The ecology and evolution of Plasmodium falciparum malaria among rural communities in Madagascar

The ecology and evolution of *Plasmodium falciparum* malaria among rural communities in Madagascar

A dissertation presented

by

Benjamin Lawrence Rice

to the

Department of Organismic and Evolutionary Biology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biology

Harvard University

Cambridge, Massachusetts

May 2019

Dissertation Advisors: Daniel L. Hartl and Christopher D. Golden Benjamin L. Rice

The ecology and evolution of *Plasmodium falciparum* malaria among rural communities in Madagascar

Abstract

Parasites in the genus *Plasmodium* annually cause hundreds of millions of cases of malaria worldwide, and more than a million cases in Madagascar alone. Alarmingly, multiple lines of evidence suggest that, counter to general trends elsewhere, the burden of malaria has increased in regions of Madagascar in recent years. Understanding the drivers and distribution of malaria among communities in Madagascar is hampered by limited health surveillance capacity and extremely high inter-regional ecological variation. This motivates the development of new approaches to studying and controlling malaria in Madagascar, especially approaches that are not reliant on the existing limited health monitoring infrastructure and that can be deployed in remote communities. One such approach is the use of population genetic data to infer trends in the parasite population from a sample of infections. To implement this approach, we completed new, multi-method field studies of rural communities in Madagascar (Chapter 2), characterized spatial variation in the prevalence of malaria infection observed at those study sites (Chapter 3), validated a panel of single nucleotide polymorphisms (SNPs) (Chapter 4), and then compared patterns in parasite genetic diversity at that SNP panel across eco-regions of Madagascar (Chapter 5). In remote communities situated within the tropical rainforest environments of northeastern Madagascar and the semi-arid, spiny forests of southwestern Madagascar, we find

evidence of: (i) high, but spatially variable malaria prevalence, (ii) high levels of within host parasite genetic diversity, and (iii) high levels of genotypic diversity in *P. falciparum* populations. Together, these data are consistent with the presence of communities with concerningly high burdens of malaria in rural Madagascar and have implications for future control efforts.

Table of Contents

Acknowledgments

This dissertation would have been far from possible without the truly invaluable support, mentorship, patience, and kindness of many people. They were instrumental and have my deepest gratitude for opening new worlds to me and aiding my exploration. Listing all would be impossible, but nine people, or groups, that it would be impossible not to list are:

First, I would like to thank my advisor Dan Hartl for being a remarkable advisor – his constant support, guidance, and the freedom he allowed me made my PhD a wonderful experience. That unwavering support, despite my insistence on taking off for the forests of Madagascar all too often, is more than I could have ever asked for. In Madagascar there is a concept, *faralahy maditra*, that says something like the last son is likely to be a troublesome one. As I may be Dan's last PhD student, and as I was certainly trouble, I think it applies here.

Second, I would like to thank my co-advisor Christopher Golden. It has been a privilege to work, travel, and learn (and sort fecal samples) under your mentorship. I am deeply grateful for you opening the world of Madagascar to me and welcoming me into your research community. Your advising left indelible marks on my worldview and my approach to how research can fit into it. Thank you for countless hours and energy spent with me in support and guidance (and eating sedap).

Third, I would like to thank my dissertation committee members Robin Hopkins and David Haig for their suggestions, kind feedback, and the time they devoted to supporting me throughout my PhD. Robin's infectious positivity and enthusiasm were an antidote at low points and David's broad evolutionary wisdom reminded me of the bigger pictures in nature and life (for e.g. endosymbiosis within Apicomplexan parasites).

Fourth, none of the work presented here would have been possible without the excellent team of colleagues I was fortunate to work with in Madagascar. *Fisaorana be ho an'ny Ekipa Mahery jiaby*: Hervet Joseph Randriamady, Andry Ny Aina Rakotmalala, Jean Frederick Randrianasolo, Dr. Mahery Rebaliha, Dr. Tafangy Ambinintsoa Nirina, Vololoniaina Ravo Rakotoarilalao, Mamy Yves Andrianantenaina, Fenoarison Faraniaina Robuste Mahonjolaza, Alex Dominique Rasamison, Raelson Hermann Paratoaly, Arisoa Vonona Miadana, Mafia Rasoanirina, Lainandrasana Faustin, Evelin Jean Gasta Anjaranirina, B. J. Rodolph Rasolofoniaina, Pascal Andriamahazoarivosoa, Jean Adolphe Manambina, Gerandine Zafindalana, Laurent, Ravaoline, Barthélémie, Toto Simplice, Delizia Martinne, and Dr. Luciano Michael Tantely. In particular, I want to state my gratitude for all the help Hervet Randriamady provided (and that we managed to pull him out of that river in Tuléar). His assistance and patience were irreplaceable. As were Fred Kely's 4AM playlists, Taffy's workout routines, Simplice's hospitality, Magibe's *sakay*, and Maman'i Aime's terrifyingly large servings of rice.

Fifth, I would like to thank the entire Hartl Lab and the descendants/affiliates that I was fortunate enough to cohabitate with at some point during my PhD: Jenny Yen Pham, Bridget Power, Kalsang Namgyal, Elena Lozovsky, Russ Corbett-Detig, Hsiao-Han Chang, Brandon Ogbunu, Julien Ayroles, Lene Martinsen, Marjorie Lienard, and Tim Sackton. I am thankful for the insightful conversations and for the friendly, caring environment you all created. It was a privilege to be a part of. One could not ask for a better lab mate than Jenny Pham. I say that while acknowledging that we may have discussed science at some point – I don't remember – but I do remember countless hours discussing more important topics like food, cooking, and other kinds of food. Without a doubt, that fantastic, food-based friendship, played a crucial role in being able to get to this point.

Sixth, much gratitude is due to the numerous collaborators, colleagues, and mentors that made my dissertation possible. Among the malaria lab(s) at the Harvard TH Chan School of Public Health: Rachel Daniels, Cory Lee Schlesener, Daniel Neafsey, Sarah Volkman, Dyann Wirth, Selina Bopp, and Katelyn Durfee warrant special mention for letting me pester them with questions, allowing me space to work, and letting me learn alongside a community of world-class malaria experts. In Madagascar, I am extremely thankful for the friendship and support (and Gastro pizza) provided by Jim Hazen at Catholic Relief Services. Additionally, I am indebted to Jess Metcalf at Princeton University, and Amy Wesolowski and Amy Winter at Johns Hopkins University, for supporting the 2017 fieldwork project and placing their trust in me to carry it out. Making bao from scratch in Morombe deserves a dissertation chapter of its own. I would also like to thank Anthonio Naunau Medard at the African Language Program for his encouragement and tolerance of my efforts to learn some Malagasy. *Misaotra vale*!

Seventh, I want to thank my previous mentors Ananias Escalante and Maria Andreína Pacheco, and members of their lab including Monica Mercedes Saldarriaga de Acosta. You introduced me to the world of research and malaria, made me feel at home in a research environment, made me feel like family, and fed me arepas.

Eighth, family – you brought me into this world and there is nothing I can write that could share how grateful for and humbled by your support and sacrifices that I am.

Ninth, inseparable from family are the close friends that made my time as a PhD student amazing; in no particular order: Bapu Vaitla (for hoopin' and heatmaps), Robert Gustafson (for bringing the noise), Ari Silver (for forgiving me for Brasilia), Eddie Duran (for food and fashion sense), Allison Andraski (for crucial hiking trips), Cristina Aggazzotti (for surviving Madagascar), Liz Ransey and Dawson Rauch (for pasta nights and single spoons), Nicholas Arisco (for many DBWs), Erika Veidis and Perri Sheinbaum (for hands videos and help ordering), Andy Kim and Zach Strecker (for being my role models), Greta Tuckute (for swerving tabooed beaches) and all members of the Nashton commune (for giving me a home), and, last, all members of the BGSL soccer team (my only regret is that we never beat HBS).

Last, it deserves mention that I am extremely grateful for the generosity, understanding, and welcome offered me by members of the communities in which these studies were performed.

Chapter 1

Introduction: Motivation to study malaria parasites and their genetic variation in Madagascar

1.1 Malaria as a major global health burden

Malaria is a parasitic disease caused by infection with species of the genus *Plasmodium* (Order: Haemospororida, Phylum: Apicomplexa). Parasites in the diverse *Plasmodium* genus are transmitted between hosts by hematophagous insects and infect a wide range of vertebrates, including reptiles, birds, humans, and other mammals (Martinsen 2008). In humans, malaria parasites are vectored by *Anopoheles* mosquitoes and, per the World Health Organization, currently result in hundreds of millions of cases and approximately 435,000 deaths per year. Approximately 97% of cases are caused by the species *Plasmodium falciparum*. Of the global burden, an estimated 92% of cases and 93% of deaths occur in Africa– a disproportionate amount (61%) of which are among children less than 5 years of age (WHO 2018).

Alarmingly, the most recent *World Malaria Report* noted that, despite gains in malaria control over the last two decades, the number of cases increased dramatically in 2016-2017 in several African countries. The trend in Madagascar was notable, with the estimated number of annual malaria cases reported increasing by over 500,000 in the last two years (WHO 2018). This motivates research to better understand the ecology and epidemiology of malaria infection in countries, such as Madagascar, currently facing difficulties in malaria control. Such studies can provide information critical to efforts to reverse setbacks and move towards malaria control and elimination.

1.2 An overview of malaria in Madagascar

Within Madagascar, existing epidemiological data on malaria primarily come from: (i) passive surveillance based on aggregating reports of cases observed at health centers (e.g. Ihantamalala 2018a), (ii) active surveillance from national surveys performed every 2-3 years that randomly sample children across administrative regions of Madagascar (e.g., Malaria Indicator Survey 2016, Kang 2018), (iii) geospatial modeling that combines prevalence data with social and environmental correlates (e.g., Howes 2016), and (iv) unsystematic localized case studies and outbreak reports (e.g., Kesteman 2016).

The most recent national Malaria Indicator Surveys (performed in 2013 and 2016) indicated that malaria prevalence was higher in many eastern and western coastal areas than it was in surveys prior to 2011. In addition, several studies have reported outbreaks in some rural districts where malaria incidence far exceeded expected regional or national rates (e.g., Ihantamalala 2018b, Randrianasolo 2010, Rabarijaona 2009). Together, these data suggest the burden of malaria in Madagascar, especially in some regions of the country, remains high and warrants attention. Chapter 3 contains further description of the distribution of malaria burden in Madagascar and our observation of foci where over 30% of individuals were malaria positive at the time of sampling.

However, existing epidemiological datasets in many malaria endemic countries, including Madagascar, have significant limitations. First, the accuracy of passive surveillance data based on clinical reporting relies on the ability of healthcare systems to adequately count infected individuals and report data to a central office. A commonly observed complication is that a substantial proportion of *P. falciparum* infected individuals remain asymptomatic and thus

unlikely to visit a clinic when infected (reviewed in Bousema *et al* 2014). Additionally, of the infected individuals who do have symptoms, only a minority have access to and seek treatment. In a large qualitative survey of caregivers in Madagascar, Do *et al* (2018) conclude that seeking treatment when showing malaria symptoms is uncommon. Likewise, Battle *et al* (2016) estimate that less than 46% of the population in Madagascar seeks and receives treatment at formal healthcare facilities and Penny *et al* (2015) estimate that less than 21% is covered by facilities offering effective treatment programs. Further, of those individuals who do seek and receive treatment, a substantial proportion are not counted by monitoring programs. Howes et al (2016) estimated that, as of 2014, less than 29% of the health districts in Madagascar were successfully reporting 50% or more of the malaria diagnostic tests that were being performed. In conclusion, clinical reporting data is unlikely to accurately characterize malaria dynamics given that a large proportion of malaria infected individuals: a) do not seek care, and b) if they do, do not have access to healthcare facilities with the capacity to reliably report infection rates.

Active surveillance programs that deploy surveys to proactively seek infected individuals are not reliant on clinical reporting; however, they require significant resources in terms of costs and labor and remain limited in Madagascar. As a result, such programs have been limited temporally or geographically. For example, sampling done for the National Malaria Control Program's semi-regular Malaria Indicator Surveys (*Enquête sur les Indicateurs du Paludisme à Madagascar*), is constrained to a certain age group (children less than 5 years of age) and lacks representativeness at sub-national, sub-regional scales.

Without reliable estimates of local trends in malaria infection it is difficult to identify areas requiring greater attention for future interventions, and to measure the impact of existing interventions. Indeed, in a summary of the progress made in malaria control between the years

2000-2015, Bhatt *et al* (2015) concluded that the major obstacle in planning more effective intervention strategies was the scarcity of reliable malaria transmission estimates from remaining higher transmission areas, especially in less developed parts of Africa. There is therefore a need for novel approaches to study *P. falciparum* parasite populations and provide the transmission estimates needed to inform intervention efforts.

1.3 Evolutionary genetic analysis as a method to study malaria transmission

Data on the genetic variation in *P. falciparum* populations can provide new insight into parasite transmission dynamics, as the pattern of genetic variation is predicted to reflect trends in the parasite population (reviewed in Volkman *et al* 2012). In areas of high, stable transmission, the parasite population size is large, and parasites are rapidly transmitted between hosts, thus increasing the probability of hosts becoming infected multiple times. Upon being taken up by the mosquito vector, recombination between the multiple parasite genomes that were in the coinfected host creates new allelic combinations and increases the haplotypic diversity of the parasite population. On the other hand, in areas of low, or unstable, transmission, malaria parasite population size is smaller and the probability of re-infecting already infected hosts is reduced. Thus, the rate at which recombination creates new haplotypes is lower. As a result, we predict differing patterns of genetic variation in parasite populations under different transmission scenarios.

In concordance with these predictions, *P. falciparum* population genetic diversity parameters have been observed to change in response to changes in malaria transmission rates in real human populations (e.g., Daniels 2013, Nkhoma 2013, Daniels 2015, Adomako-Ankomah

2017). Two genetic parameters in particular that show correlation with transmission rates can be defined as: (i) the proportion of infections that contain multiple, differing *P. falciparum* parasite genomes within that single host (termed polygenomic infections), and (ii) the proportion of infections containing a unique genotype not observed in the other genotyped infections (termed genotypic diversity). For example, in a study following a known reduction in malaria transmission in Thiès, Senegal, the estimated proportion of polygenomic infections and the genotypic diversity decreased by more than 50% and 40%, respectively (Daniels 2013). Such studies demonstrate that data on genetic variation from samples of *P. falciparum* infections can be useful towards efforts to evaluate the progress of an intervention or to provide baseline indicators upon which to measure future interventions – without relying on the existence of sufficient data from traditional epidemiological methods.

However, very little is known about the genetic variation of *Plasmodium* populations in Madagascar currently. The few existing studies have primarily focused on drug resistance (e.g., Andriantsoanirina 2010) or diversity at antigen-encoding loci (e.g., Bordbar 2014, Gendrot 2019), and as a result we have a poor understanding of neutral genetic variation. Chapters 4 and 5 of this dissertation discuss our work to study genetic variation at a panel of neutral *P. falciparum* polymorphisms.

Such information is urgently needed in Madagascar due to the inadequacies of the malaria monitoring infrastructure and the evidence of recent increases in malaria infections described above. Additionally, the rural communities of Madagascar where malaria appears to be increasing are vulnerable in other aspects. Poverty, malnutrition, and ecological degradation rates are among the highest globally (WHO 2015), and as a consequence the well-being of these communities has become a national and global priority. Many of these efforts to improve rural

health outcomes, often included in broader initiatives to reach conservation and development goals, can benefit from a better understanding of local malaria patterns.

To investigate the ecology and evolutionary genetics of *P. falciparum* in Madagascar, we performed several large field studies in rural regions of Madagascar (Chapter 2), analyzed the prevalence of malaria infection among those study populations (Chapter 3), and then sought to characterize the distribution of genotypic variation among those malaria infections (Chapters 4 and 5).

References

1. World Malaria Report 2018. Geneva: World Health Organization; 2018

2. Malaria Indicator Survey: Enquête sur les Indicateurs du Paludisme 2016. Programme National de Lutte contre le Paludisme Antananarivo, Madagascar

3. Martinsen ES, Perkins SL, Schall JJ. A three-genome phylogeny of malaria parasites (Plasmodium and closely related genera): Evolution of life-history traits and host switches. Molecular Phylogenetics and Evolution. 2008;47: 261–273. doi:10.1016/j.ympev.2007.11.012

4. Ihantamalala FA, Rakotoarimanana FMJ, Ramiadantsoa T, Rakotondramanga JM, Pennober G, Rakotomanana F, et al. Spatial and temporal dynamics of malaria in Madagascar. Malaria Journal. 2018a;17. doi:10.1186/s12936-018-2206-8

5. Kang SY, Battle KE, Gibson HS, Ratsimbasoa A, Randrianarivelojosia M, Ramboarina S, et al. Spatio-temporal mapping of Madagascar's Malaria Indicator Survey results to assess Plasmodium falciparum endemicity trends between 2011 and 2016. BMC Medicine. 2018;16. doi:10.1186/s12916-018-1060-4

6. Howes RE, Mioramalala SA, Ramiranirina B, Franchard T, Rakotorahalahy AJ, Bisanzio D, et al. Contemporary epidemiological overview of malaria in Madagascar: operational utility of reported routine case data for malaria control planning. Malaria Journal. 2016;15. doi:10.1186/s12936-016-1556-3

7. Kesteman T, Rafalimanantsoa SA, Razafimandimby H, Rasamimanana HH, Raharimanga V, Ramarosandratana B, et al. Multiple causes of an unexpected malaria outbreak in a high-transmission area in Madagascar. Malaria Journal. 2016;15. doi:10.1186/s12936-016-1113-0

8. Ihantamalala FA, Herbreteau V, Rakotoarimanana FMJ, Rakotondramanga JM, Cauchemez S, Rahoilijaona B, et al. Estimating sources and sinks of malaria parasites in Madagascar. Nature Communications. 2018b;9. doi:10.1038/s41467-018-06290-2

9. Randrianasolo L, Raoelina Y, Ratsitorahina M, Ravolomanana L, Andriamandimby S, Heraud J-M, et al. Sentinel surveillance system for early outbreak detection in Madagascar. BMC Public Health. 2010;10. doi:10.1186/1471-2458-10-31

10. Bousema T, Okell L, Felger I, Drakeley C. Asymptomatic malaria infections: detectability, transmissibility and public health relevance. Nature Reviews Microbiology. 2014;12: 833–840. doi:10.1038/nrmicro3364

11. Do M, Babalola S, Awantang G, Toso M, Lewicky N, Tompsett A. Associations between malaria-related ideational factors and care-seeking behavior for fever among children under five in Mali, Nigeria, and Madagascar. Carvalho LH, editor. PLOS ONE. 2018;13: e0191079. doi:10.1371/journal.pone.0191079

12. Battle KE, Bisanzio D, Gibson HS, Bhatt S, Cameron E, Weiss DJ, et al. Treatment-seeking rates in malaria endemic countries. Malaria Journal. 2016;15. doi:10.1186/s12936-015-1048-x

13. Penny MA, Maire N, Bever CA, Pemberton-Ross P, Briët OJT, Smith DL, et al. Distribution of malaria exposure in endemic countries in Africa considering country levels of effective treatment. Malaria Journal. 2015;14. doi:10.1186/s12936-015-0864-3

14. Howes RE, Mioramalala SA, Ramiranirina B, Franchard T, Rakotorahalahy AJ, Bisanzio D, et al. Contemporary epidemiological overview of malaria in Madagascar: operational utility of reported routine case data for malaria control planning. Malaria Journal. 2016;15. doi:10.1186/s12936-016-1556-3

15. Bhatt S, Weiss DJ, Cameron E, Bisanzio D, Mappin B, Dalrymple U, et al. The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. Nature. 2015;526: 207–211. doi:10.1038/nature15535

16. Volkman SK, Neafsey DE, Schaffner SF, Park DJ, Wirth DF. Harnessing genomics and genome biology to understand malaria biology. Nature Reviews Genetics. 2012;13: 315–328. doi:10.1038/nrg3187

17. Daniels R, Chang H-H, Séne PD, Park DC, Neafsey DE, Schaffner SF, et al. Genetic Surveillance Detects Both Clonal and Epidemic Transmission of Malaria following Enhanced Intervention in Senegal. Mayor A, editor. PLoS ONE. 2013;8: e60780. doi:10.1371/journal.pone.0060780

18. Nkhoma SC, Nair S, Al-Saai S, Ashley E, McGready R, Phyo AP, et al. Population genetic correlates of declining transmission in a human pathogen. Molecular Ecology. 2013;22: 273–285. doi:10.1111/mec.12099

19. Daniels RF, Schaffner SF, Wenger EA, Proctor JL, Chang H-H, Wong W, et al. Modeling malaria genomics reveals transmission decline and rebound in Senegal. Proceedings of the National Academy of Sciences. 2015;112: 7067–7072. doi:10.1073/pnas.1505691112

20. Adomako-Ankomah Y, Chenoweth MS, Durfee K, Doumbia S, Konate D, Doumbouya M, et al. High Plasmodium falciparum longitudinal prevalence is associated with high multiclonality and reduced clinical malaria risk in a seasonal transmission area of Mali. Craig AG, editor. PLOS ONE. 2017;12: e0170948. doi:10.1371/journal.pone.0170948

21. Rabarijaona LP, Randrianarivelojosia M, Raharimalala LA, Ratsimbasoa A, Randriamanantena A, Randrianasolo L, et al. Longitudinal survey of malaria morbidity over 10 years in Saharevo (Madagascar): further lessons for strengthening malaria control. Malaria Journal. 2009;8: 190. doi:10.1186/1475-2875-8-190

22. Bordbar B, Tuikue Ndam N, Renard E, Jafari-Guemouri S, Tavul L, Jennison C, et al. Genetic diversity of VAR2CSA ID1-DBL2Xb in worldwide Plasmodium falciparum populations: Impact on vaccine design for placental malaria. Infection, Genetics and Evolution. 2014;25: 81–92. doi:10.1016/j.meegid.2014.04.010

23. Gendrot M, Fawaz R, Dormoi J, Madamet M, Pradines B. Genetic diversity and deletion of Plasmodium falciparum histidine-rich protein 2 and 3: a threat to diagnosis of P. falciparum malaria. Clinical Microbiology and Infection. 2019;25: 580–585. doi:10.1016/j.cmi.2018.09.009

24. Andriantsoanirina V, Ratsimbasoa A, Bouchier C, Tichit M, Jahevitra M, Rabearimanana S, et al. Chloroquine Clinical Failures in P. falciparum Malaria Are Associated with Mutant Pfmdr-1, Not Pfcrt in Madagascar. Snounou G, editor. PLoS ONE. 2010;5: e13281. doi:10.1371/journal.pone.0013281

25. WHO. The global prevalence of anaemia in 2011. Geneva: World Health Organization; 2015

Chapter 2

Description of longitudinal and cross-sectional epidemiological field studies in remote communities in Madagascar

Attributions:

The contents of this chapter have been published, or prepared for publishing, as a part of four manuscripts for which I was a co-author or co-first author. The first is a manuscript published in the *International Journal of Epidemiology* (Volume 46, Issue 6, December 2017, Pages 1747–1748d): **Cohort Profile: The Madagascar Health and Environmental Research (MAHERY) study in north-eastern Madagascar** by Golden CD, Anjaranirina EJG, Fernald LCH, Hartl DL, Kremen C, Milner DA, Ralalason DH, Ramihantaniarivo H, Randriamady H, Rice BL *et al*. The second is a manuscript under review by *Frontiers in Nutrition* (February 2019): **Cohort Profile: The Madagascar Health and Environmental Research–Antongil (MAHERY–Antongil) study in north-eastern Madagascar** by Golden CD, Borgerson C, Rice BL *et al*. The third is a manuscript in preparation to be submitted to *BMC Public Health* (May 2019): **Social, demographic, and ecological variation of nutrition and disease across Madagascar: a cross-sectional study** by Golden CD and Rice BL (co-first author) *et al*. The fourth is a manuscript published in *Cell* (Volume 176, Issue 3, 24 January 2019, Pages 649-662.e20): **Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle** by Pasolli E, Asnicar F, Manara S, Zolfo M, Karcher N, Armanini F, Beghini F, Manghi P, Tett A, Ghensi P, Collado MC, Rice BL, *et al*. Components of these manuscripts relevant to subsequent analyses presented in this dissertation are selected and synthesized here.

2.1 Abstract

Madagascar is characterized by high social, cultural, and ecological variation across its geography. Seasonal cycles in local bioclimatic conditions, food availability, wealth, and human movement also drive temporal variation in health through effects on nutrition, non-communicable, and infectious disease. The purpose of the epidemiological field studies presented here was to characterize and understand the social and environmental associations to important health burdens in Madagascar. Among these important health burdens in rural communities in Madagascar is malaria. This chapter discusses three field studies carried out by the MAHERY field teams: (1) a prospective cohort study (n = 719 individuals, all ages, both sexes) of two forest adjacent communities of northeastern Madagascar, (2) another prospective cohort study (n = 878) in northeastern Madagascar, but of five coastal communities, and (3) a cross-sectional study of 24 communities (n = 6,293) distributed across the southeastern, southwestern, and central highland regions of Madagascar. In each, detailed socio-demographic surveys were paired with biological sampling (including blood and fecal samples) in order to investigate patterns in health burdens at the individual, household, and community level. Summaries of results for health indicators, including malaria prevalence by on-site rapid diagnostic tests, demonstrate a variable, but high burden of adverse health outcomes in these communities and motivates further study.

2.2 Introduction

Study of rural communities in Madagascar is motivated in part by the observation that communities, and the landscapes surrounding them, are experiencing substantial environmental changes and high rates of poor health outcomes. Madagascar has experienced significant environmental change since 1960, particularly through forest clearing for agricultural expansion (Harper 2007). From quantification from remote sensing data, forest loss (3.2 million hectares) in Madagascar was second highest in Africa between 2000-2010 and associated with this many species of wildlife are threatened with extinction (GLAAD-Global Forest Watch 2014). Variable rates of deforestation have been reported, with higher rates in certain locations in the country, particularly following periods of political upheaval (Allnutt 2013). The process of regular forest clearing and burning alters local ecologies, with downstream effects on water and soil. This, in turn, affects risks of malnutrition and infectious and non-communicable disease.

Climatic patterns are undergoing change in Madagascar as well. Historical records from 1961-2005 demonstrate an increasing temperature in 67% of locations across the island nation, with mean temperature expected to increase by 2.0-6.5 degrees Celsius by 2100. Moreover, the destructive potential of cyclones is expected to increase by 2-17% (Hannah 2008). Over the past thirty years, there have been five major droughts that have crippled the agricultural sector and more than thirty severe floods, killing hundreds of people and indirectly affecting tens of thousands. Long-term trends indicate that temperature will increase, while rainfall is expected to increase in variability (World Bank 2017).

The impact of these environmental and climatic changes will pose threats to food availability, income generation, and local ecosystems, significantly affecting disease burden, and the timing and magnitude of epidemics. These studies seeks to describe the health status of a

large sample of geographically and socially diverse Malagasy people through detailed clinical measurements and social surveys.

Studying these communities provides the opportunity to begin understanding the interactive dynamics among local ecological disturbance, reductions in populations of marine and terrestrial animals, nutritional status, the human fecal microbiome, and the incidence of intestinal parasites, zoonotic pathogens, and malaria.

2.3 Materials and Methods

2.3.1 Study Design and Enrollment

Study 1: The MAHERY Makira Cohort Study (Golden 2017)

For the first study, two communities adjacent to the Makira Natural Park (in the Analanjirofo administrative region) were selected because they have been tracked for the past 10 years to understand the year-to-year variation in income generation, natural resource use, forest reliance and food security. One of the two communities was selected because it was the original community enrolled in a 2008-2009 health study. This provided an ideal foundation to situate our in-depth case studies concerning the role of wildlife consumption in human nutrition. There were 160 households total in community A and 157 households total in community B. We assigned each household a number and created a public lottery for people to see the randomization process. We did not assign numbers to households where the head of household was older than 75 years of age. We recruited 95 households in community A. Two households were suggested for withdrawal because the head/heads of households had such a heavy alcohol dependency that they would be unable to complete the diet records and the blood draw may be too heavy of a health toll. We then randomly selected an additional two households to return the

total enrolled to 95 households. In community B, we used the same public randomization process and recruited 57 households. This household randomization process allowed the enrollment of a total 719 individuals of both sexes from ages 73 years and younger. The cohort was followed from 2013 to 2014. All households were recruited, enrolled and each individual consented and assented following our IRB approved methods (Protocol #22826, Harvard T.H. Chan School of Public Health).

Study 2: The MAHERY Coastal Cohort Study

For the second study, five coastal communities adjacent to the Makira Natural Park or Masoala National Park forested areas were selected based on their participation in resource management systems. Using a community-wide comprehensive census, we assigned each household a number and randomly selected households to be included in the research. The C1 sample contained 25 households (out of a total of 360 households), and C2-5 contained 50 households each (out of a total of 260, 634, 180, and 98 households, respectively). At the outset, we enrolled 225 households and 1031 individuals in five communities.

The cohort was followed for twenty-eight months (September 2015 until December 2017), with an initial pilot period from September 2015 until April 2016 of enrolling households and collecting survey information. During this initial period, 35 households withdrew and we continued to enroll households until reaching a total of 225 total households. Over the course of enrollment, a total of 153 individuals withdrew. Our final sample population included 878 individuals of both sexes ages 91 and younger enrolled for the clinical aspects of the study. Following our final enrollment prior to the clinical phases of the study, we had approximately 6.4% of the overall clinical enrollment withdraw due to survey fatigue and fear of needles. All

households were recruited and enrolled, and each individual consented or assented, following our IRB approved methods (Protocol #15-2230, Harvard T.H. Chan School of Public Health).

Study 3: The MAHERY cross-regional cross-sectional study

For the third study, our mixed-method approach included an observational cross-sectional study. Research subjects were men, women and their children from 1125 households evenly distributed across 24 communities in four ecologically and socio-demographically distinct regions of Madagascar (the southeast, southwest, west coast, and central plateau). These communities had already been enrolled in a Catholic Relief Services intervention to provide food security support through multiple pathways.

The regions sampled were selected to represent the major eco-regions within Madagascar. In Madagascar, these distinct eco-regions have been defined for multiple purposes, varying from subdivisions with similar malaria transmission patterns to areas with similar types of predominant vegetation, by delimiting portions of the country with similar precipitation, temperature and elevation profiles. The four regions sampled in this study (and their corresponding administrative regions) were: (i) the high rainfall, tropical rainforest southeast (within the Vatovay Fitovinany administrative region) (ii) the more arid southwest (within the Toliara II district of the Atsimo Andrefana administrative region) (iii) the seasonal rainfall, semi-deciduous forested areas of the west coast (within the Morombe district of the Atsimo Andrefana administrative region), and (iv) the sub-humid forest and grassland, colder, higher elevation central plateau (within the Amoron'i Mania administrative region). We note that the southwest and west coast regions we distinguish in our study are located within the same, large, administrative region of Madagascar (Atsimo Andrefana) but due to their geographic distance

(approximately 200 km) and marked differences in climate, vegetation, and agricultural systems we consider them as separate eco-regions in this study.

Within the 24 communities, we sampled approximately 50 households in each community for a sum of 200-300 households per region, and a total household enrollment of 1125 across all regions. For these 1125 households, all persons of both sexes and all ages therein (for a total of 6293 individuals) were enrolled into the research study. Households were defined as those with reproductive aged women and young children under five years of age. Subjects were offered no compensation for participating in interviews or providing clinical samples at the time of primary sampling, but were offered 1000 Malagasy ariary (approximately $0.28USD) at the follow-up interviews performed the following year. This amount of money was compensation for time lost from labor activities due to participation in the follow-up survey and was not deemed coercive.

We recruited individuals with a two-stage opt-out procedure. The local authority (typically a chief or community elder) accompanied the field research lead to conduct a community meeting where speeches were given to describe the work. This was the only culturally appropriate way to describe our research and allow for questions and answers to be heard by all community members. Following this meeting, we used the community census provided by the local authority to randomize households for participation. Households with no reproductive aged women and/or children under 5 years of age were excluded. Randomization occurred by assigning numbers to households in the community and then using a random number generator to select households to enroll. In some cases, community censuses were out of date or inaccurate. For these cases, community censuses were updated to include all households by conferring with community leaders and heads of local family groups. The selected households

were then visited by the investigator and the local community authority to invite participation. Interviews were performed in private and lasted no more than 1-2 hours. The local authority was not present during the interview. Not showing up for the interview was viewed by our team as the subject declining to be interviewed.

2.3.2 Biological sample collection for disease analysis

To collect biological materials (e.g., blood or feces) for disease analysis, subjects traveled (always less than a 30 minute walk) to a private room where the health assessments were conducted. Lidocaine was applied to the arm's surface to dull the pain of needle insertion when individuals feared the pain of the blood draw. Venous blood draws were collected for all participants older than 2 years of age and under 55 years of age to collect one tube (5-7 mL) of whole blood into a Sarstedt monovette with Lithium heparin. This whole blood was used to source blood for: (i) rapid diagnostic tests (RDTs) for malaria (a fingerprick was used if no venous blood draw was taken), (ii) hemoglobin status tests done using a HemoCue 201+ device, (iii) preservation of dried blood spots on Whatman filter paper FTA cards (2 spots per individual), and (iv) OmegaQuant filter paper treated with HUFASave^{TM} for preservation of fatty acids. Dried blood from the Whatman FTA was used for DNA preservation/extraction and genotyping of *Plasmodium* infections, among other disease analyses. The remaining whole blood was then separated by centrifuge into plasma and a blood pellet, with the plasma being aliquoted into 1 or 2 (depending on the quantity of plasma obtained) 1.8mL Sarstedt cryotubes that were then preserved in liquid nitrogen and then frozen at -80C.

One aliquot of frozen plasma was shipped to the Western Human Nutrition Research Center (USDA) for nutritional analyses while the other tube of plasma was stored at the Harvard

T.H. Chan School of Public Health for future disease and serology tests. The 1.8mL of plasma in the first tube was analyzed for content of ferritin, zinc, and vitamin B12 and will continue to be analyzed for other potential nutritional targets. Inflammation markers such as AGP and CRP were also measured here to control for the role of inflammation in affecting nutritional status. Dried blood from OmegaQuant filter paper was used to characterize fatty acid profiles for each individual following established protocols (Harris 2017).

Extraction of nucleic acid material from the dried blood spot was performed using the Promega Maxwell semi-automated extraction kit. Genetic analysis was performed to determine the presence or absence as well as genotypes of *Plasmodium* malaria parasites (see Chapters 4-5). For stratification purposes only, genotyping of specific human loci known to impact the likelihood of malaria infection (e.g. hemoglobin type, duffy blood group detection) will be performed.

Individuals testing positive for malaria by point-of-care rapid diagnostic test (RDT) were offered treatment and a consultation with a physician. Individuals with point-of-care hemoglobin testing results indicative of anemia were offered treatment and a consultation with a physician. Individuals deemed by physicians to require additional treatment after point-of-care testing were referred to a hospital.

In addition, we collected a fecal sample from individuals enrolled in the study by providing fecal sample tubes (with spoons) and allowing individuals to collect this themselves by defecating and then spooning a sample into a tube. These fecal samples were stored in 90% ethanol and shipped to the Harvard T.H. Chan School of Public Health for microscopic and molecular analysis for intestinal parasites.

Serum samples available from individuals will be tested for antibodies to a set of focal pathogens in order to evaluate the force of infection associated with these pathogens, and associated risk factors, interactions between nutritional and immune status; and to allow us to develop characterize the landscape of immunity for these focal pathogens.

2.3.3 Mosquito vector habitat mapping and larval collection

Our team characterized the peri-domicile larval breeding habitats of anopheline mosquitoes in each of the communities using two methods: 1) a grid-based system; and 2) stratified transects. In sites where households were tightly clustered, we conducted a systematic habitat search of the clustered area and 25m beyond the perimeter of households on the edges of the cluster. In sites with less concentrated households, we conducted habitat searches in a circular area of radius 25m around each household. All mosquito breeding habitats were mapped and geocoded using a Garmin Oregon 550t. In addition to mapping the presence/absence of breeding habitats, we also conducted transect surveys to identify the larval habitats and species composition of anopheline mosquitos endemic to each local ecology of each research site. Two 100m transects were mapped within one undisturbed area and one area that represented the dominant human-altered land-use type in each study community. All mosquito breeding habitats and distinct changes in land-use along the transect and 5m to each side of the transect line were geocoded.

In both protocols, all habitats containing water were sampled for larvae using mesh nets or dippers and pipettes in a standardized method. Larvae were collected and stored in 95% ethanol. Information on the presence of water, the presence of other animals, the habitat's dimensions, the type of habitat, and the presence and number of eggs and pupae at the time of

sampling were recorded. Additionally, all early and late instar larvae were enumerated. All larvae were sorted by genus and instar prior to identification. All 3rd and 4th instar *Anopheles* larvae were identified morphologically to the lowest possible taxonomic level at the Institut Pasteur de Madagascar.

2.3.4 Socio-demographic questionnaires

All individuals enrolled in the study, or their caregivers for individuals too young or unable to respond for themselves, were asked questions about: (i) basic demographics such as age, sex, and occupation, (ii) food security and socio-economic status, (iii) diet, and (iv) health status and disease risk associated behaviors such as mosquito bednet usage. First, censuses of households and their members were completed and GPS locations of all households were recorded. Heads of household were then asked to complete an adapted version of the Coping Strategies Index and the FAO's Household Food Insecurity Access Scale. Information was also gathered on crops grown and sold, and natural resource extraction including fishing and hunting. Then for individuals, dietary recalls (both 24 hour and one week) were completed. The health survey comprised several questions concerning morbidity recalls, bednet usage, vitamin intake, and medication usage (including deworming medicine). Women of reproductive age were asked about pregnancy and breastfeeding.

We assessed socio-economic status through: (1) recalls of each household's food-related; and (2) recalls of (a) the amount (and source) of all cash *income* earned by members of each household, (b) the number of select commercial *assets owned* by members of each household (e.g. motorcycles, bicycles, radios, laptops/tablets, flashlights, fishing nets, boats, livestock), and (c) the amount of 'luxury' *goods consumed* or used by each household during the prior day,

week, and/or month (e.g. sugar, coffee, oil, salt, petrol). We also recorded the primary economic activities of each household, their access to primary and secondary education, medical care, potable water, public safety services (e.g., police officers, court), and a list of market goods available from local stores.

2.4 Results and Discussion

Tables 1, 2, and 3 contain summaries of the characteristics of the study populations for field studies 1, 2, and 3, respectively. In total, we sampled 7,890 individuals across 31 communities in Madagascar. General observations about poverty and vulnerability of the populations across communities were: (i) monthly household income was extremely low in some households (often less than $100 USD), (ii) less than 15% of the adult population had received education at levels above elementary school, and (iii) 13-30% of the female population aged 15-49 was pregnant or breastfeeding at the time of baseline survey. For malaria in particular, pregnancy and infancy are times when the risk of severe disease and mortality are elevated.

Among health indicators, stunting is defined as the failure of children to reach standard growth rates and is associated with an environment of chronic malnutrition. More than 35% of the children surveyed in these studies were stunted or extremely stunted. Likewise the proportion of the population with anemia, often associated with insufficient iron in the diet or a high burden of parasites, was above 30%. Across the four regions of Madagascar sampled in the cross-regional study, 14% of individuals surveyed tested positive for malaria by rapid diagnostic test (RDT).

Table 1: Summary of the MAHERY Makira cohort study population (Study 1)

Sex (% female)	51.1%
Age (median years, minimum–maximum)	15.8 (0.1-73.8)
Household monthly mean income in USD (PPP[a])	23.17 (99.12)
Highest educational attainment	
None	17.1%
Elementary school	67.9%
Middle school	14.0%
High school	1.0%
Age of first marriage (years)	21.3
Stunting (severe; total)[b]	
Both sexes	14.2%; 36.8%
Females	9.6%; 38.4%
Males	11.9%; 35.6%
Underweight (severe; total)[b]	
Both sexes	4.9%; 14.6%
Females	7.1%; 19.6%
Males	3.0%; 10.4%
Wasting (severe; total)[b]	
Both sexes	2.4%; 5.3%
Females	0.9%; 4.5%
Males	3.7%; 5.9%
Reproductive-aged women[c]	31.1%
Pregnant women[d]	19.0%
Lactating women[e]	39.3%

[a] US dollars (purchasing power parity).

[b] Stunting, underweight and wasting were all defined for children under 5 years of age by the World Health Organization's standardized distributions, and the percentages here represent those falling under -2 z scores.

[c] Percentage of females 15–49 years old; [d] Percentage of reproductive-aged women reporting a pregnancy; [e] Percentage of reproductive-aged women lactating.

Table 2: Summary of the MAHERY coastal cohort study population (Study 2)

Sex (% female)	49.4
Age (median years; min-max)	16.0; (0.1 - 91.0)
Household median annual income (current international dollars)[a]	
All communities	$6840
Community 1	$4940
Community 2	$6580
Community 3	$6590
Community 4	$7910
Community 5	$4430
Stunting among children ≤ age 5 (% Severe; Total)[b]	
Both sexes (n=184)	23.9; 44.2
Females (n=99)	18.2; 37.4
Males (n=85)	28.2; 51.9
Underweight among children ≤age 5 (% Severe; Total)[b]	
Both sexes (n=184)	2.7; 19.6
Females (n=99)	2.0; 14.1
Males (n=85)	3.7; 25.9
Wasting among children ≤ age 5 (% Severe; Total)[b]	
Both sexes (n=184)	2.2; 3.6
Females (n=99)	1.0; 2.0
Males (n=85)	3.7; 6.2
Reproductive aged women (women ages 15-49 as % of all women)	46.7
Pregnant women (% of women of ages 15-49)	4.7
Lactating women (% of women of ages 15-49)	13.4

[a] Current international dollars adjusted for purchasing power parity
[b] Stunting, underweight, and wasting all based on WHO MGRSG (2006).

Table 3: Summary of the MAHERY cross-regional cross-sectional study population (Study 3)

Demographic information	
Mean age	18.58 years, *n* = 6292
Sex	52.79% female, *n* = 6292
Mean age of first marriage (women)[a]	18.25 years, *n* = 1024
Mean age of woman's spouse at first marriage[b]	23.21 years, *n* = 628
Household head's monthly income (median; mean)[c]	16.06 USD; 48.41 USD, *n* = 1121
Anemia[d]	*n* = 5540
No anemia	3848 (69.46%)
Anemia (mild, moderate, and severe combined)	1692 (30.54%)
Mild	891 (16.08%)
Moderate	732 (13.21%)
Severe	69 (1.25%)
Malaria (by RDT)[e]	*n* = 5584
Negative	4802 (86.0%)
Positive	782 (14.0%)
Reproductive aged women[f]	*n* = 1592
Pregnant	92 (5.78%)
Breastfeeding	508 (31.90%)

[a] *n* = sample size among women who recalled their age at the time of their first marriage
[b] *n* = sample size among women who recalled both their age and the age of their spouse at the time of their first marriage
[c] USD conversion from Malagasy ariary using "BANKY FOIBEN'I MADAGASIKARA" (Central Bank of Madagascar) daily average exchange rate in 2017
[d] Individuals with a hemoglobin measurement, categorized using the standard WHO thresholds for age and sex groups
[e] Among individuals with a valid rapid diagnostic test (RDT) result
[f] Among women reporting that their first menstruation event had previously occurred

Analysis of the data and samples collected as a part of these three studies is ongoing, but analyses of malaria data are presented in Chapters 3-5 and a metagenomic analysis of the fecal samples has been published by Pasolli *et al* (2019). Putative genomes assembled from metagenomic sequencing of fecal samples revealed thousands of novel bacterial species. Comparison to existing datasets from other populations showed that the Malagasy sample had twice the rate of previously unidentified bacterial genomes in the fecal microbiome as compared to populations with westernized diets. Within putative bacterial species lineages (classified as species-level genome bins), Madagascar microbiome samples often clustered into divergent subtrees – indicating that study of more isolated populations such as these communities in Madagascar is useful to better understand the total breadth of human microbiome diversity.

References

1. Harper GJ, Steininger MK, Tucker CJ, Juhn D, Hawkins F. Fifty years of deforestation and forest fragmentation in Madagascar. Environmental Conservation. 2007;34. doi:10.1017/S0376892907004262

2. Global Forest Watch. 2014. World Resources Institute. www.globalforestwatch.org.

3. Allnutt TF, Asner GP, Golden CD, Powell GVN. Mapping Recent Deforestation and Forest Disturbance in Northeastern Madagascar. Tropical Conservation Science. 2013;6: 1–15. doi:10.1177/194008291300600101

4. Hannah L, Dave R, Lowry PP, Andelman S, Andrianarisata M, Andriamaro L, et al. Climate change adaptation for conservation in Madagascar. Biology Letters. 2008;4: 590–594. doi:10.1098/rsbl.2008.0270

5. World Bank Annual Report 2017. World Bank IBRD.

6. Golden CD, Anjaranirina EJG, Fernald LC, Hartl DL, Kremen C, Milner Jr DA, et al. Cohort Profile: The Madagascar Health and Environmental Research (MAHERY) study in north-eastern Madagascar. International Journal of Epidemiology. 2017;46: 1747–1748d. doi:10.1093/ije/dyx071

7. Harris WS, Del Gobbo L, Tintle NL. The Omega-3 Index and relative risk for coronary heart disease mortality: Estimation from 10 cohort studies. Atherosclerosis. 2017;262: 51–54. doi:10.1016/j.atherosclerosis.2017.05.007

8. Pasolli E, Asnicar F, Manara S, Zolfo M, Karcher N, Armanini F, et al. Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. Cell. 2019;176: 649–662.e20. doi:10.1016/j.cell.2019.01.001

Chapter 3

Local patterns in variation in malaria prevalence within and between rural communities in Madagascar

Attributions:

The contents of this chapter are a part of a manuscript prepared for submission to the *Journal of Infectious Diseases*: **Local patterns in variation in malaria prevalence within and between rural communities in Madagascar** by Benjamin L. Rice, Christopher D. Golden, Hervet Randriamady, Andry Ny Aina Anjaharinony Rakotomalala, Miadana Arisoa Vonona, Evelin Jean Gasta Anjaranirina, James Hazen, C. Jessica E. Metcalf, and Daniel L. Hartl.

3.1 Abstract

Introduction

Recent reports indicate malaria control in Madagascar has been variable and experienced setbacks with some parts of the country experiencing increased incidence. This is in part due to dramatic variation across Madagascar in ecological and demographic parameters, such as access to health care resources, which can drive locally heterogenous malaria transmission dynamics. A greater understanding of the distribution of malaria infections at smaller spatial scales, such as within and between local communities, and how that varies across Madagascar's distinct eco-regions, is needed to better target and evaluate future interventions. However, this has been limited by a scarcity of malaria prevalence data with resolution at the sub-regional and sub-district levels.

Methods

To investigate spatial variation in malaria prevalence across ecological settings in Madagascar, we sampled 7128 individuals (all ages) from clusters of rural communities residing within five eco-regions in Madagascar: (1) humid, tropical rainforest in the northeast (administrative region:

Analanjirofo), (2) sub-tropical rainforest in the southeast (Vatovavy Fitovinany), (3) semi-arid, spiny forest in the southwest (Atsimo Andrefana), (4) dry, deciduous forest on the west coast (Atsimo Andrefana), and (5) sub-humid forest and grassland in the central plateau (Amoron'i Mania). For each of the five regions, in multiple communities (n = 6-7), all individuals (n = 1300-1597) in randomly selected households (n = 248-379 per region) were recruited for a total of 7,890 individuals sampled overall. Malaria prevalence was determined by rapid diagnostic test (RDT) cross-sectional surveys performed over 2013-2017.

Results

In the rural, and often remote, communities sampled in our study regions in Madagascar, multiple sites, especially in southeastern and southwestern Madagascar, were observed to have prevalence 2-5 times higher than the expected regional average. In several cases, prevalence between nearby (<50 km) communities varied substantially. Comparisons of observed age and household clustering data to simulations controlling for local variation in prevalence revealed that within some communities, especially in southwestern Madagascar, there was a significant excess of certain households with multiple infections.

Discussion

These data provide much needed baselines for under-studied communities in rural and remote areas of Madagascar that complement existing clinic-based data; baselines upon which ongoing and future malaria control programs can be assessed. Significant variation in prevalence between sites and between households in a region was observed, suggesting that currently reported estimates of national and regional averages are often poor predictors of malaria burden at the

local level in rural and remote communities in Madagascar. Such communities warrant greater prioritization in future malaria control efforts.

3.2 Introduction

3.2.1 Epidemiological context of malaria in Madagascar

Among the countries with over 300,000 malaria cases per year analyzed in the most recent annual malaria report by the World Health Organization, Madagascar was estimated to have the second largest increase in malaria incidence between 2016-2017 (WHO 2018). Additionally, active surveillance data from the most recent Malaria Indicator Survey (2016) from the National Malaria Control Program in Madagascar suggest that: (i) despite declines in some areas, large zones of the country, such as the west coast and south, remained at higher prevalence than was observed in 2011, and (ii) the proportion of the population living in high transmission areas (prevalence >20%) increased more than four-fold since 2011 (Kang 2018). Further, a recent analysis of trends in rapid diagnostic test (RDT) confirmed cases reported by primary and secondary health facilities found that the number of health districts reporting high monthly incidence (>50 per 1000) varied regionally but increased overall between 2010-2014 (Ihantamalala 2018). Together, these studies show that malaria remains a major public health concern in Madagascar and indicate that earlier advances in the control of malaria have stalled or reversed in some areas.

From an ecological perspective, Madagascar is known for the dramatic variation observed within the country and the exceptionally high biodiversity that results, in part, due to this ecological complexity (Myers 2000). At a broad scale, Madagascar has typically been divided into several geographic ‘eco-regions based on variation in key bioclimatic variables such

as rainfall, elevation, and predominant vegetation type (Goodman 2004). These vary from the high rainfall, warmer northern and eastern coasts, to the more arid southern and western coasts, and to the cooler, higher elevation central plateau. At a finer scale, factors such as varying proximity to water sources, unique locally confined eco-systems, and landscape transformation due to deforestation result in a highly variable landscape within each of the broad eco-zones. High rates of deforestation, largely driven by expansion in the land used for subsistence agriculture by rural populations (Sussman 1994), have led to widespread alarm among conservation organizations as it threatens or endangers a high proportion of Madagascar's endemic biodiversity (Raik 2007). Additionally, many of these ecological parameters are related to the potential intensity of malaria transmission locally.

For malaria in Madagascar, data are usually aggregated either: (i) by administrative subdivisions such as administrative regions (*faritra*) or health districts level (e.g., Ihantamalala 2018), or (ii) into one of four to five defined epidemiological zones ('faciès épidémiologiques") (MIS 20016, Mouchet 1995). For example, the national Malaria Indicator Survey (MIS) grouped data into: (a) the "equatorial" east coast, (b) the central "high plateau", (c) the mid-elevation "fringe" bordering the high plateau, (d) the southern "sub-desert", and (e) the semi-arid "tropical" west coast. A well-characterized general trend is that malaria prevalence is typically highest on the east and west coasts, and lower in the interior fringe and high plateau. Analysis of clinical reporting data has indicated that the semi-arid south and west are generally more prone to irregular peaks in incidence, while the east is characterized by more regular seasonal cycles (Kang 2018, Ihantamalala 2018). Because of this variation in malaria transmission dynamics within Madagascar, a sub-national perspective is necessary for monitoring and intervention efforts to be successful (Howes 2016).

Rural communities in Madagascar frequently have high rates of poverty, malnutrition, and anemia, along with poor access to healthcare infrastructure (WHO 2015). Because of this, several large food security, conservation, and health surveillance projects have been launched in recent years. For susceptibility to malaria, more distant, rural communities are those that are most likely to be located at the proceeding edge of forest clearance and landscape conversion – a factor associated with increased local density of malaria mosquito vectors (Zohdy 2016, Gilman 2006).

3.2.2 Spatial variation in malaria transmission

The drivers of the current difficulties in controlling malaria in Madagascar at a national level are complex but are thought to include deteriorating infrastructure related to political instability, especially in more remote areas; increased vector larval density due to agricultural expansion; human mobility; and varying access to diagnosis and treatment (Howes 2016). However, many of these factors act at more local, sub-regional, scales. For example, the rates of deforestation due to conversion to agricultural land and the frequency with which individuals migrate on a regular basis can vary greatly between locales within the same region (Ihantamalala 2018). Likewise, access to diagnosis and treatment can differ between those communities nearer urban or peri-urban infrastructure. Moreover, individual and household variables can mediate the risk of infection (e.g., Howes 2018). As a consequence, we expect variation in malaria burden between local communities within a given region and among and between households within a local community. However, there are few previous studies from Madagascar on sub-regional or sub-district variation in malaria prevalence and hence our ability to understand the magnitude of

such variation is limited. The paucity of studies also limits our understanding of the degree to which Madagascar's unique ecological context drives local malaria dynamics.

Practically, the observation of spatial variation is relevant for control efforts as it indicates that a disproportionately small portion of the population contributes a disproportionally large share of infections (Woolhouse 1997). These geographic sub-units that contain relatively higher infection rates are sometimes termed hotspots (e.g., Stresman 2018). If the pattern of spatial variation is well understood, control efforts can be made more efficient by targeting the hotspots where more infections are concentrated. Targeting hotspots is also important so as to limit the potential for the hotspots to serve as reservoirs from where infections can spread, and undo gains made elsewhere.

In order to explore sub-regional, sub-district spatial variation in Madagascar, we analyzed a newly generated data set compiled from recently completed field surveys in this study. This data set contains infection outcome (by RDT), geo-spatial, demographic, and paired questionnaire data for over 7000 individuals at 31 sites in 5 distinct eco-zones. The total number of samples analyzed here (n = 7128) is comparable to the total number analyzed in the most recent national RDT survey (n = 6569) but two notable differences are that: (i) in the study here, sampling was concentrated at much fewer sites so the number of samples per site is much greater and (ii) the number of samples per household is greater in this study as all individuals within an enrolled household were recruited, (versus individuals 6-59 months only in the national MIS). We use the denser local sampling as an opportunity to explore site-site variation across eco-zones, and variation within sites at the household and individual level.

The observation of an alarmingly high prevalence of infections within some rural communities in northeast (*faritra*: Analanjirofo region), southeast (*faritra*: Vatovavy

Fitovinany), and southwest (*faritra*: Atsimo Andrefana) Madagascar is worthy of attention. As is evidence that, especially in the southwest, infections are not randomly distributed among households. As these communities also reside in or near hotspots of poverty, malnutrition, and biodiversity, we argue that the co-location of malaria hotspots in these areas necessitates including urgent action on local malaria burdens in broader development and conservation objectives.

3.3 Materials and Methods

3.3.1 Sampling: Regions

Sampling was performed at 6-7 sites each for 5 distinct eco-regions in Madagascar. See Figure 1 for their locations, the administrative divisions they are within, sample size, and the time of sampling. Details of enrollment, ethical approval, and study design are explained by Golden et al 2017, Golden et al 2019, and Golden, Rice et al 2019. Briefly, sites in Region 1 (Northeast) were sampled as a part of prospective cohort studies performed between July 2013 – April 2014 and August 2016 – March 2017. Individuals were sampled at 3-4 month intervals in these cohorts. Due to not being sampled simultaneously, direct comparison between regions is complicated by inter- and intra-annual fluctuations in malaria transmission. However, sites within a region were sampled within the same seasonal window and on average sites were sampled within 3 weeks of each other. Regions 2-5 were sampled consecutively between January and August 2017 and at approximately the same season (following shortly after a region's peak rains).

Regions were selected for two primary reasons: (i) foremost, to capture the diversity of ecological contexts in Madagascar, and (ii) as they were within priority zones for large

conservation and food security project interventions. The regions sampled represent a majority of the eco-regions that have been described in Madagascar, but do not include two eco-regions: the interior mid-elevation fringe and the sub-desert of the deep south of Madagascar. However, these two unsampled eco-regions are estimated to contribute a small minority of the national malaria burden.

Of the regions sampled, four of the five regions reside near or within buffer zones for protected natural areas that are globally acknowledged biodiversity hotspots. Region 1 (Northeast) sites were adjacent to the Makira Natural Park or Masoala National Park, which together contain some of the largest remaining tracts of forest in Madagascar and the last remaining littoral forests. Region 2 (Southeast) contains some remaining tracts of the Eastern Coast Rainforest corridor that has been recognized as a UNESCO heritage natural site, although the sites studied were located farther from intact forests than in other regions. Communities in Region 3 (Southwest) and Region 4 (West Coast) reside near the Ifaty and Mikea forests, respectively, where remaining tracts of spiny and dry, deciduous forest have received attention recently due to being intensely threatened by charcoal production. All sampling was performed in rural areas with difficult or absent road access and where foraging for food, small-scale charcoal production, and subsistence agriculture were primary economic activities.

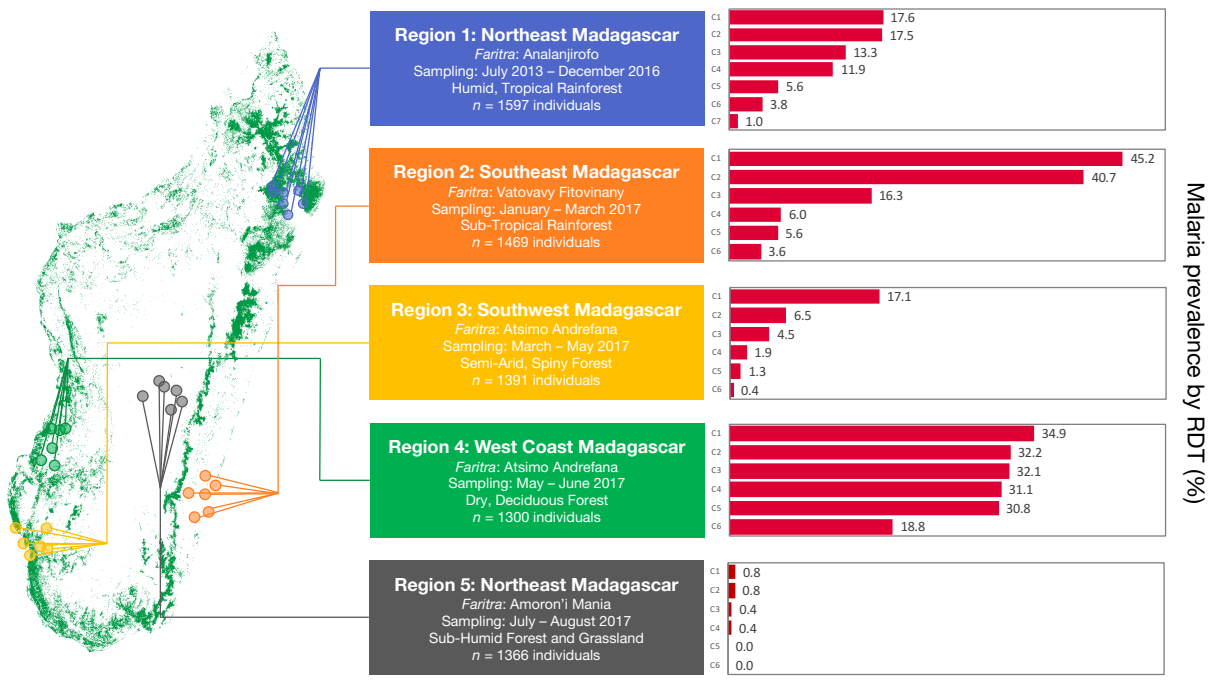

Figure 1

Sampling in Madagascar by eco-region

Regions as described by predominant vegetation type, their corresponding administrative region (*faritra*), sample size, *n*, and the dates of sampling are shown. The map shown is the approximate locations of study sites on the estimated remaining forested areas of Madagascar as of 2014 [CIRAD]. To the right, prevalence of malaria infection by rapid diagnostic test (RDT) is shown. Communities are ordered by prevalence within a region. The number of individuals sampled per site, *n*, ranged from 151 to 284 and the number of individuals positive for malaria ranged from 0 to 127.

3.3.2 Sampling: Communities

Within a region, rural communities were stratified by two principal variables in order to attempt to capture the ecological and epidemiological variation present in an area: proximity to the coast and proximity to urban infrastructure (e.g. the nearest town with road access). We defined roads as paths regularly (i.e. daily) passable by public or private vehicles. Due to inadequate census data or maps for these remote areas, sites were defined as clusters of approximately >30 households visible by satellite imagery. Due to the settlement pattern of rural communities in these areas, household clusters, their peripheral agricultural (mainly rice) fields, and their surrounding cleared or deforested lands were easily distinguishable in satellite imagery. Within strata, for each region, communities were randomly selected so as to have communities that were coastal (<3km from the coast) or inland (>3km from the coast), as well as communities that were within 15 km of an urban center or more than 15km from an urban center. For Region 5, the inland high plateau is far removed from the coast so by definition there were no coastal communities in this region. The urban centers of the districts studied were Maroantsetra, Mananjary, Toliara, Morombe, and Amobositra, respectively, for Regions 1-5. As communities were intentionally stratified to capture variation that may be present within regions, the communities selected are likely not a random sample of variation within a region. Observed site-site differences instead can be interpreted as estimates of the minimum range of variation that may be present within a region.

3.3.3 Sampling: Households

At each site, approximately 50 households were recruited for enrollment. Households were defined as groups of individuals that regularly cohabited and shared meals containing at

least one reproductive aged female and at least one child 5 years of age or younger. Households were randomly selected for recruitment from a census of all households matching these criteria that was updated prior to sampling by consultation with community leaders and heads of family groups. Some communities had less than 50 total households present, in which case all households were approached for enrollment. Household size (the current number of full-time and part-time residents within a household per the head of household) ranged from 2 to 19 (mean 5.2). Latitude and longitude of households were recorded as well as age, sex, and other socio-demographic data.

Notably, individuals of all sexes and ages within an enrolled household were recruited. Comparing reported household size to the number of individuals with a malaria RDT result and matching age and sex data showed that 86% of individuals within the recruited households were sampled for a total of 7198 individuals analyzed here. The percentage of individuals within households sampled is similar to that obtained in other studies of household prevalence in Madagascar (Howes 2018). Individuals not sampled were either physically absent from the community at the time of sampling, often due to hunting, fishing, or charcoal production activity, or declined to participate in blood sampling. Males aged 15-30 were the most commonly observed to not participate in the blood sampling for those reasons.

3.3.4 Sampling: Malaria prevalence by cross-sectional RDT survey

After obtaining informed consent or assent, a blood sample was obtained by veni-puncture or finger prick. Blood sampling procedures were performed by physicians and phlebotomists – either medical personnel from the local district health office or a trained team member from local partner development organizations with long-standing ties to local

communities. Aside from individual consent or assent, efforts to obtain community-wide understanding of the purpose and protocol of blood sampling included public meetings with the community as a whole, with local officials, with traditional community leadership, and with individual family groups (for details, see Golden et al 2017, Golden et al 2019, and Golden and Rice et al 2019; Chapter 2 in this dissertation).

Following the manufacturer's protocol, a portion of the blood sample was transferred to a SD BIOLINE® Pf/Pan (Standard Diagnostics, Gyeonggi-do, Korea, cat. no. 05FK60) rapid diagnostic test (RDT) which uses the antigens HRP2 and pLDH. These RDTs and markers have been assessed in Madagascar previously (Rice 2016). For RDTs with an invalid result, some were repeated if sufficient volume remained from the blood sample after reading the first RDT. Of the 7198 individuals sampled, 811 were positive by RDT (11.1%), and 30 had an invalid RDT result (0.4%) where the control band did not display properly. Individuals testing positive by RDT were offered a consultation with a physician and standard antimalarial medication.

3.3.5 Analysis of the distribution of infections among households

To explore the distribution of infections among households within a site, we performed simulations to compare the observed frequency of multiple infections in a household to that expected under a scenario with no household clustering. In the absence of clustered infections at the household level, the expected number of households with a given number of infections (e.g. the expected number of households with 0, 1, 2, or 3 etc. infections) depends solely on household size and an individual's independent probability of infection. A significant excess or rarity of households with multiple infections in comparison to the random expectation would suggest that some variable acting at the household level significantly affects the probability of infection. We

performed simulations using the prevalence of infection at a site and the observed household size distribution (the number of households of a given size) using 10,000 bootstrap resamples to assess significance.

3.3.6 Ethical approval

Details are described in Golden et al 2017, Golden et al 2019, and Golden and Rice et al 2019. Briefly, approval was obtained from the IRB committee at the Harvard TH Chan School of Public Health's Committee on the Use of Human Subjects, the Malagasy Ministry of Health (MOH), the ethical committee of the Institut National Santé Publique et Communitaire (INSPC) in Madagascar, district medical inspectors, and local community leaders (e.g. *Presidént-fokontany*). Adults provided informed consent for themselves and surrogate consent for infants and children, while other older children provided assent prior to participation.

3.4 Results

3.4.1 Variation in prevalence between communities

Within certain regions of Madagascar, we observed wide variation between sites in malaria prevalence by RDT (see Figure 1). For the 31 rural sites studied here, prevalence varied from 0% at two sites in the high plateau (sites 5.4 and 5.6) to over 30% at seven sites (site 2.5 and 2.6 in the southeast and sites 4.1, 4.2, 4.3, 4.4, and 4.5 in the west coast region). Prevalence was, as expected, lowest in the higher elevation, cooler Region 5 (high plateau). Prevalence was highest in the southeast sites, where site 2.5 (prevalence 40.7%), a site near the Pangalana Canal in Mahatsara-Sud commune in the Mananjary district, and site 2.6 (prevalence 45.2%), a site in

the Antsenavolo commune in the Mananjary district, had very high prevalence. Despite having two sites with prevalence over 40%, the southeast region also contained three sites with prevalence less than 10% (sites 2.1, 2.2, and 2.3). For comparison, estimated regional average prevalence for the east coast as a whole per the national MIS was 9.0%. In the west coast region, among rural communities in Morombe district, prevalence was consistently high and was above 30% in five of the six surveyed communities. For comparison, regional average prevalence across the West Coast was 8.8% in 2016 (MIS 2016).

In several cases, large variation was observed between pairs of geographically close sites (see Figure 2). For example, site 2.1 (prevalence 5.6%) and site 2.5 (prevalence 40.7%) are less than 15 km apart but varied almost 8-fold in prevalence. Another example, site 3.2 (prevalence 1.3%) and site 3.3 (prevalence 17.1%) were less than 5 km apart yet also had a large difference in prevalence. Indeed, the variation in malaria prevalence seen between sites within 15km was often greater than that seen between distant and ecologically distinct eco-regions such that some sites in the more arid southwest had higher prevalence than sites in the wetter and generally more malaria prone east coast.

Atsimo Andrefana, a large administrative region in western and southwestern Madagascar, is worthy of mention. Our sampling regions Region 3 (Toliara II district in southwest Madagascar) and Region 4 (Morombe district in the west coast) both lie within with the Atsimo Andrefana administrative region. However, the pattern of consistently high prevalence seen among the rural sites in the Morombe district differs greatly from that seen in the Toliara II district.

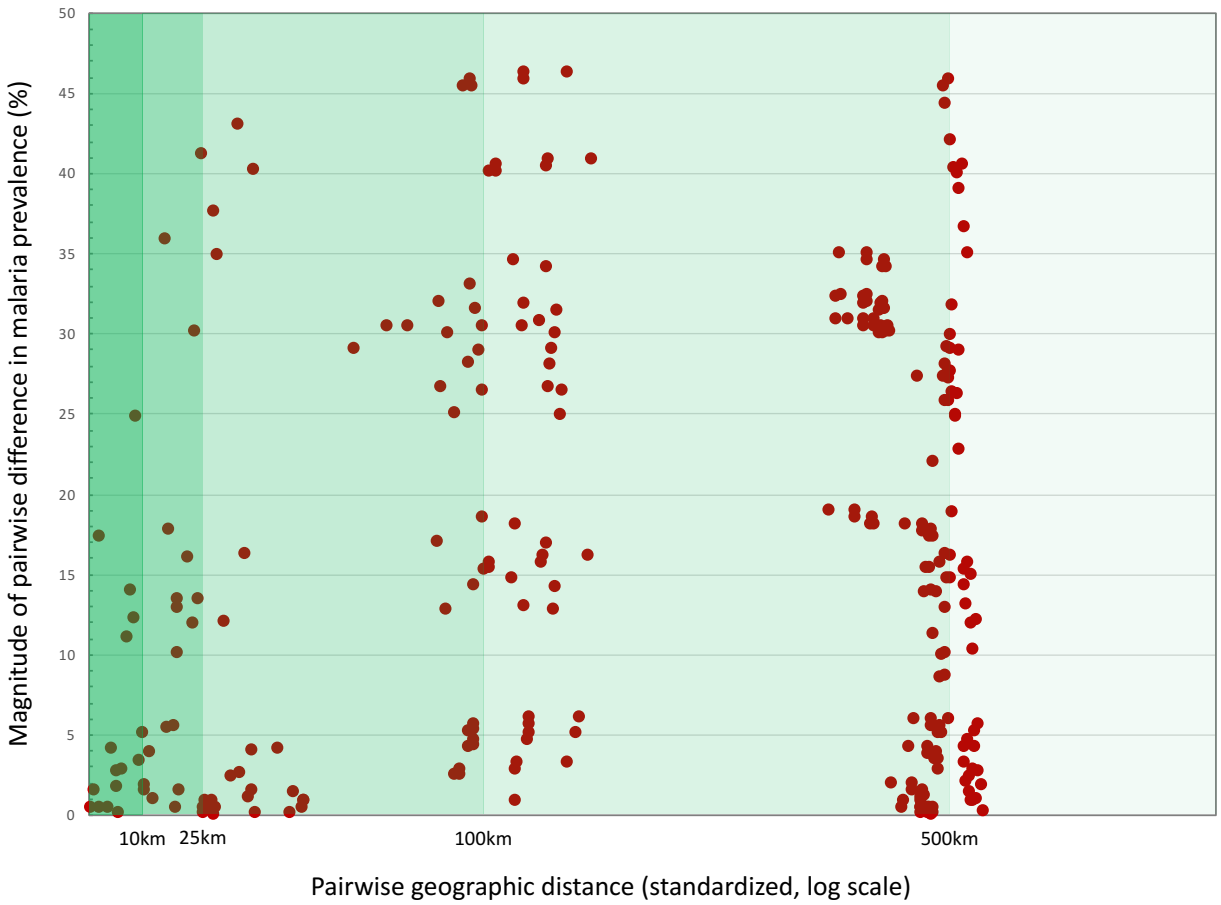

Figure 2

Comparing variation in prevalence between communities versus their geographical distance

Difference in prevalence is shown as the magnitude of the difference between all possible pairwise comparisons between communities. Geographical distance is shown as the log-scaled Euclidean distance between the latitude and longitude of the midpoints of communities. The sections of the plot corresponding to distances of 10, 25, 100, and 500+ kilometers are shaded.

3.4.2 Variation within communities: Age and sex

To explore variation in the distribution of infections within the rural communities in Madagascar, we first characterized the age and sex distribution of infections. For age, a disproportionately high percentage of infections were among individuals 4 to 20 years of age (see Figure 3). The reduced percentage of infections among very young children (ages 0-3) and the persistence of an increased probability of infection into years over 12 are noteworthy. The age and sex distribution of infections in Region 5 (high plateau) is disjoint due to the limited RDT positive individuals ($n = 6$). For sex, there were small differences in the percent of infections that were male or female among some age groups in some regions, but no consistent pattern was apparent. Overall, a similar percentage of infections was observed among males and females (52.8% and 47.2%, respectively).

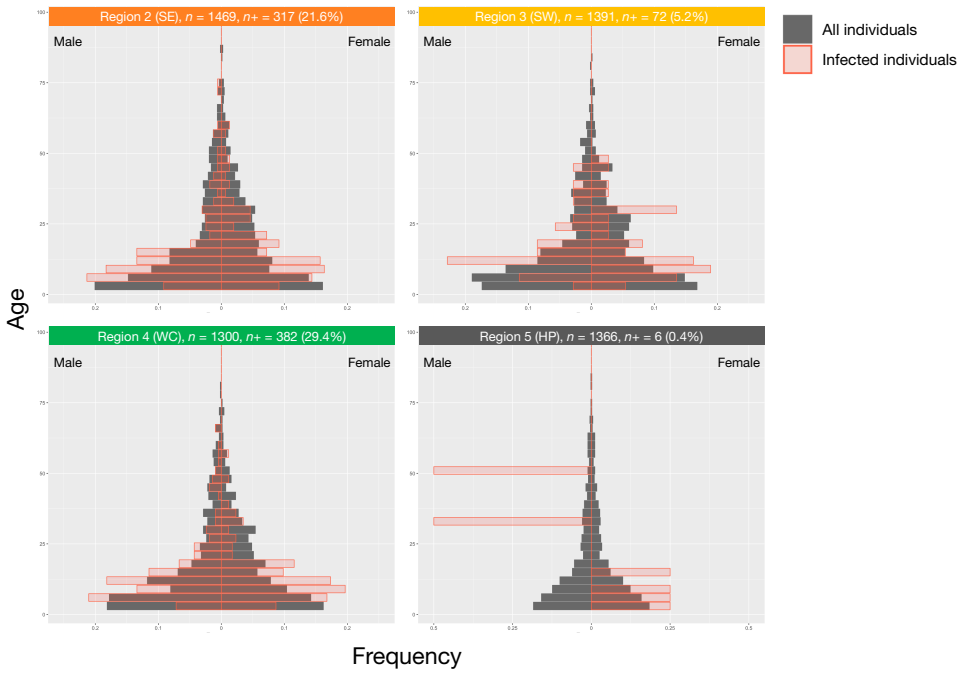

Figure 3

Age and sex distribution of infections

The distributions of all individuals (gray) and infected individuals (red) are shown by region. The proportion of individuals within a 3-year age bin is shown with males on the left and females on the right. Sample size (*n*) and the number of individuals positive (*n*+) are shown by region (southeast, SE; southwest, SW; west coast, WC; high plateau, HP).

3.4.3 Variation within communities: distribution of infections among households

In order to explore whether the distribution of infections among households was random, we determined the proportions of households found with 0, 1, 2, 3, or 4+ individuals testing positive for malaria by RDT by region (see Figure 4). Notably, high proportions of the households sampled contained at least one individual infected at the time of sampling in Regions 2, 3, and 4. This varied from 17% of households containing a malaria infection in Region 3

(Toliara II district, southwest Madagascar) to 63% of households containing a malaria infection in Region 4 (Morombe district, west coast). Alarmingly, a substantial proportion of households in Region 2 (Mananjary district, east coast) (7%) and Region 4 (Morombe district, west coast) (9%) were burdened with 4 or more infections at the time of sampling.

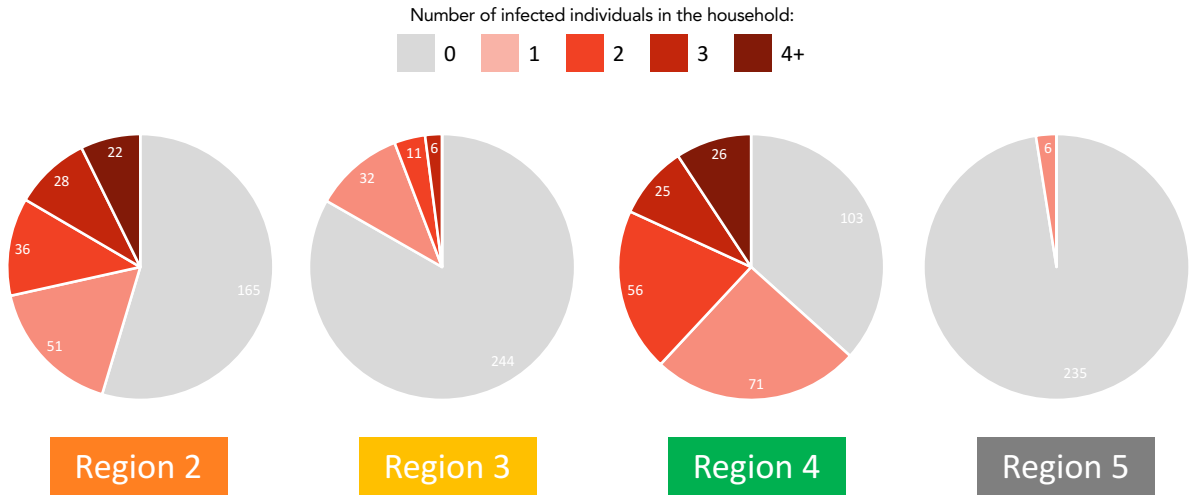

Figure 4

Households by the number of infected individuals in the household.

Within slices, the number of households observed to have the given number of infections is shown.

Regardless of the determining factors, observing a substantial proportion of households with multiple infections is noteworthy and has implications for control efforts. For example, 8 of

8 individuals in a household at site 2.5 were positive for malaria by RDT at the time of sampling. Such high burden at the household level is perhaps unexpected in a region of Madagascar that, from previous estimates, has a regional average prevalence of approximately 9%.

However, to determine whether the observed level of household clustering of infections was significantly different from that expected by prevalence and the distribution of household size alone, we performed simulations where infected individuals were randomly re-assigned to households. Such permutation tests allow comparing the observed number of households with 0 infections to the number of households expected. Six of the twenty-four sites analyzed, including four of the six sites in Region 4 and two of the six sites in Region 2, had counts of households with 0 infections that significantly deviated from the expected range (see Figure 5A). An excess of households with 0 infections requires that a disproportionately small number of households contain a disproportionately high number of infections and indicates that infections are clustering nonrandomly at the household level in those sites.

Even within sites where the household distribution of infections observed could be expected by random chance, there was wide variation in the proportion of individuals infected in a household. For example, at site 2.5, the level of household clustering was not significant but 24/50 households (46%) had at least half of the individual members of the household test positive (see Figure 5B). This illustrates the perhaps counter-intuitive result of the simulations that due to variation in prevalence between sites, and variation in household size, infections are expected to be highly concentrated in some households for some sites by random chance alone. The implication is that high malaria burden within some households in Madagascar, which likely has broad negative effects on the household, can be expected even in areas with low average prevalence.

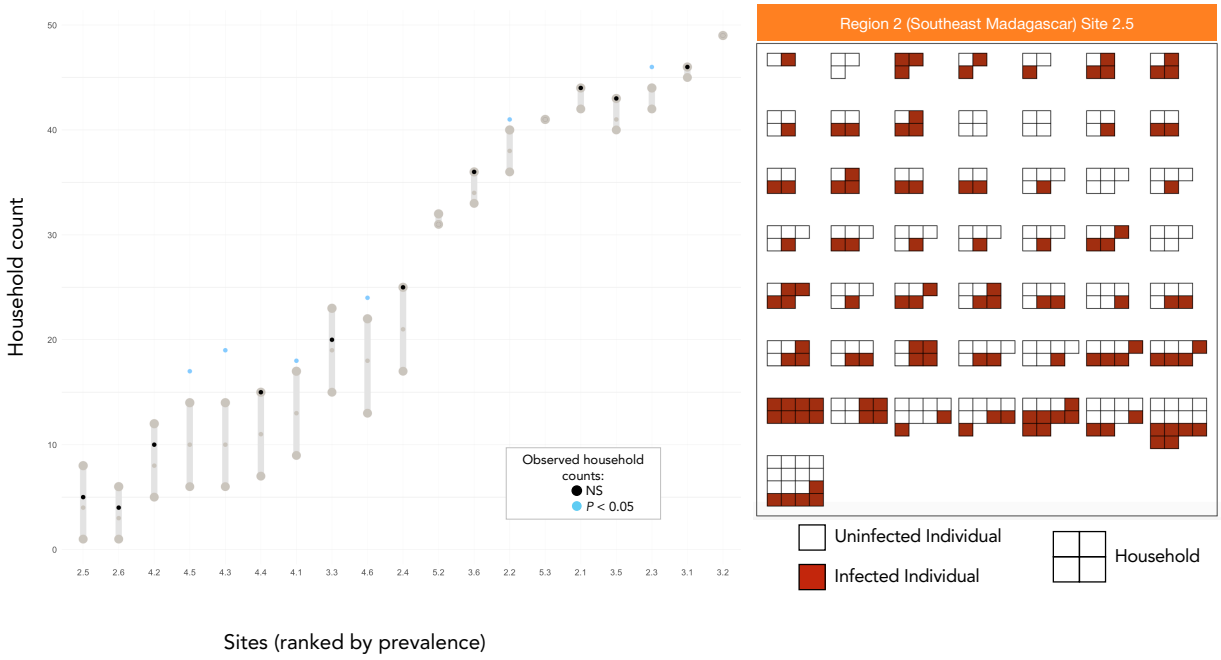

Figure 5

Distribution of infections among households

(5A, left) Comparing the observed number of households with 0 infections for a community (shown in black) to the distribution expected from simulations using the site's prevalence and observed household structure. 95% confidence intervals, as determined by permutation testing (10,000 replicates), and the median are shown in gray and sites with significant deviations from the null expectation obtained from simulations are shown in blue. (5B, right) The observed household structure and distribution of infections among households for Site 2.5 from Southeast Madagascar is shown. Households are listed by number of individuals.

3.5 Discussion

3.5.1 Providing a malaria prevalence baseline in remote, rural communities in Madagascar

Due to the limited data available on the sub-regional and sub-district variation in malaria prevalence in Madagascar, trends in malaria control across distinct ecological settings have been difficult to ascertain. At the sites sampled within the 5 eco-regions sampled here we provide high resolution estimates of malaria prevalence at the site and household level. These estimates can be compared against future data to assess progress in the control of malaria. We highlight some sites in particular, especially in the southeast and the west coast of Madagascar, where multiple sites with prevalence over 30% were observed that warrant greater attention from malaria control programs.

Future studies can also determine whether the sites sampled here, selected in part with the intention to capture distal ends of the variance present with a region, are outliers or if they signify the presence of many remote, rural sites with high malaria burden in these areas. Regardless, such baseline data is useful to malaria control plans that are increasingly focusing on better surveillance to more efficiently identify and target local areas with elevated malaria burden.

3.5.2 National and regional averages do not predict local burden

At a minimum, the observation of multiple sites with greatly elevated prevalence provides convincing evidence that relative hotspots of malaria prevalence exist among the rural and remote sites in Madagascar. That these hotspots of malaria prevalence co-locate with development and conservation priorities, especially in northeast and southwest Madagascar, warrants greater attention to local malaria prevalence in ongoing and future efforts.

The observation of an excess of infections among the households containing one or more infection among a majority of the sites sampled in the west coast of Madagascar indicates that burden at the household level may also not be predicted from average prevalence rates. Two implications are that (1) interventions that target affected households may be an effective way to target a large share of total infections and (2) hierarchical models that characterize the proportion of variation that is explained at different spatial scales are needed to understand the relative importance of individual, household, community, and regional factors in these regions of Madagascar.

3.5.3 Limitations of cross-sectional prevalence surveys

Conclusions drawn from the data analyzed here are limited by the fact that temporal fluctuations, while common in malaria, can't be studied from the cross-sectional sampling that produced a majority of the data analyzed. Cross-sectional data does not allow an understanding of historical trends at a site and as a result we cannot easily exclude that observations of high prevalence are a result of transient peaks in transmission.

However, multiple lines of evidence suggest that the large variation observed between sites is not due to recent temporal fluctuations alone. First, nearby sites that were sampled within 1-2 weeks of each other (equivalent to a single generation in the parasite life cycle) were observed to vary greatly and in some cases had a greater magnitude of variation than sampling performed at different seasons. Second, sampling at the regions with the highest observed prevalence occurred after (albeit shortly) the typical seasonal peaks in incidence (Kang 2018). This would suggest that prevalence at the sites sampled reached even higher levels prior to the time of sampling. An exception is region 5 where sampling occurred during colder months (July-

August) where local transmission is severely limited. Third, if site-site variation was explained by random fluctuations arising from the sampling strategy, then we would expect a similar magnitude of variation in all regions. However, prevalence at sites in Region 4 (west coast) was determined with the same sampling strategy but was found to be much less variable than in other regions. Fourth, if the high prevalence observations were a result of false positive RDT results than we would expect uniformly high prevalence results when applying the same methods across sites. That prevalence was, as expected, much lower in the higher elevation Region 5 sampled at the coldest season of the year makes it unlikely that an artifact of our RDT sampling explains the results.

One possible hypothesis that could explain the observed fine scale spatial variation is that variation in local prevalence is correlated with temporal variation, such that the variation observed between locales is a consequence of roving local outbreaks causing sporadic peaks in local transmission. Analysis of genetic data can be of use to test this hypothesis as sporadic, temporary outbreaks of malaria, as opposed to stable local transmission, are expected to result in different patterns of genetic variation. If the large site-site variation observed in Region 2 and Region 3 is a consequence of temporary, mobile increases in transmission than we would expect fewer unique parasite genotypes to be circulating at a given site. Repeated bottlenecking in the local parasite population as transmission rises and falls rapidly over time would reduce the number of distinct genotypes that could arise via recombination. On the other hand, less site-site variation in Region 4 may correlate with more local stability in malaria transmission over time at a given site, in which case more opportunities would accumulate for recombination among parasite lineages to produce novel genotypes.

Curiously, recent analysis of clinical reporting trends indicate that temporal instability is greater in the more arid southwest and west coast while seasonal cycling is more regular on the east coast (Kang 2018), where we observed high site-site variation over our sampling interval at our sites on the east coast. It is likely that untangling temporal and spatial variation in malaria prevalence requires additional data sources.

Follow-up studies to determine the temporal stability of observed hotspots are needed to assess their significance to malaria control efforts. If hotspots are ephemeral, then interventions should focus on rapid response, while stable, entrenched reservoirs of high transmission indicate that interventions should focus on extending coverage to remaining un-resolved hotspots.

Regardless of the contribution of temporal or spatial drivers to local variation in malaria prevalence, these data highlight that some sites and households among rural populations in Madagascar have an alarmingly high burden of malaria infection. For example, at the time of sampling at site 4.4 in Basibasy commune in Morombe district adjacent to the Mikea Forest, over 70% of households had at least one individual infected with malaria and over 40% of households had two or more individuals infected. It seems unlikely that the community's ability to respond to conservation and development efforts is not negatively affected by that burden. Indeed, a recent modeling study shows that controlling malaria burdens is expected to have a significant effect on the ability of rural households in Africa to rise from poverty over the coming decades (Willis 2018). We conclude by arguing that conservation and development efforts should account for the presence of local hotspots in malaria prevalence in rural communities in Madagascar.

Acknowledgements

First, we would like to thank members of the local communities and local community leaders for their participation. The MAHERY field teams provided invaluable assistance in completing the field sampling: Mamy Yves Andrianantenaina, Rebaliha Mahery, Fenoarison Faraniaina Robuste Mahonjolaza, Raelson Hermann Paratoaly, Vololoniaina Ravo Rakotoarilalao, Jean Frederick Randrianasolo, Alex Dominique Rasamison, Ambinintsoa Nirina Tafangy, Adolphe, Rivo, Gerandine, and Papan'i Aime. Amy Wesolowski, Amy Winter, and Nicholas Arisco provided fruitful discussions while in the field and while analyzing data.

Author Contributions

BLR, CDG, JH, CJE, and DLH conceived and designed the study. BLR, CDG, HR, ANAAR, and MAV performed blood sample processing and collected the data. HR and EJGA supervised survey teams and the collection of demographic data. BLR analyzed the data. BLR, CDG, and DLH contributed to the interpretation of the data. BLR drafted the manuscript with assistance from CDG and DLH. All authors contributed to and approved the final manuscript.

References

1. World Malaria Report 2018. Geneva: World Health Organization; 2018

2. Malaria Indicator Survey: Enquête sur les Indicateurs du Paludisme 2016 (MIS 2016). Programme National de Lutte contre le Paludisme Antananarivo, Madagascar

3. Kang SY, Battle KE, Gibson HS, Ratsimbasoa A, Randrianarivelojosia M, Ramboarina S, et al. Spatio-temporal mapping of Madagascar's Malaria Indicator Survey results to assess Plasmodium falciparum endemicity trends between 2011 and 2016. BMC Medicine. 2018;16. doi:10.1186/s12916-018-1060-4

4. Ihantamalala FA, Rakotoarimanana FMJ, Ramiadantsoa T, Rakotondramanga JM, Pennober G, Rakotomanana F, et al. Spatial and temporal dynamics of malaria in Madagascar. Malaria Journal. 2018;17. doi:10.1186/s12936-018-2206-8

5. Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J. Biodiversity hotspots for conservation priorities. Nature. 2000;403: 853–858. doi:10.1038/35002501

6. Goodman, Steven & P. Benstead, Jonathan & Schutz, Harald. (2004). The Natural History of Madagascar.

7. Sussman RW, Green GM, Sussman LK. Satellite imagery, human ecology, anthropology, and deforestation in Madagascar. Human Ecology. 1994;22: 333–354. doi:10.1007/BF02168856

8. Raik D. Forest Management in Madagascar: An Historical Overview. Madagascar Conservation & Development. 2007;2. doi:10.4314/mcd.v2i1.44123

9. Mouchet J, Blanchy S. Particularities and stratification of malaria in Madagascar. Sante 1995;5(6):386-8

10. Howes RE, Mioramalala SA, Ramiranirina B, Franchard T, Rakotorahalahy AJ, Bisanzio D, et al. Contemporary epidemiological overview of malaria in Madagascar: operational utility of reported routine case data for malaria control planning. Malaria Journal. 2016;15. doi:10.1186/s12936-016-1556-3

11. WHO. The global prevalence of anaemia in 2011. Geneva: World Health Organization; 2015

12. Zohdy S, Derfus K, Headrick EG, Andrianjafy MT, Wright PC, Gillespie TR. Small-scale land-use variability affects Anopheles spp. distribution and concomitant Plasmodium infection in humans and mosquito vectors in southeastern Madagascar. Malaria Journal. 2016;15. doi:10.1186/s12936-016-1164-2

13. Gilman RH, Tielsch J, Glass G, Shields T, Vittor AY, Lozano WS, et al. THE EFFECT OF DEFORESTATION ON THE HUMAN-BITING RATE OF ANOPHELES DARLINGI,

THE PRIMARY VECTOR OF FALCIPARUM MALARIA IN THE PERUVIAN AMAZON. The American Journal of Tropical Medicine and Hygiene. 2006;74: 3–11. doi:10.4269/ajtmh.2006.74.3

14. Ihantamalala FA, Herbreteau V, Rakotoarimanana FMJ, Rakotondramanga JM, Cauchemez S, Rahoilijaona B, et al. Estimating sources and sinks of malaria parasites in Madagascar. Nature Communications. 2018b;9. doi:10.1038/s41467-018-06290-2

15. Howes RE, Zikursh M, Cramer EY, Ramiranirina B, Ratsimbasoa A, Kang SY, Zimmerman PA, et al. Risk Factors for Malaria Infection in Central Madagascar: Insights from a Cross-Sectional Population Survey. The American Journal of Tropical Medicine and Hygiene. 2018;99: 995–1002. doi:10.4269/ajtmh.18-0417

16. Woolhouse MEJ, Dye C, Etard J-F, Smith T, Charlwood JD, Garnett GP, et al. Heterogeneities in the transmission of infectious agents: Implications for the design of control programs. Proceedings of the National Academy of Sciences. 1997;94: 338–342. doi:10.1073/pnas.94.1.338

17. Stresman GH, Mwesigwa J, Achan J, Giorgi E, Worwui A, Jawara M, et al. Do hotspots fuel malaria transmission: a village-scale spatio-temporal analysis of a 2-year cohort study in The Gambia. BMC Medicine. 2018;16. doi:10.1186/s12916-018-1141-4

18. Rice BL, Golden CD, Anjaranirina EJG, Botelho CM, Volkman SK, Hartl DL. Genetic evidence that the Makira region in northeastern Madagascar is a hotspot of malaria transmission. Malaria Journal. 2016;15. doi:10.1186/s12936-016-1644-4

19. Willis DW, Hamon N. Impact of eliminating malaria by 2040 on poverty rates among agricultural households in Africa. Gates Open Res 2018, 2:69 (https://doi.org/10.12688/gatesopenres.12849.1)

Chapter 4

Genetic evidence that the Makira region in northeastern Madagascar is a hotspot of malaria transmission

Attributions:

The contents of this chapter were published in the *Malaria Journal* ((2016) 15:596): **Genetic evidence that the Makira region in northeastern Madagascar is a hotspot of malaria transmission** by Benjamin L. Rice, Christopher D. Golden, Evelin Jean Gasta Anjaranirina, Carolina Mastella Botelho, Sarah K. Volkman, Daniel L. Hartl.

4.1 Abstract

Background

Encouraging advances in the control of *Plasmodium falciparum* malaria have been observed across much of Africa in the past decade. However, regions of high relative prevalence and transmission that remain unaddressed or unrecognized provide a threat to this progress. Difficulties in identifying such localized hotspots include inadequate surveillance, especially in remote regions, and the cost and labor needed to produce direct estimates of transmission. Genetic data can provide a much-needed alternative to such empirical estimates, as the pattern of genetic variation within malaria parasite populations is indicative of the level of local transmission. Here, genetic data were used to provide the first empirical estimates of *P. falciparum* malaria prevalence and transmission dynamics for the rural, remote Makira region of northeastern Madagascar.

Methods

Longitudinal surveys of a cohort of 698 total individuals (both sexes, 0-74 years of age) were performed in two communities bordering the Makira Natural Park protected area. Rapid diagnostic tests, with confirmation by molecular methods, were used to estimate *P. falciparum* prevalence at three seasonal time points separated by 4-month intervals. Genomic loci in a panel

of polymorphic, putatively neutral markers were genotyped for 94 *P. falciparum* infections and used to characterize genetic parameters known to correlate with transmission levels.

Results

Overall, 27.8% of individuals tested positive for *P. falciparum* over the 10-month course of the study, a rate approximately 7-fold higher than the countrywide average for Madagascar. Among those *P. falciparum* infections, a high level of genotypic diversity and a high frequency of polygenomic infections (68.1%) were observed, providing a pattern consistent with high and stable transmission.

Conclusions

Prevalence and genetic diversity data indicate that the Makira region is a hotspot of *P. falciparum* transmission in Madagascar. This suggests that the area should be highlighted for future interventions and that additional areas of high transmission may be present in ecologically similar regions nearby.

4.2 Introduction

Approximately 300,000 cases of *Plasmodium falciparum* malaria are recorded annually in Madagascar, an incidence rate that, at the reported level, places Madagascar and its population of 24.9 million among the lower tier of African countries in terms of malaria burden [1]. Updated estimates of transmission across Africa based on large databases of field surveys and sophisticated model-based geostatistics are now available and are less dependent on clinical reporting. These model-based estimates indicate that the number of reported cases in Madagascar is likely a large underestimate of the true burden, but still rank Madagascar as a low transmission setting relative to much of mainland Africa [2].

Contrasting with these low estimates for Madagascar are recent longitudinal and cross-sectional cohort studies that have repeatedly found localities with transmission and prevalence rates 3-10 fold higher than the national average (for example, see [3-5]). Further, these estimates are comparable with recognized high transmission areas globally. The localities of higher transmission and prevalence observed in Madagascar were often found in rural areas in the eastern part of the country, which is primarily characterized by lowland tropical rainforest. For example, prevalence of *P. falciparum* was 20.5% in a cohort of pregnant women reporting to antenatal clinics near Manakara [5] and varied temporally from 19.7 to 35.2% among adults in the Alaotra-Mangoro region [6]. For comparison, the average prevalence estimated from monitoring sentinel health sites was 3.1% nationally and peaked at 4.9% in the eastern part of the country [7]. Additionally, transmission was found to be newly increasing in some areas of eastern Madagascar, with a large outbreak in the Vatovavy-Fitovinany and Atsimo-Atsinanana regions in 2011-2012 as an example [8].

In terms of malaria transmission, Madagascar is typically divided into four to five zones based on broad climatic and demographic regions: the south, the eastern coast, the western coast, the high central plateau and sometimes with the margins/periphery of the high plateau included as a separate zone [7,15,38]. Transmission rates have been found to be highest along the eastern coast (4.9-11.7%) or the western coast (4.6-14.3%), with estimates varying with the age group and seasonality of sampling [7,15].

The continued reports of localized areas of elevated transmission, sometimes termed foci [9,10] or hotspots [11], which far exceed those national or regional level estimates suggest that monitoring efforts, and progress in reducing the burden of malaria, have been unevenly distributed. If unaddressed, such hotspots have the potential to undermine the effect of the recent scaling-up of malaria control in Madagascar [11,12]. Given the historical context of low intervention coverage and low funding for interventions in Madagascar in comparison to other African countries [1,13,14], identifying such hotspots of transmission is also valuable in order to efficiently direct resources to the areas where scale-up is most urgently needed. However, to do this, a better understanding of both the geographic distribution and transmission dynamics of such hotspots within Madagascar is required.

In terms of geographic distribution, a majority of the cohort studies that have measured malaria transmission and prevalence were performed in the southern parts of eastern Madagascar and, although the areas are climatically similar, it is unknown if comparable results are to be expected in the Northeast. Additionally, much of the population of the Northeast lives in rural and remote forested districts that are not well monitored by national surveillance efforts [3,15].

In terms of transmission dynamics, it is also unknown if local hotspots within Madagascar are the result of short-term spikes in transmission triggered by transient changes in

ecological or epidemiological conditions, or are due temporally stable transmission. Genetic data can be of great utility in discriminating between the two transmission scenarios without the need for labor-intensive direct estimates of transmission. The discrimination is possible because different levels of transmission intensity and stability are expected to produce different patterns of genetic variation in the *Plasmodium* parasite population [16-18]. Specifically, high, stable transmission is expected to result in a diverse assemblage of parasite genomes circulating locally, and a high percentage of infections that contain multiple, distinct *P. falciparum* genomes concurrently (termed polygenomic infections). On the other hand, in situations with low or sporadic transmission, few hosts on average are expected to encounter multiple, distinct *P. falciparum* multi-locus genotypes (MLGs) during the course of an infection, reducing the proportion of infections that are polygenomic. As a result, when transmission is unstable, opportunities for re-assortment and recombination between parasite MLGs are rare and few distinct MLGs circulate. Sporadic outbreaks, although capable of infecting a large percentage of the population when they do occur, are then expected to be composed of relatively few MLGs (albeit with some at potentially high frequency).

Here, the observation of an apparent hotspot of malaria prevalence in northeastern Madagascar is reported and a pattern of genetic variation consistent with high, stable transmission observed. This is inferred from rapid diagnostic test (RDT) surveys and single nucleotide polymorphism (SNP) genotyping of 94 *P. falciparum* positive samples from a prospective cohort study of 698 individuals living in the remote Makira region. These data highlight the need for greater monitoring of a previously understudied area in Madagascar and provide a baseline against which future interventions can be measured.

4.3 Materials and Methods

4.3.1 Study sites and study population

The Makira region is a large rural area mostly within the Analanjirofo administrative region (*faritra*) in northeastern Madagascar. It contains the protected Makira Natural Park (372,470 hectares) and buffer zones (350,000 hectares) in the surrounding low- and medium-altitude forest [19]. The park and buffer zones combined have an area of approximately 722,000 hectares and support a population of more than 127,000 people [20]. This area has been previously studied as a part of an effort to understand linkages between the conservation of its extraordinary biodiversity and human health (for e.g., see [19,21]). Communities in the Makira region consist of small, dense clusters of approximately 50-200 households. Two communities (*fokontany*) were selected. For the privacy of study participants, the names of these communities are withheld per the recommendations of the ethical body that reviewed the study and in accordance with previous studies that have been performed in these communities [21]. See Figure 1 for the approximate geographic locations. The two communities are separated by 7.2 km; however, community 2 is at a notably higher elevation (350 m) than community 1 (20 m). Neither community has road access and the distance to the nearest reporting sentinel health site (see ref. [7]) is 23-32km. Households were randomly selected within the communities. All individuals in 97 and 57 households, in community 1 and community 2, respectively, were enrolled for a total of 698 individuals.

Fig. 1

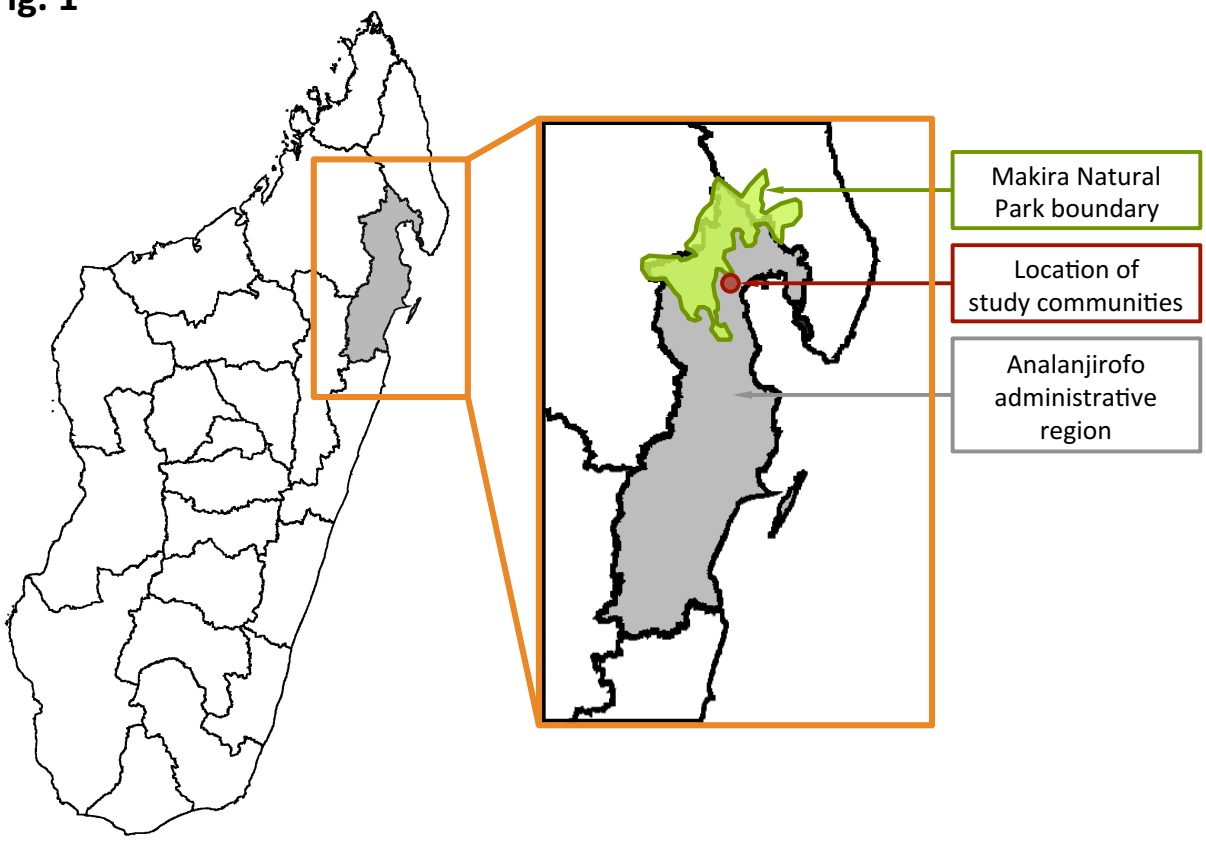

Figure 1

Location of the Makira Region and the study communities in northeastern Madagascar

The Analanjirofo administrative region (*faritra*) is colored gray. The approximate boundary of the Makira Natural Park is shown in green. Approximate location of the study communities is shown in red.

4.3.2 Ethical approval

Ethical approval was given by the Madagascar Ministry of Health and the Office for the Protection of Human Subjects at the Harvard T.H. Chan School of Public Health (IRB 22826).

Approval for the study was also sought and obtained from the Maroantsetra regional medical inspector, the mayor covering the region including both communities, and the leader (president-fokontany) in each community. All adults provided informed consent for themselves and surrogate consent for infants and children, while other older children provided assent prior to interaction with the local research staff.

4.3.3 Sample collection

For each individual, a blood sample was obtained by venous blood draw, or by finger-prick or heel-prick in a few cases for younger children. Five microlitres of blood were added to an RDT and approximately 200 microlitres of blood were added to an FTA card (Whatman FTA Classic, GE Healthcare, Marlborough, MA, USA, cat. no. WB120312) to preserve DNA for later genetic analyses. The RDTs used were: First Response® Combo Pf/Pan (Premier Medical Corporation Limited, Kachigam, UT, India, cat. no. I 16 FRC), CareStart® Pf/Pan (ACCESS BIO, Inc., Somerset, NJ, USA, cat. no. G0131), and SD BIOLINE® Pf/Pan (Standard Diagnostics, Gyeonggi-do, Korea, cat. no. 05FK60)– all of which used the antigens HRP2 and pLDH. All individuals testing positive by RDT were offered treatment with the standard first-line treatment (artesunate + amodiaquine, AS+AQ). Blood spots on FTA cards were allowed to dry and then stored at ambient temperature with desiccant in plastic bags.

Individuals were sampled at up to three time points: time point 1 (July/August 2013), time point 2 (November/December 2013), and time point 3 (March/April 2014). These sampling times correspond to the medium, low and high seasons of rainfall in the area, respectively. The number of individuals with a valid RDT result varied from 495 to 660 per time point; 698

individuals had an RDT result for at least one time point and 408 had an RDT result for all three time points.

4.3.4 DNA extraction and molecular detection of Plasmodium species

For a randomly chosen subset (n = 166) of the samples positive by RDT for *P. falciparum*, DNA was extracted from the corresponding dried blood spots using the DNA IQ™ Casework Pro for Maxwell® 16 kit. Following DNA extraction, real-time PCR (RT-PCR) assays for *P. falciparum, Plasmodium vivax, Plasmodium malariae,* and *Plasmodium ovale* were performed in order to confirm the presence of *Plasmodium* parasite genetic material. The assay for *P. falciparum* was as described previously [22], while the assays for the other three species followed a similar protocol targeting the conserved plasmepsin gene (Daniels *et al* 2016, manuscript submitted). *Plasmodium falciparum, P. vivax, P. malariae,* and *P. ovale* are known to be distributed throughout Madagascar, although *P. falciparum* is estimated to account for 96% of cases [1]. Assessing the agreement between RDT results and molecular detection of *Plasmodium* species is warranted, as the specificity of RDTs is known to be variable, especially if multiple *Plasmodium* species are present [23-26].

4.3.5 Genetic sampling: SNP-genotyping analysis

Calling alleles

For the confirmed *P. falciparum* infections, alleles from a panel of 24 putatively neutral and unlinked genomic loci were typed to produce multi-locus genotypes (MLGs) [22]. Twelve or more loci were typed for 94 samples (see results). These loci and the SNP-genotyping protocol used have been termed a *P. falciparum* "molecular barcode" and are described elsewhere

[17,22,27]. The loci are distributed on 12 of the 14 *P. falciparum* chromosomes and are identical to those of Daniels *et al* [22]. Briefly, the presence of a variant allele at the targeted locus causes a difference in the melting temperature of an RT-PCR probe that can be detected using high-resolution melting (HRM) analysis. If, for a sample, two alleles are present at the probed locus, two melting peaks, corresponding to the two alleles present, are observed. As *Plasmodium* parasites are haploid in the bloodstream stage of their life cycle, observing two alleles for a locus indicates that multiple distinct parasite genomes were present in the host at the time of sampling. Such infections are reported as polygenomic infections. To be conservative, and consistent with previous conventions, an infection is reported as polygenomic if at least two of the loci analyzed for that sample exhibit the two melting peaks indicative of two alleles present [17,22].

As the concentration of parasite DNA in blood samples can be quite low relative to other co-infecting parasites, or relative to host DNA, a pre-amplification step was performed prior to SNP-genotyping. Pre-amplification was previously shown to increase the success rate and sensitivity of SNP-genotyping (method same as used in ref. [33]).

Genetic correlates of transmission

Two genetic parameters hypothesized to reflect transmission dynamics and empirically shown to co-vary with transmission intensity are (i) the percentage of infections that are polygenomic and (ii) the frequency distribution of MLGs [17][18]. When transmission is low or unstable, fewer distinct MLGs are present locally and more individuals contain a MLG identical to one that has already been sampled. This leads to some MLGs being observed at elevated frequencies within the sample. In contrast, at higher and more stable transmission levels, few infections contain repeatedly observed MLGs.

Estimating minor allele frequencies and resolution

In order to evaluate the resolution of our genetic sampling, we first estimated the minor allele frequency (MAF) at each of the 24 loci. To account for the fact that polygenomic samples were present and that the multiplicity of infection within polygenomic samples was unknown, a form of the "predominant allele" approach was used [29]. For polygenomic samples, those loci where multiple alleles were present were not considered, but the loci where only one allele was observed (indicating that the MLGs forming the polygenomic infection had the same allele at the locus) were considered. Tallying predominant alleles has the advantage of allowing the alleles in polygenomic infection samples (which can be a large proportion of the sample) to be taken into consideration when determining MAF. This approach has also been empirically shown to correlate closely with other measures of allele frequency [29,30], but does include an assumption that allele frequencies at the observed "monogenomic" loci are not strongly biased.

Assuming alleles at a locus are neutral and segregate independently, the probability that two unrelated parasite lineages would be observed to have the same allele by chance is given by $MAF^2 + (1\text{-}MAF)^2$. For multiple loci, the probability of observing an identical MLG by chance (identity by state, IBS) is simply the product of IBS probabilities for each locus. As the average MAF for the loci analysed approaches its maximum of 0.5 (indicating 50% of the samples have the alternate allele), the number of loci required for a significant level of resolution decreases. Using a significance cut-off of $\alpha = 0.01$ and the derived estimate of the MAF for each locus, the number of loci needed to have a significantly small probability (i.e., < 0.01) of randomly observing identical MLGs was determined if: (i) the loci with the highest observed MAFs were used, or (ii) the loci with the lowest observed MAFs were used. For (i) and (ii), these represent

the fewest and the most loci needed, respectively, to achieve the chosen level of confidence given the observed MAFs.

Estimating multi-locus genotype frequencies and genetic similarity

To identify any samples with identical MLGs and to estimate the genetic similarity among samples, pairwise percent identity was calculated (using a Python script, available upon request) for all samples and for subsets of samples of interest. For polygenomic infections, we calculated genetic similarity among the *P. falciparum* MLGs within the infection as the proportion of typed loci for that infection that shared the same allele. To account for differences in the number of pairwise comparisons used to calculate mean genetic similarity for different subsets of samples, bootstrap resampling was used to test for significance with subsamples equal in size to the number of comparisons made in the smallest subset (in R, script available upon request). *P*-values are reported as the proportion of subsamples of a given size that had a mean within the standard error of the mean of the other subset of samples.

4.4 Results

4.4.1 High prevalence of P. falciparum infection by RDT

For both communities, prevalence by RDT varied temporally, from a low of 5.7-10.1% in July-August to a high of 17-18% in March-April for each community (see Figure 2). Overall, 26.3% and 29.9% of individuals for community 1 and community 2, respectively, tested positive for *P. falciparum* by RDT during the course of the 10-month follow-up. For both communities combined, there were 504 individuals (72.2%) with a negative *P. falciparum* RDT result at all three time points. Among individuals with a positive RDT result, 169 (24.2%), 20 (2.9%), 4

(0.6%), and 1 (0.1%) individuals were positive at one time point, two consecutive time points, two nonconsecutive time points, and all three time points, respectively. Among children under five, 34 of 134 children sampled (25.4%) were RDT positive. At the household level, 63.0% of households had at least one member of the household test positive for *P. falciparum* by RDT.

For the randomly chosen subset (n = 166) of the samples positive by RDT for *P. falciparum* that were analysed the presence of *P. falciparum* was identified by RT-PCR in 88% of the RDT positive samples. This supports general agreement between RDT and RT-PCR results and is comparable to other analyses of RT-PCR and RDT concordance in Madagascar (e.g. 79.5% in the national 2013 Malaria Indicator Survey [15]). Possible explanations for the 12% of RDT positive samples that were not positive for *P. falciparum* by RT-PCR include: (i) false positive RDTs, (ii) a failure to preserve DNA between the time of sampling and the later RT-PCR analysis for some samples, (iii) failed RT-PCR reactions and (iv) infections due to other *Plasmodium* species. Apart from *P. falciparum*, two samples were positive for *P. vivax* by RT-PCR, one from each community at the third time point. The *P. vivax* positive sample from community 1 was also positive for *P. falciparum* and thus a mixed-species infection. Furthermore, two samples from community 1 were positive for *P. malariae*. The *P. malariae* sample from time point 1 was found to also harbour a polygenomic infection of *P. falciparum*. This sample was therefore both a mixed-species infection as well as a complex *P. falciparum* infection (see more discussion of polygenomic *P. falciparum* infections below).

Fig. 2

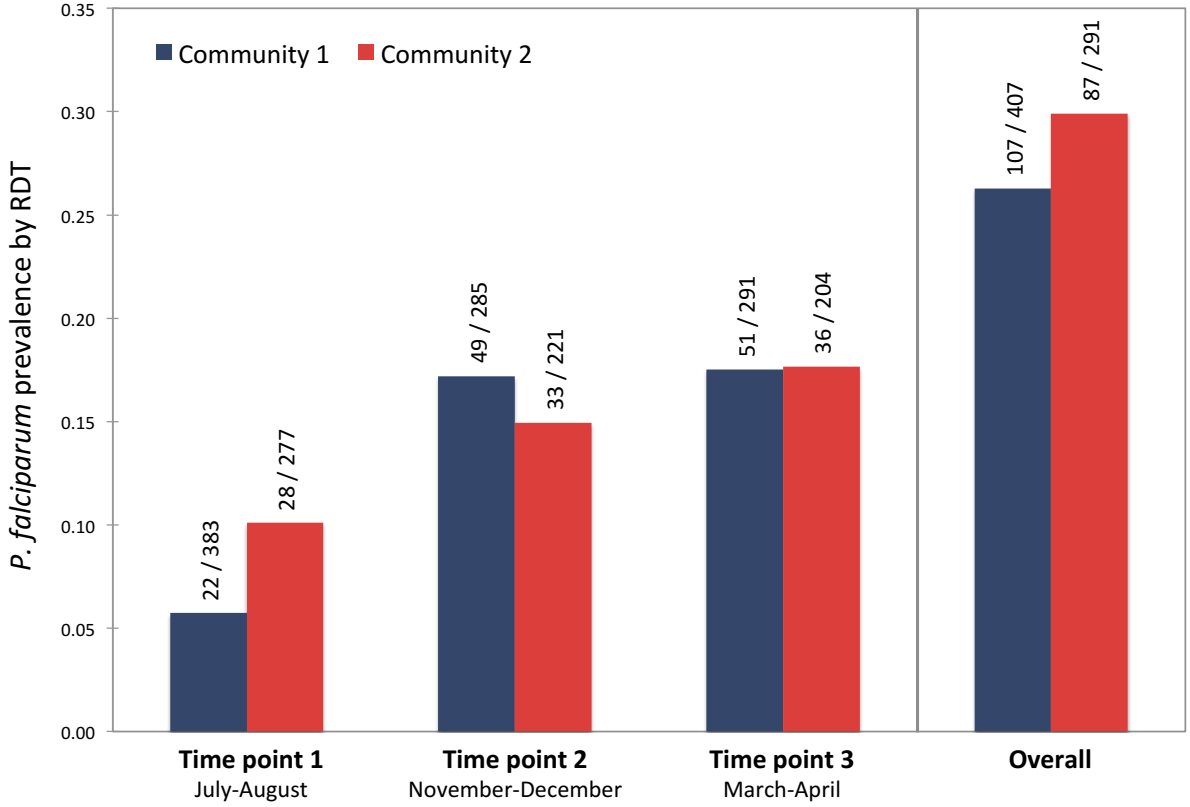

Figure 2

High relative prevalence and temporal variation of *P. falciparum* in the Makira region of Madagascar

Prevalence of *P. falciparum* by RDT is shown with community 1 in blue and community 2 in red. The number of positive samples out of the number of individuals surveyed is shown above the bars. Shown to the right is the overall proportion of individuals that tested positive at one or more time point during the course of the study. The overall percentage of individuals positive for both communities combined was 27.8%.

4.4.2 Frequency of polygenomic infections

In addition to observing high prevalence, we also observed that a high frequency of infections were polygenomic (68.1%) among the 94 *P. falciparum* infections that were typed at 12 or more loci (see Table 1). This observation was replicated across communities, across time points, across sexes, and across subjects' age groups (range 59.6-79.5%). The percentage of infections found to be polygenomic was also similar among the samples typed at 12-17 loci (68.5%) and the samples that were typed at 18 or more loci (66.7%). The number of SNP positions that were polymorphic within infections (i.e. polygenomic) ranged from 0 to 11. The mean and median number of polymorphic loci among polygenomic samples was 3.95 and 3, respectively.

Table 1

Frequency of polygenomic *P. falciparum* infections in the Makira region of Madagascar

Category	**Group**	***n*[a]**	**% Polygenomic**
Overall	All samples	94	68.1
Community	Community 1	60	61.7
	Community 2	34	79.4
Time point	Time point 1	29	65.5
	Time point 2	43	72.1
	Time point 3	22	63.6
Sex[b]	M	41	78.0
	F	52	59.6
Age[b]	2-12 years	27	70.4
	13+ years	66	66.7
Genetic sampling	12-17 loci typed	73	68.5
	18+ loci typed	21	66.7

[a]*n* is the number of infections genotyped

[b]Age and sex data was not recorded for one individual

Although the frequency of polygenomic infections was higher in community 2 (79.4%) than in community 1 (61.7%), this difference was not statistically significant ($p = 0.612$, Fisher exact test for count data). Neither was the difference between polygenomic infection frequency among males (78.0%) and females (59.6%) ($p = 0.075$) or that between any pair of time points ($p > 0.05$). The failure to observe a significant difference between time points is notable, as rainfall differs 4-fold between time points 2 and 3 and prevalence was observed to change approximately 3-fold between time points 1 and 3. These data indicate that some factor other than precipitation drives temporal changes in prevalence and that the frequency of polygenomic infections may be seasonally stable.

4.4.3 SNP allele frequency and validation

In order to permit downstream genetic analysis, it was first validated that the loci assayed were polymorphic. Then the number of loci needed for a given level of resolution in distinguishing parasite MLGs was estimated. All 24 loci analysed were found to be polymorphic, with a high MAF on average (mean MAF 0.310, range 0.083 to 0.500). The mean number of loci typed per sample was 15.6, with a range from 12 to 22 loci (see below for the use of 12 loci as the minimum cut-off). For no sample were all 24 loci typed, but MAF was calculated from a mean of 61.3 samples per locus for all 24 loci (range 15 to 87). Loci A6 and A10, with alleles being callable from only 15 and 16 samples each, respectively, had significantly lower reporting rates than the other loci. Inefficient binding of primers or probes for these loci, possibly due to a

variant sequence circulating in the population, may explain the lower number of samples successfully typed for these loci.

From the allele frequency data, and assuming independence, we determined the mean number of loci needed for the probability to be < 0.01 for observing an identical MLG for unrelated parasite genomes (see Figure 3). If two parasite genomes were compared at only the loci with the highest observed MAF (and thus the highest resolution), then only seven loci would be needed. Conversely, comparing the parasite genomes at the loci with the lowest MAFs would require typing at five additional loci (12 total) to have the same level of resolution. To be conservative, our analyses subsequently focused on the samples for which 12 or more loci were genotyped (n = 94).

Fig. 3

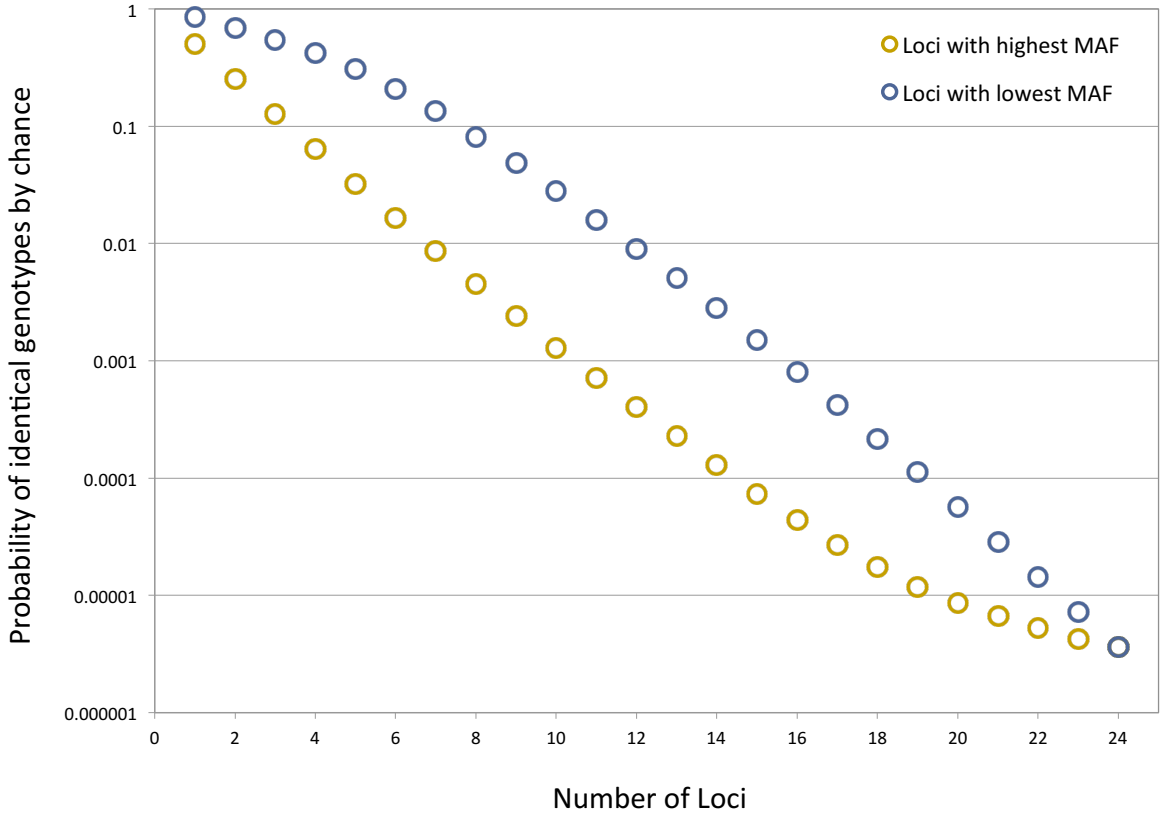

Figure 3

The probability of observing identical multi-locus genotypes by random chance as a function of MAF and the number of loci considered

The probability that two unrelated parasites would have an identical multi-locus genotype by chance (often referred to as identity by state) is shown on the *y*-axis on a logarithmic scale. The probability was calculated using loci with the highest minor allele frequencies (MAFs) (shown in yellow), loci with the lowest MAFs (blue). See Methods for details on calculation (independence and neutrality assumed).

4.4.4 Frequency distribution of multi-locus genotypes: Limited evidence of repeat multi-locus genotypes observed between infections

After validating that the loci analyzed could provide a sufficiently high probability of distinguishing unrelated parasite lineages, the frequency of occurrence for the MLGs in the sample set was characterized. The observed MLGs of all samples were compared in all possible pairs, and the samples with the fewest pairwise genetic differences were analyzed further. Notably, there were no instances in which the same MLG was found to be present in more than one individual in our sample. For one individual (individual 1.68.02, see Supplementary File 1), an identical MLG was observed at consecutive time points. While treatment was offered and provided after a positive RDT result, it is possible that the treatment failed to clear parasites from the bloodstream or was not adhered to by the individual. As a result, we find it more likely that this represents the persistence of a single infection between time points and not a separate

infection of this individual with an identical MLG. MLGs were generated at multiple time points for seven other individuals and in each case differed at multiple loci.

Among polygenomic infections, in comparisons between three samples (2.33.04.2, 2.33.08.2, and 1.92.01.2, see Supplementary File 1), one or more samples were polygenomic but did not exhibit any differences among the loci for which both samples had a single allele. For these cases, it is possible that the MLGs contributing to the polygenomic infection have the same allelic configuration as one present in the other sample. Regardless, such a situation arose in only a small number of comparisons. Due to the lack of definitive instances where independent infections contained identical MLGs, these data strongly indicate that the pattern of repeated clusters of identical MLGs expected from low or sporadic transmission [31] is not present in the Makira region.

4.4.5 Genetic similarity among polygenomic infections

In order to investigate levels of genetic similarity among the parasite lineages within polygenomic infections, the proportion of loci that differed within polygenomic infections was compared to the average number of differences between randomly selected monogenomic infections. As seen in Figure 4, a small number of polygenomic infections had significantly fewer shared loci than the average pairwise identity for all monogenomic samples. A possible explanation is that these exceptionally variable polygenomic infections contain more than two parasite MLGs, which would reduce the number of alleles shared by all parasite genomes present in the infection. Despite these few outliers, the distribution of genetic similarity among polygenomic infections was significantly skewed towards higher levels of genetic similarity in comparison to that among monogenomic samples ($p < 0.0001$, bootstrap subsampling to correct

for differences in the number of comparisons). This result suggests that parasite genomes in polygenomic infections are genetically more closely related than would be expected, and it is consistent with the co-transmission of multiple, related parasites by the bite of a single infectious vector. It is also noteworthy that no significant difference was observed in the average pairwise similarity between all monogenomic samples from community 1 or community 2. Hence this metric suggests an absence of parasite population structure at the community level.

Fig. 4

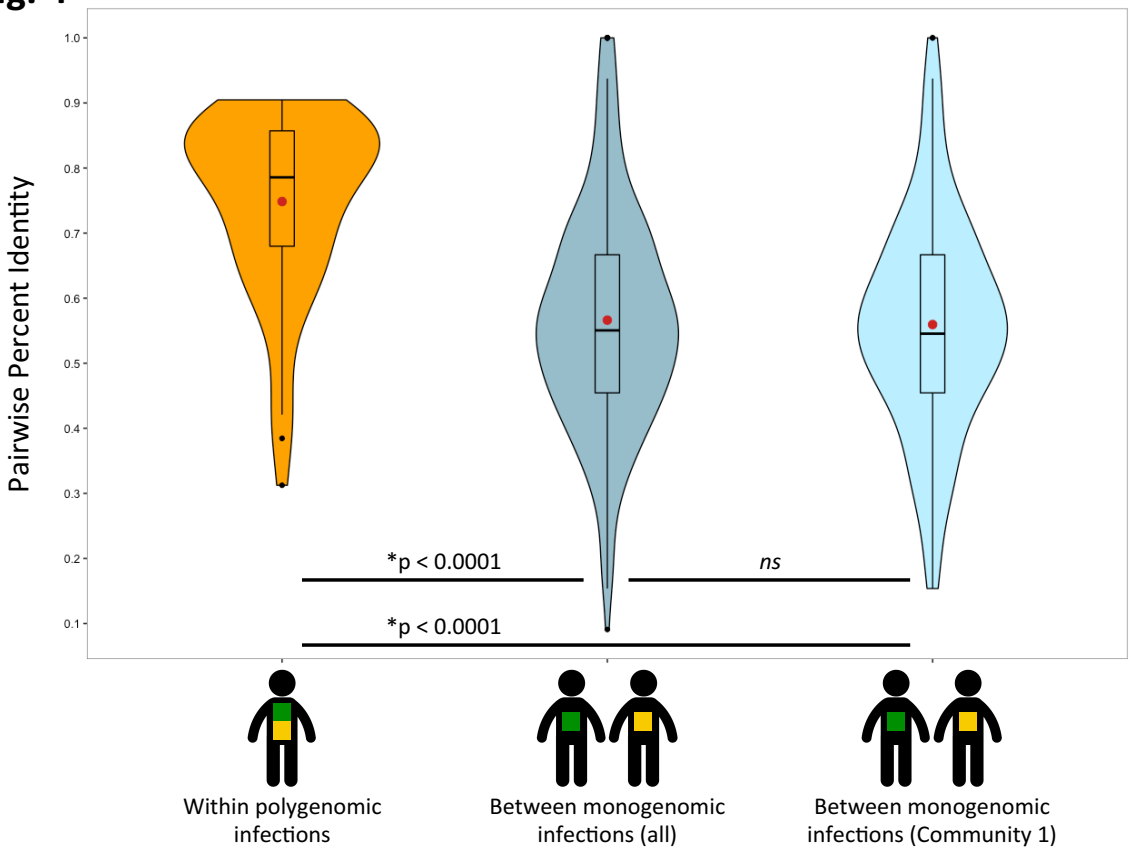

Figure 4

The distribution of genetic similarity among polygenomic and monogenomic infections

Violin plots of the distribution of pairwise percent identity among polygenomic and monogenomic infections. Means are marked with a red point. Boxplots are shown within the violin plots with the median, interquartile range (IQR), and data points more than 1.5 × IQR from the upper or lower quartiles (outliers) marked. *P*-values determined by bootstrap sub-sampling (see Methods). *ns*, not significant.

4.5 Discussion

4.5.1 Plasmodium falciparum is prevalent and genetically diverse in the Makira region

Prevalence of *P. falciparum* was 5.7-17.6% at each time point for both sites studied, with 27.8% of individuals and more than 60% of households having at least one positive result over the course of the study. Notably, prevalence in both communities, even at its lowest point (5.7%), was higher than the highest prevalence captured by a recent analysis of national sentinel health sites located in eastern Madagascar (4.9%)[7]. Additionally, the observed prevalence among children under 5 years old seen here (25.4%) was much higher than that found in the national 2013 Malaria Indicator Survey study of 1725 children in eastern Madagascar in the same age group (11.7%) [15]. These observations of prevalence and incidence indicate that transmission frequency, and hence the burden of *P. falciparum*, is much higher at the two sites than would be expected from previous estimates based on regional and national surveillance in Madagascar [1,2,7,32].

Apart from the two study sites, approximately 127,000 people live in the Makira region, all in rural settings with comparable ecological conditions. Additionally, an estimated 86.8% of the 1 million people living in the broader Analanjirofo administrative region (containing the

Makira region) also live within similar rural districts [15]. Further, the Analanjirofo region is just one of four such large regions in northeastern Madagascar, which together have a population of over 4.4 million [20]. Accordingly, if the prevalence pattern we observed applies widely across communities/fokontany in the area and the northeastern regions, then this part of Madagascar would be expected to be a significant contributor to the national malaria burden (estimated at a total of 750,000 to 2,100,000 cases annually [1]). The policy implications are significant; it is possible that government surveillance has focused on coastal and peri-urban sites that are easier to access, or has relied on clinic-based statistics, which may be misleading for remote regions.

An alternative possibility is that the prevalence and incidence seen at the two Makira region sites studied here merely reflects a transient spike. This possibility seems unlikely, because if transmission were sporadic and we simply happened to observe a temporary, unstable peak in transmission in this study, then a pattern of low diversity, frequently repeated MLGs, and fewer polygenomic infections would be expected in the genetic data. In settings of low, unstable transmission, sometimes termed clonal transmission, few MLGs predominate and the reshuffling of those MLGs by recombination or re-assortment is rare. This pattern of clonal transmission has been observed in locations at the periphery of *P. falciparum* transmission such as Panama [32] and in areas of Africa and Asia after interventions have successfully reduced transmission [17,18].

The pattern observed here is the opposite, as we see very limited evidence of repeated MLGs and a high frequency of polygenomic infections for all subgroups of the sample set (see Table 1). The frequency of polygenomic infections observed in the Makira region was similar, or even higher, than that seen in other high transmission, pre-intervention settings (some examples from studies using SNP-genotyping shown in Table 2).

Table 2

Examples of the frequency of polygenomic infections inferred using SNP-genotyping compared across geographic regions and transmission settings

Country	Year	Polygenomic (%)	Transmission setting	Reference
Madagascar	2013-14	68.1	Present study	Present study
Senegal	2006	41	Pre-intervention	Daniels *et al* (2015)
	2011	26	Post-intervention	
Malawi	2006	76	Pre-intervention	Sisya *et al* (2015)
	2012	68	Post-intervention[a]	
Thailand	2001	63	Pre-intervention[b]	Nkhoma *et al* (2013)
	2012	14	Post-intervention[b]	

[a]The intervention in Malawi was deemed to be ineffective at reducing transmission

These indicate that the localities in the Makira region studied here are currently unrecognized foci of high transmission and signal the possibility that other such localities exist in Madagascar. In addition to representing a previously unrecognized source of malaria burden, these localities might serve as sources of reintroduction to areas where interruption of transmission has seen progress. For instance, since the early 1990s the malaria burden has fallen substantially in much of the interior, central plateau of Madagascar [15,34-37]. Unmonitored hotspots could provide refuges for parasite lineages that could later spillover and compromise these gains. Moreover, recent reports suggest that the burden of malaria has indeed risen broadly since 2011 [15]. This highlights the need for further research and greater surveillance of northeastern Madagascar, and other remote regions currently lacking comprehensive surveillance. Such efforts could determine the extent to which neighboring areas and larger regions also contain foci of elevated prevalence and transmission.

4.5.2 Potential confounding effects of co-transmission and spatial heterogeneity

Two potential confounding effects that may limit our ability to infer relative levels of transmission from the frequency of polygenomic infections are co-transmission and small-scale spatial heterogeneity in transmission intensity. Co-transmission, where multiple parasite MLGs are taken up by the same mosquito vector and transmitted to the next host, is a mechanism by which polygenomic infections can result from a single transmission event. If co-transmission is common, then the probability of observing polygenomic infections is less dependent on the number of transmission events. Likewise, high spatial heterogeneity in transmission could also cause an increase in the frequency of polygenomic infections without an increase in the number of transmission events by concentrating transmission events in a subset of the population. Hosts within the subset would have a higher probability of being reinfected during the course of an infection and thus more polygenomic infections would be produced.

If these two confounders were present in this study, then it would be expected to see (i) a higher average level of genetic similarity among polygenomic infections than between monogenomic infections and (ii) evidence that *P. falciparum* infection was not randomly distributed among the sample population. For co-transmission, serial co-passage of two parasite lineages would provide those lineages with multiple opportunities to recombine with each other. It would be expected that repeated recombination between the parasite genomes would result in progeny with MLGs that were more similar to each other than for parasites that had not recombined as often. Accordingly, if co-transmission were prevalent, it would be expected that the average similarity among the MLGs comprising polygenomic infections would be higher than the average similarity between MLGs chosen randomly from the population. Conversely, if

polygenomic infections arose from superinfection (the re-infection of an infected host), then it would be expected that the average similarity among polygenomic infections would be similar to the average between MLGs randomly chosen from the population. Genetic similarity among polygenomic infections was significantly higher than the average between monogenomic infections, consistent with the presence of co-transmission.

As evidence against unusually extreme spatial heterogeneity in our sample, a high percentage (63.0%) of households contained an individual infected at one time point or more. Additionally, there was no evidence that genetic similarity was partitioned between communities. It seems more likely that co-transmission and/or spatial heterogeneity are widespread phenomena in *P. falciparum* populations and therefore their effect is accounted for when comparing populations. Further study of co-transmission and spatial heterogeneity, especially in Madagascar, would serve to better characterize this relationship.

4.5.3 Conclusion

Despite the abundance of estimates for low to moderate levels of malaria transmission on average regionally and countrywide, we show evidence that the Makira region in northeastern Madagascar is a hotspot of malaria transmission. First, RDT surveys, with RT-PCR confirmation generally validating the results, were performed and an overall level of infection approximately 7-fold higher than the national average and 2 to 5-fold higher than existing regional estimates was observed. Next, a panel of SNP markers was validated as variable and informative in Madagascar, and these SNPs were used to estimate the allele and MLG frequencies as well as the frequency of polygenomic infections. A pattern consistent with high, stable transmission was observed. Because few previous studies of *P. falciparum* genetic diversity in Madagascar have

been performed, the identification of variable markers, the estimates of minor allele frequencies, and the calculation of the number of loci needed for a given level of resolution will be useful in future studies. Such studies are needed to better define the relative contributions of co-transmission and small-scale spatial heterogeneity to patterns of parasite genetic diversity. Remaining hotspots of *P. falciparum* transmission pose a threat to the recent gains in malaria control in Madagascar, and across Africa, and a better understanding of the genetic correlates of transmission can provide a better method for their identification and monitoring.

List of abbreviations

Multi-locus genotype (MLG); *Faritra* – the largest administrative regions of Madagascar; Rapid diagnostic test (RDT); Single nucleotide polymorphism (SNP); Real-time polymerase chain reaction (RT-PCR); High resolution melting (HRM); Minor allele frequency (MAF); Identity by state (IBS); Interquartile range (IQR)

Ethics approval and consent to participate

Ethical approval was given by the Madagascar Ministry of Health, the Maroantsetra regional medical inspector, the mayor covering the region capturing both communities, the chief in each community, and the Office for the Protection of Human Subjects at the Harvard T.H. Chan School of Public Health (IRB#22826).

Author contributions

BLR performed experiments and drafted the manuscript. CDG and SKV designed the field research; EJGA and CDG performed the field research. BLR, CDG, SKV and DLH conceived

and designed the analytical methods. CMB performed some of the speciation assays and HRM analysis. CDG and DLH reviewed and edited the manuscript. SKV contributed materials and analysis tools. All authors read, contributed to, and approved the final manuscript.

Acknowledgements

We thank Be Jean Rodolph Rasolofoniaina, Miadana Arisoa Vonona, Rijaniaina Ambinintsoa Ralaiarison, Abdy Roland Berobia, Laurent, Ravaoliny, Berthelemie, and many other MAHERY (Madagascar Health and Environmental Research) collaborators for their research assistance. We also thank the research subjects whose dedication and participation made this work possible. For experimental advice and technical assistance we thank Soa Andrian, Rachel Daniels, Katelyn Durfee, Elizabeth Hamilton and Kalsang Namgyal. We thank Wesley Wong for help with a script to calculate pairwise genetic similarity and for critical discussion of analysis. We thank Danny Milner Jr for critical discussion of study design and analysis and members of the Planetary Health Alliance for general feedback.

References

1. WHO. World Malaria Report 2015: Madagascar. Geneva: World Health Organization; 2015. http://www.who.int/malaria/publications/country-profiles/profile_mdg_en.pdf?ua=1. Accessed 5 May 2016.
2. Bhatt S, Weiss DJ, Cameron E, Bisanzio D, Mappin B, Dalrymple U, et al. The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. Nature. 2015;526:207–11.
3. Robert V, Le Goff G, Andrianaivolambo L, Randimby FM, Domarle O, Randrianarivelojosia M, et al. Moderate transmission but high prevalence of malaria in Madagascar. Int J Parasitol. 2006;36:1273–81.
4. Marrama L, Jambou R, Rakotoarivony I, Leong Pock Tsi JM, Duchemin JB, Laventure S, et al. Malaria transmission in Southern Madagascar: influence of the environment and hydro-agricultural works in sub-arid and humid regions. Part 1. Entomological investigations. Acta Trop. 2004;89:193–203.
5. Maïga-Ascofaré O, Rakotozandrindrainy R, Girmann M, Hahn A, Randriamampionona N, Poppert S, et al. Molecular epidemiology and seroprevalence in asymptomatic *Plasmodium falciparum* infections of Malagasy pregnant women in the highlands. Malar J. 2015;14:188.
6. Rabarijaona LP, Randrianarivelojosia M, Raharimalala LA, Ratsimbasoa A, Randriamanantena A, Randrianasolo L, et al. Longitudinal survey of malaria morbidity over 10 years in Saharevo (Madagascar): further lessons for strengthening malaria control. Malar J. 2009;8:190.
7. Kesteman T, Randrianarivelojosia M, Mattern C, Raboanary E, Pourette D, Girond F, et al. Nationwide evaluation of malaria infections, morbidity, mortality, and coverage of malaria control interventions in Madagascar. Malar J. 2014;13:465.

8. Kesteman T, Rafalimanantsoa SA, Razafimandimby H, Rasamimanana HH, Raharimanga V, Ramarosandratana B, et al. Multiple causes of an unexpected malaria outbreak in a high-transmission area in Madagascar. Malar J. 2016;15:57.
9. WHO. Global Malaria Control and Elimination: Report of a technical review. Geneva: World Health Organization; 2008. http://www.who.int/malaria/publications/atoz/9789241596756/en/ Accessed 5 May 2016.
10. Ruktannonchai NW, DeLeenheer P, Tamem AJ, Alegana VA, Caughlin TT, Zu Erbach-Schoenberg E, et al. Identifying malaria transmission foci for elimination using human mobility data. PLoS Comput Biol. 2016;12:e1004846.
11. Bousema T, Drakeley C, Gesase S, Hashim R, Magesa S, Mosha F et al. Identification of hotspots of malaria transmission for targeted malaria control. J Infect Dis. 2010;201:1764-74.
12. Ministère de la Santé Publique, Programme National de Lutte contre le Paludisme. Revue du programme paludisme à Madagascar. Antananarivo; 2011.
13. Kesteman T, Randrianarivelojosia M, Raharimanga V, Randrianasolo L, Piola P, Rogier C. Effectiveness of malaria control interventions in Madagascar: a nationwide case-control survey. Malar J. 2016;15:83.
14. Battle KE, Bisanzio D, Gibson HS, Bhatt S, Cameron E, Weiss DJ, et al. Treatment-seeking rates in malaria endemic countries. Malar J. 2016;15:20.
15. MIS 2013: Institut National de la Statistique (INSTAT), Programme National de lutte contre le Paludisme (PNLP), Institut Pasteur de Madagascar (IPM) et ICF International. 2013. *Enquête sur les Indicateurs du Paludisme (EIPM) 2013*. Calverton, MD, USA : INSTAT, PNLP, IPM et ICF International.

16. Volkman SK, Neafsey DE, Schaffner SF, Park DJ, Wirth DF. Harnessing genomics and genome biology to understand malaria biology. Nat Rev Genet. 2012;13:315–28.
17. Daniels R, Chang H-H, Séne PD, Park DC, Neafsey DE, Schaffner SF, et al. Genetic surveillance detects both clonal and epidemic transmission of malaria following enhanced intervention in Senegal. PLoS ONE. 2013;8:e60780.
18. Nkhoma SC, Nair S, Al-Saai S, Ashley E, McGready R, Phyo AP, Nosten F, et al. Population genetic correlates of declining transmission in a human pathogen. Mol Ecol. 2013; 22:273–285.
19. Wildlife Conservation Society (WCS). Makira Forest Protected Area Project: 2005-2013 Project Implementation Report. 31 August, 2015 v3.0.
20. Institut National de la Statistique (INSTAT). Madagascar - Enquête Démographique et de Santé 2008-2009. USAID.
21. Golden CD, Fernald LC, Brashares JS, Rasolofoniaina BJ, Kremen C. Benefits of wildlife consumption to child nutrition in a biodiversity hotspot. Proc Natl Acad Sci USA. 2011;108:19653-6.
22. Daniels R, Volkman SK, Milner DA, Mahesh N, Neafsey DE, Park DJ, et al. A general SNP-based molecular barcode for *Plasmodium falciparum* identification and tracking. Malar J. 2008;7:223.
23. Mariette N, Barnadas C, Bouchier C, Tichit M, Ménard D. Country-wide assessment of the genetic polymorphism in *Plasmodium falciparum* and *Plasmodium vivax* antigens detected with rapid diagnostic tests for malaria. Malar J. 2008;7:219.
24. Ratsimbasoa A, Ravony H, Vonimpaisomihanta J-A, Raherinjafy R, Jahevitra M, Rapelanoro R, et al. Management of uncomplicated malaria in febrile under five-year-old children by

community health workers in Madagascar: reliability of malaria rapid diagnostic tests. Malar J. 2012;11:85.

25. Ratsimbasoa A, Randriamanantena A, Raherinjafy R, Rasoarilalao N, Ménard D. Which malaria rapid test for Madagascar? Field and laboratory evaluation of three tests and expert microscopy of samples from suspected malaria patients in Madagascar. Am J Trop Med Hyg. 2007;76:481–5.
26. Ehtesham R, Fazaeli A, Raeisi A, Keshavarz H, Heidari A. Detection of mixed-species infections of *Plasmodium falciparum* and *Plasmodium vivax* by nested PCR and rapid diagnostic tests in southeastern Iran. Am J Trop Med Hyg. 2015;93:181–5.
27. Daniels R, Hamilton EJ, Durfee K, Ndiaye D, Wirth DF, Hartl DL, et al. Methods to Increase the sensitivity of high resolution melting single nucleotide polymorphism genotyping in malaria. J Vis Exp 2015:e52839.
28. Mharakurwa S, Daniels R, Scott A, Wirth DF, Thuma P, Volkman SK. Pre-amplification methods for tracking low-grade *Plasmodium falciparum* populations during scaled-up interventions in Southern Zambia. Malar J 2014;13:89.
29. Anderson TJC, Nair S, Sudimack D, Williams JT, Mayxay M, Newton PN, et al. Geographical distribution of selected and putatively neutral SNPs in Southeast Asian malaria parasites. Mol Biol Evol. 2005;22:2362–74.
30. Maïga-Ascofaré O, Le Bras J, Mazmouz R, Renard E, Falcão S, Broussier E, et al. Adaptive differentiation of *Plasmodium falciparum* populations inferred from single-nucleotide polymorphisms (SNPs) conferring drug resistance and from neutral SNPs. J Infect Dis. 2010;202:1095–103.

31. Daniels RF, Schaffner SF, Wenger EA, Proctor JL, Chang HH, Wong W et al. Modeling malaria genomics reveals transmission decline and rebound in Senegal. Proc Natl Acad Sci USA. 2015;112:7067-72

32. Obaldia N, Baro NK, Calzada JE, Santamaria AM, Daniels R, Wong W, Chang H-H, et al. Clonal outbreak of *Plasmodium falciparum* infection in eastern Panama. J Infect Dis. 2015;211:1087–1096.

33. Sisya TJ, Kamn'gona RM, Vareta JA, Fulakeza JM, Mukaka MFJ, Seydel KB, et al. Subtle changes in *Plasmodium falciparum* infection complexity following enhanced intervention in Malawi. Acta Trop. 2015;142:108–14.

34. Ralisoa O, Rasamindrakotroka A, Ravelojana B, Rafarasoa L, Randrianarisoa E, Rakotoson R, et al. *Plasmodium falciparum* malaria transmission indices in a highland village of the Ikopa River Valley near Antananarivo, Madagascar. Parassitologia. 1991;33:219–23.

35. Romi R, Razaiarimanga MC, Raharimanga R, Rakotondraibe EM, Ranaivo LH, Pietra V, et al;. Impact of the malaria control campaign (1993-1998) in the highlands of Madagascar: parasitological and entomological data. Am J Trop Med Hyg. 2002;66:2–6.

36. Ratovonjato J, Randrianarivelojosia M, Rakotondrainibe ME, Raharimanga V, Andrianaivolambo L, Le Goff G, et al. Entomological and parasitological impacts of indoor residual spraying with DDT, alphacypermethrin and deltamethrin in the western foothill area of Madagascar. Malar J. 2014;13:21.

37. Roll Back Malaria: Focus on Madagascar. Progress and Impact Series-Country Reports. Geneva; 2013.

38. Mouchet J, Blanchy S. Particularities and stratification of malaria in Madagascar. Sante 1995;5(6):386-8.

Chapter 5

Cross-regional comparisons of *Plasmodium falciparum* genetic diversity in Madagascar by SNP genotyping

Attributions:

The contents of this chapter are a manuscript in preparation to be submitted to the journal *BMC Public Health*: **Cross-regional comparisons of *Plasmodium falciparum* genetic diversity in Madagascar by SNP genotyping** by Benjamin L. Rice, Christopher D. Golden, Hervet J. Randriamady, and Daniel L. Hartl.

5.1 Abstract

Background

Different regions of Madagascar have distinct ecologies with differing precipitation patterns and mosquito vector communities. However, the extent to which this regional ecological variation is associated with different malaria transmission patterns is largely unknown.

Methods

To better understand local malaria transmission patterns across ecological settings in Madagascar, we assayed a panel of single nucleotide polymorphisms (SNPs), termed a molecular barcode, for a newly available sample of *Plasmodium falciparum* infections. Samples (n = 112) from rural communities in two distinct eco-regions in the Atsimo Andrefana administrative region were analyzed: (i) Toliara II district and (ii) Morombe district. We characterized within host parasite genetic diversity (the frequency of hosts containing multiple differing *P. falciparum* genomes at the time of sampling) and the genotypic diversity (the frequency of unique versus repeated genotypes) for these study populations.

Results

From the subset of samples initially analyzed for Toliara II and Morombe districts, we observed a high diversity of multi-locus genotypes and a high frequency of polygenomic infections. Despite differing in precipitation pattern, other ecological parameters, and the level of site-site variation in malaria prevalence, we failed to observe a significant difference in the frequency of repeated genotypes in either region. Polygenomic infection frequency among the samples analyzed for Toliara II district (78.3%) was higher than in Morombe (33.3%) or Maroantsetra districts (68.1%, reported in Chapter 4).

Discussion

The observation of high within host genetic diversity and high genotypic diversity in the west coast and southwest study regions suggests recombination among circulating *P. falciparum* genomes is frequent and clonal stability over time is low. This pattern is inconsistent with temporary, sporadic spikes in incidence being responsible for the observations of high parasite prevalence in some rural communities in these areas. Further study of these districts and their greater prioritization in future malaria control efforts is warranted.

5.2 Introduction

5.2.1 Spatial variation of malaria prevalence in Madagascar

Within Madagascar, there is extreme variation in ecological parameters, such as predominant vegetation and rainfall, and human population parameters, such as access to infrastructure and agricultural systems (Goodman 2004, Howes 2016). This is expected to contribute to differences in malaria transmission patterns in different eco-regions of the country, but data limitations have prevented in-depth study of variation in trends in malaria transmission at smaller spatial scales.

Previous studies (including chapter 3 of this dissertation, MIS 2016) have largely been cross-sectional studies. These studies have identified regions with high prevalence and regions with high variation in prevalence between nearby communities. At a coarse level, some regions have been observed to have higher average prevalence, as, for example, Rice et al 2019 (Chapter 3 of this dissertation) found mean prevalence among communities studied in the west coast region of Madagascar to be above 30%, while it was less in the southwest (5.3%). Within regions, prevalence between sites within 15 km often varied greatly.

However, such data lacks the longitudinal sampling needed to observe temporal trends. When comparing areas, variation in the prevalence observed could be explained by inter- or intra-annual fluctuations in malaria transmission or it could be explained by a spatially heterogenous, but stable, distribution of transmission intensity (see Figure 1). Discriminating between these two transmission patterns, where spatial variation in malaria infection is predominantly driven by either: (i) transient changes in transmission intensity, or (ii) stable differences in transmission intensity between areas, can provide information useful to malaria control efforts.

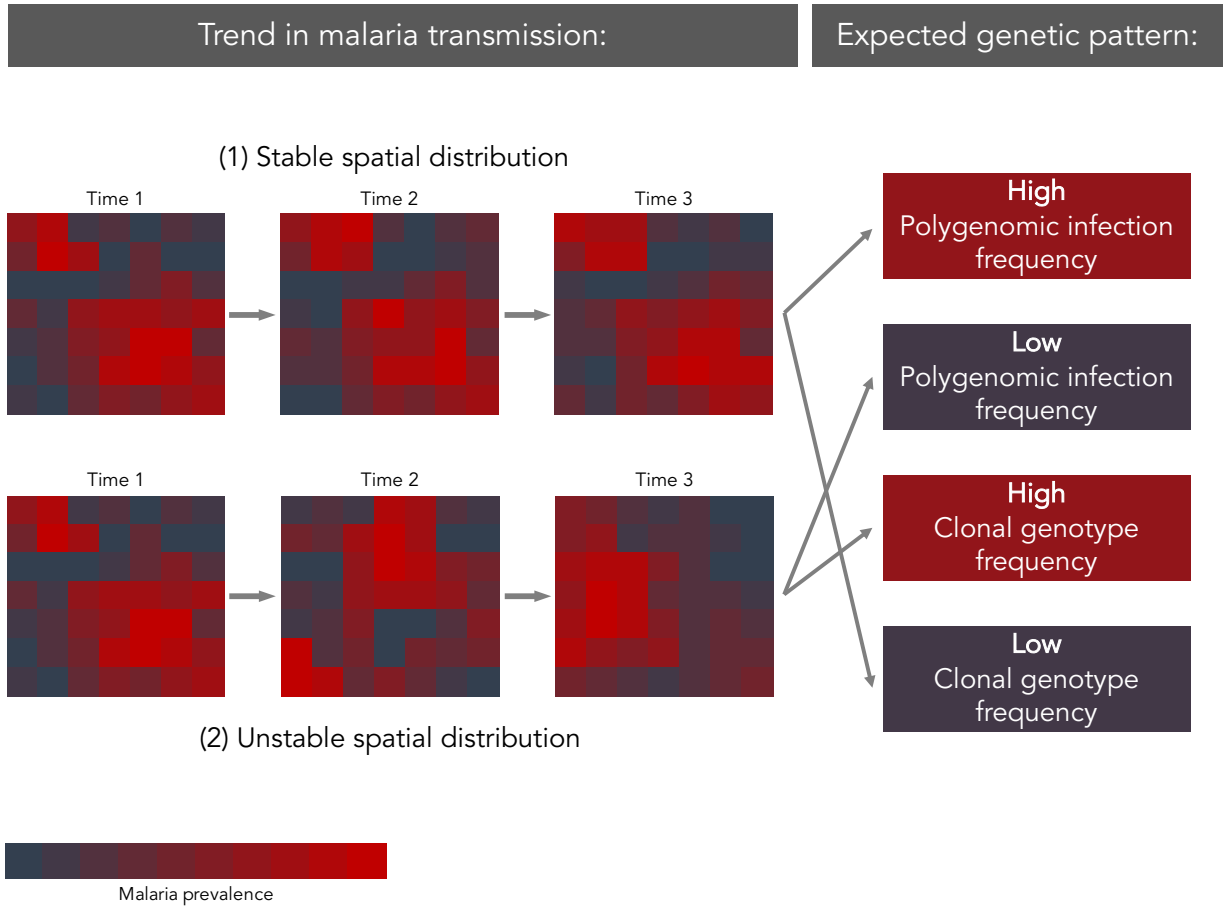

Figure 1

Using insight from genetic data to infer likely malaria transmission trends

A schematic of malaria transmission trends and the pattern of genetic variation they are hypothesized to produce. Changes in the intensity of malaria transmission in an area over time are shown per the color scale shown below the figure.

If variation in malaria prevalence is driven by rapid changes in malaria transmission over time, rapid identification of, and response to, emerging transmission hotspots would be an effective control strategy. On the other hand, if variation in malaria prevalence is driven by persistent pockets of higher transmission, prioritizing these areas for increased, sustained interventions would be a more effective control strategy.

In Madagascar, an analysis of infections reported by rural health clinics suggests there are regional differences in the temporal stability of malaria incidence, with western coastal regions having highly fluctuating incidence reports (Howes 2016, 2018, Ihantamalala 2018). For local hotspots of malaria transmission, Rice *et al* 2016 (Chapter 4 of this dissertation) report evidence that northeastern Madagascar contains rural, forested communities with higher than expected malaria transmission. Outside of Madagascar, local hotspots of malaria transmission have been observed in a wide variety of geographical settings and are receiving increasing attention from malaria control efforts (e.g., Mogeni 2017a, 2017b, Saita 2019, Pringle 2019, Stresman 2018)

To explore local malaria transmission patterns across ecological settings in Madagascar, approaches are needed that can provide insight into trends using cross-sectional sampling due to the scarcity of intra-annual, or inter-annual, longitudinal studies in many regions of Madagascar. Another complication in investigating spatio-temporal trends is that a substantial portion of malaria infected individuals can remain asymptomatic (Boussema 2014). In this study we used two approaches in an attempt to circumvent these challenges: (i) we used population genetic analysis of newly available cross-sectional samples from multiple regions of Madagascar, and (ii) we analyzed the available paired data on host temperature at the time of blood draw, a proxy for symptomatic state, to account for the presence of an asymptomatic reservoir.

5.2.2 Studying transmission dynamics in P. falciparum using population genetic data

As reviewed by Volkman *et al* 2012, and Chapters 1 and 4 of this manuscript, the pattern of genetic variation among the *P. falciparum* genomes in a sample can be used to gain insight into parasite transmission trends for a population. Rice *et al* 2016 (Chapter 4) previously validated a panel of polymorphic markers in Madagascar. Here we seek to deploy that SNP genotyping panel to the newly available samples of malaria infections from regions of Madagascar previously shown to contain communities with high relative prevalence (Chapter 3).

We aim to use genetic data to discriminate between two general transmission scenarios (see Figure 1). In the case of a temporally unstable spatial distribution of malaria transmission intensity, we expect to observe a pattern of a lower frequency of polygenomic infections and a higher frequency of repeatedly sampled clonal genotypes. However, in the case of a stable distribution of transmission intensity, we expect to observe a higher frequency of polygenomic infections and a lower frequency of repeatedly sampled clonal genotypes. Observing the latter pattern would indicate that the observation of prevalence levels much higher than national and regional averages at these sites (Chapter 3) is not likely explainable by a transient increase.

For this chapter, we extracted DNA from a subset of *P. falciparum* infections (n = 112) and from genotyping we observed a high frequency of asymptomatic infections with little evidence of clonal transmission. Genotypes were observed to be shared among multiple infections in the sample with low probability. The frequency of polygenomic infections differed between the west coast region (33%) and the southwestern region (78%), but additional sampling in the west coast region is likely needed. Implications for malaria control strategies are discussed.

5.3 Materials and Methods

5.3.1 Study regions and sample set

Putative *P. falciparum* infections were selected from previously completed rapid diagnostic test (RDT) surveys in two studyregions: (i)communities situated in the semi-arid, spiny forest and scrubland environment of Toliara II district (referred to as the southwest region of Madagascar in this manuscript), and (ii) communities situated in the dry, deciduous forest environment of Morombe district (referred to as the west coast region of Madagascar). Average annual rainfall (458 mm) and temperature (24.5C) in the west coast region is higher than in the southwest (343 mm, 24.1C) (Climate-Data Model 2014). Genetic data generated from analysis of samples from the west coast and southwest regions were compared to previously generated estimates from study communities in northeastern Madagascar (Maroantsetra district) (Rice 2016).

For each infection, paired data was available on location and temperature as a proxy for symptomatic versus asymptomatic status. Temperature was recorded using an infrared thermometer scan of the individual's temple. External body temperature readings require correction to accurately reflect internal body temperature; we use uncorrected data here to simply compare temperature profiles of infected and uninfected individuals at the time of blood sampling as a proxy for fever symptoms (see Figure 2).

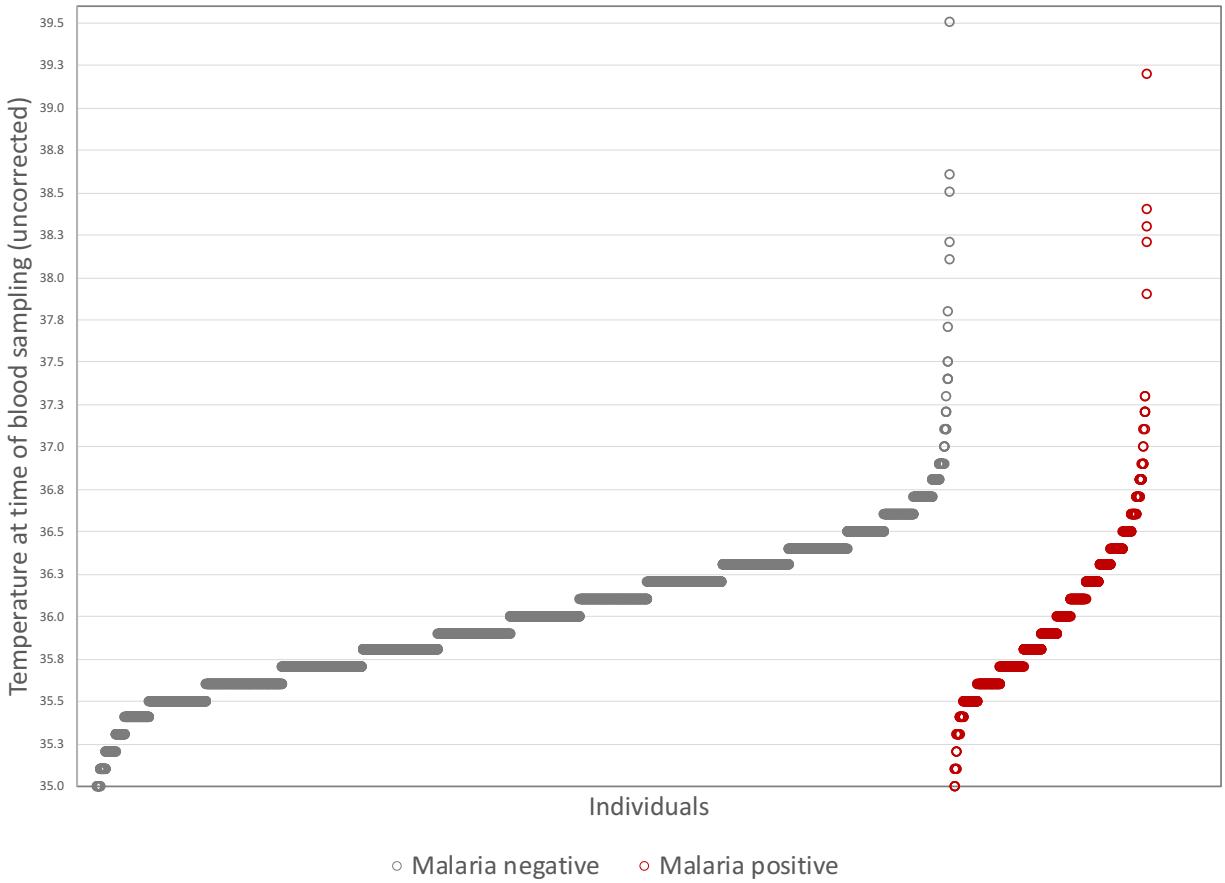

Figure 2

Temperature profile of individuals negative or positive for malaria by rapid diagnostic test

Uncorrected temperatures as taken by infrared thermometer at the temple. Individuals positive for malaria by RDT are shown in red.

5.3.2 DNA extraction and SNP genotyping

Methods for DNA extraction and SNP genotyping have been described previously (Rice 2016). Briefly, for a randomly chosen subset (n = 112) of the samples positive by RDT for *P. falciparum*, DNA was extracted from the corresponding dried blood spots using the DNA IQ™

Casework Pro for Maxwell® 16 kit. Following DNA extraction, alleles from a panel of 24 putatively neutral and unlinked genomic loci were typed to produce multi-locus genotypes. The SNP-genotyping protocol is described elsewhere (Rice 2016, Daniels 2008, Daniels 2013, Daniels 2015).

For determining the allele present at a locus, observing two alleles for a locus indicates that multiple, non-identical parasite genomes were present in the host at the time of sampling. Such infections are termed polygenomic infections. Conservatively, an infection is reported as polygenomic if multiple alleles are observed at least two of the loci analyzed (Daniels 2008, Daniels 2013). See Figure 3 for a representation of the SNP genotypes obtained for an example subset of monogenomic and polygenomic samples.

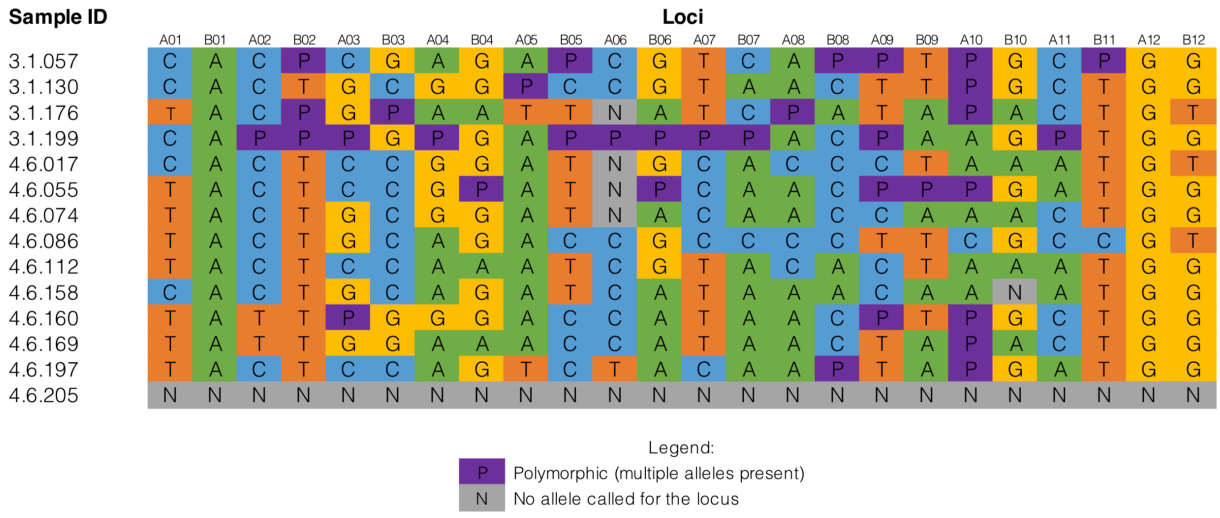

Figure 3

SNP genotyping results shown as a molecular barcode

A subset of the infections genotyped from southwestern and west coast of Madagascar is shown, with alleles colored by nucleotide. Polymorphic loci, where multiple alleles were observed, are

labeled P (shown in purple). Assays that failed to identify an allele are labeled N (shown in gray). One sample that was RDT positive but failed to amplify *P. falciparum* DNA at any locus (sample 4.6.205) is shown.

5.4 Results

5.4.1 SNP genotyping

First, in order to confirm that the panel of genomic markers analyzed were polymorphic and informative in these study populations, we calculated the allele frequency at each locus (see Figure 4). For three of the 24 loci, all monogenomic infections we sampled were fixed for the major allele. For two of these three fixed loci, we observed multiple individual infections that were polymorphic at these loci, indicating the minor allele was present within a parasite genome contributing to the polygenomic infection. The median minor allele frequency was 0.24 with 18 of the 24 loci having a minor allele frequency greater than 0.10. Among the 18 assayed loci with high minor allele frequency (>0.10), alleles could be successfully determined for an average of 86% of the samples analyzed for a given locus. For two of those 18 loci (loci 11 and 15), alleles could be determined for substantially less samples. Approximately 40% of samples did not yield a readable allele signal for those two loci, leaving 16 of the 24 loci that were found to have a high minor allele frequency and a high rate of successful allele calling.

Despite those problematic loci, sufficient resolution to begin investigating genotypic diversity in the populations remains as the 16 validated loci are: (i) distributed across 9 different *P. falciparum* chromosomes, and (ii) have a high minor allele frequency. The probability of observing an identical genotype among unrelated *P. falciparum* genomes decreases exponentially as a function of the number of loci and their minor allele frequency. For 16 loci

with an average minor allele (MAF) frequency of 0.24 the probability of an identical genotype being observed by chance is given by $[MAF^2 + (1\text{-}MAF)^2]^{16} = 7.0 \times 10^{-4}$.

There are several possible reasons for the failure of specific assayed loci to yield a readable allele. Experimental error, including PCR failure could explain some of the missing allele calls. To ameliorate this, failed assays will be repeated. Additionally, for those loci where a substantial number of samples failed, the assay for that locus could be more sensitive to a variant in the study population that interferes with primer binding or more sensitive to low concentrations of parasite DNA. Pre-amplification methods to increase the abundance of parasite DNA in the samples will be used to increase the probability of a higher allele calling success rate. Evidence supporting that low parasite DNA concentration may be present in these samples is that even among the successfully called allele assays, the real-time PCR based read outs of the assays often did not detect parasite DNA until the last cycles of amplification.

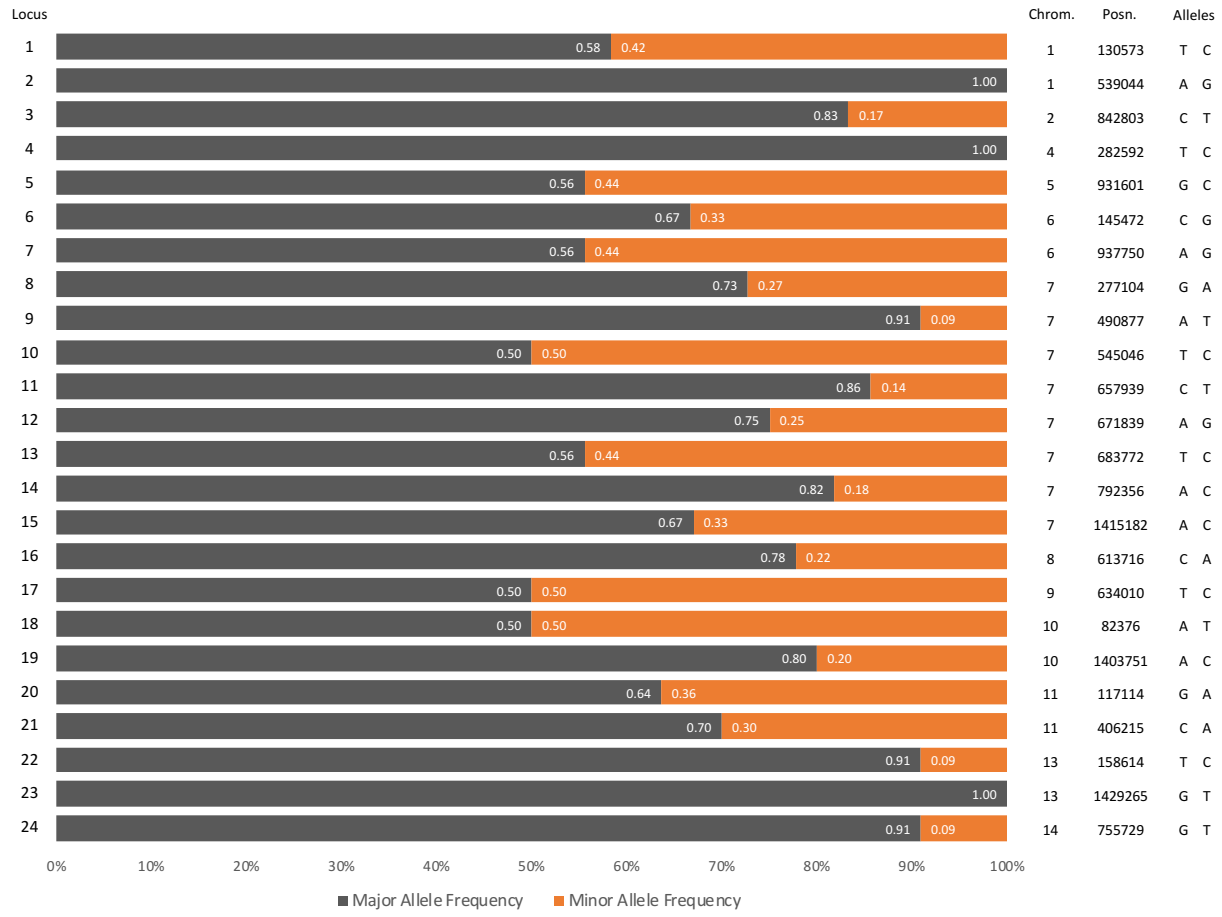

Figure 4

Minor allele frequency at the 24 loci analyzed

For each locus analyzed, the frequency of the major allele (shown in gray), frequency of the minor allele (shown in orange), chromosome (chrom.), genomic position on the chromosome (posn.), and major and minor alleles, are shown, respectively.

Among the samples analyzed, a small percentage (7.7%) failed to amplify *P. falciparum* DNA for any of the loci assayed (see sample 4.6.205 in Figure 3 for example). Likely explanations for these samples that were RDT positive but failed to show *P. falciparum* DNA are: (i) a failure to preserve or extract parasite DNA from the dried blood spots, (ii) a false positive RDT result for a malaria negative individual, (iii) the infection was caused by a non-*P. falciparum Plasmodium* species, and (iv) the infections were sampled late in the infectious course and parasite DNA had been cleared from the host blood stream prior to the clearance of the *P. falciparum* antibodies that RDTs detect. Regardless, this affected a small percentage of samples and is consistent with observations of previous studies (Rice 2016, Chapter 4).

5.4.2 Genetic correlates of transmission dynamics

We next analyzed two genetic parameters known to correlate with transmission intensity: within host parasite genetic diversity (the frequency polygenomic *P. falciparum* infections) and the genotypic diversity (the frequency of unique versus repeated genotypes) for these study populations. The proportion of infections that were polygenomic in the southwestern region (Toliara II district) (78.3%) was higher than was observed for the northeastern region (Maroantsetra district) (68.1%) and the west coast region (Morombe district) (33.3%) (see Figure 5).

From the subset of samples analyzed here, a noteworthy result was that the region with the lowest average prevalence at the time of sampling (the southwest) had the highest frequency of polygenomic infections. In all of the regions, none of the samples we determined to have monogenomic infection had a genotype that was identical to that of another monogenomic infection. This indicates that genotypic diversity and recombination rates were high in these

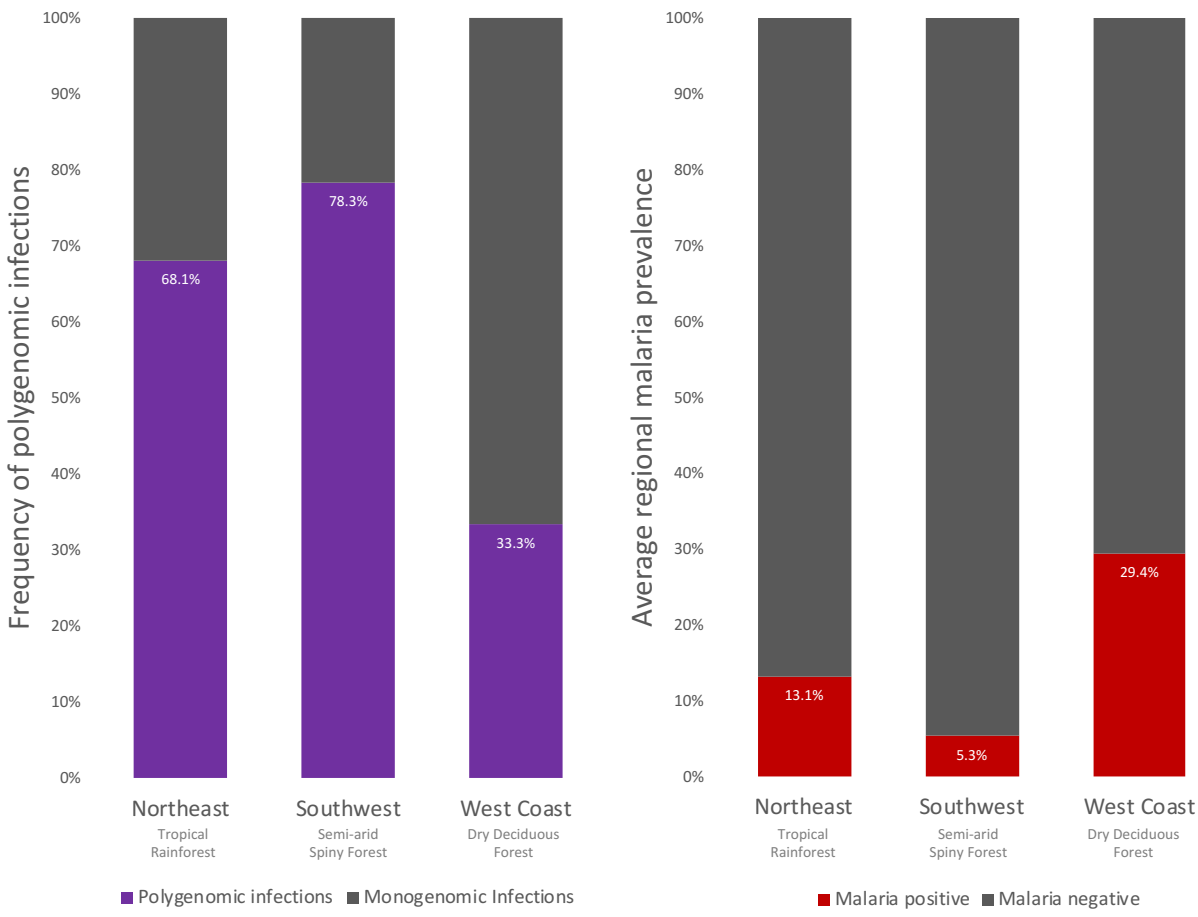

Figure 5

Frequency of polygenomic infections compared to local prevalence estimates

To the left is shown the frequency of *P. falciparum* infections that were polygenomic as defined by infections that contained multiple loci observed to be polymorphic. For comparison, mean prevalence is shown to the right for the sites in the Maroantsetra, Toliara II, and Morombe districts of the northeastern, southwestern, and the western coast regions of Madagascar, respectively.

populations, as there was no evidence of certain genotypes propagating clonally. Among the samples we determined to have polygenomic infections, there was a small proportion that were identical at the loci that were not polymorphic within those infections.

To investigate the potential relevance of asymptomatic individuals, we next analyzed the paired temperature data (shown in Figure 2). High temperatures, indicative of active fever, were commonly observed among neither the infected nor uninfected individuals, and more than 53% of infected individuals had a temperature below the median value for uninfected individuals.

5.5 Discussion

The failure to commonly observe the same multi-locus genotype(s) in multiple sampled infections suggests that in both the Toliara II and Morombe districts, recombination among circulating *P. falciparum* genomes is frequent and clonal stability over time is low. This suggests that the previous estimates of high parasite prevalence within some communities is unlikely to be explained by sporadic clonal outbreaks from a single source. Consistent with this observation, a large number of the infected individuals did not have evidence of symptoms at the time of sampling indicating historical transmission leading to the presence of some clinical immunity among individuals in these areas.

The implication of these findings is that residual pockets, where malaria transmission is sufficiently high to allow frequent recombination, likely exist in southwestern and west coast regions of Madagascar. Interventions aimed at reducing overall malaria burden in Madagascar may benefit from accounting for and targeting such pockets.

While it seems extremely unlikely that these genetic data could have been produced by very low malaria transmission intensities, limitations exist in our ability to confidently infer

trends in malaria transmission dynamics from a 24 SNP barcode. Foremost, it is difficult to distinguish between multiple possible explanations for observed variation in the polygenomic infection frequency between regions.

Polygenomic infections could predominantly arise from superinfection, where the polygenomic infection frequency is a function of the probability of infected hosts becoming re-infected. A lack of access to treatment for a subset of the population, local deforestation or water management increasing vector abundance near some households, or some other locally confined risk factor may explain the higher polygenomic infection frequency observed in southwestern Madagascar infections despite the lower average prevalence in that region overall.

Conversely, varying rates of co-infection, where polygenomic infections are co-transmitted between hosts (and superinfections are not needed to achieve within host parasite diversity), could also contribute to the variation in the frequency of polygenomic infections. It is possible that despite a lower parasite prevalence in the southwestern region, some aspect of mosquito biting exposure increases the probability of co-transmission. If co-infection is more common, clonality may decrease more rapidly after sudden outbreaks as more co-infections allow opportunities for recombination to rapidly introduce new genotypes into circulation. A larger panel of SNP markers to look at longer haplotypes and regions of the genome identical by descent may have higher resolution for these samples (Schaffner 2018, Wong 2017, 2018).

References

1. Goodman, Steven & P. Benstead, Jonathan & Schutz, Harald. (2004). The Natural History of Madagascar.

2. Howes RE, Mioramalala SA, Ramiranirina B, Franchard T, Rakotorahalahy AJ, Bisanzio D, et al. Contemporary epidemiological overview of malaria in Madagascar: operational utility of reported routine case data for malaria control planning. Malaria Journal. 2016;15. doi:10.1186/s12936-016-1556-3

3. Malaria Indicator Survey: Enquête sur les Indicateurs du Paludisme 2016 (MIS 2016). Programme National de Lutte contre le Paludisme Antananarivo, Madagascar

4. Kang SY, Battle KE, Gibson HS, Ratsimbasoa A, Randrianarivelojosia M, Ramboarina S, et al. Spatio-temporal mapping of Madagascar's Malaria Indicator Survey results to assess Plasmodium falciparum endemicity trends between 2011 and 2016. BMC Medicine. 2018;16. doi:10.1186/s12916-018-1060-4

5. Ihantamalala FA, Herbreteau V, Rakotoarimanana FMJ, Rakotondramanga JM, Cauchemez S, Rahoilijaona B, et al. Estimating sources and sinks of malaria parasites in Madagascar. Nature Communications. 2018b;9. doi:10.1038/s41467-018-06290-2

6. Howes RE, Zikursh M, Cramer EY, Ramiranirina B, Ratsimbasoa A, Kang SY, Zimmerman PA, et al. Risk Factors for Malaria Infection in Central Madagascar: Insights from a Cross-Sectional Population Survey. The American Journal of Tropical Medicine and Hygiene. 2018;99: 995–1002. doi:10.4269/ajtmh.18-0417

7. Mogeni P, Williams TN, Omedo I, Kimani D, Ngoi JM, Mwacharo J, et al. Detecting Malaria Hotspots: A Comparison of Rapid Diagnostic Test, Microscopy, and Polymerase Chain Reaction. The Journal of Infectious Diseases. 2017a;216: 1091–1098. doi:10.1093/infdis/jix321

8. Saita S, Pan-ngum W, Phuanukoonnon S, Sriwichai P, Silawan T, White LJ, et al. Human population movement and behavioural patterns in malaria hotspots on the Thai–Myanmar border: implications for malaria elimination. Malaria Journal. 2019;18. doi:10.1186/s12936-019-2704-3

9. Pringle JC, Tessema S, Wesolowski A, Chen A, Murphy M, Carpi G, et al. Genetic Evidence of Focal Plasmodium falciparum Transmission in a Pre-elimination Setting in Southern Province, Zambia. The Journal of Infectious Diseases. 2019;219: 1254–1263. doi:10.1093/infdis/jiy640

10. Stresman GH, Mwesigwa J, Achan J, Giorgi E, Worwui A, Jawara M, et al. Do hotspots fuel malaria transmission: a village-scale spatio-temporal analysis of a 2-year cohort study in The Gambia. BMC Medicine. 2018;16. doi:10.1186/s12916-018-1141-4

11. Mogeni P , The Hotspot Group Authors, Omedo I, Nyundo C, Kamau A, Noor A, et al. Effect of transmission intensity on hotspots and micro-epidemiology of malaria in sub-Saharan Africa. BMC Medicine. 2017b;15. doi:10.1186/s12916-017-0887-4

12. Bousema T, Okell L, Felger I, Drakeley C. Asymptomatic malaria infections: detectability, transmissibility and public health relevance. Nature Reviews Microbiology. 2014;12: 833–840. doi:10.1038/nrmicro3364

13. Volkman SK, Neafsey DE, Schaffner SF, Park DJ, Wirth DF. Harnessing genomics and genome biology to understand malaria biology. Nature Reviews Genetics. 2012;13: 315–328. doi:10.1038/nrg3187

14. Climate model open source 2019. Climate Data.org (https://en.climate-data.org/info/sources/)

15. Rice BL, Golden CD, Anjaranirina EJG, Botelho CM, Volkman SK, Hartl DL. Genetic evidence that the Makira region in northeastern Madagascar is a hotspot of malaria transmission. Malaria Journal. 2016;15. doi:10.1186/s12936-016-1644-4

16. Daniels R, Volkman SK, Milner DA, Mahesh N, Neafsey DE, Park DJ, et al. A general SNP-based molecular barcode for Plasmodium falciparum identification and tracking. Malaria Journal. 2008;7: 223. doi:10.1186/1475-2875-7-223

17. Daniels R, Chang H-H, Séne PD, Park DC, Neafsey DE, Schaffner SF, et al. Genetic Surveillance Detects Both Clonal and Epidemic Transmission of Malaria following Enhanced Intervention in Senegal. Mayor A, editor. PLoS ONE. 2013;8: e60780. doi:10.1371/journal.pone.0060780

18. Daniels R, Hamilton EJ, Durfee K, Ndiaye D, Wirth DF, Hartl DL, et al. Methods to Increase the sensitivity of high resolution melting single nucleotide polymorphism genotyping in malaria. J Vis Exp 2015:e52839.

19. Schaffner SF, Taylor AR, Wong W, Wirth DF, Neafsey DE. hmmIBD: software to infer pairwise identity by descent between haploid genotypes. Malaria Journal. 2018;17. doi:10.1186/s12936-018-2349-7

20. Wong W, Wenger EA, Hartl DL, Wirth DF. Modeling the genetic relatedness of Plasmodium falciparum parasites following meiotic recombination and cotransmission. Pascual M, editor. PLOS Computational Biology. 2018;14: e1005923. doi:10.1371/journal.pcbi.1005923

21. Wong W, Griggs AD, Daniels RF, Schaffner SF, Ndiaye D, Bei AK, et al. Genetic relatedness analysis reveals the cotransmission of genetically related Plasmodium falciparum parasites in Thiès, Senegal. Genome Medicine. 2017;9. doi:10.1186/s13073-017-0398-0