

# Population Genomics Of Plasmodium Falciparum And Malaria Control Implications In Abidjan (Cote D'ivoire)

**Désiré N'Da Ehouni**

Jomo Kenyatta University of Agriculture and Technology

**Abibatou Konaté**

Universite Felix Houphouet-Boigny

**Steven G Nyanjom**

Jomo Kenyatta University of Agriculture and Technology

**Amed Ouattara**

University of Maryland School of Medicine

**Hervé Eby I. Menan**

Universite Felix Houphouet-Boigny

**William Yavo** (✉ [yavowilliam@yahoo.fr](mailto:yavowilliam@yahoo.fr))

Universite Felix Houphouet-Boigny <https://orcid.org/0000-0002-2711-5073>

---

## Research

**Keywords:** Malaria, Population Genomics, Whole Genome Sequencing, Côte d'Ivoire, Plasmodium falciparum

**Posted Date:** September 8th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-72315/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# POPULATION GENOMICS OF PLASMODIUM FALCIPARUM AND MALARIA CONTROL IMPLICATIONS IN ABIDJAN (COTE D'IVOIRE)

Désiré N Ehouni<sup>1,2</sup>, Abibatou Konaté<sup>2,3</sup>, Steven G Nyanjom<sup>1</sup>, Amed Ouattara<sup>4</sup>, Hervé Menan<sup>3</sup> and William Yavo<sup>2,3\*</sup>

<sup>1</sup>Department of Biochemistry, College of Health Sciences, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

<sup>2</sup>Malaria Research and Control Center, National Institute of Public Health, Abidjan, Côte d'Ivoire

<sup>3</sup>Department of Parasitology and Mycology, Pharmaceutical and Biological Sciences Training and Research Unit, University Félix Houphouët-Boigny, Abidjan, Côte d'Ivoire

<sup>4</sup>Malaria Research Program, Center for Vaccine Development and Global Health, University of Maryland School of Medicine, Baltimore, USA

\*Correspondence: [yavowilliam@yahoo.fr](mailto:yavowilliam@yahoo.fr)

## Abstract

**Background:** The sudden onset of *P. falciparum* resistance to antimalarial drugs requires careful surveillance of African parasite population. Genomics tools are implemented to detect evolutionary changes that could impact on malaria control and elimination strategies. Here, we evaluated the genome-wide pattern of selection and sequence variation of *P. falciparum* populations from Côte d'Ivoire.

**Methods:** The study was carried out in three localities of Abidjan from 2013 to 2014. We collected 70 blood samples following a written consent from patients above two years of age. After extracting *P. falciparum* DNA from isolates, we performed Whole Genome Sequencing and used population genomics approaches to investigate genetic diversity, complexity of infection and identify loci under positive directional selection.

**Results:** We observed an excess of rare variants in the population showing a clear mutation process in the isolates. Moderate *Fst* estimates (0.3) was detected for *surfin* genes expressed at the surface of infected erythrocytes and released merozoites. Seven iHS regions that had at least two SNPs with a

score > 3.2 were identified. These regions code for genes that have been under strong directional selection. Two of these genes were the chloroquine resistance transporter (*crt*) on chromosome 7 and the dihydropteroate reductase (*dhps*) on chromosome 8 despite their official proscription for malaria treatment since 2007. There was also a recent selective sweep in the erythrocyte membrane protein (*Pfemp1*).

**Conclusion:** Population genomics revealed selective drug pressure and balancing selection on protective immunity target genes, demonstrating his effectiveness to follow up adaptive evolution of parasite populations and adopt appropriate strategies to control malaria in Côte d'Ivoire.

**Keywords:** Malaria, Population Genomics, Whole Genome Sequencing, Côte d'Ivoire, *Plasmodium falciparum*

## **Background**

Despite considerable efforts from the World Health Organization (WHO) and other funding agencies including the Gates Foundation and the United States President Malaria Initiative (PMI), *Plasmodium falciparum* (*P. falciparum*) malaria remains a major public health issue in the world, and mostly in sub-Saharan Africa [1]. In Côte d'Ivoire, malaria is endemic and the entire population is at risk of contracting the disease especially pregnant women and children under the age of five [2]. Malaria incidence has plateaued during the past few years. In addition, the emergence of malaria parasite resistance to artemisinin derivatives highlighted the need to monitor parasite population [3,4]. Investigating the genetic profile of *P. falciparum* may detect evolutionary changes that can have an impact on malaria control and elimination efforts.

During the past decade, genomics and genetics studies have been conducted to identify *P. falciparum* markers associated with disease severity, resistance to drugs and escape to the human immune system. These studies provide valuable informations for malaria control [5–7]. Hence, a population genomics study of *P. falciparum* in a single endemic population of Gambia identified new genes under balancing selection, such as the apical membrane antigen 1 gene (*ama1*) which encodes a prime vaccine candidate

[8]. In addition, a study in Nigeria detected regions that were actively maintained in the gene presumably a signature of adaptation to drug pressure and host immunity. [9]. No evidence or very little sub-population of the *P. falciparum* population has been found across West Africa [5]. These findings suggest gene flow between regions, despite differences in transmission seasonality and local vector species abundances [10–12] . However, studies have also shown that malaria parasite population structure is increasing in low transmission areas [13,14] . The emergence of multigenic drug resistance has been favored by high rates of inbreeding that can be estimated by within-host diversity [14–16] . Thus, malaria transmission intensity and parasite genetic diversity vary greatly across West Africa regions due to variation in rainfall seasonality indicating more highly mixed genotype infections in Guinea than in The Gambia [17]. In Côte d’Ivoire, transmission of malaria in Côte d’Ivoire occurs all year round, with seasonal peaks during the rainy seasons. It is imperative to assess the genetic diversity of individual infections relative to the genetic diversity of the parasite population as a whole. Recent genome-wide scan of *P. falciparum* revealed loci under selection in known drug targets as *crt* (chloroquine resistance transporter), *dhps* (dihydropteroate synthase), and *dhfr* (dihydrofolate reductase) in two localities of Ghana [12]. Other studies performed in Senegal, The Gambia and Guinea [7,10,18,19] reported evidence in signatures of selection surrounding genes involved with chloroquine (*mdr1* and *crt*) and antifolate (*dhfr* and *dhps*) resistance. This suggests that Côte d’Ivoire could report similar patterns since it reflects the same historical drug use with previous described countries. Although population genomics investigated across West Africa have shown strong positive selection around known drug resistance genes, inferences of local mechanisms require looking more distinctively at individual populations such as Côte d’Ivoire.

This study assess the genetic profile of circulating *P. falciparum* strains and aims to understand the genome-wide patterns of selection in Côte d’Ivoire. Using samples from different localities (Abobo, Koumassi and Yopougon), we assessed whether there is a significant population substructure in Côte

d'Ivoire and measure the complexity of infection between three localities of the country. Finally, we identified chromosomes and genes that could be under recent positive selection and balancing selection.

## **Methods**

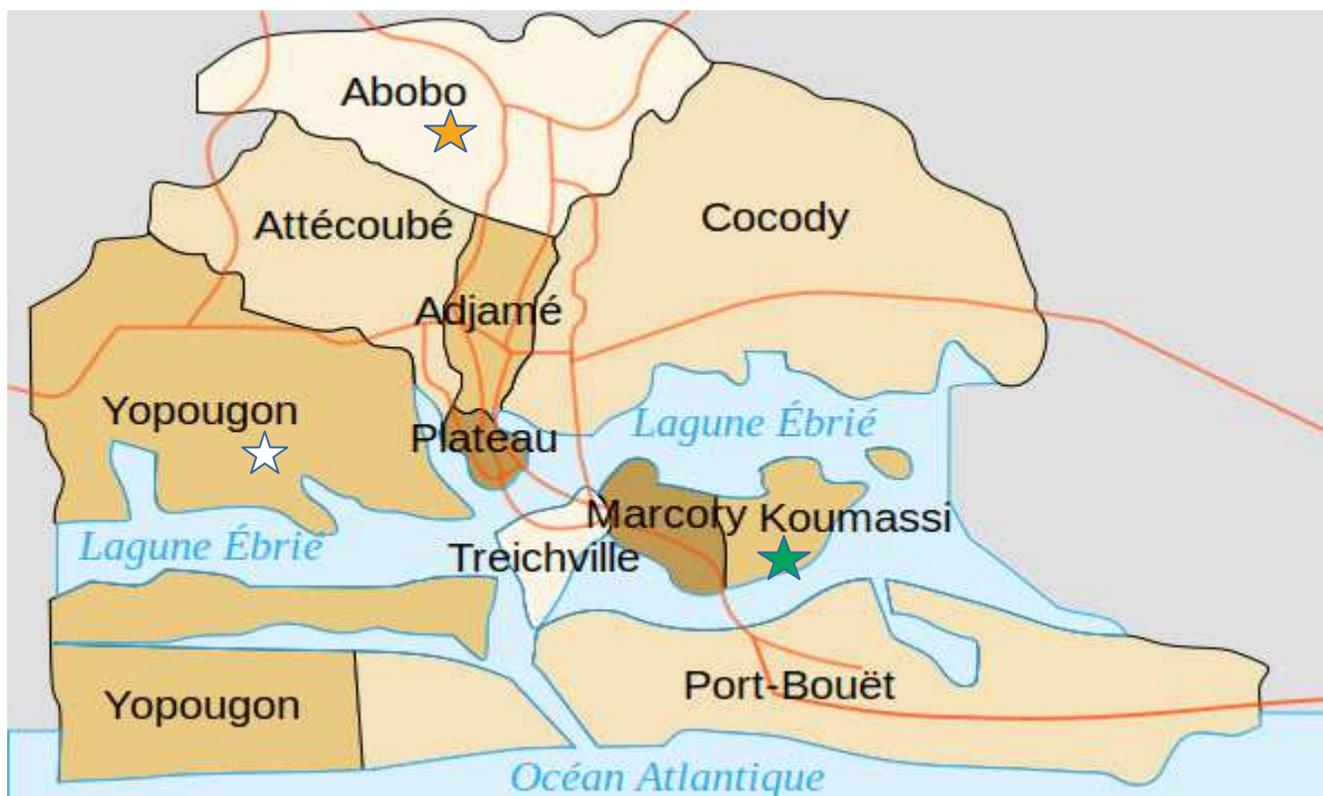
### **Ethics Statement**

An approval to collect and analyze clinical isolates was granted by the National Ethics and Research Committee (CNER-CI) according to protocols and standards operating procedures of Good Clinical Practices of the ICH harmonized Triplicate Guide Lines for Good Clinical Practice made in 1996 and the Helsinki Declaration on human being research. Samples were collected following a written informed consent from patients or their legal guardians for study participants under 18 years old.

### **Study sites and sampling of *Plasmodium falciparum* from clinical isolates**

Abidjan is an urban malaria endemic city of the southern Atlantic coast of Côte d'Ivoire, on the Gulf of Guinea, in West Africa. Malaria transmission in Abidjan is intense with recrudescence during the rainy season.

Blood samples were collected from September 2013 to February 2014 during the rainy and dry seasons, in local health facilities located across three localities of Abidjan (Fig. 1) including Abobo (5°25'N, 4°1'W), Koumassi (5°18'N, 3°57'W) and Yopougon (5°19'N, 4°4'W). These samples were collected in the context of the Plasmodium Diversity Network in Africa (PDNA) project. This study involved 70 individuals including 31 in Abobo, 19 in Koumassi and 20 in Yopougon. Patients were eligible for recruitment if they were children more than 2 years old, presenting an axillary temperature of 37.5 °C or history of fever during the last 24 hours, with *P. falciparum* mono-infection. After administrating the informed consent, 5 mL of venous blood sample was collected in ethylenediaminetetraacetic acid vacutainers (EDTA), depleted of leukocytes using CF11 cellulose columns, and frozen at - 20 °C. *P. falciparum* DNA was later extracted using the QI Amp blood mini kit (Qiagen, UK) followed by the whole genome sequencing of *P. falciparum*.



**Fig. 1 Map of Abidjan district. The three study localities (Abobo, Koumassi and Yopougon) are identified by stars**

### **Whole Genome Sequencing of *Plasmodium falciparum* from clinical isolates**

DNA from the 70 clinical samples collected in Abidjan were prepared for library preparation and Whole Genome Sequencing at the Wellcome Trust Sanger Institute using an Illumina HiSeq platform. Standard laboratory protocols were used to determine DNA quantity and proportion of human DNA in each sample. All the 70 samples were above the “good quality” thresholds and were put forward for whole genome Illumina paired-end sequencing [14,20]. Sequencing reads were mapped to the *P. falciparum* 3D7 reference genome using bwa mem version 0.7.15 with -M parameter to mark shorter split hits as secondary [21]. Standard alignment metrics were generated for each sample using the stats utility from samtools version 1.2 [22]. We also used GATK’s CallableLoci (version 3.5) to determine the proportion of genomic positions callable in each sample [23]. Potential SNPs (Single Nucleotide Polymorphisms) and indels were discovered by running GATK’s HaplotypeCaller (version 3.6) independently across each of the 70 sample-level bam files. This resulted in genotype calls for both SNPs and short indels. SNPs

and indels were filtered separately. Each variant was assigned a quality score using GATK's Variant Quality Score Recalibration (VQSR) version 3.6. The tools VariantRecalibrator and ApplyRecalibration were used for this purpose. Regions of the genome which we previously identified as being enriched for errors were masked out. VariantRecalibrator was run using the PASS variants from the *P. falciparum* crosses 1.0 release as a training set. For SNPs we used 15.0 as a prior for the training set variants. ApplyRecalibration was then run to assign each variant a quality score named VQSLOD (Variant quality score). Higher values of VQSLOD indicate higher quality. Variants (both SNPs and indels) with a VQSLOD score  $\leq 0$  were filtered out. Variants were excluded from the analysis if they were positioned within subtelomeric regions, located within the hypervariable *Var*, *Rifin*, and *Stevor* gene families, or were positioned within repetitive sequences as identified by Tandem Repeat Finder. Data were then filtered out to extract 89578 bi-allelic SNPs in the core genome with a VQSLOD score  $> 6$ . In addition, we filtered out those SNPs to exclude isolates with missing calls at  $> 10\%$  of all positions and SNPs with calls missing in  $>10\%$  of isolates. A total of 89211 SNPs remained after filtering.

### **Determination of allele frequencies and balancing selection**

Analysis of allele frequency distributions, including within population Tajima's D test was performed using Vcftools and custom R scripts to identify genes under balancing selection [24]. For Tajima's D test, we extracted bi-allelic SNPs that were segregating within our population. Missing data were excluded by removal of individual isolates on a gene by gene basis due to the observation that most of missing data clustered within a small number of isolates. The allele frequency spectrum for each gene was assessed with at least 3 SNPs using custom R scripts. Typable SNPs in this study were classified as synonymous or non-synonymous based on amino-acids changes when compared to the 3D7 reference genome sequence. The ancestral state of SNPs has been determined by comparing each sequence to homologous sequences in *Plasmodium reichenowi* (*Pr*), a parasite with recent common ancestry.

## Identification of intra-population signatures of recent selection sweep

To detect loci under recent positive selection, we computed the standardized integrated Haplotype Score ( $|iHS|$ ) for each SNP with no missing data and a minimum minor allele frequency of 0.05 using the REHH R software package.  $iHS$  has been calculated for each SNP with no missing data and a minor allele frequency of  $>0.05$  [25]. The genetic map distance between markers inferred with LDhat 2.2 [26] was measured using a block penalty of 10 million rjMCMC iterations, and a burn-in of 100 000 iterations. Selection windows were defined by calculating the distance required for the extended haplotype homozygosity of each SNP to decay to a level of 0.05 in each direction. Overlapping EHH windows from individual high-scoring SNPs ( $|iHS| > 3.29$ , suggestive line) were combined into continuous windows, and windows supported by only a single SNP position were subsequently discarded. Bonferoni correction was applied for genome wide significance. The reference and non-reference allele was described as ancestral and derived alleles, respectively. For significance the REHH package generated a two-sided p-value as

$$-\log_{10}(2\Phi(-|iHS|)),$$

Where  $\Phi(iHS)$  represents the Gaussian cumulative distribution function.

## Complexity of Infection with $F_{ws}$

Infection complexity was determined by  $F_{ws}$  fixation index [14, 27] using the moimix package in R and computed with R scripts. For all bi-allelic coding SNPs,  $F_{ws}$  was calculated using the formula  $F_{ws} = 1 - (H_w / H_s)$  where  $H_w$  is the within-individual heterozygosity and  $H_s$  is the within-population heterozygosity. At each bi-allelic SNPs, heterozygosity was estimated using the formula  $H = 1 - (p^2 + q^2)$ , where  $p$  and  $q$  are the frequencies of the two alleles ( $p = 1 - q$ ). At each SNP,  $p$  and  $q$  were estimated for each individual as the proportions of sequencing reads that carried each allele in the individual sample. At the population level, the allele frequencies at the SNPs level were estimated as the mean of the allele frequencies in the individual composing the population sample. The minor allele frequency (MAF) was reported as the frequency of the least common allele at that SNP. Within-individual and within-population

heterozygosity were computed and assigned to ten equal sized MAF bins [0.0-0.05] ... [0.45-0.5] and for each bin the mean within sample and population heterozygosity was computed.

### **Population structure, Principal Component Analysis (PCA) and *Fst*-metric**

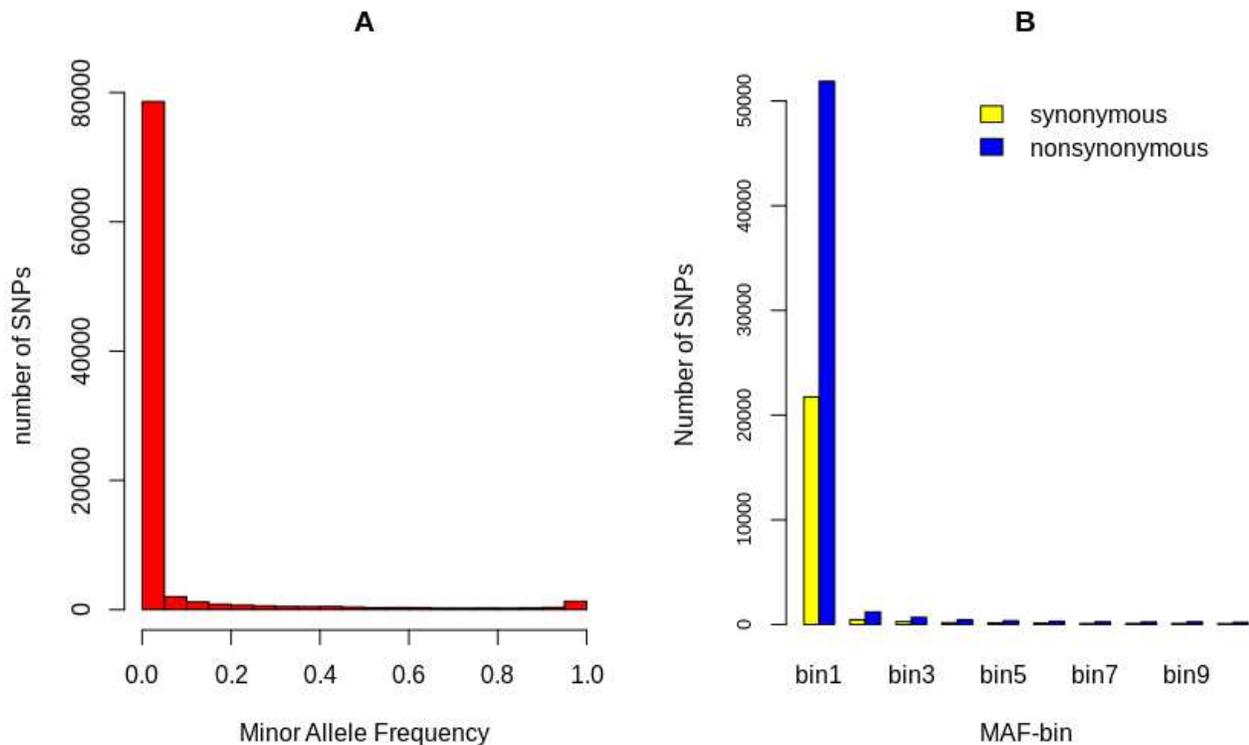
We conducted a PCA and *Fst* as implemented in PLINK1.9 and Vcftools, respectively, to assess the structure of study populations. We used Weir and Cockerham's population genetic differentiation estimator *Fst* between study sites [28]. For PCA, we applied the Linkage Disequilibrium correction to remove the correlated pairs of SNPs and identity by descent (IBD) to identify and remove any closely-related samples before computing principal components (PCs). We further calculated the top 10 eigenvectors from the population genotype. R statistical package was used to analyze our data. For *Fst* analysis missing data for some isolates were excluded on a per SNP basis.

### **Results**

Following a quality control of all the 70 clinical samples, we extracted 89578 bi-allelic SNPs in the core genome with a VQSLOD score greater than 6 SNPs. Missing calls greater than 10% were excluded (367 SNPs) from the analysis leaving 89211 SNPs for all isolates.

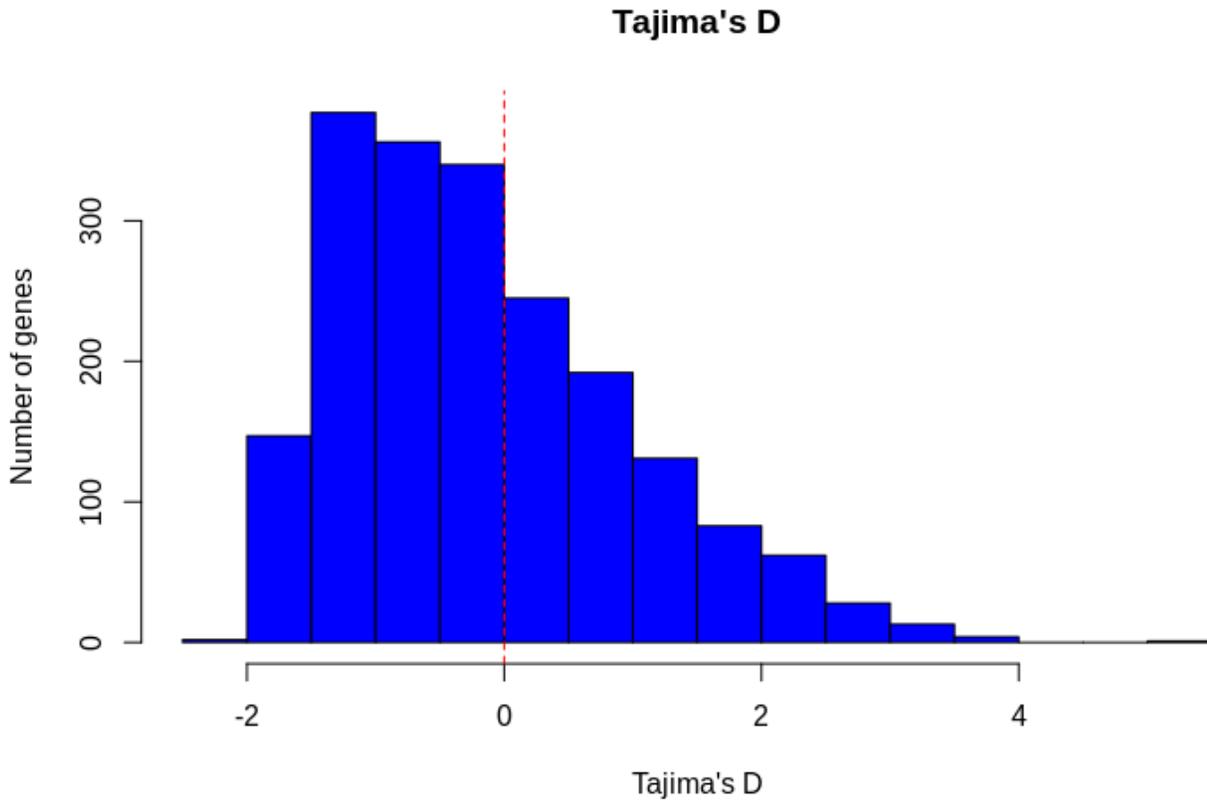
### **Allele frequency distribution and balancing selection**

There was an excess of rare variants in the population, with the majority (80%) being single isolates (**Fig. 2A**). Coding sequences had higher coverage than intergenic regions, probably due to A + T allelic richness. There was an excess of non-synonymous SNPs compared to the synonymous SNPs, showing a clear mutation process in the isolates (**Fig. 2B**).

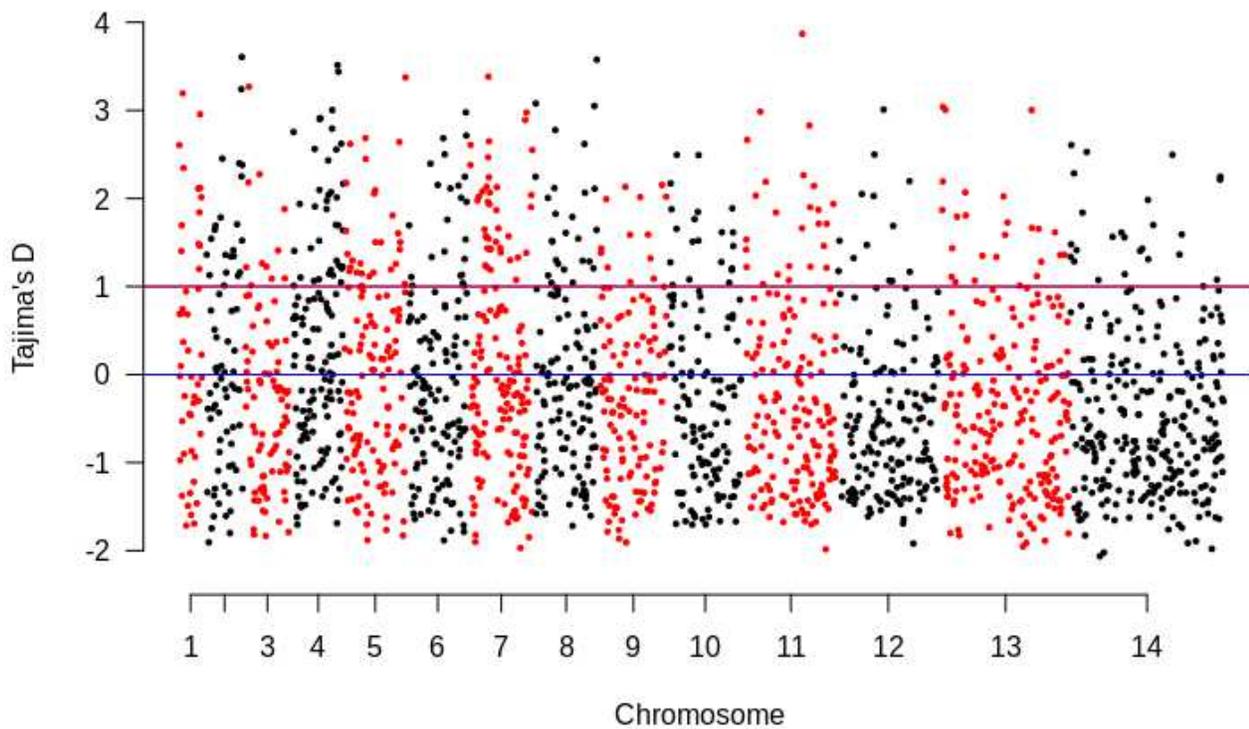


**Fig. 2 (A) Minor allele frequency distribution across 89578 bi-allelic SNPs of 70 clinical isolates from Côte d'Ivoire. (B) Distribution of non-synonymous and synonymous SNPs in the population. Within-individual and within-population heterozygosity are assigned to ten equal sized MAF bins [0.0-0.05] ... [0.45-0.5].**

Genes with at least 3 SNPs had their allele frequency distribution assessed by computing Tajima's D statistic over 16697 SNPs that are segregating in the population. For the 262 genes analyzed, Tajima's D values were mostly negative (mean = - 0.15) with only 47 genes (18%) having positive Tajima's D values, indicating an excess of low frequency compared with that expected on a mutation-drift equilibrium population (**Fig. 3 and Fig. 4**). The whole genome level, we identified 18 genes with Tajima's D values greater than 1 (**Table 1**). Significant representation of genes was included (*PfEMPI*, *SURFIN 4.2*, *GLURP*, *mSP7-like*).



**Fig. 3** Frequency distribution of Tajima's D values for genetic loci as annotated in Pf3D7 genome.



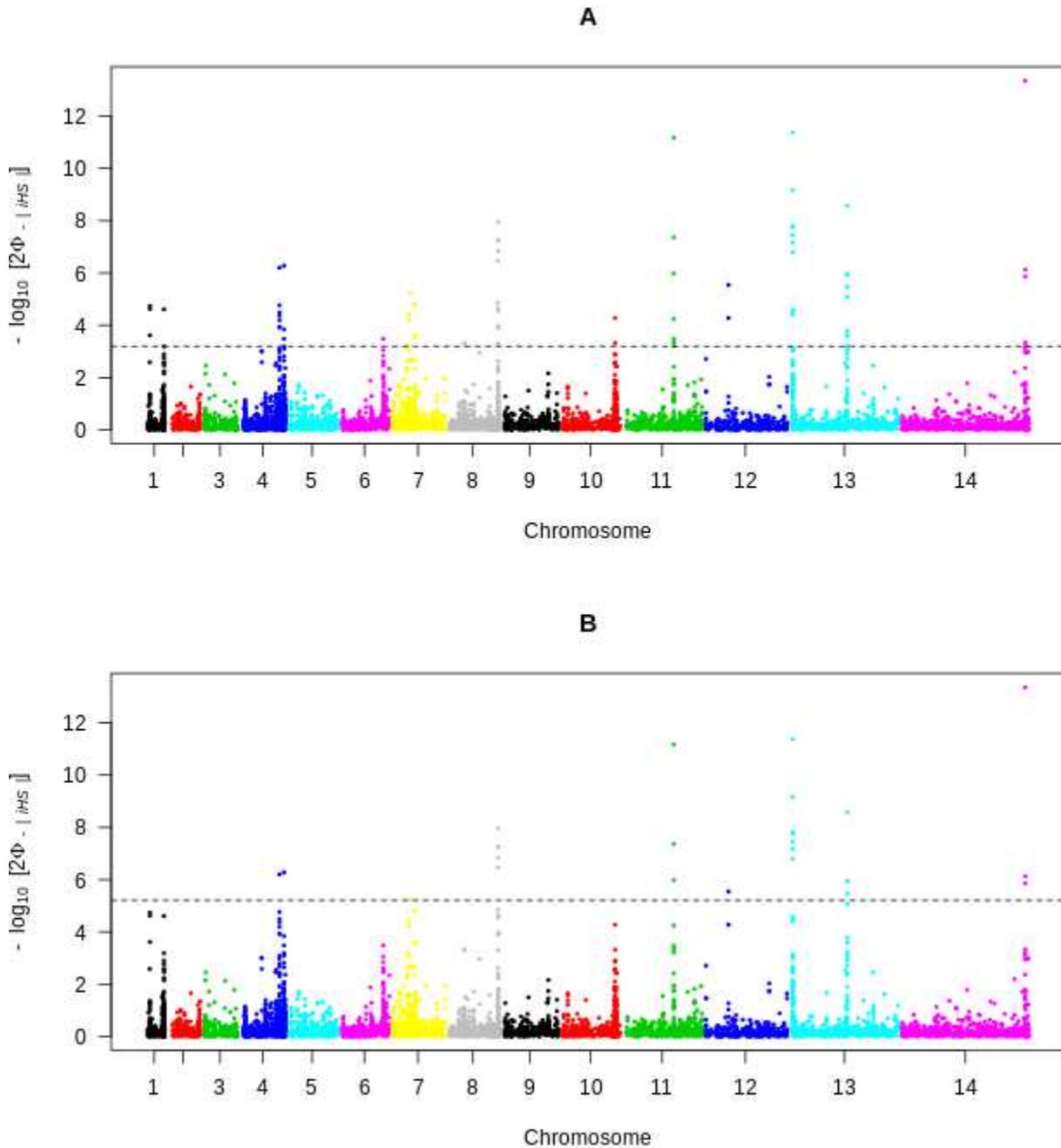
**Fig. 4** Genome-wide Manhattan plot of Tajima's D values for *P. falciparum* genes with at least three SNPs. Chromosomes are identified by the alternate red and black coloring with genes plotted as individual points based on their position within each chromosome. The blue horizontal line discriminate negative from positive values. Genes with a Tajima's D value greater than 1 in the three populations were represented above the horizontal purple line (genome wide line).

**Table 1 Genes (18) with > 3 SNPs showing highest Tajima's D values (> 1) in a genome wide analysis of 70 Ivorian *P. falciparum* isolates.**

Gene ID	Number of SNPs	Tajima's D	Protein encoded by gene
PF3D7_0102200	4	1,77	Ring-infected erythrocyte surface antigen
PF3D7_0103100	243	1,97	vacuolar protein sorting-associated protein 51, putative
PF3D7_0220000	131	3,19	liver stage antigen 3
PF3D7_0413100	97	1,39	erythrocyte membrane protein 1, PfEMP1
PF3D7_0420000	73	1,35	zinc finger protein, putative
PF3D7_0424400	35	1,98	surface-associated interspersed protein 4.2 (SURFIN 4.2)
PF3D7_0526600	83	1,18	conserved Plasmodium protein
PF3D7_0532100	30	2,11	early transcribed membrane protein 5
PF3D7_0629200	122	1,43	DnaJ protein, putative
PF3D7_0708900	217	1,97	protein SCO1, putative
PF3D7_0711400	33	1,40	histone deacetylase complex sub-unit SAP18, putative
PF3D7_0712900	196	1,12	erythrocyte membrane protein 1, PfEMP1
PF3D7_0807800	140	1,07	26S proteasome regulatory sub-unit RPN10, putative
PF3D7_0830600	120	3,51	Plasmodium exported protein (PHISTc)
PF3D7_1032300	147	1,08	Conserved Plasmodium protein
PF3D7_1035300	171	2,4	glutamate-rich protein GLURP
PF3D7_01301700	95	1,45	CX3CL1-binding protein 2
PF3D7_01334300	91	1,29	MSP7-like protein

## Evidence of signature of positive directional selection in Côte d'Ivoire

We conducted a scan of the *P. falciparum* whole genome by using REHH package in R. An assessment of the standardized integrated Haplotype score (| iHS |) identified a recent positive selection in seven loci that had four or more SNPs with a standardized | iHS | value greater than 3.2 (suggestive line) (**Fig. 5A**) and four loci that had at least one SNP with an | iHS | greater than 5 (**Fig. 5B and Table 2**). We identified windows containing genes that have been under strong and recent positive selection. Indeed, there was a strong signature surrounding one of the major chloroquine resistance gene, *Pfcr1* on chromosome 7 highlighted by his value. In addition, a strong signature pressure was also observed around genes encoding the erythrocyte membrane protein 1 (*PfEMP1*), a gene involved into the adhesion of *P. falciparum* to endothelial cells. *Pfdhps*, a drug resistance gene located on chromosome 8, was also found to be under positive selection. No detectable signature was found around the multi-drug resistance and the anti-folate drug target genes, *mdr1* and *dhfr* respectively. Another region of very high | iHS | value on chromosome 11 had SNPs within the *ama1* gene. Windows spanning regions of elevated | iHS | values were also observed on chromosome 10 and 13, covering genes encoding for the merozoite surface protein 7 like antigen, *Pftrap* (Thrombospondin related adhesive protein), *GLURP* (Glutamate Rich Protein) and *msp3* (merozoite surface protein 3).



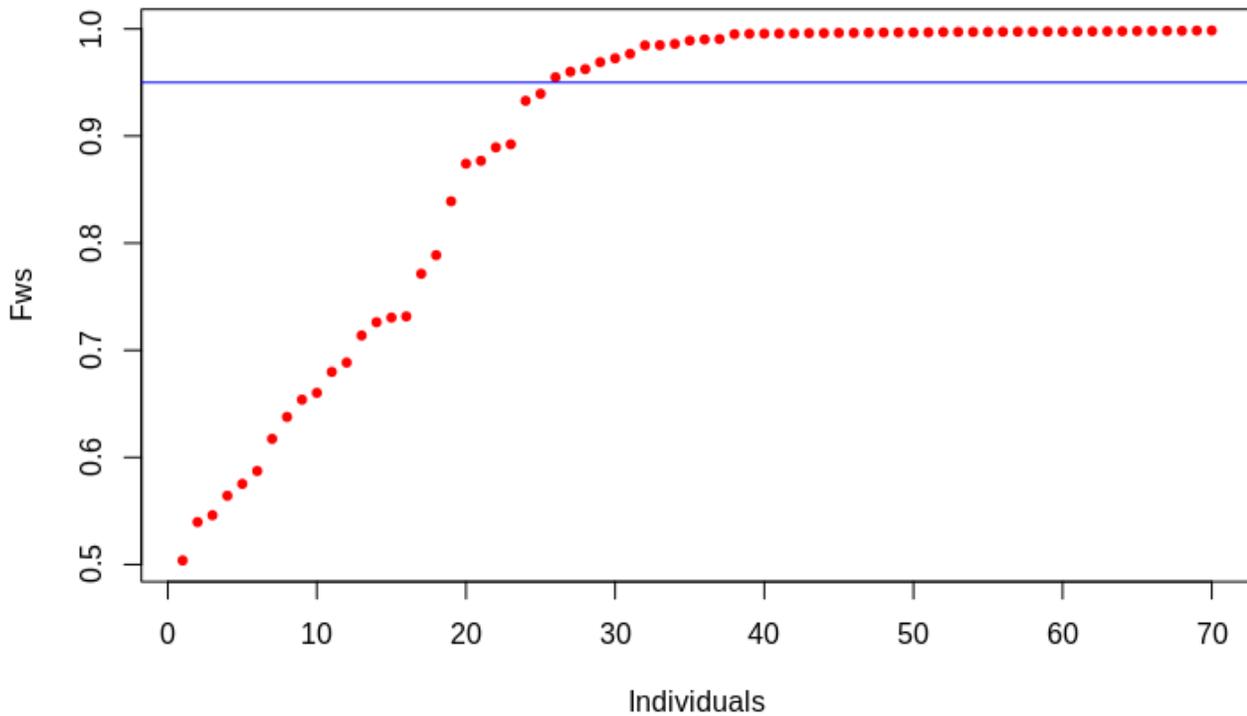
**Fig. 5** Genome wide Manhattan plot scan of standardized  $|iHS|$  for SNPs with MAF equal or greater than 5% in 70 clinical isolates from Cote d'Ivoire. Chromosomes are represented by rainbow coloring of their SNPs. High scoring SNPs are above the dashed grey line for A ( $|iHS| > 3.20$ ; suggestive line) and B ( $|iHS| > 5.20$ ; Bonferroni corrected threshold) indicating loci under recent positive selection.

**Table 2 Genomic Windows of | iHS | values greater than 3.20 (suggestive) line and greater than 5 (Bonferoni corrected threshold). Windows overlapping *Pfprt* and *Pfdhps* are highlighted in bold.**

Chromosome	Window start (kb)	Window end (kb)	Region size (kb)	Number of supporting SNPs	Genes within region
1	180	530	350	4	PF3D7_0104100- PF3D7_0113800
4	987	1100	113	10	PF3D7_0421700- PF3D7_0424300
7	<b>400</b>	<b>610</b>	<b>210</b>	<b>8</b>	<b>PF3D7_0708900- PF3D7_0713300</b>
8	<b>468</b>	<b>1312</b>	<b>844</b>	<b>14</b>	<b>PF3D7_0809200- PF3D7_0830800</b>
10	1396	1410	14	2	PF3D7_1035200- PF3D7_1035400
11	1274	1295	21	7	PF3D7_1132900- PF3D7_1133400
13	106	1466	1360	18	PF3D7_1301800- PF3D7_1335900

#### **Assessing the genome wide complexity of infection ( $F_{ws}$ )**

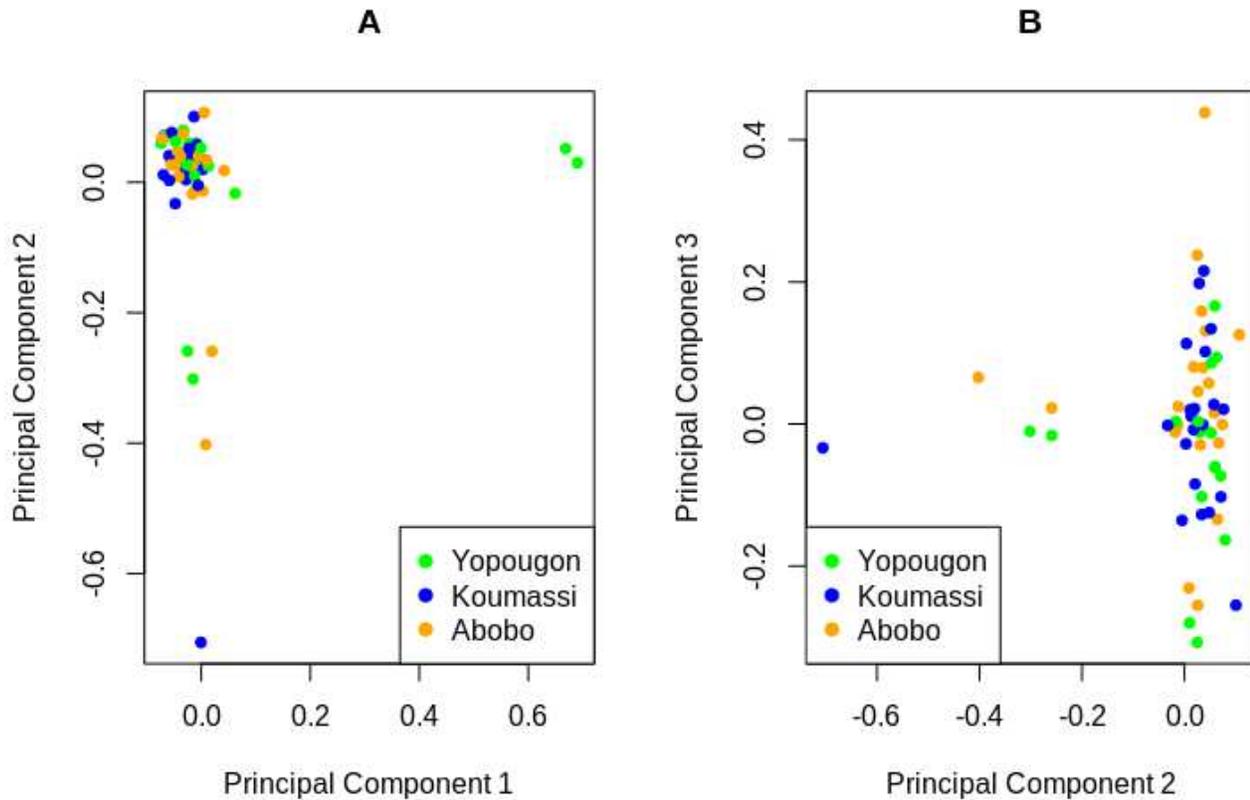
Within-infection  $F_{ws}$  fixation index describes the relationship between the diversity observed within a patient to that of the population using estimates of heterozygosity. It provides a measure of the risk of outcrossing between parasites within an individual to generate new genotypes during recombination in the mosquito host [27]. In our analysis,  $F_{ws}$  scores ranged from 0.50 to 0.99 (mean 0.89, median 0.98) with 64% samples presenting high  $F_{ws}$  estimates (ie > 0.95; **Fig. 6**). These high  $F_{ws}$  values were an evidence of infections dominated by single predominant genotypes at the time of sampling across our population.



**Fig.6 Ordered between sites (Yopougon, Abobo and Koumassi) within-infection  $Fws$  fixation index for each clinical sample with SNPs data. The horizontal blue line indicates  $Fws$  estimates  $> 0.95$  (isolate with a single predominant genotype).**

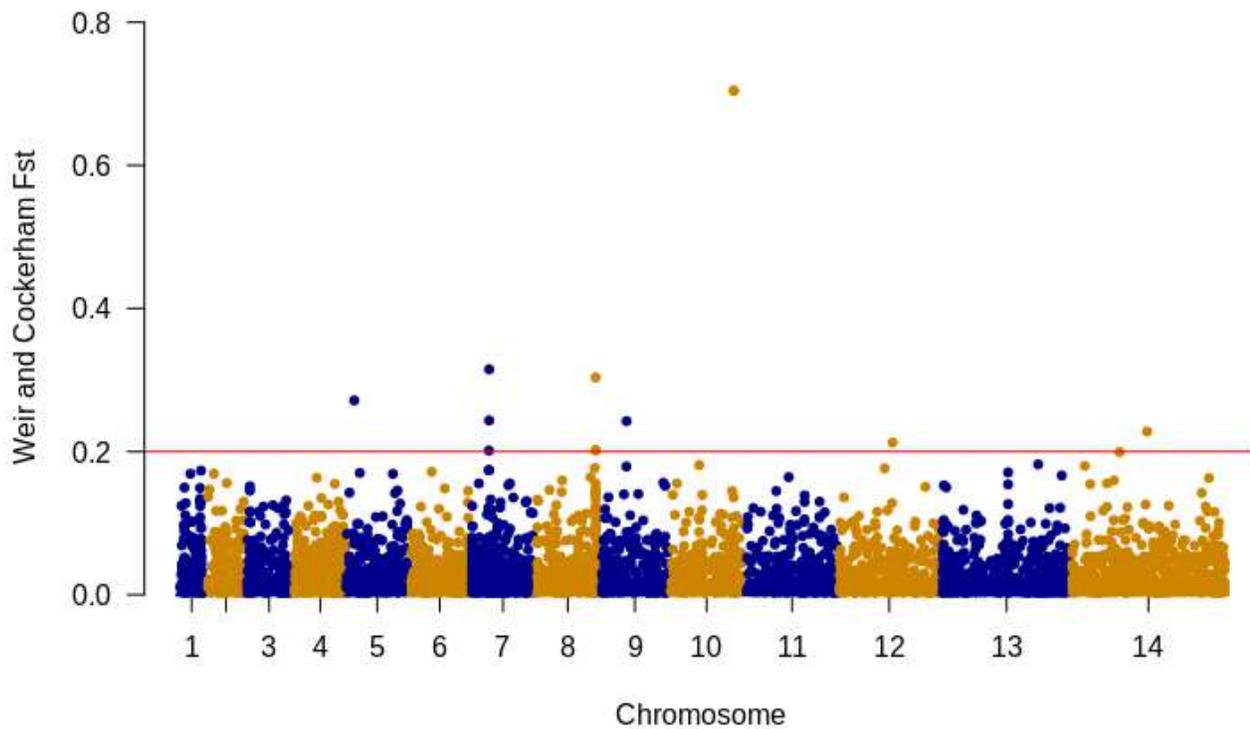
### **Population Structure and Differentiation**

We conducted a principal component analysis using the 89578 SNPs with no missing data in all 70 clinical samples. The first three principal components (10.8% of the total variation) were plotted and showed no evidence of population structure in most samples from the three populations. Only few isolates appeared as outliers that were not very divergent (**Fig. 7**).



**Fig.7 Principal components analysis of 70 clinical isolates from Abidjan (Yopougon, Koumassi, Abobo) using 89578 SNPs. A. The first and second components represent 3.2% and 4.1% of the total variation respectively. B. The second and third components cover 4.1% and 3.5% of the total variation respectively.**

We measured  $F_{st}$  to assess the genetic differentiation and evaluate the overall effect of population substructure on parasite populations. There was a minimal differentiation between population highlighted by a very low value for mean  $F_{st}$  estimate (mean 0.001) (**Fig. 8**). Only 12 SNPs had  $F_{st}$  values equal or greater to 0.2 and the highest  $F_{st}$  value was 0.7 (**Table 3**).



**Fig. 8** Manhattan plot of genome wide *Fst* for differentiation between populations for 89578 SNPs with each chromosome identified by the alternating blue/orange coloring. The red line indicates all the SNPs with *Fst* values above 0.2 (Table 3). The genome-wide *Fst* estimate is 0.001.

**Table 3 Most differentiated SNPs allele frequencies between the three populations.**

Chromosome	SNP Position	Gene	Reference Allele Frequency	<i>Fst</i>	Coding Effect	Amino Acid Change
5	192222	PF3D7_0504800	0.36	0.27	Non-synonymous	N4171D
7	430364	PF3D7_0709600	0.57	0.2	Synonymous	-
7	435368	PF3D7_0709700	0.53	0.24	Non-synonymous	T968A
7	435497	PF3D7_0709700	0.54	0.31	Synonymous	-
8	1311574	PF3D7_0830800	0.88	0.2	Non-synonymous	F313Y
8	1311584	PF3D7_0830800	0.88	0.2	Synonymous	-
8	1311619	PF3D7_0830800	0.89	0.3	Non-synonymous	E328A
9	600149	PF3D7_0914000	0.24	0.24	Non-synonymous	T819I
10	1394930	PF3D7_1035200	0.27	0.7	Non-synonymous	G31E
10	1394950	PF3D7_1035200	0.28	0.7	Non-synonymous	A38T
12	1183319	PF3D7_1228900	0.81	0.21	Non-synonymous	V369E
14	1618596	Intergenic	0.43	0.22	-	-

Three of the differentiated SNPs were located on chromosome 7. The SNP at position 435368 span the region coding for *P. falciparum* esterase gene *Pfpare* (*P. falciparum* prodrug activation and resistance esterase) that is responsible for resistance to pepstatin esters [29]. Furthermore, a SNP located on chromosome 8 encodes an amino-acid within the SURFIN<sub>8.2</sub> gene (PF3D7\_0830800), a polymorphic antigen that is expressed to the surface of *P. falciparum* infected erythrocyte (IE) and released merozoites [30]. PF3D7\_1035200 gene containing two SNPs with highest genome wide *Fst* values were located within a single region on chromosome 10, encoding a conserved protein with an unknown function.

## Discussion

*P. falciparum* genomic studies are the ideal tools to assess selection processes as well as evolution patterns of parasite populations [7,31]. In this study, we sequenced the whole genome of 70 clinical isolates from Côte d'Ivoire to identify signature of selection and migration.

Our findings suggest a strong and recent positive selection on *Pfcr* and *Pfdhps*, two malaria drug resistance genes. These findings have been corroborated by previous studies conducted in Ghana, Guinea, The Gambia and Senegal [7,12,17]. These data are reflective of the selective pressure induced by antimalarial drug treatment on West African parasite population. Chloroquine was the first line of malaria

treatment in Cote d'Ivoire until 2007 and the selective pressure may have occurred when chloroquine and sulfadoxine-pyrimethamine were widely used as first lines of antimalarial therapies in Côte d'Ivoire. In contrast to our findings, Konaté *et al* observed a local decay the chloroquine mutation rate, rejecting the hypothesis of a recent positive pressure [32]. The difference in our findings may be explained by our smaller sample size from a single sentinel site for antimalarial drug efficacy in Côte d'Ivoire which doesn't really reflect the drug pressure on *Pfcr*t in the country. While other studies [10,19] conducted in West Africa, have found a drug-induced selective pressure on *Pfdhfr*; *Pfmdr1*, our study did not corroborate these findings. We hypothesize that the timing of samples collection and malaria incidence may have been responsible for these differences. In fact our study samples were collected from 2013 to 2014 in a urban setting while samples from Senegal were collected from 2002 to 2009 in a peri-urban area [31]. *P. falciparum* virulence is attributed to the parasite's ability to modify the erythrocyte surface to adhere and invade the host immune system [33]. Therefore, *PfEMP1* encoding gene have been very polymorphic. We detected high extended haplotype score around *PfEMP1* highlighting the selective pressure on this gene. Evidence of positive directional selection in *msh3*, *glurp*, *ama1*, *msh7* and *pftrap* on chromosomes 10, 11 and 13 respectively, known as non-drug related drivers of directional selection, as well as antigenic loci that leads drug resistance [34, 35]. Some of these antigens are malaria vaccine candidates that are usually expressed in merozoites and are thought to be targets of protective immunity and under balancing selection as described in Asian and African population [10, 36].

In malaria-endemic areas, people are usually infected with multiple variants of *P. falciparum* [13]. Inbreeding levels determine the rates of an effective recombination and play a central role in understanding the population genomics of the parasite [37]. Within host diversity fixation index (*F<sub>ws</sub>*) measure the risk of inbreeding for parasites within an isolate compared to that in the whole population [14,27]. *F<sub>ws</sub>* values in our study ranged from 0.5 to 0.99 (mean 0.89) with most of the isolates having *F<sub>ws</sub>* estimates greater than 0.95. These results parallel findings from several studies conducted in West Africa [7,12]. The results showed high inbreeding levels of parasite within the host with no threat to

malaria control efforts. However, a follow-up study conducted in areas bordering Côte d'Ivoire (Burkina and Mali) with a bigger sample size shown low  $F_{ws}$  values [27]. Further exploration of the complexity of infection is required through different epidemiological settings to enable more effective interpretation of the relative out-crossing risks associated with different  $F_{ws}$  scores.

To assess gene flow between the three study sites, we conducted a Principal Component Analysis of 70 clinical samples did not show any clustering between populations, as reported recently [5]. We noticed only few outliers that were not very divergent. Indeed, the three areas are located in the same district. Computation of *Weir and Cockerham's Fst* to measure allele frequency differentiation between the three sites with a stringent genome wide cut-off ( $F_{st} \geq 0.2$ ) identified 12 highly differentiated SNPs between populations. The mean  $F_{st}$  was very low in our population. Similar findings were made in studies conducted in West Africa [7,11,12], suggesting high levels of gene. Located in chromosome 7, one of these SNPs is highly differentiated around regions coding for the parasite esterase gene *Pfpare*, conferring resistance to pepstatin ester that is a potent peptidyl inhibitor of various malarial aspartic proteases and also has parasitocidal activity [29]. Although this compound is not used in our region, this finding could be physiologically relevant because many children in malaria endemic regions are malnourished and contain low, even undetectable, levels of plasma amino acids [38]. The family of *surfin* genes also ranked high  $F_{st}$  estimate. This protein, which is expressed at the surface of the IE and the released merozoite throughout blood stage development may be crucial for parasite survival [30]. Further investigation on this protein might be needed since it is a threat to malaria intervention efforts.

To find genes under balancing selection, Tajima's D statistic test was computed. We identified genes with at least three SNPs under balancing selection. Most genes had negative Tajima's D values (mean -0,15), in agreement with observation in Senegal, Nigeria, Guinea and the Gambia [7,9,31]. This may be explained by the historical bottleneck discovered before in Africa [39]. Moreover, balancing selection over genes with Tajima's D values greater than 1 were evident in a subset of genes encoding known antigens (*Pfemp1*, *glurp*) and targets of immune selection (*msp7-like*, *surfin*) consistent with evident

decrease of population size shown in Kenya and other geographical locations in Africa [40,41] . However, it is notable to consider that not all genes with positive values are effectively under balancing selection due to the evolution of our population after a recent bottleneck.

## **Conclusion**

This study reports the first WGS of *P. falciparum* strains in Côte d'Ivoire to assess signature of selection and gene flow between study areas. We identified regions of the genome under selective pressure in drugs and vaccine encoding genes. We have also shown the lack of structure of the parasite populations. A detailed understanding of *P. falciparum* genomics could facilitate malaria control. Identification of balancing selection also reinforced the theory of population expansion in Africa. It is important to notify that further investigation should be made in a larger sample size of the country, taking account temporal and spatial factors to monitor the evolution of gene flow and genetic diversity of the parasite in Côte d'Ivoire.

## **List of abbreviations**

**WHO:** World Health Organization

**PMI:** President Malaria Initiative

***ama1*:** apical membrane antigen 1

***Pfmdr1*:** *P. falciparum* multi-drug resistance protein 1

***Pfprt*:** *P. falciparum* chloroquine resistance transporter

***Pfdhfr*:** *P. falciparum* dihydrofolate reductase

***Pfdhps*:** *P. falciparum* dihydropteroate synthase

**CNER-CI:** National Ethics and Research Committee

**PDNA:** Plasmodium Diversity Network Africa

**EDTA:** Ethyl Di-amino Tetra Acetic

**DNA:** Desoxyribo Nucleic Acid

**VQSR:** Variant Quality Score Recalibration

**VQSLOD:** Variant Quality Score

**SNP:** Single Nucleotide Polymorphism

**Pr:** *Plasmodium reichenowi*

**iHS:** integrated Haplotype Score

**MAF:** Minor Allele Frequency

**PCA:** Principal Component Analysis

***PfEMP1:*** *P. falciparum* Erythrocyte Membrane Protein 1

**SURFIN:** Surface associated interspersed protein

***msp7-like:*** merozoite surface protein 7

**GLURP:** Glutamate Rich Protein

***Pftrap:*** *P. falciparum* Thrombospondin related adhesive protein

***msp3:*** merozoite surface protein 3

***Pfpare:*** *P. falciparum* Prodrug Activation and Resistance Esterase

**IE:** Infected Erythrocyte

## **Ethics Declaration**

### **Ethics approval and consent to participate**

This study followed protocols and standards operating procedures of Good Clinical Practices of the ICH harmonized Triplicate Guide Lines for Good Clinical Practice made in 1996 and the Helsinki Declaration on human being research. The study protocol was reviewed and approved by the National Ethics and Research Committee (CNER). Samples were collected following a written informed consent from patients or their legal guardians for study participants under 18 years old.

### **Consent for publication**

Not applicable

### **Availability of data and materials**

The dataset analyzed during the current study is available from the last author on reasonable request.

## **Competing interests**

The authors declare that they have no competing interests

## **Funding**

Samples collection and Sequencing were undertaken in partnership with MalariaGEN at the Wellcome Sanger Institute and Plasmodium Diversity Network Africa with funding under the Medical Research Council Centenary Award to the MRC Centre for Genomics and Global Health, University of Oxford (G0600718 to D.K.; M006212).

Désiré Ehouni's MSc studies was funded by the Developing Excellence in Leadership and Genomics for Malaria Elimination (DELGEME) program (Grant 107740/Z/15/Z) sponsored by the Developing Excellence in Leadership and Training (DELTAS) Africa Initiative, of the African Academy of Sciences (AAS).

## **Authors' contribution**

AK supervised the clinical studies. DE analyzed the data. WY was the principal investigator. DE drafted the first manuscript. WY, AO and SN provided guidance on revision of the manuscript. All authors read and approved the final manuscript.

## **Acknowledgments**

We are grateful for the hospitality and generous collaboration of the staff of Health centers where the study took place. We thank also the patients and parents of the patients included in this study for their collaboration.

## **Authors' information**

### **Affiliation**

**Department of Biochemistry, College of Health Sciences, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya**

Désiré Ehouni & Steven G Nyanjom

**Malaria Research and Control Center, National Institute of Public Health, Abidjan, Côte d'Ivoire**

Désiré Ehouni, Abibatou Konaté & William Yavo

**Department of Parasitology and Mycology, Pharmaceutical and Biological Sciences Training and Research Unit, University of Cocody, Abidjan, Côte d'Ivoire**

Abibatou Konaté, Hervé Menan & William Yavo

**Malaria Research Program, Center for Vaccine Development and Global Health, University of Maryland School of Medicine, Baltimore, USA**

Amed Ouattara

### **Corresponding author**

Correspondence to William Yavo

### **References**

1. WHO | World malaria report 2018. (2019). WHO [Internet]. Who. World Health Organization; 2019. Available from: <http://www.who.int/malaria/publications/world-malaria-report-2018/en/>
2. USAID, CDC. U . S . President 's Malaria Initiative Cambodia: Malaria Operational Plan FY 2020. 2020;
3. Dondorp AM, Fairhurst RM, Slutsker L, MacArthur JR, M.D. JGB, Guerin PJ, et al. The Threat of Artemisinin-Resistant Malaria. *N Engl J Med.* 2011;365:1073–5.
4. Noedl H, Se Y, Schaefer K, Smith BL, Socheat D, Fukuda MM. Evidence of Artemisinin-Resistant Malaria in Western Cambodia. *N Engl J Med.* 2008;359:2619–20.
5. Amambua-Ngwa A, Amenga-Etego L, Kamau E, Amato R, Ghansah A, Golassa L, et al. Major subpopulations of *Plasmodium falciparum* in sub-Saharan Africa. *Science* (80- ) [Internet]. American Association for the Advancement of Science; 2019 [cited 2019 Aug 23];365:813–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/31439796>
6. Duffy CW, Amambua-Ngwa A, Ahouidi AD, Diakite M, Awandare GA, Ba H, et al. Multi-population genomic analysis of malaria parasites indicates local selection and differentiation at the *gdf1* locus regulating sexual development. *Sci Rep.* Nature Publishing Group; 2018;8.
7. Mobegi VA, Duffy CW, Amambua-Ngwa A, Loua KM, Laman E, Nwakanma DC, et al. Genome-wide analysis of selection on the malaria parasite *plasmodium falciparum* in West African populations of differing infection endemicity. *Mol Biol Evol.* 2014;31:1490–9.
8. Amambua-Ngwa A, Tetteh KKA, Manske M, Gomez-Escobar N, Stewart LB, Deerhake ME, et al. Population Genomic Scan for Candidate Signatures of Balancing Selection to Guide Antigen Characterization in Malaria Parasites. *PLoS Genet.* 2012;8.

9. Oyebola KM, Idowu ET, Olukosi YA, Awolola TS, Amambua-Ngwa A. Pooled-DNA sequencing identifies genomic regions of selection in Nigerian isolates of *Plasmodium falciparum*. *Parasites and Vectors*. 2017;10.
10. Amambua-Ngwa A, Park DJ, Volkman SK, Barnes KG, Bei AK, Lukens AK, et al. SNP genotyping identifies new signatures of selection in a deep sample of west African *Plasmodium falciparum* malaria parasites. *Mol Biol Evol* [Internet]. 2012 [cited 2019 Jul 29];29:3249–53. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22688945>
11. Duffy CW, Ba H, Assefa S, Ahouidi AD, Deh YB, Tandia A, et al. Population genetic structure and adaptation of malaria parasites on the edge of endemic distribution. *Mol Ecol*. 2017;26:2880–94.
12. Duffy CW, Assefa SA, Abugri J, Amoako N, Owusu-Agyei S, Anyorigiya T, et al. Comparison of genomic signatures of selection on *Plasmodium falciparum* between different regions of a country with high malaria endemicity. *BMC Genomics* [Internet]. *BMC Genomics*; 2015;16:1–11. Available from: <http://dx.doi.org/10.1186/s12864-015-1746-3>
13. Anderson TJC, Haubold B, Williams JT, Estrada-Franco JG, Richardson L, Mollinedo R, et al. Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. *Mol Biol Evol*. Oxford University Press; 2000;17:1467–82.
14. Manske M, Miotto O, Campino S, Auburn S, Zongo I, Ouedraogo J, et al. Europe PMC Funders Group Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. 2013;487:375–9.
15. Dye C, Williams BG. Multigenic drug resistance among inbred malaria parasites. *Proc R Soc B Biol Sci*. Royal Society; 1997;264:61–7.
16. Paul REL, Packer MJ, Walmsley M, Lagog M, Ranford-Cartwright LC, Paru R, et al. Mating patterns in malaria parasite populations of Papua New Guinea. *Science* (80- ) [Internet]. *Science*; 1995 [cited 2019 Aug 9];269:1709–11. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/7569897>
17. Mobegi VA, Loua KM, Ahouidi AD, Satoguina J, Nwakanma DC, Amambua-Ngwa A, et al. Population genetic structure of *Plasmodium falciparum* across a region of diverse endemicity in West Africa. *Malar J*. 2012;11.
18. Nwakanma DC, Duffy CW, Amambua-Ngwa A, Oriero EC, Bojang KA, Pinder M, et al. Changes in malaria parasite drug resistance in an endemic population over a 25-year period with resulting genomic evidence of selection. *J Infect Dis*. 2014;209:1126–35.
19. Park DJ, Lukens AK, Neafsey DE, Schaffner SF, Chang H-H, Valim C, et al. Sequence-based association and selection scans identify drug resistance loci in the *Plasmodium falciparum* malaria parasite. *Proc Natl Acad Sci*. *Proceedings of the National Academy of Sciences*; 2012;109:13052–7.
20. Miotto O, Almagro-Garcia J, Manske M, MacInnis B, Campino S, Rockett KA, et al. Multiple populations of artemisinin-resistant *Plasmodium falciparum* in Cambodia. *Nat Genet*. Europe PMC Funders; 2013;45:648–55.

21. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* [Internet]. Oxford University Press; 2009 [cited 2020 Aug 10];25:1754–60. Available from: [/pmc/articles/PMC2705234/?report=abstract](https://pubmed.ncbi.nlm.nih.gov/19505943/)
22. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* [Internet]. *Bioinformatics*; 2009 [cited 2020 Aug 10];25:2078–9. Available from: <https://pubmed.ncbi.nlm.nih.gov/19505943/>
23. Depristo MA, Banks E, Poplin R, Garimella K V., Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* [Internet]. NIH Public Access; 2011 [cited 2020 Aug 10];43:491–501. Available from: [/pmc/articles/PMC3083463/?report=abstract](https://pubmed.ncbi.nlm.nih.gov/21243432/)
24. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 1989;123:585–95.
25. Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. *PLoS Biol*. 2006;4:0446–58.
26. McVean G, Awadalla P, Fearnhead P. A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics*. Genetics Society of America; 2002;160:1231–41.
27. Auburn S, Campino S, Miotto O, Djimde AA, Zongo I, Manske M, et al. Characterization of within-host *Plasmodium falciparum* diversity using next-generation sequence data. *PLoS One*. 2012;7.
28. Weir BS, Cockerham CC. Estimating F-Statistics for the Analysis of Population Structure. *Evolution* (N Y) [Internet]. 1984 [cited 2019 Aug 19];38:1358. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28563791>
29. Istvan ES, Mallari JP, Corey VC, Dharia N V., Marshall GR, Winzeler EA, et al. Esterase mutation is a mechanism of resistance to antimalarial compounds. *Nat Commun*. Nature Publishing Group; 2017;8:1–8.
30. Winter G, Kawai S, Haeggström M, Kaneko O, Von Euler A, Kawazu SI, et al. SURFIN is a polymorphic antigen expressed on *Plasmodium falciparum* merozoites and infected erythrocytes. *J Exp Med*. The Rockefeller University Press; 2005;201:1853–63.
31. Chang HH, Park DJ, Galinsky KJ, Schaffner SF, Ndiaye D, Ndir O, et al. Genomic sequencing of *Plasmodium falciparum* malaria parasites from Senegal reveals the demographic history of the population. *Mol Biol Evol*. 2012;29:3427–39.
32. Konaté A, Gnagne PA, Bédia-Tanoh VA, Amiah-Droh M, Tano DK, Ignace Eby Menan H, et al. Low rates of *Plasmodium falciparum* Pfcrt K76T mutation in three sentinel sites of malaria monitoring in Côte d’Ivoire. *Acta Parasitol*. De Gruyter; 2018;63:795–801.
33. Pasternak ND, Dzikowski R. PfEMP1: An antigen that plays a key role in the pathogenicity and immune evasion of the malaria parasite *Plasmodium falciparum* [Internet]. *Int. J. Biochem. Cell Biol*.

2009 [cited 2020 May 20]. p. 1463–6. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1357272508005098>

34. Nguitragool W, Bokhari AAB, Pillai AD, Rayavara K, Sharma P, Turpin B, et al. Malaria parasite *clag3* genes determine channel-mediated nutrient uptake by infected red blood cells. *Cell*. 2011;145:665–77.

35. van Tyne D, Park DJ, Schaffner SF, Neafsey DE, Angelino E, Cortese JF, et al. Identification and functional validation of the novel antimalarial resistance locus PF10\_0355 in *Plasmodium falciparum*. *PLoS Genet*. Public Library of Science; 2011;7.

36. Borrmann S, Straimer J, Mwai L, Abdi A, Rippert A, Okombo J, et al. Genome-wide screen identifies new candidate genes associated with artemisinin susceptibility in *Plasmodium falciparum* in Kenya. *Sci Rep*. Nature Publishing Group; 2013;3.

37. Nkhoma SC, Nair S, Cheeseman IH, Rohr-Allegrini C, Singlam S, Nosten F, et al. Close kinship within multiple-genotype malaria parasite infections. *Proc R Soc B Biol Sci*. Royal Society of London; 2012;279:2589–98.

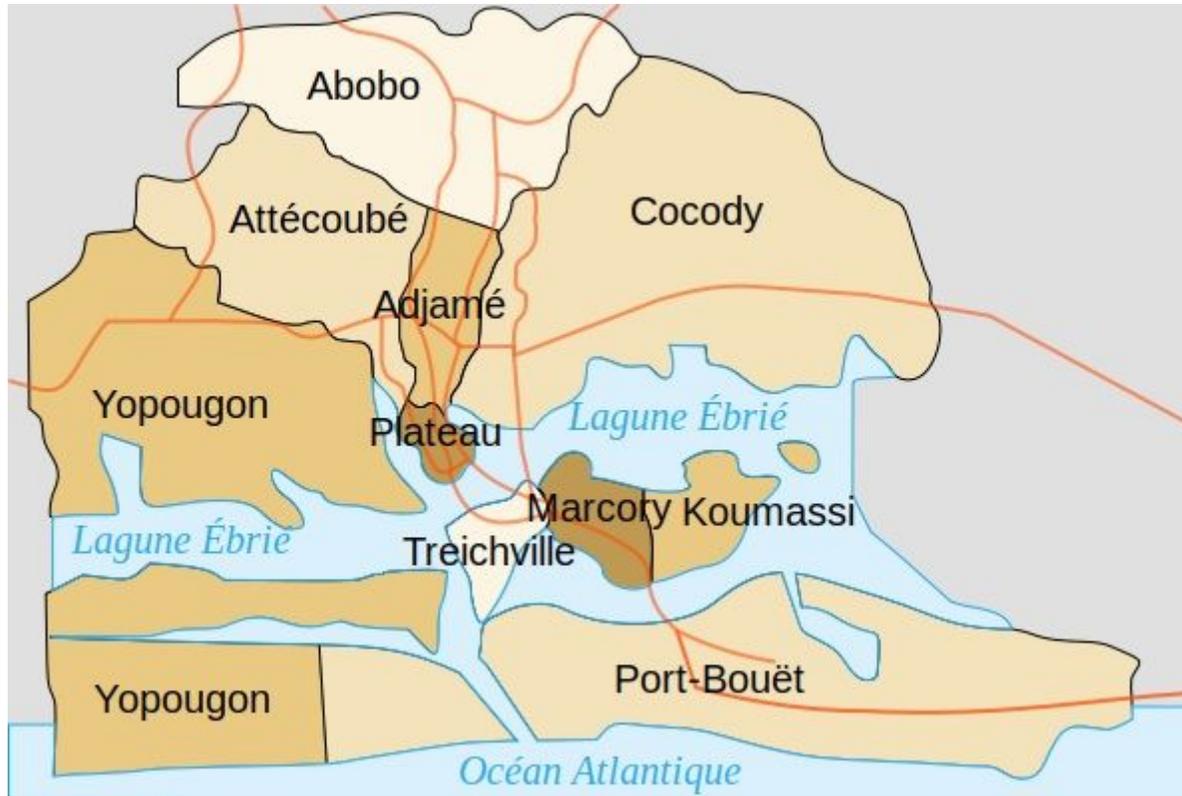
38. Liu J, Istvan ES, Gluzman IY, Gross J, Goldberg DE. *Plasmodium falciparum* ensures its amino acid supply with multiple acquisition pathways and redundant proteolytic enzyme systems. *Proc Natl Acad Sci U S A*. National Academy of Sciences; 2006;103:8840–5.

39. Joy DA, Feng X, Mu J, Furuya T, Chotivanich K, Krettli AU, et al. Early origin and recent expansion of *Plasmodium falciparum*. *Science* (80- ). 2003;300:318–21.

40. Ochola LI, Tetteh KKA, Stewart LB, Riitho V, Marsh K, Conway DJ. Allele frequency-based and polymorphism-versus-divergence indices of balancing selection in a new filtered set of polymorphic genes in *Plasmodium falciparum*. *Mol Biol Evol*. 2010;27:2344–51.

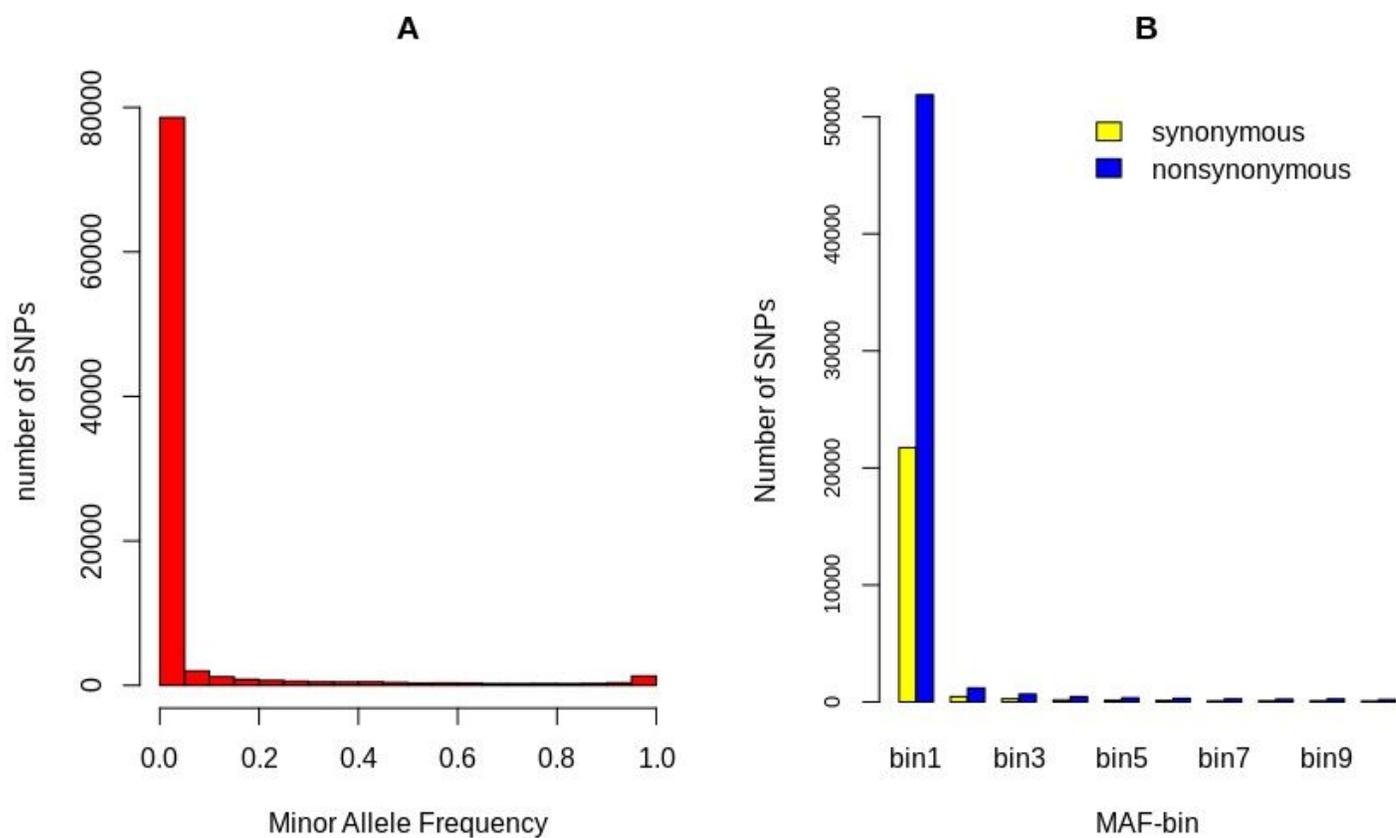
41. Weedall GD, Conway DJ. Detecting signatures of balancing selection to identify targets of anti-parasite immunity. *Trends Parasitol*. 2010;26:363–9.

## Figures



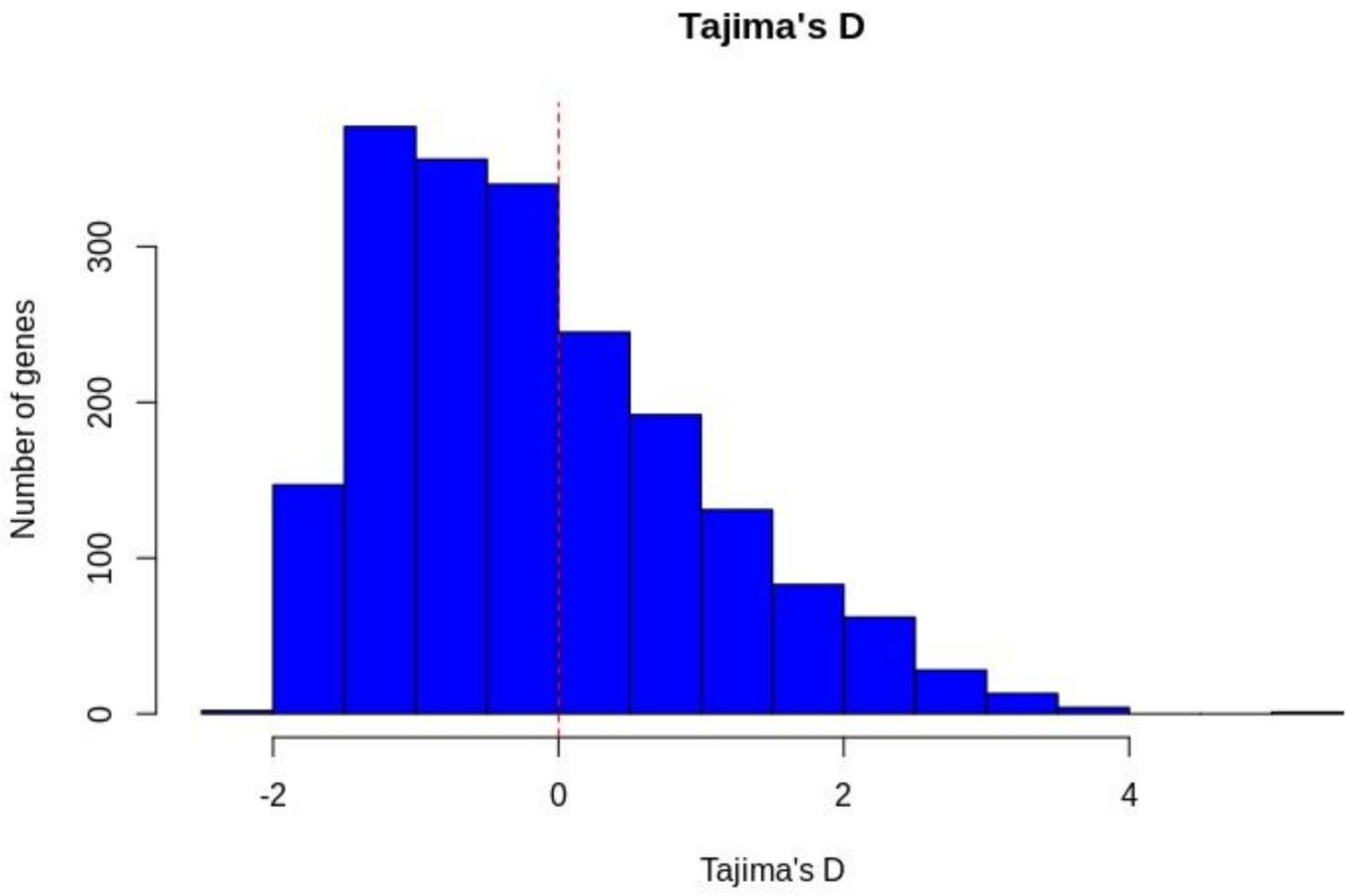
**Figure 1**

Map of Abidjan district. The three study localities (Abobo, Koumassi and Yopougon) are identified by stars



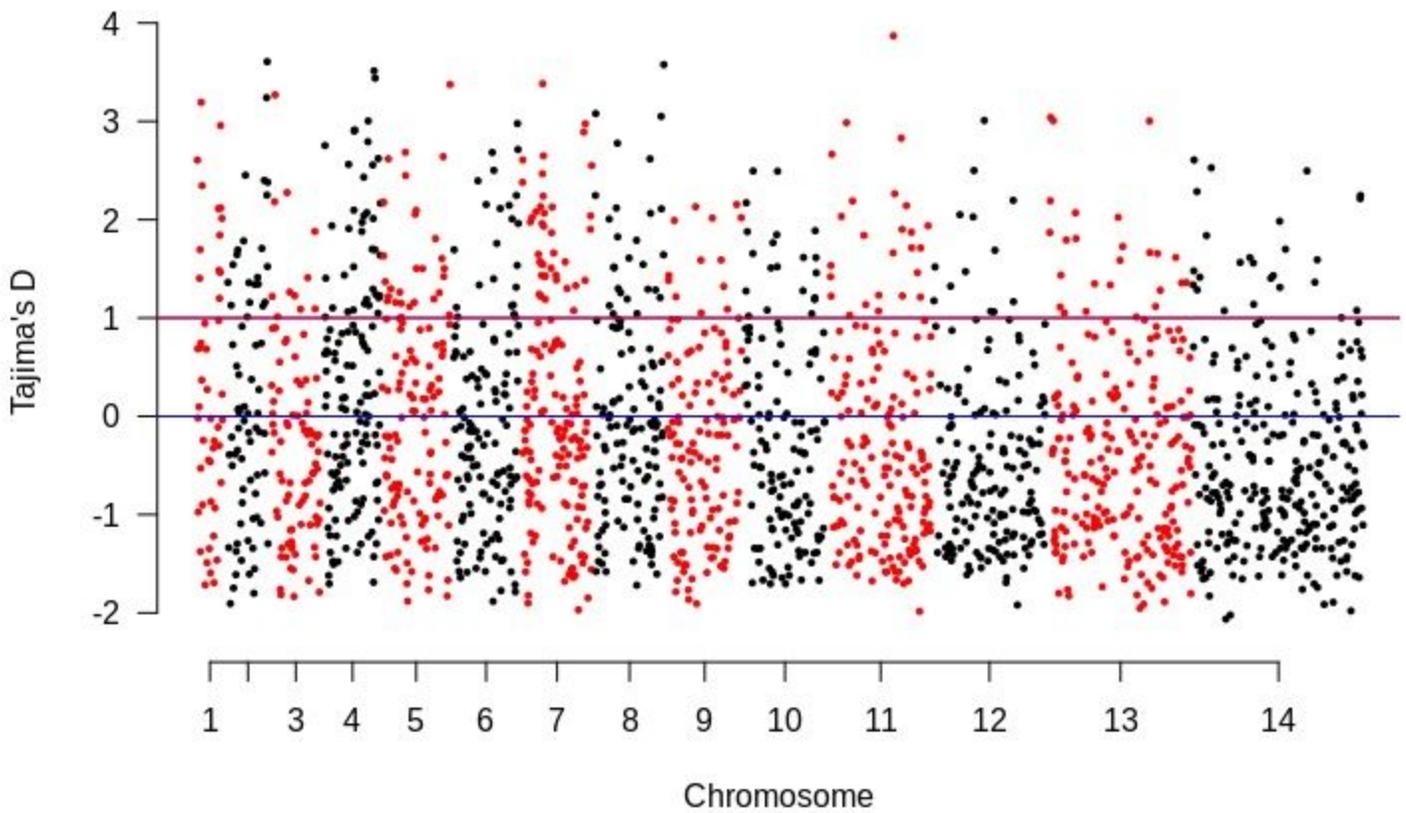
**Figure 2**

(A) Minor allele frequency distribution across 89578 bi-allelic SNPs of 70 clinical isolates from Côte d'Ivoire. (B) Distribution of non-synonymous and synonymous SNPs in the population. Within-individual and within-population heterozygosity are assigned to ten equal sized MAF bins [0.0-0.05] ... [0.45-0.5].



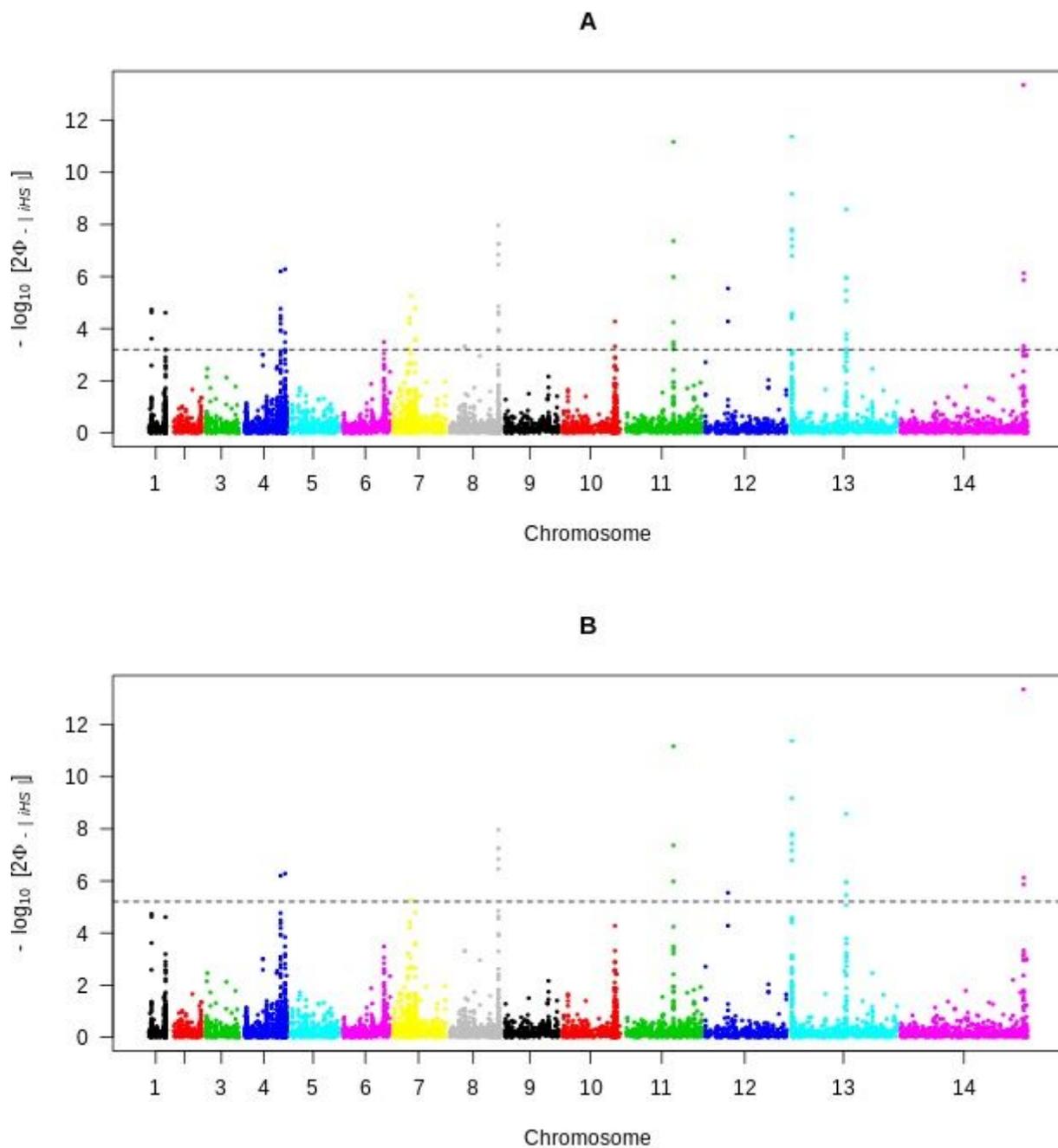
**Figure 3**

Frequency distribution of Tajima's D values for genetic loci as annotated in Pf3D7 genome.



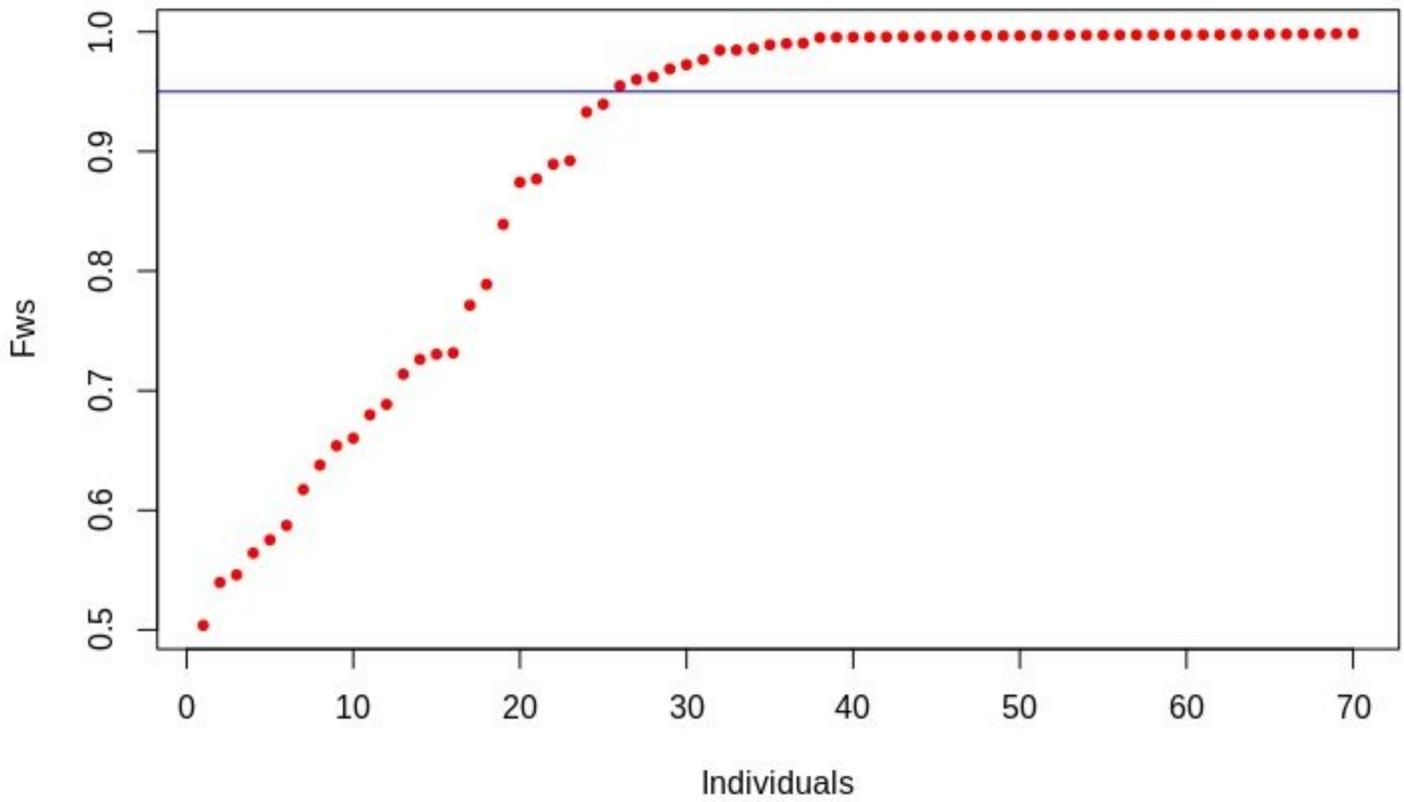
**Figure 4**

Genome-wide Manhattan plot of Tajima's D values for *P. falciparum* genes with at least three SNPs. Chromosomes are identified by the alternate red and black coloring with genes plotted as individual points based on their position within each chromosome. The blue horizontal line discriminates negative from positive values. Genes with a Tajima's D value greater than 1 in the three populations were represented above the horizontal purple line (genome wide line).



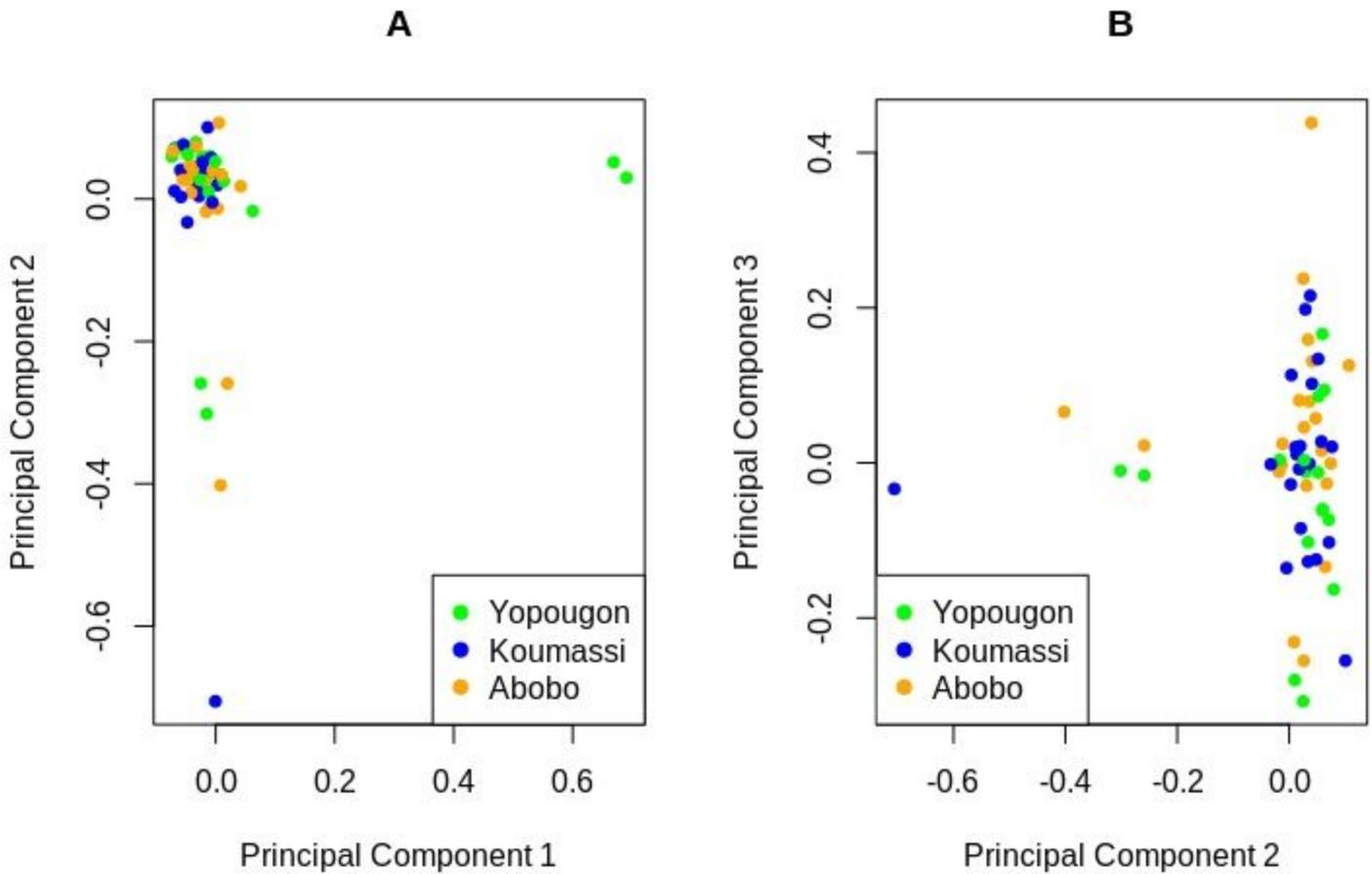
**Figure 5**

Genome wide Manhattan plot scan of standardized  $|iHS|$  for SNPs with MAF equal or greater that 5% in 70 clinical isolates from Cote d'Ivoire. Chromosomes are represented by rainbow coloring of their SNPs. High scoring SNPs are above the dashed grey line for A ( $|iHS| > 3.20$ ; suggestive line) and B ( $|iHS| > 5.20$ ; Bonferoni corrected threshold) indicating loci under recent positive selection.



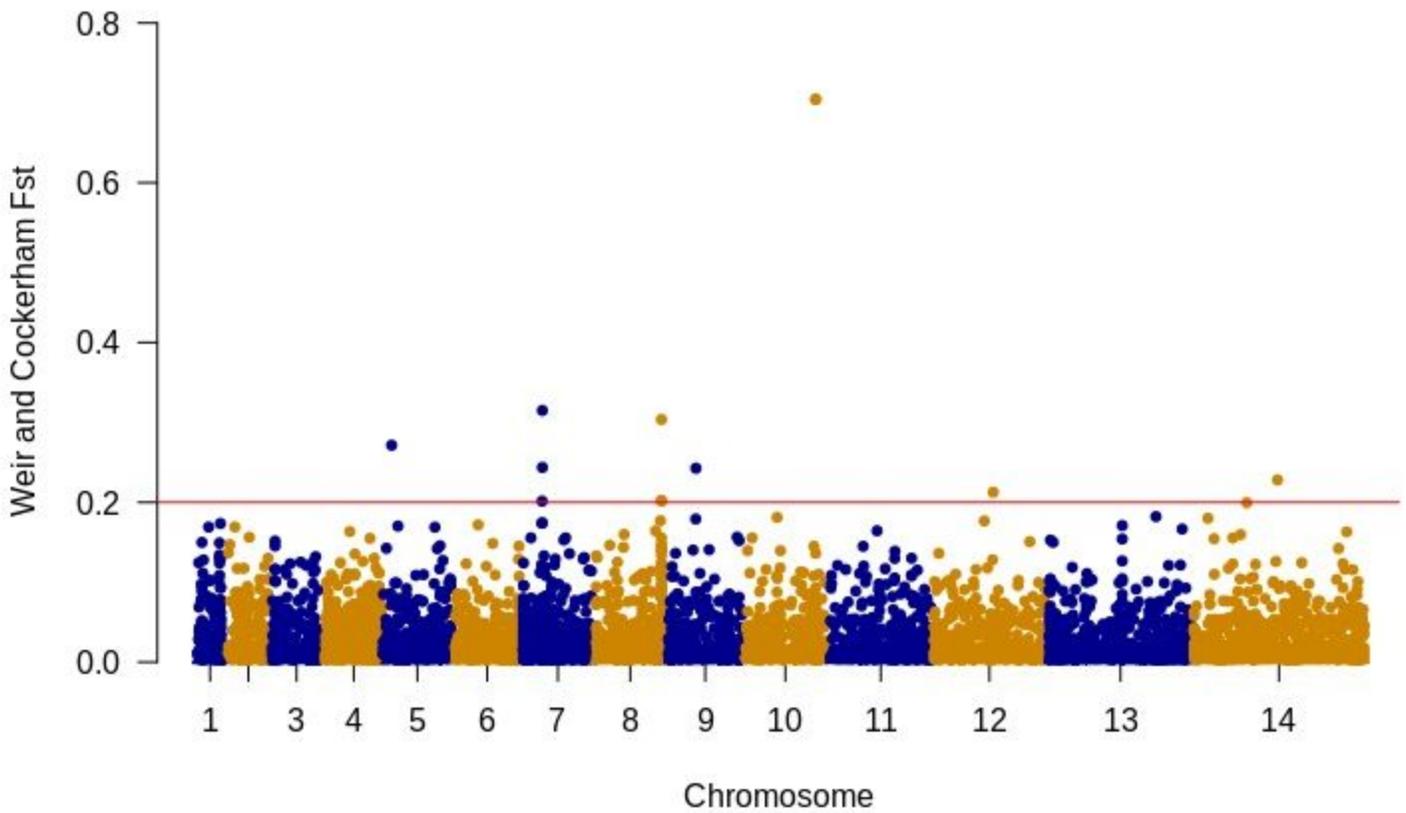
**Figure 6**

Ordered between sites (Yopougon, Abobo and Koumassi) within-infection Fws fixation index for each clinical sample with SNPs data. The horizontal blue line indicates Fws estimates > 0.95 (isolate with a single predominant genotype).



**Figure 7**

Principal components analysis of 70 clinical isolates from Abidjan (Yopougon, Koumassi, Abobo) using 89578 SNPs. A. The first and second components represent 3.2% and 4.1% of the total variation respectively. B. The second and third components cover 4.1% and 3.5% of the total variation respectively.



**Figure 8**

Manhattan plot of genome wide  $F_{st}$  for differentiation between populations for 89578 SNPs with each chromosome identified by the alternating blue/orange coloring. The red line indicates all the SNPs with  $F_{st}$  values above 0.2 (Table 3). The genome-wide  $F_{st}$  estimate is 0.001.