Using Mitochondrial Genome Sequences to Track the Origin of Imported *Plasmodium vivax* Infections Diagnosed in the United States

Priscila T. Rodrigues, João Marcelo P. Alves, Ana Maria Santamaria, José E. Calzada, Manijet Xayavong, Monica Parise, Alexandre J. da Silva, and Marcelo U. Ferreira

Department of Parasitology, Institute of Biomedical Sciences, University of São Paulo, São Paulo, Brazil; Department of Parasitology, Gorgas Memorial Institute of Health, Panama City, Panama; Center for Global Health, Division of Parasitic Diseases and Malaria, Centers for Disease Control and Prevention, Atlanta, Georgia

**Abstract.** Although the geographic origin of malaria cases imported into the United States can often be inferred from travel histories, these histories may be lacking or incomplete. We hypothesized that mitochondrial haplotypes could provide region-specific molecular barcodes for tracing the origin of imported *Plasmodium vivax* infections. An analysis of 348 mitochondrial genomes from worldwide parasites and new sequences from 69 imported malaria cases diagnosed across the United States allowed for a geographic assignment of most infections originating from the Americas, southeast Asia, east Asia, and Melanesia. However, mitochondrial lineages from Africa, south Asia, central Asia, and the Middle East, which altogether contribute the vast majority of imported malaria cases in the United States, were closely related to each other and could not be reliably assigned to their geographic origins. More mitochondrial genomes are required to characterize molecular barcodes of *P. vivax* from these regions.

**INTRODUCTION**

Although malaria transmission has been interrupted in most of Europe and North America, new infections continue to be imported into developed countries, mainly by migrants and travelers coming from areas with ongoing malaria transmission. Accordingly, approximately 6,000 and 1,700 imported malaria cases are reported each year in France and the United Kingdom, respectively. The most recent malaria surveillance data from the United States is from 2010, when 1,688 imported cases (the largest number since 1980) were reported.

The geographic origin of most imported infections can be inferred from travel histories, which were available for 88% of the malaria cases (all parasite species considered) reported in US residents in 2010. Of the cases where a travel history was reported, 64% of cases were acquired in Africa, 16% of cases were acquired in the Americas, and 15% of cases were acquired in Asia. Nevertheless, Asia contributed most of the 325 laboratory-confirmed imported *Plasmodium vivax* infections reported in 2010 (222 cases or 68%) followed by Africa (18%) and the Americas (11%). Most *P. vivax* infections imported from Africa were acquired in countries that are known for being endemic for this species (such as Ethiopia, with 31 cases). However, some originated from countries such as Uganda (five cases), Ghana (two cases), and Kenya (two cases), where transmission of *P. vivax* was thought to be exceedingly rare until recently.

Determining the geographic origin of an imported infection poses a challenge when travel history is not recorded, such as in 44% and 12% of the United States imported cases diagnosed in 2009 and 2010, respectively. Such gaps might have a long-term negative impact on the malaria surveillance system. In addition, travel histories may bias the investigation of *P. vivax* and *P. ovale* infections in travelers, because primary blood-stage parasitemia may have been suppressed by chemoprophylaxis and relapses may occur several months after exposure. As a result, any vivax malaria episode diagnosed in a recently returning traveler may, in fact, be a relapse of an infection acquired during previous travel, further complicating the attribution of infections to particular endemic sites. This information is important from the public health perspective, because local mosquito-borne transmission must be distinguished from real imported infections in malaria-receptive areas, such as southern Europe, and endemic countries with malaria elimination programs (e.g., Zanzibar, Saudi Arabia, and Sri Lanka).

Limited sequence analyses of antigen-coding nuclear loci have revealed differences between local and imported parasites as well as between parasite strains imported from different endemic regions. However, the 6-kb mitochondrial genome of malaria parasites may have some advantages over currently available nuclear markers as a source of information on *P. vivax* polymorphisms for epidemiological studies for a few reasons. (1) Because they are uniparentally inherited, mitochondrial genome lineages do not recombine with each other and can be resolved by phylogenetic analysis. (2) Mitochondrial haplotypes, at least for *P. falciparum*, cluster according to the geographic origin of the samples. (3) Most genetic polymorphisms in the mitochondrial genome are evolutionarily neutral and may reflect the population history of lineages more precisely than those genetic polymorphisms in antigen-coding genes. (4) Several hundreds of complete mitochondrial genome sequences from worldwide-collected *P. vivax* are currently available in public databases. In this study, we sought to determine whether mitochondrial genomes provide region-specific molecular barcodes for tracing the geographic origin of imported *P. vivax* infections diagnosed in the United States.

**PATIENTS, MATERIALS, AND METHODS**

**Study population.** We analyzed 69 cases of imported *P. vivax* infection originally diagnosed between 2004 and 2008 by state and local health departments across the United States. Paired (ethylenedinitrilo)tetraacetic acid (EDTA) blood samples and Giemsa-stained slides had been submitted to the Division of Parasitic Diseases and Malaria of the Centers for Disease Control and Prevention, Atlanta, Georgia.
Disease Control and Prevention (CDC) for routine laboratory confirmation of diagnosis by conventional light microscopy and species-specific nested polymerase chain reaction (PCR) targeting the 18S rRNA gene of human malaria parasites. DNA was extracted from 200 µL EDTA blood with the QIAamp blood kit (Qiagen, Chatsworth, CA) according to the manufacturer’s instructions and stored at 4°C until use for PCR identification. Informative travel histories were available for 52 (75.4%) cases, indicating that 23 infections had been acquired in the Americas (3 cases from Mexico, 17 cases from Central America, and 3 cases from South America), 15 infections had been acquired in India, 7 infections had been acquired in Africa, 3 infections had been acquired in Melanesia (2 cases from Papua New Guinea and 1 case from Vanuatu), 2 infections had been acquired in Afghanistan, 1 infection had been acquired in southeast Asia, and 1 infection had been acquired in Korea.

**PCR amplification and sequencing of mitochondrial genomes.** We used two pairs of previously described oligonucleotide primers (PvMit1F/PvMit1R and PvMit2F/PvMit2R) (Supplemental Table 1) to amplify two overlapping fragments (2,911- and 3,497-bp long, respectively) of the mitochondrial genome of *P. vivax*. Long-range, high-fidelity PCR amplification was performed using PrimeSTAR HS DNA polymerase (Takara, Otsu, Shiga, Japan), which has efficient 3’ → 5’ exonuclease proofreading activity. PCR reactions were performed in 50 µL total reaction volume containing 6 µL DNA template, 0.3 µM each oligonucleotide primer, 10× PrimeSTAR HS DNA Buffer, 2.5 mM each deoxynucleoside (dNTP), and 5 units PrimeSTAR HS DNA polymerase. PCR was performed on a GeneAmp PCR 9700 thermal cycler (Applied Biosystems, Foster City, CA) at 94°C for 1 minute followed by 30 cycles at 98°C for 10 seconds, 55°C for 5 seconds, and 72°C for 1 minute. A final extension was done at 72°C for 10 minutes. PCR products were purified with the Illustra GFX PCR and Gel Band Purification Kit (GE Healthcare Biosciences, Pittsburgh, PA) and sequenced using the BigDye kit (version 3.1, Applied Biosystems, Foster City, CA) on an ABI 3100 DNA sequencer (Applied Biosystems, Foster City, CA). We used 13 oligonucleotide primer pairs (Supplemental Table 2) to fully sequence both strands of each amplicon, with reads ranging between 303 and 566 bp. To remove nucleotide ambiguities, some sequences were further confirmed on two independent PCR reactions from the same DNA template. The sequenced fragments were assembled into complete mitochondrial genome sequences using the DNASTAR software (version 8.1.13, Madison, WI) and deposited in the GenBank database (accession numbers KF668430–KF668442).

**Sources of additional mitochondrial genome sequences.** We have analyzed complete mitochondrial genome sequences from 417 *P. vivax* isolates collected worldwide. This dataset comprises mitochondrial genomes from 69 imported infections sequenced as part of this study and 335 complete mitochondrial genome sequences that were publicly available as of December of 2012 (GenBank accession numbers AY791517–AY791692, AY598035–AY598140, AB550270–AB550280, JN788737–JN788776, and DQ96547–DQ96548). Except for three sequences deposited by Cox-Singh and Lau in 2008, all sequence data had been previously published. To augment the number of samples from the Americas in the database, we used the methods described above to amplify and sequence the complete mitochondrial genome of seven additional *P. vivax* isolates collected between 2010 and 2012 in Remansinho, a rural community in northwestern Brazil, and six isolates collected between 2007 and 2008 as part of routine surveillance activities across three provinces (Eastern Panama, Darien, and Bocas del Toro) of Panama, Central America (GenBank accession numbers KF668430–KF668442).

**Data analysis.** Mitochondrial genome sequences were aligned using ClustalX (version 2.0.12) and manually edited using MEGA software (version 4.0.2). For geographic analysis, sequences with known origin were assigned to the following regions: east Asia (Korea and China), southeast Asia (India, Pakistan, Sri Lanka, and Bangladesh), southeast Asia (Myanmar, Thailand, Cambodia, Vietnam, Philippines, Indonesia, Borneo, and East Timor), Melanesia (Papua New Guinea, Solomon Islands, and Vanuatu), Africa (Madagascar, Mozambique, Mauritania, Nigeria, São Tomé, Ethiopia, Tanzania, Rwanda, Angola, Uganda, Sudan, and Namibia), the Middle East and central Asia (Turkey, Iran, and Afghanistan), and the Americas (Belize, Costa Rica, El Salvador, Guatemala, Honduras, Nicaragua, Panama, Dominican Republic, Mexico, Colombia, Guyana, Ecuador, Venezuela, Brazil, and Peru). Bayesian phylogenetic analysis was carried out using MrBayes (version 3.2.1) and performing four runs of two chains each (one heated and one cold) for 5 million generations. The phylogenetic tree was drawn using Dendroscope, and minor manual edits were done for aesthetic purposes. Only unique haplotypes were included in the Bayesian phylogenetic analysis of GenBank-derived sequences; when identical haplotypes came from different regions, the geographic origin of one of them was randomly assigned to the haplotype. However, all imported malaria samples sequenced in this study are represented in the tree shown in Figure 1, even when two or more of them had identical haplotypes. Additionally, median-joining phylogenies were generated using Network software (version 4.6; www.fluxus-engineering.com) with default parameters and transversions weighted two times as much as transitions. This analysis was carried out using all 417 complete mitochondrial genome sequences.

**Ethical considerations.** The study protocol was approved by the Ethical Committee for Research with Human Subjects of the Institute of Biomedical Sciences, University of São Paulo (960/CEP) on September 27, 2010. Specimens were used in accordance with a CDC human subject-approved protocol; no additional specimens were obtained for the molecular analysis.

**RESULTS**

**Bayesian phylogeny.** Figure 1 shows the consensus tree generated by the Bayesian phylogenetic analysis. Most clades had low (< 50) estimated posterior probabilities, and many of them included samples from different geographic regions. However, despite the overall lack of strong genetic and geographic structure in the phylogeny, we found 10 well-supported clades with four or more sequences each and estimated posterior probabilities > 50. We examined whether the mitochondrial lineages were segregated according to their known or presumed geographic origin. Clades A, B, D, E, H, and J were particularly informative to trace the origin of imported malaria infections, because they comprise mitochondrial lineages predominantly from...
a single geographic location. For example, clade A contained 54 lineages largely restricted to Melanesia (except for two samples from southeast Asia and one sample each from Africa and the Middle East) and was split into several well-supported subclades. Samples 3CDC and 24CDC (from Papua New Guinea) and 64CDC (unknown origin) clustered together, and 50CDC (from Vanuatu) was placed in another subclade of clade A. This clustering pattern was consistent with the presumed origin of samples 3CDC, 24CDC, and 50CDC (from Vanuatu) and suggested that sample 64CDC also originated from Melanesia. Overall, these results indicate that Melanesian isolates have had enough evolutionary time to diverge from most other parasite populations. Clade B, with 52 lineages, included 22 samples from imported malaria cases; 18 of them were putatively from the Americas and 4 were of undetermined origin. They clustered together with mitochondrial lineages predominantly from the Americas, except for two samples each from southeast Asia and Melanesia and one sample from south Asia, suggesting that samples 40CDC, 32CDC, 59CDC, and 63CDC were most likely imported from the Americas. Clades D (10 lineages) and H (15 lineages), both separated into several subclades, comprised only haplotypes from China and Korea, including one United States-diagnosed imported infection (43CDC) with history of recent travel to South Korea. Clade E contained only four haplotypes, all from south Asia, including one United States-diagnosed infection putatively imported from India (47CDC). Finally, clade J (35 lineages), also

Figure 1. Bayesian phylogenetic analysis of a worldwide collection of *P. vivax* mitochondrial genome sequences, including 69 samples from imported malaria cases diagnosed in the United States between 2004 and 2008 (labeled with numbers followed by CDC). The posterior probabilities of selected clades (those clades with values > 50) are listed next to the corresponding branches. We labeled 10 of 10 well-supported clades with four or more sequences each with letters from A to J. Terminal branches were colored to indicate the origin of samples; the Middle East group also comprised imported malaria samples (labeled as 4CDC and 48CDC; putatively from Afghanistan). Yellow indicates samples from imported malaria cases of unknown origin. Only unique haplotypes from worldwide collected samples were included in the phylogenetic tree; when identical haplotypes came from different regions, the geographic origin of one of them was randomly assigned to the haplotype. However, all 69 imported malaria samples sequenced in this study were represented in the tree, even when two or more of them had identical haplotypes.
separated into several subclades, contained two imported malaria haplotypes. One sample (49CDC) was presumably imported from southeast Asia (Myanmar or Thailand), and the other (4726CDC) was from India; both clustered together with lineages known or presumed to have originated mostly from southeast Asia (N = 25) or Melanesia (N = 7). However, this clade also contained three samples from south Asia (including 4726CDC). Altogether, clades A, B, D, E, H, and J include 25 lineages from United States-diagnosed imported infections with a putative place of acquisition inferred from travel histories. Only one of them (4726CDC in clade J) did not cluster according to its presumed geographic origin. Five imported infections of unknown origin, included in these clades, could be reliably traced back to a geographic population.

Nevertheless, the small clades C, F, G, and I (25 samples altogether) were much less informative, because they combined samples with distinct geographic origins. In clades C and F (six lineages each), imported malaria samples of unknown origin (58CDC, 65CDC, and 67CDC) clustered together with mitochondrial lineages from south Asia and Africa (including the imported cases 44CDC, 4891CDC, and CDC4879 [all of presumed Indian origin]); as a result, their precise geographic origin could not be confirmed. Cluster G contained only imported malaria samples: four samples from India, three samples from Africa, one sample from Afghanistan, and one sample from Mexico. Finally, clade I comprised three sequences from southeast Asia and one sequence from south Asia. As a result, we were unable to confirm the presumed origin of 12 imported infections and assign origins to 3 imported infections included in clades C, F, G, or I.

Finally, 24 United States-diagnosed imported malaria samples were placed in the star-like component of the tree instead of well-supported clades. These samples include six infections putatively acquired in India (6CDC, 52CDC, 5075CDC, 4797CDC, 27CDC, and 9CDC), four infections putatively acquired in the Americas (23CDC, 41CDC, 11CDC, and 14CDC), four infections putatively acquired in Africa (28CDC, 10CDC, 18CDC, and 38CDC), one infection putatively acquired in Afghanistan (4CDC), and nine infections of undetermined origin (2CDC, 53CDC, 54CDC, 56CDC, 55CDC, 57CDC, 60CDC, 62CDC, and 66CDC). No inference could be made about the origin of these samples. In conclusion, only a minority of mitochondrial lineages from United States-diagnosed imported malaria cases clustered in the phylogenetic tree according to their geographic origin inferred

**Figure 2.** Median-joining network of a worldwide collection of *P. vivax* mitochondrial genome haplotypes, including 69 samples from imported malaria cases diagnosed in the United States between 2004 and 2008 (labeled with numbers followed by CDC). Circles represent haplotypes, and their sizes are proportional to haplotype frequencies. Colors indicate the region of origin of the samples like in Figure 1. Blue = the Americas; brown = the Middle East (including two imported malaria cases putatively from Afghanistan); green = southeast Asia; orange = South Asia; pink = Melanesia; purple = China and Korea; red = Africa; yellow = imported malaria cases of unknown origin. Each line connecting the circles represents a mutational step, and grey circles represent hypothetical missing intermediates.
by the travel histories, with inconclusive or inconsistent results for 28 (53.8%) samples. Only 5 of 17 (29.4%) imported infections lacking informative travel histories could be unambiguously assigned to geographic populations.

**Median-joining phylogeny.** The topology of the haplotype network generated by the median-joining methodology (Figure 2) mirrored the one of the tree obtained by Bayesian phylogenetic analysis (Figure 1). There were several apparent haplotype clusters, many of them comprising lineages from predominantly a single geographical location. Nearly all Melanesian lineages (including samples 3CDC, 24 CDC, and 50CDC of the putative Melanesian origin and 64CDC of undetermined origin) are included in the three clusters in the upper right of Figure 2, which roughly corresponds to clade A of the phylogenetic tree (Figure 1). Proceeding in a clockwise direction, the next cluster roughly corresponds to clade B of the phylogenetic tree and contains most American lineages, including 17 imported malaria samples of putative American origin. The four imported malaria samples of unknown origin found in this cluster could be assigned to the American population. However, the imported malaria sample 8CDC (putatively from South America) was placed ambiguously midway between the American cluster and one of the Melanesian clusters. The next cluster in the network (still in a clockwise direction) contains lineages from diverse origins: mostly from Africa and south Asia but also from the Middle East, central Asia ( Afghanistan), and the Americas. All African samples (including those samples from imported malaria infections with presumed African origin) are placed in this cluster. A large number of samples from United States-diagnosed imported malaria cases of unknown origin are included in this cluster and could not be unambiguously assigned to any particular geographic region. The small cluster in the lower right corner of Figure 2 is comprised predominantly of samples from east Asia (China and Korea), including the imported malaria sample 43CDC, which is of putative Korean origin. A small secondary cluster (only five haplotypes), which is related to the African/south Asian cluster and one of the southeast Asian groups, also contained predominantly east Asian samples. The three remaining clusters included mostly southeast Asian samples intermingled with some Melanesian, south Asian, and east Asian lineages. One of these clusters contains the United States-diagnosed malaria sample 49CDC, which was putatively imported from either Thailand or Myanmar.

In conclusion, the network analysis was unable to reveal additional geographic structure in this worldwide sample of mitochondrial genome lineages that had been missed by the Bayesian phylogenetic analysis. Most haplotypes from Africa, south Asia, and the Middle East were connected to each other or shared among populations, creating a major cluster in the network that does not correspond to a single geographic population. Ongoing parasite population admixture, caused by human migrations across the Indian Ocean, is a likely explanation for these findings.

**DISCUSSION**

In this study, we showed many mitochondrial genome sequences of *P. vivax* segregating according to the geographic origin of the isolates, but the resolution obtained was not sufficient to assign most individual samples of undetermined origin to specific geographic populations. Previous microsatellite analyses of worldwide *P. vivax* populations revealed a similar picture; although most samples clustered according to their continental origin, they could not be readily assigned to a single geographic population. Accordingly, samples from Ethiopia were clearly segregated from samples from Myanmar and Sri Lanka, but the two Asian populations could not be differentiated from each other based on allelic frequencies of microsatellite loci, despite the vast geographic distance between the collection sites. In addition, analysis of microsatellites has been unable to clearly differentiate southeast Asian strains from Melanesian strains of *P. vivax*.

The most challenging samples for geographic tracking are those samples from south Asia, central Asia, Africa, and the Middle East, which are comprised of mitochondrial lineages that are closely related to each other. For these locations, mitochondrial genomes clearly fail to provide population-specific molecular barcodes useful for epidemiological studies. One can speculate that a large proportion of *P. vivax* strains currently circulating in east Africa has been imported by migrants from south Asia, allowing for an extensive genetic admixture between parasites from these locations. The number of currently available mitochondrial genome sequences from the most challenging populations remains limited. Africa and south Asia (India, Pakistan, Bangladesh, and Sri Lanka) together contributed 74% of the imported *P. vivax* malaria cases diagnosed in the United States in 2010; however, of 417 mitochondrial sequences analyzed here, only 39 sequences came from India, and 60 sequences came from Africa (including 15 and 7 United States-diagnosed imported malaria infections with Indian and African origin that were inferred by travel histories, respectively). *P. vivax* populations from Africa and India display extensive polymorphism and genetic substructure, which does not necessarily correspond to sampling locations. A very dense sampling of these populations will be required for a more accurate geographic tracking of imported infections from Africa and India.

Nevertheless, most of the imported malaria samples from the Americas, southeast Asia, east Asia, and Melanesia could be traced to their origins by the analysis of their mitochondrial genomes. These regions are the presumed origin for 15% of the imported *P. vivax* infections diagnosed in the United States in 2010. Informative polymorphisms were scattered across the 6-kb sequence, implying that nearly full-length mitochondrial genome sequencing is required to determine the continental origin of these isolates. Despite the relatively low levels of local malaria transmission, populations of *P. vivax* are extremely diverse genetically and geographically structured in the Amazon Basin (which accounts for 90% of the malaria recorded in the Americas and the Caribbean). The parasites from these regions show a significant genetic divergence from other populations separated by relatively short distances. A very recent analysis of mitochondrial lineages of *P. vivax* sampled across the Americas confirmed the high level of genetic diversity and substructuring of local parasite populations, with at least four well-defined haplotype clusters. These data are consistent with the introduction of *P. vivax* in the Americas from multiple sources and at different times since the first humans reached the continent in pre-Columbian times. This hypothetical scenario may explain why some mitochondrial lineages of American parasites cluster...
together with the lineages from Africa and south Asia and a few lineages are connected to Melanesian lineages, whereas most lineages have clearly diverged from all other \textit{P. vivax} populations (Figures 1 and 2). The complex population history of extant \textit{P. vivax} populations in the Americas can further complicate the assignment of imported infections from the Americas to one of the highly divergent \textit{P. vivax} subpopulations that circulate in the continent; assignment may require an additional sampling effort to cover the vast repertoire of local mitochondrial lineages.

In conclusion, currently available mitochondrial sequence data allow for a precise assignment of most imported \textit{P. vivax} infections from the Americas, southeast Asia, east Asia, and Melanesia to their regions of origin. The accumulation of new complete mitochondrial genome data from \textit{P. vivax} strains collected worldwide over the next few years should improve our ability to trace the origin of imported infections from other locations, such as south Asia, Africa, and the Middle East, which contribute the vast majority of malaria cases diagnosed in the United States.

Received October 11, 2013. Accepted for publication February 6, 2014. Published online March 17, 2014.


Acknowledgments: The authors thank Márcio M. Yamamoto, Melissa da Silva Bastos, Michelle C. C. Brandi, and María José Menezes (Department of Parasitology, University of São Paulo) for technical support.

Financial support: This research was supported by funds from the National Institutes of Health, United States (Grant RO1 AI 075416), the Centers for Disease Control and Prevention, United States, the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Brazil (Grant 2010/50333-8), and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil (Grant 590106/2011-2). P.T.R. received a scholarship from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazil and is currently supported by CNPq. J.M.P.A. is currently supported by FAPESP. M.U.F. received a senior research scholarship from CNPq, Brazil.

Disclaimer: The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors’ addresses: Priscila T. Rodrigues, João Marcelo P. Alves, and Marcelo U. Ferreira, Department of Parasitology, Institute of Bio-medical Sciences, University of São Paulo, São Paulo, Brazil, E-mails: priscilathihara@usp.br, alvesjmp@gmail.com, and mulferreir@usp.br. Ana Maria Santamaria and José E. Calzada, Department of Parasitology, Gorgas Memorial Institute of Health, Panama City, Panama, E-mails: asantamaria@gorgas.gob.pa and jcalzada@gorgas.gob.pa. Maniphet Xayavong, Monica Parise, and Alexandre J. da Silva, Division of Parasitic Diseases and Malaria, Center for Global Health, Centers for Disease Control and Prevention, Atlanta, GA, E-mails: maxi1@cdc.gov, mep01@cdc.gov, and abs8@cdc.gov.

REFERENCES


