Plasmodium falciparum Genetic Diversity Maintained and Amplified Over 5 Years of a Low Transmission Endemic in the Peruvian Amazon

OraLee H. Branch,*†,[1] Patrick L. Sutton,†,[1] Carmen Barnes,[2] Juan Carlos Castro,[3] Julie Hussin,[4] Philip Awadalla,[4] and Gisely Hijar[2]

[1]Department of Medical Parasitology, New York University

[2]Laboratorio de Biotecnologia y Biologia Molecular, Instituto Nacional de Salud, Lima, Peru

[3]Laboratorio Investigaciones Productos Naturales Antiparasitarios de la Amazonia, Universidad Nacional Amazonia Peruana, Iquitos, Peru

[4]Sainte-Justine Research Centre, University of Montreal, Montréal, Québec, Canada

†These authors contributed equally to this work.

***Corresponding author:** E-mail: oralee.branch@nyumc.org.

Associate editor: Daniel Falush

Research article

Abstract

Plasmodium falciparum entered into the Peruvian Amazon in 1994, sparking an epidemic between 1995 and 1998. Since 2000, there has been sustained low *P. falciparum* transmission. The Malaria Immunology and Genetics in the Amazon project has longitudinally followed members of the community of Zungarococha ($N = 1{,}945$, 4 villages) with active household and health center-based visits each year since 2003. We examined parasite population structure and traced the parasite genetic diversity temporally and spatially. We genotyped infections over 5 years (2003–2007) using 14 microsatellite (MS) markers scattered across ten different chromosomes. Despite low transmission, there was considerable genetic diversity, which we compared with other geographic regions. We detected 182 different haplotypes from 302 parasites in 217 infections. Structure v2.2 identified five clusters (subpopulations) of phylogenetically related clones. To consider genetic diversity on a more detailed level, we defined haplotype families (hapfams) by grouping haplotypes with three or less loci differences. We identified 34 different hapfams identified. The F_{st} statistic and heterozygosity analysis showed the five clusters were maintained in each village throughout this time. A minimum spanning network (MSN), stratified by the year of detection, showed that haplotypes within hapfams had allele differences and haplotypes within a cluster definition were more separated in the later years (2006–2007). We modeled hapfam detection and loss, accounting for sample size and stochastic uctuations in frequencies overtime. Principle component analysis of genetic variation revealed patterns of genetic structure with time rather than village. The population structure, genetic diversity, appearance/disappearance of the different haplotypes from 2003 to 2007 provides a genome-wide "real-time" perspective of *P. falciparum* parasites in a low transmission region.

Key words: malaria, genetic diversity, immunity, low transmission, Peru, microsatellite.

Background

Plasmodium falciparum malaria causes more than 1 million deaths annually in Sub-Saharan Africa alone. Although malaria is being reduced in some historically high-malaria transmission regions, it is spreading into new geographic regions. Malaria epidemics begin when infected mosquitoes enter into a susceptible population of humans. Recurrent transmission of malaria parasites by mosquitoes within a human population may establish a long-term, persistent, endemic transmission cycle. During endemic transmission, parasites that have been successfully transmitted over time will inevitably undergo some genetic changes by random point mutation, replication slippage, or recombination. The redetection or loss of these parasites defines the parasite population structure.

In regions of historically high transmission, the population structure is obscured by frequent overlapping infections. In contrast, the discrete, nonoverlapping infections in low and recent malaria transmission areas promise a less ambiguous characterization of the malaria parasite population structure. For that reason, we began our study in the Peruvian Amazon Jungle. *Plasmodium falciparum* was first detected in perimeter regions of the jungle city, Iquitos, in 1994, with an epidemic occurring between 1995 and 1998. This epidemic was curbed, likely by effective intervention efforts of fumigation and free, highly controlled drug treatment. Since 2000, there has been sustained low *P. falciparum* transmission (Roberts et al. 1997; Aramburu Guarda et al. 1999; Roshanravan et al. 2003; Branch et al. 2005).

© The Author(s) 2010. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.5), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Open Access

The Malaria Immunology and Genetics in the Amazon project, established in 2003, follows parasite genetic diversity and human host immune responses over time. In 2003–2004, combined active and passive case detection described a transmission rate of less than 0.5 *P. falciparum* infections/person/year in Zungarococha, a community of approximately 1,900 individuals located south of Iquitos in one of the highest *P. falciparum* transmission regions (Branch et al. 2005). Our previous work showed that although the frequency of having more than one parasite clone (genotype) per infection was low, there is significant population-level diversity (PLD) in *P. falciparum* parasites circulating in Zungarococha (Branch et al. 2005; Chenet et al. 2008; Sutton et al. 2009).

The current study characterizes *P. falciparum* parasite population structure, temporally and spatially, using 14 microsatellite (MS) markers scattered across ten chromosomes. In general, previous MS studies have demonstrated that the number of alleles and diversity of alleles (heterozygosity) across various MS markers decrease with decreasing *P. falciparum* transmission (Anderson et al. 2000; Leclerc et al. 2002; Hoffmann et al. 2003; Machado et al. 2004; Bogreau et al. 2006; Dalla Martha et al. 2007; Zhong et al. 2007; Bonizzoni et al. 2009). Such decreases in genetic diversity are generally is attributable to both low endemicity and fewer opportunities for sexual recombination between genetically distinct parasites circulating in low transmission, resulting in linkage of markers across chromosomes (Conway et al. 1997; Anderson et al. 2000; Durand et al. 2003). Genetic linkage in low transmission settings can be particularly valuable for tracking parasites in a population temporally and spatially. Only one prior *P. falciparum* MS study considered diversity over time and reported that parasites defined by MS haplotypes appeared/disappeared in a way consistent with random neutral events, that is, immigration of new clones and loss of clones due to genetic drift (Orjuela-Sanchez et al. 2009).

This study had two main objectives: 1) Determine the overall parasite population structure and subgrouping of similar genotypes temporally and spatially and 2) Determine how allele frequencies change with time in this low-endemic region. Each of the alleles had between 3 and 8 polymorphic forms. Parasites were defined by concatenating alleles detected in 14 MS loci into individual haplotypes. Considering infections detected within the four networked villages of Zungarococha over 5 years, we found 182 different haplotypes from 302 parasites in 217 discrete infections. Although this is a hypoendemic region, the maintenance of parasite diversity over time is consistent with a high-effective population size of *P. falciparum* in this low-endemic region. Tracing genetic diversity over time and space might help us to understand how *P. falciparum* has persisted in this human population from 2003 to present.

Materials and Methods

Study Design

This longitudinal study employed both active and passive case detection surveillance methods of 1,945 individuals from Zungarococha, in the San Juan District south of Iquitos, Peru from 2003 to 2007. The San Juan district is an epicenter of *P. falciparum* and *P. vivax* transmission, with infections typically occurring during the rainy season from January to July. The community of Zungarococha is a network of four villages: Zungarococha town (ZG), Puerto Almendra (PA), Ninarumi (NR), and Llanchama (LL). There is a distance of approximately 1 km between ZG, PA, and NR (well within *Anopheles darlingi* immigration range) and approximately 5 km between ZG and LL (although possible, it is not likely to have mosquito migration between ZG and LL). Zungarococha was selected based upon existing reports of locally acquired *P. falciparum* malaria infections, a high *P. falciparum* incidence rate compared with communities in and surrounding Iquitos, the acceptance by the community, and the fact that the community is composed of four villages each serviced by the same health center.

Active case detection (ACD) consisted of one blood sample collected per consenting individual in the beginning and ending of the malaria season (January to July). ACD also included weekly blood sampling during at least 1 of the 7 transmission months. Sample collection in ACD is orchestrated by a physician who obtains fingerprick blood samples or BD Vacutainer (Becton, Dickinson and Co., Franklin Lakes, NJ) blood samples along with a comprehensive epidemiology/demographic and clinical questionnaire.

In addition to ACD, all individuals who presented with a fever or reported fever were tested for malaria by microscopy upon visiting the community health center staffed by our team and the Peruvian Ministry of Health. Details of ACD (asymptomatic and symptomatic) and passive case detection (symptomatic) are described in Branch et al. (2005).

Whether asymptomatic or symptomatic, treatments are given at no cost to the patient. All treatments were given through the MINSA authorities, following the MINSA National Drug Policy Guidelines. *Plasmodium vivax* treatment is chloroquine (10 mg/kg for 3 days) with primaquine (0.5 mg/kg for 7 days). *Plasmodium falciparum* treatment is meoquine (12.5 mg/kg daily for 2 days) with artesunate (4 mg/kg daily for 3 days) in nonpregnant patients older than 1 year of age. These are given as observed drug treatment therapy; all malaria cases are diagnosed and reported so that there is limited access to these drugs outside of the MINSA system.

All blood samples, positive or negative by microscopy, underwent DNA extraction by Qiagen DNeasy Blood and Tissue Kits (Qiagen Inc., Valencia, CA) and were tested for presence of *Plasmodium* species (Rubio et al. 1999) using a seminested multiplex polymerase chain reaction (PCR) method targeting DNA encoding the ssrDNA (Branch et al. 2005; Sutton et al. 2009).

DNA samples were selected by including an average of 28.9% (minimum = 16.6, maximum = 46.0; standard deviation [SD] = 11.1) of the malaria infections detected in each year. Malaria infections were considered on a monthly basis, such that if there were PCR positive samples detected within 1 month of sampling, this was considered one infection (infection-month). Any *P. falciparum* clones

detected within the month were concatenated, possibly resulting in some infections having more than one clone detected throughout the infection (a multiclone, "mixed infection") (Sutton et al. 2009). As shown in Sutton et al. (2009), these mixed infections were likely due to the simultaneous inoculation of more than one clone during a single mosquito biting event, due to the extreme low transmission (<0.5 *P. falciparum* infections/person/year).

Human Subjects Ethical Approval

All protocols were reviewed and approved by the Institutional Review Boards at New York University, the Peruvian Ministry of Health Institutes of National Health, and the University of Alabama at Birmingham (former affiliation). We obtained written informed consent from all participants in this study. In the case of minors less than 7 years old, the parents or guardians gave consent. In the case of minors between 7 and 18, both assent from the minor and consent from the parents or guardians were obtained prior to enrollment.

Microsatellite Marker Selection and Amplification

Fourteen MS markers were chosen from the MS linkage map for amplification. Six of these MS markers (ARA2, PfPK2, POLYA, TA42, TA1, and TA109) were chosen because they were used in previous studies in South America, including Bolivia, Brazil, the Brazilian Amazon, Columbia, and Peru (Anderson et al. 2000; Hoffmann et al. 2003; Machado et al. 2004; Dalla Martha et al. 2007). The same six markers were also used in regions of higher transmission, like the Democratic Republic of Congo, Papua New Guinea, Uganda, Western Kenya, and Zimbabwe; this enabled comparison of genetic diversity between low- and high-transmission regions (Anderson et al. 2000; Leclerc et al. 2002; Bogreau et al. 2006; Zhong et al. 2007). We used eight new MS markers (B5M5, BM17, C1M4, C1M67, C4M69, C9M11, C13M13, and Pf2802) because they are known to be highly polymorphic markers (suggestion courtesy of Dr Xin-zhuan Su). These eight markers were originally tested in the National Institutes of Health (NIH), Department of Infectious Disease laboratory with a test set of 20 DNA samples from our study population. These eight MS markers (B5M5, BM17, C1M4, and C13M13) have been used in recent antimalaria drug studies (Vieira et al. 2004; Liu et al. 2008).

PCR protocols were developed by Dr Su's team at the NIH Infectious Disease laboratory. PCR master mixes were comprised of uorescently labeled primers at 5 μM (Integrated DNA Technologies, Inc, Coralville, IA), 1 mM concentration of deoxynucleoside triphosphate mixture (Invitrogen, Carlsbad, CA), $MgCl_2$ at a concentration of 2 mM, 10 PCR buffer, Platinum Taq (Invitrogen), molecular grade water, and extracted DNA (genomic DNA adjusted to 10–20 ng/μl) or whole-genome amplified DNA. All PCRs were performed in an Eppendorf Mastercycler ep (Westbury, NY). The PCR program was as follows: denature at 94 °C for 2 min, 1 cycle; denature at 94 °C for 20 s, anneal at 52 °C for 10 s, anneal at 47 °C for 10 s, extension time of 30 s at 60 °C, 42 cycles; extension time of 5 min at 60 °C, 1 cycle.

PCR product repeat-length sizes were determined on an ABI 3130xl genetic analyzer (Applied Biosystems, Foster, CA). The use of three different uorescently labeled primers enabled three MS markers to be measured at the same time in a multiplexed assay. Each MS marker allele length was determined by using internal size standards (GeneScan 500 LIZ Size Standard, Applied Biosystems) with the GeneMapper v4.0 software. Only alleles detected with a peak height $\times$200 uorescent units were considered.

During every microsatellite reaction, a Dd2 reference isolate was included as a positive control, *P. vivax* DNA, human DNA, and water were utilized as negative controls. Negative controls were used to ensure against PCR contamination and to rule out spurious bands that may be generated by nonspecific hybridization from human DNA. Moreover, the system was tested for *P. falciparum* specificity by attempting amplification of each primer with various *P. vivax* and *P. falciparum* samples.

Data Analysis

Assembling the Haplotypes

The specificity of having multiple single copy loci (14 MS markers) results in a fine characterization of parasite clone frequencies in the population, defined by a MS marker haplotype. Methods from Anderson et al. (2000) were used to interpret individual clonal haplotypes from infections identified as having more than one clone present (complex). Complex infections from microsatellite haplotype profiles can be identified by detection of two or more alleles at a single locus. However, it is possible that nonspecific marker binding will cause noise in the form of minor peak heights during the capillary electrophoresis reaction. Therefore, the standard method is to exclude minor peaks less than 1/3 the height of the major peaks (alleles) (Anderson et al. 2000). If only one locus had multiple peaks, then separating the clones of complex infections was unambiguous. In this method, single-clone haplotype profiles were assembled (phased). In the few instances where there was more than one locus with multiple alleles (35 samples from 217 infections), major peaks were tentatively paired with major peaks and minor peaks were tentatively paired with minor peaks due to the peak height approximately representing the density of major and minor clones in an infection. Final haplotype profiles determination for the 35 complex infections was made using an iteration analysis that determined agreement with the most parsimonious haplotype profile of each constituent of all complex infections.

Population Genetic Diversity and Structure

Expected heterozygosity (H_e) was used to quantify the amount of genetic diversity considering the haplotypes. Additionally, H_e was calculated for each locus and then compared with the allele count observed at each locus. When H_e is greater than expected based upon the observed number of alleles detected at each locus, it suggests that the population may have recently undergone a bottleneck

(Cornuet and Luikart 1996; Garza and Williamson 2001; Schultz et al. 2009). Using Bottleneck v1.2.02, we tested if this phenomenon was occurring in this Peruvian Amazon study, considering 1,000 simulations for mutations models of infinite alleles (IAM) and stepwise mutation (SMM). Ultimately, we use SMM as it has been identified as a more stringent model when using microsatellite data (Luikart and Cornuet 1998; Iwagami et al. 2009).

For population structure analysis, we used the unsupervised Bayesian clustering algorithm implemented in Structure 2.2 to group individual MS haplotype profiles into genetically related clusters (Pritchard et al. 2000). Each haplotype was given an estimated membership coefficient (Q) that translates to the fraction of relatedness (or ancestry) within/between each cluster (Rosenberg et al. 2005). The number of related clusters (K) assumed in this program must be predicted, we predicted there would be between 1 and 10 clusters within this study population. For each value of K, the clustering algorithm was performed 5 times for 10,000 Monte Carlo Markov Chain iterations, preceded by a burn-in period of 10,000 iterations. Population differentiation of each cluster was confirmed using Arlequin v3.5 (Excoffier and Slatkin 1995; Excoffier and Lischer 2010). Inferred ancestry values from Structure v2.2 were exported into Microsoft Excel and used to create individual bar plots for each sampling epoch in this longitudinal study.

Linkage disequilibrium (LD), a measurement of the nonrandom association of alleles at different loci, was calculated across all 14 loci comprising the MS haplotypes for each infection grouped according to hapfam and also by year of appearance (Arlequin 3.5, Excoffier and Lischer 2010). Fisher's exact tests were used to determine statistical significance of LD and average percentage of linked loci per locus.

Haplotype Family Analysis

Each parasite clone was compared with each of the other clones (pairwise comparison). We grouped infections into haplotypes families based on the number of common loci. A haplotype family (hapfam) was defined by any haplotype profiles that were within three loci deviations of other haplotypes (79.0% concordance). Population differentiation of hapfams was tested using Arlequin v3.5 (Excoffier and Slatkin 1995; Excoffier and Lischer 2010). We tested the average number of pairwise differences between hapfams and within hapfams, with Nei's distance as an additional measurement for genetic distance between populations (Nei and Li 1979; Reynolds et al. 1983; Peterson et al. 1995; Slatkin 1995). These data are reported in supplementary figure 1 (Supplementary Material online).

We developed an estimator, n_s, to compute redetection of hapfams within a fixed time frame (2003–2007). The estimator is based on the idea that a hapfam with more individuals at the start of the sampling period will have a higher probability of redetection at the end of the study period. When a sampled hapfam is not observed within a given year ($n = 0$), the probability that we sample this hapfam the following year is low. In contrast, when five or more individuals are observed in a given year ($n \times 5$), the probability is high. Let $P(R)$ be the probability of redetection from year $y1$ to year $y2$ given that the redetection depends on the number of individual haplotypes observed in $y1$ ($ny1$). $P(R)$ is defined as:

$$P(R) = P(R|n_{y1} = 0)P(n = 0) + P(R|0<n_{y1}<5)P(0<n_{y1}<5) + P(R|n_{y1} \times 5)P(n>5). \quad (1)$$

P ($n = 0$) is the proportion of families that were not observed in a given year. Similarly, P ($0 < n_{y1} < 5$) and P ($n_{y1} \times 5$) are the proportion of families that were observed with $0 < n < 5$ and $n > 5$, respectively.

The conditional probabilities, $P(R|n_{y2}=0)$, $P(R|0<n_{y2}<5)$, and $P(R|n_{y1} \times 5)$, are the probability of redetecting a family at $y2$ when $n_{y1} = 0$, when $0 < n_{y1} < 5$, and when $n_{y1} \times 5$, respectively. Let yf be any year from $y2$ to 2007 and let E be one of the following events: $n_{y1} = 0$, $0 < n_{y1} < 5$, or $n_{y1} \times 5$. Then:

$$P(R|E) = P(n_{yf}>0|E) = \frac{P(E|n_{yf}>0) \quad P(n_{yf}>0)}{P(E)}. \quad (2)$$

The quantities $P(E|n_{yf}>0)$, $P(n_{yf}>0)$, and $P(E)$ can be estimated from the data. If N different families are present in the sample, the estimate of the number redetected between 2003 and 2007 is $n_s = P(R) \quad N$.

We estimated the expected number of redetected families n_s overall (for the entire sample), and we used a chi-square test to compare it with the observed estimate of the number of redetected families computed separately for each category (monomorphic and polymorphic). The null hypothesis is that the number of redetected families in a category is the same as the expected number given all the data.

We assessed whether there is significant variation in haplotype diversity among communities and years by analysis of variance (ANOVA). Patterns of genetic diversity were inferred by principal component analysis (PCA). The number of repeats is standardized to have a mean of zero and variance of 1, and the PCA is conducted on the matrix of standardized number of repeats using the svdPca method from the R package "pcaMethods." We tested for significant differentiation of the regional (ZG, PA, NR, LL) and temporal (2003–2005, 2006–2007) samples on the first three PCs through ANOVA using the R statistical package.

Data and Results

Genetic Diversity Characterized by Haplotype Profiles

In total, this study included 217 different *P. falciparum* infections occurring between 2003 and 2007 in the four villages comprising the community of Zungarococha. The 217 infections were randomly selected to obtain a sampling density of approximately 25% of all infections detected

Table 1. All Infections Detected in Study by Year, by Village, and Number of *Plasmodium falciparum* Clones Identified.

Village	Infection Breakdown	2003, (N = 34)	2004, (N = 63)	2005, (N = 114)	2006, (N = 70)	2007, (N = 21)
PA (N = 89)	Number of infections identified	25	56	70	18	17
	Number of infections tested	10	9	25	12	5
	Number of clones	18	10	39	15	7
NR (N = 99)	Number of infections identified	28	94	118	38	8
	Number of infections tested	6	13	26	19	5
	Number of clones	12	22	34	25	6
LL (N = 67)	Number of infections identified	2	31	102	50	4
	Number of infections tested	0	13	18	20	1
	Number of clones	0	15	26	22	4
ZG (N = 47)	Number of infections identified	11	66	68	18	15
	Number of infections tested	2	11	13	6	3
	Number of clones	4	16	15	8	4

in each year and village (table 1). Using our 14 MS marker-defined haplotypes to define clones, there were 144 single-clone infections and 73 complex infections detected. Of the 73 complex infections detected, 38 were easily translated into single-clone infections due to having polymorphisms in only one locus, whereas 23 had polymorphisms at two different loci and 12 had polymorphisms at more than two loci. Therefore, potential haplotype mischaracterization (incorrect phasing, "misalignment," of the polymorphisms defining each clone) would be limited to 35 infections. Performing our analyses (presented below) with and without these 35 complex infections showed that inferring the phasing of these haplotypes did not alter the general results.

There were 182 unique MS haplotypes detected in the 302 parasite infections (within 217 individuals).

Global-Level Genetic Diversity

PLD, alleles detected, allele counts, and expected heterozygosity (H_e) were calculated (table 2). Previous studies reported basic measurements of PLD using eight of the same markers (C4M69, PolyA, PF2802, TA42, TA1, TA109, ARA2, and PFPK2) in different geographic regions (compared in table 3 and fig. 1). Most of the allele size ranges at each MS marker observed in this study were within the range of those previously reported. The only exception was the range of allele sizes detected at the C4M69 locus, which were approximately 50–65 bp longer than the allele sizes reported in Bogreau et al. (2006).

We noted two distinctions comparing our Peru data with the prior studies. First, the Peruvian Amazon population had an elevated mean number of alleles per locus (by approximately a factor of 2) when compared with regions of similar hypoendemicity in Central and Eastern Brazil, Bolivia, and Columbia (Anderson et al. 2000; Hoffmann et al. 2003) (see table 3 and fig. 1). The mean number of alleles detected in Peru was more comparable with reports from the Brazilian Amazon (Machado et al. 2004; Dalla Martha et al. 2007) and South east Asia which included sites in Vietnam, Thailand, and Papua New Guinea (PNG) (Anderson et al. 2000; Hoffmann et al. 2003) (fig. 1). Secondly, the difference between the mean H_e and allele count was lower than the difference observed in the Brazil, Bolivia, and Vietnam data (Anderson et al. 2000). Generally, a high difference between H_e and allele count suggests population

Table 2. Microsatellite Marker Loci and Basic Genetic Diversity Analyses.

		Alleles, Named "a"–"h" in Order of Frequency									
Locus Name	Chromosome Number	a	b	c	d	e	f	g	h	H_e	Stepwise Mutation Model[a]
C1M4	1	206	192	220	210					0.567	Excess
B5M5	3	194	176	171	209					0.229	Deficient
C4M69[b]	4	413	425	429	420					0.388	Deficient
POLYA[b]	4	209	206	230	212					0.352	Deficient
PF2802[b]	5	140	167	156	146	126	164	177	183	0.585	Deficient
TA42[b]	5	185	190	199	176					0.205	Deficient
TA1[b]	6	171	177	155	183					0.598	Excess
TA109[b]	6	160	163	153						0.338	Deficient
BM17	8	162	174	152	164					0.606	Excess
C9M11	9	130	140	122	158	185	132			0.183	Deficient
ARA2[b]	11	60	72	75	54	63				0.651	Excess
PFPK2[b]	12	178	175	166	180	170	160			0.617	Deficient
C1M67	13	232	253	190	211	220	200	263		0.255	Deficient
C13M13	13	144	148	132	136	174	153	162		0.631	Deficient
P value testing if there is an excess H_e[a]											P = 0.9

NOTE.—[a]$H_e$was not in excess, indicating there was not a recent population constriction (bottleneck) leading to loss of alleles. The IAM and SMM was tested, the SMM computation more suitable for MS markers (Bottleneck v1.2.02) were tested by a Wilcoxon signed-rank test.

[b] Microsatellite loci used in previous studies. Table 3 further considers these 8 markers used in previous studies.

Table 3. Number of Alleles Detected Globally.

Study Location	Endemicity	Number of Alleles Per Locus								Mean	References
		C4M69	POLYA	Pf2802	TA42	TA1	TA109	ARA2	PfPK2		
Africa											
Senegal	Hyper	—	—	—	5	—	8	—	—	6.5	Leclerc et al. (2002)
Djibouti, Senegal	Hyper	7	—	8	5	—	—	—	—	6.7	Bogreau et al. (2006)
Dakar, Senegal	Hyper	7	—	7	5	—	—	—	—	6.3	
Niamey, Senegal	Hyper	10	—	9	—	—	—	—	—	9.5	
Zouan Hounien, Senegal	Hyper	10	—	17	—	—	—	—	—	13.5	
Kombewa, W. Kenya	Hyper	—	14	—	5	10	10	8	9	9.3	Zhong et al. (2007)
Kakamega, W. Kenya	Hyper	—	9	—	4	9	8	7	8	7.5	
Kisii, W. Kenya	Hyper	—	6	—	2	4	6	7	7	5.3	
Congo	Hyper	—	16	—	10	11	10	10	10	11.2	Anderson et al. (2000)
Uganda	Hyper	—	18	—	6	11	11	9	11	11.0	
Zimbabwe	Hyper	—	18	—	10	12	13	10	11	12.3	
South America											
Peruvian Amazon	Hypo	4	4	8	4	4	3	5	5	4.6	Current study
Brazilian Amazon	Hypo	—	12	—	4	8	7	8	—	7.8	Machado et al. (2004)
Brazilian Amazon	Hypo	—	10	—	6	5	4	2	5	5.3	Dalla Martha et al. (2007)
Brazil	Hypo	—	—	—	—	—	—	—	—	—	Vieira et al. (2004)
Brazil	Hypo	—	—	—	3	5	3	2	7	4.0	Hoffmann et al. (2003)
Brazil	Hypo	—	5	—	2	3	3	2	2	2.8	Anderson et al. (2000)
Bolivia	Hypo	—	2	—	2	3	3	2	1	2.2	
Columbia	Hypo	—	4	—	1	1	1	2	3	2.0	
South East Asia and Oceania											
Vietnam	Meso	—	—	—	—	4	—	7	5	5.3	Hoffmann et al. (2003)
Kalinga, Philippines	Hypo	—	4	—	3	2	1	—	4	2.8	Iwagami et al. (2009)
Palawan, Philippines	Meso	—	7	—	5	8	1	—	3	4.8	
Davao del Norte, Philippines	Hypo—Meso	—	8	——	2	10	2	—	3	5.0	
Thailand	Hypo	—	6	—	2	5	2	5	6	4.3	Anderson et al. (2000)
Buksak, PNG	Hyper	—	10	—	5	8	2	5	6	6.0	
Mebat, PNG	Hyper	—	10	—	5	8	4	5	7	6.5	

bottlenecks (constrictions in alleles in prior years). The *He* not being in excess in Peru did not support a bottleneck or decrease in effective population size.

To test if this Peruvian Amazon population had recently undergone bottleneck events, we calculated H_e for each locus and then compared with the allele count observed at each locus using the program Bottleneck v1.2.02 (Cornuet and Luikart 1996; Garza and Williamson 2001; Schultz et al. 2009). H_e was in significant excess in only four of the 14 different loci (ARA2, BM17, C1M4, and TA1) (table 2). Using a Wilcoxon signed-rank test and mode-shift analysis, we found that this does not support a population that has recently undergone a bottleneck ($P = 0.9$, with normal L-shaped distribution).

Characterizing the Population Structure

To determine the most accurate number of clusters of related clones (K) circulating in this population, a range between 1 and 10 clusters was tested (see Materials and Methods). A $K = 5$ consistently had the highest clusteredness value over five simulations for each prespecified K value, with $K = 0.81$. A clear increase in clusteredness values approaching $K = 5$ in both directions, maximizing at $K = 5$, indicated a population defined by five different clusters. We tested for population differentiation between clusters using the fixation index, F_{st} (Excoffier and Slatkin 1995; Excoffier and Lischer 2010). We found that each of the five clusters represented independent subpopulations of a larger circulating parasite population ($P < 0.01$, Pairwise F_{st}).

For a more exact definition of clones, we defined haplotype families (hapfams) using a haplotype-to-haplotype pairwise comparison method. The 182 different haplotypes (found in the 217 infections) could be grouped into 34 hapfams. Many of the hapfams (14 of 34) included between 3 and 15 of the 302 clones identified in this study. Three of these 34 hapfams described the majority of infections: hapfam #14 at 27.2% ($n = 82$), #8 at 20.5% ($n = 62$), and #11 at 10.6% ($n = 32$). Sixteen of the 34 hapfams were detected only one or two times.

We correlated the results of both population structuring methods (clusters of related clones and hapfams of related clones). Of 302 parasite clones, 286 clones (94.7%) were concordant between the hapfam and cluster groupings. In 26 of the 34 hapfams defined ($n = 231$ parasite clones of 302, 76.5% of total), all members (100%) were identified within one specific cluster (fig. 2). In six of the eight instances where all members of a hapfam

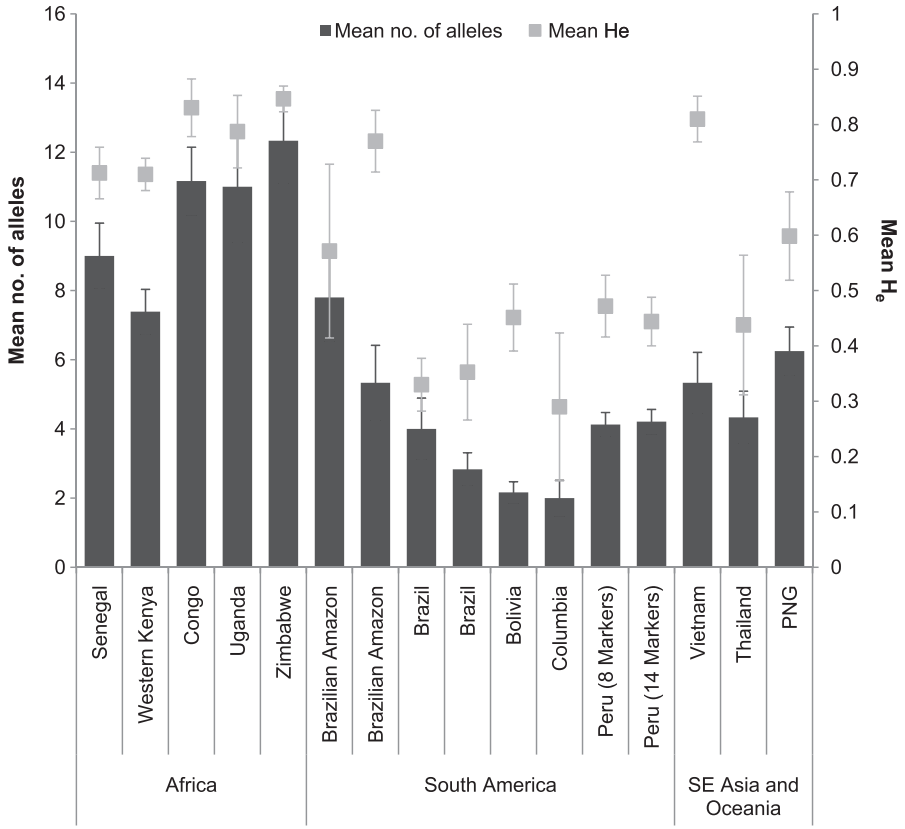

Fig. 1. Global perspective of the mean number of alleles per locus versus the mean expected heterozygosity. The mean number of alleles per locus is represented by the dark gray bars measured on the primary *y* axis, and the mean expected heterozygosity (H_e) is represented by the light gray squares measured on the secondary *y* axis (error bars indicated standard error of the mean). On the *x* axis are locations of studies that have also examined PLD by utilizing microsatellite markers. Some studies are missing microsatellite markers that were reported in other studies (including ours), but this is simply an adaptation from what has been previously reported on a global perspective. As well, some studies report observed heterozygosity (H_o) rather than H_e but are just described synonymously. See table 3 for references.

did not identify to one specific cluster, more than 85% of the members did identify with a specific cluster. Only 16 (5.3%) clones were in hapfams that grouped to more than one cluster.

Each haplotype with its corresponding hapfam characterization and cluster characterization is shown in figure 2 using a minimum spanning network (MSN) algorithm. The MSN is divided by early years of the study (2003–2005, fig. 2*A*) versus the later years of the study (2006–2007, fig. 2*B*). In 2003–2005, there were approximately 2/3 more infections detected in this population than in 2006–2007 (table 1). Despite the lower number of infections (and haplotypes) detected in the later years, the overall area of the MSN spanning area was similar. The haplotypes within a given hapfam generally had more allele differences (shown by the length of the line in the MSN) in the later years. Additionally, it is clear that rather than having high-frequency clusters, as indicated by the circle size, the clusters were more expansive and defined a more diverse set of haplotypes (and hapfams) in the later years. Overall, the agreement between cluster and hapfam was more apparent in the early years of the study (2003–2005) versus the later years of the study (2006–2007), a likely indication of temporal diversification (fig. 2*A* and *B*).

Temporal and Spatial Variation of Clusters

At a level of year of infection and village of detection, we considered the cluster characterization to determine the overall maintenance of genetic diversity. Despite the incidence of infections varying between years, with 2005 having nearly doubled the number of infections as years 2004 and 2006, and five times the number of infections as in 2007, most clusters were detected in each year within each village of this study (fig. 3). There were only two exceptions: cluster 5 was not detected in 2003 and cluster 4 was not detected in one village. It is possible that the number of infections tested per village/year could explain these few absences. Overall, we observed that the clusters were maintained in this population both temporally and spatially.

Expected heterozygosity (H_e) values within each village (H_e range = 0.40–0.48) and over time (H_e range = 0.33–0.45, excluding year 2003) remained relatively constant, despite some uctuation in the proportion of each cluster temporally and spatially. F_{st} was used to determine if there was divergence between these parasite subpopulations. Significant divergence was detected between each village ($P < 0.01$, pairwise F_{st}) except PA and NR and also between each year ($P < 0.01$, pairwise F_{st}) of this study (fig. 3). We tested if

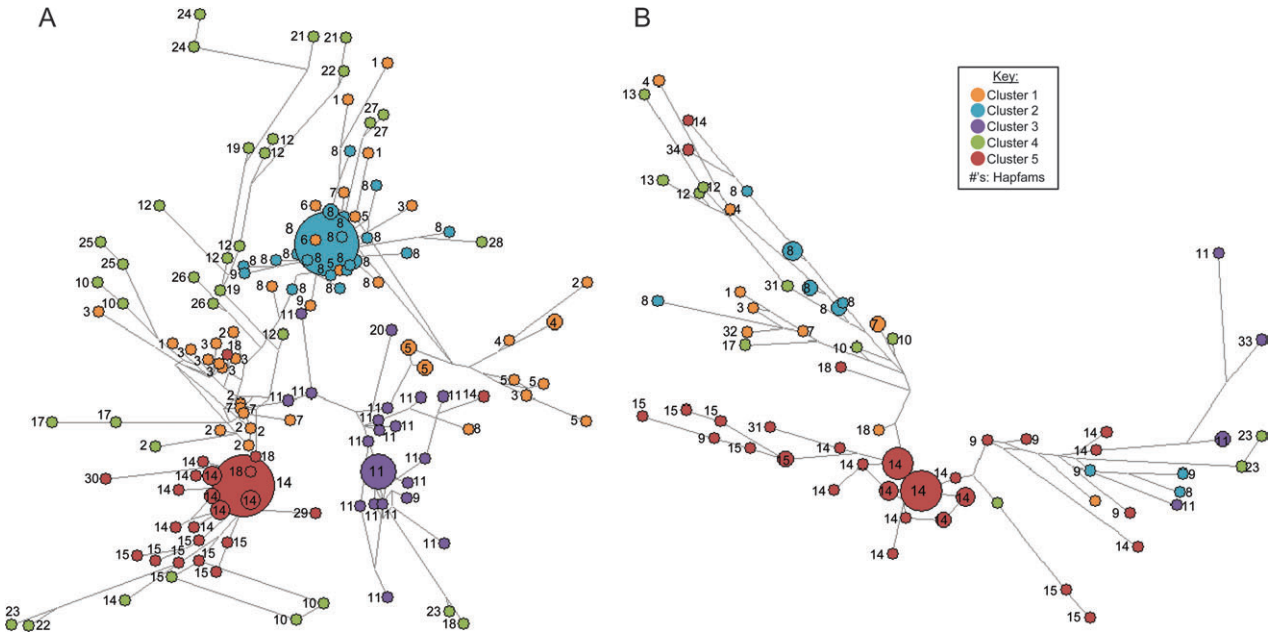

Fig. 2. Minimum spanning network of parasites collected in 2003–2005 (*A*) and 2006–2007 (*B*). The MS haplotype is shown with its cluster and hapfam characterization. The size of the circle re ects the frequency of the given haplotype in the MSN (*A* and *B*, separately). Haplotypes with different in less than 3 MS markers were defined as the same hapfam. There were 34 hapfams, and each haplotype is indicated by its hapfam (numbered 1–34). Colored circles represent clusters (1–5) as classified by Structure v2.2: orange = cluster 1, light blue = cluster 2, purple = cluster 3, green = cluster 4, and magenta = cluster 5. The clusters and hapfams generally agreed, particularly in 2003–2005. In 2006–2007, despite the lower number of infections, the genetic differentiation between haplotypes tended to increase (overall area similar in *A* and *B*; length of lines longer in *B*).

the number of hapfams detected was a function of the size and frequency of the hapfams parasites (haplotypes) over time. First, we established whether the hapfams were monomorphic or polymorphic based on the original criteria used to define hapfams, where members of a hapfam could have variation in three or fewer different loci. Hapfams were considered monomorphic if there was a single haplotype that represented the entire family. A hapfam was considered polymorphic if the haplotype sequences within the hapfam did have variation.

The number of hapfams detected for the first time in the study (new), detected from one year to the next (redetected), and failing to be detected again (lost) were determined while considering whether the hapfams were monomorphic or polymorphic (fig. 4). Totaling the instances of hapfams redetection, the monomorphic hapfams were redetected in 15 of 33 (45.5%) possible transitions from one year to another, while the polymorphic hapfams were redetected 32 of 36 (88.9%) possible transitions from one year to another. Moreover, the redetection of polymorphic hapfams relative to monomorphic hapfams was significantly greater ($P < 0.0001$, χ^2), with the odds of polymorphic hapfam redetection being 8.47-fold higher than monomorphic hapfams (odds ratio = 8.47). To control for size in families, we used an estimator that takes into account the fact that a hapfam with more individuals has a higher probability of redetection (eqs. 1 and 2). This estimator is used to compute the redetection of families for the entire data set and in each category separately (table 4). The expectation of survival for the monomorphic hapfams is not significantly different from that computed for all hapfams ($P = 0.53$, χ^2). However, the expected number of redetecting polymorphic hapfams is significantly greater than expected for all families ($P = 0.0229$, χ^2). This result suggests that families that are redetected over the years have 1–3 differences.

Maintenance of Subpopulations and Diversification

We examined variation in haplotype diversity among years and villages and found the presence of significant variation in villages ($P = 0.017$) and time ($P < 0.0001$). This ANOVA suggests that more variation is explained by the temporal component. To confirm this result, we performed a PCA of genetic variation that revealed patterns of genetic structure with time (fig. 5*A*) rather than village (fig. 5*B*). Assessing the significance of PCs using Tracy–Widom distribution is not applicable here because of the small number of markers. However, the scores for the first (26.47% of extracted variance), second (16.04%), and third (13.31%) principal components are significantly different between infections detected in years 2003–2005 and in years 2006–2007 ($P = 5.9 \quad 10^{-10}$ for PC1, $P = 7.4 \quad 10^{\ 8}$ for PC2, and $P = 0.031$ for PC3). In contrast, among the different villages, the scores are significantly different but only for the first principal component ($P = 0.032$).

To investigate the possible emergence of new hapfams by immigration, we divided infections by village and then by year of detection: 1) those detected in years 2003–2005 only, 2) those detected in 2003–2005 and also 2006–2007, and 3) those detected only in 2006–2007 (table 5). An

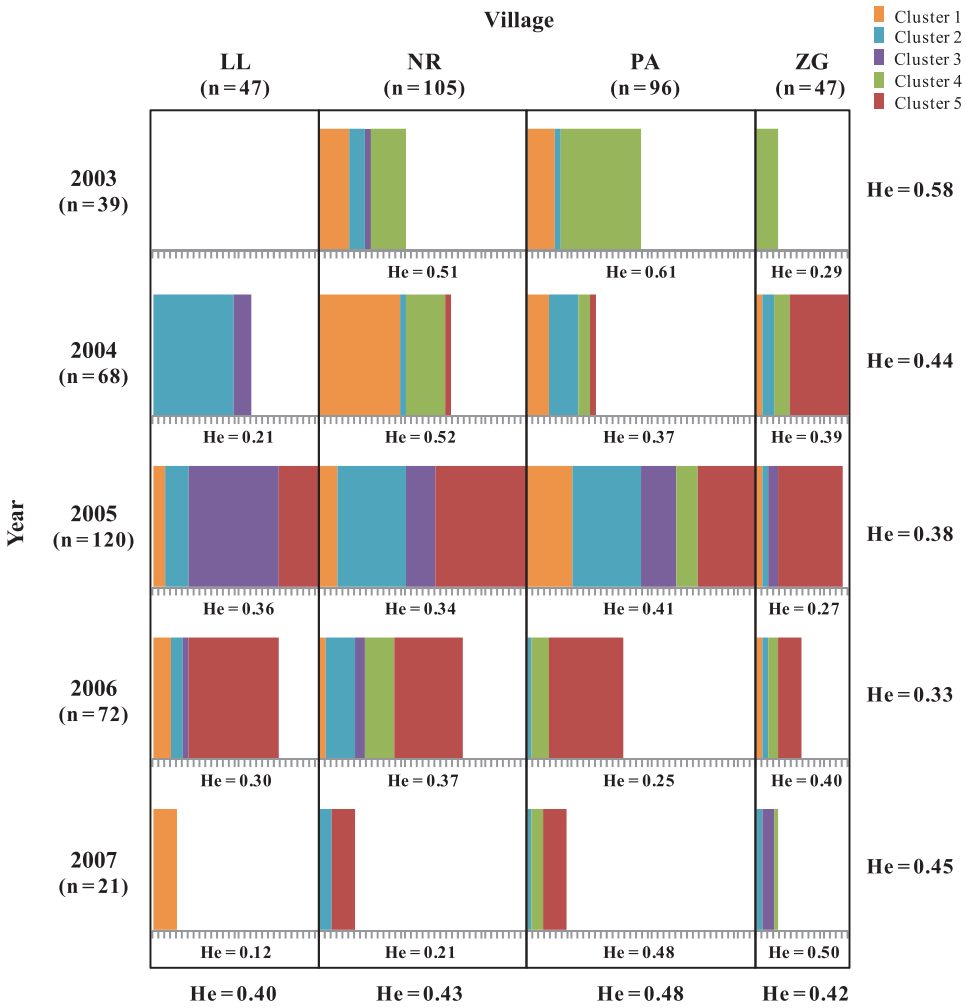

Fig. 3. (*a*, *b*) Bar plot of clustered infections examined by village–year. Each of the five horizontal *x* axes represent the year that infections were detected, whereas the location of the infection by village, in which the infection was detected, intersects in four locations on the *y* axes. Where each village intersects with a corresponding year is referred to as a village–year, indicative of the infections detected in a particular village during a particular year. Individual infections are identified by gray tick marks along the corresponding year *x* axis. Colored bars represent clusters (1–5) as classified by Structure v2.2: orange = cluster 1, light blue = cluster 2, purple = cluster 3, green = cluster 4, and magenta = cluster 5. The overall H_e values are shown to the right for each study–year plotted and individual villages along the bottom most *x* axis. Individual village–year H_e values are shown beneath the infections reported during each respective village–year. Incomplete gaps within village–years indicate a lack of infections detected during that village–year.

immigration event from a village outside the community of Zungarococha might be observed if specific hapfams were detected de novo in villages throughout the 5 years of this study. Of the 19 hapfams that were detected in 2006–2007, only five hapfams (8 infections) were not previously detected between 2003 and 2005. Therefore, we can only suggest that 8 of the 264 infections detected in the later years were possibly attributable to immigration from outside of the parasites circulating in this population.

To consider if hapfams detected in the later years evolved from parasites detected in the earlier years, LD was calculated for each of the three temporal groups above. The average percentage of linked loci per locus for those hapfams detected only in years 2003–2005, for those detected in 2003–2005, and also 2006–2007, and for those only detected in 2006–2007 was 41.8%, 72.7%, and 3.0%, respectively. There was significant decay in LD observed in years 2006–2007 versus 2003–2005 ($P < 0.0001$, Fishers Exact). The 5 hapfams only detected in 2006–2007 were comprised of only eight infections. These 5 hapfams could have arisen in the villages due to recombination between existing haplotypes or due to immigration. The percent of linked loci in 2003–2005, when most infections were detected, suggests that the majority of diversity is not introduced via immigration of different clones from outside these villages.

Discussion

Longitudinal cohort studies in recent and low-malaria transmission regions enable us to study the characteristics associated with the maintenance of endemics. In this study, we investigated the population structure dynamics of *P. falciparum* in the community of Zungarococha near

A

Number of infections (n) monomorphic families detected each year										
Hapfam ID	2003		2004		2005		2006		2007
6	3	x	0		0		0		0
7	2	>	1	>	1	>	1	>	1
16	2	x	0		0		0		0
17	2	>	0	>	0	>	0	>	1
19	2	x	0		0		0		0
20	1	x	0		0		0		0
21	1	x	0		0		0		0
22	1	x	0		0		0		0
23	1	>	1	>	0	>	2	x	0
24	2	x	0		0		0		0
25	2	x	0		0		0		0
4			3	>	0	>	2	x	0
5			6	>	4	x	0		0
26			2	x	0		0		0
27					2	x	0		0
28					1	x	0		0
29					1	x	0		0
30					1	x	0		0
13							1	>	2
31							2		0
34							1		0
32									1
33									1
Total n redetected		3		5		4		3	
Total n Lost		8		1		5		4	

B

(n) polymorphic families detected each year									
Hapfam ID	2003		2004		2005		2006		2007
2	5	>	2	x	0		0		0
8	3	>	20	>	28	>	7	>	4
10	2	>	2	>	0	>	3	x	0
11	1	>	3	>	24	>	3	>	1
12	4	>	2	>	0	>	1	x	0
1			1	>	3	>	0	>	1
3			4	>	5	>	0	>	1
9			2	>	1	>	4	>	3
14			6	>	34	>	34	>	3
15			6	>	1	>	6	>	2
18			2	>	3	>	3	x	0
Total n Redetected		5		10		10		7	
Total n Lost		0		1		0		3	

> Re-detected: n>0 in the following year
> Lost: n>0 in any following year
x Lost: n=0 for all following years

Fig. 4. (*A*, *B*) Detection, redetection, and loss of monomorphic versus polymorphic hapfams over time. Hapfams are represented within green boxes at time of entry. The monomorphic hapfams are shown in the top panel (*A*) and the polymorphic below (*B*) with the name of the hapfam (numbered 1–34). In both monomorphic and polymorphic hapfams, there can be different haplotypes in each hapfam. Hapfams are called the same if $\leq$3 MS loci makers are different. Therefore, even in monomorphic hapfams, mutations can occur. We report the number of individuals in hapfams (*n*) for years 2003–2006 was used to compute parameters to estimate the redetection probabilities (see table 4).

Iquitos, Peru over a 5-year period (2003–2007) of sustained low transmission (<0.5 infections/person/year) that occurred 5 years postepidemic (1994–1998). Using 14 MS loci scattered throughout the genome of *P. falciparum*, we found markedly high population-level genetic diversity when one considers the relatively low transmission history of these villages. Among 217 different infections, we found 182 different MS haplotypes. Using an unsupervised Bayesian clustering program, called Structure v2.2, we found five clusters of related parasites (population substructure). The haplotypes could be subgrouped into 34 hapfams, which were a fine-level description of the subpopulations but were consistent with the population defined by Structure. Having 182 haplotypes that could be grouped into 34 hapfams was surprising given the low and recent transmission (Mackinnon and Marsh 2010).

Our first objective was to characterize this genetic diversity and population structure, in comparison with other geographic regions. Prior MS studies of *P. falciparum* genetic diversity in other regions of low transmission report a low genetic diversity and evidence for parasite populations having undergone genetic bottlenecks (fig. 1). If

Table 4. Hapfams Parameters and Estimate of Expected Number of Redetecting Families Over Time.

	Monomorphic	Polymorphic	Total
Families[a]	21	11	32
Redetected families	15	32	47
Observations[a]	84	44	160
Observations $0 < n < 5$[a]	27	21	47
Observations $n \geq 5$[a]	1	11	12
Observations $n = 0$[a]	56	12	101
Redetected from $0 < n < 5$	9	17	26
Redetected from $n \geq 5$	1	11	12
Redetected from $n = 0$	5	4	9
Probability of redetection	0.18	0.73	0.29
Expected No. Redetected	3.75	8	9.4
Expected No. Lost	17.25	3	22.6

NOTE.—[a]Computed without taking observations for year 2007 into account because the redetection rates from 2007 to 2008 are unknown. These observations were only considered to count the number of families redetected from 2006 to 2007. Therefore, the two families that entered in 2007 were not considered in the analysis.

a parasite population was primarily shaped by an epidemic or if there were bottlenecks and/or genetic drift, we would expect to observe few parasite haplotypes in the population (a clonal or low diversity population). For example, Urdaneta et al. (2001) observed little diversity in Venezuela, consistent with this type of epidemic or population demography. The allele count and H_e in our Peruvian cohort parasite population gives no evidence of such a bottleneck. We found the allele count and H_e similar to endemic regions of the Brazilian Amazon, Southeast Asia, and Oceania (fig. 1).

In our Peru cohort, the failure to detect a population bottleneck and the accumulation of genetic diversity over time would lead us to predict a large effective population size. The MSN graphics show this diversity. Approximately, 2/3 of all the infections detected occurred in years 2003–2005 (table 1 and fig. 2*A* and *B*) and even with only 1/3 of infections observed, the MSN shows much of this genetic diversity in years 2006–2007. Our second objective was to consider the population over time, redetection of alleles over time and space. The MSN graphic (fig. 2*A* and *B*) showed that in the later years of the study (fig. 2*B*), haplotypes were more diverged with more allelic changes, haplotypes, within hapfams. The polymorphic hapfams in the population were more likely to be redetected over successive years relative to the monomorphic families (fig. 4). We consider that monomorphic families were lost more readily due to genetic drift. On the other hand, redetecting the polymorphic families with possible 1–3 allele changes, is consistent with large effective population size. In agreement, the PCA and the ANOVA found significant variation of genetic diversity by year and village, with more variation being explainable by year (fig. 5*a* and *b*). Although immigration events were possible, they appear unlikely to explain all the instances of detecting parasites different by 1–3 alleles over the years of this study (table 5). Agreements of haplotype clustering in early versus later years of the study (fig. 2*A* and *B*) also provide evidence that diversity is higher than expected given the transmission rate and that this diversity occurs within the re-detected hapfams.

The maintenance of diversity over time suggests we have a large population size. The effective population size (N_e) can be estimated by a formula with H_e of the allele frequencies and mutation rates in MS markers (see Iwagami et al. 2009) or by the conventional methods using nucleotide polymorphisms in noncoding regions (Mu et al. 2002). However, the N_e calculation assumes that the mutation rates and transmission rates are constant over time and homogenous over space. In this study, we found that the N_e is most likely impacted by the maintenance and amplification of genetic diversity over time.

There has been one longitudinal MS study published, which considered 44 infections that occurred between

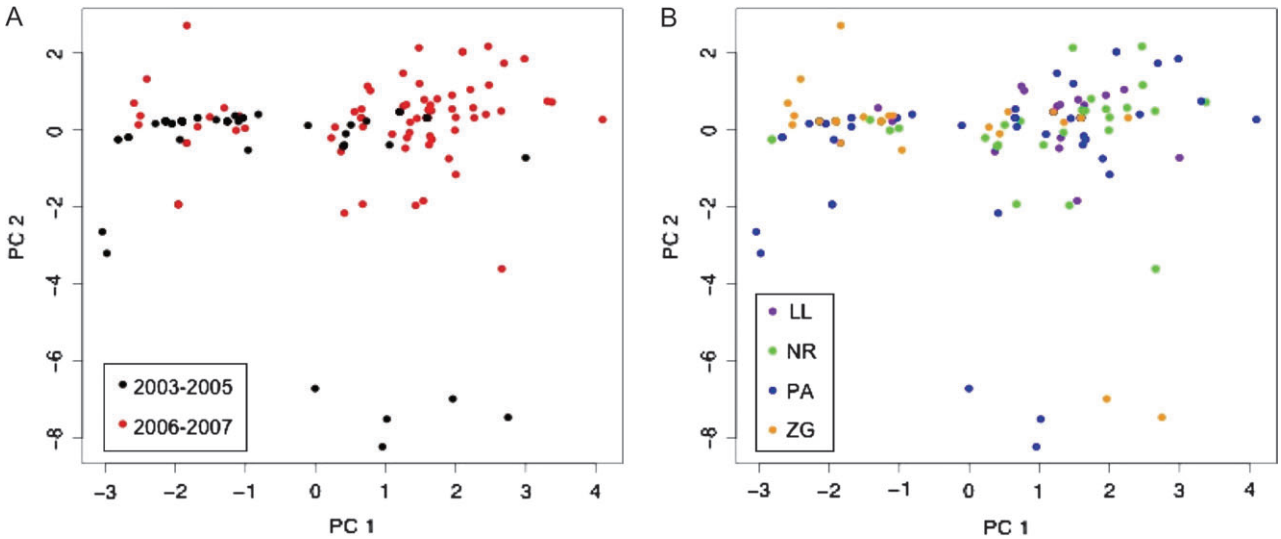

Fig. 5. (*A*, *B*) Graphic representation of the first two principle components for 206 individual infections genotyped with 14 microsatellites markers. Color code shows subgroups of infections partitioned (*A*) by year of sampling and (*B*) by villages. The level of genetic variation detected is more in uenced by the year of collection, than the village of collection.

Table 5. Hapfam Persistence in Villages.

	Detected in 2006–07 and Detected in 2003–05																		
	Same Village										Not Same Village					Not Detected Before			
Village																			
LL	3		8			11		14	15	18		1	4			31	32		
NR		7	8	9	10	11		14			23					31			34
PA		7	8	9	10			14	15	18					13				
ZG			8			11	12	14	15	18				17				33	
Total haplotypes (*n*)	10	6	62	10	7	32	7	82	15	8	4	5	5	3	3	2	1	1	1

NOTE.—The 19 hapfams (named by number and shown by numbers in columns in order below) that were detected in 2006–2007 were divided into two groups: those that were detected earlier (2003–2005), and those that were not detected earlier in Zungarococha.

2004 and 2006 in Grenada, Brazil and 11 infections that occurred in 2008 in Acre, Brazil (Orjuela-Sanchez et al. 2009). Using Structure, they found three different populations of parasites that appeared, disappeared, and/or reappeared in the population in a manner consistent with genetic drift. In contrast, in our longitudinal study, we redetect each of the five main clusters in each of the 5 years. The smaller sample size and the distance separating Grenada and Acre (60 km) may have contributed their detecting bottlenecks, genetic drift with only certain quite different alleles being detected over the successive years.

Our finding maintenance of hapfams with 1–3 allele differences over time might be re ective of differences in the immunoepidemiology within these regions in Brazil versus Zungarococha, Peru. We define immunoepidemology as the interaction of parasite infection and host susceptibility to infection. In Peru, we find asymptomatic infections, polymorphisms in antigen encoding genes, and strong human host antibody responses (Branch et al. 2005; Torres et al. 2008; Sutton et al. 2010). It is plausible that this may contribute to the maintenance and/or amplification of genetic diversity by impacting which parasites infect or are transmitted from host to host. We propose that sequencing antigen encoding genes near the MS markers is needed to understand the immunoepidemology on genetic diversity in this population.

Considering the presumed neutral MS markers, there is a direct relationship between endemicity, recombination rates, and population size (Mackinnon and Marsh 2010). However, if there was a nonhomogeneously distributed infection frequency perhaps due to endemicity and/or changing mixed-clone infection prevalence, we could imagine detecting a large effective population size even in an overall average low-transmission region. Recombination is obviously limited by the opportunity for outcrossing between different clones (Anderson et al. 2000; Durand et al. 2003). Here, we report that only 33.6% (73/217) of infections were mixed-clone infections (table 1). A 33.6% mixed infection rate would predict more than 80% linkage of genes distanced by more than 1/3 of the length of most *P. falciparum* chromosomes (Conway et al. 1999; Sakihama et al. 1999). However, the frequency or types of mixed-clone infections could uctuate or change over time. For example, the low LD in the last year suggests recombination in the last years, which could result in a larger effective population size in the last years despite the overall decreasing epidemiologic transmission in the last years.

Conclusions

We find that *P. falciparum* emergence and then establishment of low continued transmission in Iquitos, Peru, is characterized by a considerably high genetic diversity and maintenance of this diversity temporally and spatially. Although malaria researchers have often proposed that there is maintenance of genetic diversity in endemic regions where parasites reinfect the human hosts, this study provides the first genome-wide real-time insight into such a population structure that might impact our global efforts of malaria elimination.

Supplementary Material

Supplementary figure 1 is available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This study was supported by R01 grant AI064831 from the NIH/National Institute of Allergy and Infectious Disease (year 2005 to present). C.B. was supported by a fellowship award from the Howard Hughes Medical Institute: this fellowship propelled the research and the technology transfer in Peru. J.H. and Dr P.A. contributed the probability analysis and PCA while they received support from an NSERC student fellowship and Canadian Institute of Health Research (grant number 200183). Ms Hijar was supported by the Peruvian INS and conducted experiments, organized data, mentored our Howard Hughes fellow, and was lead in the INS laboratory in Peru. We would like to thank the Zungarococha community members and authorities for their ongoing commitment to the Malaria Immunology and Genetics in the Amazon (MIGIA) Study. We thank Dr Jean Hernandez for overall coordination and Dr Crystyan Siles for patient care. The field, laboratory and data researchers, and staff were all essential for this investigation and continue to make the MIGIA study possible. We thank Freddy Alava for enrollment follow-up, Ever Alvarez, and Anibal Sanchez for microscopy, Jey Montenegro and Aldo Montenegro, Zoila Reategui, and Elva Sanchez for sample collection during patient visits. We thank Lindsay Prado, Claudia Silva, Sory Vasquez, and Nolberto Tangoa for laboratory sample

processing, and Odilo Alava, Dania Vela, and Noelia for data entry.

References

Anderson TJ, Haubold B, Williams JT, et al. (16 co-authors). 2000. Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. *Mol Biol Evol*. 17:1467–1482.

Aramburu Guarda J, Ramal Asayag C, Witzig R. 1999. Malaria reemergence in the Peruvian Amazon region. *Emerg Infect Dis*. 5:209–215.

Bogreau H, Renaud F, Bouchiba H, et al. (15 co-authors). 2006. Genetic diversity and structure of African *Plasmodium falciparum* populations in urban and rural areas. *Am J Trop Med Hyg*. 74:953–959.

Bonizzoni M, Afrane Y, Baliraine FN, Amenya DA, Githeko AK, Yan G. 2009. Genetic structure of *Plasmodium falciparum* populations between lowland and highland sites and antimalarial drug resistance in Western Kenya. *Infect Genet Evol*. 9:806–812.

Branch O, Casapia WM, Gamboa DV, Hernandez JN, Alava FF, Roncal N, Alvarez E, Perez EJ, Gotuzzo E. 2005. Clustered local transmission and asymptomatic *Plasmodium falciparum* and *Plasmodium vivax* malaria infections in a recently emerged, hypoendemic Peruvian Amazon community. *Malar J*. 4:27.

Chenet SM, Branch OH, Escalante AA, Lucas CM, Bacon DJ. 2008. Genetic diversity of vaccine candidate antigens in *Plasmodium falciparum* isolates from the Amazon basin of Peru. *Malar J*. 7:93.

Conway DJ. 1997. Natural selection on polymorphic malaria antigens and the search for a vaccine. *Parasitol Today*. 13:26–29.

Conway DJ, Roper C, Oduola AM, Arnot DE, Kremsner PG, Grobusch MP, Curtis CF, Greenwood BM. 1999. High recombination rate in natural populations of *Plasmodium falciparum*. *Proc Natl Acad Sci U S A*. 96:4506–4511.

Cornuet JM, Luikart G. 1996. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* 144:2001–2014.

Dalla Martha RC, Tada MS, Ferreira RG, da Silva LH, Wunderlich G. 2007. Microsatellite characterization of *Plasmodium falciparum* from symptomatic and non-symptomatic infections from the Western Amazon reveals the existence of non-symptomatic infection-associated genotypes. *Mem Inst Oswaldo Cruz*. 102: 293–298.

Durand P, Michalakis Y, Cestier S, Oury B, Leclerc MC, Tibayrenc M, Renaud F. 2003. Significant linkage disequilibrium and high genetic diversity in a population of *Plasmodium falciparum* from an area (Republic of the Congo) highly endemic for malaria. *Am J Trop Med Hyg*. 68:345–349.

Excoffier L, Lischer HE. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour*. 10:564–567.

Excoffier L, Slatkin M. 1995. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol Biol Evol*. 12:921–927.

Garza JC, Williamson EG. 2001. Detection of reduction in population size using data from microsatellite loci. *Mol Ecol*. 10:305–318.

Hoffmann EH, Ribolla PE, Ferreira MU. 2003. Genetic relatedness of *Plasmodium falciparum* isolates and the origin of allelic diversity at the merozoite surface protein-1 (MSP-1) locus in Brazil and Vietnam. *Malar J*. 2:24.

Iwagami M, Rivera PT, Villacorte EA, Escueta AD, Hatabu T, Kawazu S, Hayakawa T, Tanabe K, Kano S. 2009. Genetic diversity and population structure of *Plasmodium falciparum* in the Philippines. *Malar J*. 8:96.

Leclerc MC, Durand P, de Meeus T, Robert V, Renaud F. 2002. Genetic diversity and population structure of *Plasmodium falciparum* isolates from Dakar, Senegal, investigated from microsatellite and antigen determinant loci. *Microbes Infect*. 4:685–692.

Liu S, Mu J, Jiang H, Su XZ. 2008. Effects of *Plasmodium falciparum* mixed infections on in vitro antimalarial drug tests and genotyping. *Am J Trop Med Hyg*. 79:178–184.

Luikart G, Cornuet JM. 1998. Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. *Conserv Biol*. 12:228–237.

Machado RL, Povoa MM, Calvosa VS, Ferreira MU, Rossit AR, dos Santos EJ, Conway DJ. 2004. Genetic structure of *Plasmodium falciparum* populations in the Brazilian Amazon region. *J Infect Dis*. 190:1547–1555.

Mackinnon MJ, Marsh K. 2010. The selection landscape of malaria parasites. *Science* 328(5980):866–871.

Mu J, Duan J, Makova KD, Joy DA, Huynh CQ, Branch OH, Li W-H, Su XZ. 2002. Chromosome-wide SNPs reveal an ancient origin for *Plasmodium falciparum*. *Nature* 418:323–324.

Nei M, Li WH. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci U S A*. 76:5269–5273.

Orjuela-Sanchez P, Da Silva-Nunes M, Da Silva NS, Scopel KK, Goncalves RM, Malafronte RS, Ferreira MU. 2009. Population dynamics of genetically diverse *Plasmodium falciparum* lineages: community-based prospective study in rural Amazonia. *Parasitology* 136:1097–1105.

Peterson AC, Di Rienzo A, Lehesjoki AE, de la Chapelle A, Slatkin M, Freimer NB. 1995. The distribution of linkage disequilibrium over anonymous genome regions. *Hum Mol Genet*. 4:887–894.

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

Reynolds J, Weir BS, Cockerham CC. 1983. Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics* 105:767–779.

Roberts DR, Laughlin LL, Hsheih P, Legters LJ. 1997. DDT, global strategies, and a malaria control crisis in South America. *Emerg Infect Dis*. 3:295–302.

Rosenberg NA, Mahajan S, Ramachandran S, Zhao C, Pritchard JK, Feldman MW. 2005. Clines, clusters, and the effect of study design on the inference of human population structure. *PLoS Genet*. 1:e70.

Roshanravan B, Kari E, Gilman RH, et al. (11 co-authors). 2003. Endemic malaria in the Peruvian Amazon region of Iquitos. *Am J Trop Med Hyg*. 69:45–52.

Rubio JM, Benito A, Roche J, Berzosa PJ, Garcia ML, Mico M, Edu M, Alvar J. 1999. Semi-nested, multiplex polymerase chain reaction for detection of human malaria parasites and evidence of *Plasmodium vivax* infection in Equatorial Guinea. *Am J Trop Med Hyg*. 60:183–187.

Sakihama N, Kimura M, Hirayama K, Kanda T, Na-Bangchang K, Jongwutiwes S, Conway D, Tanabe K. 1999. Allelic recombination and linkage disequilibrium within MSP-1 of *Plasmodium falciparum*, the malignant human malaria parasite. *Gene* 230:47–54.

Schultz JK, Baker JD, Toonen RJ, Bowen BW. 2009. Extremely low genetic diversity in the endangered Hawaiian Monk Seal (*Monachus schauinslandi*). *J Hered*. 100:25–33.

Slatkin M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139:457–462.

Sutton PL, Neyra V, Hernandez JN, Branch OH. 2009. *Plasmodium falciparum* and *Plasmodium vivax* infections in the Peruvian Amazon: propagation of complex, multiple allele-type infections without super-infection. *Am J Trop Med Hyg*. 81:950–960.

Sutton PL, Clark EH, Silva C, Branch OH. 2010. The *Plasmodium falciparum* merozoite surface protein-1 19 KD antibody

response in the Peruvian Amazon predominantly targets the non-allele specific, shared sites of this antigen. *Malar J*. 9:3.

Torres KJ, Clark EH, Hernandez JN, Soto-Cornejo KE, Gamboa D, Branch OH. 2008. Antibody response dynamics to the *Plasmodium falciparum* conserved vaccine candidate antigen, merozoite surface protein-1 C-terminal 19kD (MSP1-19kD), in Peruvians exposed to hypoendemic malaria transmission. *Malar J*. 7:173.

Urdaneta L, Lal A, Barnabé C, Oury B, Goldman I, Ayala FJ, Tibayrenc M. 2001. Evidence for clonal propagation in natural isolates of *Plasmodium falciparum* from Venezuela. *Proc Natl Acad Sci U S A*. 98:6725–6729.

Vieira PP, Ferreira MU, Alecrim MG, Alecrim WD, da Silva LH, Sihuincha MM, Joy DA, Mu J, Su XZ, Zalis MG. 2004. pfcrt polymorphism and the spread of chloroquine resistance in *Plasmodium falciparum* populations across the Amazon Basin. *J Infect Dis*. 190:417–424.

Zhong D, Afrane Y, Githeko A, Yang Z, Cui L, Menge DM, Temu EA, Yan G. 2007. *Plasmodium falciparum* genetic diversity in western Kenya highlands. *Am J Trop Med Hyg*. 77:1043–1050.