



Genetic Markers of Adaptation of *Plasmodium falciparum* to Transmission by American Vectors Identified in the Genomes of Parasites from Haiti and South America

Massimiliano S. Tagliamonte,^{a,f,g} Charles A. Yowell,^a Maha A. Elbadry,^{b,g} Jacques Boncy,^c Christian P. Raccurt,^d Bernard A. Okech,^{b,g} Erica M. Goss,^{e,g} Marco Salemi,^{f,g} John B. Dame^{a,g}

^aDepartment of Infectious Diseases and Immunology, College of Veterinary Medicine, University of Florida, Gainesville, Florida, USA

^bDepartment of Environmental and Global Health, College of Public Health and Health Professions, University of Florida, Gainesville, Florida, USA

^cLaboratoire National de Santé Publique, Ministère de la Santé Publique et de la Population, Port-au-Prince, Haiti

^dDepartment of Tropical Medicine and Infectious Diseases, Faculty of Medicine, University of Quisqueya, Port-au-Prince, Haiti

^eDepartment of Plant Pathology, College of Agricultural and Life Sciences, University of Florida, Gainesville, Florida, USA

^fDepartment of Pathology, Immunology, and Laboratory Medicine, College of Medicine, University of Florida, Gainesville, Florida, USA

^gEmerging Pathogens Institute, University of Florida, Gainesville, Florida, USA

ABSTRACT The malaria parasite, *Plasmodium falciparum*, was introduced into Hispaniola and other regions of the Americas through the slave trade spanning the 16th through the 19th centuries. During this period, more than 12 million Africans were brought across the Atlantic to the Caribbean and other regions of the Americas. Since malaria is holoendemic in West Africa, a substantial percentage of these individuals carried the parasite. St. Domingue on Hispaniola, now modern-day Haiti, was a major port of disembarkation, and malaria is still actively transmitted there. We undertook a detailed study of the phylogenetics of the Haitian parasites and those from Colombia and Peru utilizing whole-genome sequencing. Principal-component and phylogenetic analyses, based upon single nucleotide polymorphisms (SNPs) in protein coding regions, indicate that, despite the potential for millions of introductions from Africa, the Haitian parasites share an ancestral relationship within a well-supported monophyletic clade with parasites from South America, while belonging to a distinct lineage. This result, in stark contrast to the historical record of parasite introductions, is best explained by a severe population bottleneck experienced by the parasites introduced into the Americas. Here, evidence is presented for targeted selection of rare African alleles in genes which are expressed in the mosquito stages of the parasite's life cycle. These genetic markers support the hypothesis that the severe population bottleneck was caused by the required adaptation of the parasite to transmission by new definitive hosts among the *Anopheles* (*Nyssorhynchus*) spp. found in the Caribbean and South America.

IMPORTANCE Historical data suggest that millions of *P. falciparum* parasite lineages were introduced into the Americas during the trans-Atlantic slave trade, which would suggest a paraphyletic origin of the extant isolates in the Western Hemisphere. Our analyses of whole-genome variants show that the American parasites belong to a well-supported monophyletic clade. We hypothesize that the required adaptation to American vectors created a severe bottleneck, reducing the effective introduction to a few lineages. In support of this hypothesis, we discovered genes expressed in the mosquito stages of the life cycle that have alleles with multiple, high-frequency or fixed, nonsynonymous mutations in the American populations which are rarely found in African isolates. These alleles appear to be in gene products critical for transmission through the anopheline vector. Thus, these results may inform efforts to develop novel transmission-blocking vaccines by identifying parasite proteins functionally interacting with the vector that are important for successful transmission. Further, to the best of our knowledge,

Citation Tagliamonte MS, Yowell CA, Elbadry MA, Boncy J, Raccurt CP, Okech BA, Goss EM, Salemi M, Dame JB. 2020. Genetic markers of adaptation of *Plasmodium falciparum* to transmission by American vectors identified in the genomes of parasites from Haiti and South America. *mSphere* 5:e00937-20. <https://doi.org/10.1128/mSphere.00937-20>.

Editor Lars Hviid, University of Copenhagen

Copyright © 2020 Tagliamonte et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Marco Salemi, salemi@pathology.ufl.edu, or John B. Dame, damej@ufl.edu.

Received 21 September 2020

Accepted 1 October 2020

Published 21 October 2020

these are the first whole-genome data available from Haitian *P. falciparum* isolates. Defining the genome of these parasites provides genetic markers useful for mapping parasite populations and monitoring parasite movements/introductions.

KEYWORDS Haiti, *Plasmodium falciparum*, adaptive mutations, evolutionary biology, malaria, phylogenetics, vector-borne diseases

Accumulated evidence suggests that prior to the colonization of the Americas by nations of Western Europe, malaria caused by *Plasmodium falciparum* was not a disease in the Americas (1). *P. falciparum* infection is holoendemic in West Africa; thus, a large, diverse population of this parasite was introduced into the Americas via the trans-Atlantic slave trade; records describe the disembarkation of more than 12 million African slaves in the Americas during the period of 1500 to 1875 (2, 3). African mosquitos [*Anopheles (Cellia)* spp.], to which the parasite was highly adapted for efficient transmission, did not become established in the Americas (4); thus, transmission was sustained in the diverse geographical regions of the Western Hemisphere via native *Anopheles (Nyssorhynchus)* spp. Nearly half of the African slaves (>5.5 million) disembarked in the Caribbean islands, and more than 1 million of those disembarked on the island of Hispaniola, where ~900,000 arrived between 1700 and 1800 (2). Approximately 5.8 million slaves disembarked in diverse ports in South America, where currently *P. falciparum* is actively transmitted in the 10 countries surrounding the Amazon River basin (5).

Vector control and better access to medical care in the 20th century, plus improvements to economic and housing conditions, have eliminated malaria from all Caribbean islands except Hispaniola, which encompasses Haiti and the Dominican Republic. In Haiti, this disease is endemic and becomes epidemic in the rainy seasons (6). Malaria is also transmitted in the Dominican Republic, with the highest risk in the western part of the country, near the Haitian border, where many cases are considered imported malaria, resulting in local transmission (7). Haiti is considered a low-transmission area, although it has foci with higher risk (8, 9). The Haiti Malaria Elimination Consortium (HaMEC), led by the CDC, is aiming to eliminate malaria from Hispaniola by 2022, but there are major obstacles to reaching this goal. Several studies have highlighted the high prevalence of asymptomatic infections, and their contribution to transmission is yet to be properly estimated (10–12). The malarious countries in South America have a higher incidence of malaria, and among them only Paraguay is in a pre-elimination phase (5).

Recent studies have investigated the Haitian *P. falciparum* population structure using microsatellite markers (13) or a limited number of single nucleotide polymorphisms (SNPs) (14) and found low diversity and evidence of focal transmission. Isolation of the Haitian parasite population from other populations in the Americas was suggested (14) using data from a limited number of SNP markers.

In the present study, we analyzed the phylogenetic relationship of the Haitian *P. falciparum* population to those from South America and other continents, utilizing data from whole-genome sequencing. The *P. falciparum* reference genome was made available in 2002 (15), providing a template for mapping the genomic sequences of field isolates for comparative studies. The analyses reported here were performed using newly obtained whole-genome sequence data from 21 Haitian isolates plus genome sequence data previously obtained for 16 Colombian and 11 Peruvian isolates.

The much larger number of informative characters available from whole-genome sequence data enhances the analytic power of the tools of population genetics, compared to microsatellites or a limited number of SNPs (14, 16), particularly for organisms, such as *P. falciparum*, which are characterized by low intraspecies diversity. The large sample of SNPs available from genomic data was used to perform a higher-resolution analysis of the ancestral relationship of the Haitian isolates to others from South America and Africa. In contrast to historical data, our results with data from Haitian and South American isolates describe an ancestral relationship within a well-supported monophyletic clade with para-

sites from Africa. Since a severe population bottleneck experienced by the parasites was likely responsible for the results, the hypothesis was tested that the American vectors, having evolved in isolation for ~100 million years separately from those hosting *P. falciparum* in Africa (17), have exerted powerful selective pressures on the parasite. Thus, the data set was examined for the targeted selection of rare African alleles in genes expressed in the mosquito stages of the parasite's life cycle.

RESULTS

Twenty-two *P. falciparum* isolates were collected from active malaria cases, 21 of which came from the Department of Grand'Anse, with one from the Sud-Est (South-East) Department (Fig. 1A). Parasite DNA isolated from the patients' blood samples was utilized for nearly all analyses, but four of the isolates from Grand'Anse were also adapted to long-term cultures. The primary samples for this study were thus obtained from the region of Haiti where clinical malaria is prevalent and where sufficient genetic material could be obtained for whole-genome sequencing.

Whole-genome sequencing. Parasite genomic DNA was recovered from 22 of the blood samples, and each was amplified to obtain sufficient DNA for Illumina library preparation for sequencing. Amplified material from six isolates was evaluated for possible uneven whole-genome amplification by quantitative PCR (qPCR) of six single-copy marker genes with varying A+T content and genomic locations (AMA1, CRT, GEX06, GST, MRP1, and SOAP). There was a 4- to 5-fold variation in the concentrations of these marker genes (data not shown), reflecting a known problem in obtaining uniform amplification of the *P. falciparum* genome (18). Since the concentrations of these different single-copy genes were on the same order of magnitude, the amplified DNA was deemed acceptable for genome sequencing, and a segment of the AMA1 gene was used for estimating by qPCR *P. falciparum* DNA content in all amplified samples (Text S1).

DNA obtained from parasites in low-passage-number *in vitro* culture was compared with amplified DNA from the primary isolate for two of the isolates. Less than 2% of SNPs (1.4% and 1.8%) differed between the amplified primary genomic DNA and that of the cultured parasites, and in these cases, the differences were homozygous versus heterozygous genotypes. These differences were considered to be minor; thus, in two instances when cultured parasites were available and the primary isolate had a relatively high human mitochondrial DNA (mtDNA) contamination, unamplified genomic DNA from cultured parasites was used for library preparation, sequencing, and variant calling. The final whole-genomic sequence data set consisted of data from 21 Haitian isolates, from which we obtained at least 40× genome coverage (Table S1).

SNP marker analysis. A recent study on Haitian *P. falciparum* (14) used 24 SNP markers as a barcode to investigate its population structure and its relationship to South American strains. Their spatial principal-component analysis (sPCA) results indicated that Haiti strains mostly cluster independently of the continental strains, with possible gene flow between Colombia and Hispaniola. The 24-SNP analysis was repeated by obtaining these SNPs from whole-genome sequencing data from the Haitian isolates in this study, plus 11 isolates from Peru and 16 from Colombia, retrieved from the MalariaGEN database, compiled by the MalariaGEN Community Project (19, 20) on www.malariagen.net/ and stored on the European Nucleotide Archive (<http://www.ebi.ac.uk/ena>). To these data were added 40 published barcode sequences from Haitian parasites from the study by Charles et al. (14). For consistency with the original pipeline, isolates missing data from more than 5 markers or containing more than 1 multiallelic site were discarded. This led to dropping two Haitian isolates. Similar to Charles et al. (14), we found that two Colombian isolates clustered with the Haitian population (Fig. 1B), indicating the potential for gene flow between the two areas. These findings are consistent with previous analyses that compared other Haitian departments (13, 14), and together, the results indicate that the parasite population in Haiti is a single unstructured population with focal transmission. Thus, the isolates utilized for genomic SNP analyses appear to be suitably representative of *P. falciparum* in Haiti.

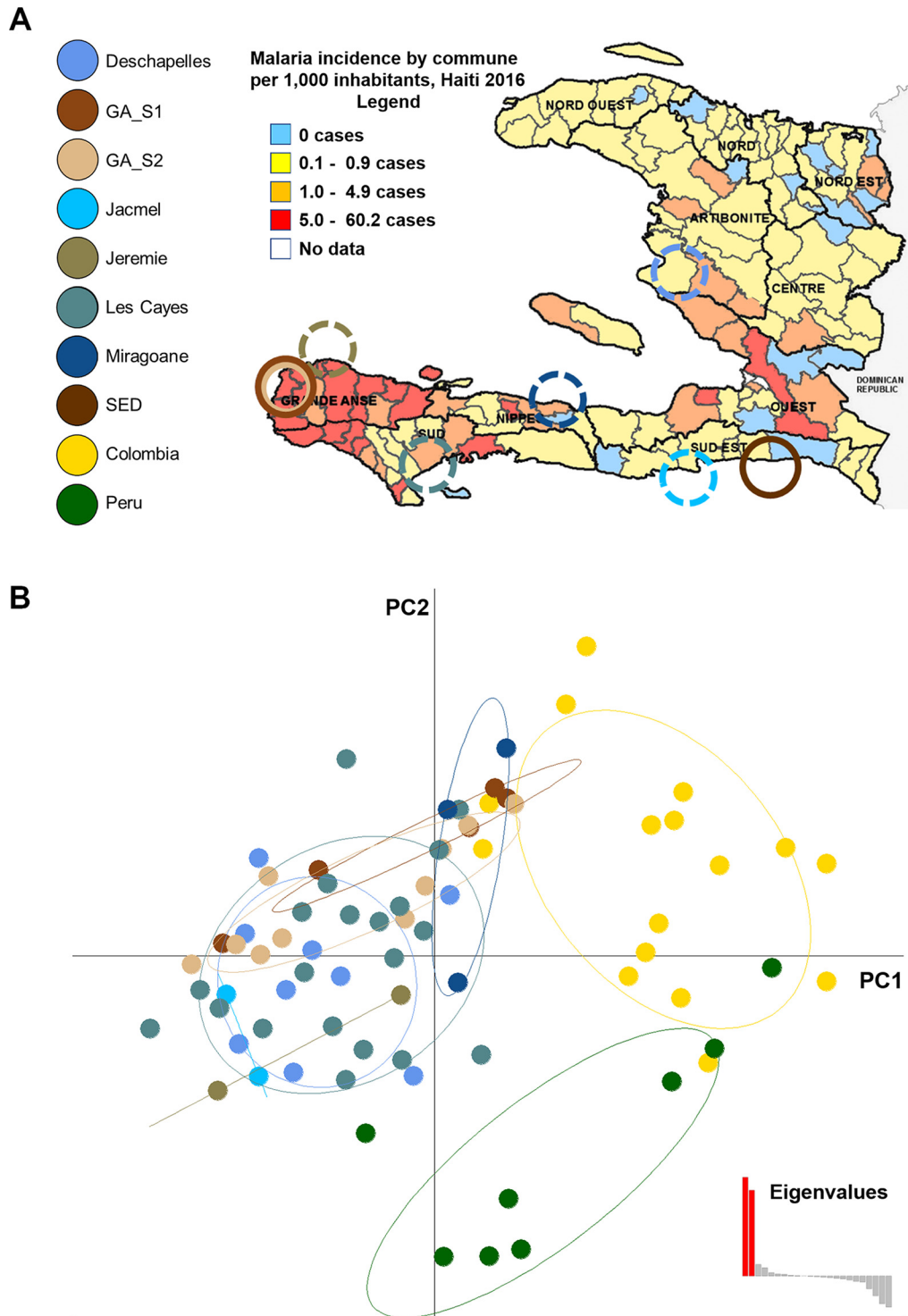


FIG 1 Clustering analyses on 24 SNP markers. (A) Map of Haiti and sampling sites. Incidence data are from reference 74. Dotted lines indicate origins of samples from reference 14; solid lines indicate origins of our samples, for which whole-genome data are available. (B) sPCA using 24 SNP barcodes from reference 14; following original paper filtering protocol for consistency, one of the isolates that were filtered out was our sample from the Sud-Est Department.

Variant calling for population genetic analyses. Two variant calling iterations were performed, one using the 21 Haitian isolates only (Haitian data set) and one using these data plus whole-genome sequencing data for 149 isolates from numerous malarious regions worldwide retrieved from the MalariaGEN database, for a total of 170

isolates (WW data set). Genome sequence data from a total of 27 samples were available from South America, 16 of which originated from Colombia and 11 from Peru. Based on historical data regarding the trans-Atlantic slave trade (3, 21), we downloaded genome sequence data from isolates recovered in West Africa: Gambia, Ghana, and Cameroon (10 samples from each country). We also downloaded data from 10 samples from Central Africa (Democratic Republic of Congo) and 10 from each Kenya, Malawi, and Tanzania to represent East Africa. Data on 22 genomes from Papua New Guinea were downloaded as representative of Oceania and 10 each from Cambodia, Myanmar, and Thailand to represent Southeast Asia. Data sets chosen were from paired-end Illumina libraries, with a minimum 100-nucleotide read length. The largest data sets from each country were chosen and whenever possible were larger than 2 Gb. An exception to the read length rule was made for South American data sets, due to the scarcity of samples available; thus, these libraries have 54- to 100-nucleotide paired-end reads. Details regarding the downloaded data are in the supplemental material (Table S2).

When variant calling was done on Haitian isolates versus the coding regions of the 3D7 genome, 447,339 variants were obtained. A filtering pipeline based on the one published by Manske et al. (19) was applied, and after filtering, 22,044 variant loci were retained as reliable for further analyses. These were located in 3,189 different genes. The *P. falciparum* genome is particularly difficult to map; thus, reliably mapped reads come from just 60% of its ~5,300 annotated genes (15). About 90% (19,901) of the variants were SNPs, while the rest were indels, most of which were in frame. Almost three-quarters of the SNPs (14,148) were nonsynonymous mutations resulting in an amino acid change in the translated gene product. This is a known phenomenon, and it has been attributed to continuous positive selection exerted by the host (22). The complexity of infection was assessed using THE REAL McCOIL (23), and all Haitian isolates represented single infections with comparatively few heterozygous alleles consistent with the low-transmission setting.

For phylogenetic purposes, variant calling was repeated on the WW data set. Variant calling produced almost 1,400,000 variants, which were reduced to about 139,000 variants by the filtering process. In order to determine the ancestral relationship of the Haitian *P. falciparum* to isolates from Africa, Asia, and South America, sites with indels, conserved mutations (those found in all isolates), and singletons were removed, since they provide no phylogenetic information, leaving 50,469 sites in 3,106 genes representing almost 60% of the genes in the nuclear genome.

Population genetics using whole-genome SNPs. Principal-component analysis was used to study the clustering of the samples without assuming an evolutionary model. The PCA shows a clear separation between isolates from different continents (Fig. 2). Haitian isolates clustered separately from South American isolates, while the Peruvian and Colombian clusters partially overlapped. The Haitian isolates grouped into two different clusters, one of which consisted of nine virtually identical isolates, evidence of an apparent epidemic expansion (Fig. 2A). Discriminant analysis of the principal components (DAPC) was also performed, which minimizes within-group variance, while maximizing between-group variance (24). The results are similar to those of the PCA, yielding evidence for seven different population clusters (Fig. S1).

The sample set was reduced from 170 to 149 sequences by removing samples resulting from apparent clonal (epidemic) expansion (Fig. S2) or having complexity of infection (COI). The Haitian data set was reduced from 21 to nine isolates by selecting the six independent isolates and one representing each of the three epidemic expansions (Table S1). Three isolates from Peru were also removed from two epidemic expansions. When only one isolate representing each epidemic expansion among Haitian and Peruvian samples was utilized in the PCA analysis, similar results were obtained (Fig. 2B and C).

Phylogenetic relationship of the Haitian parasite to the African, Asian, and South American strains using whole-genome SNPs. Based on historical data, we

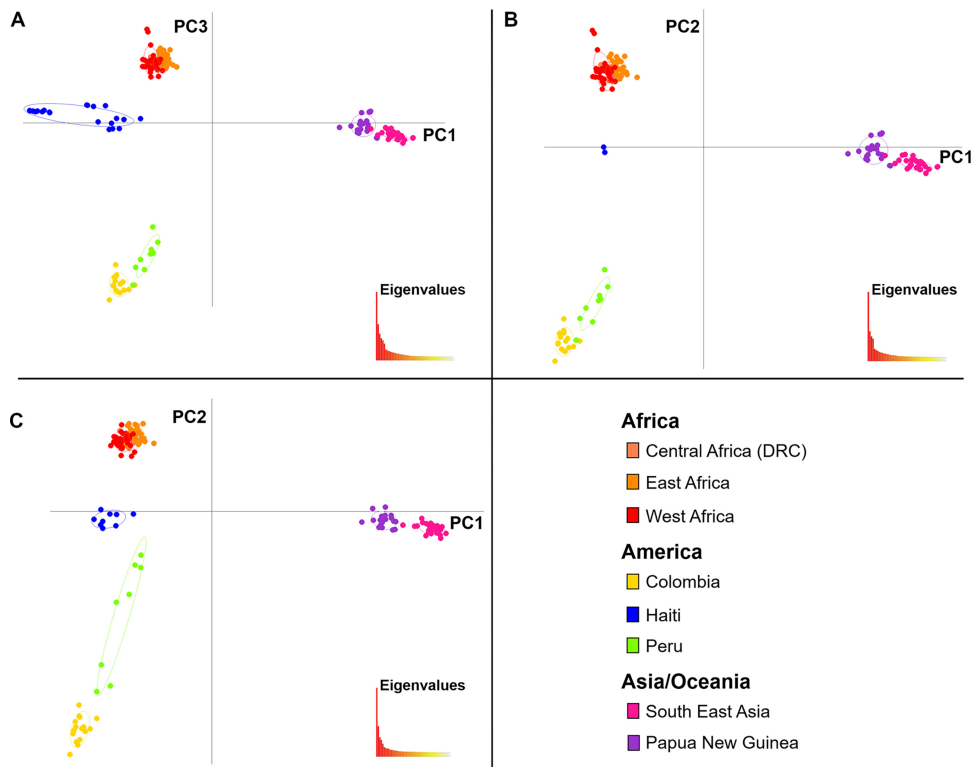


FIG 2 Two-dimensional plot of the principal-component analysis using 50,469 SNPs. (A) PCA performed using all samples. (B) To account for the possibility that covariation among parasites that share recent ancestry has a disproportionate impact on PC weightings, the PCA was repeated keeping only a single representative of the two dominant Haitian lineages. (C) PCA performed using a subset of 149 sequences created by removing samples resulting from apparent clonal (epidemic) expansion identified in Fig. S2 or having complexity of infection.

would expect the *P. falciparum* strains found in the Americas to have a paraphyletic origin. This scenario was affirmed by Joy et al., using mtDNA sequences (25), and by Yalcindag et al. (26), using genomic SNPs consisting of a mix of coding and noncoding loci. The first data set we used to investigate the ancestry of American parasite populations was the whole-genome SNP alignment (50,469 loci). From these we removed SNPs under strong positive selection in multiple ways. The first alignment subset included putatively neutral SNPs only, as determined by Bayescan v.2.1 (27–29). This program identifies candidate loci under selection using differences in allele frequencies between populations. After removal of these loci, 48,194 remained. A subset consisting of 15,020 sites was generated by keeping synonymous mutations only. Multinucleotide variants (MNVs) were also eliminated, as different combinations of SNPs in the same codon might result in a nonsynonymous mutation in part of the samples. It is recognized that this approach does not ensure the exclusive selection of neutral mutations, since different synonymous codons could have an impact on gene expression regulation (30). Recognizing that one of the main drivers of selection on *P. falciparum* upon migration to new areas is the change of vector species (31–34), we also removed from the data sets for analysis the data from genes which are ≥ 10 -fold upregulated in the late gametocyte (gametocyte V), ookinete, and sporozoite stages (rather than the asexual blood stages) (35, 36). The phylogenetic signal of these alignments was verified as shown in Fig. S3.

The neighbor-joining (NJ) tree from the synonymous SNPs minus genes upregulated in the sexual and mosquito stages (13,597 loci) is depicted in Fig. 3. To calculate the tree, we used the log-det model, which has been shown to be robust to biased base composition (37). The other trees are reported in the supplemental material and have similar topologies (Fig. S4), with the separation of populations between continents

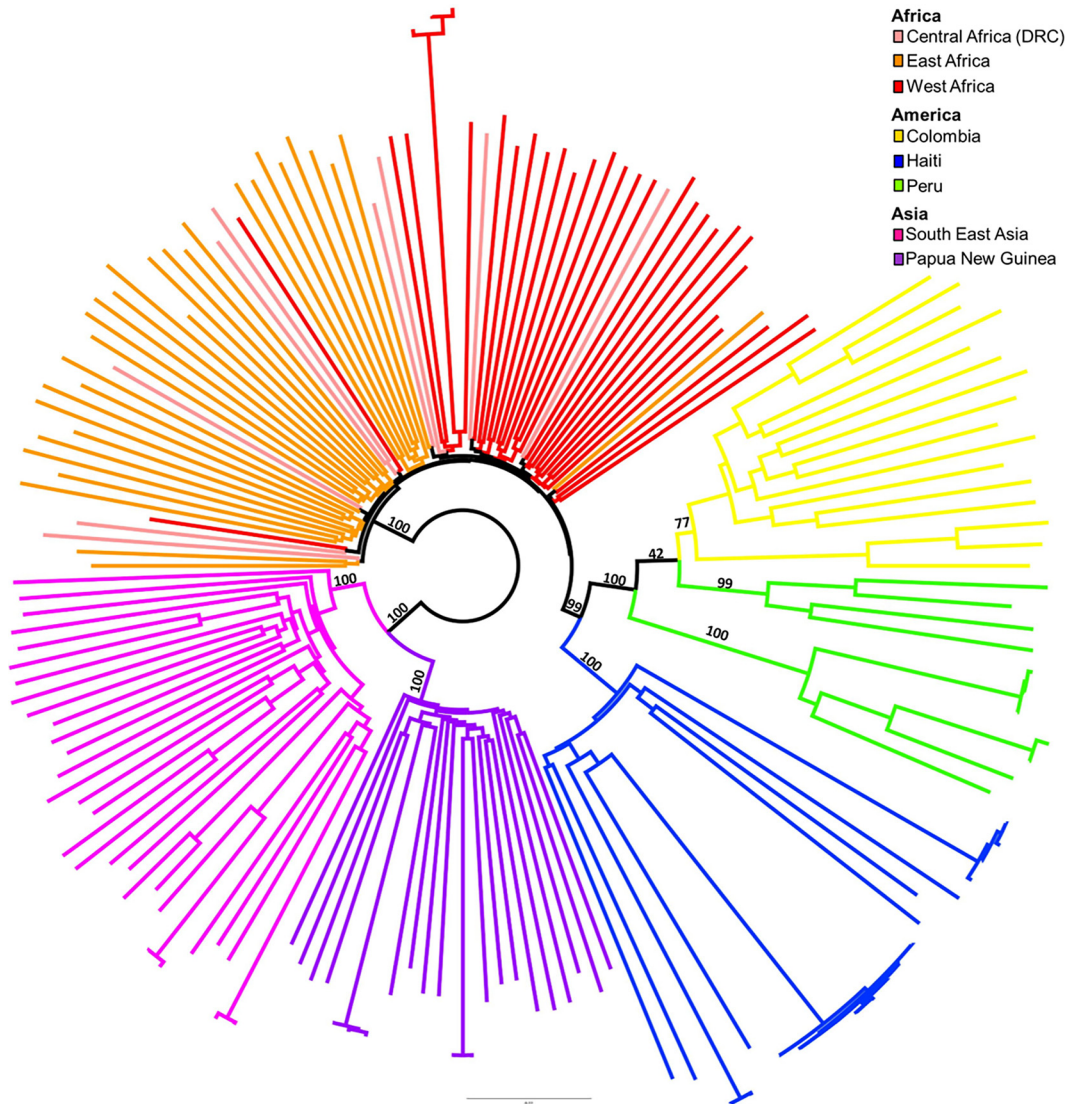


FIG 3 Neighbor-joining tree calculated using synonymous SNPs. MNVs and SNPs in genes upregulated in the mosquito stages of the parasite life cycle were removed (13,627 sites were retained). The tree was calculated using the log-det model.

having strong bootstrap support. The low support for African clades matches the parasite's known high diversity and high transmission rates on that continent (19, 38). The monophyly of the American clade is always well supported, and the Haitian *P. falciparum* population is isolated. Looking at the different trees and bootstrap support values, the population structure of the Colombian and Peruvian parasites is not fully resolved and is probably partially mixed. The isolation of the Haitian population has important epidemiological implications, should a larger sample size, covering other areas of the Western Hemisphere, confirm these findings.

Selective pressure of American mosquito vectors on *P. falciparum* genes. The results of the phylogenetic analyses, showing a monophyletic American clade, are counterintuitive, because of the scenario involving continuous introduction from different parts of the African continent through the slave trade. Considering the extent to which we went to remove loci potentially under selection, these results likely derive from the use of coding SNPs and reflect a genuine bottleneck which the parasite went through after it was introduced into the Americas. The most dramatic change to which *P. falciparum* had to adapt was transmission by novel definitive hosts, as the American *Anopheles* (*Nyssorhynchus*) spp. diverged from the African *Anopheles* (*Cellia*) spp. ~100

TABLE 1 *P. falciparum* genes upregulated in the mosquito stage

Gene name	Gene ID	Product description	No. of variant codons in American strains ^a	Upregulation (fold) in stage ^b		
				Gametocyte V	Ookinete	Sporozoite
TRAP	PF3D7_1335900	Thrombospondin-related anonymous protein	7	0	0	3,235.7
CTRP	PF3D7_0315200	Circumsporozoite- and TRAP-related protein	5	0	582.1	0
PSOP26	PF3D7_1244500	Conserved <i>Plasmodium</i> protein, unknown function	5	0	133	0
P47	PF3D7_1346800	6-cysteine protein	5	37.6	27.8	0
NA	PF3D7_0511400	Conserved <i>Plasmodium</i> protein, unknown function	2	0	0	147.4
SIAP1	PF3D7_0408600	Sporozoite invasion-associated protein 1	1	0	0	270.7
NA	PF3D7_0515500	Amino acid transporter, putative	1	27.2	0	0
PBLP	PF3D7_0818600	BEM46-like protein, putative	1	0	0	33.4
CRMP1	PF3D7_0911300	Cysteine repeat modular protein 1	1	0	0	14
ICP	PF3D7_0911900	Falstatin	1	0	0	11.4
NA	PF3D7_0924600	Conserved <i>Plasmodium</i> protein, unknown function	1	29.4	19.8	0
NA	PF3D7_1020200	Conserved <i>Plasmodium</i> protein, unknown function	1	86.5	223.6	0
CRMP3	PF3D7_1208200	Cysteine repeat modular protein 3	1	0	0	16.8
P48/45	PF3D7_1346700	6-cysteine protein	1	25	15.8	0
NA	PF3D7_1348400	Conserved <i>Plasmodium</i> membrane protein, unknown function	1	20	0	0
NA	PF3D7_1403200	Conserved <i>Plasmodium</i> protein, unknown function	1	158.9	82.1	0
SOAP	PF3D7_1404300	Secreted ookinete adhesive protein, putative	1	0	70.8	0

^aVariant codons which have ≥ 0.7 frequency in America and ≤ 0.3 frequency in Africa.

^bDetermined by comparison to blood stages.

million years ago (17). Evidence has recently emerged regarding the impact that novel vectors have had on the Pfs47 gene (31–34, 39, 40). Allelic changes in Pfs47 partially control infectivity for different vectors, suggesting that this and perhaps additional genes are under selection during adaptation to novel vector species.

A preliminary test was performed on our alignment as a way to identify potential genes with mutations necessary for adaptation to the American vectors. For these analyses, the Haitian data set was reduced from 21 to the nine isolates representative of the Haitian subclades (Fig. 3) by selecting the six independent isolates and one representing each of three apparent epidemic expansions (Table S1). Three isolates from Peru were also removed from two apparent epidemic expansions. *P. falciparum* expression data (35, 36) were downloaded from PlasmoDB26 database (41), and we identified those genes that are upregulated in the gametocyte, ookinete, and sporozoite stages as opposed to the blood stages. We then split the alignment in two, depending on SNPs belonging to genes upregulated in the mosquito stages compared to the blood stages. We compared the ratio of nonsynonymous to synonymous substitutions (dN/dS); dN/dS is statistically higher in the subset data from genes upregulated in the mosquito stages (1.01 versus 0.53; $P < 0.01$), further suggestive of the impact that the vector might have had on the parasite population.

As evolutionary rates and substitution patterns may vary between genes, we tried to narrow our data set to a few more likely candidate genes for further, in-depth analyses. Genes under strong selection were preliminarily identified by filtering to identify nonsynonymous mutations having a frequency of ≥ 0.7 in the 33 American isolates (9 Haitian, 16 Colombian, and 8 Peruvian) and ≤ 0.3 in the 70 African isolates; 68 variant genes were retained of the 3,106-gene data set. In this data set, 397 genes (13.1%) were upregulated in one or more of these sexual/mosquito stages of the parasite life cycle, but among the retained 68 genes, the frequency of sexual/mosquito-stage genes was almost 2-fold higher, with 17 (25%) being upregulated in one or more of these stages (Table 1). Twelve of these genes contained a single variant codon with differential frequency; one had two such mutations, three had five, and one had seven. The four genes with the most mutations were TRAP (PF3D7_1335900), CTRP (PF3D7_0315200), PSOP26 (PF3D7_1244500), and Pfs47 (PF3D7_1346800).

The consensus sequences of these genes were compiled for a representative subset of the isolates, as described in Text S1. No sign of recombination was found by RDP4

TABLE 2 FUBAR results for CTRP^a

Codon no.	Amino acids	Amino acid frequencies in:				FUBAR posterior probability of positive selection
		Haiti	South America	America	Africa	
17	H, P	1, 0	0.96, 0.04	0.97, 0.03	0, 1	0.95
319	N, D	1, 0	0.96, 0.04	0.97, 0.03	0, 1	0.82
659	R, Q	1, 0	1, 0	1, 0	0.3, 0.7	0.94
1046	K, N	1, 0	0.88, 0.12	0.91, 0.09	0, 1	0.83
1260	N, S	1, 0	0.92, 0.08	0.94, 0.06	0.27, 0.73	0.96
2093	P, S	1, 0	0.62, 0.38	0.73, 0.27	0, 1	0.94
2098	Q, E	1, 0	0.96, 0.04	0.97, 0.03	0.17, 0.83	0.96

^aPF3D7_0315200 (circumsporozoite and TRAP-related protein; 2,114 amino acids [aa]). Only codons with differential frequencies between Haiti and Africa were retained.

analysis (42), performed as described by Mavian et al. (43). Selection analysis was performed by a fast, unconstrained Bayesian approximation algorithm (FUBAR) (44). Using a posterior probability cutoff of 0.8 and codon frequencies of ≥ 0.7 in the American isolates and ≤ 0.3 in African isolates, we identified 24 codons among these four genes under positive selection (Tables 2 to 5). Complete FUBAR results are shown in Table S3.

DISCUSSION

Whole-genome analysis for *P. falciparum* is particularly complicated, due to the repetitive nature and high AT content of the genome, requiring an intense effort to eliminate analysis artifacts. Further, *P. falciparum* also regularly undergoes sexual recombination, which constitutes an additional obstacle to phylogenetic analyses. Despite these challenges and potential limitations, the resulting data can help answer a multiplicity of questions, including shedding light on the variety of evolutionary drivers acting on the parasite (19, 20, 45). The whole-genome sequence data reported here are the first available from Haitian *P. falciparum* isolates. These data were obtained from isolates obtained in Grand'Anse plus one in Sud-Est, two regions with the highest rates of transmission of malaria in Haiti. Comparisons of the genotypes of these isolates to those obtained by Charles et al. (14) indicate that data obtained from the isolates analyzed reasonably represent the entire Haitian *P. falciparum* population.

A large proportion of the coding SNPs common to the Haitian *P. falciparum* parasite population are now known, and there are many SNPs that are unique and private to the Haiti population, based on our analysis. This knowledge will aid in the elimination of malaria from the island, by offering the ability to discern imported infections from indigenous ones. Reassessing this scenario with data from additional strains from Central America and the eastern areas of South America will be essential to monitor movements of the parasite between these regions and consequent possible reintroductions.

The shape of the NJ tree (Fig. 3), with long terminal branches, could derive from bottlenecks resulting from multiple strategies to fight malaria but are more likely due to less recent ones resulting from the necessity of adapting to new definitive hosts following migration to Asia and the Americas, followed by genetic drift. The presence

TABLE 3 FUBAR results for PSOP26^a

Codon no.	Amino acids	Amino acid frequencies in:				FUBAR posterior probability of positive selection
		Haiti	South America	America	Africa	
209	F, V	1, 0	0.88, 0.12	0.91, 0.09	0, 1	0.84
494	P, R	1, 0	1, 0	1, 0	0, 1	0.80
664	R, S	1, 0	1, 0	1, 0	0, 1	0.93
722	K, N	1, 0	1, 0	1, 0	0, 1	0.87
736	N, K	1, 0	1, 0	1, 0	0.07, 0.93	0.96

^aPF3D7_1244500 (conserved *Plasmodium* protein, unknown function; 810 aa). Only codons with differential frequencies between Haiti and Africa were retained.

TABLE 4 FUBAR results for TRAP^a

Codon no.	Amino acids	Amino acid frequencies in:				FUBAR posterior probability of positive selection
		Haiti	South America	America	Africa	
66	K, N	0.56, 0.44	0.92, 0.08	0.82, 0.18	0.1, 0.9	1.00
83	E, D	0.78, 0.22	1, 0	0.94, 0.06	0.23, 0.77	0.96
92	I, V	0.78, 0.22	0.92, 0.08	0.88, 0.12	0, 1	0.98
277	L, I, T	0.89, 0.11, 0	0.92, 0.08, 0	0.91, 0.09, 0	0.27, 0.7, 0.03	0.99
297	Q, H, D	0.67, 0.33, 0	0.92, 0.08, 0	0.85, 0.15, 0	0, 0.53, 0.47	0.97
509	R, K	0.67, 0.33	0.83, 0.17	0.79, 0.21	0, 1	0.92
541	F, Y	0.67, 0.33	0.92, 0.08	0.85, 0.15	0, 1	0.84

^aPF3D7_1335900 (thrombospondin-related anonymous protein; 574 aa). Only codons with differential frequencies between Haiti and Africa were retained.

of such bottlenecks represents an obstacle difficult to overcome in any coalescent analysis. However, coding SNPs have been valuable for differentiating among populations (19, 20, 46, 47), and these data show that the Haitian population is clearly distinct from Colombian and Peruvian populations.

Based upon historical data, we would expect the American parasite population to be paraphyletic; however, this is not seen. The monophyly of the American clade as seen in our analyses likely derives from the use of coding SNPs. Our synonymous SNPs may be largely neutral (30, 48), but they have probably been subjected to repeated selective sweeps and were fixed along with mutations that are advantageous in Haiti and other regions of the Americas. Since the American isolates appear as a single clade despite millions of introductions in widely separated geographical regions, this suggests a common selective bottleneck. Branching within this clade into distinct subclades is then created by local bottlenecks and geographical isolation followed by genetic drift.

The greatest challenge to the parasite in the transition to the Americas was the sudden change in the definitive host, resulting from the transoceanic migration of the intermediate host. Finding the proportion of genes under strong selection expressed in the sexual/mosquito stages to be twice that which was expected provided further impetus to evaluate the hypothesis that the common bottleneck experienced was the required adaptation to transmission by *Anopheles (Nyssorhynchus)* spp. in all locales. This was followed by further adaptation to additional species-specific selective pressures presented by local vectors of this subgenus. Genes which are upregulated in the mosquito stages of the life cycle of the parasite are thus potentially under selective drive when the parasite shifts from one vector species to another, and they are scattered across the *P. falciparum* genome on different chromosomes, which would exacerbate the genetic bottleneck acting upon the parasite population.

It is estimated that the African and American anopheline vectors evolved independently for nearly 100 million years (17), offering an evolutionary basis for major differences in the genus. Further, evidence has recently emerged regarding the impact that the immune systems of novel vectors have had on the parasite genome. Key mutations in the Pfs47 gene allow *P. falciparum* to escape the complement-like immune system of its definitive host, where different optimal combinations of the amino acid substitutions are necessary for the successful infection of vector species in different

TABLE 5 FUBAR results for Pfs47^a

Codon no.	Amino acids	Amino acid frequencies in:				FUBAR posterior probability of positive selection
		Haiti	South America	America	Africa	
178	V, I	1, 0	0.83, 0.17	0.87, 0.13	0, 1	0.86
236	I, T	1, 0	1, 0	1, 0	0, 1	0.85
242	L, S	1, 0	1, 0	1, 0	0, 1	0.86
247	A, V	1, 0	1, 0	1, 0	0, 1	0.86
248	L, I	1, 0	1, 0	1, 0	0.17, 0.83	0.94

^aPF3D7_1346800 (6-cysteine protein; 439 aa). Only codons with differential frequencies between Haiti and Africa were retained.

regions of the world (31–34, 39, 40). The crucial role of this gene product was first identified through linkage analysis of the progeny of a cross between GB4 (an African isolate) and 7G8 (a Brazilian isolate). While multiple selective pressures have shaped the low-diversity American parasite populations, our research correctly identified Pfs47 and the four codons (codons 236, 242, 247, and 248) for which there is *in vivo* evidence of being under selection by the vector (31). The fifth codon identified here (codon 178) might play an additional role in the successful infection of *Anopheles albimanus*, the dominant vector on Hispaniola, since presumably progenitor strains giving rise to 7G8 were adapted for successful transmission via *Anopheles darlingi*, the dominant vector in Brazil.

The other three genes we identified have not been previously reported to contribute to adaptation to specific vectors. TRAP (PF3D7_1335900) is a protein with two adhesive domains (the A and TSR domains) and is essential for trafficking to the salivary glands of the mosquito (49). Five of the seven SNPs under selection in American isolates (three in the A domain and two in TSR) are found in these two regions of the protein. CTRP (PF3D7_0315200) is a conserved protein essential to ookinete motility and invasion of mosquito midgut epithelium (50). This protein has a COOH-terminal transmembrane domain and a short cytoplasmic domain with a possible rhomboid protease cleavage site adjacent to the external face of the transmembrane domain (50). The function of the fourth gene, PSOP26 (PF3D7_1244500), is unknown; however, our results suggest an important role for this protein, which is expressed in the ookinete stage, in the interface with the vector. Interestingly, both the Honduran isolate HB3 (NCBI accession no. GCA_900631985.1) and the Salvadoran isolate Santa Lucia (NCBI accession no. GCA_000150455.3) present all 24 of the mutations located in these four genes, which we found under selection in the American isolates evaluated here. Conversely, the isolate 7G8, originally from Brazil (NCBI accession no. GCA_000150435.3), presents only 17 of the 24 mutations, where the predominant African alleles are found: codon 178 in the Pfs47 gene; codons 1046, 1260, and 2093 in the CTRP gene; and codons 66, 509 and 541 in the TRAP gene. This suggests that different American vectors may exert different selective pressures on some of the codons. Another possibility is that not all of the codons that were positive in the *in silico* analyses are actually under selection. Some SNPs might be associated by chance with other positively selected codons and thus have “hitchhiked” on the selective drive. Also, low diversity and isolation of various American populations will have had an impact on selection analyses, thus potentially resulting in some false positives.

The present results constitute a starting point for investigating gene products which have a high likelihood of having a crucial interaction with the mosquito vector. The next step would be to assess the effect of such mutations *in vivo* on transmission efficiency in different vectors. This would help not only define the effect of specific alleles but also illustrate the functions of protein products that are poorly characterized thus far. This knowledge of relevant protein variants also might help in developing transmission blocking vaccines by identifying genes products critical for transmission. Such vaccines may target either a crucial parasite protein, as recently demonstrated for Pfs47 (51), or the functional contact between that parasite protein and the vector (52, 53).

Other genes, expressed in the human host, might also be under mosquito selective pressure, indirectly. As an example, *A. albimanus*, the vector on Hispaniola, has a strong (20:1) preference for livestock versus humans and is exophilic and exophagic (54), greatly reducing the intensity of transmission and severity of infection. Parasite alleles promoting enhanced transmission through the Haitian vector and persistent human infections with long-term production of gametocytes were probably under immediate selection. Further, these selective pressures would be expected to limit the productive gene flow from *P. falciparum* from other hemispheres, including drug resistance alleles, as these strains would not be competitive with the adapted local strains.

Our comparatively small sample and the small number of published sequences from American parasites limit the current phylogenetic reconstruction. Recent

technologies, such as selective whole-genome amplification (sWGA) and single-cell sequencing, have been successfully implemented with *Plasmodium* spp. (55–57), and blood spots collected on filter paper are finally usable for whole-genome analysis. This will facilitate expanding access to isolates in the future, adding greatly to the data set of whole-genome sequences available from *P. falciparum* populations in the Americas.

MATERIALS AND METHODS

Sample collection and processing. This study was conducted in accordance with institutional review board guidelines and requirements of the University of Florida and the ethical review board of the Haitian Ministry of Health, after all permits and approvals had been obtained (IRB201400225; MSPP reference no. 1314-62). Blood samples were collected with informed consent from patients who were positive for malaria by a rapid diagnostic test during the period of September 2014 through February 2015 and subsequently deidentified. Aliquots of some of the samples were also placed in *in vitro* culture (58). Leukocytes were removed using either CF11 cellulose columns (59) or Plasmodipur filters (catalog no. 8011Filter25u; EuroProxima BV). Following DNA extraction (genomic DNA midikit; Zymo, Inc.) and in preparation for constructing a sequencing library, the amount of parasite and human DNA recovered from each sample was estimated by TaqMan-based qPCR as described in Text S1. DNA from each primary isolate was utilized for microsatellite analysis and for sequencing library preparation, as described in Text S1.

Genomic sequencing and data quality analysis. Sequencing was performed on an Illumina MiSeq system using the Illumina MiSeq reagent kit v3 reagents according to the manufacturer's instructions to generate 300-nucleotide paired-end reads.

The general quality of the sequence data was assessed using Fastqc v. 0.11.4 (60). Primer dimers and leftover insert sequences were removed from the Haitian sample sequences with Trimmomatic v.0.36 (61), and Trim Galore (62) was used for downloaded data sets.

Variant calling and filtering. Variant calling was performed by following directions from the SAMtools 1.3.1 pipeline (63) as described at http://www.htslib.org/workflow/#mapping_to_variant and in reference 64. Results were annotated using SnpEff v.4.2 (65). Since the *P. falciparum* genome has a high AT bias (~82% AT content) and is rich in repetitive segments, it was necessary to further refine these results, using a protocol defined by Manske et al. (19), which we implemented with custom scripts in R language v.3.3.1 (66) through the RStudio shell (67) and Shustring (68). Details of this pipeline and the impact of filtering steps on the two data sets are reported at https://drive.google.com/file/d/1dA_TPvuJGEiz41w2fU8Q82bOCJNU4JJ/view?usp=sharing.

Principal-component analysis (PCA) and spatial principal-component analysis (sPCA) were performed on the resulting data as described in Text S1. Genetic distance was calculated with the general time-reversible model versus transitions and transversions and the Xia test (69) was performed using DAMBE v. 6.4.81 (70). The final alignment (13,627 sites) was scanned for recombination using GARD (71) as implemented in HyPhy (72).

Phylogenetic analyses were performed as described in Text S1. *dN* and *dS* were calculated using DnaSP v.6.12.03 (73).

Data availability. The sequence data obtained from the 21 Haitian isolates used in this study were filtered to remove sequences not mapping to the *P. falciparum* genome. Read pairs for which at least one read mapped to the *P. falciparum* genome were uploaded to the SRA database with project number [PRJNA603776](https://www.ncbi.nlm.nih.gov/sra/PRJNA603776).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

TEXT S1, DOCX file, 0.04 MB.

FIG S1, TIF file, 0.2 MB.

FIG S2, TIF file, 0.3 MB.

FIG S3, TIF file, 0.5 MB.

FIG S4, TIF file, 1.4 MB.

TABLE S1, DOCX file, 0.02 MB.

TABLE S2, DOCX file, 0.04 MB.

TABLE S3, DOCX file, 0.6 MB.

ACKNOWLEDGMENTS

Research was supported by a grant from the Emerging Pathogens Institute of the University of Florida (14-6).

This publication uses data from the MalariaGEN *Plasmodium falciparum* Community Project as described in reference 47. We thank E. Yalcindag for the helpful discussion of his data early in this study.

REFERENCES

- Rodrigues PT, Valdivia HO, de Oliveira TC, Alves JMP, Duarte AMRC, Cerutti-Junior C, Buey JC, Brito CFA, de Souza JC, Hirano ZMB, Bueno MG, Catão-Dias JL, Malafronte RS, Ladeia-Andrade S, Mita T, Santamaria AM, Calzada JE, Tantular IS, Kawamoto F, Rajmakers LRJ, Mueller I, Pacheco MA, Escalante AA, Felger I, Ferreira MU. 2018. Human migration and the spread of malaria parasites to the New World. *Sci Rep* 8:1993. <https://doi.org/10.1038/s41598-018-19554-0>.
- Anonymous. 2013. Trans-atlantic slave trade database. Emory University. <https://slavevoyages.org/voyage/database#statistics>. Accessed 11 August 2017.
- Beckles H. 2002. Slave voyages: the transatlantic trade in enslaved Africans. UNESCO, Paris, France.
- Parmakelis A, Russello MA, Caccone A, Marcondes CB, Costa J, Forattini OP, Sallum MA, Wilkerson RC, Powell JR. 2008. Historical analysis of a near disaster: *Anopheles gambiae* in Brazil. *Am J Trop Med Hyg* 78:176–178. <https://doi.org/10.4269/ajtmh.2008.78.176>.
- WHO. 2017. World Malaria Report 2016. World Health Organization, Geneva, Switzerland.
- Boncy PJ, Adrien P, Lemoine JF, Existe A, Henry PJ, Raccurt C, Brasseur P, Fenelon N, Dame JB, Okech BA, Kaljee L, Baxa D, Prieur E, El Badry MA, Tagliamonte MS, Mulligan CJ, Carter TE, Beau de Rochars VM, Lutz C, Parke DM, Zervos MJ. 2015. Malaria elimination in Haiti by the year 2020: an achievable goal? *Malar J* 14:237. <https://doi.org/10.1186/s12936-015-0753-9>.
- Clinton Health Access Initiative. 2013. The feasibility of malaria elimination on the island of Hispaniola, with a focus on Haiti: an assessment conducted January–June 2013.
- Eisele TP, Keating J, Bennett A, Londono B, Johnson D, Lafontant C, Krogstad DJ. 2007. Prevalence of *Plasmodium falciparum* infection in rainy season, Artibonite Valley, Haiti, 2006. *Emerg Infect Dis* 13:1494–1496. <https://doi.org/10.3201/eid1310.070567>.
- von Fricken ME, Weppelmann TA, Lam B, Eaton WT, Schick L, Masse R, Beau De Rochars MV, Existe A, Larkin J, Ill, Okech BA. 2014. Age-specific malaria seroprevalence rates: a cross-sectional analysis of malaria transmission in the Ouest and Sud-Est departments of Haiti. *Malar J* 13:361. <https://doi.org/10.1186/1475-2875-13-361>.
- Elbadry MA, Al-Khedry B, Tagliamonte MS, Yowell CA, Raccurt CP, Existe A, Boncy J, Weppelmann TA, Beau De Rochars VE, Lemoine JF, Okech BA, Dame JB. 2015. High prevalence of asymptomatic malaria infections: a cross-sectional study in rural areas in six departments in Haiti. *Malar J* 14:510. <https://doi.org/10.1186/s12936-015-1051-2>.
- Raccurt CP, Brasseur P, Lemoine F, Ciceron M, Existe A, Boncy J. 2014. Epidemiological characteristics of malaria in the village of Corail, Grand'Anse, Haiti. *Bull Soc Pathol Exot* 107:337–341. (In French.) <https://doi.org/10.1007/s13149-014-0391-4>.
- Raccurt CP, Ciceron M, Existe A, Boncy J, Brasseur P, Lemoine F. 2015. Gametocyte carriage in asymptomatic *Plasmodium falciparum* infections in Haiti (2010–2013). *Bull Soc Pathol Exot* 108:14–16. (In French.) <https://doi.org/10.1007/s13149-014-0367-4>.
- Carter TE, Malloy H, Existe A, Memnon G, St Victor Y, Okech BA, Mulligan CJ. 2015. Genetic diversity of *Plasmodium falciparum* in Haiti: insights from microsatellite markers. *PLoS One* 10:e0140416. <https://doi.org/10.1371/journal.pone.0140416>.
- Charles M, Das S, Daniels R, Kirkman L, Delva GG, Destine R, Escalante A, Villegas L, Daniels NM, Shigyo K, Volkman SK, Pape JW, Golightly LM. 2016. *Plasmodium falciparum* K76T pfcrt gene mutations and parasite population structure, Haiti, 2006–2009. *Emerg Infect Dis* 22:786–793. <https://doi.org/10.3201/eid2205.150359>.
- Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, Paulsen IT, James K, Eisen JA, Rutherford K, Salzberg SL, Craig A, Kyes S, Chan MS, Nene V, Shallom SJ, Suh B, Peterson J, Angiuoli S, Pertea M, Allen J, Selengut J, Haft D, Mather MW, Vaidya AB, Martin DM, Fairlamb AH, Fraunholz MJ, Roos DS, Ralph SA, McCadden GI, Cummings LM, Subramanian GM, Mungall C, Venter JC, Carucci DJ, Hoffman SL, Newbold C, Davis RW, Fraser CM, Barrell B. 2002. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 419:498–511. <https://doi.org/10.1038/nature01097>.
- Downing T, Stark O, Vanaerschoot M, Imamura H, Sanders M, Decuyper S, de Doncker S, Maes I, Rijal S, Sundar S, Dujardin JC, Berriman M, Schonian G. 2012. Genome-wide SNP and microsatellite variation illuminate population-level epidemiology in the *Leishmania donovani* species complex. *Infect Genet Evol* 12:149–159. <https://doi.org/10.1016/j.meegid.2011.11.005>.
- Moreno M, Marinotti O, Krzywinski J, Tadei WP, James AA, Achee NL, Conn JE. 2010. Complete mtDNA genomes of *Anopheles darlingi* and an approach to anopheline divergence time. *Malar J* 9:127. <https://doi.org/10.1186/1475-2875-9-127>.
- Oyola SO, Manske M, Campino S, Claessens A, Hamilton WL, Kekre M, Drury E, Mead D, Gu Y, Miles A, MacInnis B, Newbold C, Berriman M, Kwiatkowski DP. 2014. Optimized whole-genome amplification strategy for extremely AT-biased template. *DNA Res* 21:661–671. <https://doi.org/10.1093/dnares/dsu028>.
- Manske M, Miotto O, Campino S, Auburn S, Almagro-Garcia J, Maslen G, O'Brien J, Djimde A, Doumbo O, Zongo I, Ouedraogo JB, Michon P, Mueller I, Siba P, Nzila A, Borrmann S, Kiara SM, Marsh K, Jiang H, Su XZ, Amaratunga C, Fairhurst R, Sochat D, Nosten F, Imwong M, White NJ, Sanders M, Anastasi E, Alcock D, Drury E, Oyola S, Quail MA, Turner DJ, Ruano-Rubio V, Jyothi D, Amenga-Etego L, Hubbard C, Jeffreys A, Rowlands K, Sutherland C, Roper C, Mangano V, Modiano D, Tan JC, Ferdig MT, Amambua-Ngwa A, Conway DJ, Takala-Harrison S, Plowe CV, Rayner JC, et al. 2012. Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. *Nature* 487:375–379. <https://doi.org/10.1038/nature11174>.
- Miotto O, Almagro-Garcia J, Manske M, MacInnis B, Campino S, Rockett KA, Amaratunga C, Lim P, Suon S, Sreng S, Anderson JM, Duong S, Nguon C, Chuor CM, Saunders D, Se Y, Lon C, Fukuda MM, Amenga-Etego L, Hodgson AV, Asoala V, Imwong M, Takala-Harrison S, Nosten F, Su XZ, Ringwald P, Ariey F, Dolecek C, Hien TT, Boni MF, Thai CQ, Amambua-Ngwa A, Conway DJ, Djimde AA, Doumbo OK, Zongo I, Ouedraogo JB, Alcock D, Drury E, Auburn S, Koch O, Sanders M, Hubbard C, Maslen G, Ruano-Rubio V, Jyothi D, Miles A, O'Brien J, Gamble C, Oyola SO, et al. 2013. Multiple populations of artemisinin-resistant *Plasmodium falciparum* in Cambodia. *Nat Genet* 45:648–655. <https://doi.org/10.1038/ng.2624>.
- UNESCO. 1979. The African slave trade from the fifteenth to the nineteenth century: reports and papers of the meeting of experts. UNESCO, Paris, France.
- Nygaard S, Braunstein A, Malsen G, Van Dongen S, Gardner PP, Krogh A, Otto TD, Pain A, Berriman M, McAuliffe J, Dermitzakis ET, Jeffares DC. 2010. Long- and short-term selective forces on malaria parasite genomes. *PLoS Genet* 6:e1001099. <https://doi.org/10.1371/journal.pgen.1001099>.
- Chang HH, Worby CJ, Yeka A, Nankabirwa J, Kanya MR, Staedke SG, Dorsey G, Murphy M, Neafsey DE, Jeffreys AE, Hubbard C, Rockett KA, Amato R, Kwiatkowski DP, Buckee CO, Greenhouse B. 2017. THE REAL MCOIL: a method for the concurrent estimation of the complexity of infection and SNP allele frequency for malaria parasites. *PLoS Comput Biol* 13:e1005348. <https://doi.org/10.1371/journal.pcbi.1005348>.
- Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet* 11:94. <https://doi.org/10.1186/1471-2156-11-94>.
- Joy DA, Feng X, Mu J, Furuya T, Chotivanich K, Krettli AU, Ho M, Wang A, White NJ, Suh E, Beerli P, Su XZ. 2003. Early origin and recent expansion of *Plasmodium falciparum*. *Science* 300:318–321. <https://doi.org/10.1126/science.1081449>.
- Yalcindag E, Elguero E, Arnathau C, Durand P, Akiana J, Anderson TJ, Aubouy A, Balloux F, Besnard P, Bogreau H, Carnevale P, D'Alessandro U, Fontenille D, Gamboa D, Jombart T, Le Mire J, Leroy E, Maestre A, Mayxay M, Menard D, Musset L, Newton PN, Nkoghe D, Noya O, Ollomo B, Rogier C, Veron V, Wide A, Zakeri S, Carme B, Legrand E, Chevillon C, Ayala FJ, Renaud F, Prugnolle F. 2012. Multiple independent introductions of *Plasmodium falciparum* in South America. *Proc Natl Acad Sci U S A* 109:511–516. <https://doi.org/10.1073/pnas.1119058109>.
- Fischer MC, Foll M, Excoffier L, Heckel G. 2011. Enhanced AFLP genome scans detect local adaptation in high-altitude populations of a small rodent (*Microtus arvalis*). *Mol Ecol* 20:1450–1462. <https://doi.org/10.1111/j.1365-294X.2011.05015.x>.
- Foll M, Gaggiotti O. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* 180:977–993. <https://doi.org/10.1534/genetics.108.092221>.
- Foll M, Fischer MC, Heckel G, Excoffier L. 2010. Estimating population

- structure from AFLP amplification intensity. *Mol Ecol* 19:4638–4647. <https://doi.org/10.1111/j.1365-294X.2010.04820.x>.
30. Chan S, Ch'ng JH, Wahlgren M, Thutkawkorapin J. 2017. Frequent GU wobble pairings reduce translation efficiency in *Plasmodium falciparum*. *Sci Rep* 7:723. <https://doi.org/10.1038/s41598-017-00801-9>.
 31. Canepa GE, Molina-Cruz A, Barillas-Mury C. 2016. Molecular analysis of Pfs47-mediated *Plasmodium* evasion of mosquito immunity. *PLoS One* 11:e0168279. <https://doi.org/10.1371/journal.pone.0168279>.
 32. Molina-Cruz A, DeJong RJ, Ortega C, Haile A, Abban E, Rodrigues J, Jaramillo-Gutierrez G, Barillas-Mury C. 2012. Some strains of *Plasmodium falciparum*, a human malaria parasite, evade the complement-like system of *Anopheles gambiae* mosquitoes. *Proc Natl Acad Sci U S A* 109:E1957–E1962. <https://doi.org/10.1073/pnas.1121183109>.
 33. Molina-Cruz A, Garver LS, Alabaster A, Bangiolo L, Haile A, Winikor J, Ortega C, van Schaijk BC, Sauerwein RW, Taylor-Salmon E, Barillas-Mury C. 2013. The human malaria parasite Pfs47 gene mediates evasion of the mosquito immune system. *Science* 340:984–987. <https://doi.org/10.1126/science.1235264>.
 34. Molina-Cruz A, Canepa GE, Kamath N, Pavlovic NV, Mu J, Ramphul UN, Ramirez JL, Barillas-Mury C. 2015. *Plasmodium* evasion of mosquito immunity and global malaria transmission: the lock-and-key theory. *Proc Natl Acad Sci U S A* 112:15178–15183. <https://doi.org/10.1073/pnas.1520426112>.
 35. Hoffman S. 2013. Mosquito or cultured sporozoites and blood stage transcriptome (NF54). PlasmoDB. https://plasmodb.org/plasmo/app/record/dataset/DS_ef0bba5e5b.
 36. Lopez-Barragan MJ, Lemieux J, Quinones M, Williamson KC, Molina-Cruz A, Cui K, Barillas-Mury C, Zhao K, Su XZ. 2011. Directional gene expression and antisense transcripts in sexual and asexual stages of *Plasmodium falciparum*. *BMC Genomics* 12:587. <https://doi.org/10.1186/1471-2164-12-587>.
 37. Massingham T, Goldman N. 2007. Statistics of the log-det estimator. *Mol Biol Evol* 24:2277–2285. <https://doi.org/10.1093/molbev/msm160>.
 38. Anderson TJ, Haubold B, Williams JT, Estrada-Franco JG, Richardson L, Mollinedo R, Bockarie M, Mokili J, Mharakurwa S, French N, Whitworth J, Velez ID, Brockman AH, Nosten F, Ferreira MU, Day KP. 2000. Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. *Mol Biol Evol* 17:1467–1482. <https://doi.org/10.1093/oxfordjournals.molbev.a026247>.
 39. Molina-Cruz A, Barillas-Mury C. 2014. The remarkable journey of adaptation of the *Plasmodium falciparum* malaria parasite to New World anopheline mosquitoes. *Mem Inst Oswaldo Cruz* 109:662–667. <https://doi.org/10.1590/0074-0276130553>.
 40. Molina-Cruz A, Canepa G, Alves E, Silva T, Williams A, Nagyal S, Yenokidok-Douti L, Nagata B, Calvo E, Andersen J, Boulanger M, Barillas-Mury C. 2020. *Plasmodium falciparum* evades immunity of anopheline mosquitoes by interacting with a Pfs47 midgut receptor. *Proc Natl Acad Sci U S A* 117:2597–2605. <https://doi.org/10.1073/pnas.1917042117>.
 41. Aurrecochea C, Brestelli J, Brunk BP, Dommer J, Fischer S, Gajria B, Gao X, Gingle A, Grant G, Harb OS, Heiges M, Innamorato F, Iodice J, Kissinger JC, Kraemer E, Li W, Miller JA, Nayak V, Pennington C, Pinney DF, Roos DS, Ross C, Stoeckert CJ, Jr, Treatman C, Wang H. 2009. PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Res* 37:D539–D543. <https://doi.org/10.1093/nar/gkn814>.
 42. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. 2015. RDP4: detection and analysis of recombination patterns in virus genomes. *Virus Evol* 1:vev003. <https://doi.org/10.1093/ve/vev003>.
 43. Mavian C, Rife BD, Dollar JJ, Cella E, Ciccozzi M, Prospero MCF, Lednický J, Morris JG, Capua I, Salemi M. 2017. Emergence of recombinant Mayaro virus strains from the Amazon basin. *Sci Rep* 7:8718. <https://doi.org/10.1038/s41598-017-07152-5>.
 44. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. 2012. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet* 8:e1002764. <https://doi.org/10.1371/journal.pgen.1002764>.
 45. Miotto O, Amato R, Ashley EA, MacInnis B, Almagro-Garcia J, Amaratunga C, Lim P, Mead D, Oyola SO, Dhorda M, Imwong M, Woodrow C, Manske M, Stalker J, Drury E, Campino S, Amenga-Etego L, Thanh TN, Tran HT, Ringwald P, Bethell D, Nosten F, Phyto AP, Pukrittayakamee S, Chotivanich K, Chuo C, Nguon C, Suon S, Sreng S, Newton PN, Mayxay M, Khanthavong M, Hongvanthong B, Htut Y, Han KT, Kyaw MP, Faiz MA, Fanello CI, Onyamboko M, Mokuolu OA, Jacob CG, Takala-Harrison S, Plowe CV, Day NP, Dondorp AM, Spencer CC, McVean G, Fairhurst RM, White NJ, Kwiatkowski DP. 2015. Genetic architecture of artemisinin-resistant *Plasmodium falciparum*. *Nat Genet* 47:226–234. <https://doi.org/10.1038/ng.3189>.
 46. Maiga-Ascofare O, Le Bras J, Mazmouz R, Renard E, Falcao S, Broussier E, Bustos D, Randrianarivelosia M, Omar SA, Aubouy A, Lepere JF, Jean-Francois V, Djimde AA, Clain J. 2010. Adaptive differentiation of *Plasmodium falciparum* populations inferred from single-nucleotide polymorphisms (SNPs) conferring drug resistance and from neutral SNPs. *J Infect Dis* 202:1095–1103. <https://doi.org/10.1086/656142>.
 47. MalariaGEN *Plasmodium falciparum* Community Project, Amato R, Miotto O, Woodrow CJ, Almagro-Garcia J, Sinha I, Campino S, Mead D, Drury E, Kekre M, Sanders M, Amambua-Ngwa A, Amaratunga C, Amenga-Etego L, Andriananjaka V, Apinijoh T, Ashley E, Auburn S, Awandare GA, Baraka V, Barry A, Boni MF, Borrmann S, Bousema T, Branch O, Bull PC, Chotivanich K, Conway DJ, Craig A, Day NP, Djimde A, Dolecek C, Dondorp AM, Drakeley C, Duffy P, Echeverry DF, Egwang TG, Fairhurst RM, Faiz MA, Fanello CI, Hien TT, Hodgson A, Imwong M, Ishengoma D, Lim P, Lon C, Marfurt J, Marsh K, Mayxay M, Michon P, Moberg V, et al. 2016. Genomic epidemiology of artemisinin resistant malaria. *eLife* 2016:e08714. <https://doi.org/10.7554/eLife.08714>.
 48. Saul A, Battistutta D. 1988. Codon usage in *Plasmodium falciparum*. *Mol Biochem Parasitol* 27:35–42. [https://doi.org/10.1016/0166-6851\(88\)90022-9](https://doi.org/10.1016/0166-6851(88)90022-9).
 49. Ghosh A, Devenport M, Jethwaney D, Kalume D, Pandey A, Anderson V, Sultan A, Kumar N, Jacobs-Lorena M. 2009. Malaria parasite invasion of the mosquito salivary gland requires interaction between the *Plasmodium* TRAP and the *Anopheles* saglin proteins. *PLoS Pathog* 5:e1000265. <https://doi.org/10.1371/journal.ppat.1000265>.
 50. Ramakrishnan C, Dessens JT, Armson R, Pinto SB, Talman AM, Blagborough AM, Sinden RE. 2011. Vital functions of the malarial ookinete protein, CTRP, reside in the A domains. *Int J Parasitol* 41:1029–1039. <https://doi.org/10.1016/j.ijpara.2011.05.007>.
 51. Canepa GE, Molina-Cruz A, Yenokidok-Douti L, Calvo E, Williams AE, Burkhardt M, Peng F, Narum D, Boulanger MJ, Valenzuela JG, Barillas-Mury C. 2018. Antibody targeting of a specific region of Pfs47 blocks *Plasmodium falciparum* malaria transmission. *NPJ Vaccines* 3:26. <https://doi.org/10.1038/s41541-018-0065-5>.
 52. Atkinson SC, Armistead JS, Mathias DK, Sandeu MM, Tao D, Borhani-Dizaji N, Tarimo BB, Morlais I, Dinglasan RR, Borg NA. 2015. The *Anopheles*-midgut APN1 structure reveals a new malaria transmission-blocking vaccine epitope. *Nat Struct Mol Biol* 22:532–539. <https://doi.org/10.1038/nsmb.3048>.
 53. Dinglasan RR, Jacobs-Lorena M. 2008. Flipping the paradigm on malaria transmission-blocking vaccines. *Trends Parasitol* 24:364–370. <https://doi.org/10.1016/j.pt.2008.05.002>.
 54. Weathersbee AA. 1944. Observations on the relative attractiveness of man and horse for *Anopheles albimanus* Weideman. *Am J Trop Med Hyg* 1:24:25–28. <https://doi.org/10.4269/ajtmh.1944.s1-24.25>.
 55. Oyola SO, Ariani CV, Hamilton WL, Kekre M, Amenga-Etego LN, Ghansah A, Rutledge GG, Redmond S, Manske M, Jyothi D, Jacob CG, Otto TD, Rockett K, Newbold CI, Berriman M, Kwiatkowski DP. 2016. Whole genome sequencing of *Plasmodium falciparum* from dried blood spots using selective whole genome amplification. *Malar J* 15:597. <https://doi.org/10.1186/s12936-016-1641-7>.
 56. Sundararaman SA, Plenderleith LJ, Liu W, Loy DE, Learn GH, Li Y, Shaw KS, Ayoub A, Peeters M, Speede S, Shaw GM, Bushman FD, Brisson D, Rayner JC, Sharp PM, Hahn BH. 2016. Genomes of cryptic chimpanzee *Plasmodium* species reveal key evolutionary events leading to human malaria. *Nat Commun* 7:11078. <https://doi.org/10.1038/ncomms11078>.
 57. Trevino S, Nkhoma S, Nair S, Daniel B, Moncada K, Khoswe S, Banda R, Nosten F, Cheeseman I. 2017. High-resolution single-cell sequencing of malaria parasites. *Genome Biol Evol* 9:3373–3383. <https://doi.org/10.1093/gbe/evx256>.
 58. Moll K, Kaneko A, Scherf A, Wahlgren M. 2013. *Methods in malaria research*, 6th ed. EVIMaLaR, Glasgow, United Kingdom.
 59. Sriprawat K, Kaewpongsri S, Suwanarusuk R, Leimanis ML, Lek-Uthai U, Phyto AP, Snounou G, Russell B, Renia L, Nosten F. 2009. Effective and cheap removal of leukocytes and platelets from *Plasmodium vivax* infected blood. *Malar J* 8:115. <https://doi.org/10.1186/1475-2875-8-115>.
 60. Andrews S. 2010. A quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
 61. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
 62. Krueger F. 2012. A wrapper tool around Cutadapt and FastQC to con-

- sistently apply quality and adapter trimming to FastQ files, with some extra functionality for MspI-digested RRBS-type (Reduced Representation Bisulfite-Seq) libraries. https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.
63. Li R, Li Y, Fang X, Yang H, Wang J, Kristiansen K, Wang J. 2009. SNP detection for massively parallel whole-genome resequencing. *Genome Res* 19:1124–1132. <https://doi.org/10.1101/gr.088013.108>.
 64. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernysky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43:491–498. <https://doi.org/10.1038/ng.806>.
 65. Cingolani P, Platts A, Wang Le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 6:80–92. <https://doi.org/10.4161/fly.19695>.
 66. RCoreTeam. 2016. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
 67. RStudioTeam. 2015. RStudio: integrated development for R. RStudio, Inc., Boston, MA. <http://www.rstudio.com/>.
 68. Haubold B, Pierstorff N, Moller F, Wiehe T. 2005. Genome comparison without alignment using shortest unique substrings. *BMC Bioinformatics* 6:123. <https://doi.org/10.1186/1471-2105-6-123>.
 69. Xia X, Xie Z, Salemi M, Chen L, Wang Y. 2003. An index of substitution saturation and its application. *Mol Phylogenet Evol* 26:1–7. [https://doi.org/10.1016/s1055-7903\(02\)00326-3](https://doi.org/10.1016/s1055-7903(02)00326-3).
 70. Xia X. 2017. DAMBE6: new tools for microbial genomics, phylogenetics, and molecular evolution. *J Hered* 108:431–437. <https://doi.org/10.1093/jhered/esx033>.
 71. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD. 2006. Automated phylogenetic detection of recombination using a genetic algorithm. *Mol Biol Evol* 23:1891–1901. <https://doi.org/10.1093/molbev/msl051>.
 72. Pond SL, Frost SD, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21:676–679. <https://doi.org/10.1093/bioinformatics/bti079>.
 73. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio J, Guirao-Rico S, Librado P, Ramos-Onsins S, Sánchez-Gracia A. 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol* 34:3299–3302. <https://doi.org/10.1093/molbev/msx248>.
 74. Lemoine JF, Boncy J, Filler S, Kachur SP, Fitter D, Chang MA. 2017. Haiti's commitment to malaria elimination: progress in the face of challenges, 2010–2016. *Am J Trop Med Hyg* 97(Suppl 4):43–48. <https://doi.org/10.4269/ajtmh.16-0902>.