
Ta87	NA	53
Pfpk2	12	45
Polya	4	51
Pfpg377	12	51
ARA2	11	45
2490	NA	45
TA42	NA	45
TA81	5	45
TA109	6	45

Table 4.3 Predictor variables used for landscape genetic analysis.

Category	Variable	Source
Climate	Average Temperature	WorldClim BIO1
	Annual Precipitation	WorldClim BIO12
Social	Human Population Density	Worldpop
	Distance to Roads	Global Roads Open Access Data Set, version 1
Landscape	Vegetation Index	NASA MCD12Q1
	Topographic Wetness Index	NERC Environmental Information Data Centre

Table 4.4. Genetic diversity indices for *P. falciparum* in Kenya. A_R is allelic richness; H_E is expected heterozygosity; G-W is the Garza-Williamson Index.

Site	A_R	H_E	G-W
Por	3.11	0.62	0.14
Hom	4.84	0.69	0.16
Ken	4.61	0.78	0.17
Emu	5.20	0.43	0.07
May	5.19	0.73	0.14
Eld	3.47	0.73	0.17
Kab	4.24	0.45	0.07
Mar	4.99	0.59	0.10
Nak	3.92	0.53	0.07
Gil	4.11	0.61	0.09
Kak	3.60	0.54	0.11
Jar	4.41	0.62	0.14
Jun	5.26	0.78	0.17
Mus	4.64	0.65	0.15

Table 4.5. Pairwise F_{ST} values for *P. falciparum* populations in Kenya. Lower triangle indicates pairwise F_{ST} value and upper triangle indicates statistical significance ($P<0.05$).

	Por	Ken	Hom	Emu	May	Eld	Kab	Gil	Nak	Mar	Kak	Jar	Mus	Jun
Por	0	*		*			*				*			*
Ken	0.09	0	*	*			*							
Hom	0.08	0.04	0	*	*	*	*							
Emu	0.28	0.21	0.23	0	*	*	*	*	*	*	*	*	*	*
May	0.09	0.03	0.05	0.21	0		*			*				*
Eld	0.08	0.03	0.03	0.20	0.01	0	*							*
Kab	0.21	0.17	0.20	0.43	0.20	0.18	0	*		*	*	*	*	*
Gil	0.12	0.05	0.06	0.27	0.03	0.04	0.27	0		*	*	*	*	*
Nak	0.13	0.07	0.10	0.32	0.06	0.05	0.25	0.10	0					
Mar	0.06	0.02	0.02	0.11	0.03	0.02	0.16	0.05	0.06	0				
Kak	0.10	0.03	0.03	0.20	0.04	0.04	0.16	0.07	0.10	0.01	0	*		
Jar	0.09	0.02	0.04	0.21	0.03	0.03	0.14	0.04	0.01	0.02	0.03	0		
Mus	0.07	0.02	0.01	0.19	0.03	0.02	0.16	0.04	0.05	0.01	0.03	0.02	0	
Jun	0.11	0.03	0.05	0.20	0.06	0.06	0.19	0.08	0.09	0.00	0.03	0.06	0.03	0

Table 4.6. Single surface models for *P. falciparum* in Kenya.

	AICc	ΔAIcc	weight
Western Kenya			
1) Null	-189.05	0.00	0.34
2) Annual Precipitation	-188.00	1.05	0.20
3) Average Temperature	-187.18	1.87	0.13
4) Human Population Density	-186.28	2.77	0.09
5) Vegetation Index	-186.15	2.90	0.08
6) Distance to Roads	-186.13	2.92	0.08
7) Geographic Distance	-186.10	2.95	0.08
Coastal Kenya			
1) Null	-23.93	0.00	0.82
2) Annual Precipitation	-19.15	4.79	0.07
3) Distance to Roads	-18.78	5.15	0.06
4) Average Temperature	-17.28	6.66	0.03
5) Vegetation Index	-14.05	9.89	0.01
6) Human Population Density	-13.42	10.51	0.00
7) Geographic Distance	-13.08	10.85	0.00

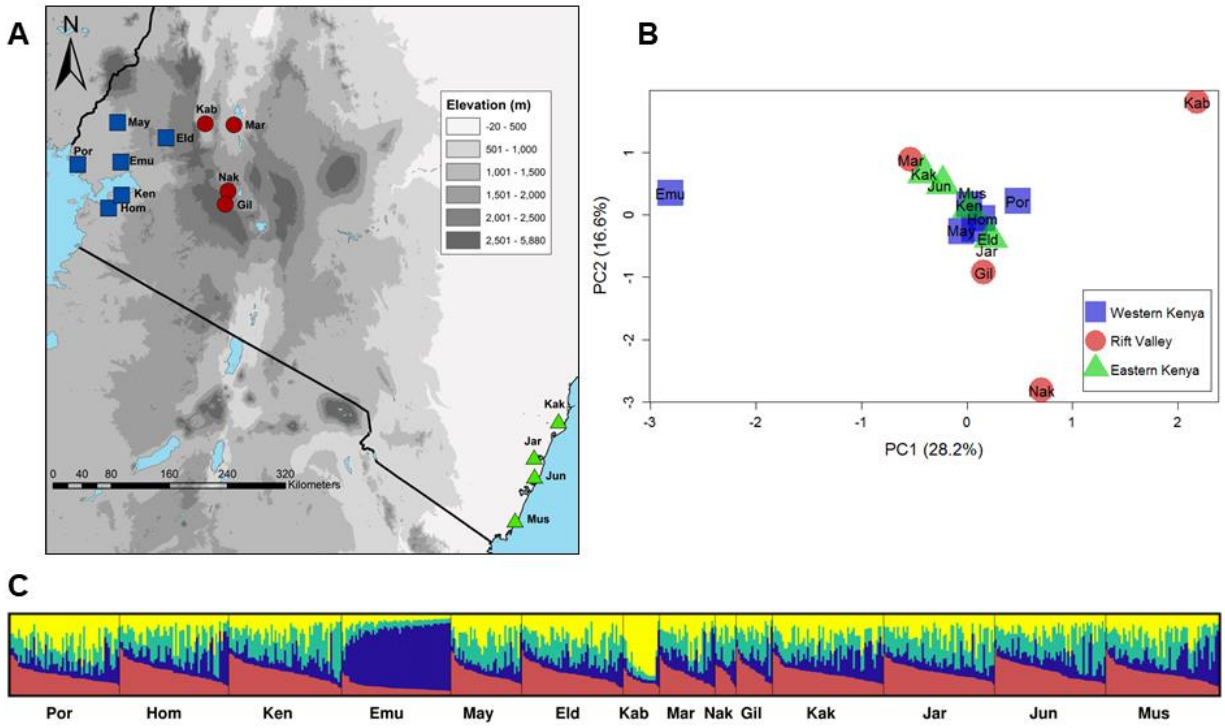


Figure 4.1. Population structure of *P. falciparum* in Kenya. A) Map of *An. gambiae s.s.* sites. B) Principal component analysis (PCA) of genetic relatedness. C) Bayesian population structuring analysis from STRUCTURE.

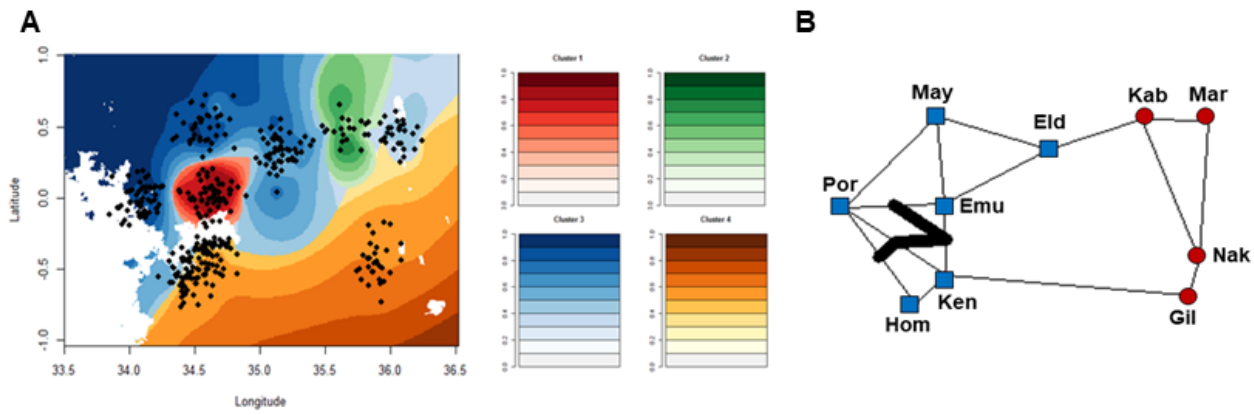


Figure 4.2. Spatial structure of *P. falciparum* populations in Western Kenya. A) Spatial distribution of ancestry coefficients using TESS and the admixture model B) Results of barrier analysis. The thick black line indicates a significant barrier to gene flow between populations.

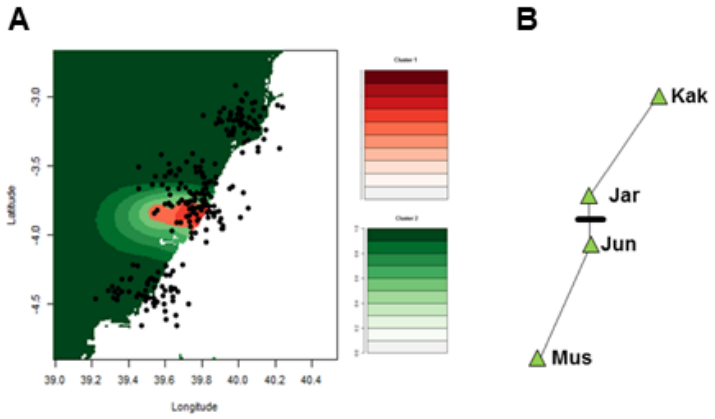


Figure 4.3. Spatial structure of *P. falciparum* populations in Eastern Kenya. A) Spatial distribution of ancestry coefficients using TESS and the admixture model B) Results of barrier analysis. Thick black line indicates a significant barrier to gene flow between populations.

References

1. Service MW. Mosquito (Diptera: Culicidae) dispersal—the long and short of it, 1997. *J Med Entomol 34*(6): 579-88.
2. Kamau L, Mukabana WR, Hawley WA, Lehmann T, Irungu LW, Orago AAS, Collings FH, 1999. Analysis of genetic variability in *Anopheles arabiensis* and *Anopheles gambiae* using microsatellite loci. *Insect Mol Biol 8*(2): 287-97.
3. Lynch C, Roper C, 2011. The transit phase of migration: circulation of malaria and its multidrug-resistant forms in Africa. *PLoS Med 8*(5): e1001040.
4. Carrel M, et al, 2015. The geography of malaria genetics in the Democratic Republic of Congo: a complex and fragmented landscape. *Soc Sci Med 33*: 233-41.
5. Bonizzoni M, Afrane Y, Baliraine FN, Amenya DA, Githeko AK, Yan G, 2009. Genetic structure of *Plasmodium falciparum* populations between lowland and highland sites and antimalarial drug resistance in Western Kenya. *Infect Genet Evol 9*(5): 806-12.
6. Mulenge FM, Hunja CW, Magiri E, Culleton R, Kaneko A, Aman RA, 2016. Genetic diversity and population structure of *Plasmodium falciparum* in Lake Victoria Islands, a region of intense transmission. *Am J Trop Med Hyg 95*(5): 1077-85.
7. Zhong D, et al, 2007. *Plasmodium falciparum* genetic diversity in western Kenya highlands. *Am J Trop Med Hyg 77*(6): 1043-50.
8. Vardo-Zalik AM, Zhou G, Zhong D, Afrane YA, Githeko AK, Yan G, 2013. Alterations in *Plasmodium falciparum* genetic structure two years after increased malaria control efforts in western Kenya. *Am J Trop Med Hyg 88*(1): 29-36.

9. Gatei W, et al, 2015. Genetic diversity of *Plasmodium falciparum* parasite by microsatellite markers after scale-up of insecticide-treated bed nets in western Kenya. *Malar J 14*(1): 495.
10. Lo E, et al, 2017. Transmission dynamics of co-endemic *Plasmodium vivax* and *P. falciparum* in Ethiopia and prevalence of antimalarial resistant genotypes. *PLoS Negl Trop Dis 11*(7): e0005806.
11. Biek R, Real LA, 2010. The landscape genetics of infectious disease emergence and spread. *Mol Ecol 19*(17): 3515-31.
12. Holderegger R, Wagner HH, 2006. A brief guide to landscape genetics. *Landscape Ecol 21*(6): 793-6.
13. Manel S, Holderegger R, 2013. Ten years of landscape genetics. *Trends Ecol Evol 28*(10): 614-21.
14. Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP, 2010. Landscape genetics: where are we now?. *Mol Ecol 19*(17): 3496-514.
15. Storfer A, et al, 2007. Putting the 'landscape' in landscape genetics. *Heredity 98*(3): 128.
16. Bereczky S, MÅrtensson A, Gil JP, FÄrnert A, 2005. Rapid DNA extraction from archive blood spots on filter paper for genotyping of *Plasmodium falciparum*. *Am J Trop Med Hyg 72*(3):249-51.
17. Rougemont M, Van Saanen M, Sahli R, Hinrikson HP, Bille J, Jaton K, 2004. Detection of four Plasmodium species in blood from humans by 18S rRNA gene subunit-based and species-specific real-time PCR assays. *Journal Clin Microbiol 42*(12):5636-43.

18. Anderson TJ, SU XZ, Bockarie M, Lagog M, Day KP, 1999. Twelve microsatellite markers for characterization of *Plasmodium falciparum* from finger-prick blood samples. *Parasitology 119*(2): 113-25.
19. Oetting WS, Lee HK, Flanders DJ, Wiesner GL, Sellers TA, King RA, 1995. Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13 tailed primers. *Genomics 30*(3): 450-8.
20. Excoffier L, Lischer HE, 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour 10*(3):564-7.
21. Goudet J, Hierfstat, 2005. A package for R to compute and test hierarchical F- statistics. *Mol Ecol Resour 5*(1):184-6.
22. Jombart T, 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics 24*(11): 1403-5.
23. Pritchard JK, Stephens M, Donnelly P, 2000. Inference of population structure using multilocus genotype data. *Genetics 155*(2): 945-59.
24. Evanno G, Regnaut S, Goudet J, 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol 14*(8): 2611-20.
25. Jakobsson M, Rosenberg NA, 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics 23*(14):1801-6.
26. Rosenberg NA, 2004. DISTRUCT: a program for the graphical display of population structure. *Mol Ecol Resour 4*(1): 137-8.

27. Durand E, Jay F, Gaggiotti OE, François O, 2009. Spatial inference of admixture proportions and secondary contact zones. *Mol Biol Evol 26*(9): 1963-73.

28. Jay F, et al, 2012. Forecasting changes in population genetic structure of alpine plants in response to global warming. *Mol Ecol 21*(10): 2354-68.

29. Caye K, Deist TM, Martins H, Michel O, François O, 2016. TESS3: fast inference of spatial population structure and genome scans for selection. *Mol Ecol Resour 16*(2): 540-8.

30. Manni F, Guerard E, Heyer E, 2004. Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by using Monmonier's algorithm. *Hum Biol 76*(2): 173-90.

31. Monmonier MS, 1973. Maximum- difference barriers: An alternative numerical regionalization method. *Geogr Anal 5*(3): 245-61.

32. Oyler-McCance SJ, Fedy BC, Landguth EL, 2013. Sample design effects in landscape genetics. *Conserv Genet 14*(2): 275-85.

33. Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL, 1994. High resolution of human evolutionary trees with polymorphic microsatellites. *Nature 368*: 455.

34. Adamack AT, Gruber B, 2014. PopGenReport: simplifying basic population genetic analyses in R. *Methods Ecol Evol 5*(4): 384-7.

35. Cushman S, Storfer A, Waits L, 2015. Landscape genetics: concepts, methods, applications. *John Wiley & Sons*.

36. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A, 2005. Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol 25*: 1965-1978.

37. Channan S, Collins K, Emanuel WR, 2014. Global mosaics of the standard MODIS land cover type data. University of Maryland and the Pacific Northwest National Laboratory, College Park, Maryland, USA.

38. Friedl MA, Sulla-Menashe D, Tan B, Schneider A, Ramankutty N, Sibley A, Huang X, 2010. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sens Environ 114*(1): 168-82.

39. McRae BH, Dickson BG, Keitt TH, Shah VB, 2008. Using circuit theory to model connectivity in ecology, evolution, and conservation. *Ecology 89*(10): 2712-24.

40. Richardson JL, Brady SP, Wang IJ, Spear SF, 2016. Navigating the pitfalls and promise of landscape genetics. *Mol Ecol 25*(4): 849-63.

41. Peterman WE, Anderson TL, Ousterhout BH, Drake DL, Semlitsch RD, Eggert LS, 2015. Differential dispersal shapes population structure and patterns of genetic differentiation in two sympatric pond breeding salamanders. *Conserv Genet 16*(1):59-69.

42. Peterman WE, 2018. ResistanceGA: An R package for the optimization of resistance surfaces using genetic algorithms. *Methods Ecol Evol 00*: 1-10.

43. Erlander S, Stewart NF, 1990. The Gravity Model in Transportation Analysis – Theory and Extensions (VSP Books, Zeist, The Netherlands).

44. Chao DL, Halloran ME, Longini IM, Jr., 2011 Vaccination strategies for epidemic cholera in Haiti with implications for the developing world. *Proc Natl Acad Sci USA 108*: 7081–7085.

45. Mari L, et al, 2012, Modelling cholera epidemics: The role of waterways, human mobility and sanitation. *J R Soc Interface 9*: 376–388.

46. Omedo I, et al 2017. Micro-epidemiological structuring of *Plasmodium falciparum* parasite populations in regions with varying transmission intensities in Africa. *Wellcome open research 2*:29.

47. Zhou G, et al, 2011. Changing patterns of malaria epidemiology between 2002 and 2010 in Western Kenya: the fall and rise of malaria. *PloS one 6*(5): e20318.

48. Hemming-Schroeder E, et al, 2018. Impacts of antimalarial drugs on *Plasmodium falciparum* drug resistance markers, Western Kenya, 2003–2015. *Am J Trop Med Hyg 98*(3): 692-9.

49. Anderson TJ, et al, 2000. Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. *Mol Biol Evol 17*(10): 1467-82.

50. Pumpaibool T, et al, 2009. Genetic diversity and population structure of *Plasmodium falciparum* in Thailand, a low transmission country. *Malar J 8*(1): 155.

51. World Health Organization, 2016. World malaria report 2015. WHO, Geneva.

52. Anthony TG, Conway DJ, Cox-Singh J, Matusop A, Ratnam S, Shamsul S, Singh B, 2005. Fragmented population structure of *Plasmodium falciparum* in a region of declining endemicity. *J Infect Dis 191*(9): 1558-64.

53. Mobegi VA, Loua KM, Ahouidi AD, Satoguina J, Nwakanma DC, Amambua-Ngwa A, Conway DJ, 2012. Population genetic structure of *Plasmodium falciparum* across a region of diverse endemicity in West Africa. *Malar J 11*(1): 223.

54. Lum JK, Kaneko A, Tanabe K, Takahashi N, Björkman A, Kobayakawa T, 2004. Malaria dispersal among islands: human mediated *Plasmodium falciparum* gene flow in Vanuatu, Melanesia. *Acta Trop 90*(2):181-5.

55. Yalcindag E, et al, 2012. Multiple independent introductions of *Plasmodium falciparum in* South America. *Proc Natl Acad Sci USA 109*(2): 511-6.

56. Contamin H, Fandeur T, Rogier C, Bonnefoy S, Konate L, Trape JF, Mercereau-Puijalon O, 1996. Different genetic characteristics of *Plasmodium falciparum* isolates collected during successive clinical malaria episodes in Senegalese children. *Am J Trop Med Hyg* 54(6):632-43.

57. Gray KA, Dowd S, Bain L, Bobogare A, Wini L, Shanks GD, Cheng Q, 2013. Population genetics of *Plasmodium falciparum* and *Plasmodium vivax* and asymptomatic malaria in Temotu Province, Solomon Islands. *Malar J 12*(1): 429.

58. Medley KA, Jenkins DG, Hoffman EA, 2015. Human- aided and natural dispersal drive gene flow across the range of an invasive mosquito. *Mol Ecol 24*(2): 284-95.

59. Hemming-Schroeder E, Lo E, Salazar C, Puente S, Yan G, 2018. Landscape Genetics: A Toolbox for Studying Vector-Borne Diseases. *Front Ecol Evol 6*: 21.