Plasmodium Population Structure in the Context of Malaria Control and Elimination

by

Stella M. Chenet

A Dissertation Presented in Partial Fullfilment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2014 by the
Graduate Supervisory Committee

Ananias A. Escalante, Chair
Josephine Clark-Curtiss
Michael Rosenberg
Jesse E. Taylor

ARIZONA STATE UNIVERSITY

May 2014

ABSTRACT

Malaria is a vector-borne parasitic disease affecting tropical and subtropical regions. Regardless control efforts, malaria incidence is still incredible high with 219 million clinical cases and an estimated 660,000 related deaths (WHO, 2012). In this project, different population genetic approaches were explored to characterize parasite populations. The goal was to create a framework that considered temporal and spatial changes of *Plasmodium* populations in malaria surveillance. This is critical in a vector borne disease in areas of low transmission where there is not accurate information of when and where a patient was infected. In this study, fragment analysis data and single nucleotide polymorphism (SNPs) from South American samples were used to characterize *Plasmodium* population structure, patterns of migration and gene flow, and discuss approaches to differentiate reinfection vs. recrudescence cases in clinical trials. A Bayesian approach was also applied to analyze the *Plasmodium* population history by inferring genealogies using microsatellites data. Specifically, fluctuations in the parasite population and the age of different parasite lineages were evaluated through time in order to relate them with the malaria control plan in force. These studies are important to understand the turnover or persistence of "clones" circulating in a specific area through time and consider them in drug efficacy studies. Moreover, this methodology is useful for assessing changes in malaria transmission and for more efficiently manage resources to deploy control measures in locations that act as parasite "sources" for other regions. Overall, these results stress the importance of monitoring malaria demographic changes when assessing the success of elimination programs in areas of low transmission.

TABLE OF CONTENTS

Page

CHAPTER Page

CHAPTER Page

CHAPTER Page

APPENDIX

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1

INTRODUCTION

Malaria is a major global health problem estimated to cause 219 million clinical cases and an estimated 660 000 deaths yearly in tropical and subtropical regions (WHO, 2012a). Although new strategies and efforts for malaria control have been deployed, malaria still represents a major health burden. This vector borne disease caused by parasites of the genus *Plasmodium* is endemic in 99 countries with transmission intensities varying from very low (parasite prevalence<10% or hypoendemic), to moderate transmission (10-50% parasite prevalence), to extremely high (parasite prevalence in children aged 2-9 years >50% or holoendemic) (WHO, 2012b).

Due to the wide range of malaria transmission rates, control programs focus on reducing the disease burden until it is maintained at a reasonable level to proceed towards the elimination step. Nevertheless, local heterogeneity in malaria incidence reduces the efficacy of control strategies and profiling these cases is important to determine the kind of health care services needed (Carter et al., 2000). For control programs to be successful, research on useful and efficient modes of detection, monitoring and response must be undertaken. By understanding the parasite genetic diversity and gene flow dynamics with and between human populations, effective spatial and temporal boundaries for intervention can be implemented. Accordingly, one of the main strategies should focus on characterizing changes in *Plasmodium* population genetics to determine fluctuations in local transmission intensity and provide valuable tools for malaria surveillance and the evaluation of malaria elimination programs.

1.1. *Plasmodium* life cycle

Malaria is a vector borne disease caused by apicomplexan parasites of the genus *Plasmodium: Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium ovale*, and *Plasmodium malariae* (outbreaks of the non-human malaria, *P. knowlesi*, in humans have also been reported). All *Plasmodium* species examined to date have 14 chromosomes, with one mitochondrion and one plastid genome. The chromosomes vary in length from 500 kilobases to 3.5 megabases (Su et al., 2007).

Infections in humans begin with the bite of an infected female Anopheline mosquito. Sporozoites released from the salivary glands of the mosquito enter the bloodstream during feeding and quickly invade hepatocytes. During the next two weeks, the liver-stage parasites differentiate and undergo asexual multiplication resulting in tens of thousands of merozoites which burst from the hepatoctye. However, in *P. vivax* and *P. ovale* not all infected liver cells burst. Parasites in the liver cells which do not burst (hypnozoites) may remain for years and are responsible for new episodes of the disease if reactivated. Individual merozoites invade red blood cells and undergo an additional round of multiplication within a schizont. The length of the erythrocytic stage depends on the parasite species: 48 hours for *P. falciparum*, *P. vivax*, and *P. ovale* and 72 hours for *P. malariae*. In total, the incubation interval of *P. falciparum* (23-39 days) is longer than in *P. vivax* (20 days), making the control in *P. vivax* more complex, especially if we consider possible late relapses (WHO, 2007). Clinical manifestations of malaria, such as, fever and chills, are associated with the synchronous rupture of the infected erythrocyte. The released merozoites invade additional erythrocytes. However, some merozoites

differentiate into sexual forms. These gametocytes are taken up by the mosquito during a blood meal. Within the mosquito midgut, the male gametocyte undergoes further division, producing 8 flagellated microgametes which fertilize the female macrogamete. The resulting ookinete encysts on the exterior of the gut wall as an oocyst. Then, the oocyst ruptures, releasing hundreds of sporozoites into the mosquito body cavity where they eventually migrate to the mosquito salivary gland (CDC, 2012).

1.2.Malaria control and elimination

In high transmission settings, the first objective of malaria control programs is to decrease the overall malaria burden by reducing transmission intensity with preventive methods such as vector control and insecticide impregnated bed nets (WHO 2012a; Alonso et al., 2011). Under this scenario, a sensitive malaria case definition is required (i.e. detection of febrile cases) to accomplish high treatment coverage and reduce mortality. A different approach is considered in hypoendemic or mesoendemic areas, where a methodology to monitor malaria foci and correctly identify infected people is crucial for malaria control programs. These areas mimic the epidemiologic conditions during the transition from control to elimination with no more than five new cases per 1000 population at risk per year (The malERA Consultative Group on Diagnoses and Diagnostics, 2011).

For evaluating the success of malaria control programs, it is necessary to have a good estimation of malaria incidence by country. Currently, the most common methods rely on routine surveillance reports of active malaria cases from health centers and on

population-based surveys of parasite prevalence and case incidence from selected locations to generate risk maps across malaria endemic areas (Hay et al., 2009). Whereas both methods have limitations, estimates derived from routine surveillance data are usually lower than those derived from surveys of parasite prevalence due to incomplete case detection and the way patients use public and private health services (Cibulskis et al., 2011). Since traditional case detection depends on several factors such as drug use, drug resistance, acquired immunity and parasite seasonal variation, the number of circulating parasites in the bloodstream is clearly affected. As a consequence, transmission intensity could be underestimated (Corran et al., 2007). Another standard measure of transmission intensity is the entomological inoculation rate (EIR: mean number of infectious mosquito bites per individual per year); however, this method also lacks precision because mosquito distributions are markedly heterogeneous (Drakeley et al., 2003) and the mosquito biting rates are difficult to estimate (Smith et al., 2004).

A more robust, relatively low tech and high throughput malaria detection tool relies on seroconversion measure (Corran et al., 2007; Bejon et al., 2010). *Plasmodium* exposed individuals remain seropositive for antimalarial antibodies for long periods of time and the seroconversion rate is related to the force of malaria infection (Corran et al., 2007). However, these assays need to be adapted according to the transmission settings and considering areas where *P. falciparum* and *P. vivax* are found in sympatry. Plus, there are notorious differences among antigens so the definition of force of infections still remains elusive and need to be further studied since it is relative to the antigen used.

1.3. Malaria foci and *Plasmodium* hotspots

A good estimate of malaria prevalence is helpful to identify foci of residual transmission over time and space. These malaria foci are well defined locations containing the epidemiological and climatic factors necessary for malaria transmission to occur (WHO, 2007). Within foci there are specific areas or groups of people where transmission intensity is much stronger due to the most favorable conditions (hotspots); for instance, a neighborhood in close proximity to mosquito breeding sites or a specific highly vulnerable group of people such as infants carrying high loads of parasitemia (Bousema et al., 2012) or pregnant women that cannot receive primaquine (for eradication of *P. vivax* or *P. ovale* hypnozoites) as part of their treatment. People that do not develop the disease but carry gametocytes or hypnozoites (that could be later activated), also constitute a hotspot since they participate in the transmission process. Once such malaria hotspots have been identified, establishing an appropriate case definition based on diagnosis (considering parasitemia, gametocytemia and *Plasmodium* species) is critical for malaria treatment. If infections are left untreated, they can last for many months and cause recurring clinical episodes with asymptomatic periods (Crowell et al., 2013).

Nowadays, malaria diagnosis relies on microscopy and/or rapid diagnostic tests (RDTs). Whereas both methods have major drawbacks such as the difficulty to identify mixed infections and/or low levels of parasitemia and the extremely parasite variability or lack of antigens (Gamboa et al., 2010; Koita et al., 2012); they have proven effective in areas with high malaria incidence. Nonetheless, when the goal is to obtain information other than just the record keeping of cases, molecular diagnostic tools provide a better picture

of the type of infection, clarifying patterns of transmission (in conjunction with epidemiological data) and drug resistance. Thus, in low transmission areas, malaria control programs need to ascertain not only that antimalarial drugs are being appropriately used, but also that the effectiveness of their actions is reducing transmission (WHO, 2012b). Under such circumstances, tools and measures that characterize changes in *Plasmodium* populations are required. Furthermore, associations can be established between malaria infections and the individuals driving transmission or hot-pops (Cotter et al., 2013). Such studies considering heterogeneities in time and space are necessary to characterize and determine which particular subpopulations or clonal lineages adapted to specific microenvironments and are maintained through time.

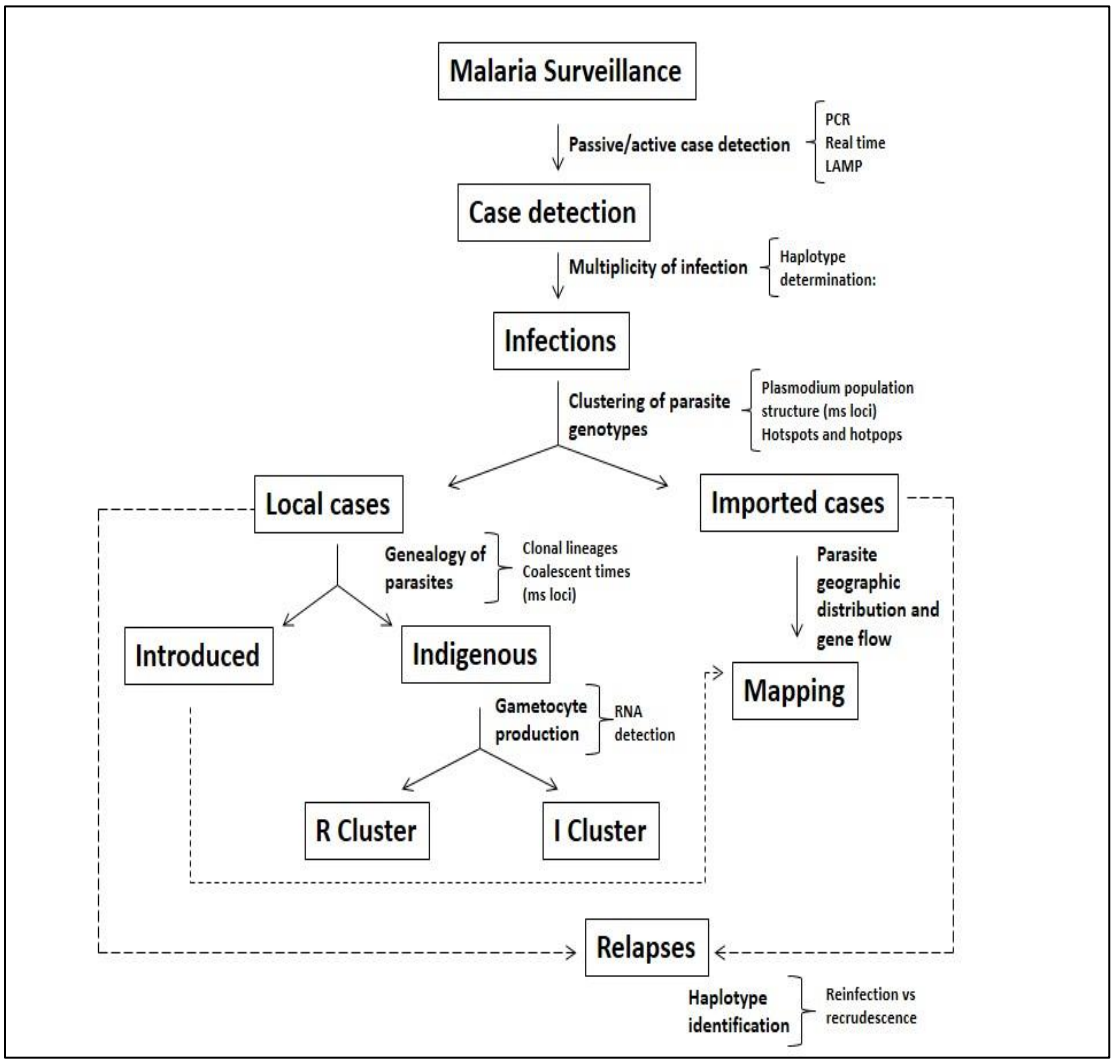
Integrating mapping activities of hotspots with the outputs of epidemiological surveillance systems allows defining the limits of adverse conditions for transmission. To maintain interrupted transmission or elimination, malaria control programs require an accurate estimate of the effective population size of *Plasmodium* and their number of genotypes circulating among human mobile populations. When focusing on migrants, time-location estimation needs at least two points in time and space to accurately identify mobile hot-spots also known as hot-pops (e.g. miners in South America) and *Plasmodium* haplotypes associated (Blair et al, 2006; da Silva-Nunes et al., 2012). Mapping the routes of parasite dispersal by human carriers will allow targeting both regions: where imported infections originate (source population) and where they substantially contribute to transmission (sink population). Particularly, in low transmission countries, imported cases comprise a significant proportion of all cases and pose a risk for malaria resurgence

in such areas. On the other hand, if local cases are reported, then it is important to differentiate between introduced (first-generation local transmission) and indigenous (second or higher generation) cases since these indicate the origin of active transmission (WHO, 2007). Thus, the main focus for a successful malaria control strategy is to integrate spatial epidemiology with molecular epidemiology and identify the specific *Plasmodium* lineages associated with malaria hotspots.

1.4. *Plasmodium* population monitoring and evaluation framework

Although many programmatic responses to detect malaria infections exist, there is not a systematic description on how genotyping and molecular studies could contribute towards developing tools to control and manage malaria cases and how these could be included in a surveillance program. The ultimate goal would be to integrate novel monitoring and malaria evaluation metrics to an accurate mapping of foci including circulating *Plasmodium* genotypes and human individuals driving transmission. According to this objective, we describe the steps to maximize the interpretation of molecular epidemiology data including *Plasmodium* population genetic analysis for proper management decisions in control programs (Figure 1).

Figure 1. Molecular methodology to incorporate parasite genetic information in epidemiological studies (ms: microsatellite, R: resistant and I: infectious clusters).



First of all, adequate sampling of *Plasmodium* infections in a locality is necessary to perform genetic population studies. In low transmission areas, such as in South America, sampling from different localities within a country, as well as from different time points, are necessary to characterize the dynamics of patchy, structured populations since they may need different control strategies at a national level. Such sampling should include not only infections from symptomatic patients but also asymptomatic people and people

living in areas of difficult accessibility that are less likely to receive health care. Thus, active case detection and follow up of malaria cases are needed to predict and control any possible outbreaks. The requirements for case screening in areas in which malaria has been eliminated but are still vulnerable to malaria transmission are similar to those for countries in the elimination phase. Under elimination programs, it is necessary to include genotyping methods for identifying morphologically similar species and for distinguishing new infections from relapses and recrudescence cases (WHO, 2007). The use of PCR-based methods to detect malaria parasites in blood samples increases the sensitivity of detection compared with microscopy. Nowadays, qualitative and quantitative PCR protocols that are robust, sensitive and species specific are not only available in the lab but there are also alternative methods that could be used in the field (Lucchi et al., 2010). Moreover, it is possible to determine the prevalence of gametocyte density and their potential risk in different transmission settings by the use of Reverse transcription polymerase chain reaction (RT-PCR), QT-NASBA (Real-time quantitative nucleic acid sequence-based amplification) and RT-LAMP (Reverse-transcription, loop-mediated isothermal amplification). Infected individuals with gametocytes in their bloodstream represent natural reservoirs for malaria. Mature gametocytes are detectable in the bloodstream at days 7 to 15 after the initial wave of asexual parasites from which they are derived. Hence, interventions to monitor gametocyte load in individuals are necessary to differentiate between potential reservoirs and susceptible individuals. This is not an easy process because of the highly specialized nature of gametocytes and the individual roles of both sexes. Around 250 to 300 genes are upregulated at the mRNA level during gametocyte development (Bousema and Drakeley, 2011). So far, there is

limited evidence on how differences in gametocytogenesis and sex ratio might affect transmission intensity.

On the other hand, complementary to the clustering analysis and the identification of hotspots, population genetics allows us to infer the genetic relationship between *Plasmodium* lineages to estimate time, their persistence and their concentration in specific areas or groups of people. In addition, studying the demographic history of parasites is useful to obtain inferences about the magnitude of *Plasmodium* population size changes to evaluate the effectiveness of malaria control programs. In general, a *Plasmodium* population genetic study must start with a decision regarding appropriate genetic markers. Markers must have the correct sensitivity for the research question. It is possible to have too much information (if entities are too different there is nothing to link them) or too little information (no signal). Genetic polymorphisms such as single nucleotide polymorphisms (SNPs) and short tandem repeats (STRs) are valuable to explore the dynamics of parasites with important phenotypes. Such genetic information is used as molecular barcodes to track the origin of infections (Campino et al., 2011). However, it is important to acknowledge how each of the different genetic markers could provide or not precise information for control and elimination purposes. For instance, SNPs level of polymorphism and rates of evolution may not be suitable for operational investigations that require short evolutionary time scales, such as seasonal malaria episodes (Chenet et al., 2012). Microsatellites, on the other hand, would be more suitable for this purpose due to their mutation rate; they gain and lose repeat units by DNA-replication slippage, a mutation mechanism that is specific to tandemly repeated

sequences (Schlötterer et al., 2000). Although mutation patterns of microsatellites are far more complex than the frequently assumed stepwise mutation model (Wu and Drummond, 2011), they are still valuable to provide information for epidemiological time scales. Microsatellites have been extensively used to investigate the population structure of *Plasmodium* (Anderson et al., 1999, Imwong et al., 2004; Karunaweera et al., 2008; Van den Eede et al., 2010; Chenet et al 2012) and optimized protocols for microsatellite genotyping and analysis in *Plasmodium falciparum* and *Plasmodium vivax* are currently available (Chenet et al 2012; Orjuela-Sánchez et al., 2013) as well as guidelines for data interpretation according to the specific research question. Considering the evaluation time frame of three years to certify an area as malaria free, the most suitable markers would be those evolving more quickly such as microsatellites.

Moreover, in low transmission areas the identification of sensitive or resistant clusters of parasites (or subpopulations) based on few molecular markers is more plausible than trying to accomplish individual identification since there is strong clonal structure in these areas (Griffing et al., 2011; Echeverry, 2013). Protocols to track antimalarial drug resistance are extremely important in the context of malaria control. Thus, as elimination is approached, therapy and follow up should be monitored directly to ensure that individuals have cleared all parasites. Although control measures have significantly reduced *P. falciparum* transmission, *P. vivax* detection and management still represent a major issue because of the presence of hypnozoites, a potential parasite reservoir. Since *P. vivax* can remain latent in the liver, and produce relapses, its effective management requires the use of 8-aminoquinolones (e.g. primaquine) (Shanks, 2012). However,

further studies are needed for a reliable and easiest way to detect glucose 6-phosphate dehydrogenase (G6PD) deficiency in areas with *P. vivax* to prevent hemolysis in patients treated with primaquine (Howes et al., 2013).

Hence, by considering the malaria heterogeneity in transmission through time and space and the correct identification of mobile hot-pops using epidemiological data (i.e. age, sex, economic activity, frequency of travels, clinical symptoms/asymptomatic) plus *Plasmodium* genetic analysis (higher probability of mixed infections, persistent/drug resistant lineages, gametocytemia), proper surveillance of high malaria risk groups can be accomplished.

1.5. Summary

Accurate measurement of malaria transmission is essential for the monitoring and evaluation of malaria control programs targeting elimination. Such programs need rapid, sensitive, standardized, and reproducible transmission measurement methods focused on *Plasmodium* population studies to monitor the program progress. The assumption that the parasite population definition is the same as the human or the mosquito populations bias the true location of hotspots (and more importantly hot-pops) and mislead control efforts. When achieving malaria elimination, new infections are a direct measure of ongoing transmission and require labor-intensive, active surveillance studies, particularly where immune individuals are unlikely to experience symptomatic disease. Targeting these hot-pops could be challenging if we consider that the follow up of asymptomatic or mobile human groups represents a greater investment of time, money and human resources but it

is necessary to maintain malaria free areas. In conclusion, it is important to utilize the genetic information gather in each of the *Plasmodium* population surveillance steps to answer specific questions about the success of the malaria control program and how this should be oriented towards the monitoring of cases. Thus, integrated approaches for malaria control programs that consider not only the biology of the vector and the epidemiological data but also the *Plasmodium* population and human behavior dynamics are needed to enforce control and prevent reintroduction of malaria.

1.6.References

Alonso, P.L., Brown, G., Arevalo-Herrera, M., Binka, F., Chitnis, C., Collins, F., Doumbo, O.K., Greenwood, B., Hall, B.F., Levine, M.M., Mendis, K., Newman, R.D., Plowe, C.V., Rodríguez, M.H., Sinden, R., Slutsker, L., Tanner, M. (2011). A research agenda to underpin malaria eradication. PLoS Medicine, 8: e1000406.

Anderson, T.J., Haubold, B., Williams, J.T., Estrada-Franco, J.G., Richardson, L., Mollinedo, R., Bockarie, M., Mokili, J., Mharakurwa, S., French, N., Whitworth, J., Velez, I.D., Brockman, A.H., Nosten, F., Ferreira, M.U., Day, K.P. (2000). Microsatellite markers reveal a spectrum of population structures in the malaria parasite Plasmodium falciparum. Molecular Biology Evolution, 17(10),1467-82.

Anderson, T.J., Su, X.Z., Bockarie, M., Lagog, M., Day, K.P. (1999). Twelve microsatellite markers for characterization of Plasmodium falciparum from finger-prick blood samples. Parasitology, 119:113-25.

Bejon, P., Williams, T.N., Liljander, A., Noor, A.M., Wambua, J., Ogada, E., Olotu, A., Osier, F.H., Hay, S.I., Färnert, A., Marsh, K. (2010). Stable and Unstable Malaria Hotspots in Longitudinal Cohort Studies in Kenya. PLoS Med, 7(7): e1000304. doi:10.1371/journal.pmed.1000304.

Blair, S., Carmona-Fonseca, J., Piñeros, J.G., Ríos, A., Alvarez, T., Alvarez, G., Tobón, A. (2006). Therapeutic efficacy test in malaria falciparum in Antioquia, Colombia. Malaria Journal, 5:14.

Bousema, T., Griffin, J.T., Sauerwein, R.W., Smith, D.L., Churcher, T.S., Takken, W., Ghani, A., Drakeley, C., Gosling, R. (2012). Hitting hotspots: spatial targeting of malaria for control and elimination. PLoS, 9:e1001165. doi: 10.1371/journal.pmed.1001165.

Bousema, T., & Drakeley, C. (2011). Epidemiology and infectivity of *Plasmodium falciparum* and *Plasmodium vivax* gametocytes in relation to malaria control and elimination. Clinical Microbiology Review, 24: 377–410.

Campino, S., Auburn, S., Kivinen, K., Zongo, I., Ouedraogo, J.B., Mangano, V., Djimde, A., Doumbo, O.K., Kiara, S.M., Nzila, A., Borrmann, S., Marsh, K., Michon, P., Mueller, I., Siba, P., Jiang, H., Su, X.Z., Amaratunga, C., Socheat, D., Fairhurst, R.M., Imwong, M., Anderson, T., Nosten, F., White, N.J., Gwilliam, R., Deloukas, P., MacInnis, B., Newbold, C.I., Rockett, K., Clark, T.G., Kwiatkowski, DP. (2011). Population genetic analysis of Plasmodium falciparum parasites using acustomized Illumina GoldenGate genotyping assay. PLoS One, 6(6):e20251. doi: 10.1371/journal.pone.0020251.

Carter, R, Mendis KN, Roberts, D. (2000). Bulletin of the World Health Organization. World Health Organization, 78 (12), 1401-1411.
Chenet, S.M., Schneider, K.A., Villegas, L., Escalante, A.A. (2012). Local population structure of Plasmodium: impact on malaria control and elimination. Malaria Journal, 11:412. doi: 10.1186/1475-2875-11-412.

Cibulskis, R.E., Aregawi, M., Williams, R., Otten, M., Dye, C. (2011). Worldwide incidence of malaria in 2009: estimates, time trends, and a critique of methods. PLoS Medicine, 8(12):e1001142. doi: 10.1371/journal.pmed.1001142.

Corran, P., Coleman, P., Riley, E., Drakeley, C. (2007) Serology: a robust indicator of malaria transmission intensity? Trends in Parasitology, 23:575-82.
Cotter, C., Sturrock, H.J., Hsiang, M.S., Liu, J., Phillips, A.A., Hwang, J., Gueye, C.S., Fullman, N., Gosling, R.D., Feachem, R.G. (2013). The changing epidemiology of malaria elimination: new strategies for new challenges. Lancet, 13: 60310-4. doi: 10.1016/S0140-6736(13)60310-4.

Crowell, V., Yukich, J.O., Briët, O.J.T., Ross, A., Smith, T.A. (2013). A Novel Approach for Measuring the Burden of Uncomplicated Plasmodium falciparum Malaria: Application to Data from Zambia. PLoS ONE, 8(2): e57297. doi:10.1371/journal.pone.0057297.

da Silva-Nunes, M., Moreno, M., Conn, J.E., Gamboa, D., Abeles, S., Vinetz, J.M., Ferreira, M.U. (2012). Amazonian malaria: Asymptomatic human reservoirs, diagnostic challenges, environmentally driven changes in mosquito vector populations, and the mandate for sustainable control strategies. Acta Tropica, 121: 281-291.

Drakeley, C., Schellenberg, D., Kihonda, J., Sousa, C.A., Arez, A.P., Lopes, D., Lines, J., Mshinda, H., Lengeler, C., Armstrong, Schellenberg, J., Tanner, M., Alonso, P. (2013). An estimation of the entomological inoculation rate for Ifakara: a semi-urban area in a region of intense malaria transmission in Tanzania. Tropical Medicine & International Health, 8(9):767-74.

Echeverry, D.F., Nair, S., Osorio, L., Menon, S., Murillo, C., Anderson, T.J. (2013). Long term persistence of clonal malaria parasite Plasmodium falciparum lineages in the Colombian Pacific region. BMC Genetics, 14:2. doi: 10.1186/1471-2156-14-2.

Gamboa, D., Ho, M.F., Bendezu, J., Torres, K., Chiodini, P.L., Barnwell, J.W., Incardona, S., Perkins, M., Bell, D., McCarthy, J., Cheng, Q. (2010). A large proportion of P. falciparum isolates in the Amazon region of Peru lack pfhrp2 and pfhrp3: implications for malaria rapid diagnostic tests. PLoS One, 5(1):e8091. doi: 10.1371/journal.pone.0008091.

Griffing, S.M., Mixson-Hayden, T., Sridaran, S., Alam, M.T., McCollum, A.M., Cabezas, C., Marquiño Quezada, W., Barnwell, J.W., De Oliveira, A.M., Lucas, C., Arrospide, N., Escalante, A.A., Bacon, D.J., Udhayakumar, V. (2011). South American Plasmodium falciparum after the malaria eradication era: clonal population expansion and survival of the fittest hybrids. PLoS One, 6:e23486. doi: 10.1371.

Hay, S.I., Guerra, C.A., Gething, P.W., Patil, A.P., Tatem, A.J., Noor, A.M., Kabaria, C.W., Manh, B.H., Elyazar, I.R., Brooker, S., Smith, D.L., Moyeed, R.A., Snow, R.W. (2009). A world malaria map: Plasmodium falciparum endemicity in 2007. PLoS Med, 6:e1000048. doi: 10.1371/journal.pmed.1000048.

Howes, R.E., Battle, K.E., Satyagraha, A.W., Baird, J.K., Hay, S.I. (2013). G6PD deficiency: global distribution, genetic variants and primaquine therapy. Advances in Parasitology, 81:133-201.

Imwong, M., Nair, S., Pukrittayakamee, S., Sudimack, D., Williams, J.T., Mayxay, M., Newton, P.N., Kim, J.R., Nandy, A., Osorio, L., Carlton, J.M., White, N.J., Day, N.P., Anderson, T.J. (2004). Contrasting genetic structure in Plasmodium vivax populations from Asia and South America. Journal of Infectious Diseases, 190:1547-55.

Karunaweera, N.D., Ferreira, M.U., Munasinghe, A., Barnwell, J.W., Collins, W.E., King, C.L., Kawamoto, F., Hartl, D.L., Wirth, D.F. (2008). Extensive microsatellite diversity in the human malaria parasite *Plasmodium vivax*. Gene, 410:105-12.

Koita, O.A., Doumbo, O.K., Ouattara, A., Tall, L.K., Konaré, A., Diakité, M., Diallo, M., Sagara, I., Masinde, G.L., Doumbo, S.N., Dolo, A., Tounkara, A., Traoré, I., Krogstad, D.J. (2012). False-negative rapid diagnostic tests for malaria and deletion of the histidine-rich repeat region of the hrp2 gene. American Journal of Tropical Medicine & Hygiene, 86: 194–198.

Lucchi, N.W., Demas, A., Narayanan, J., Sumari, D., Kabanywanyi, A., Kachur, S.P., Barnwell, J.W., Udhayakumar, V. (2010). Real-time fluorescence loop mediated isothermal amplification for the diagnosis of malaria. PLoS One, 5:e13733. doi: 10.1371/journal.pone.0013733.

Orjuela-Sánchez, P., Brandi, M.C., Ferreira, M.U. (2013). Microsatellite Analysis of Malaria Parasites. Methods in Molecular Biology, 1006: 247-258.

Schlötterer, C. (2000). Evolutionary dynamics of microsatellite DNA. Chromosoma, 109(6):365-71.

Shanks, G.D. (2012). Control and elimination of Plasmodium vivax. Advances in Parasitology, 80:301-41.

Smith, D.L., Dushoff, J., McKenzie, F.E. (2004). The risk of a mosquito-borne infection in a heterogeneous environment. PLoS Biol, 2(11):e368.

Su, X., Hayton, K., Wellems, T.E. (2007). Genetic linkage and association analyses for trait mapping in Plasmodium falciparum. Nature Reviews Genetics, 8(7):497-506.
The malERA Consultative Group on Diagnoses and Diagnostics. (2011). A research agenda for malaria eradication: diagnoses and diagnostics. PLoS Medicine, 8: e1000396.

Van den Eede, P., Van der Auwera, G., Delgado, C., Huyse, T., Soto-Calle, V.E., Gamboa, D., Grande, T., Rodriguez, H., Llanos, A., Anné, J., Erhart, A., D'Alessandro, U. (2010). Multilocus genotyping reveals high heterogeneity and strong local population structure of the Plasmodium vivax population in the Peruvian Amazon. Malaria Journal, 9:151. doi: 10.1186/1475-2875-9-151.

WHO (2007). Malaria elimination. A field manual for low and moderate endemic countries.

WHO (2012a). World Malaria Report. Geneva: World Health Organization.

WHO (2012b). Disease surveillance for malaria control: operational manual.

Wu, C.H. & Drummond, A.J. (2011). Joint inference of microsatellite mutation models, population history and genealogies using transdimensional Markov Chain Monte Carlo. Genetics, 188(1):151-64.

www.cdc.gov/malaria

CHAPTER 2

LOCAL POPULATION STRUCTURE: IMPACT ON MALARIA CONTROL AND ELIMINATION

Regardless of the growing interest in detecting population structures in malarial parasites, there have been limited discussions on how to use this concept in control programs. In such context, the effects of the parasite population structures will depend on interventions' spatial or temporal scales. This investigation explores the problem of identifying genetic markers, in this case microsatellites, to unveil *Plasmodium* genetic structures that could affect decisions in the context of elimination. The study was performed in a low-transmission area, which offers a good proxy to better understand problems associated with surveillance at the final stages of malaria elimination.

Plasmodium vivax samples collected in Tumeremo, Venezuela, between March 2003 and November 2004 were analysed. Since *Plasmodium falciparum* also circulates in many low endemic areas, *P. falciparum* samples from the same locality and time period were included for comparison. *Plasmodium vivax* samples were assayed for an original set of 25 microsatellites and *P. falciparum* samples were assayed for 12 microsatellites.

Not all microsatellite loci offered reliable local data. A complex temporal-cluster dynamics was found in both *P. vivax* and *P. falciparum*. Such dynamics affect the numbers and the type of microsatellites required for identifying individual parasites or parasite clusters when performing cross-sectional studies. The minimum number of microsatellites required to differentiate circulating *P. vivax* clusters differs from the minimum number of hyper-variable microsatellites required to distinguish individuals

within these clusters. Regardless the extended number of microsatellites used in *P. vivax*, it was not possible to separate all individual infections.

Molecular surveillance has great potential; however, it requires preliminary local studies in order to properly interpret the emerging patterns in the context of elimination. Clonal expansions and cluster turnovers need to be accounted for when using molecular markers. These affect the number and type of microsatellite markers, as well as the expected genetic patterns in the context of operational investigations. By considering the local dynamics, elimination programs could cost-effectively use molecular markers. However, population level studies need to consider the local limitations of a given set of loci in terms of providing epidemiologically relevant information.

2.1. Background

There has been growing interest in the population structure of malarial parasites (Imwong et al., 2007; Van den Eede et al., 2010; Pumpaibool et al., 2009; Griffing et al., 2011; Arnott et al., 2012; Chenet et al., 2012; Iwagami et al., 2012). Population structure (deviation from random mating) is a common phenomenon in nature; it is the result of several processes including inbreeding, epidemic population expansions, and geographic isolation (isolation by distance) (Wright, 1943; Meirmans, 2012). In most cases, identifying the population structure of a given infection agent relies on a blind approach that considers “sufficient” gene sampling and a “good” survey of the natural populations (Gauthier and Tibayrenc, 2005), where “good” and “sufficient” depend on the underlying question. However, evaluating the importance of genetic structures for control and

elimination programs requires understanding how they provide information with epidemiologic value. As examples, whereas sometimes the population structure is informative *per se* (e.g. gene flow or separating domestic from imported cases); in other situations, it becomes a problem, e.g. clonal expansions may hamper the ability of separating a new infection from a recrudescence case in a drug efficacy study. Thus, population structures affect the interpretation of patterns emerging from molecular markers in contexts that are relevant for control and elimination programs.

It is worth noting that genetic structures are detectable at the time scales determined by the loci mutation rates. Thus, whereas neutral single nucleotide polymorphisms (SNPs) are widely used for genetic analyses in *Plasmodium falciparum* (Mu et al., 2005; Volkman et al., 2007; Neafsey et al., 2008; Cheeseman et al., 2012) and are currently established in *Plasmodium vivax* (Orjuela-Sánchez et al., 2010), their rates of evolution may identify genetic structures that are not informative for elimination programs in some contexts; e.g. SNPs may not describe the dynamics of haplotypes carrying drug-resistant associated mutations (Griffing et al., 2012). Those more programmatic applications require markers evolving at mutation rates that can identify events at relatively recent (epidemiologically relevant) time scales.

Microsatellites are useful markers under such circumstances, they allow detection of genetic structures at recent divergence times (Payseur et al., 2011) and they are abundant in the *P. falciparum* and *P. vivax* genomes (Russell et al., 2006). In addition, these loci are considered to be selectively neutral, unless they are located near genes under selection

(linked), e.g. genes with mutations conferring drug resistance (Cheeseman et al., 2012; Wootton et al., 2002; McCollum et al., 2008; McCollum et al., 2012). These characteristics make them useful in large-scale population genetic studies (including genome scans looking for mutations under selection) and for answering operational relevant questions such as separating new infections from a recrudescent case (Arnott et al., 2012; Nyachieo et al., 2005; Greenhouse et al., 2006; Orjuela-Sánchez et al., 2009; Restrepo et al., 2011). Even though this has been the case in *P. falciparum*, the production of hypnozoites (homologous or heterologous) in *P. vivax* and the high rate of multiple clone infections represent a more complex dynamics which may alter the results of a study. Moreover, regardless their high mutation rate, epidemiologic uses for microsatellite loci should consider the effect of the population structures due to demographic processes such as clonal expansions (clonal genetic structure) in areas with low transmission (Van den Eede et al., 2010; Pumpaibool et al., 2009; Griffing et al., 2011; Sunnucks, 2000; Mwangi et al., 2006; Schneider et al., 2010; McCollum et al., 2007**)**. Since the population genetic structures are usually unknown *a priori*, this investigation focuses on the importance of pilot studies.

For this study, a *P. vivax* population was characterized in a single low transmission area using isolates collected in two consecutive years. Then, the genetic structure in *P. vivax* was ascertained by using 25 microsatellite loci. Although such an extended number of loci is rarely used in operational research (Imwong et al., 2007; Van den Eede et al., 2010; Pumpaibool et al., 2009; Griffing et al., 2011; Arnott et al., 2012), this panel allowed evaluating the variation per locus in the study area and their reproducibility in

the laboratory. The data obtained were then used to determine the number and properties of loci required to discriminate different epidemiologic scenarios: identification of individual infections (applicable when trying to separate between recrudescence and new infections) and discrimination between clusters of isolates. In order to better understand the dynamics of genetic structures in *P. vivax*, a sympatric *P. falciparum* population sampled during the same time period was included for comparison. Most studies only consider one of the two species (Imwong et al., 2007; Van den Eede et al., 2010; Pumpaibool et al., 2009; Griffing et al., 2011) limiting the comparisons that can be made in areas where both parasites are endemic. Since these two malarial parasite populations come from a low-transmission area, strong linkage disequilibrium was expected (Imwong et al., 2007; Van den Eede et al., 2010; Pumpaibool et al., 2009; Griffing et al., 2011; Schneider et al., 2010; McCollum et al., 2007). However, in the case of *P. falciparum*, the population underwent strong drug selection (McCollum et al., 2007). Thus, if the expansion of a resistant lineage determined the clonal structure (Griffing et al., 2011, McCollum et al., 2007; Schneider and Kim, 2011) of *P. falciparum*, then a more stable cluster structure through time in this parasite was expected (Griffing et al., 2011) compared to the one likely to be observed in *P. vivax* where no drug resistance has been documented.

This investigation found that the number and type of microsatellite loci needed for operational research should be tailored given the objective. Indeed, the number of microsatellite loci required to differentiate circulating *P. vivax* clusters differs from the minimum number of hyper-variable microsatellites required to distinguish individuals

within these clusters. It was also found that the capacity of separating individual infections with a manageable number of microsatellite loci is limited in low transmission areas such as the one under investigation; hence, the use of these loci for separating new infections from recrudescent cases need to take into account such limitations. Both parasites follow similar patterns with alternation of clonal lineages at loci non-liked with drug resistant mutations. Thus, whereas drug resistant haplotypes could be fixed in *P. falciparum* (Griffing et al., 2011, McCollum et al., 2007), cyclical clonal replacements are still taking place. Finally, the inclusion of a temporal sampling scheme in the pilot investigation allowed detecting changes in the genetic structure due to migration or other demographic processes. Overall, this study emphasizes the need to evaluate the local diversity of loci and demography by using a pilot study before designing a molecular epidemiologic investigation.

2.2.Methods

2.2.1. Study area and *Plasmodium* isolates

Two hundred fifteen blood samples with *Plasmodium* monoinfections (107 with *P. vivax* and 108 infected with *P. falciparum*) were used from a surveillance study in Tumeremo (Bolivar State), Venezuela, between March 2003 and November 2004. Any patient positive for malaria by microscopy was invited to provide a sample for further characterization of the parasite and then treated according to national guidelines. The study protocol was approved by the bioethics commission of the *Instituto de Altos Estudios* Dr. Arnoldo Gabaldon in Venezuela.

Malaria transmission in Venezuela has fluctuated since 2000, and peaked in 2004, 2005 and 2007. In 2004, the number of cases in the country rose to almost 42,000, primarily as a result of an increase in *P. vivax* malaria in Bolivar State. The endemic nature of malaria in the State of Bolivar is determined by migration related to gold mining, particularly on the country's border with Guyana. Bolivar state has the greatest number of reported cases of malaria (85% of the nationally reported cases), mainly among miners, agricultural workers and indigenous groups (McCollum et al., 2007). *Plasmodium vivax* is the predominant malaria parasite species in the country followed by *P. falciparum* and *Plasmodium malariae*. *Anopheles darlingi* and *Anopheles marajoara* are the main vectors identified in the study area (Moreno et al., 2007).

2.2.2. Microsatellite analysis

Plasmodium genomic DNA was isolated from whole blood using the QIAamp DNA mini kit (QIAGEN, Valencia, CA). *Plasmodium vivax* samples were assayed for 25 microsatellites (Imwong et al., 2006; Karunaweera et al., 2008) and *P. falciparum* samples were assayed for 12 microsatellites (Anderson et al., 1999). Fluorescently labelled microsatellite PCR products were separated on an Applied Biosystems 3730 capillary sequencer and scored using Gene Marker v1.95 (SoftGenetics LLC). The finding of one or more additional alleles was interpreted as a co-infection with two or more genetically distinct clones in the same isolate (multiple-clone infection, transmitted by one or several mosquitoes). Additional alleles generating peaks of at least one third the height of the predominant allele were also scored. Missing data (no amplifications) are reported by loci but not considered for defining haplotypes.

2.2.3. Population genetic analysis

Microsatellite data was formatted using the Microsatellite tool kit (Park et al., 2001). The heterozygosity estimate (H_E) was used as a measure of overall genetic diversity. This was defined as $H_E = [n/(n-1)][1 - \sum_{i=1}^{L} p_i^2]$, where n is the number of isolates analysed and p_i is the frequency of the i-th allele (i=1, …, L) in the population. H_E gives the average probability that a pair of alleles randomly selected from the population is different. The sampling variance for H_E was calculated as $2(n-1)/n^3[2(n-2)][\sum_{i=1}^{L} p_i^3 - (\sum_{i=1}^{L} p_i^2)^2]$ (Nash et al., 2005). To calculate allele frequencies per locus, samples with mixed infections were included. (If a sample contained, e.g., three microsatellite alleles, each was weight by 1/3 to calculate frequencies.) H_E was calculated using Mathematica 8 (Wolfram Research, Inc.) (code available on request).

To test whether microsatellite haplotypes clustered as a single geographic population, the model-based clustering algorithm implemented in the Structure 2.1 software was applied (Pritchard et al., 2006). This software uses a Bayesian clustering approach to assign isolates to *K* populations or clusters characterized by a set of allele frequencies at each locus. The number of such populations may be either previously known or unknown. In the context of this investigation, this approach allows for the identification of groups or populations of parasites that could be circulating in this area (Tumeremo, Venezuela). The observed genetic diversity was evaluated at different *K* values (*K*=2 to 10). Given that this clustering algorithm incorporates stochastic simulations, each *K* value was run

independently ten times with a burn-in period of 10,000 iterations followed by 50,000 iterations. The admixture model was used in all analysis which allows for the presence of individuals with ancestry in two or more of the *K* populations (Pritchard et al., 2006). Structure harvester v0.6.8 was used to visualize the output from Structure (Dent et al., 2012). To facilitate the interpretation of population-genetic clustering results, CLUMPP (Cluster Matching and Permutation Program) was also used. CLUMPP strips away the 'label switching' heterogeneity so that the 'genuine multimodality' can be detected and quantified (Jakobsson et al., 2007). In addition, *distruct* 1.1 was used to graphically display the clustering results (Rosenberg et al., 2004). The posterior probability for each number of populations or clusters (*K*) is computed and then the *K* value that better explains the genetic data is an estimate of the number of circulating clusters or populations circulating. Whereas such genetic structures are a deviation from the expectations under one population undergoing random mating, each cluster cannot be considered a random mating population on its own since such clusters could represent clonal lineages. Moreover, Fstat 2.9.3.2 (Goudet et al., 1995) was used to calculate F_{st} between clusters.

Evidence of linkage disequilibrium between alleles from different loci in parasite populations was analysed with Arlequin 3.11 (Laval et al., 2005). A standardized index of association (I^S_A) was also used to test for evidence of overall multilocus linkage disequilibrium in the Venezuelan population. This test compares the variance (V_D) of the number of alleles shared between all pairs of haplotypes observed in the population (*D*) with the variance expected under random association of alleles (V_E) as follows: $I^S_A = (V_D$

/ $V_E - 1$) $(r - 1)$, where r is the number of loci analysed. V_E is derived from 10,000 simulated data sets in which alleles were randomly reshuffled among haplotypes. Significant linkage disequilibrium is detected if V_D is greater than 95% of the values derived from the reshuffled data sets. Data were analysed with LIAN 3.1 (Haubold et al., 2000). Pairwise linkage disequilibria (LD) were also calculated. In particular, R (Maruyama et al., 1982) was calculated. For each pairwise comparison only those samples were included in which data were not missing and only one allele occurred at one of the two loci. If multiple alleles for a locus were found in a sample, e.g., four alleles, the respective haplotypes were weighted by ¼ to calculate haplotype frequencies. Allele frequencies were calculated as described above (p_i) from all included samples. Calculations were implemented in Mathematica 8 (Wolfram Research, Inc.).

Additionally, deviations from pairwise linkage-equilibrium were explored by using Freeman-Halton's test (Freeman et al., 1951) (generalized Fisher's exact test for kxm contingency tables) at a significance level of 5%. Of particular note, contingency tables, resulting from pairwise comparison of microsatellite loci, typically have sparse entries, and in many cells the count is less than 5. This renders asymptotic tests (e.g. G-test) inappropriate. Since exact tests require integer-valued cell counts, for each pair of microsatellite loci, only samples that had exactly one microsatellite allele at both loci could be included in the analysis. (Therefore, for some pairwise comparisons, fewer samples were included than for calculating R.) As the number of tests performed and size of resulting contingency tables becomes numerically challenging, Monte-Carlo estimates for exact p-values were calculated using 10,000 random iterations. However, for 2x2

tables Fisher's exact test was always performed. Computations were performed with SAS® (Version 9.2).

2.2.4. Statistical analysis

To test the hypothesis that the probability of sampling parasites from different clusters differs over time, the Kruskal-Wallis test was performed with the sampling date as the dependent variable and the clusters as the independent variable. The null hypothesis was that the probability of sampling from each cluster was the same during the two years. To test if the chronological sequence of clusters would not appear randomly, the asymptotic test of Barton and David (Barton et al., 1957) was performed ("generalization" of the Wald-Wolfowitz runs test). In addition, associations of time and cluster-membership were further explored. For this purpose, cluster-membership was arranged as a nominal variable, and sample date - grouped in monthly ranges - as an ordinal variable in a k x l contingency table. The χ^2 and likelihood-χ^2 tests were used. The Pearson's $\emptyset$-coefficient, Pearson's contingency coefficient (C_p), and Cramer's V were also obtained. All statistical tests were performed using SAS® (Version 9.2).

2.3. Results

Twelve *P. falciparum* imperfect microsatellites reported by Anderson *et al.* (Anderson et al., 1999), were tested for their ability to amplify in this population and to demonstrate polymorphism on a set of Venezuelan field isolates from 2003–2004 (Table 1). Suitable markers were considered as those loci that amplified at least 80% of the samples and showed evidence of polymorphism with a clear peak pattern. Microsatellites TA40 and

TA87 did not meet the criteria and were not considered in the analysis, but loci TA1, Polyα, TA60, ARA2, Pfg377, PfPK2, TA109, TA81, TA42 and 2490 were included for the population genetic analysis. These markers are distributed across seven chromosomes and no two loci are physically closely linked. Markers TA1 and TA109 are 14.4 cM apart on chromosome 6; markers Pfg377 and PfPk2 are on chromosome 12 and separated by a distance of 45.8 cM; and markers TA81 and TA42 are 68.9 cM apart on chromosome 5. All loci had repeat units of 3bp or more.

For *P. vivax*, a panel of 11 dinucleotide and 14 markers of 3bp or more was tested. A total of 22 microsatellites that consistently amplified the samples and showed polymorphism were selected (Table 2). These markers were distributed in 13 of the 14 chromosomes and linkage disequilibrium has been previously reported between microsatellites MS4-MS5 in chromosome 6 and MS7-MS8 in chromosome 12.

Table 1. Characterization of the *P. falciparum* microsatellite loci in all samples analyzed from Venezuela

Locus	Chr	Size range	N. of Alleles	HE	SD
POLYa	4	160-187	4	0.44	0.06
TA81	5	118-121	2	0.16	0.04
TA42	5	185-202	3	0.59	0.03
TA1	6	169-181	3	0.66	0.02
TA109	6	162-177	3	0.58	0.03
2490	10	81-84	2	0.39	0.04
ARA2	11	62-65	2	0.11	0.04
Pfg377	12	92-95	2	0.31	0.05
PfPK2	12	162-165	3	0.62	0.03
TA60	13	70-79	2	0.43	0.04

2.3.1. Multiple infections

In *P. falciparum*, 2.8% of the 108 samples had multiple infections while in *P. vivax* 15.9% out of 107 samples were also found with multiple infections harbouring two or three alleles at two or more loci. When grouping samples according to their date of collection, a greater proportion of mixed clone infections were found in *P. vivax* samples from 2004 (at the peak of malaria transmission in Venezuela), with more than double the amount of mixed infections found in *P. vivax* samples from 2003.

2.3.2. Genetic diversity

The frequency distribution of alleles at each locus was determined using all samples; however, three samples each were excluded in *P. falciparum* and *P. vivax* since they only amplified two or no loci. Most of the *P. vivax* loci but MS3, MS7 and MS10 followed (almost) a normal distribution (histogram of microsatellite repeats), which indicates an appropriate sampling of the alleles present in Tumeremo. However, many of the *P. vivax* markers showed a considerable number of (rare) alleles or also referred as minor allele frequency (MAF) (frequency lower than 5%, see Figure 1 A and B, Table 2). In *P. falciparum*, the same trend was not observed since only a small number of alleles per loci were found suggesting low genotypic diversity due to recent strong selection by drug pressure and low recombination (McCollum et al., 2007). Furthermore, note that *P. vivax* relapses from hypnozoites occur, which increases the potential to harbor higher genetic diversity as compared with *P. falciparum* if patients do not comply with the primaquine treatment and as a result it hampers the ability to differentiate between reinfection and recrudesce cases. The expected heterozygosity (H_E) was calculated in both species using all microsatellites (Figure 1 C and Figure 1 D). *P.vivax* was found to be highly polymorphic (H_E = 0.73 ± 0.032) in comparison to *P. falciparum* (H_E = 0.44 ± 0.035). Moreover, the year by year analysis (Additional file 1) revealed higher levels of genetic diversity in samples from 2004 (H_E = 0.79 ± 0.0221) than from 2003 (H_E = 0.67 ± 0.0301). In *P. falciparum*, the level of genetic diversity in 2003 (H_E = 0.51 ± 0.0378) was higher than in 2004 (H_E = 0.39 ± 0.0641) and all mixed samples were found during 2003.

Figure 1. Genetic diversity in Microsatellite loci. A) Number of alleles per loci and B) percentage of alleles in low frequency (<5%) in *P. vivax*. Expected heterozygosity (H_E) in C) *P. falciparum* and in D) *P. vivax* using all samples from Tumeremo

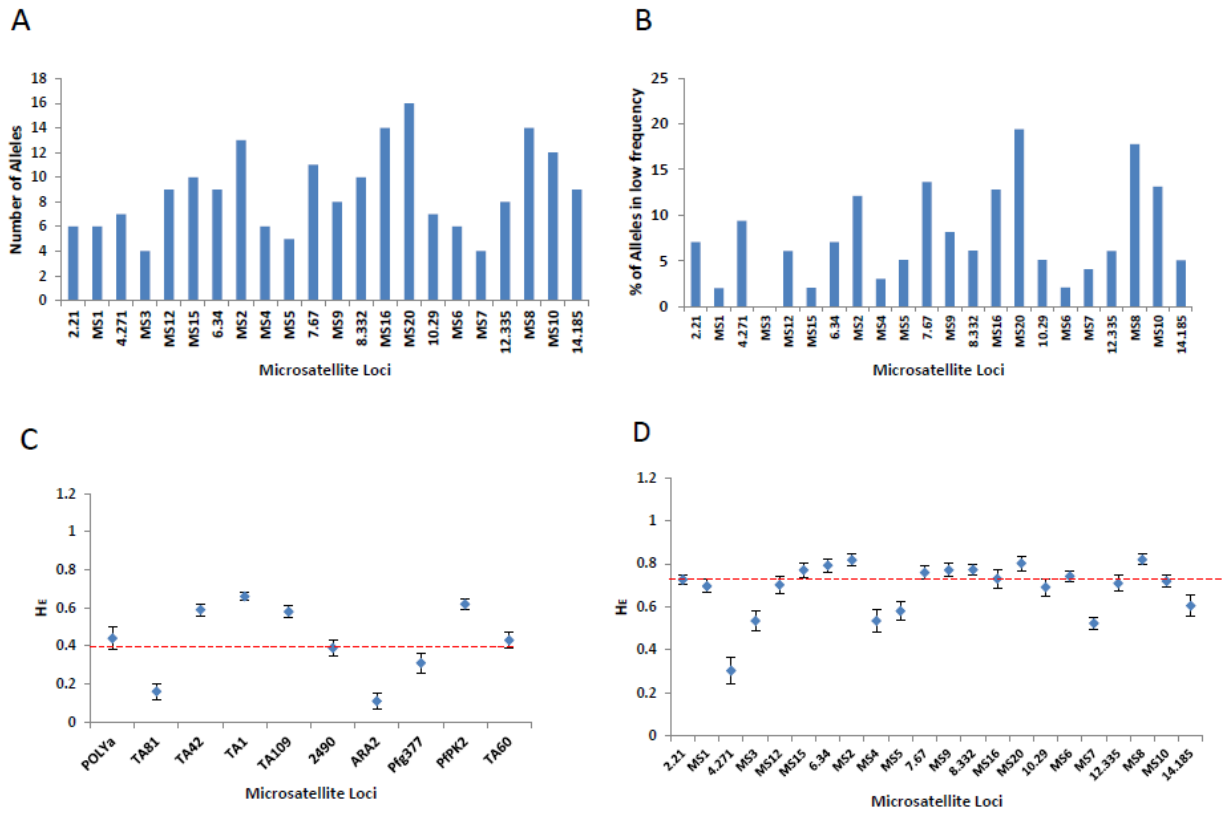


Table 2. Characterization of the *P. vivax* microsatellite loci in all samples analyzed from Venezuela

Locus	Chr	Core sequence*	Size range	N. of Alleles	%MAF	HE	SD
Dinucleotides							
2.21	2	AC	92-104	6	7.07	0.73	0.02
4.271	4	AT	86-122	7	9.38	0.30	0.06
6.34	6	AC	138-158	9	7.07	0.79	0.03
7.67	7	AT	98-124	11	13.64	0.76	0.03
8.332	8	AT	216-260	10	6.12	0.77	0.03
10.29	10	AT	110-130	7	5.10	0.69	0.04
12.335	12	AT	156-180	8	6.06	0.71	0.04
14.185	14	AT	266-289	9	5.05	0.61	0.05
Trinucleotides							
MS1	3	$(GAA)_{11}$	231-246	6	2.02	0.70	0.03
MS3	4	$(GAA)_{11}$	184-193	4	0	0.54	0.05
MS15	5	$(TCT)_{10}$	236-284	10	2.04	0.77	0.03
MS4	6	$(AGT)_{18}$	189-210	6	3.03	0.53	0.06
MS9	8	$(GGA)_{18}$	155-176	8	8.16	0.77	0.03
MS7	12	$(GAA)_{9}$	145-157	4	4.08	0.52	0.03
Imperfect							
MS12	5	$(TTC)_{10}(TGC)_{4}$	185-269	9	6.06	0.70	0.04
MS2	6	$(TAAA)_{2}TATA(TAAA)_{6}TATA(TAAA)_{19}$	181-285	13	12.12	0.82	0.03
MS5	6	$CCTCTT(CCT)_{11}$	172-187	5	5.10	0.58	0.04
MS16	9	$(ACA)_{9}GCA(ACA)_{3}GCA(ACA)_{7}GCA(ACA)_{3}GCAATC$	238-400	14	12.79	0.73	0.04
MS20	10	$(GAA)_{11}GAG(GAA)_{13}(CAA)_{4}$ $GAA(CAA)_{5}$	193-259	16	19.39	0.80	0.03
MS6	11	$(TCC)_{2}(TCT)_{3}(CCT)_{2}(TCC)_{2}$ $GCTTCT(TCC)_{10}$	211-244	6	2.08	0.74	0.03
MS8	12	$(CAG)_{2}(CAA)_{11}$	201-285	14	17.78	0.82	0.02
MS10	13	$GAA(GGA)_{2}AGA(GGA)_{9}AGA(GGA)_{4}AGAGGAAGA(GGA_{3})$	189-249	12	13.13	0.72	0.03

*Descibed by Imwong et al., 2006 and Karunaweera et al., 2008

2.3.3. Population structure

The population structures of *P. falciparum* and *P. vivax* were determined using the Structure 2.1 software. Single infections and mixed infections with only one polymorphic locus were included in the analysis with complete data or missing data in no more than three loci. The posterior probability for each cluster (*K*) was computed and the clustering patterns obtained with $K = 4$ was associated with the highest Delta *K* values (Delta K = mean (|L''(*K*)|)/sd(L(*K*))) for both species (Figure 2) (Dent et al., 2012). Most parasites had their origin clearly assigned to a single cluster; few of them were fairly admixed, showed by bars partitioned into *K* coloured segments. Each fragment in the bar represents the estimated membership fractions of each isolate or strain in *K* clusters. However, the algorithm implemented in Structure might detect K-1 reasonable clusters, and subsume remaining low frequency variants in the K^{th} cluster. This might be indicated by higher genetic diversity in just one cluster. In each of the four *P. falciparum* clusters, similar numbers of samples were found (Table 3a); however, in *P. vivax* most of the samples were grouped either in the clusters labeled as "A" or "D" (Table 3b). In addition, when grouping *Plasmodium* haplotypes by year, *P. falciparum* samples from either 2003 or 2004 were also grouped in four clusters. Still, in 2003 haplotypes from *P. falciparum* cluster C were overrepresented while in 2004 haplotypes from falciparum cluster A were more abundant. Clusters B and D were present in almost the same frequency during both years. On the other hand, *P. vivax* samples from 2003 did group in four clusters whereas samples from 2004 were grouped in only three clusters; nevertheless, these results might be biased due to the lower number of 2004 samples included in the analysis. Due to the

high prevalence of multi-clonal infections (13 out of 41) in the 2004 *P. vivax* samples, it was difficult to determine the total number of haplotypes present in the sample so an estimate using the single infections was considered. From this description, it seems that the probability of sampling parasites from different clusters differs over time. The Kruskal-Wallis test was performed and showed to be significant for *P. falciparum* and *P. vivax* (p-values <0.01), indicating that the probabilities of sampling a parasite belonging to a given cluster changes over time. This is an indication of clonal expansions.

Figure 2. Population structure of A) *P. falciparum* (cluster A – pink, cluster B – yellow, cluster C – blue and cluster D – orange) and B) *P. vivax* (cluster A – orange, cluster B – yellow, cluster C – blue and cluster D – pink) inferred from microsatellite typing of Tumeremo haplotypes from 2003 and 2004 using STRUCTURE.

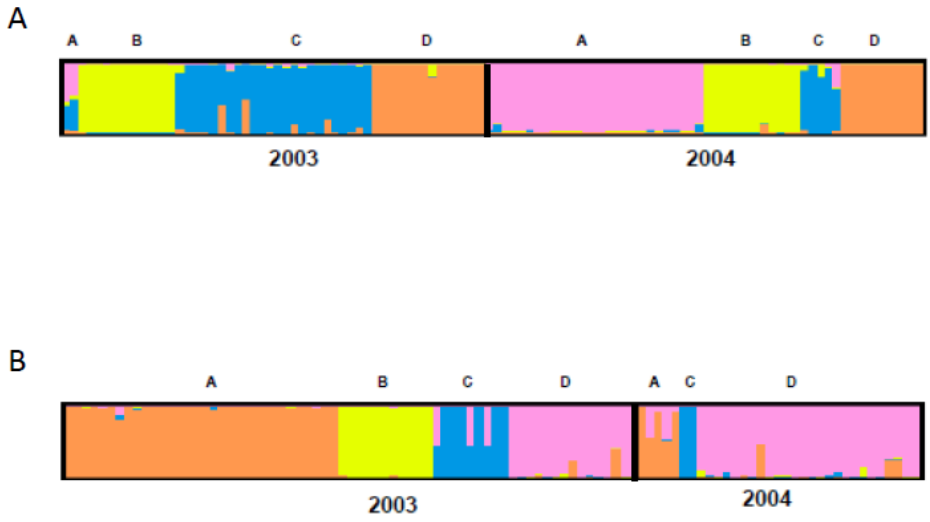


Table 3. Genetic diversity by clusters

a).P. falciparum					
Cluster	Samples	Haplotypes	N of Alleles	HE	SD
A	29	3	1.4	0.12	0.07
B	24	3	1.3	0.07	0.05
C	29	13	2.3	0.38	0.05
D	24	2	1.1	0.01	0.01
b). P. vivax					
A	37	15	2.5	0.15	0.04
B	11	7	1.4	0.11	0.04
C	11	4	1.8	0.32	0.05
D	41	37	7.9	0.77	0.03

If parasite prevalence is due to events of clonal expansion, different clusters would be predominant at different time periods. Hence, the chronological sequence of clusters would not appear randomly. Since some samples were taken on the same day, the chronological order on the sample labels was followed. The Barton-David test was significant ($p<10^{-11}$ for *P. falciparum*, $p<10^{-7}$ for *P. vivax*) indicating that cluster-membership is not random over time, which is another indication for clonal expansion.

Furthermore, associations of time and cluster-membership were explored. The X^2 and likelihood-X^2-tests were both significant on an alpha level of 5% (all p-values <0.01). For *P. vivax*, $\phi = 0.1790, \times C_p = 0.5838$, $C_p = 0.5838$, and $V = 0.4151$, were obtained, indicating an association between time and cluster-membership. Likewise, for *P. falciparum*, $\phi = 0.8043$, $C_p = 0.6267$; and $V = 0.4643$ were obtained. However, it should be mentioned that the 85% of the cells in the contingency had expected counts less than 5, which renders the application of X^2 tests and X^2-based association measures problematic. It should be mentioned that the sample size was too large to perform Fisher's exact test, or rather the generalization of Freeman and Halton (Freeman et al., 1951). The statistics indicate that there is some association between time and cluster membership; however, this association is not very strong. Notably, the association appears stronger in *P. falciparum* than in *P. vivax*, which can be explained by the hypothesis of clonal expansion due to drug resistance. It may be possible that due to relapses in *P. vivax* the pattern of clonal expansion is distorted; however, data on the follow-up primaquine treatments are not available.

Using only six (Polya, TA60, ARA2, Pfg377, TA81 and 2490) of the 10 *P. falciparum* loci is enough to differentiate the four clusters with a 98% confidence interval. In addition, at least eight loci (TA1, Polya, TA60, Pfg377, PfPK2, TA109, TA42 and 2490) are needed to correctly identify the 41 unique haplotypes found in the 104 *P. falciparum* isolates determined with all ten loci. Thus, the number of loci used does not allow discriminating all individuals.

Plasmodium vivax samples were also reanalysed in order to obtain the four clusters and the different identifiable haplotypes from the 94 *P. vivax* samples (100 clones in total within single infections or mixed infections in one locus only). The set of loci used to identify *P. vivax* clusters (four representative loci from the population, which harbours average values of H_E and percentage of alleles in low frequency, and two highly polymorphic loci to differentiate maximum number of individuals within cluster) were: locus 2.21, locus 12.335, locus MS1, locus MS6, locus MS8, and locus MS16, whereas the set of loci needed to differentiate the maximum number of haplotypes were: locus 6.34, locus 7.67, locus 8.332, locus MS2, locus MS8, locus MS9, locus MS10, locus MS15, locus MS16 and locus MS20. The F_{st}values between clusters in *P. falciparum* and *P. vivax* populations were also calculated to determine if there was strong genetic differentiation within these subpopulations. The values were significant and ranged between 0.4 to 0.9 in *P. falciparum* and 0.3 to 0.8 in *P. vivax*. These results confirm that there is a strong genetic differentiation within these subpopulations (Table 4).

Table 4. Microsatellite-based genetic differentiation (Fst) between *Plasmodium* clusters (P-value<0.05).

a). P.falciparum			
Clusters	A	B	C
B	0.68		
C	0.48	0.67	
D	0.79	0.91	0.66
b). P.vivax			
B	0.85		
C	0.79	0.76	
D	0.42	0.36	0.28

2.3.4. Linkage disequilibrium

In order to calculate linkage disequilibrium, the data were organized to follow the stepwise mutation model when possible. Significant multilocus linkage disequilibrium was found in *P. falciparum* and *P. vivax* isolates according to the standard index of association ($I^S_{A2003} = 0.1236$, $p<0.01$; $I^S_{A2004} = 0.172$, $p<0.01$ and $I^S_{A2003} = 0.3776$, $p<0.01$; $I^S_{A2004} = 0.1448$, $p<0.01$, respectively) and to pairwise comparisons using Arlequin 3.11. Linkage disequilibrium was also calculated considering the time of collection of the isolates, with unique haplotypes and excluding microsatellites in the same chromosome. In all analyses performed significant linkage disequilibrium was found. Further analysis to calculate R in *P. falciparum* (Figures 3, A and B) and *P. vivax* (Figure 4, A and B)

conducted between pairs of microsatellites and using all information, including mixed infections, also revealed linkage disequilibrium in both species. Additionally, we tested for pairwise linkage disequilibrium using Freeman-Halton's ("Fisher's exact") test (Figures 3, C and D and Figure 4, C and D). To rule out the possibility that association resulted from combining highly different subpopulations in *P. vivax*, linkage disequilibrium was also calculated using pairwise analysis within the genetic clusters present in both years and a significant result was found for cluster D (I^S_A = 0.1136, p-value <0.001). Furthermore, significant LD (I^S_A = 0.172, p-value <0.05) was also found in *P. vivax* cluster B, which was only present in 2003.

Figure 3. R values of linkage disequilibrium in *P. falciparum* samples from A) 2003 and B) 2004 and the respective outcomes of Freeman-Halton's test for significant deviations from linkage equilibrium (C and D) at a significance level of 0.05.

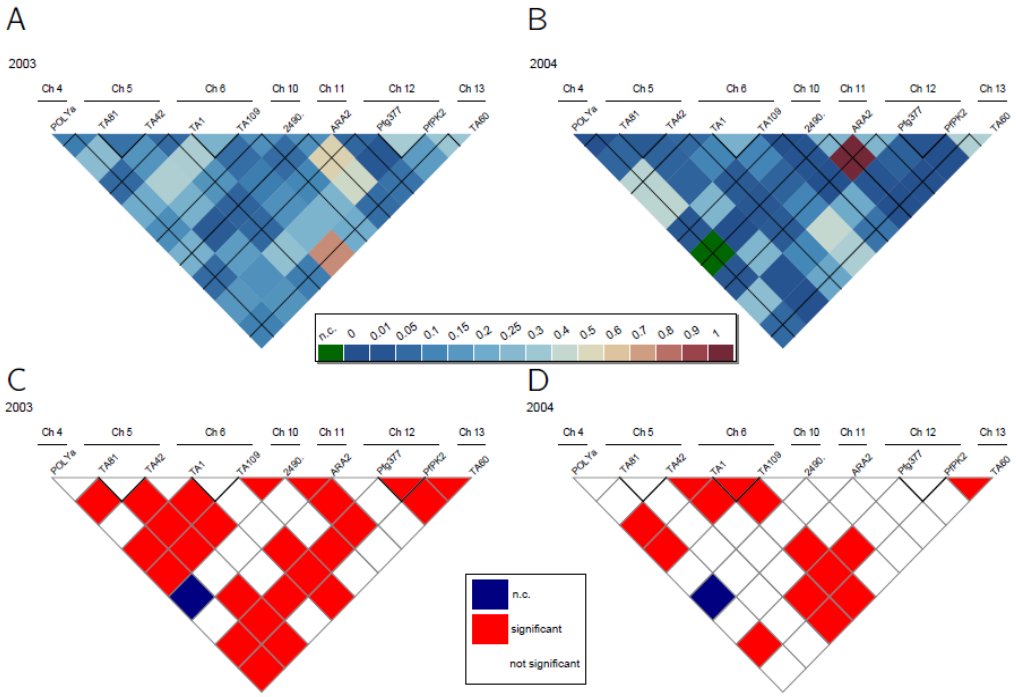
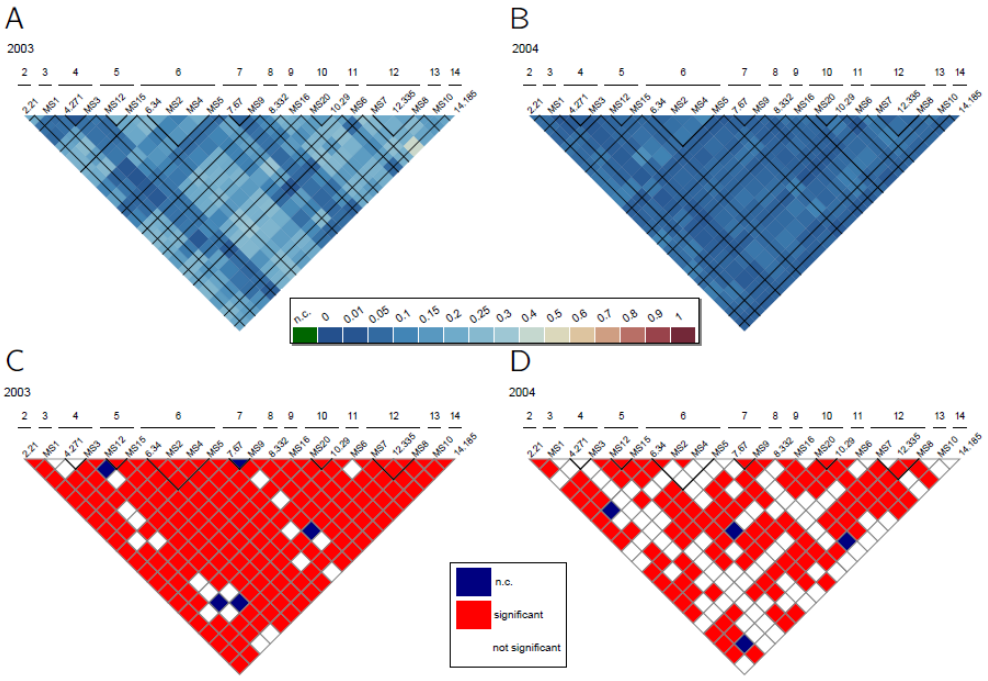


Figure 4. R values of linkage disequilibrium in *P. vivax* samples from A) 2003 and B) 2004 the respective outcomes of Freeman-Halton's test for significant deviations from linkage equilibrium (C and D) at a significance level of 0.05.



2.4. Discussion

Low-transmission areas are of great interest because they are potential targets for malaria elimination. Even further, they allow for testing how to deploy resources effectively at the final stages of the elimination process. Under such conditions, the precise monitoring of malaria infections is indispensable. Whereas molecular marker-based approaches could enrich malaria surveillance (Arnott et al., 2012), it needs to be ascertained whether such methods - besides their general popularity - are cost-effective for detecting substructures in parasite populations that are valuable for elimination programs, especially in these low-transmission settings. One step in the process of translating population-genetic concepts into epidemiology is to determine whether an available set of

markers has the sensitivity to detect patterns that are epidemiologically informative, especially when performing cross-sectional studies in a narrow period of time that simply allow to see a predominant clonal expansion. Whereas, hypothetically, the number of markers could be increased up to levels that can rule out many confounding factors; in reality, malaria elimination programs face limited funds, samples, and human resources that can be invested in molecular surveillance.

In general, highly polymorphic loci are suitable to identify *Plasmodium* individuals within populations whereas more conserved (less polymorphic) loci are useful to establish clusters or parasite populations within a given endemic area. For the first purpose, imperfect microsatellites would be more appropriate due to the complexity on their core sequence which results in greater allelic diversity, while dinucleotides and trinucleotides are more conserved and best represent the local population diversity. However, there are always exceptions in each of the categories established. Distinctions between loci have practical consequences. Highly polymorphic markers have many alleles in low frequency that are more likely to fluctuate by random events (e.g. random genetic drift but also inadequate sampling); thus, a local genotype could not be detected at a given time and then "appear" as new (e.g. being identified as an introduction) after deploying control interventions. Those cases could be considered "imported" rather than the result of residual transmission, misleading the program to conclude that the control strategy was more efficacious than in reality. This problem is particularly important if the only baseline information is provided by a cross-sectional study that simply detected a predominant clonal expansion. On the contrary, the number of treatment "failures"

(putative recrudescent cases) could be over-estimated and conclude that a drug treatment has reduced its efficacy by using low polymorphic marker that cannot properly differentiate genotypes within a cluster. Thus, in order to get a fine scale fingerprint of the epidemiological relevant events targeted by a given molecular surveillance program, the number and kind of microsatellites used should consider the characteristics of the population under study. Those characteristics include the replacement of clusters at least between two transmission seasons.

This investigation provides some basic information about how to design a pilot (baseline) investigation that could support prospective molecular population-based studies, especially in areas with low genetic diversity and low transmission as the one found in Tumeremo, Venezuela. Those steps are summarized in a flow chart (Figure 5). First, an extended set of loci should be evaluated in order to determine how many reliable work (e.g. can be consistently amplified or the pattern in the electropherogram is unambiguously interpretable) in that area. Then, based on that set of loci, the number of mixed infections, haplotypes and clusters circulating in a given geographic area during at least two transmission seasons can be evaluated. The assumption is that the information collected from the original set of loci (in this case, 22 loci in *P. vivax*) provides the best possible resolution given the available resources. Based on the identified structures using the extended set of loci, it is possible to estimate the minimum number of microsatellite loci needed to detect such structure and differentiate *Plasmodium* populations (clusters) in that specific geographic area. The same baseline data can be used to estimate the minimum number of microsatellites needed to differentiate individuals within these

clusters. Hypothetically, since each of the samples analyzed originated from a distinct individual, many confounding effects could be reduced by simply increasing the number of markers. However, the number of loci that can be used in a given study is finite, simply because they are tied to the available resources. Thus, pilot studies are needed in order to determine the best resolution that can be obtained using molecular markers in a given context.

Considering all data analysed for *P. falciparum* and *P. vivax*, only 45% of the *P. falciparum* and 85% of the *P. vivax* samples at the individual level were correctly identified. In order to increase the percentage of individual identification for *P. falciparum* in a place like Tumeremo that underwent a strong drug selective pressure, a greater number of polymorphic loci are needed since the heterozygosity per locus is much lower compared with *P. vivax*. This represents a problem since the analysis of more than 10 loci per sample could be costly. It is also important to emphasize that there are clusters of individuals infected by highly related parasites (clonal expansions) and those simply cannot be easily separated. Thus, a manageable number of microsatellites will only identify haplotypes or clonal lineages per population, some of them stable through time (Griffing et al., 2011; Urdaneta et al., 2001; Razakandrainibe et al., 2005). Population-based studies should be designed considering such restrictions by targeting groups (e.g. arms in a drug efficacy study) or geographic areas where the turnover of such clones (or their persistence) could be informative at the population level. In the case of drug efficacy studies, the proportions of undifferentiated genotypes from prior and post treatment samples can be compared between groups providing a way to reduce the

confounding effect of the local parasite demography. Unfortunately, such an approach may require high sample sizes.

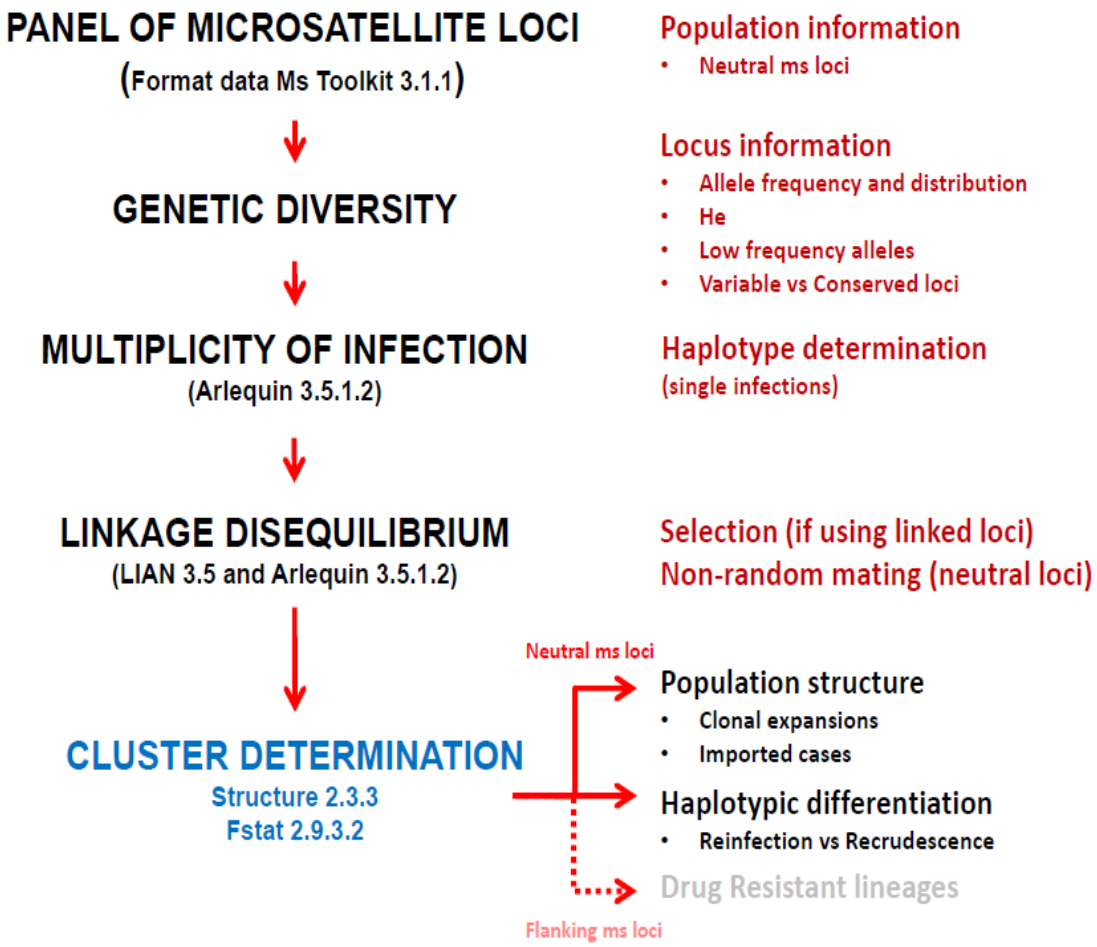
Plasmodium vivax and *P. falciparum* circulating in Tumeremo differ according to the prevalence of mixed clone infections and the overall levels of genetic diversity. Since in *P. falciparum* drug resistant haplotypes are fixed, it was expected that such early drug pressure will result in few clones present year-round (Griffing et al., 2011). However, both species exhibited significant linkage disequilibrium (LD), similar number of clusters/clones when considering haplotypes only, and a similar pattern of temporal replacement of clusters. In *P. falciparum* the four clusters found in 2003 were also found the following year. In *P. vivax*, however, three of the four clusters persisted through time. This could indicate a sporadic event in 2003 of a parasite lineage (cluster B) coming into the Tumeremo population from one of the shared borders with Guyana and Brazil. Moreover, significant LD was found in *P. vivax* within cluster B, which strengthens the idea of a group of minor frequency lineages that are not stable in time and that could be the result of imported cases. Detecting introductions is important especially in the context of control and elimination of the disease; those migration events could explain a local increase in incidence indicating that the deployed strategies/policies were effective in controlling the previously circulating parasites. Ideally, data from other populations should allow identifying the source of the migrant cases providing additional information about how malaria is maintained at a regional level.

A higher level of LD in *P. vivax* compared to *P. falciparum* was not expected since this could not be accounted by recent epidemic expansions of drug resistant lineages like in *P.*

falciparum (Griffing et al., 2011; McCollum et al., 2007). Thus, the pattern observed in *P. vivax* is indicative of ongoing reduced levels of recombination due to the parasite demography. Overall, the local ecology appears to explain the turnover of clones/clusters in both parasites at this time scale. The temporal interval used in this study is appropriate in the context of surveillance when monitoring malaria prevalence from one malaria season to the next one. Plus, the effectiveness of local malaria control can be evaluated by observing changes (which could be the result of incoming gene flow) in the parasite population structure.

It is worth noting that in *P. vivax*, also cluster D had a significant level of LD among lineages within that cluster. When explored closely, this cluster D included different lineages found in low frequency that are stable through time. This cluster D then contrasts with cluster B; where the first seems to be stable in the population and composed by several minor haplotypes, the later (B) may indicate migration/introduction of haplotypes in lower frequency into the endemic area. Thus, a close observation of the linkage disequilibrium within clusters provides additional information that could explain the observed patterns of malaria transmission.

Figure 5. Flow chart illustrating the steps to follow in a *Plasmodium* population genetic study. The goal is to identify appropriate microsatellite markers for haplotypic determination (to differentiate between reinfection and recrudescence cases) and to determine clusters or subpopulations (to evaluate clonal expansions and possible imported cases).



2.5. Conclusions

Molecular epidemiological studies should consider temporal heterogeneities in the parasite population structures. In the context of this investigation, clonal expansions in

South America affect estimates of the local genetic diversity. Such dynamic will generate a distorted view if studies are based in a sample collected in a single time point. A second observation is that, with a reasonable number of loci, there is uncertainty in the separation of infections at individual level. Such uncertainties need to be incorporated whenever recrudescent cases need to be separated from new infections in the context of drug efficacy studies. Third, it is relatively easy to identify major clones or sub-populations with a small number of polymorphic microsatellites. Such information could be valuable to better understand gene flow and patterns of migration/re-introduction. It seems logical to contrast samples from at least two transmission seasons, even in gene flow studies, and use loci that allow detecting spatial connectivity so they are useful to track reintroductions at a local level or spatial movements at a regional level. Finally, it was observed that both *P. falciparum* and *P. vivax* exhibit similar patterns of clonal expansions. Whereas previous studies have found strong linkage disequilibrium in both parasites (Imwong et al., 2007; Van den Eede et al., 2010; Pumpaibool et al., 2009; Griffing et al., 2011; Arnott et al; Iwagami et al., 2012), this is the first time that such dynamics are described in sympatric populations through time. Overall, it seems that such clonal temporal replacements take place even when drug resistant mutations could be fixed, as is the case for *P. falciparum* (Griffing et al., 2011; McCollum et al., 2007).

Even though only a limited number of loci could be considered for routine analysis, the use of well-chosen microsatellite loci represent a highly sensitive method for cluster differentiation and provides a way to deeper analyze patterns of gene flow as well as parasite lineages maintained through time. Overall, this investigation highlights the need

of locally evaluating the diversity of microsatellite loci during at least two transmission seasons before starting molecular epidemiologic investigations.

2.6. References

Anderson, T.J., Su, X.Z., Bockarie, M., Lagog, M., Day, K.P. (1999). Twelve microsatellite markers for characterization of Plasmodium falciparum from finger-prick blood samples. Parasitology, 119:113–125.

Arnott, A., Barry, A.E., Reeder, J.C. (2012). Understanding the population genetics of Plasmodium vivax is essential for malaria control and elimination. Malaria Journal, 10:11.

Barton, D.E., David, F. (1957). Multiple runs. Biometrika, 44:168–170.

Cheeseman, I.H., Miller, B.A., Nair, S., Nkhoma, S., Tan, A., Tan, J.C., Al Saai, S., Phyo, A.P., Moo, C.L., Lwin, K.M., McGready, R., Ashley, E., Imwong, M., Stepniewska, K., Yi, P., Dondorp, A.M., Mayxay, M., Newton, P.N., White, N.J., Nosten, F., Ferdig, M.T., Anderson, T.J. (2012). A major genome region underlying artemisinin resistance in malaria. Science, 336:79–82.

Chenet, S.M., Tapia, L.L., Escalante, A.A., Durand, S., Lucas, C., Bacon, D.J. (2012). Genetic diversity and population structure of genes encoding vaccine candidate antigens of Plasmodium vivax. Malaria Journal, 11:68.

Earl, D.A., von Holdt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources, 4:359–361. Structure harvester available from:http://taylor0.biology.ucla.edu/struct_harvest/

Ferreira MU, Karunaweera ND, da Silva-Nunes M, da Silva NS, Wirth DF, Hartl DL. (2007). Population structure and transmission dynamics of Plasmodium vivax in rural Amazonia. Journal of Infectious Diseases, 195:1218–1226.

Freeman, G.H., Halton, J.H. (1951). Note on an Exact Treatment of Contingency, Goodness of Fit and Other Problems of Significance. Biometrika, 38:141–149.

Gauthier, C., Tibayrenc, M. (2005). Population structure of malaria parasites: the driving epidemiological forces. Acta Tropica, 3:241–250.

Goudet, J. (1995). FSTAT Version 1.2: a computer program to calculate F-statistics. J. Heredity, 86:485–486. FSTAT 2.9.3.2 available at: http://www2.unil.ch/popgen/softwares/fstat.htm.

Greenhouse, B., Myrick, A., Dokomajilar, C., Woo, J.M., Carlson, E.J., Rosenthal, P.J., Dorsey, G. (2006). Validation of microsatellite markers for use in genotyping

polyclonal Plasmodium falciparum infections. American Journal of Tropical Medicine & Hygiene, 75:836–842.

Griffing, S.M., Mixson-Hayden, T., Sridaran, S., Alam, M.T., McCollum, A.M., Cabezas, C., Marquiño Quezada, W,, Barnwell, J.W., De Oliveira, A.M., Lucas, C., Arrospide, N., Escalante, A.A., Bacon, D.J., Udhayakumar, V. (2011). South American Plasmodium falciparum after the malaria eradication era: clonal population expansion and survival of the fittest hybrids. PLoS One, 6:e23486.

Haubold, B., Hudson, R.R. (2000). LIAN 3.0: detecting linkage disequilibrium in multilocus data. Bioinformatics, 16:847–848. LIAN 3.5 available from: http://adenine.biz.fh-weihenstephan.de/cgi-bin/lian/lian.cgi.pl.

Imwong, M., Nair, S., Pukrittayakamee, S., Sudimack, D., Williams, J.T., Mayxay, M., Newton, P.N., Kim, J.R., Nandy, A., Osorio, L., Carlton, J.M., White, N.J., Day, N.P., Anderson, T.J. (2007). Contrasting genetic structure in Plasmodium vivax populations from Asia and South America. International Journal for Parasitology, 37:1013–1022.

Imwong, M., Sudimack, D., Pukrittayakamee, S., Osorio, L., Carlton, J.M., Day, N.P., White, N.J., Anderson, T.J. (2006). Microsatellite variation, repeat array length, and population history of Plasmodium vivax. Molecular Biology and Evolution, 23:1016–1018.

Iwagami, M., Fukumoto, M., Hwang, S.Y., Kim, S.H., Kho, W.G., Kano, S. (2012). Population structure and transmission dynamics of Plasmodium vivax in the Republic of Korea based on microsatellite DNA analysis. PLoS Neglected Tropical Diseases, 6:e1592.

Jakobsson, M., Rosenberg, N.A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics, 23:1801–1806. CLUMPP 1.1.2 available from: http://www.stanford.edu/group/rosenberglab/clumpp.html.

Karunaweera, N.D., Ferreira, M.U., Munasinghe, A., Barnwell, J.W., Collins, W.E., King, C.L., Kawamoto, F., Hartl, D.L., Wirth, D.F. (2008). Extensive microsatellite diversity in the human malaria parasite Plasmodium vivax. Gene, 410:105–112.

Laval, E.G., Schneider S. (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evolutionary Bioinformatics, 1:47–50.
Maruyama, T. In: Molecular Evolution, Protein Polymorphism and the Neutral Theory. Kimura, M., editor. Tokyo: Japan Scientific Societies Press; 1982. Stochastic integrals and their application to population genetics, 151–166.

McCollum, A.M., Basco, L.K., Tahar, R., Udhayakumar, V., Escalante, A.A. (2008). Hitchhiking and selective sweeps of Plasmodium falciparum sulfadoxine and pyrimethamine resistance alleles in a population from central Africa. Antimicrobial Agents and Chemotherapy, 52:4089–4097.

McCollum, A.M., Mueller, K., Villegas, L., Udhayakumar, V., Escalante, A.A. (2007). Common origin and fixation of Plasmodium falciparum dhfr and dhps mutations associated with sulfadoxine-pyrimethamine resistance in a low-transmission area in South America. Antimicrobial Agents and Chemotherapy, 51:2085–2091.

McCollum, A.M., Schneider, K.A., Griffing, S.M., Zhou, Z., Kariuki, S., Ter-Kuile, F., Shi, Y.P., Slutsker, L., Lal, A.A., Udhayakumar, V., Escalante, A.A. (2012). Differences in selective pressure on dhps and dhfr drug resistant mutations in western Kenya. Malaria Journal, 11:77.

Meirmans, P.G. (2012) The trouble with isolation by distance. Molecular Ecology, 21:2839–46.

Moreno, J.E., Rubio-Palis, Y., Páez, E., Pérez, E., Sánchez, V. (2007). Abundance, biting behaviour and parous rate of anopheline mosquito species in relation to malaria incidence in gold-mining areas of southern Venezuela. Medical and Veterinary Entomology, 21:339–349.

Mu, J., Awadalla, P., Duan, J., McGee, K.M., Joy, D.A., McVean, G.A., Su, X.Z. (2005) Recombination hotspots and population structure in Plasmodium falciparum. PLoS Biology, 3:e335.

Mwangi, J.M., Omar, S.A., Ranford-Cartwright, L.C. (2006). Comparison of microsatellite and antigen-coding loci for differentiating recrudescing Plasmodium falciparum infections from reinfections in Kenya. International Journal of Parasitology, 36:329–336.

Nash, D., Nair, S., Mayxay, M., Newton, P.N., Guthmann, J.P., Nosten, F., Anderson, T.J. (2005). Selection strength and hitchhiking around two anti-malarial resistance genes. Proceedings of the Royal Society B: Biological Sciences, 272:1153–1161.

Neafsey, D.E., Schaffner, S.F., Volkman, S.K., Park, D., Montgomery, P., Milner. D.A. Jr, Lukens, A., Rosen, D., Daniels, R., Houde, N., Cortese, J.F., Tyndall, E., Gates, C., Stange-Thomann, N., Sarr, O., Ndiaye, D., Ndir, O., Mboup, S., Ferreira, M.U., Moraes Sdo, L., Dash, A.P., Chitnis, C.E., Wiegand, R.C., Hartl, D.L., Birren, B.W., Lander, E.S., Sabeti, P.C., Wirth, D.F. (2008) Genome-wide SNP genotyping highlights the role of natural selection in Plasmodium falciparum population divergence. Genome Biology, 9:R171.

Nyachieo, A., Van Overmeir, C., Laurent, T., Dujardin, J.C., D'Alessandro, U. (2005). Plasmodium falciparum genotyping by microsatellites as a method to distinguish between recrudescent and new infections. American Journal of Tropical Medicine & Hygiene, 73:210–3.

Orjuela-Sánchez, P., da Silva, N.S., da Silva-Nunes, M., Ferreira, M.U. (2009). Recurrent parasitemias and population dynamics of Plasmodium vivax polymorphisms in rural Amazonia. American Journal of Tropical Medicine & Hygiene, 81:961–968.

Orjuela-Sánchez, P., Karunaweera, N.D., da Silva-Nunes, M., da Silva, N.S., Scopel, K.K., Gonçalves, R.M., Amaratunga, C., Sá, J.M., Socheat, D., Fairhust, R.M., Gunawardena, S., Thavakodirasah, T., Galapaththy, G.L., Abeysinghe, R., Kawamoto, F., Wirth, D.F., Ferreira, M.U. (2010). Single-nucleotide polymorphism, linkage disequilibrium and geographic structure in the malaria parasite Plasmodium vivax: prospects for genome-wide association studies. BMC Genetics, 11:65.

Park, S.D.E. Trypanotolerance in West African Cattle and the Population Genetic Effects of Selection. PhD thesis. University of Dublin; 2001. Microsatellite tool kit version 3.1.1 available from: http://www.animalgenomics.ucd.ie/sdepark/ms-toolkit/

Payseur, B.A., Jing, P., Haasl, R.J. (2011). A genomic portrait of human microsatellite variation. Molecular Biology and Evolution, 28:303–312.

Pritchard, J.K., Stephens, M., Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics, 155:945–959. Structure 2.3.3 available from:http://pritch.bsd.uchicago.edu/structure.html.

Pumpaibool, T., Arnathau, C., Durand, P., Kanchanakhan, N., Siripoon, N., Suegorn, A., Sitthi-Amorn, C., Renaud, F., Harnyuttanakorn, P. (2009). Genetic diversity and population structure of Plasmodium falciparum in Thailand, a low transmission country. Malar Journal, 8:155.

Razakandrainibe FG, Durand P, Koella JC, De Meeüs T, Rousset F, Ayala FJ, Renaud F. (2005). "Clonal" population structure of the malaria agent Plasmodium falciparum in high-infection regions. Proceedings of the National Academy of Sciences, 102:17388–93.

Restrepo, E., Imwong, M., Rojas, W., Carmona-Fonseca, J., Maestre, A. (2011). High genetic polymorphism of relapsing P. vivax isolates in northwest Colombia. Acta Tropica, 119:23–29.

Rosenberg, N.A. (2004). Distruct: a program for the graphical display of population structure. Molecular Ecology Notes, 4:137–138. Distruct 1.1 available at:http://www.stanford.edu/group/rosenberglab/distruct.html.

Russell, B., Suwanarusk, R., Lek-Uthai, U. (2006). Plasmodium vivax genetic diversity: microsatellite length matters. Trends in Parasitology, 22:399–401.

Schneider, K.A., Kim, Y. (2010). An analytical model for genetic hitchhiking in the evolution of antimalarial drug resistance. Theoretical Population Biology, 78:93–108.

Schneider, K.A., Kim, Y. (2011). Approximations for the hitchhiking effect caused by the evolution of antimalarial-drug resistance. Journal of Mathematical Biology, 62:789–832.

Sunnucks, P. (2000). Efficient genetic markers for population biology. Trends in Ecology and Evolution, 15:199–203.

Urdaneta L, Lal A, Barnabe C, Oury B, Goldman I, Ayala FJ, Tibayrenc M. (2001). Evidence for clonal propagation in natural isolates of Plasmodium falciparum from Venezuela. Proceedings of the National Academy of Sciences, 98:6725–6729.

Van den Eede, P., Van der Auwera, G., Delgado, C., Huyse, T., Soto-Calle, V.E., Gamboa, D., Grande, T., Rodriguez, H., Llanos, A., Anné, J., Erhart, A., D'Alessandro, U. (2010) Multilocus genotyping reveals high heterogeneity and strong local population structure of the Plasmodium vivax population in the Peruvian Amazon. Malaria Journal, 9:151.

Volkman, S.K., Sabeti, P.C., DeCaprio, D., Neafsey, D.E., Schaffner, S.F., Milner, D.A. Jr, Daily, J.P., Sarr, O., Ndiaye, D., Ndir, O., Mboup, S., Duraisingh, M.T., Lukens, A., Derr, A., Stange-Thomann, N., Waggoner, S., Onofrio, R., Ziaugra, L., Mauceli, E., Gnerre, S., Jaffe, D.B., Zainoun, J., Wiegand, R.C., Birren, B.W., Hartl, D.L., Galagan, J.E., Lander, E.S., Wirth. D.F. (2007). A genome-wide map of diversity in Plasmodium falciparum. Nature Genetics, 39:113–119.

Wootton, J.C., Feng, X., Ferdig, M.T., Cooper, R.A., Mu, J., Baruch, D.I., Magill, A.J., Su, X.Z. (2002). Genetic diversity and chloroquine selective sweeps in Plasmodium falciparum. Nature, 418:320–323.

Wright S. Isolation by distance. (1943). Genetics, 2:114–38.

CHAPTER 3

ANALYZING LOCAL PLASMODIUM POPULATION DYNAMICS USING A BAYESIAN APPROACH

In America, it is estimated that 137 million people live in areas at risk of *Plasmodium* transmission. However, malaria incidence has significantly decreased during the past decade. In such endemic areas with low malaria transmission, control programs consider a transition from control to elimination. Nevertheless, fluctuations in local transmission intensity as well as local heterogeneity in malaria incidence reduce the efficacy of elimination programs. Malaria incidence is the summary of several processes affecting the parasite population. Thus, malaria surveillance can optimize the deployment of control strategies by considering temporal and spatial changes in the parasite *Plasmodium* populations by using their genetic information. In this study, we explore methodologies to assess *Plasmodium falciparum* demographic changes. Specifically, we studied the effect of changes in the drug policy implementation, as well as estimated the temporal changes of local population clusters and drug resistant lineages in Turbo, Colombia. We used microsatellite loci because they are expected to be neutral if they are not linked to loci under directional selection and their mutation rates allow tracking events at a time scale that is relevant for epidemiologic investigations. In addition, all samples were genotyped for drug resistant mutations (crt, dhfr and dhps) by direct sequencing. We used a Bayesian framework to determine the best supported mutation model and to estimate the observed microsatellite mutation rate. Furthermore, the information was used to evaluate historical patterns in Plasmodium heterozygosity as it is affected by changes in

the parasite population size. We observed a clear turnover of clonal lineages in this population that correlates with major drug policy changes. Our analyses estimate that the average mutation rate for these microsatellite loci lies between $5.35x10^{-3}$ and $3.77x10^{-2}$ substitutions/locus/month and that the root height lies between 16 and 30 years. If we consider that there is a strong population structure in the continent, local malaria control programs could easily detect changes in the malaria transmission patterns and re-introductions by using markers evolving at these estimated rates.

3.1.Background

Malaria causes around 219 million cases every year (WHO, 2012). During the 1970s and 1980s, malaria prevalence and incidence significantly increased in South America due to the rapid and disorganized development and settlement process (Cruz et al., 2013). Fortunately, during the past ten years, the trend has reversed and malaria has significantly decreased. Furthermore, there are many areas with limited spatial connectivity. Such endemic areas in South America, where the number of cases have been controlled and *Plasmodium* populations are considerably patchy, are potential targets for malaria elimination.

Earlier studies of *P. falciparum* population structure in Latin America identified clonal parasite populations due to inbreeding and bottlenecks caused by drug control and epidemic expansions (McCollum et al. 2007). In these areas, local populations exhibit high genetic differentiation when compared with others (Griffing et al., 2012; Echeverry et al., 2013). Although population structure and the persistence of clonal lineages have

been described, there have been few developments directed to translate information derived from the *Plasmodium* demographic into programs aiming to evaluate the efficacy of malaria control and elimination tools. For instance, we can monitor the turnover of clones or clusters of haplotypes circulating in a specific area through time (Chenet et al., 2012) and/or elucidate the acquisition, evolution and dissemination of drug resistance genotypes as well as changes in the population gene frequency that could reduce its population size.

In our study, we used samples from Colombia, one of the two most malaria-endemic countries in South America (Rodriguez et al., 2011). Like in many other countries, malaria in Colombia is a disease that affects mostly rural areas. Indeed, approximately ten million people in this country live in rural areas with potential malaria risk. The northwestern region of Colombia harbors almost 60% of the total malaria cases in the country distributed in both Antioquia and Cordoba departments (Rodriguez et al., 2011). Furthermore, two regions in Antioquia (Uraba and Bajo Cauca) are responsible for almost 90% of malaria cases of this state. The main economic activity of Bajo Cauca is outwash mining followed by cattle farming and rice and sorghum cultivation. In Urabá, the banana growing industry is the main activity, mostly located in the central zone, followed by cattle farming in the northern part and timber exploitation from natural forests in the southern zone. Due to these economic activities, there is substantial migration of workers from highly exposed malaria areas (deeper in the forest near mosquito breeding sites) to towns. Migrations from these areas have highly contributed to the malaria problem in Colombia, spreading *P. falciparum* drug resistance lineages (Blair et al, 2006).

The rapid increase of sulfadoxine-pyrimethamine (SP) resistant cases led to changes in the national drug policy. In 2006, the Colombian government implemented an Artemisinin Combined Therapy (ACT) scheme (Hernandez et al., 2013). The current distribution of SP resistance in Colombia varies significantly. There are high prevalence of SP resistance levels in the Colombian Amazon basin (Osorio et al., 2007), moderate prevalence in the northwestern regions and zones surrounding Sinú and Cauca rivers (Blair et al., 2006 and Pérez et al., 2008), and only a few cases of treatment failure in the southern Pacific coast (González et al., 2003; Hernandez et al., 2013).

Our main objective was to describe *Plasmodium* demographic changes before and after drug policy changes, as well as to estimate the age of population clusters in a region of Colombia with relatively high malaria incidence. We used microsatellite data since their mutation rate is suitable for investigations of recent demographic changes using methods based on the coalescent theory and can provide unique insights into the demographic history of the parasite population.

3.2. Materials and Methods

3.2.1. *Plasmodium* isolates

In this study, we used a total of two hundred sixty-three samples of Plasmodium falciparum from a surveillance study in Turbo (Antioquia department), Colombia. One hundred samples were collected between October 2002 and July 2003, eighty between January 2004 to December 2005, forty-seven between April to October 2007 and thirty-six between March 2008 and July 2009. In order to analyze the data, we grouped all

samples into three time frames, before (Group I: 2002-2003) and after the ACT treatment implementation (Group II: 2004-2005 and Group III: 2007-2009).

All samples came from patients with a history of fever for two to three days, symptoms compatible with malaria, and a unique infection with *P. falciparum* confirmed by thick smear. These patients attended the malaria diagnosis clinic at their local hospital. Males and non-pregnant females (confirmed by a fast pregnancy test), above one year of age, living in the rural or urban zone of their municipality, where they normally reside, were included in the study. Patients with severe malaria or any other disease or with appearance of serious adverse effects due to antimalarial drugs were excluded from the study. Patients infected with P. falciparum were treated with amodiaquine (AQ) plus SP until 2006. In 2007, the effective treatment was artesunate (AS) plus mefloquine (MQ), which in 2008 changed to artemether plus lumefantrine (Coartem) for uncomplicated malaria.

3.2.2. DNA isolation and genotyping methods

Genomic DNA was isolated from whole blood using the QIAamp DNA mini kit (QIAGEN, Valencia, CA) and whole genome amplification was performed using the Repli-g Mini Kit (QIAGEN, Valencia, CA). The samples were genotyped for P. falciparum mutations at crt codons 72-76, dhfr codons 50, 51, 59, 108 and 164 and dhps codons 436, 437, 540, 581 and 613 by direct sequencing using an Applied Biosystems 3730 capillary sequencer. Samples from single infections as determined by microsatellite loci (see next section) were only considered. The primers used for the amplification and sequencing of dhfr were: AE721: 5'- ATGATGGAACAAGTCTGCG -3' and AE773: 5'-GAATTCTTCTACTTGTAGGATC -3'. The cycling conditions were as followed: 94°C

for 7 minutes; 40 cycles of 94°C for 1 minute, 54°C for 1 minute, and 72°C for 2 minutes; and a final extension of 72°C for 10 minutes. For the amplification and sequencing of crt and dhps the primers and conditions used have been previously described (Griffing et al., 2010; Vinayak et al., 2010).

3.2.3. Microsatellite analysis

Twelve neutral microsatellites (POLYa, TA60, ARA2, Pfg377, PfPK2, TA109, TA81, TA87, TA40, TA1, TA42 and 2490) scattered in chromosomes 4, 5, 6, 10, 11, 12 and 13 were used (Anderson et al., 1999) in this study. In addition, all samples were assayed for 15 microsatellite loci that span 129 Kb on chromosome 4 around *Pfdhfr*, 15 loci that span 139 Kb on chromosome 9 around *Pfdhps* (McCollum et al., 2012) and 10 loci that span 500Kb on chromosome 7 around *Pfcrt* (Griffing et al., 2010). Fluorescently labeled PCR products were separated on an Applied Biosystems 3730 capillary sequencer and scored using Gene Marker v1.95 (SoftGenetics LLC). The discovery of one or more additional alleles was interpreted as a co-infection with two or more genetically distinct clones in the same isolate. Missing data (no amplifications) are reported by loci but not considered for defining haplotypes.

3.2.4. Population genetic analysis

Microsatellite data was formatted using the Microsatellite tool kit (Park, 2001). The heterozygosity estimate (H_E) was used as a measure of overall genetic diversity. This was defined as $H_E = [n/(n-1)][1 - \sum_{i=1}^{L} p_i^2]$, where n is the number of isolates analyzed

and p_i is the frequency of the i-th allele (i=1, …, L) in the sample. H_E estimates the average probability that a pair of alleles randomly selected from the population is different. The sampling variance for H_E was calculated as $2(n-1)/n^3[2(n-2)][\sum_{i=1}^{L} p_i^3 - (\sum_{i=1}^{L} p_i^2)^2]$ (Nash, 2005). To test whether microsatellite haplotypes clustered as a single geographic population, the model-based clustering algorithm implemented in the Structure 2.1 software was applied (Pritchard et al., 2000). This software uses a Bayesian clustering approach to assign isolates to K populations or clusters characterized by a set of allele frequencies at each locus. The observed genetic diversity was evaluated at different K values (K=2 to 10). Given that this clustering algorithm relies on stochastic simulations, each K value was run independently ten times with a burn-in period of 10,000 iterations followed by 50,000 iterations. The admixture model was used in all analyses and allows for the presence of individuals with ancestry in two or more of the K populations (Pritchard et al., 2000). We used Structure harvester v0.6.8 to visualize the output from Structure (Earl and vonHoldt, 2012). Although such genetic structures are a deviation from the expectations under one population undergoing random mating, each cluster cannot be considered a random mating population on its own since such clusters could represent clonal lineages. We used Network 4.6.1.1 to construct (Bandelt et al., 1999) median-joining networks using neutral and microsatellite-link data.

Evidence of pairwise linkage disequilibrium between different loci and Fst comparisons between clusters were calculated using Arlequin 3.11 (Excoffier, G et al., 2005). A standardized index of association (I^S_A) was also used to test for evidence of overall

multilocus linkage disequilibrium in the population. This test compares the variance (V_D) of the number of alleles shared between all pairs of haplotypes observed in the population (D) with the variance expected under random association of alleles (V_E) as follows: $I^S_A = (V_D / V_E - 1) (r - 1)$, where r is the number of loci analyzed. V_E is derived from 10,000 simulated data sets in which alleles are randomly reshuffled among haplotypes. Significant linkage disequilibrium is detected if V_D is greater than 95% of the values derived from the reshuffled data sets. Data were analyzed with LIAN 3.5 (Haubold et al., 2000).

3.2.5. Microsatellite mutation models

Microsatellites gain and lose repeat units by unequal crossing over or DNA-replication slippage, a mutation mechanism that is specific to tandemly repeated sequences (Schlötterer et al., 2000). Two extreme mutation models for microsatellites have been previously described: the infinite allele model (IAM; Kimura and Crow, 1964) and the stepwise mutation model (SMM; Kimura and Ohta, 1978). In the IAM, each mutation creates a novel allele at a given rate, μ. Consequently, this model does not allow for homoplasy. Identical alleles share the same ancestry and are identical by descent (IBD). The second model is the SMM (Kimura and Ohta, 1978). Under this scenario, each mutation creates a novel allele either by adding or deleting a single repeated unit of the microsatellite, with an equal probability μ/2 in both directions. Consequently, alleles of very different sizes will be more distantly related than alleles of similar sizes.

More recently, a framework for Bayesian coalescent inference from microsatellite data has been implemented in Beast v.1.7.5 (Wu and Drummond, 2011) using 12 different microsatellite mutation models (Sainudiin et al., 2004). These are classified according to 3 criteria:

a) Mutation rate independence from allele length. The mutation rate could be independent of length (equal rate = E) or the mutation rate could be proportional to microsatellite length (proportional rate = P).
b) The probability of a contraction versus an expansion. Both probabilities are equal (unbiased = U), the probabilities are unequal but independent of allele length (constant = C) or the probabilities are unequal but dependent on allele length (linear = L).
c) Effect of mutation on allele length. Either only mutations of a single repeat (one phase = 1) or mutations of more than one repeat (two phase = 2) are allowed.

Microsatellite mutation rates range from 10^{-6} to 10^{-2} per generation and are thus significantly higher than nucleotide substitution rates (Schlötterer et al., 2000). In *P. falciparum* an average microsatellite mutation rate of 1.59×10^{-4} has been estimated from the observation of five unique non-parental alleles in 35 progeny of a genetic cross that were genotyped for 901 microsatellites (Su et al., 1999; Anderson et al., 2000). However, microsatellite mutation rates can also be estimated from population data. The underlying rationale is that the observed variance in repeat number is determined by the effective population size and the microsatellite mutation model. Hence, the mutation rate can be determined if these parameters are known. The strength of this approach is that a

relatively moderate sample size is sufficient to estimate mutation rates. The drawback of this approach is that it rests on the assumption of mutation drift equilibrium, a condition that is rarely met. Furthermore, effective population sizes are very difficult to determine in natural populations (Schlötterer et al., 2000).

3.2.6. Coalescent analysis

We used BEAST version 1.7.5 (Drummond et al., 2011) to estimate the microsatellite mutation rate by testing the 12 different microsatellite models (Wu and Drummond 2011). Each analysis was performed including tip dates in months (collection date) and using the simplest coalescent constant population size model with a strict clock and a 1/x prior. We assumed a parasite generation time of one month for our analysis considering the sexual and asexual stage of *Plasmodium* inside the mosquito and the human host with a mutation rate that should correspond to the *Plasmodium* meiotic and mitotic rate in the population. We then calculated the Bayes Factor (the harmonic mean of the log likelihood) to compare microsatellite models and chose the one that best fitted the data. We also used the Extended Bayesian Skyline Plots (EBSP) to estimate the population size through time. This is a non-parametric piecewise linear model, which combines information from multiple unlinked neutral loci and estimates the number of population size changes. Each analysis was performed with at least three independent runs based on $8x10^8$ generations and samples taken every $1x10^6$ steps after an appropriate burn-in lasting at least 80 million generations. Convergence of the Markov chains was checked in TRACER (Rambaut and Drummond, 2007) with ESS>100.

3.3.Results

3.3.1. Genotyping of neutral microsatellite loci

A total of 254 samples correctly amplified for nine loci (clear single peak), from which 57 samples were mixed infections. We identified 80 different haplotypes (APPENDIX A) from all single infections as well as from mixed infections for one locus only. From these haplotypes, four of them (No. 13, 19, 36 and 38) were consistently found since 2003. It is possible that we did not recover all haplotypes from 2002-2003 on the following years due to the sampling process and differences in sample size (Table 1a). We further analyzed all samples by grouping them according to their collection dates, and found a significant difference between all years, except between 2007 and 2009, and between 2008 and 2009 (Table 1b). The heterozygosity in the first group and in the second group were slightly higher than in samples from the third group; additionally, a greater number of infections and haplotypes were found in the first group (Table 1a). When evaluating the genetic diversity per locus, we observed that four of the nine loci were highly conserved through time (Table 2).

We also found significant LD between different loci when performing a pairwise LD analysis per year. Moreover, when performing the multilocus LD association analysis, we found a significant LD result in all groups of samples per year ($I^S_{A2003} = 0.058$, $p<0.01$; $I^S_{A2004} = 0.056$, $p<0.01$; $I^S_{A2005} = 0.14$, $p<0.01$, $I^S_{A2007} = 0.072$, $p<0.01$; $I^S_{A2008} = 0.143$, $p<0.01$; $I^S_{A2009} = 0.163$, $p<0.01$) except in 2002, most likely because of the small number of samples collected in this year.

Table 1a. Plasmodium genetic diversity per year

Year	Number of cases[+]	Sample size*	Mixed infections[++]	Haplotypes	He	SD	No Alleles	SD
Group I	1270	79	22	39	0.33	0.1	2.33	1.22
2002	864	12	3	12	0.35	0.11	2.44	1.24
2003	406	67	19	31	0.31	0.09	3	1.12
Group II	714	88	17	32	0.31	0.1	2.7	2.18
2004	365	51	11	23	0.29	0.1	2.44	1.74
2005	349	37	6	16	0.28	0.09	2.11	1.62
Group III	1230	79	18	23	0.23	0.07	3	1.14
2007	856	56	10	14	0.22	0.08	3	1.41
2008	220	22	3	10	0.3	0.07	2.44	0.88
2009	154	15	5	4	0.17	0.06	2.33	0.71

*Corresponds to the total number of infections using complete haplotypes (mixed infections on multiple loci are not included).
[+]Approximate number of *P. falciparum* cases in Turbo
[++]All mixed infections included (complete and incomplete haplotypes)

Table 1b. Microsatellite-based genetic differentiation (Fst) between years

Year	2002	2003	2004	2005	2007	2008
2003	0.08*					
2004	0.20*	0.07*				
2005	0.19*	0.11*	0.11*			
2007	0.27*	0.18*	0.19*	0.27*		
2008	0.18*	0.08*	0.09*	0.19*	0.08*	
2009	0.29*	0.15*	0.15*	0.29*	0.02	0.02

*P-value<0.05

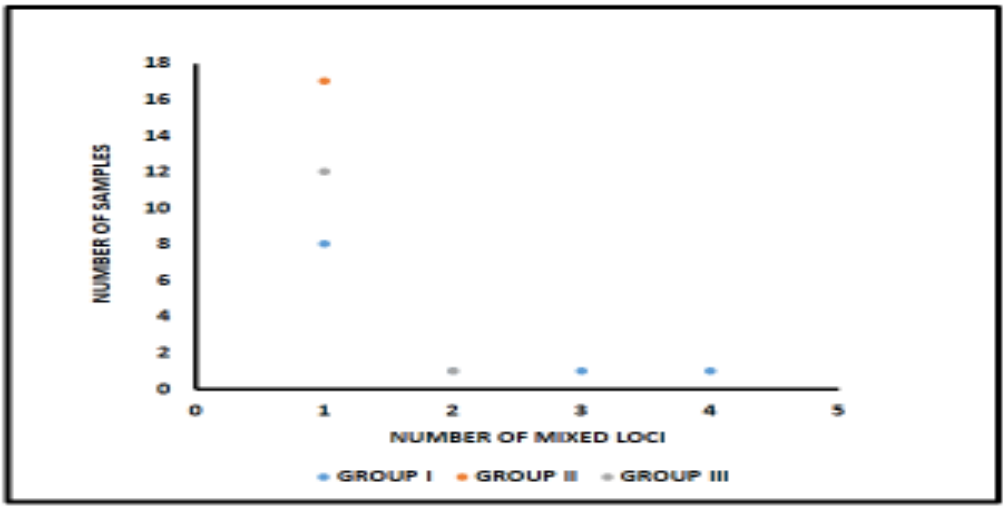
Table 2. Plasmodium genetic diversity per year in each of the microsatellite locus

Year	2002	2003	2004	2005	2007	2008	2009	
He	+/- 0.35	+/- 0.31	+/- 0.29	+/- 0.28	+/- 0.22	+/- 0.30	+/- 0.17	+/-
SD	0.11	0.09	0.10	0.09	0.08	0.07	0.06	
Polya	0.71	0.62	0.57	0.41	0.45	0.59	0.47	
TAA81	0.55	0.51	0.44	0.47	0.19	0.47	0.18	
TA42	0	0	0	0	0.08	0.19	0	
TAA109	0	0.03	0	0	0	0.1	0	
2490	0.64	0.49	0.27	0.48	0.54	0.28	0.18	
ARA2	0.71	0.65	0.61	0.47	0.04	0.44	0.18	
pfg377	0	0	0	0	0	0.1	0	
PfPK2	0.53	0.52	0.75	0.72	0.6	0.56	0.51	
TAA60	0	0	0	0	0.04	0	0	

3.3.2. Mixed infections

Most of the samples with more than one infection (different strains co-infecting the same individual) were different at one locus only, which implies that the lineage present on the samples have very similar haplotypes (Figure 1). However, in the most diverse group (Group I: He = 0.33+/-0.1, mixed infections=22, haplotypes=39) we found mixed infections with different alleles in 3 and 4 loci. On the other hand, we observed a clear pattern of similar haplotypes circulating as co-infections in samples from 2004, where the same pair of haplotypes were found in different samples.

Figure 1. Mixed infections found per locus in different group sampling (I: 2002-2003, II: 2004-2005, III: 2007-2009) using complete multi-locus haplotypes.



3.3.3. Genotyping of *Pfcrt*, *Pfdhfr* and *Pfdhps*

The genotyping results showed that for *Pfcrt* most of the samples were double mutants (75E and 76T) with only one triple mutant (74I, 75E and 76T) and one wild type genotype. For *dhfr*, pyrimethamine resistance-associated mutations N51I and S108N were found in all samples except in one that harbored a wild genotype (migrant from Central America). For *dhps*, only the wild type and the single mutant A437G genotypes were found. A slightly greater amount (15%) of the sulfadoxine-sensitive genotype was found in the Group III samples compared to those (12%) in Group I. We also explored the allelic variation in the microsatellite loci flanking these drug resistant genes and characterized the haplotypes found in single infections or in a single locus mixed infections. A total of 29 different haplotypes were described for *Pfcrt,* 28 were found for *dhfr* and 29 for *dhps* (APPENDIX B-D). For *dhps*, 15 haplotypes were associated with the single mutant and 14 with the wild type.

3.3.4. Cluster analysis

Neutral microsatellite data was also used to characterize the population structure of the *P. falciparum* samples in Turbo. Our results showed that all samples can be categorized in six clusters, some of them maintained through time (Figure 2A; Table 3a; Table 3b). While all six clusters were found in samples from 2002-2003, only five were detected in samples from 2007-2009, and a clear expansion of cluster D was observed in 2005. Cluster E was not present in 2007, 2008 and 2009, which might indicate the introduction of sporadic migrants during 2002 to 2004 or a clonal lineage associated with AQ

sensitivity that was eventually removed from the population due to severe drug pressure. On the other hand, clusters C and F did not showed a significant genetic differentiation, which could be explained to the high level of admixture in the population; specially, if we consider the high percentage of mixed infections (around 22%) present in the samples. We also used Network to infer the relationship between haplotypes (Figure 2B) and clusters maintained through time. Haplotypes from cluster F were highly dispersed in the network without a clear grouping pattern. Plus, from all samples analyzed, only one did not show a clear connection within the network and was considered a migrant from out of the country. This sample had a wild genotype for *Pfcrt*, *Pfdhps* and *Pfdhfr*, which is seen in Central America.

Table 3a. Microsatellite-based genetic differentiation (Fst) between clusters

	A	B	C	D	E
B	0.08*				
C	0.05*	0.11*			
D	0.18*	0.15*	0.31*		
E	0.08*	0.09*	0.17*	0.12*	
F	0.02*	0.07*	0.00	0.24*	0.10*

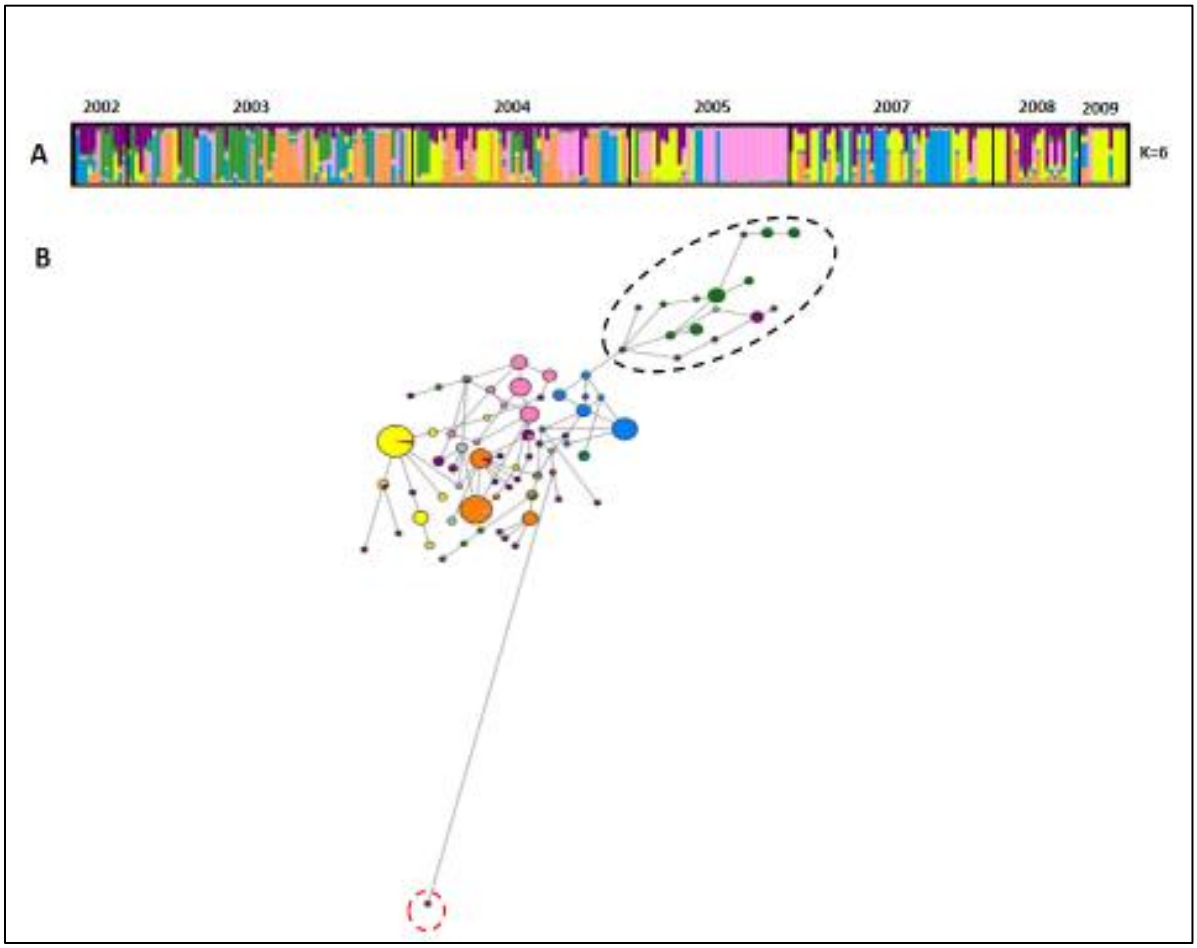
*P-value<0.05

Table 3b. Genetic diversity by cluster

Cluster	Sample size	Haplotypes	He	SD	No Alleles	SD	LD
A (Orange)	45	22	0.28	0.08	3.11	1.45	0.03
B (Blue)	32	19	0.32	0.10	2.33	1.22	0.09
C (Yellow)	46	19	0.23	0.08	2.56	1.81	0.06
D (Pink)	40	22	0.28	0.09	2.44	1.59	0.04
E (Green)	31	24	0.32	0.11	2.44	1.59	0.01
F (Purple)	28	16	0.29	0.09	2.44	1.33	0.03

*P-value<0.05

Figure 2. Population structure results from 9 neutral microsatellite loci. A) Clustering per year using Structure 2.3. Each color represents a different population cluster. B) Median joining haplotype network. The haplotypes are represented by circles with width being proportional to their frequencies. The links are character differences and the red circles represent median vectors required to connect sequences within the network with maximum parsimony. The sample in a dashed red circle represents a migrant most likely from Central America and the group of samples in a black dashed circle refers to a more distant cluster of samples that got lost after 2005.



Additionally, the *dhps* linked-microsatellites results showed three different clusters found in all years with one of them associated with the *dhps* wild genotype, while the other two represent mutant *dhps* haplotypes (Figure 3). We also compared our results with the ones obtained in Peru (Griffing et al., 2013) and Venezuela (McCollum et al., 2007) and constructed haplotype networks with the microsatellite data (Figure 4). We clearly observed how the microsatellite haplotypes linked to resistant genotypes were grouped according to their geographic origin as it is expected for independent origins of resistance. However, the haplotypes that were not associated with the resistant genotype but with the wild type in *Pfdhps*, showed greater diversity and were spread in the network (Figure 4b).

Figure 3. Median joining haplotype network using *Pfdhps*-linked microsatellites. Haplotypes in the red dashed circle are single mutants grouped in color clusters according to their genotypes (using Structure). Haplotypes in green are wild type genotypes.

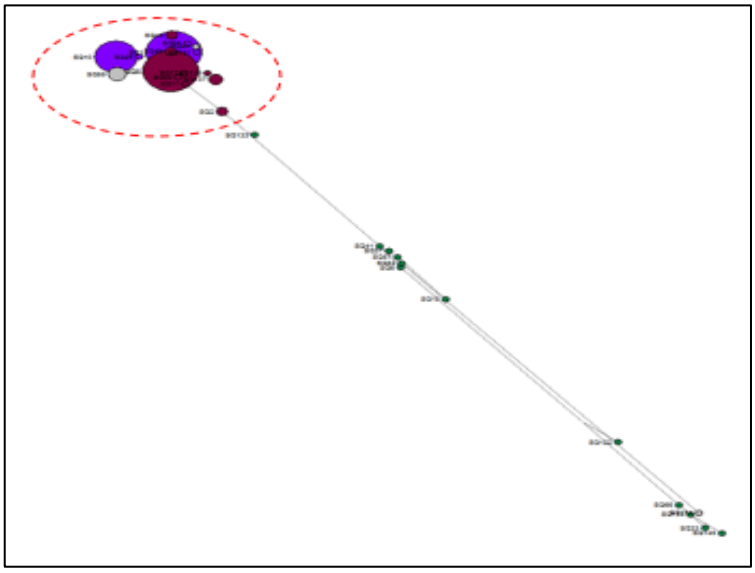
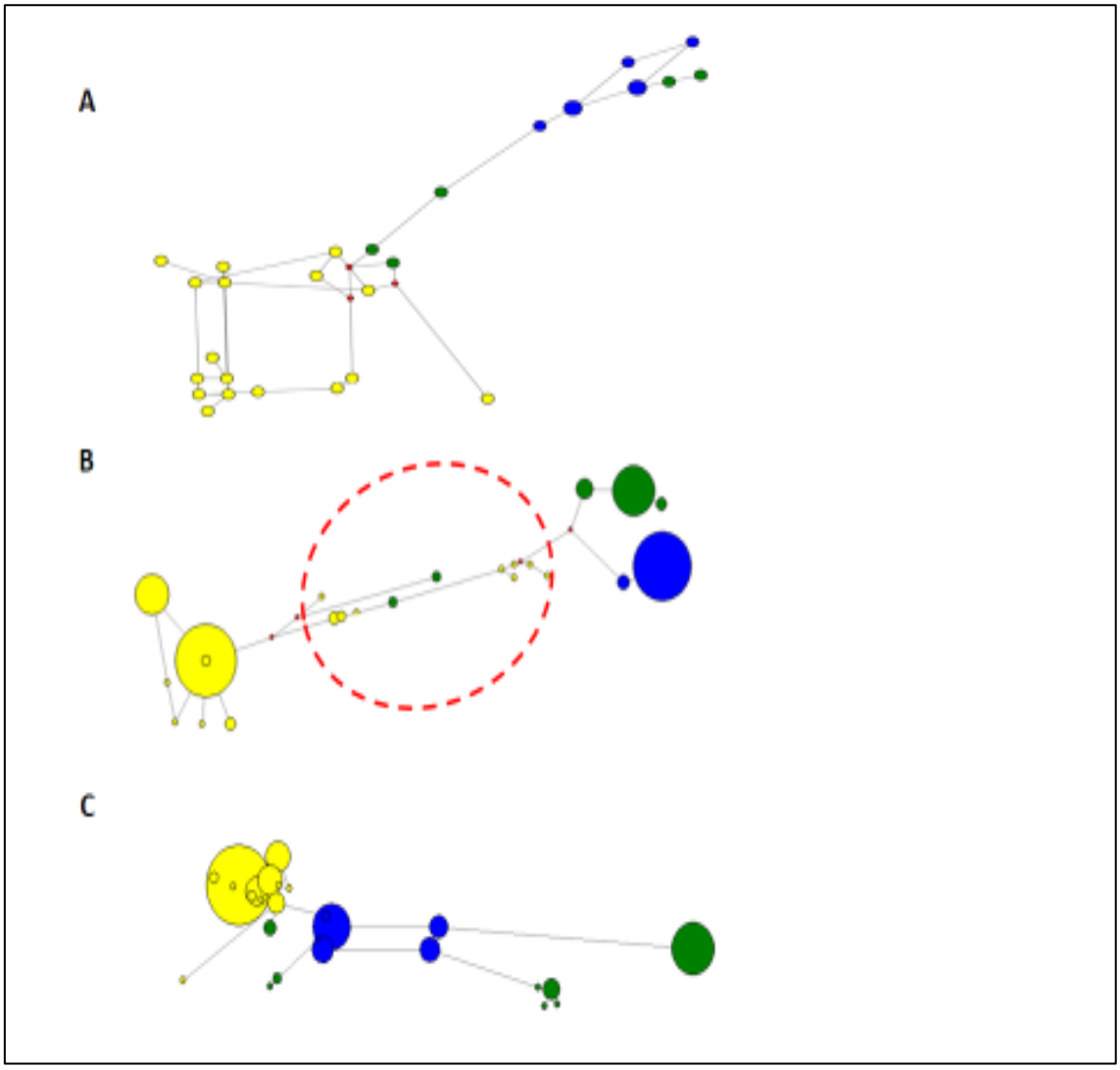


Figure 4. Median joining haplotype networks using a) *Pfcrt*-linked microsatellites b) *Pfdhps*-linked microsatellites and c) *Pfdhfr*-linked microsatellites. Different colors are used for the haplotypes according to the countries from which they belong (yellow=Turbo, Colombia; green=Iquitos, Peru and; blue=Tumeremo, Venezuela). The *Pfdhps* haplotypes in the red dashed circle are wild type.



3.3.5. Time estimation and mutation rates

In order to estimate a mutation rate that could best reflect the change patterns in allele frequencies per month, we tested the 12 different microsatellite models implemented in Beast 1.7.5 and identified the most strongly supported model according to the likelihood of the data under each model. We determined that the best fitting model for the neutral microsatellite loci corresponded to EC1, although we obtained similar results with PL1 with less computational time and, as a result, we decided to proceed with this model for convenience (Table 4). The posterior distribution of the average mutation rate per locus was between 5.35 $x10^{-3}$ and 3.77 $x10^{-2}$ substitutions/month considering the most informative loci (Table 5) with a root height between 19.36±5.11 and 28.37±8.7 years using the most informative neutral microsatellites. We also used the most informative linked microsatellite data to estimate the mutations rates of *Pfcrt* (1.08 $x10^{-3}$ - 3.34 $x10^{-2}$), *Pfdhfr* (7.27 $x10^{-4}$ – 2.93 $x10^{-3}$) and *Pfdhps* (5.95 $x10^{-3}$ – 2.13 $x10^{-2}$) microsatellite flanking loci and obtained similar values as the ones found for the neutral loci. For the mutation rates estimations of these genes, we used the best microsatellite mutation models (PU2, EC1 and PU1, respectively).

According to the EBSP results, we observed fluctuations on the effective population size (Ne) somewhat similar with the ones observed by the number of cases during the years of data collection (Figure 5). However, the pattern is not exactly the same most likely due to the underlying population structure and clonal expansions that increased the number of individuals in the population but not necessarily the Ne in the same magnitude.

Moreover, the number of cases significantly dropped in 2003 and kept this trend until 2006, when the Ne and incidence started to increase again.

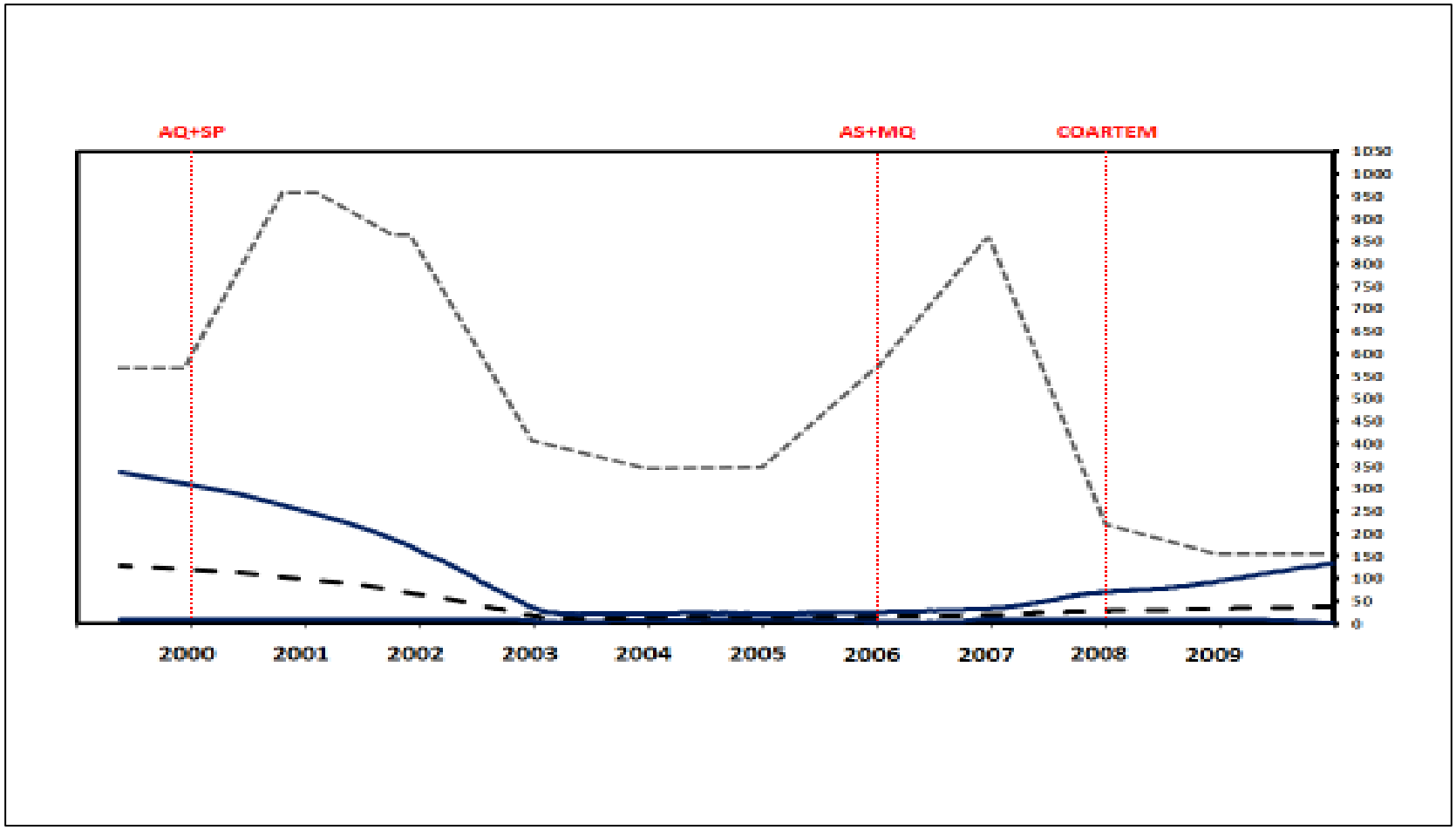
Table 4. Models of microsatellite evolution using neutral microsatellite loci.

Mutation Model	ln P(model | data)	S.E.
EU1	-293.01	+/- 1.986
EU2	-302.502	+/- 2.594
EC1	-270.358	+/- 2.919
EC2	-300.767	+/- 1.168
EL1	-289.253	+/- 1.263
EL2	-324.75	+/- 2.011
PU1	-303.904	+/- 2.6
PU2	-300.213	+/- 3.14
PC1	-865.385	+/- 140.695
PC2	-802.399	+/- 67.545
PL1	-281.057	+/- 1.672
PL2	-328.27	+/- 4.27

Table 5. Summary statistics of mutation rates per neutral locus showing the 6 more informative loci.

Loci	POLYa	ARA2	PfPK2	TA81	TA42	2490
Mean	3.77E-02	1.06E-02	3.25E-02	5.35E-03	8.00E-03	2.18E-02
stderr of mean	1.11E-03	3.89E-04	7.09E-04	1.26E-04	3.77E-04	5.91E-04
Median	3.40E-02	9.57E-03	3.00E-02	4.72E-03	6.39E-03	1.99E-02
geometric mean	3.38E-02	9.27E-03	2.98E-02	4.56E-03	5.81E-03	1.94E-02
95% HPD lower	9.01E-03	1.78E-03	1.04E-02	8.26E-04	1.42E-04	4.68E-03
95% HPD upper	7.37E-02	2.13E-02	6.09E-02	1.16E-02	2.07E-02	4.29E-02
auto-correlation time (ACT)	3.93E+06	5.34E+06	2.74E+06	1.81E+06	3.77E+06	3.22E+06
effective sample size (ESS)	275.038	202.309	393.838	595.85	286.73	335.739

Figure 5. The EBSP represents *P. falciparum* population changes through time. The median population size is represented by the black dashed line, and the 95% HPD by the area between the blue lines. The number of *P. falciparum* cases in Turbo is represented by a dashed gray line and the implemented treatments are in red.



3.4.Discussion

The complex haplo-diploid malaria life cycle poses potential problems when analyzing the *Plasmodium* population genetic data. Because of the unique parasite features such as multiple asexual generations, population expansion within hosts, and stochastic transmission, both natural selection and random genetic drift can increase at the same time and are intensified simultaneously (Chang et al., 2013). It has been argued that recombination is also a major force increasing diversity in the *Plasmodium* population (Sutton et al., 2011) and breaking the clonal structure within a local population. However, *Plasmodium* meiotic recombination takes place only once per life cycle within the mosquito host with a high proportion of inbred individuals; such a dynamic reduces the chances of recombination events in a low transmission area. Moreover, when more deeply analyzing the multi-allelic haplotypes present in mixed infections in Turbo, we found that most of them were highly similar with only one allele different, which could be the result of mitotic mutations rather than sexual recombination. Nevertheless, the relatively high proportion of co-infections present in Turbo could account for the high admixture pattern observed when performing the clustering analysis. Changes in the allele frequencies of *Plasmodium* populations may be driven by several forces; however, in the context of malaria control where a high drug selective pressure is administered, selection and drift might be coupling together and shaping changes in allele frequencies through time.

In areas of Latin America, where a strong clonal population is expected due to inbreeding, the local effective population size is small and drug resistance mutations could go to fixation by genetic drift when they reach high frequency even after removing

the drug pressure. Under such circumstances, any fitness cost that the mutations associated with resistance may have in the absence of the drug is not sufficient for the sensitive parasite to increase in frequency. For instance, our observations of *Pfcrt*, *Pfhdr* and *Pfdhps* in Turbo revealed fixed haplotypes associated with chloroquine and pyrimethamine resistance. This pattern has been observed in other areas with low transmission (Griffing et al., 2010; McCollum et al., 2007). Although genetic drift, especially during epidemic expansions, could fix neutral alleles in isolated populations, neutral mechanisms cannot fix the same phenotype as observed for the resistant parasites across many *Plasmodium* populations (Wootton, 2002). According to this knowledge, the expectation is that there are multiple origins for drug resistance that develop locally (Wellems, 2001) as has been observed in different parts of South America (Corredor et al., 2010; Griffing et al 2010; McCollum et al 2007). Moreover, when comparing the microsatellite haplotype data linked to resistant genotypes from Colombia, Peru and Venezuela, we found that these haplotypes mainly grouped according to their geographic origin which supports the hypothesis of many independent origins of mutations associated with drug resistance in South America. However, lineages from Tumeremo (Venezuela) and Iquitos (Peru) seemed to be more closely related that the ones found in Turbo (Colombia) which could be explained by the Andes acting as a geographic barrier and limiting the gene flow with the other countries. As a result, the highly local haplotypic differentiation observed in South America, implies that there is not enough migration between these areas and that specific local malaria control strategies (considering drug efficacy) should be effective in the elimination goal. A similar pattern has been observed in *P. vivax* (Taylor et al. 2013).

Nevertheless, appropriate tools for *Plasmodium* population surveillance are still necessary. In this context, we used a Bayesian software analysis to evaluate changes in Ne in *Plasmodium* populations through time. This approach is valuable considering that incidence and the parasite effective population size does not necessarily have a linear relationship. By estimating fluctuations on Ne, we can evaluate if the control measures are being effective or not. For instance, if the number of cases increases in the population, but the estimated Ne is still low (single gentoype or a cluster of highly related individuals are circulating in the area), we can attribute this increase in incidence to a faulty in the control policy. Under this scenario, the number of cases could be reduced again once the proper control measures are taken. On the other hand, if we observe an increase in incidence followed by an increase on Ne, we can infer that there is an increase in diversity most likely due to the introduction of individuals or "migrants" from the surrounding areas into the original population.

When applying the analysis to a local population in Colombia, we observed that the incidence in Turbo significantly dropped in 2003, remaining low for the next couple of years. This bottleneck resulted in a decrease in the effective population size as shown by the EBSP. Whereas genetic drift due to changes in transmission intensity could account for the decrease in genetic diversity in *Plasmodium*; it is more plausible that the dramatic drop in prevalence and Ne were driven by strong control measures and subsequent parasite reduction that resulted from changes in drug policies after replacing SP by combined therapy with AQ and by later including an ACT treatment. Thus, our study provided an opportunity to observe allele changes in a local *Plasmodium* population over

many generations (considering a parasite generation of one month, which is approximately the time elapsed for *Plasmodium* to complete its life cycle) and in relation to specific public health interventions. It is worth noting that getting a representative sampling also plays a major role in evaluating genetic diversity fluctuations and changes of Ne through time. However, in this particular case, the increase of Ne after 2006, which also correlated with an increase in the number of malaria cases in the area, it is most likely the result of migrants into the population, which increased the genetic diversity.

3.5.Conclusion

Our main objective was to describe *Plasmodium* demographic changes before and after a new drug policy implementation in a low transmission area in Colombia. We estimated the age of the most recent ancestor of the population clusters and evaluate any fluctuations through time respective to specific malaria drug policy programs implemented. On the other hand, the mutation rate of microsatellites calculated in this study allowed us to estimate the *P. falciparum* effective population size and its fluctuations by representing a mitotic rate of one *Plasmodium* generation per month. These studies provide the scientific bases for the evaluation of treatment policies, control and possible elimination of malaria which translates in a significant Ne drop in areas of low transmission.

3.6.References

Anderson TJ, Su XZ, Bockarie M, Lagog M, Day KP (1999). Twelve microsatellite markers for characterization of Plasmodium falciparum from finger-prick blood samples. Parasitology 119 (Pt 2): 113-125.

Anderson TJ, Haubold B, Williams JT, Estrada-Franco JG, Richardson L, Mollinedo R et al (2000). Microsatellite markers reveal a spectrum of population structures in the malaria parasite Plasmodium falciparum. Molecular biology and evolution 17: 1467-1482.

Bandelt HJ, Dress AW (1992). Split decomposition: a new and useful approach to phylogenetic analysis of distance data. Molecular phylogenetics and evolution 1: 242-252.

Blair S, Carmona-Fonseca J, Pineros JG, Rios A, Alvarez T, Alvarez G et al (2006). Therapeutic efficacy test in malaria falciparum in Antioquia, Colombia. Malaria journal 5: 14.

Corredor V, Murillo C, Echeverry DF, Benavides J, Pearce RJ, Roper C et al (2010). Origin and dissemination across the Colombian Andes mountain range of sulfadoxine-pyrimethamine resistance in Plasmodium falciparum. Antimicrobial agents and chemotherapy 54: 3121-3125.

Cruz LR, Spangenberg T, Lacerda MV, Wells TN (2013). Malaria in South America: a drug discovery perspective. Malaria journal 12: 168.

Chang HH, Moss EL, Park DJ, Ndiaye D, Mboup S, Volkman SK et al (2013). Malaria life cycle intensifies both natural selection and random genetic drift. Proceedings of the National Academy of Sciences of the United States of America 110: 20129-20134.

Chenet SM, Schneider KA, Villegas L, Escalante AA (2012). Local population structure of Plasmodium: impact on malaria control and elimination. Malaria journal 11: 412.

Drummond AJ, Rambaut A (2007). BEAST: Bayesian evolutionary analysis by sampling trees. BMC evolutionary biology 7: 214.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular biology and evolution 29: 1969-1973.

Earl, D.A., von Holdt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno

method. Conservation Genetics Resources, 4:359–361. Structure harvester available from: http://taylor0.biology.ucla.edu/struct_harvest/

Echeverry DF, Nair S, Osorio L, Menon S, Murillo C, Anderson TJ (2013). Long term persistence of clonal malaria parasite Plasmodium falciparum lineages in the Colombian Pacific region. BMC genetics 14: 2.

Excoffier, L. and H.E. L. Lischer (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources. 10: 564-567.

Gonzalez IJ, Padilla JO, Giraldo LE, Saravia NG (2003). [Efficacy of amodiaquine and sulfadoxine/pyrimethamine in the treatment of malaria not complicated by Plasmodium falciparum in Narino, Colombia, 1999-2002]. Biomedica : revista del Instituto Nacional de Salud 23: 38-46.

Griffing S, Syphard L, Sridaran S, McCollum AM, Mixson-Hayden T, Vinayak S et al (2010). pfmdr1 amplification and fixation of pfcrt chloroquine resistance alleles in Plasmodium falciparum in Venezuela. Antimicrobial agents and chemotherapy 54: 1572-1579.

Griffing SM, Viana GM, Mixson-Hayden T, Sridaran S, Alam MT, de Oliveira AM et al (2013). Historical shifts in Brazilian P. falciparum population structure and drug resistance alleles. PloS one 8: e58984.

Haubold B, Hudson RR (2000). LIAN 3.0: detecting linkage disequilibrium in multilocus data. Linkage Analysis. Bioinformatics 16: 847-848.

Hernandez DC, Guerra AP, Cucunuba ZM, Nicholls RS, Barrera SM (2013). Haplotypes associated with resistance to sulfadoxine-pyrimethamine in Plasmodium falciparum in two malaria endemic locations in Colombia. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases 18: 183-190.

Kimura M, Crow JF (1964). The Number of Alleles That Can Be Maintained in a Finite Population. Genetics 49: 725-738.

Kimura M, Ohta T (1978). Stepwise mutation model and distribution of allelic frequencies in a finite population. Proceedings of the National Academy of Sciences of the United States of America 75: 2868-2872.

McCollum AM, Mueller K, Villegas L, Udhayakumar V, Escalante AA (2007). Common origin and fixation of Plasmodium falciparum dhfr and dhps mutations associated with sulfadoxine-pyrimethamine resistance in a low-transmission area in South America. Antimicrobial agents and chemotherapy 51: 2085-2091.

McCollum AM, Schneider KA, Griffing SM, Zhou Z, Kariuki S, Ter-Kuile F et al (2012). Differences in selective pressure on dhps and dhfr drug resistant mutations in western Kenya. Malaria journal 11: 77.

Nash D, Nair S, Mayxay M, Newton PN, Guthmann JP, Nosten F et al (2005). Selection strength and hitchhiking around two anti-malarial resistance genes. Proceedings Biological sciences / The Royal Society 272: 1153-1161.

Osorio L, Perez Ldel P, Gonzalez IJ (2007). [Assessment of the efficacy of antimalarial drugs in Tarapaca, in the Colombian Amazon basin]. Biomedica : revista del Instituto Nacional de Salud 27: 133-140.

Park, S.D.E. Trypanotolerance in West African Cattle and the Population Genetic Effects of Selection. PhD thesis. University of Dublin; 2001. Microsatellite tool kit version 3.1.1 available from: http://www.animalgenomics.ucd.ie/sdepark/ms-toolkit/

Perez MA, Cortes LJ, Guerra AP, Knudson A, Usta C, Nicholls RS (2008). [Efficacy of the amodiaquine+sulfadoxine-pyrimethamine combination and of chloroquine for the treatment of malaria in Cordoba, Colombia, 2006]. Biomedica : revista del Instituto Nacional de Salud 28: 148-159.

Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. Genetics 155: 945-959.

Rodriguez JC, Uribe GA, Araujo RM, Narvaez PC, Valencia SH (2011). Epidemiology and control of malaria in Colombia. Memorias do Instituto Oswaldo Cruz 106 Suppl 1: 114-122.

Sainudiin R, Durrett RT, Aquadro CF, Nielsen R (2004). Microsatellite mutation models: insights from a comparison of humans and chimpanzees. Genetics 168: 383-395.

Schlotterer C (2000). Evolutionary dynamics of microsatellite DNA. Chromosoma 109: 365-371.

Su X, Ferdig MT, Huang Y, Huynh CQ, Liu A, You J et al (1999). A genetic map and recombination parameters of the human malaria parasite Plasmodium falciparum. Science 286: 1351-1353.

Sutton PL, Torres LP, Branch OH (2011). Sexual recombination is a signature of a persisting malaria epidemic in Peru. Malaria journal 10: 329.

Taylor JE, Pacheco MA, Bacon DJ, Beg MA, Machado RL, Fairhurst RM et al (2013). The evolutionary history of Plasmodium vivax as inferred from mitochondrial

genomes: parasite genetic diversity in the Americas. Molecular biology and evolution 30: 2050-2064.

Vinayak S, Alam MT, Mixson-Hayden T, McCollum AM, Sem R, Shah NK et al (2010). Origin and evolution of sulfadoxine resistant Plasmodium falciparum. PLoS pathogens 6: e1000830.

Wellems TE, Plowe CV (2001). Chloroquine-resistant malaria. The Journal of infectious diseases 184: 770-776.

Wootton JC, Feng X, Ferdig MT, Cooper RA, Mu J, Baruch DI et al (2002). Genetic diversity and chloroquine selective sweeps in Plasmodium falciparum. Nature 418: 320-323.

Wu CH, Drummond AJ (2011). Joint inference of microsatellite mutation models, population history and genealogies using transdimensional Markov Chain Monte Carlo. Genetics 188: 151-164.

CHAPTER 4

GENETIC VARIATION IN *Plasmodium vivax* POPULATIONS

Plasmodium vivax is the most widely distributed human malaria species in the world causing an estimated 80-90 million cases each year. In the Americas, *P. vivax* is the most prevalent malaria species; and it represents more than 70% of clinical cases reported annually in the Amazon region. Although *P. vivax* is only rarely lethal, features of its transmission biology give this species greater resilience than *Plasmodium falciparum* to face adverse conditions and succeed in the transmission process. Moreover, relapses and broader temperature tolerance in *P. vivax* are expected to increase gene flow between different geographic regions and greatly contribute to the malaria spreading.

In this project, we studied *P. vivax* populations from Brazil to evaluate population structure and significant gene flow within the Amazon basin and compared these results with what has been reported on *P. falciparum*. The analysis of malaria parasite population structure and gene flow facilitate meaningful control and preventive measures to be implemented since new mutations leading to drug resistance could rapidly spread given these parasite dynamics.

4.1.Background

Malaria is one of the most prevalent parasitic diseases in the world and causes high morbidity and death, mainly in *Plasmodium falciparum*-infected, non-treated patients (WHO, 2012). Although transmission of both *P. falciparum* and *P. vivax* occur

in Latin America (with rare *P. malariae* infections), *P. vivax* causes intense morbidity and contributes to significant political, social and economic instability in developing countries of Latin America.

There are distinct pathologic and epidemiologic characteristics of *P. vivax,* which differ from *P. falciparum*. *P. vivax* is characterized by the presence of a hypnozoite stage, which can persist in the liver for months to years, causing recurrent disease although treatment and clearance of the initial blood stage infection is achieved (Schneider and Escalante, 2013; McCollum et al., 2014). Also, a difference in *Anopheles* mosquito dynamics, which provides *P. vivax* with a broader temperature tolerance, enables its transmission in temperate climates not tolerated by *P. falciparum.*

The so-called benign tertian malaria caused by *P. vivax* is by no means harmless. It is associated with low birth weight during pregnancy and thrombocytopenia, and new research suggests that severe complications from *P. vivax* malaria may be more common than previously thought (Anstey et al., 2007; Kochar et al., 2005). In the last few years a pattern of unusual clinical complications with fatal cases associated with *P. vivax* has been reported in Brazil (Leal-Santos et al., 2013). Moreover, there have been some reports on the emergence of chloroquine resistant *P. vivax* (Baird et al., 2004; Oliveira-Ferreira J et al., 2010). In Brazil, the annual incidence of malaria greatly increased due to massive human migration and the establishment of frontier agricultural settlements and open mining enclaves in the Amazonian rainforest during the 1970s (Cruz Marques, 1987). Later on, the intensification of malaria control measures, including the use of DDT for mosquito control, substantially decreased malaria cases in the Amazon basin

(Oliveira-Ferreira et al., 2010). Currently, malaria transmission hotspots in Brazil are intermingled with areas of low or moderate risk. Most malaria transmission typically occurs in mining and logging camps and new farming settlements (da Silva-Nunes et al., 2008, de Castro et al., 2006 and de Castro et al., 2007), which increase the abundance of the main malaria vector, *Anopheles darlingi* and favors malaria transmission (da Silva et al., 2010). This mosquito species breeds profusely in sunlit pools of water in areas of natural floodplain. In addition, it is highly anthropophilic and exhibits a predominantly exophilic behavior (it feeds on humans when individuals are carrying outdoor activities or it enters houses to feed and leaves to rest on nearby vegetation) (Oliveira-Ferreira et al., 1990).

Local *P. vivax* populations are extremely diverse and geographically structured in the Amazon Basin (Ferreira et al., 2007, Orjuela-Sánchez et al., 2010 and Van den Eede et al., 2010). However, when comparing sympatric parasites of *P. vivax* and *P. falciparum*, much more microsatellite diversity is found in *P. vivax* populations (Ferreira et al., 2007, Orjuela-Sánchez et al., 2009a, Chenet et al., 2012). Thus, this genetic variation in *P. vivax* severely delays the development of the variant-specific component of naturally acquired immunity (Bastos et al., 2007 and Souza-Silva et al., 2010).

In this study, we examined *P. vivax* isolates collected in Amapá, Acre, Pará, Rondônia and Mato Grosso States. We aimed to determine the extent of genetic diversity, linkage disequilibrium and population structure within Brazil and compare the prevalent lineages with those circulating in Colombia and Venezuela during the same time period. Our hypothesis was that extensive *Plasmodium* genetic diversity was going to be found in the

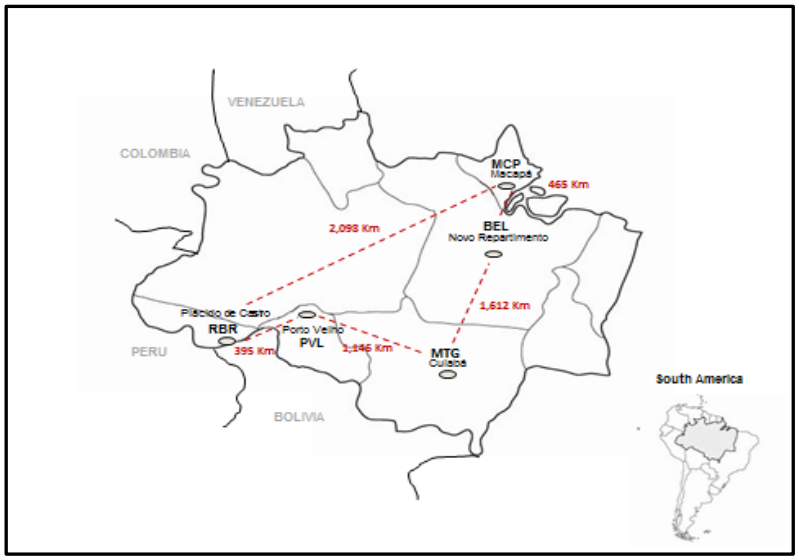
Amazon basin with highly structured populations but with similar lineages shared between other countries in South America due to human migration. In addition, we gave a description of genetic differentiation and the rate of gene flow among populations from the Brazilian Amazon basin based on microsatellite independent markers.

4.2.Methods

4.2.1. *Plasmodium* isolates

Two hundred twenty four DNA samples of *Plasmodium vivax* were obtained from a surveillance study of 5 sites (Macapá, Plácido de Castro, Novo Repartimento, Porto Velho, Cuiabá) in the Brazilian Amazon, from May 2005 to August 2006 (Figure 1). The collection of samples was approved by the Institutional Review Board (IRB) of the Faculty of Medicine, São José do Rio Preto. All participants were enrolled according to the following criteria: sought medical assistance for clinical malaria symptoms, were > 18 years of age, and had a positive *P. vivax* malaria diagnosis by thick blood film. People with other illnesses, pregnant women and patients < 18 year were excluded.

Figure 1. A map of the 5 study sites in the Brazilian Amazon. The dotted red line represents distance in Km between locations.



Macapá (MCP) is the capital of the State of Amapá and is located in the Brazilian Eastern Amazon. It has 381, 215 inhabitants in a territorial area of 6,407 km^2. Plácido de Castro (RBR) is a city located 90 km away from the capital of the State of Acre, on the Western Brazilian Amazon. It has a population of 17,334 inhabitants and a territorial area of 2,047 km^2. Novo Repartimento (BEL) is a gold mining area in southeastern Pará, located in the Brazilian Eastern Amazon. It has 47,197 inhabitants and a territorial area of 11,407 km^2. Porto Velho (PVL), capital of the Rondônia State, is located in the Western Brazilian Amazon, with a population of 360,068 inhabitants and a territorial area of 34,082 km^2. Cuiabá (MTG) is the capital of the Mato Grosso State, with a territorial area of 3,538 km^2 and 516,008 inhabitants. The main economic activities in these areas are agriculture, commerce of manufactured products and farming. The climate in these areas is characterized as tropical with no dry season; thus, malaria transmission is active all year

and *An. darlingi* is the major vector. It is worth mentioning that over 90% of the cases reported in MTG are imported (Storti-Melo et al., 2011; Ministério da Saúde, Resumo Epidemiológico, 2009).

We also included 40 *P. vivax* samples from Turbo, Colombia and 96 samples from Tumeremo, Venezuela (Chenet et al., 2012) collected during the same time period in order to compare the *P. vivax* lineages circulating and evaluate gene flow between these areas in South America.

4.2.2. Microsatellite analysis

Plasmodium vivax samples were assayed for 12 microsatellites (Imwong et al., 2006; Karunaweera et al., 2008). Fluorescently labelled PCR products were separated on an Applied Biosystems 3730 capillary sequencer and scored using Gene Marker v1.95 (SoftGenetics LLC). The finding of one or more additional alleles was interpreted as a co-infection with two or more genetically distinct clones in the same isolate (multiple-clone infection, transmitted by one or several mosquitoes). Additional alleles generating peaks of at least one third the height of the predominant allele were also scored. Missing data (no amplifications) are reported by loci but not considered for defining haplotypes.

4.2.3. Population genetic analysis

Microsatellite data was formatted using the Microsatellite tool kit (Park, 2001). The heterozygosity estimate (H_E) was used as a measure of overall genetic diversity. This was defined as $H_E = [n/(n-1)][1 - \sum_{i=1}^{L} p_i^2]$, where n is the number of isolates analyzed

and p_i is the frequency of the i-th allele (i=1, …, L) in the sample. H_E estimates the average probability that a pair of alleles randomly selected from the population is different. The sampling variance for H_E was calculated as $2(n-1)/n^3[2(n-2)][\sum_{i=1}^{L} p_i^3 - (\sum_{i=1}^{L} p_i^2)^2]$ (Nash, 2005). To test whether microsatellite haplotypes clustered as a single geographic population, the model-based clustering algorithm implemented in the Structure 2.1 software was applied (Pritchard et al., 2000). This software uses a Bayesian clustering approach to assign isolates to *K* populations or clusters characterized by a set of allele frequencies at each locus. The observed genetic diversity was evaluated at different *K* values (*K*=2 to 10). Given that this clustering algorithm relies on stochastic simulations, each *K* value was run independently ten times with a burn-in period of 10,000 iterations followed by 50,000 iterations. The admixture model was used in all analyses and allows for the presence of individuals with ancestry in two or more of the *K* populations (Pritchard et al., 2000). We used Structure harvester v0.6.8 to visualize the output from Structure (Earl and vonHoldt, 2012). Although such genetic structures are a deviation from the expectations under one population undergoing random mating, each cluster cannot be considered a random mating population on its own since such clusters could represent clonal lineages. We used Network 4.6.1.1 to construct (Bandelt et al., 1999) median-joining networks using neutral and microsatellite-link data.

Evidence of pairwise linkage disequilibrium between different loci and Fst comparisons between populations were calculated using Arlequin 3.11 (Excoffier, G et al., 2005). A standardized index of association (I^S_A) was also used to test for evidence of overall

multilocus linkage disequilibrium in the population. This test compares the variance (V_D) of the number of alleles shared between all pairs of haplotypes observed in the population (D) with the variance expected under random association of alleles (V_E) as follows: $I^S_A = (V_D / V_E - 1)\ (r - 1)$, where r is the number of loci analyzed. V_E is derived from 10,000 simulated data sets in which alleles are randomly reshuffled among haplotypes. Significant linkage disequilibrium is detected if V_D is greater than 95% of the values derived from the reshuffled data sets. Data were analyzed with LIAN 3.5 (Haubold et al., 2000).

4.2.4. Migration patterns between populations

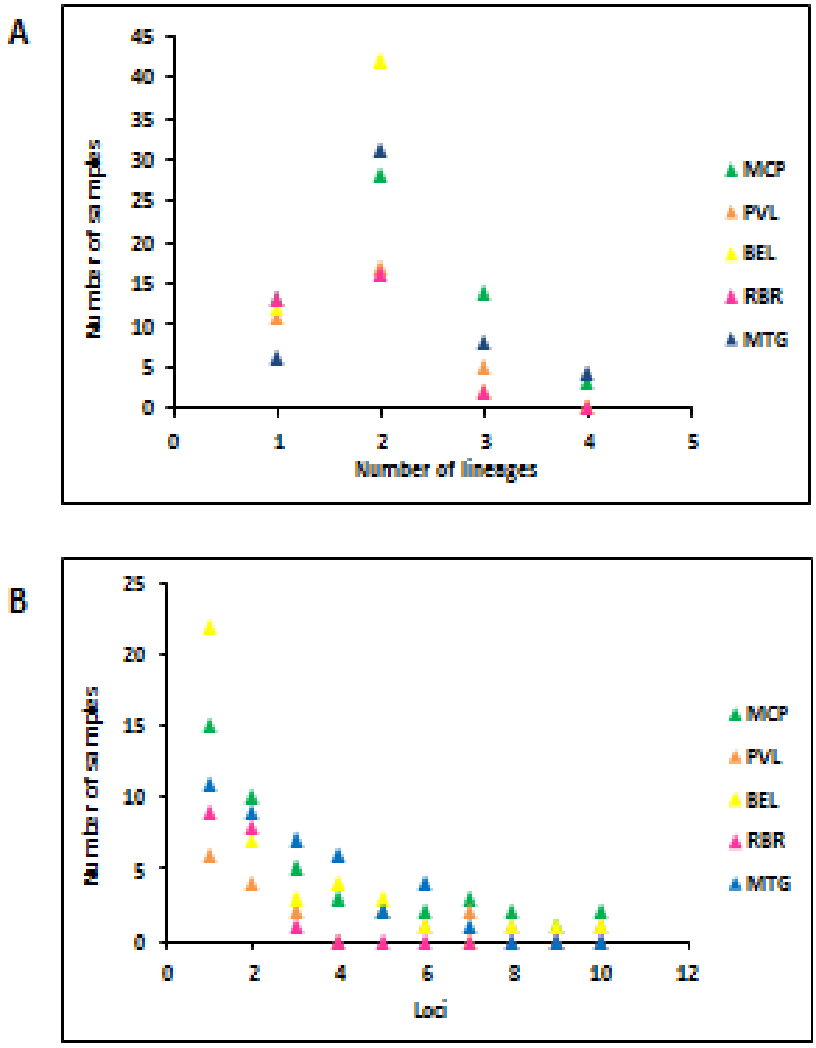
The coalescence-based program MIGRATE-N 3.2.6. (Beerli P and Felsenstein J, 2001; Beerli P, 2006; Beerli P, Palczewski M, 2010) was used to estimate gene flow between populations in Brazil. We also used this software to estimate the mutation-scaled effective population size $\Theta = 2N_e\mu$ (haploid), where N_e is the effective population size and μ is the mutation rate per generation per locus, as well as mutation-scaled migration rates $M = m/\mu$, where m is the immigration rate per generation among populations. The model comparisons were done using the Bayes factors that need the accurate calculation of marginal likelihoods. These likelihoods were calculated using thermodynamic integration in MIGRATE-N 3.2.6 (Beerli P, Palczewski M, 2010).

4.3.Results

4.3.1. Multiple infections

A total of one hundred sixty-six samples from all Brazilian populations had more than one *P. vivax* lineage. Most of the samples had two different lineages co-infecting the same individual with some (from Macapá and Cuiabá) having up to 4 different *P. vivax* parasites circulating at the same time (Figure 2A). When analyzing the number of loci with different alleles – mixed loci – between co-infecting lineages per sample, we found up to 10 loci with at least two alleles (Figure 2B), which implies that the lineages present on the samples are very diverse and that mutations due to mitosis events could not explain such heterogeneity. Almost each lineage (constructed with phase microsatellite gentoypes) corresponds to a different haplotype circulating in these Brazilian populations.

Figure 2. Mixed infections found in the samples. A) Number of different lineages co-infecting an individual. B) Number of different alleles found between lineages.



4.3.2. Genetic diversity

From all populations analyzed, samples from MCP, BEL and MTG had the greatest percentage of mixed infections (Table 1a). Although MTG presented several mixed infections, most of the cases reported in this area are imported from the Amazon basin. Moreover, PVL showed the greatest number of malaria cases in the Amazonic region of

Brazil and migration from this area to other locations is expected. In general, the genetic diversity within each population was very similar (He= 0.75-0.8 +/- 0.02) and when evaluating the genetic differentiation per population, we found little genetic differentiation between populations (Table 1b). These Fst values found between PVL, BEL, RBR, MCP and MTG suggest significant gene flow between these locations. We also calculated Rst values between all populations to consider variations in the underlying mutation rate and found the highest values when comparing between MTG and the other populations (Table 1c). The important factor to evaluate is the magnitude of the ratio of mutation over migration since these could underestimate Fst values; nevertheless, mutation is unlikely to matter when levels of gene flow are high.

Table 1a. *Plasmodium vivax* genetic diversity per population.

Population	N. of cases	Sample size	Mixed infections	N. of haps*	He	SD	No Alleles	SD
MCP	3506	58	45	40	0.79	0.02	9.42	1.98
PVL	36307	33	19	15	0.79	0.02	7.42	1.83
BEL	5764	56	44	34	0.79	0.02	7.83	2.66
RBR	774	31	18	16	0.75	0.02	6.67	2.02
MTG	133	46	40	21	0.80	0.02	9.33	3.03

*Number of haplotypes considering complete information from all loci.

Table 1b. Microsatellite-based genetic differentiation (Fst) per population

Fst	MCP	PVL	BEL	RBR
PVL	0.004*			
BEL	0.002*	0.003*		
RBR	0.004*	0.005*	0.003*	
MTG	0.004*	0.005*	0.003*	0.005*

*P-value<0.05

Table 1c. Microsatellite-based genetic differentiation (Rst) per population

Rst	MCP	PVL	BEL	RBR
PVL	0.013*			
BEL	0.101*	0.063*		
RBR	0.077*	0.024*	0.076*	
MTG	0.229*	0.249*	0.19*	0.19*

*P-value<0.05

4.3.3. Population structure

As a starting point to infer the relationships among populations we used the clustering algorithm implemented in STRUCTURE and explored different numbers of populations (K) to uncover hierarchical population structure (Figure 3; Table 2a; Table 2b). Although pairwise Fst and Rst comparisons of the geographically defined populations did not reveal strong population structure, we managed to group all samples in 5 clusters. According to the color pattern, we observed highly admixed populations and none of the particular clusters could be attributed to a specific geographic location.

Figure 3. Local population structure using neutral microsatellite loci. Each color represents a different population cluster and each bar represents a haplotype.

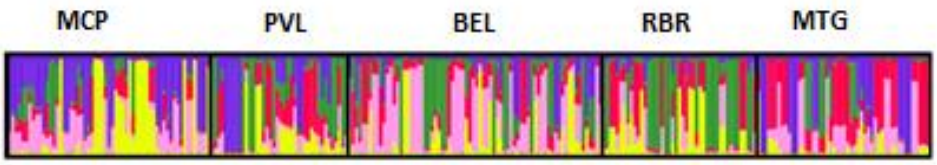


Table 2a. Microsatellite-based genetic differentiation (Fst) per cluster

Fst	1(red)	2(green)	3(purple)	4(pink)
2(green)	0.003*			
3(purple)	0.005*	0.004*		
4(pink)	0.005*	0.004*	0.006*	
5(yellow)	0.008*	0.006*	0.009*	0.009*

*P-value<0.05

Table 2b. Microsatellite-based genetic differentiation (Rst) per cluster

Rst	1 (red)	2(green)	3(purple)	4(pink)
2 (green)	0.15*			
3 (purple)	0.22*	0.23*		
4(pink)	0.11*	0.16*	0.22*	
5 (yellow)	0.16*	0.22*	0.010*	0.19*

*P-value<0.05

4.3.4. Effective population size (Ne) and migration rates

Estimations of mutation-scaled effective population size (θ) and migration rates (M) among populations were performed using Bayesian inference and the Brownian motion mutation model in MIGRATE-N 3.2.6. This model allowed estimating locus specific relative mutation rates by using the number of alleles per locus. We used 9 microsatellites for the analysis since they provided the most complete information to reconstruct haplotypes with phase information. All estimates of θ per location were scaled using a microsatellite mutation rate of $1.59x10^{-4}$ per locus per generation (Anderson et al., 2000) to calculate Ne (Figure 4). Effective population sizes varied among the different Brazilian populations (Ne from 2201 to 6541). Migration rates between populations were very similar within 0.002 and 0.004. This agrees with calculated Fst values which did not show great genetic difference between parasites coming from these geographical locations (Figure 5). Our estimations of migration rates and effective population sizes supported the hypothesis of extensive gene flow between *P. vivax* populations in Brazil.

Figure 4. Estimates of effective population size (Ne) based on Bayesian inferences of migration rates and population sizes. The central box of the plots represents the values from the lower to upper quartile (25 to75 percentile). The horizontal line extends from the 2.5% percentile to the 97.5% percentile.

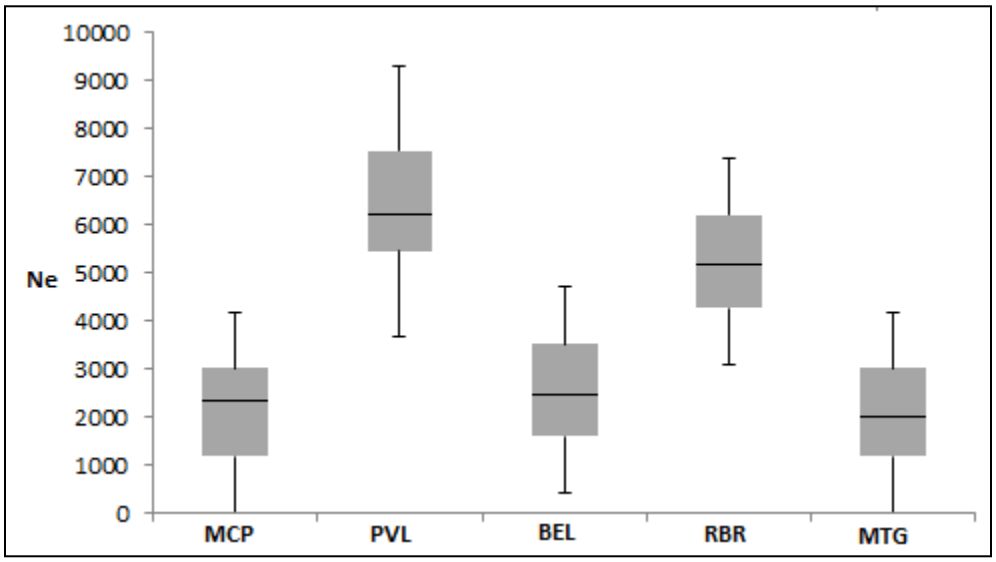


Figure 5. Graph representing the median mutation-scaled effective population size (Θ) and the nmutation-scaled migration rates (M) calculated by MIGRATE (brownian motion model/1 million generations) generated from Fst calculations. The arrows represent the direction of most of the migrantion.

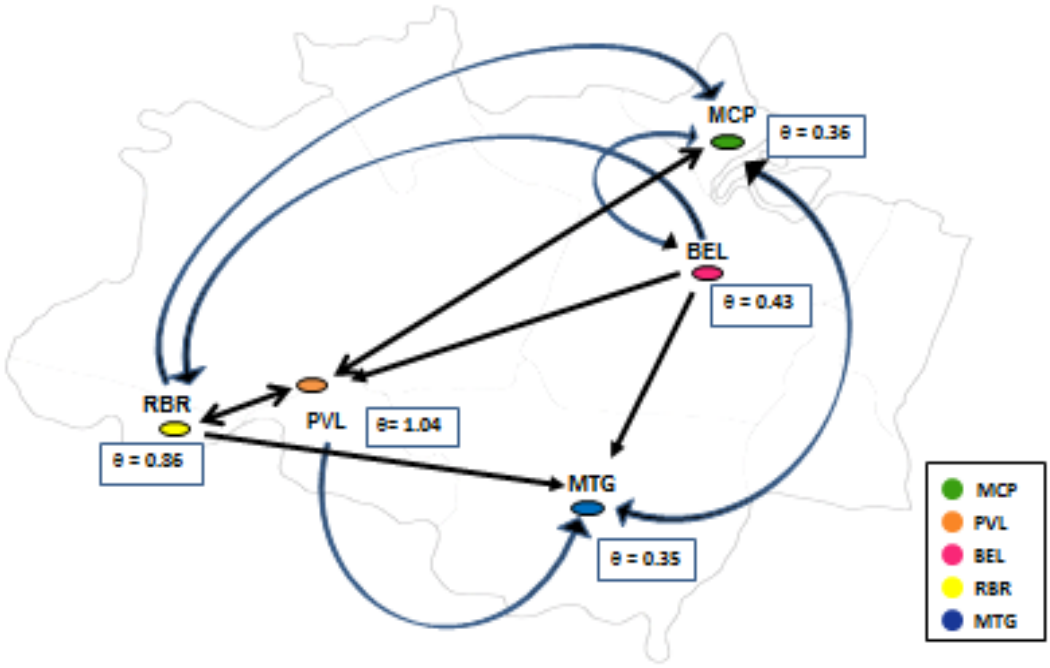


Figure 6. Population structure between South American countries using neutral microsatellite loci. Each color represents a different population cluster and each bar represents a haplotype.

4.3.5. Linkage disequilibrium

Comparable levels of genetic diversity (He=0.5-0.8) per microsatellite locus were found in all samples. Interestingly, we did not find shared haplotypes between different geographical locations which strengthens the idea of high local genetic diversity. However, as it was mentioned before, most of the lineages did share the same alleles in most of the loci (Figure 2b) so haplotypes that differ in one locus could be the result of mitotic mutation and further expansion rather than meiotic recombination.

We also calculated the overall association index (I^S_A) within localities and significant LD was detected ($I^S_{A\ MCP} = 0.12$, $p<0.01$; $I^S_{A\ PVL} = 0.08$, $p<0.01$; $I^S_{A\ BEL} = 0.10$, $p<0.01$, $I^S_{A\ RBR} = 0.2$, $p<0.01$; $I^S_{A\ MTG} = 0.16$, $p<0.01$) in all data sets. Next, we examined patterns of linkage disequilibrium between every pair of markers in all populations and found significant LD between most pairs.

4.4.Discussion

In the Amazon basin, *P. vivax* populations are highly diverse and, unlike *P. falciparum*, are more prone to present different lineages co-infecting the same individual. Particular biological characteristics in *P. vivax* such as relapses and broader temperature tolerance could explain the difference in prevalence and gene flow with sympatric *P. falciparum* populations. Indeed, clear clonal population expansions in *P. vivax* are not as frequently found as in *P. falciparum* due to more recombination chances and efficient parasite spreading (Carlton et al., 2008).

In our study, we confirmed that *P. vivax* isolates from different States of Brazil in rural Amazonia exhibited extensive genetic diversity and frequent multiple clone infections coexisting with strong multilocus linkage disequilibrium. However, the parasites circulating in these locations were different from the ones circulating in other nearby countries such as Colombia and Venezuela (Figure 6), presenting high levels of admixture in the Amazonic area. These results suggest the presence of highly local structured populations in South America with low gene flow between countries. Particularly, we did not observe any shared haplotypes with Turbo, Colombia and the other two countries. This could be explained by the Andes acting as a geographical barrier between the Amazon and the pacific Coast.

Regarding local Brazilian *P. vivax* populations, our results clearly showed extensive migration between amazonic regions in Brazil and Cuiabá (Mato Grosso). In this area, 99% of the cases are reported to be imported (Ministério da Saúde, Resumo Epidemiológico, 2009). Malaria prevalence in Mato Grosso is largely explained by the high mobility of infected individuals between areas with high malaria prevalence (such as mining camps) and others with low prevalence. Furthermore, the region from which the miners originally migrated and the duration of residence in the *garimpo* (gold mining operation camps) may also reflect different responses to malaria (biological resistance based on acquired immunity) (Barbieri et al., 2007). In 2004, a diamond mining rush attracted over 2000 illegal miners from different parts of the Brazilian Amazonia into a reserve in Rondônia adjacent to Mato Grosso. These miners caused extensive changes on the area due to deforestation and disruption in river beds. Morevoer, many of them could

have been already infected with Plasmodia and as a consequence the number of cases in this state increased dramatically. The environmental capacity for malaria transmission in Rondônia is high. We hypothesize that due to the optimal transmission conditions in Rondônia, not only high recombination with the incoming lineages was favored but also clonal expansions of similar but not identical lineages that significantly increased the number of cases in Rondônia and Mato Grosso.

Particularly in Porto Velho (Rondônia's capital), we found highly effective *Plasmodium* population size and migration rate estimates, which imply a significant gene flow between PVL and all of the other sites included in this study. We also found a higher immigration rate from Novo repartimento (a gold mining area) to Porto Velho (Rondônia), as a result of the "diamond fever". Additionally, Porto Vehlo is the most important transportation and communication center of Rondônia; malaria transmission occurs all year and the floodplain of Madeira River, one of the main effluents of the Amazon River, represents a vast mosquito breeding site in close proximity to the city. Moreover, there seems to be extensive migration between PVL and cities in the adjacent states of Acre (Placido de Castro) and Mato Grosso (Cuiabá), which explains the high theta values obtained for those cities.

Similar results suggesting that either Rondônia or Pará acted as source populations for different sites have been previously reported in *P. falciparum* (Griffing et al., 2013); however, in those studies, Griffing et al did not use an explicit model that could account for the population structure like the one reported here. The low Ne and migration estimates obtained for Pará does not suggest great emigration from this area. Still, it is

important to point out that we did have a limited amount of samples coming from the areas, considering the actual number of cases in those regions and as consequence we might not be recovering all haplotype diversity. Nevertheless, we observe a correlation between the estimates of Ne and the actual number of cases per region, except for RBR and MTG, where the Ne values seemed to be overestimated. In Cuiabá (MTG) and Placido de Castro (RBR) the estimates does not represent the autochthonous number of cases since there is extensive migration to these locations; thus, the Ne estimates for Cuiabá and Placido de Castro are a composite of the different source populations. This is an important result since the number of cases is the main metric used to describe fluctuations in *Plasmodium* populations. However, the number of cases alone does not allow us to make inferences in terms of the parasite gene flow and/or migration. Thus, our approach could be used in areas that aim to eradicate malaria and where surveillance and control need to be enhanced with adequate tools. In this way, a similar surveillance approach can be used in other areas in order to distinguish malaria hotspots and/or distinguish between autochthonous and foreign cases due to extensive migration.

4.5.Conclusions

Unfortunately, Brazilian malaria control efforts still have not efficiently suppressed malaria and therefore these locations maintain high numbers of cases. Moreover, migration of infected individuals may seriously restrict containment of *P. vivax*, increasing the potential for reintroduction of *Plasmodium* into non-endemic or controlled areas outside the Amazon basin. While our data confirmed substantial migration between sites in the Brazilian Amazon Basin, we did not find any structural or clustering patterns

of *P. vivax* lineages explaining the direction of this migration but we could have an idea of the major sources due to the θ estimates and migration rates. There is still a lack of knowledge regarding transmission and spread of *P. vivax* in areas where malaria control and elimination programs have made progress. The data gathered and methodology presented in this study is useful for assessing changes in malaria transmission and for more efficiently manage resources to enforce control measures in locations that act as parasite "sources" for other regions. It is worrisome that there is enough internal migration that could rapidly spread drug resistance within Brazil and even reach other countries from the Amazon basin.

4.6.References

Anderson TJ, Haubold B, Williams JT, Estrada-Franco JG, Richardson L, Mollinedo R et al (2000). Microsatellite markers reveal a spectrum of population structures in the malaria parasite Plasmodium falciparum. Molecular biology and evolution **17:** 1467-1482.

Baird JK (2004). Chloroquine resistance in Plasmodium vivax. Antimicrobial agents and chemotherapy 48: 4075-4083.

Bandelt HJ, Forster P, Rohl A (1999). Median-joining networks for inferring intraspecific phylogenies. Molecular biology and evolution **16:** 37-48.

Bastos MS, da Silva-Nunes M, Malafronte RS, Hoffmann EH, Wunderlich G, Moraes SL et al (2007). Antigenic polymorphism and naturally acquired antibodies to Plasmodium vivax merozoite surface protein 1 in rural Amazonians. Clinical and vaccine immunology : CVI 14: 1249-1259.

Beerli P, Felsenstein J (2001). Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. Proceedings of the National Academy of Sciences of the United States of America 98: 4563-4568.

Beerli P (2006). Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. Bioinformatics 22: 341-345.

Beerli P, Palczewski M (2010). Unified framework to evaluate panmixia and migration direction among multiple sampling locations. Genetics 185: 313-326.

Carlton JM, Escalante AA, Neafsey D, Volkman SK (2008). Comparative evolutionary genomics of human malaria parasites. Trends in parasitology 24: 545-550.

Corredor V, Murillo C, Echeverry DF, Benavides J, Pearce RJ, Roper C et al (2010). Origin and dissemination across the Colombian Andes mountain range of sulfadoxine-pyrimethamine resistance in Plasmodium falciparum. Antimicrobial agents and chemotherapy 54: 3121-3125.

Chenet SM, Schneider KA, Villegas L, Escalante AA (2012). Local population structure of Plasmodium: impact on malaria control and elimination. Malaria journal 11: 412.

Cruz Marques A (1987). Human migration and the spread of malaria in Brazil. Parasitology Today 6:166-70.

da Silva-Nunes M, Codeco CT, Malafronte RS, da Silva NS, Juncansen C, Muniz PT et al (2008). Malaria on the Amazonian frontier: transmission dynamics, risk factors, spatial distribution, and prospects for control. The American journal of tropical medicine and hygiene 79: 624-635.

da Silva NS, da Silva-Nunes M, Malafronte RS, Menezes MJ, D'Arcadia RR, Komatsu NT et al (2010). Epidemiology and control of frontier malaria in Brazil: lessons from community-based studies in rural Amazonia. Transactions of the Royal Society of Tropical Medicine and Hygiene 104: 343-350.

de Castro MC, Monte-Mor RL, Sawyer DO, Singer BH (2006). Malaria risk on the Amazon frontier. Proceedings of the National Academy of Sciences of the United States of America 103: 2452-2457.

de Castro MC, Sawyer DO, Singer BH (2007). Spatial patterns of malaria in the Amazon: implications for surveillance and targeted interventions. Health & place **13:** 368-380.

Earl, D.A., von Holdt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources, 4:359–361. Structure harvester available from: http://taylor0.biology.ucla.edu/struct_harvest/

Excoffier, L. and H.E. L. Lischer (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources. 10: 564-567.

Ferreira MU, Karunaweera ND, da Silva-Nunes M, da Silva NS, Wirth DF, Hartl DL (2007). Population structure and transmission dynamics of Plasmodium vivax in rural Amazonia. The Journal of infectious diseases 195: 1218-1226.

Griffing SM, Viana GM, Mixson-Hayden T, Sridaran S, Alam MT, de Oliveira AM et al (2013). Historical shifts in Brazilian P. falciparum population structure and drug resistance alleles. PloS one 8: e58984.

Haubold B, Hudson RR (2000). LIAN 3.0: detecting linkage disequilibrium in multilocus data. Linkage Analysis. Bioinformatics 16: 847-848.

Imwong M, Sudimack D, Pukrittayakamee S, Osorio L, Carlton JM, Day NP et al (2006). Microsatellite variation, repeat array length, and population history of Plasmodium vivax. Molecular biology and evolution 23: 1016-1018.

Karunaweera ND, Ferreira MU, Munasinghe A, Barnwell JW, Collins WE, King CL et al (2008). Extensive microsatellite diversity in the human malaria parasite Plasmodium vivax. Gene 410: 105-112.

McCollum AM, Soberon V, Salas CJ, Santolalla ML, Udhayakumar V, Escalante AA et al (2014). Genetic variation and recurrent parasitaemia in Peruvian Plasmodium vivax populations. Malaria journal 13: 67.

Ministério da Saúde, Resumo Epidemiológico, 2009.

Nash D, Nair S, Mayxay M, Newton PN, Guthmann JP, Nosten F et al (2005). Selection strength and hitchhiking around two anti-malarial resistance genes. Proceedings Biological sciences / The Royal Society 272: 1153-1161.

Oliveira-Ferreira J, Lacerda MV, Brasil P, Ladislau JL, Tauil PL, Daniel-Ribeiro CT (2010). Malaria in Brazil: an overview. Malaria journal 9: 115.

Orjuela-Sanchez P, de Santana Filho FS, Machado-Lima A, Chehuan YF, Costa MR, Alecrim M et al (2009). Analysis of single-nucleotide polymorphisms in the crt-o and mdr1 genes of Plasmodium vivax among chloroquine-resistant isolates from the Brazilian Amazon region. Antimicrobial agents and chemotherapy 53: 3561-3564.

Orjuela-Sanchez P, Karunaweera ND, da Silva-Nunes M, da Silva NS, Scopel KK, Goncalves RM et al (2010). Single-nucleotide polymorphism, linkage disequilibrium and geographic structure in the malaria parasite Plasmodium vivax: prospects for genome-wide association studies. BMC genetics 11: 65.

Park, S.D.E. Trypanotolerance in West African Cattle and the Population Genetic Effects of Selection. PhD thesis. University of Dublin; 2001. Microsatellite tool kit version 3.1.1 available from: http://www.animalgenomics.ucd.ie/sdepark/ms-toolkit/

Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. Genetics 155: 945-959.

Souza-Silva FA, da Silva-Nunes M, Sanchez BA, Ceravolo IP, Malafronte RS, Brito CF et al (2010). Naturally acquired antibodies to Plasmodium vivax Duffy binding protein (DBP) in Brazilian Amazon. The American journal of tropical medicine and hygiene 82: 185-193.

Van den Eede P, Van der Auwera G, Delgado C, Huyse T, Soto-Calle VE, Gamboa D et al (2010). Multilocus genotyping reveals high heterogeneity and strong local population structure of the Plasmodium vivax population in the Peruvian Amazon. Malaria journal 9: 151.

WHO (2012). World Malaria Report. Geneva: World Health Organization.

5. CONCLUSIONS

Malaria control and elimination programs, in the absence of vaccines, rely on interventions aiming to reduce disease transmission. Traditional approaches focus on measuring malaria incidence; however, this approximation has limited power when transmission is low and driven by the movement of high risk groups (e.g. miners and loggers) or when transmission is residual and driven by clusters of cases originated from a local or introduced case. These patterns can only be unveiled by studying the genetic variation of the parasite populations.

Even though only a limited number of loci could be considered for routine analysis, the use of well-chosen microsatellite loci represents a high sensitive method for cluster differentiation and provides a way to deeper analyze patterns of gene flow as well as parasite lineages maintained through time. In general, highly polymorphic loci are suitable to identify individuals within populations whereas more conserved (less polymorphic) loci are useful to establish clusters or parasite populations within a given endemic area.

Low transmission areas exhibit strong linkage disequilibrium (LD) in both *P. falciparum* and *P. vivax* parasites. Although selection due to the fixation of drug resistant haplotypes in *P. falciparum* could account in part for the observed LD, sympatric populations of *P. falciparum* and *P. vivax* in some areas of Latin America exhibited similar characteristics in terms of linkage disequilibrium (LD), number of clusters/clones circulating, and patterns of temporal replacement of clusters. Thus, the local ecology seems to generate LD in malarial parasites, most likely due to inbreeding.

The relative high proportion of similar lineages co-infecting an individual could be explained by mitotic mutations rather than sexual recombination as observed in Turbo, Colombia. Conversely, the presence of lineages with differences in more than two loci, as the ones observed in Brazil, could be explained by either super-infections or co-infections of more dissimilar lineages resulting from sexual recombination.

The highly local haplotypic differentiation observed in South America, implies that there is not significant migration between these areas and that specific local malaria control strategies (considering drug efficacy) should be effective to achieve local elimination.

A Bayesian analysis approach to evaluate changes in effective population size (Ne) is valuable considering that incidence and the *Plasmodium* Ne does not necessarily have a linear relationship. By estimating fluctuations on Ne and evaluating its correlation with the species incidence, we can evaluate if malaria control measures have been effective or not.

In the Amazon basin, local *P. vivax* populations are highly diverse and more prone than *P. falciparum* to present different lineages co-infecting the same individual. Particular biological characteristics in *P. vivax* such as relapses and broader temperature tolerance could explain the difference in prevalence and gene flow with sympatric *P. falciparum* populations.

The estimation of Ne and migration rates allowed us to identify patterns consistent with high local incidence that likely resulted from the influx of imported (migrant) cases. Thus, a molecular surveillance approach can be used to distinguish malaria sources and also distinguish between autochthonous and foreign cases.

This study allowed us to incorporate knowledge on population clonal structure circulating through time or space in *Plasmodium* surveillance and drug efficacy studies. Moreover, the methodology presented allows us to assess changes in malaria transmission to more efficiently manage resources for control measures in locations that act as parasite "sources". Overall, our results stressed the importance of monitoring malaria demographic changes when assessing the success of elimination programs in areas of low transmission.

REFERENCES

Alonso, P.L., Brown, G., Arevalo-Herrera, M., Binka, F., Chitnis, C., Collins, F., Doumbo, O.K., Greenwood, B., Hall, B.F., Levine, M.M., Mendis, K., Newman, R.D., Plowe, C.V., Rodríguez, M.H., Sinden, R., Slutsker, L., Tanner, M. (2011). A research agenda to underpin malaria eradication. PLoS Medicine, 8: e1000406.

Anderson TJ, Haubold B, Williams JT, Estrada-Franco JG, Richardson L, Mollinedo R et al (2000). Microsatellite markers reveal a spectrum of population structures in the malaria parasite Plasmodium falciparum. Molecular biology and evolution 17: 1467-1482.

Anderson TJ, Su XZ, Bockarie M, Lagog M, Day KP (1999). Twelve microsatellite markers for characterization of Plasmodium falciparum from finger-prick blood samples. Parasitology 119 (Pt 2): 113-125.

Arnott, A., Barry, A.E., Reeder, J.C. (2012). Understanding the population genetics of Plasmodium vivax is essential for malaria control and elimination. Malaria Journal, 10:11.

Baird JK (2004). Chloroquine resistance in Plasmodium vivax. Antimicrobial agents and chemotherapy 48: 4075-4083.

Bandelt HJ, Dress AW (1992). Split decomposition: a new and useful approach to phylogenetic analysis of distance data. Molecular phylogenetics and evolution 1: 242-252.

Barton, D.E., David, F. (1957). Multiple runs. Biometrika, 44:168–170.

Bastos MS, da Silva-Nunes M, Malafronte RS, Hoffmann EH, Wunderlich G, Moraes SL et al (2007). Antigenic polymorphism and naturally acquired antibodies to Plasmodium vivax merozoite surface protein 1 in rural Amazonians. Clinical and vaccine immunology : CVI 14: 1249-1259.

Beerli P (2006). Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. Bioinformatics 22: 341-345.

Beerli P, Felsenstein J (2001). Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. Proceedings of the National Academy of Sciences of the United States of America 98: 4563-4568.

Beerli P, Palczewski M (2010). Unified framework to evaluate panmixia and migration direction among multiple sampling locations. Genetics 185: 313-326.

Bejon, P., Williams, T.N., Liljander, A., Noor, A.M., Wambua, J., Ogada, E., Olotu, A., Osier, F.H., Hay, S.I., Färnert, A., Marsh, K. (2010). Stable and Unstable Malaria Hotspots in Longitudinal Cohort Studies in Kenya. PLoS Med, 7(7): e1000304. doi:10.1371/journal.pmed.1000304.

Blair S, Carmona-Fonseca J, Pineros JG, Rios A, Alvarez T, Alvarez G et al (2006). Therapeutic efficacy test in malaria falciparum in Antioquia, Colombia. Malaria journal 5: 14.

Bousema, T., & Drakeley, C. (2011). Epidemiology and infectivity of Plasmodium falciparum and Plasmodium vivax gametocytes in relation to malaria control and elimination. Clinical Microbiology Review, 24: 377–410.

Campino, S., Auburn, S., Kivinen, K., Zongo, I., Ouedraogo, J.B., Mangano, V., Djimde, A., Doumbo, O.K., Kiara, S.M., Nzila, A., Borrmann, S., Marsh, K., Michon, P., Mueller, I., Siba, P., Jiang, H., Su, X.Z., Amaratunga, C., Socheat, D., Fairhurst, R.M., Imwong, M., Anderson, T., Nosten, F., White, N.J., Gwilliam, R., Deloukas, P., MacInnis, B., Newbold, C.I., Rockett, K., Clark, T.G., Kwiatkowski, DP. (2011). Population genetic analysis of Plasmodium falciparum parasites using acustomized Illumina GoldenGate genotyping assay. PLoS One, 6(6):e20251. doi: 10.1371/journal.pone.0020251.

Carlton JM, Escalante AA, Neafsey D, Volkman SK (2008). Comparative evolutionary genomics of human malaria parasites. Trends in parasitology 24: 545-550.

Carter, R, Mendis KN, Roberts, D. (2000). Bulletin of the World Health Organization. World Health Organization, 78 (12), 1401-1411.

Chang HH, Moss EL, Park DJ, Ndiaye D, Mboup S, Volkman SK et al (2013). Malaria life cycle intensifies both natural selection and random genetic drift. Proceedings of the National Academy of Sciences of the United States of America 110: 20129-20134.

Cheeseman, I.H., Miller, B.A., Nair, S., Nkhoma, S., Tan, A., Tan, J.C., Al Saai, S., Phyo, A.P., Moo, C.L., Lwin, K.M., McGready, R., Ashley, E., Imwong, M., Stepniewska, K., Yi, P., Dondorp, A.M., Mayxay, M., Newton, P.N., White, N.J., Nosten, F., Ferdig, M.T., Anderson, T.J. (2012). A major genome region underlying artemisinin resistance in malaria. Science, 336:79–82.

Chenet SM, Schneider KA, Villegas L, Escalante AA (2012). Local population structure of Plasmodium: impact on malaria control and elimination. Malaria journal 11: 412.

Chenet, S.M., Tapia, L.L., Escalante, A.A., Durand, S., Lucas, C., Bacon, D.J. (2012). Genetic diversity and population structure of genes encoding vaccine candidate antigens of Plasmodium vivax. Malaria Journal, 11:68.

Cibulskis, R.E., Aregawi, M., Williams, R., Otten, M., Dye, C. (2011). Worldwide incidence of malaria in 2009: estimates, time trends, and a critique of methods. PLoS Medicine, 8(12):e1001142. doi: 10.1371/journal.pmed.1001142.

Corran, P., Coleman, P., Riley, E., Drakeley, C. (2007) Serology: a robust indicator of malaria transmission intensity? Trends in Parasitology, 23:575-82.

Corredor V, Murillo C, Echeverry DF, Benavides J, Pearce RJ, Roper C et al (2010). Origin and dissemination across the Colombian Andes mountain range of sulfadoxine-pyrimethamine resistance in Plasmodium falciparum. Antimicrobial agents and chemotherapy 54: 3121-3125.

Cotter, C., Sturrock, H.J., Hsiang, M.S., Liu, J., Phillips, A.A., Hwang, J., Gueye, C.S., Fullman, N., Gosling, R.D., Feachem, R.G. (2013). The changing epidemiology of malaria elimination: new strategies for new challenges. Lancet, 13: 60310-4. doi: 10.1016/S0140-6736(13)60310-4.

Crowell, V., Yukich, J.O., Briët, O.J.T., Ross, A., Smith, T.A. (2013). A Novel Approach for Measuring the Burden of Uncomplicated Plasmodium falciparum Malaria: Application to Data from Zambia. PLoS ONE, 8(2): e57297. doi:10.1371/journal.pone.0057297.

Cruz LR, Spangenberg T, Lacerda MV, Wells TN (2013). Malaria in South America: a drug discovery perspective. Malaria journal 12: 168.

Cruz Marques A (1987). Human migration and the spread of malaria in Brazil. Parasitology Today 6:166-70.

Da Silva-Nunes M, Codeco CT, Malafronte RS, da Silva NS, Juncansen C, Muniz PT et al (2008). Malaria on the Amazonian frontier: transmission dynamics, risk factors, spatial distribution, and prospects for control. The American journal of tropical medicine and hygiene 79: 624-635.

Da Silva NS, da Silva-Nunes M, Malafronte RS, Menezes MJ, D'Arcadia RR, Komatsu NT et al (2010). Epidemiology and control of frontier malaria in Brazil: lessons from community-based studies in rural Amazonia. Transactions of the Royal Society of Tropical Medicine and Hygiene 104: 343-350.

Da Silva-Nunes, M., Moreno, M., Conn, J.E., Gamboa, D., Abeles, S., Vinetz, J.M., Ferreira, M.U. (2012). Amazonian malaria: Asymptomatic human reservoirs, diagnostic challenges, environmentally driven changes in mosquito vector populations, and the mandate for sustainable control strategies. Acta Tropica, 121: 281-291.

De Castro MC, Monte-Mor RL, Sawyer DO, Singer BH (2006). Malaria risk on the Amazon frontier. Proceedings of the National Academy of Sciences of the United States of America 103: 2452-2457.

De Castro MC, Sawyer DO, Singer BH (2007). Spatial patterns of malaria in the Amazon: implications for surveillance and targeted interventions. Health & place 13: 368-380.

Drakeley, C., Schellenberg, D., Kihonda, J., Sousa, C.A., Arez, A.P., Lopes, D., Lines, J., Mshinda, H., Lengeler, C., Armstrong, Schellenberg, J., Tanner, M., Alonso, P. (2013). An estimation of the entomological inoculation rate for Ifakara: a semi-urban area in a region of intense malaria transmission in Tanzania. Tropical Medicine & International Health, 8(9):767-74.

Drummond AJ, Rambaut A (2007). BEAST: Bayesian evolutionary analysis by sampling trees. BMC evolutionary biology 7: 214.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular biology and evolution 29: 1969-1973.

Earl, D.A., von Holdt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources, 4:359–361. Structure harvester available from:http://taylor0.biology.ucla.edu/struct_harvest/

Echeverry DF, Nair S, Osorio L, Menon S, Murillo C, Anderson TJ (2013). Long term persistence of clonal malaria parasite Plasmodium falciparum lineages in the Colombian Pacific region. BMC genetics 14: 2.

Excoffier, L. and H.E. L. Lischer (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources. 10: 564-567.

Ferreira MU, Karunaweera ND, da Silva-Nunes M, da Silva NS, Wirth DF, Hartl DL. (2007). Population structure and transmission dynamics of Plasmodium vivax in rural Amazonia. Journal of Infectious Diseases, 195:1218–1226.

Freeman, G.H., Halton, J.H. (1951). Note on an Exact Treatment of Contingency, Goodness of Fit and Other Problems of Significance. Biometrika, 38:141–149.

Gamboa, D., Ho, M.F., Bendezu, J., Torres, K., Chiodini, P.L., Barnwell, J.W., Incardona, S., Perkins, M., Bell, D., McCarthy, J., Cheng, Q. (2010). A large proportion of P. falciparum isolates in the Amazon region of Peru lack pfhrp2 and pfhrp3: implications for malaria rapid diagnostic tests. PLoS One, 5(1):e8091. doi: 10.1371/journal.pone.0008091.

Gauthier, C., Tibayrenc, M. (2005). Population structure of malaria parasites: the driving epidemiological forces. Acta Tropica, 3:241–250.

Gonzalez IJ, Padilla JO, Giraldo LE, Saravia NG (2003). [Efficacy of amodiaquine and sulfadoxine/pyrimethamine in the treatment of malaria not complicated by Plasmodium falciparum in Narino, Colombia, 1999-2002]. Biomedica : revista del Instituto Nacional de Salud 23: 38-46.

Goudet, J. (1995). FSTAT Version 1.2: a computer program to calculate F-statistics. J. Heredity, 86:485–486. FSTAT 2.9.3.2 available at: http://www2.unil.ch/popgen/softwares/fstat.htm.

Greenhouse, B., Myrick, A., Dokomajilar, C., Woo, J.M., Carlson, E.J., Rosenthal, P.J., Dorsey, G. (2006). Validation of microsatellite markers for use in genotyping polyclonal Plasmodium falciparum infections. American Journal of Tropical Medicine & Hygiene, 75:836–842.

Griffing S, Syphard L, Sridaran S, McCollum AM, Mixson-Hayden T, Vinayak S et al (2010). pfmdr1 amplification and fixation of pfcrt chloroquine resistance alleles in Plasmodium falciparum in Venezuela. Antimicrobial agents and chemotherapy 54: 1572-1579.

Griffing SM, Viana GM, Mixson-Hayden T, Sridaran S, Alam MT, de Oliveira AM et al (2013). Historical shifts in Brazilian P. falciparum population structure and drug resistance alleles. PloS one 8: e58984.

Griffing, S.M., Mixson-Hayden, T., Sridaran, S., Alam, M.T., McCollum, A.M., Cabezas, C., Marquiño Quezada, W., Barnwell, J.W., De Oliveira, A.M., Lucas, C., Arrospide, N., Escalante, A.A., Bacon, D.J., Udhayakumar, V. (2011). South American Plasmodium falciparum after the malaria eradication era: clonal population expansion and survival of the fittest hybrids. PLoS One, 6:e23486. doi: 10.1371.

Griffing, S.M., Mixson-Hayden, T., Sridaran, S., Alam, M.T., McCollum, A.M., Cabezas, C., Marquiño Quezada, W,, Barnwell, J.W., De Oliveira, A.M., Lucas, C., Arrospide, N., Escalante, A.A., Bacon, D.J., Udhayakumar, V. (2011). South American Plasmodium falciparum after the malaria eradication era: clonal population expansion and survival of the fittest hybrids. PLoS One, 6:e23486.

Haubold, B., Hudson, R.R. (2000). LIAN 3.0: detecting linkage disequilibrium in multilocus data. Bioinformatics, 16:847–848. LIAN 3.5 available from: http://adenine.biz.fh-weihenstephan.de/cgi-bin/lian/lian.cgi.pl.

Hay, S.I., Guerra, C.A., Gething, P.W., Patil, A.P., Tatem, A.J., Noor, A.M., Kabaria, C.W., Manh, B.H., Elyazar, I.R., Brooker, S., Smith, D.L., Moyeed, R.A., Snow, R.W. (2009). A world malaria map: Plasmodium falciparum endemicity in 2007. PLoS Med, 6:e1000048. doi: 10.1371/journal.pmed.1000048.

Hernandez DC, Guerra AP, Cucunuba ZM, Nicholls RS, Barrera SM (2013). Haplotypes associated with resistance to sulfadoxine-pyrimethamine in Plasmodium falciparum in two malaria endemic locations in Colombia. Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases 18: 183-190.

Howes, R.E., Battle, K.E., Satyagraha, A.W., Baird, J.K., Hay, S.I. (2013). G6PD deficiency: global distribution, genetic variants and primaquine therapy. Advances in Parasitology, 81:133-201.

Imwong M, Sudimack D, Pukrittayakamee S, Osorio L, Carlton JM, Day NP et al (2006). Microsatellite variation, repeat array length, and population history of Plasmodium vivax. Molecular biology and evolution 23: 1016-1018.

Imwong, M., Nair, S., Pukrittayakamee, S., Sudimack, D., Williams, J.T., Mayxay, M., Newton, P.N., Kim, J.R., Nandy, A., Osorio, L., Carlton, J.M., White, N.J., Day, N.P., Anderson, T.J. (2004). Contrasting genetic structure in Plasmodium vivax populations from Asia and South America. Journal of Infectious Diseases, 190:1547-55.

Iwagami, M., Fukumoto, M., Hwang, S.Y., Kim, S.H., Kho, W.G., Kano, S. (2012). Population structure and transmission dynamics of Plasmodium vivax in the Republic of Korea based on microsatellite DNA analysis. PLoS Neglected Tropical Diseases, 6:e1592.

Jakobsson, M., Rosenberg, N.A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics, 23:1801–1806. CLUMPP 1.1.2 available from: http://www.stanford.edu/group/rosenberglab/clumpp.html.

Karunaweera ND, Ferreira MU, Munasinghe A, Barnwell JW, Collins WE, King CL et al (2008). Extensive microsatellite diversity in the human malaria parasite Plasmodium vivax. Gene 410: 105-112.

Kimura M, Crow JF (1964). The Number of Alleles that can be maintained in a finite population. Genetics 49: 725-738.

Kimura M, Ohta T (1978). Stepwise mutation model and distribution of allelic frequencies in a finite population. Proceedings of the National Academy of Sciences of the United States of America 75: 2868-2872.

Koita, O.A., Doumbo, O.K., Ouattara, A., Tall, L.K., Konaré, A., Diakité, M., Diallo, M., Sagara, I., Masinde, G.L., Doumbo, S.N., Dolo, A., Tounkara, A., Traoré, I., Krogstad, D.J. (2012). False-negative rapid diagnostic tests for malaria and deletion of the histidine-rich repeat region of the hrp2 gene. American Journal of Tropical Medicine & Hygiene, 86: 194–198.

Laval, E.G., Schneider S. (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evolutionary Bioinformatics, 1:47–50.

Lucchi, N.W., Demas, A., Narayanan, J., Sumari, D., Kabanywanyi, A., Kachur, S.P., Barnwell, J.W., Udhayakumar, V. (2010). Real-time fluorescence loop mediated isothermal amplification for the diagnosis of malaria. PLoS One, 5:e13733. doi: 10.1371/journal.pone.0013733.

Maruyama, T. In: Molecular Evolution, Protein Polymorphism and the Neutral Theory. Kimura, M., editor. Tokyo: Japan Scientific Societies Press; 1982. Stochastic integrals and their application to population genetics, 151–166.

McCollum AM, Mueller K, Villegas L, Udhayakumar V, Escalante AA (2007). Common origin and fixation of Plasmodium falciparum dhfr and dhps mutations associated with sulfadoxine-pyrimethamine resistance in a low-transmission area in South America. Antimicrobial agents and chemotherapy 51: 2085-2091.

McCollum AM, Soberon V, Salas CJ, Santolalla ML, Udhayakumar V, Escalante AA et al (2014). Genetic variation and recurrent parasitaemia in Peruvian Plasmodium vivax populations. Malaria journal 13: 67.

McCollum, A.M., Basco, L.K., Tahar, R., Udhayakumar, V., Escalante, A.A. (2008). Hitchhiking and selective sweeps of Plasmodium falciparum sulfadoxine and pyrimethamine resistance alleles in a population from central Africa. Antimicrobial Agents and Chemotherapy, 52:4089–4097.

McCollum, A.M., Mueller, K., Villegas, L., Udhayakumar, V., Escalante, A.A. (2007). Common origin and fixation of Plasmodium falciparum dhfr and dhps mutations associated with sulfadoxine-pyrimethamine resistance in a low-transmission area in South America. Antimicrobial Agents and Chemotherapy, 51:2085–2091.

McCollum, A.M., Schneider, K.A., Griffing, S.M., Zhou, Z., Kariuki, S., Ter-Kuile, F., Shi, Y.P., Slutsker, L., Lal, A.A., Udhayakumar, V., Escalante, A.A. (2012). Differences in selective pressure on dhps and dhfr drug resistant mutations in western Kenya. Malaria Journal, 11:77.

Meirmans, P.G. (2012) The trouble with isolation by distance. Molecular Ecology, 21:2839–46.

Ministério da Saúde, Resumo Epidemiológico, 2009.

Moreno, J.E., Rubio-Palis, Y., Páez, E., Pérez, E., Sánchez, V. (2007). Abundance, biting behaviour and parous rate of anopheline mosquito species in relation to malaria incidence in gold-mining areas of southern Venezuela. Medical and Veterinary Entomology, 21:339–349.

Mu, J., Awadalla, P., Duan, J., McGee, K.M., Joy, D.A., McVean, G.A., Su, X.Z. (2005) Recombination hotspots and population structure in Plasmodium falciparum. PLoS Biology, 3:e335.

Mwangi, J.M., Omar, S.A., Ranford-Cartwright, L.C. (2006). Comparison of microsatellite and antigen-coding loci for differentiating recrudescing Plasmodium falciparum infections from reinfections in Kenya. International Journal of Parasitology, 36:329–336.

Nash, D., Nair, S., Mayxay, M., Newton, P.N., Guthmann, J.P., Nosten, F., Anderson, T.J. (2005). Selection strength and hitchhiking around two anti-malarial resistance genes. Proceedings of the Royal Society B: Biological Sciences, 272:1153–1161.

Neafsey, D.E., Schaffner, S.F., Volkman, S.K., Park, D., Montgomery, P., Milner. D.A. Jr, Lukens, A., Rosen, D., Daniels, R., Houde, N., Cortese, J.F., Tyndall, E., Gates, C., Stange-Thomann, N., Sarr, O., Ndiaye, D., Ndir, O., Mboup, S., Ferreira, M.U., Moraes Sdo, L., Dash, A.P., Chitnis, C.E., Wiegand, R.C., Hartl, D.L., Birren, B.W., Lander, E.S., Sabeti, P.C., Wirth, D.F. (2008) Genome-wide SNP genotyping highlights the role of natural selection in Plasmodium falciparum population divergence. Genome Biology, 9:R171.

Nyachieo, A., Van Overmeir, C., Laurent, T., Dujardin, J.C., D'Alessandro, U. (2005). Plasmodium falciparum genotyping by microsatellites as a method to distinguish between recrudescent and new infections. American Journal of Tropical Medicine & Hygiene, 73:210–3.

Oliveira-Ferreira J, Lacerda MV, Brasil P, Ladislau JL, Tauil PL, Daniel-Ribeiro CT (2010). Malaria in Brazil: an overview. Malaria journal 9: 115.

Orjuela-Sanchez P, de Santana Filho FS, Machado-Lima A, Chehuan YF, Costa MR, Alecrim M et al (2009). Analysis of single-nucleotide polymorphisms in the crt-o and mdr1 genes of Plasmodium vivax among chloroquine-resistant isolates from the Brazilian Amazon region. Antimicrobial agents and chemotherapy 53: 3561-3564.

Orjuela-Sanchez P, Karunaweera ND, da Silva-Nunes M, da Silva NS, Scopel KK, Goncalves RM et al (2010). Single-nucleotide polymorphism, linkage disequilibrium and geographic structure in the malaria parasite Plasmodium vivax: prospects for genome-wide association studies. BMC genetics 11: 65.

Orjuela-Sánchez, P., Brandi, M.C., Ferreira, M.U. (2013). Microsatellite Analysis of Malaria Parasites. Methods in Molecular Biology, 1006: 247-258.

Orjuela-Sánchez, P., da Silva, N.S., da Silva-Nunes, M., Ferreira, M.U. (2009). Recurrent parasitemias and population dynamics of Plasmodium vivax polymorphisms in rural Amazonia. American Journal of Tropical Medicine & Hygiene, 81:961–968.

Osorio L, Perez Ldel P, Gonzalez IJ (2007). [Assessment of the efficacy of antimalarial drugs in Tarapaca, in the Colombian Amazon basin]. Biomedica : revista del Instituto Nacional de Salud 27: 133-140.

Park, S.D.E. Trypanotolerance in West African Cattle and the Population Genetic Effects of Selection. PhD thesis. University of Dublin; 2001. Microsatellite tool kit version 3.1.1 available from: http://www.animalgenomics.ucd.ie/sdepark/ms-toolkit/

Payseur, B.A., Jing, P., Haasl, R.J. (2011). A genomic portrait of human microsatellite variation. Molecular Biology and Evolution, 28:303–312.

Perez MA, Cortes LJ, Guerra AP, Knudson A, Usta C, Nicholls RS (2008). [Efficacy of the amodiaquine+sulfadoxine-pyrimethamine combination and of chloroquine for the treatment of malaria in Cordoba, Colombia, 2006]. Biomedica : revista del Instituto Nacional de Salud 28: 148-159.

Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. Genetics 155: 945-959.

Pumpaibool, T., Arnathau, C., Durand, P., Kanchanakhan, N., Siripoon, N., Suegorn, A., Sitthi-Amorn, C., Renaud, F., Harnyuttanakorn, P. (2009). Genetic diversity and population structure of Plasmodium falciparum in Thailand, a low transmission country. Malar Journal, 8:155.

Razakandrainibe FG, Durand P, Koella JC, De Meeüs T, Rousset F, Ayala FJ, Renaud F. (2005). "Clonal" population structure of the malaria agent Plasmodium falciparum in high-infection regions. Proceedings of the National Academy of Sciences, 102:17388–93.

Restrepo, E., Imwong, M., Rojas, W., Carmona-Fonseca, J., Maestre, A. (2011). High genetic polymorphism of relapsing P. vivax isolates in northwest Colombia. Acta Tropica, 119:23–29.

Rodriguez JC, Uribe GA, Araujo RM, Narvaez PC, Valencia SH (2011). Epidemiology and control of malaria in Colombia. Memorias do Instituto Oswaldo Cruz 106 Suppl 1: 114-122.

Rosenberg, N.A. (2004). Distruct: a program for the graphical display of population structure. Molecular Ecology Notes, 4:137–138. Distruct 1.1 available at: http://www.stanford.edu/group/rosenberglab/distruct.html.

Russell, B., Suwanarusk, R., Lek-Uthai, U. (2006). Plasmodium vivax genetic diversity: microsatellite length matters. Trends in Parasitology, 22:399–401.

Sainudiin R, Durrett RT, Aquadro CF, Nielsen R (2004). Microsatellite mutation models: insights from a comparison of humans and chimpanzees. Genetics 168: 383-395.

Schlotterer C (2000). Evolutionary dynamics of microsatellite DNA. Chromosoma 109: 365-371.

Schneider, K.A., Kim, Y. (2010). An analytical model for genetic hitchhiking in the evolution of antimalarial drug resistance. Theoretical Population Biology, 78:93–108.

Schneider, K.A., Kim, Y. (2011). Approximations for the hitchhiking effect caused by the evolution of antimalarial-drug resistance. Journal of Mathematical Biology, 62:789–832.

Shanks, G.D. (2012). Control and elimination of Plasmodium vivax. Advances in Parasitology, 80:301-41.

Smith, D.L., Dushoff, J., McKenzie, F.E. (2004). The risk of a mosquito-borne infection in a heterogeneous environment. PLoS Biol, 2(11):e368.

Souza-Silva FA, da Silva-Nunes M, Sanchez BA, Ceravolo IP, Malafronte RS, Brito CF et al (2010). Naturally acquired antibodies to Plasmodium vivax Duffy binding protein (DBP) in Brazilian Amazon. The American journal of tropical medicine and hygiene 82: 185-193.

Su X, Ferdig MT, Huang Y, Huynh CQ, Liu A, You J et al (1999). A genetic map and recombination parameters of the human malaria parasite Plasmodium falciparum. Science 286: 1351-1353.

Su, X., Hayton, K., Wellems, T.E. (2007). Genetic linkage and association analyses for trait mapping in Plasmodium falciparum. Nature Reviews Genetics, 8(7):497-506.

Sunnucks, P. (2000). Efficient genetic markers for population biology. Trends in Ecology and Evolution, 15:199–203.

Sutton PL, Torres LP, Branch OH (2011). Sexual recombination is a signature of a persisting malaria epidemic in Peru. Malaria journal 10: 329.

Taylor JE, Pacheco MA, Bacon DJ, Beg MA, Machado RL, Fairhurst RM et al (2013). The evolutionary history of Plasmodium vivax as inferred from mitochondrial genomes: parasite genetic diversity in the Americas. Molecular biology and evolution 30: 2050-2064.

The malERA Consultative Group on Diagnoses and Diagnostics. (2011). A research agenda for malaria eradication: diagnoses and diagnostics. PLoS Medicine, 8: e1000396.

Urdaneta L, Lal A, Barnabe C, Oury B, Goldman I, Ayala FJ, Tibayrenc M. (2001). Evidence for clonal propagation in natural isolates of Plasmodium falciparum from Venezuela. Proceedings of the National Academy of Sciences, 98:6725–6729.

Van den Eede, P., Van der Auwera, G., Delgado, C., Huyse, T., Soto-Calle, V.E., Gamboa, D., Grande, T., Rodriguez, H., Llanos, A., Anné, J., Erhart, A., D'Alessandro, U. (2010) Multilocus genotyping reveals high heterogeneity and strong local population structure of the Plasmodium vivax population in the Peruvian Amazon. Malaria Journal, 9:151.

Vinayak S, Alam MT, Mixson-Hayden T, McCollum AM, Sem R, Shah NK et al (2010). Origin and evolution of sulfadoxine resistant Plasmodium falciparum. PLoS pathogens 6: e1000830.

Volkman, S.K., Sabeti, P.C., DeCaprio, D., Neafsey, D.E., Schaffner, S.F., Milner, D.A. Jr, Daily, J.P., Sarr, O., Ndiaye, D., Ndir, O., Mboup, S., Duraisingh, M.T., Lukens, A., Derr, A., Stange-Thomann, N., Waggoner, S., Onofrio, R., Ziaugra, L., Mauceli, E., Gnerre, S., Jaffe, D.B., Zainoun, J., Wiegand, R.C., Birren, B.W., Hartl, D.L., Galagan, J.E., Lander, E.S., Wirth. D.F. (2007). A genome-wide map of diversity in Plasmodium falciparum. Nature Genetics, 39:113–119.

Wellems TE, Plowe CV (2001). Chloroquine-resistant malaria. The Journal of infectious diseases 184: 770-776.

WHO (2007). Malaria elimination. A field manual for low and moderate endemic countries.

WHO (2012a). World Malaria Report. Geneva: World Health Organization.

WHO (2012b). Disease surveillance for malaria control: operational manual.

Wootton, J.C., Feng, X., Ferdig, M.T., Cooper, R.A., Mu, J., Baruch, D.I., Magill, A.J., Su, X.Z. (2002). Genetic diversity and chloroquine selective sweeps in Plasmodium falciparum. Nature, 418:320–323.

Wright S. Isolation by distance. (1943). Genetics, 2:114–38.

Wu CH, Drummond AJ (2011). Joint inference of microsatellite mutation models, population history and genealogies using transdimensional Markov Chain Monte Carlo. Genetics 188: 151-164.

www.cdc.gov/malaria

APPENDIX A

HAPLOTYPES INFERRED FROM THE NEUTRAL MICROSATELLITE LOCI

No.	Frequency	S.D	Polya	TAA60	ARA2	pfg377	PfPK2	TAA109	TAA81	TAA42	2490
1	0.02	0.009	158	79	65	95	173	162	121	185	74
2	0.004	0.004	158	79	65	95	173	162	121	185	78
3	0.004	0.004	158	79	65	95	173	162	127	185	78
4	0.004	0.004	148	79	69	95	162	162	127	185	86
5	0.004	0.004	162	79	69	95	173	162	127	185	78
6	0.037	0.012	148	79	69	95	173	162	121	185	81
7	0.004	0.004	162	79	55	95	173	162	121	185	86
8	0.004	0.004	162	79	55	95	173	162	127	185	86
9	0.008	0.006	162	79	55	95	162	162	121	185	86
10	0.008	0.006	148	79	65	95	162	162	127	185	86
11	0.004	0.004	158	79	65	95	171	162	121	185	86
12	0.004	0.004	162	79	69	95	173	162	127	185	86
13	0.008	0.006	158	79	65	95	173	162	127	185	81
14	0.004	0.004	148	79	65	95	162	177	127	185	86
15	0.004	0.004	148	79	69	95	173	162	127	185	86
16	0.004	0.004	148	79	69	95	173	162	121	185	86
17	0.016	0.008	162	79	69	95	162	162	121	185	81
18	0.106	0.02	148	79	69	95	173	162	127	185	81
19	0.142	0.022	148	79	65	95	162	162	127	185	81
20	0.004	0.004	158	79	69	95	162	162	127	185	74
21	0.028	0.011	162	79	55	95	162	162	121	185	81
22	0.02	0.009	148	79	65	95	168	162	121	185	74
23	0.004	0.004	158	79	65	95	171	162	121	185	74
24	0.012	0.007	148	79	55	95	162	162	121	185	81
25	0.004	0.004	158	79	65	95	173	162	121	185	81
26	0.004	0.004	162	79	55	95	162	162	121	185	74
27	0.004	0.004	162	79	55	95	162	162	121	185	78
28	0.024	0.01	162	79	69	95	173	162	127	185	81
29	0.012	0.007	162	79	55	95	173	162	121	185	74
30	0.008	0.006	162	79	65	95	162	162	121	185	81
31	0.008	0.006	148	79	55	95	173	162	127	185	81
32	0.004	0.004	148	79	65	95	162	162	127	185	74
33	0.004	0.004	162	79	55	95	173	162	127	185	81
34	0.004	0.004	148	79	69	95	162	162	121	185	81
35	0.004	0.004	162	79	65	95	173	162	121	185	74
36	0.012	0.007	162	79	65	95	173	162	127	185	81
37	0.004	0.004	162	79	65	95	162	162	127	185	81
38	0.065	0.016	162	79	65	95	173	162	127	185	74
39	0.004	0.004	158	79	65	95	173	162	127	185	74

40	0.024	0.01	148	79	65	95	158	162	127	185	81
41	0.012	0.007	148	79	55	95	158	162	121	185	81
42	0.008	0.006	151	79	65	95	162	162	127	185	81
43	0.008	0.006	151	79	65	95	158	162	127	185	81
44	0.012	0.007	148	79	69	95	171	162	127	185	81
45	0.012	0.007	151	79	69	95	171	162	127	185	81
46	0.008	0.006	146	79	69	95	171	162	127	185	81
47	0.004	0.004	162	79	69	95	171	162	127	185	81
48	0.008	0.006	151	79	69	95	171	162	121	185	81
49	0.004	0.004	151	79	55	95	162	162	121	185	81
50	0.004	0.004	162	79	65	95	158	162	127	185	74
51	0.004	0.004	151	79	69	95	173	162	121	185	81
52	0.008	0.006	148	79	69	95	171	162	121	185	81
53	0.004	0.004	155	79	57	95	171	162	127	185	74
54	0.008	0.006	162	79	65	95	171	162	121	185	74
55	0.004	0.004	148	79	65	95	160	162	127	185	81
56	0.004	0.004	148	79	65	95	175	162	127	185	81
57	0.004	0.004	148	79	65	95	171	162	127	185	81
58	0.045	0.013	148	79	69	95	173	162	121	185	74
59	0.028	0.011	148	79	69	95	171	162	121	185	74
60	0.016	0.008	162	79	65	95	171	162	127	185	74
61	0.004	0.004	151	79	69	95	173	162	121	185	74
62	0.008	0.006	151	79	69	95	171	162	121	185	74
63	0.008	0.006	148	79	65	95	162	162	127	185	78
64	0.045	0.013	148	79	65	95	173	162	127	185	81
65	0.004	0.004	148	79	65	95	173	162	127	185	84
66	0.004	0.004	148	79	65	95	162	162	115	185	81
67	0.004	0.004	158	79	69	95	162	162	121	185	74
68	0.004	0.004	148	81	65	95	173	162	127	185	81
69	0.004	0.004	148	79	65	95	173	162	127	182	81
70	0.016	0.008	148	79	65	95	173	162	121	185	81
71	0.004	0.004	162	79	65	95	173	162	127	182	74
72	0.004	0.004	148	79	65	95	165	162	121	185	74
73	0.016	0.008	158	79	69	95	162	162	127	185	81
74	0.004	0.004	158	79	69	95	162	162	127	182	81
75	0.004	0.004	148	79	65	95	173	162	121	182	81
76	0.004	0.004	162	79	65	95	162	162	121	185	74
77	0.004	0.004	148	79	69	95	173	162	118	185	81
78	0.004	0.004	148	79	65	95	162	162	124	185	78
79	0.004	0.004	148	79	65	92	162	162	127	185	78
*			181	79	65	102	191	189	115	185	84

*Migrant form Central America. Haplotypes in gray were found in mixed infections.

APPENDIX B

CRT LINKED MICROSATELLITE LOCI AND MUTATION SITES

No.	Freq.	-257	-200	-45	-4.8	-4.5	4.6	7	48	60	245	72	73	74	75	76
1	0.005	195	180	114	183	233	157	307	114	126	191	C	V	M	E	T
2	0.01	195	182	111	183	228	157	307	114	126	191	C	V	M	E	T
3	0.005	195	180	114	183	228	157	307	114	123	191	C	V	M	E	T
4	0.005	195	180	114	183	228	157	307	114	123	181	C	V	M	E	T
5	0.168	195	180	114	183	228	157	307	114	140	181	C	V	M	E	T
6	0.021	181	177	114	183	228	157	307	114	140	191	C	V	M	E	T
7	0.225	195	177	114	183	228	157	307	114	126	191	C	V	M	E	T
8	0.016	195	177	114	183	228	157	307	114	126	181	C	V	M	E	T
9	0.089	195	177	114	185	228	157	307	114	126	191	C	V	M	E	T
10	0.304	195	180	114	183	228	157	307	114	126	191	C	V	I	E	T
11	0.005	179	177	114	183	228	157	307	114	140	191	C	V	M	E	T
12	0.016	195	177	114	185	230	157	307	114	126	191	C	V	M	E	T
13	0.01	195	182	114	183	228	157	307	114	126	191	C	V	M	E	T
14	0.01	179	177	114	183	228	157	307	114	126	181	C	V	M	E	T
15	0.005	195	177	114	183	228	157	307	114	140	191	C	V	M	E	T
16	0.005	195	177	114	183	228	157	307	114	126	192	C	V	M	E	T
17	0.01	195	180	114	183	230	157	307	114	140	181	C	V	M	E	T
18	0.01	195	180	114	183	228	157	307	114	126	181	C	V	M	E	T
19	0.005	181	180	114	183	228	157	307	114	140	191	C	V	M	E	T
20	0.005	196	180	114	183	228	157	307	114	126	181	C	V	M	E	T
21	0.016	195	180	114	183	230	157	307	114	126	191	C	V	M	E	T
22	0.021	195	180	114	183	228	157	307	114	147	181	C	V	M	E	T
23	0.005	195	180	114	183	228	157	307	114	137	182	C	V	M	E	T
24	0.005	195	180	111	183	228	157	307	114	140	181	C	V	M	E	T
25	0.005	181	177	114	183	228	157	307	114	126	181	C	V	M	E	T
26	0.005	189	177	114	183	228	157	307	114	126	191	C	V	M	E	T
27	0.005	189	177	114	183	228	160	307	114	126	187	C	V	M	E	T
28	0.005	189	177	114	185	228	157	307	114	126	181	C	V	M	E	T
*		179	180	?	166	233	167	307	126	147	181	C	V	M	N	K

*Migrant from Central America.

APPENDIX C

DHFR LINKED MICROSATELLITE LOCI AND MUTATION SITES

No.	Freq.	-89	-58	-7.058	-4.973	-4.423	-4.12	-3.771	-1.139	-0.064	0.105	0.45	1.33	5.782	18.209	39.91	50	51	59	108	164
1	0.50323	139	161	258	200	197	175	216	223	96	159	100	203	107	282	185	C	**I**	C	**N**	I
2	0.00645	139	161	258	200	197	175	216	223	96	162	100	203	107	282	185	C	**I**	C	**N**	I
3	0.00645	137	161	258	200	197	175	216	232	96	159	100	203	107	282	185	C	**I**	C	**N**	I
4	0.07742	139	161	258	200	197	175	212	223	96	159	100	203	107	282	188	C	**I**	C	**N**	I
5	0.05161	139	161	258	200	197	175	214	223	96	159	100	203	107	282	185	C	**I**	C	**N**	I
6	0.01936	139	161	258	200	200	175	212	223	96	159	100	203	107	282	188	C	**I**	C	**N**	I
7	0.05161	139	161	258	200	200	175	216	223	96	159	100	203	107	282	185	C	**I**	C	**N**	I
8	0.0129	124	161	258	200	197	175	216	223	96	159	100	203	107	282	185	C	**I**	C	**N**	I
9	0.01936	124	161	258	200	197	175	214	223	96	159	100	203	107	282	185	C	**I**	C	**N**	I
10	0.02581	139	161	258	200	197	175	214	223	96	159	100	203	107	282	188	C	**I**	C	**N**	I
11	0.00645	139	161	258	200	200	175	212	223	94	159	100	203	107	282	188	C	**I**	C	**N**	I
12	0.00645	139	161	258	200	200	175	216	223	96	159	100	203	107	282	188	C	**I**	C	**N**	I
13	0.10968	139	161	258	200	197	175	210	223	96	159	100	203	107	282	185	C	**I**	C	**N**	I
14	0.00645	139	161	258	200	197	175	216	232	96	159	100	203	107	282	185	C	**I**	C	**N**	I
15	0.00645	137	161	258	204	197	175	216	223	96	159	100	203	107	282	185	C	**I**	C	**N**	I
16	0.00645	139	161	258	200	197	175	216	223	94	159	100	203	107	282	188	C	**I**	C	**N**	I
17	0.00645	139	161	258	200	197	175	212	223	94	159	100	203	107	282	185	C	**I**	C	**N**	I
18	0.00645	139	161	258	200	197	175	210	223	94	159	100	203	107	282	185	C	**I**	C	**N**	I
19	0.00645	137	158	258	200	197	175	216	223	96	159	100	203	107	282	185	C	**I**	C	**N**	I
20	0.00645	139	161	258	200	197	175	208	223	96	159	100	203	107	282	183	C	**I**	C	**N**	I
21	0.0129	137	161	258	200	197	175	216	223	96	159	100	203	107	282	185	C	**I**	C	**N**	I
22	0.00645	139	161	258	200	197	175	216	200	96	159	100	203	107	282	185	C	**I**	C	**N**	I
23	0.0129	139	161	258	200	197	175	208	223	96	159	100	203	107	282	185	C	**I**	C	**N**	I
24	0.00645	137	161	258	200	197	175	214	223	96	159	100	203	107	282	185	C	**I**	C	**N**	I
25	0.00645	139	163	258	200	197	175	208	223	96	159	100	203	107	282	185	C	**I**	C	**N**	I
26	0.00645	139	163	258	200	197	179	216	223	96	159	100	203	107	282	195	C	**I**	C	**N**	I
27	0.00645	139	161	258	200	197	175	186	206	96	159	100	203	107	282	185	C	**I**	C	**N**	I
*		?	?	260	222	201	186	212	211	100	177	81	203	111	282	173	C	**N**	C	**S**	I

*Migrant from Central America

APPENDIX D

DHPS LINKED MICROSATELLITE LOCI AND MUTATION SITES

No.	Freq	-72.8	-34.6	-18.8	-11.1	-7.5	-2.8	-1.5	-0.1	0.5	1.4	3.4	6.4	9	49.6	66	436	437	540	581	613
1	0.31	178	206	101	219	169	181	168	132	140	238	108	291	105	300	329	S	**G**	K	A	A
2	0.297	178	206	101	219	169	181	168	132	140	238	108	291	105	296	329	S	**G**	K	A	A
3	0.007	178	206	95	219	173	181	165	132	140	265	108	287	107	283	329	S	**A**	K	A	A
4	0.179	178	206	101	219	169	181	168	132	140	238	108	301	105	300	329	S	**G**	K	A	A
5	0.007	178	206	101	219	169	181	168	132	140	238	108	291	105	293	329	S	**G**	K	A	A
6	0.007	178	206	101	219	**167**	181	168	132	140	238	108	291	105	296	329	S	**G**	K	A	A
7	0.007	178	206	95	219	169	181	165	134	140	265	106	287	107	296	329	S	**A**	K	A	A
8	0.007	178	206	101	219	169	181	168	134	140	238	106	291	105	300	329	S	**G**	K	A	A
9	0.007	178	206	101	219	169	181	168	132	140	238	108	291	105	283	329	S	**G**	K	A	A
10	0.007	178	206	101	219	173	181	168	132	140	238	108	301	105	300	329	S	**G**	K	A	A
11	0.007	178	206	101	219	169	181	168	134	140	238	108	291	105	296	329	S	**G**	K	A	A
12	0.007	178	206	101	228	177	197	165	132	140	269	108	291	133	283	329	S	**A**	K	A	A
13	0.007	178	206	95	219	169	181	165	132	140	265	106	287	107	283	329	S	**A**	K	A	A
14	0.007	178	206	95	219	169	181	165	132	140	265	108	287	107	283	329	S	**A**	K	A	A
15	0.014	178	206	101	219	169	181	168	132	136	238	108	291	105	300	329	S	**G**	K	A	A
16	0.007	178	206	101	219	169	181	168	132	136	238	108	291	105	296	329	S	**G**	K	A	A
17	0.034	178	206	101	219	169	181	168	132	140	238	108	301	105	296	329	S	**G**	K	A	A
18	0.007	178	206	95	219	169	181	165	134	140	265	106	287	107	283	329	S	**A**	K	A	A
19	0.007	180	206	101	228	177	197	168	134	140	269	108	291	133	283	329	S	**A**	K	A	A
20	0.007	178	206	95	219	169	181	165	134	140	265	104	287	107	283	329	S	**A**	K	A	A
21	0.007	178	206	101	219	169	181	168	132	140	238	111	291	105	300	329	S	**G**	K	A	A
22	0.007	180	206	101	228	177	197	168	134	140	275	108	291	105	283	329	S	**A**	K	A	A
23	0.007	178	206	101	219	173	181	168	132	140	238	108	291	105	300	329	S	**G**	K	A	A
24	0.007	178	206	95	219	169	181	168	132	140	238	108	291	105	296	329	S	**G**	K	A	A
25	0.007	180	206	95	219	169	181	168	132	140	238	108	291	105	283	329	S	**A**	K	A	A
26	0.007	180	206	101	228	177	197	165	134	140	269	108	291	133	283	329	S	**A**	K	A	A
27	0.007	180	206	95	219	169	181	168	132	140	238	108	291	105	296	329	S	**A**	K	A	A
28	0.007	180	206	101	228	177	197	165	134	140	262	108	291	133	283	329	S	**A**	K	A	A
*	0.007	204	208	88	219	169	197	165	134	140	248	102	287	101	287	322	S	**A**	K	A	A

*Migrant from Central America