

# Molecular Epidemiology of Plasmodium Falciparum by Multiplexed Amplicon Deep Sequencing in Senegal

Tolla NDIAYE (✉ [ndiayetola@gmail.com](mailto:ndiayetola@gmail.com))

CHU Le Dantec de Dakar

**Mouhamad Sy**

Universite Cheikh Anta Diop de Dakar/ Hopital Aristide Le Dantec de Dakar

**Amy Gaye**

Universite Cheikh Anta Diop de Dakar/ Hopital Aristide Le Dantec de Dakar

**Katherine J. Siddle**

Broad institute of MIT and Harvard, Cambridge, MA, USA

**Daniel J. Park**

Broad Institute of MIT and Harvard, Cambridge, MA, USA

**Amy K. Bei**

Yale University School of Public Health

**Awa B. Deme**

Universite Cheikh Anta Diop de Dakar/ Hopital Aristide Le Dantec de Dakar

**Aminata Mbaye**

Universite Cheikh Anta Diop de Dakar/ Hopital Aristide Le Dantec de Dakar

**Baba Dieye**

Universite Cheikh Anta Diop de Dakar/ Hopital Aristide Le Dantec de Dakar

**Yaye Die Ndiaye**

Universite Cheikh Anta Diop de Dakar/ Hopital Aristide Le Dantec de Dakar

**Ibrahima Mbaye Ndiaye**

Universite Cheikh Anta Diop de Dakar/ Hopital Aristide Le Dantec de Dakar

**Mamadou Alpha Diallo**

Universite Cheikh Anta Diop de Dakar/ Hopital Aristide Le Dantec de Dakar

**Khadim Diongue**

Universite Cheikh Anta Diop de Dakar/ Hopital Aristide Le Dantec de Dakar

**Sarah K. Volkman**

Departments of Immunology and Infectious Diseases, Harvard University, Cambridge, MA, USA

**Aida Sadikh Badiane**

Universite Cheikh Anta Diop de Dakar/ Hopital Aristide Le Dantec de Dakar

**Daouda Ndiaye**

## Research

**Keywords:** Molecular epidemiology, Plasmodium falciparum, PfmSP1, PfmSP2, multiplexed amplicon deep sequencing, MOI

**Posted Date:** July 28th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-41863/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

**Version of Record:** A version of this preprint was published on November 10th, 2020. See the published version at <https://doi.org/10.1186/s12936-020-03471-7>.

# Abstract

## Background

Molecular epidemiology can provide important information regarding the genetic diversity and transmission of *Plasmodium falciparum*, which can assist in designing and monitoring elimination efforts. However, malaria molecular epidemiology including understanding the genetic diversity of the parasite and performing molecular surveillance of transmission has been poorly documented in Senegal. Next Generation Sequencing (NGS) offers a practical, fast and high-throughput approach to understand malaria population genetics. This study aims to unravel the population structure of *P. falciparum* and to estimate the allelic diversity, multiplicity of infection (MOI), and evolutionary patterns of the malaria parasite using the NGS platform.

## Methods

Multiplex amplicon deep sequencing of merozoite surface protein 1 (PfMSP1) and merozoite surface protein 2 (PfMSP2) genes in fifty-three *P. falciparum* isolates from two epidemiologically different areas in the South and North of Senegal, was carried out.

## Results

A total of 76 *Pfmsp1* and 116 *Pfmsp2* clones were identified and 135 different alleles were found, 56 and 79 belonged to the *pfmsp1* and *pfmsp2* genes, respectively. K1 and IC3D7 allelic families were most predominant in both sites. The local haplotype diversity ( $H_d$ ) and nucleotide diversity ( $\pi$ ) were higher in the South than in the North for both genes. For *pfmsp1*, a high positive Tajima's D (TD) value was observed in the South ( $D = 2.0453$ ) while negative TD value was recorded in the North ( $D = -1.46045$ ) and F-Statistic ( $F_{st}$ ) was 0.19505. For *pfmsp2*, non-directional selection was found with a highly positive TD test in both areas and  $F_{st}$  was 0.02111. The mean MOI for both genes was 3.07 and 1.76 for the South and the North, respectively, with a statistically significant difference between areas ( $p = 0.001$ ).

## Conclusion

This study revealed a high genetic diversity of *pfmsp1* and *pfmsp2* genes and low genetic differentiation in *P. falciparum* population in Senegal. The MOI means were significantly different between the Southern and Northern areas. Findings also showed that multiplexed amplicon deep sequencing is a useful technique to investigate genetic diversity and molecular epidemiology of *P. falciparum* infections.

## Background

*Plasmodium falciparum* is the deadliest human malaria species (99.7%) and the most prevalent in Africa, responsible for 92% of malaria cases and 93% of deaths cases [1]. A better understanding of malaria epidemiology could be helpful for disease control and prevention. The extent of genetic diversity and multiplicity of infection (MOI) are essential to understand malaria epidemiological patterns, providing insight into the dynamics of malaria transmission, human exposure to mosquito bites, acquisition of malaria immunity [2–5] and assessing of malaria control interventions [6–8].

*Plasmodium falciparum* merozoite surface protein 1 (PfMSP1) and merozoite surface protein 2 (PfMSP2) are polymorphic antigens which have been extensively studied in malaria molecular epidemiology [9–11]. Polymerase chain reaction (PCR)-based genotyping of the polymorphic regions of these genes (block 2 of *Pfmsp1* and block 3 of *Pfmsp2*) can be used to assess the allelic diversity and to determine MOI [12]. Therefore, many previous studies have used these two markers in order to evaluate genetic diversity, population structure and MOI [13–16]. However, these PCR-based methods may underestimate the extent of allelic diversity in co-infected hosts due to the use of size-polymorphisms to infer alleles; a particular problem when using short fragment amplification [17–18 19]. Microsatellite typing is an alternative tool to assess genetic diversity but might fail to detect many minor clones as it requires a cut-off of 33% of the predominant peak for a minimal peak height [20–22]. Similarly, high resolution melting (HRM) has been used for genotyping of single nucleotide polymorphisms (SNP), but shows limitations in the detection of indels [23]. Advances in next generation sequencing (NGS) technologies and bioinformatic analysis has enabled the accurate detection of clones and even minor clones, overcoming the limitations of these other methods [24]. These techniques provide a more accurate estimation of MOI compared to the standard PCR-based methods [25–26] and accurate assessment of genetic diversity in *P. falciparum* populations [27]. Indeed, using amplicon deep sequencing of *Pfmsp1* using Illumina MiSeq/HiSeq Zhong *et al.*, reported 307 haplotypes with 99 predominant and an average MOI of 2.68 from 222 samples collected in Ethiopia, East Africa [24, 27].

In Senegal, previous investigations of *P. falciparum* genetic diversity and malaria molecular epidemiology [7, 8, 28–30] have mainly focused in one region, Thies low malaria endemicity area. Thus, limited information on malaria molecular epidemiology are available for other areas of the country and further detailed studies are needed to understand the genetic diversity and population structure country-wide. Here, multiplexed amplicon deep sequencing of *Pfmsp1* and *Pfmsp2* genes in *P. falciparum* isolates sampled from a malaria hyper-endemic area in the South (Kedougou region) and a pre-elimination areas in the North (Podor and Matam) was performed. This data was used to unravel the population structure, to estimate allelic diversities and MOI, and to understand the population dynamics, gene flow and evolutionary pattern of the malaria parasite *P. falciparum* in Senegal.

## Methods

### Study sites, sample collection

This study was conducted in Kédougou region, (KG, Southern Senegal) and Podor and Matam (PM, Northern Senegal). Kédougou (685 Km from the capital, Dakar) (**Fig. 1**) is hyper-endemic for malaria with an incidence higher than fifteen malaria cases per 1000 habitants. In this area, malaria transmission is seasonal from July to December with a high entomological inoculation rate (EIR) ranging from 20 to 100 infectious bites/person/ year [31]. In contrast, Northern Senegal (Podor and Matam, **Fig. 1**) is a malaria pre-elimination area with an incidence of less than five cases per 1,000 habitants and an EIR equal to 1.3 infectious bites/person/ year [31]. Samples were collected from patients with malaria symptoms attending the National Malaria Control Programme's (NMCP) sentinel site health facilities during malaria transmission season between September and December 2016 (**Fig. 1**). In the South, venous blood samples were collected on filter paper from patients fulfilling the following inclusion criteria: having fever (axillary temperature  $\geq 37.5$  C) or history of fever in the previous 48 hours, age ranging from 6 months to 75 years and suffering from uncomplicated *P. falciparum* malaria with parasite density  $\geq 1000$  asexual forms per microlitre. Patients who presented signs or symptoms of severe malaria as defined by the WHO [32] and pregnant women were not included in the study. In the North, *P. falciparum* positive Rapid Diagnostic Tests (RDTs) were collected in health facilities (NMCP sentinel sites) in Matam and Podor.

### **Amplification and sequence analysis of Pfmosp1 and Pfmosp2**

The experimental workflow from amplicon preparation to data analysis are detailed in **Fig. 2**.

### **Multiplex PCR amplification of Pfmosp1 and Pfmosp2 genes**

Parasite genomic deoxyribonucleic acid (DNA), from filter paper and RDTs was extracted using QIAamp DNA Mini kit (Qiagen, QIAGEN, USA) according to the manufacturer's instructions. *P. falciparum* molecular identification was performed using the photo-induced electron transfer (PET)-PCR assay [33] on a Roche LightCycler 96 instrument (Roche Molecular Systems,

Inc). Each experimental run included both a negative (no template) and a positive (3D7 *P. falciparum* strain) control. Samples with a cycle threshold (CT) of 40 or less were scored as positive [33–34]. The two *Pfmosp1* and *Pfmosp2* polymorphic genes were amplified by multiplex PCR. All PCR reactions were conducted in a 20  $\mu$ l reaction mixture containing 2  $\mu$ l of template DNA, 4  $\mu$ l Phusion high-fidelity (HF) PCR master mix, 0.2  $\mu$ l HF Phusion Taq Polymerase, 10 mM deoxyribonucleotide triphosphate (dNTPs), 10  $\mu$ M of each forward and reverse primer of each gene. Cycling conditions were as follows: initial denaturation at 94 °C for 5 min, followed by 40 cycles of denaturation at 94 °C for 30 s, annealing at 58 °C for 1 min and extension at 68 °C for 1 min 40 s; a final extension was done at 68 °C for 5 min [23, 35]. The PCR products were revealed by electrophoresis on 2% agarose gels stained with ethidium bromide and visualized under ultraviolet (UV) trans-illumination (VersaDoc®, BIORAD, Hercules, USA). The size of PCR fragments was estimated using a 100 bp molecular weight ladder (**Fig. 2A**). The length of PCR product varied from 800 bp-1000 bp and from 600 bp-800 bp for *Pfmosp1* and *Pfmosp2*, respectively.

### **Multiplex amplicon deep sequencing of Pfmosp1 and Pfmosp2**

Amplicons from each sample were purified using a 0.6X DNA SPRI (Agencourt AMPure XP beads Beckman Coulter®, CA, USA). Samples were normalized to 0.25 ng/μl concentration using the Qubit® 3.0 Fluorometer and Invitrogen™ Qubit® Quantitation Kit (Life Technologies, Carlsbad, California). The Nextera XT DNA Library Preparation kit (Illumina, USA) was used for library preparation of 1 ng of purified PCR product from each sample [36]. Sequencing libraries were cleaned using a 0.6X DNA SPRI and quantified by qPCR using the KAPA Library Quantification kit (KapaBiosystems, MA, USA) on a Roche LightCycler 96 instrument. The concentration of sequencing libraries was normalized and all samples were pooled. The concentration and mean fragment length of the pool were determined using the Agilent High Sensitivity DNA Kit (2100 Expert Software) and the concentration was confirmed by KAPA qPCR. The pool of all samples was sequenced on one run of an Illumina Miseq sequencer with a Miseq v2 reagent Kit (VEROGEN, North America) and 100nt paired end sequencing (Fig. 2B).

## Data analysis

Sequencing data was analysed using viral-ngs, version 1.19.2 pipeline implemented on the DNAnexus cloud-based platform [37]. Briefly, raw reads were demultiplexed to individual sample libraries and reads mapping to the human genome or to other known technical contaminants (e.g., sequencing adapters) were removed. After filtering the reads, they were assembled based on *Pfmsp1* or *Pfmsp2* gene mapping using the Trinity software [38]. For every sample, the number of clones of each gene were determined from the number of contigs generated by Trinity software. Following de-novo assembly, clones were filtered considering an average of k-mer coverage of 2 with length of at least 200 bp [ $2*(k - 1)$ ,  $k = 101$ ]. The average read coverage of the individual contig was 1000 X, the percentage of the average read coverage of individual contig was 1% of the average read coverage of overall contigs and allele frequencies > 1%. The clones found were classified into allelic families according to the polymorphism of block 2 of *Pfmsp1*, and block 3 of *Pfmsp2* using the Geneious software (Fig. 2.C). For *Pfmsp1*, K1-like were characterized by amino acid tripeptide repeat units SAQ, SGT, SGA, SGP, MAD20-like by SVA, SGG, SKG, SVT and R033-like by non-synonymous amino acid change. For *Pfmsp2*, IC3D7-like presented GA or SG amino acid dipeptides repeat units and a poly-threonine repeat while FC27-like were characterized by 32 amino acid, ADTIASGSQSSTNSASTSTTNGESQTTTPTA or its variants and 12 amino acid: ESISPSPPIITTT.

For genetic diversity and population structure analysis, DNA sequences were first aligned using BioEdit Sequence Alignment Editor V.7.0.5.3 [39]. DnaSP v. 6.0 software was used to estimate haplotype diversity (Hd), nucleotide diversity ( $\pi$ ) and Tajima's D (TD) [40–41]. Genetic differentiation between populations, Wright's F-statistic (Fst), was estimated using Arlequin software version 3.5 [42]. The Geneious software was used to create the neighbor-joining tree in order to identify the number of parasite clusters and parental links between strains.

The MOI mean was calculated by dividing the total number of alleles detected by the total number of samples for each gene [43]. The online Biostatgv Student's test (<https://biostatgv.sentiweb.fr/?module=tests/student>) was used to compare MOI between localities. For all tests, comparisons were considered statistically significant when P-values < 0.05.

# Results

Fifty-three (53) samples (28 from the South and 25 from the North) were successfully amplified and sequenced for the two polymorphic markers *Pfmsp1* and *Pfmsp2*. All Fastq sequencing data were submitted to the SRA/NCBI database with the following accession numbers: from SRX8632497 to SRX8632549 (<https://www.ncbi.nlm.nih.gov/sra/PRJNA642844>)

## Allelic polymorphisms and population structure of *Pfmsp1*

Fifty-one *Pfmsp1* gene were successfully assembled from 53 *P. falciparum* isolates. A total of 76 clones were found of which 47 and 29 in the South and the North, respectively (Additional file 1). Of the detected *Pfmsp1* clones, 56 different alleles were identified with 33 in the South area and 23 in the North area.

In the South, the 33 alleles found were classified into 3 different allele types: 18 of K1-like, 8 of MAD20-like, and 7 of RO33-like (Table 1). In the K1-like, each allele had different numbers and arrangements of amino acid tripeptide repeat unit SGASAQSGT, SGT or SGPSGT. The MAD20-like were varied by the number of SGGVA amino acid tripeptide motif. Meanwhile RO33-like was less polymorphic, with only limited numbers of amino acid substitutions. The neighbor-joining tree of different allele types showed two parasite clusters (Fig. 3A). *Plasmodium falciparum* population presented high level of genetic diversity with high haplotype diversity (Hd) of 0.93 and a nucleotide diversity ( $\pi$ ) of 0.28306. The TD test was performed in each area in order to identify any deviation from the neutral evolution of parasites and a highly positive TD value was observed in the South ( $D = 2.0453$ ) with a statistically significantly different ( $p < 0.05$ ) (Table 2).

Table 1  
number of alleles per allelic family of *Pfmsp1* and *Pfmsp2* in the South (Kedougou) and the North (Podor and Matam).

Localities	Kedougou	North area
Gene / Allelic families	N	
Msp1		
K1-like	18	8
MAD20-like	8	6
R033-like	7	9
Total alleles	33	23
Msp2		
IC/3D7-like	35	19
FC27-like	18	7
Total alleles	53	26
N: number of alleles		

Table 2  
Genetic diversity of *mSP1* and *mSP2* genes of *P. falciparum* isolates from the South (Kedougou) and North (Podor and Matam).

Genes	<i>MSP1</i>		<i>MSP2</i>	
	Kedougou	North area	Kedougou	North area
No of Isolate	28	23	28	25
Clones (N)	47	29	76	40
Haplotype diversity (Hd)	0.93	0.761	0.82	0.843
Nucleotide diversity	$\pi$	0.28306	0.09457	0.48339
	$\theta$	0.24245	0.16236	0.3534
Tajima's D $\square$ test of neutrality	2.0453*	-1.46045	5.21073*	3.46684*
FSTs	0.19505*		0.02111	
*indicates the significance at P,0.05 level.				

In the North, the 23 alleles found were classified into: 8 of K1-like, 6 of MAD20-like and 9 of RO33-like (Table 1). The K1-like were different based on the number of amino acid tripeptide repetition SGASAQSGT or SGPSGT. The MAD20-like varied with SGGSSVA repeat unit and RO33-like was relatively well conserved sequences. The neighbor-joining tree of different allele types revealed two parasites clusters (Fig. 3B). Parasite population exhibited less genetic variation with  $\pi = 0.09457$  and  $Hd = 0.76$  and a negative TD value was recorded ( $D = -1.46045$ ), which was not statistically significant (Table 2). A significant genetic differentiation between the South and the North *P. falciparum* populations was observed with  $F_{st} = 0.19505$  ( $p < 0.05$ ) (Table 2).

### Allelic polymorphisms and population structure of *Pfmsp2*

All 53 *Pfmsp2* gene of *P. falciparum* isolates were successfully assembled. A total of 116 clones were found, 76 in the South and 40 in the North (Additional file 2). Among these clones 79 different alleles were identified of which 53 and 26 in the South and the North, respectively.

In the South, among the 53 different alleles detected, 35 belonged to the IC3D7-like while the other 18 were classified into the FC27-like (Table 1). IC3D7-like showed different dimorphic structures based on GA or SG amino acid dipeptides repeat units and poly-threonine repeats. Meanwhile, FC27-like varied with (ADTIASGSQSSTNSASTSTTNNNGESQTTTPTA) or its variant amino acid succession. The neighbour-joining tree of parasites showed four clusters (Fig. 4.A).

In the North, of the detected 26 different alleles, 19 were classified into IC3D7-like and 7 belonged to FC27-like (Table 1). IC3D7-like varied with GA or SG amino acid dipeptides repeat units and poly-threonine repeat. FC27-like were different based on the amino acid succession (ADTIASGSQSSTNSASTSTTNNNGESQTTTPTA) and the typical non-synonymous substitution of 6 amino acids (SSGNAP) was found at the C-terminal region of 5 of the 7 FC27-like identified. The neighbour-joining tree of different alleles revealed three clusters (Fig. 4B).

The observed genetic diversity was similarly elevated in both regions ( $Hd = 0.82$  in the South and  $Hd = 0.843$  in the North), and nucleotide diversity was also very similar ( $\pi = 0.48339$  in the South and  $\pi = 0.42473$  in the North). Non-directional selection was found with a highly positive TD test in the two areas,  $D = 5.21073$  in the South and  $D = 3.46684$  in the North with statistically different ( $p < 0.05$ ). The  $F_{st}$  between the two localities showed an almost panmictic population ( $F_{st} = 0.02111$ ) (Table 2).

## Multiplicity Of Infection (moi)

The mean MOI for *Pfmsp1*, *Pfmsp2* and both genes together in each area are shown in (Fig. 5). The mean MOI for *Pfmsp1* was 1.5 in the South and 1.24 in the North, the number of clones per isolate ranged from 1 to 5 in these two regions and no statistically significant difference was observed between regions ( $p = 0.159$ ). For *Pfmsp2*, mean MOI was significantly different between the South and the North ( $p = 0.010$ ). The number of clones per samples ranged from 1 to 7 in the South with a MOI = 2.71 while in

the North a MOI = 1.6 and the average number of clones varying between 1 and 4. Mean MOI for both genes was 3.07 and 1.76 in the South and North respectively with statistically significant differences between areas ( $p = 0.001$ ).

## Discussion

This study aimed to compare genetic diversity of *P. falciparum* and parasites population structure of the Southern (a malaria hyper-endemic area) and Northern Senegal (a malaria pre-elimination area) by using multiplexed amplicon deep sequencing of *Pfmsp1* and *Pfmsp2* genes. To date, a little data is available on the genetic diversity of *P. falciparum* particularly in the North area where local monitoring on parasite population structure could help to achieve malaria elimination. Therefore, this study provides insight of the molecular characteristics of the *P. falciparum* population, in order to gain in-depth understanding of the genetic nature of parasites circulating in these areas.

A high degree of polymorphism of the genes *Pfmsp1* and *Pfmsp2* was found in Senegal. A total of one hundred thirty-five different alleles were identified; 56 *Pfmsp1* alleles (26 K1-like, 14 MAD20-like and 16 RO33-like) and 79 *Pfmsp2* alleles (54 IC3D7-like and 25 FC27-like). These results are consistent with previous observations of an extremely polymorphic *Pfmsp1* (8K1-like, 14MAD20-like, 6 RO33-like) and *Pfmsp2* (41 IC3D7-like, 18 FC27-like) using a similar amplicon deep sequencing technique in Myanmar [44] and Aspelg Jones *et al.* who reported 225 different K1-like, 123 different MAD20-like and 9 distinct RO33-like in Africa and Asia [45]. Most of these alleles were detected in the South, hyper-endemic area, 33 and 53 distinct alleles were observed for *Pfmsp1* (18 K1-like, 8 MAD20-like, and 7 RO33-like) and *Pfmsp2* (35 IC3D7-like and 18 FC27-like), respectively. Comparable results from elsewhere in sub-Saharan Africa (Gabon and Ivory-Cost) reported by Yavo *et al.* underlined the same diversity with 27 K1-like, 22 MAD20-like and 18 RO33-like alleles for *Pfmsp1* and 28 IC3D7-like and 20 FC27-like for *Pfmsp2* [16]. In areas with high malaria transmission intensity the probability of genetic recombination between circulating strains is high and could affected the variation, the numbers and arrangements of amino acid repeats units of *Pfmsp1* and *Pfmsp2* which are the main factors driving the increased allelic diversity [44, 46–48]. It is known that a large number of mutations in parasite populations arms the organism to successfully battle out adverse environmental conditions through adaptation. Indeed, it has been shown that *P. falciparum* genetic diversity is indicative of the ability of malaria parasites to adapt to their hosts by selection of advantageous traits, such as drug resistance and antigenic variability [49].

For *Pfmsp1*, despite a haplotype diversity of  $Hd = 0.76$  in the North, the negative value of TD suggests there is clonal expansion of the parasite. A similar finding was also reported by the team of Daniels *et al.* in 2013 in Thiès, using a 24 SNP molecular barcode approach [30]. This situation suggests a high self-fertilization rate between genetically identical parasites during the sexual stages in the mosquito in these areas where malaria transmission is declining. However, in the South the *P. falciparum* population exhibits a high  $Hd$  ( $Hd = 0.93$ ) with an important nucleotide substitution rate ( $\pi = 0.28306$ ) and a  $D$  value significantly different from expected under the neutral model of molecular evolution ( $D = 2.0453$ ). This variability in parasite population genetics observed between studies sites is consistent with the difference

in malaria incidence in these two regions [2]. These findings suggest that the genetic diversity of *P. falciparum* is greater in high malaria transmission areas and decreases when transmission regresses [2, 4]. The neighbour-joining tree of different allele types revealed two parasite clusters in the South as in the North and many clones sharing same patterns in amino acid level showing parental link between strains. Finally, the low  $F_{st}$  value of *Pfmsp1* between parasite population from the South and the North indicates the presence of gene flow between these populations, likely facilitated by extensive human migration events between regions causing the vector's displacement.

For *Pfmsp2*, three clusters with a high  $H_d$  and  $\pi$  were observed associated with a balancing selection (Positive TD value) in each area and an almost panmitic parasite population (low  $F_{st}$  and genetic relatedness) between the two studies sites. No major difference was observed between the two areas and is also alarming as it was reported that malaria parasites use genetic diversity as an escape mechanism either from human immunity or from anti-malarial treatment [50–51]. Indeed, the larger the clones and different allele types in the parasite's population are, the more easily the parasite population may be able to adapt to the environmental conditions, potentially affecting malaria control measures in the area [52–53]. Similarly, a low genetic differentiation between the parasite populations from the two study sites indicates a gene flow within the population probably deriving from the frequent and high mobility of infected human hosts. This situation could cause malaria outbreaks in residual foci (unpublished data). These areas are also vulnerable to a rebound of *P. falciparum* genetic diversity as previously reported by Daniels *et al.* using a 24 SNP molecular barcodes molecular in 2015 in Thiès, a low-transmission area in Senegal [7]. Such events are a major challenge for malaria control and elimination in low-transmission areas.

Multiclonal infections were higher in the South, suggesting that this is more common in areas of high malaria endemicity. This is consistent with previous observations for high transmission areas, such as in Ethiopia [24]. Allelic polymorphism was also higher in the South, which could increase the likelihood of recombination (genetic crossing-over) in these areas [46]. Recombination can generate novel parasite variants, giving rise to new genetic diversity that may exhibit phenotypic differences such as virulence, drug resistance or immune evasion [54–55]. MOI is an indicator of malaria transmission level because: it has been found to be higher in high malaria transmission areas and decreases when transmission declines [2, 56–61]. MOI mean was significantly higher in the South than in the North ( $p = 0.001$ ), suggesting that malaria transmission remains still active in this area. Moreover, the low mean MOI found in the North (1.76), may suggest a decline in malaria transmission levels, underlining the effectiveness of the scale-up in malaria control measures since 2006 [32]. The results of this study contribute insight into the genetic diversity of *P. falciparum* population in two malaria endemic areas in Senegal. In addition, this study argues that multiplexed amplicon deep sequencing represents an important advance for surveillance of parasites population within the country. This method allows the sequencing of several genes of interest in one run and can be readily adapted to other genes as the sequencing of anti-malaria resistance markers.

# Conclusion

This study revealed a high genetic diversity of *Pfmsp1* and *Pfmsp2* genes in Senegal. Findings also showed an almost panmitic *P. falciparum* population with an important gene flow and MOI disparities between a malaria hotspot (Kedougou, Southern area) and malaria pre-elimination areas (Podor and Matam, Northern areas), a situation that may hamper malaria elimination in these Northern regions. Therefore, continuous molecular epidemiological surveillance to monitor the genetic diversity and parasite population structure in Senegal would be necessary for malaria control in Senegal. Additionally, this study showed that multiplexed amplicon deep sequencing is a useful technique to investigate genetic diversity and molecular epidemiology of *P. falciparum* parasite infections.

# Abbreviations

NGS

Next Generation Sequencing

MOI

Multiplicity of infection

PfMSP1

Merozoite surface protein 1 of *Plasmodium falciparum*

PfMSP2

Merozoite surface protein 2 of *Plasmodium falciparum*

*pfmsp1*

*Plasmodium falciparum msp1* gene

*pfmsp2*

*Plasmodium falciparum msp2* gene

Hd

Haplotype diversity

$\pi$

Nucleotide diversity

TD

Tajima's D

PCR

Polymerase chain reaction

HRM

High resolution melting

SNP

Single nucleotide polymorphism

NMCP

National Malaria Control Programme

WHO

World Health Organization  
EIR  
Entomological inoculation rate  
DNA  
Deoxyribonucleic acid  
RDTs  
Rapid Diagnostic Tests  
PET  
Photo-induced electron transfer  
CT  
Cycle threshold  
HF  
High-fidelity  
DNTPs  
Deoxyribonucleotide triphosphate  
UV  
Ultraviolet  
Fst  
F-Statistic  
NCBI  
National Center for Biotechnology Information  
PM  
Podor and Matam,  
KG  
Kédougou

## **Declarations**

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

The study was approved by the National Ethics Committee for Health Research of Senegal. Before patients were enrolled in the studies sites, benefits and any perceived risks were explained to all participants in French or local languages. Informed consent for adults or guardians consent for children less than 18 years old were obtained before participant recruitment and sample collection.

### **Consent for publication**

The participants in this study are consent for publication

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

The data supporting the findings of this article are included within the article.

## Competing interests

The authors declare that they have no competing interests.

## Funding

The work was supported by the International Centers of Excellence for Malaria Research, (ICEMR) West Africa (U19AI089696).

## Authors' contributions

TN, MS, AG, KS and DP conceived and designed the study. TN, MS and AG carried out the experiments and collected data. TN, MS, AG and KS analysed the data. TN, MS, AG, KS, DP, AKB, ASB, ABD, MAD and KD contributed to writing the manuscript. All authors read and approved the final manuscript.

## Acknowledgements

We acknowledge all participants of the study in Senegal. We thank the staffs from Broad Institute of MIT and Harvard, Cambridge, MA, USA for assisting with Illumina sequencing and data analysis pipelines. We thank also Mamane Nassirou Garba, Younouss Diedhiou, Lamine Ndiaye, Amadou Mactar Mbaye and staff of Cheikh Anta Diop University for their contribution in this study.

## References

1. WHO.: World malaria report 2018. Geneva, World Health Organization.  
<http://www.who.int/malaria/publications/world-malaria-report-2018/report/en/>.
2. Anderson TJ, 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.* 2000;17:1467–82.
3. Khaireh BA, Assefa A, Guessod HH, Basco LK, Khaireh MA, Pascual A, et al. Population genetics analysis during the elimination process of *Plasmodium falciparum* in Djibouti. *Malar J.* 2013;12:201.
4. Nabet C, Doumbo S, Jeddi F, Konaté S, Manciuilli T, Fofana B, et al. Genetic diversity of *Plasmodium falciparum* in human malaria cases in Mali. *Malar J.* 2016;15:353.
5. Mohd Abd Razak MR, Sastu UR, Norahmad NA, Abdul-Karim A, Muhammad A, Muniandy PK, et al. Genetic diversity of *Plasmodium falciparum* populations in malaria declining areas of Sabah, East Malaysia. *PLoS One.* 2016;11:e0152415.

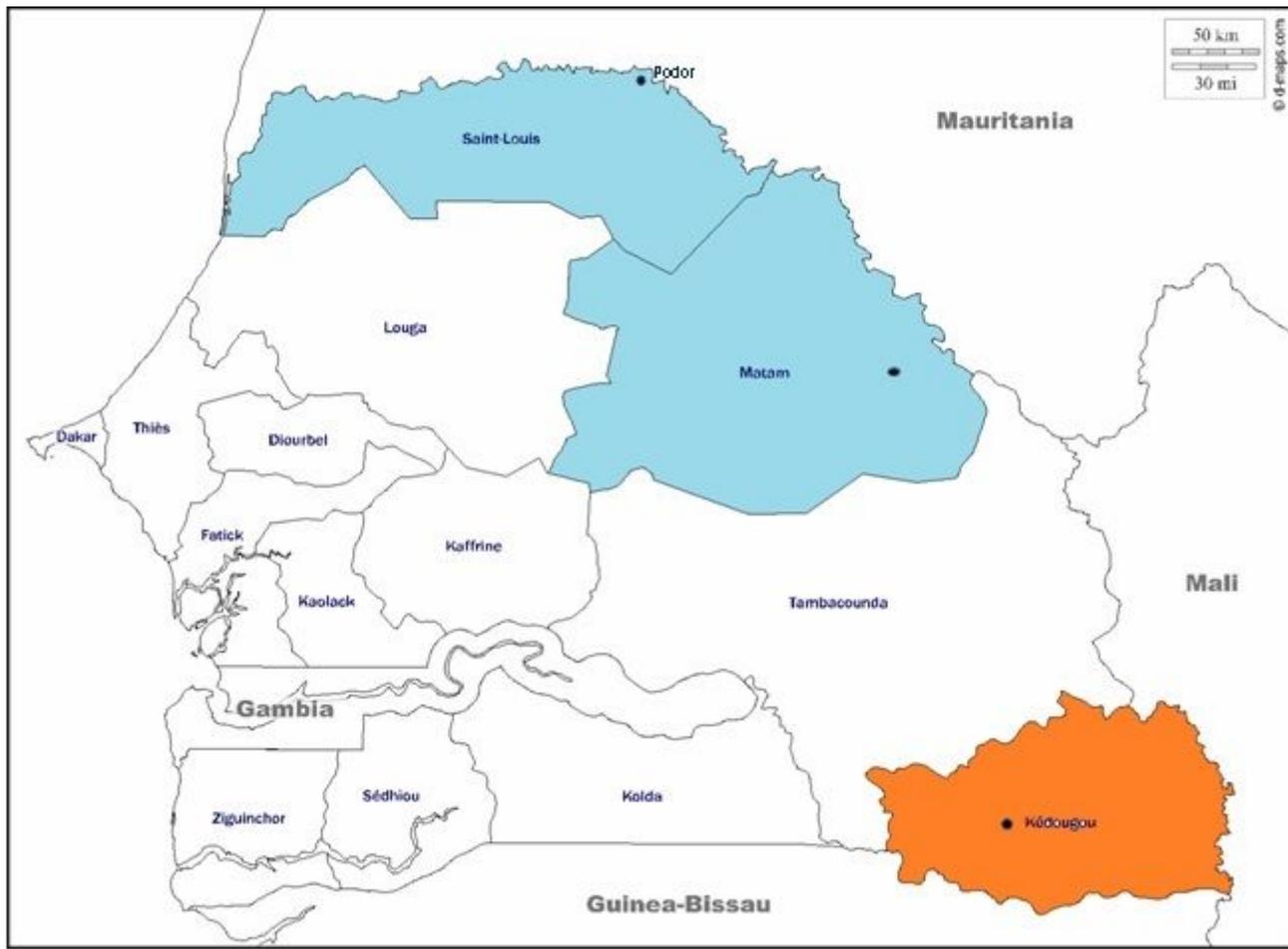
6. Arnott A, Barry AE, Reeder JC. Understanding the population genetics of *Plasmodium vivax* is essential for malaria control and elimination. *Malar J.* 2012;11:14.
7. Daniels RF, Schaffner SF, Wenger EA, Proctor JL, Chang H-H, Wong W, et al. Modeling malaria genomics reveals transmission decline and rebound in Senegal. *Proc Natl Acad Sci USA.* 2015;112:7067–72.
8. Niang M, Thiam LG, Loucoubar C, Sow A, Sadio BD, Diallo M, et al. Spatio-temporal analysis of the genetic diversity and complexity of *Plasmodium falciparum* infections in Kédougou, Southeastern Senegal. *Parasit Vectors.* 2017;10:33.
9. Gilson PR, Nebl T, Vukcevic D, Moritz RL, Sargeant T, Speed TP, et al. Identification and stoichiometry of glycosylphosphatidylinositol-anchored membrane proteins of the human malaria parasite *Plasmodium falciparum*. *Mol Cell Proteomics.* 2006;5:1286–99.
10. Lin CS, Uboldi AD, Epp C, Bujard H, Tsuboi T, Czabotar PE, et al. Multiple *Plasmodium falciparum* Merozoite Surface Protein 1 complexes mediate merozoite binding to human erythrocytes. *J Biol Chem.* 2016;291:7703–15.
11. Baldwin MR, Li X, Hanada T, Liu S-C, Chishti AH. Merozoite surface protein 1 recognition of host glycophorin A mediates malaria parasite invasion of red blood cells. *Blood.* 2015;125:2704–11.
12. Tanabe K, Mackay M, Goman M, Scaife JG. Allelic dimorphism in a surface antigen gene of the malaria parasite *Plasmodium falciparum*. *J Mol Biol.* 1987;195:273–87.
13. Mwingira F, Nkwengulila G, Schoepflin S, Sumari D, Beck H-P, Snounou G, et al. allele frequency and diversity in sub-Saharan Africa. *Malar J.* 2011;10:79.
14. Ahmedou Salem MSO, Ndiaye M, OuldAbdallahi M, Lekweiry KM, Bogreau H, Konaté L, et al. Polymorphism of the merozoite surface protein-1 block 2 region in *Plasmodium falciparum* isolates from Mauritania. *Malar J.* 2014;13:26.
15. Soulama I, Nébié I, Ouédraogo A, Gansane A, Diarra A, Tiono AB, et al. *Plasmodium falciparum* genotypes diversity in symptomatic malaria of children living in an urban and a rural setting in Burkina Faso. *Malar J.* 2009;8:135.
16. Yavo W, Konaté A, Mawili-Mboumba DP, Kassi FK, Tshibola Mbuyi ML, Angora EK, et al. Genetic polymorphism of msp1 and msp2 in *Plasmodium falciparum* isolates from Côte d'Ivoire versus Gabon. *J Parasitol Res.* 2016;2016:3074803.
17. Khaminsou N, Kritpetcharat O, Daduang J, Charentanyarak L, Kritpetcharat P. Genetic analysis of the merozoite surface protein-1 block 2 allelic types in *Plasmodium falciparum* clinical isolates from Lao PDR. *Malar J.* 2011;10:371.
18. Bouyou-Akotet MK, M'Bondoukwé NP, Mawili-Mboumba DP. Genetic polymorphism of merozoite surface protein-1 in *Plasmodium falciparum* isolates from patients with mild to severe malaria in Libreville, Gabon. *Parasite.* 2015;22:12.
19. Messerli C, Hofmann NE, Beck H-P, Felger I. Critical evaluation of molecular monitoring in malaria drug efficacy trials and pitfalls of length-polymorphic markers. *Antimicrob Agents Chemother.* 2017;61:01500–16.

20. Greenhouse B, Myrick A, Dokomajilar C, Woo JM, Carlson EJ, Rosenthal PJ, et al. Validation of microsatellite markers for use in genotyping polyclonal *Plasmodium falciparum* infections. *Am J Trop Med Hyg.* 2006;75:836–42.
21. Hosseinzadeh-Colagar A, Haghghatnia MJ, Amiri Z, Mohadjerani M, Tafrihi M. Microsatellite (SSR) amplification by PCR usually led to polymorphic bands: evidence which shows replication slippage occurs in extend or nascent DNA strands. *Mol Biol Res Commun.* 2016;5:167–74.
22. Van Hong N, Delgado-Ratto C, Thanh PV, Van den Eede P, Guetens P, Binh NTH, et al. Population genetics of *Plasmodium vivax* in four rural communities in Central Vietnam. *PLoS Negl Trop Dis.* 2016;10:e0004434.
23. Papa Mze N, Ndiaye YD, Diedhiou CK, Rahamatou S, Dieye B, Daniels RF, et al. RDTs as a source of DNA to study *Plasmodium falciparum* drug resistance in isolates from Senegal and the Comoros Islands. *Malar J.* 2015;14:373.
24. Zhong D, Lo E, Wang X, Yewhalaw D, Zhou G, Atieli HE, et al. Multiplicity and molecular epidemiology of *Plasmodium vivax* and *Plasmodium falciparum* infections in East Africa. *Malar J.* 2018;17:185.
25. Juliano JJ, Porter K, Mwapasa V, Sem R, Rogers WO, Arie F, et al. Exposing malaria in-host diversity and estimating population diversity by capture-recapture using massively parallel pyrosequencing. *Proc Natl Acad Sci USA.* 2010;107:20138–43.
26. Koepfli C, Mueller I. Malaria epidemiology at the clone level. *Trends Parasitol.* 2017;33:974–85.
27. Lerch A, Koepfli C, Hofmann NE, Messerli C, Wilcox S, Kattenberg JH, et al. Development of amplicon deep sequencing markers and data analysis pipeline for genotyping multi-clonal malaria infections. *BMC Genom.* 2017;18:64.
28. Leclerc MC, Durand P, de Meeûs T, Robert V, Renaud F. Genetic diversity and population structure of *Plasmodium falciparum* isolates from Dakar, Senegal, investigated from microsatellite and antigen determinant loci. *Microb Infect.* 2002;4:685–92.
29. Ahouidi AD, Bei AK, Neafsey DE, Sarr O, Volkman S, Milner D, et al. Population genetic analysis of large sequence polymorphisms in *Plasmodium falciparum* blood-stage antigens. *Infect Genet Evol.* 2010;10:200–6.
30. Daniels R, Chang H-H, Séné PD, Park DC, Neafsey DE, Schaffner SF, et al. Genetic surveillance detects both clonal and epidemic transmission of malaria following enhanced intervention in Senegal. *PLoS One.* 2013;8:e60780.
31. WHO. Communicable diseases cluster: severe falciparum malaria. *Trans R Soc Trop Med Hyg.* 2000;94:0S1–9.
32. Plan stratégique national de lutte contre le paludisme au Sénégal. 2016–2020. [http://www.pnlp.sn/wp-content/uploads/2016/08/PNLP\\_PSN\\_VFF\\_03-02-2016.pdf](http://www.pnlp.sn/wp-content/uploads/2016/08/PNLP_PSN_VFF_03-02-2016.pdf).
33. Lucchi NW, Karell MA, Journal I, Rogier E, Goldman I, Ljolje D, et al. PET-PCR method for the molecular detection of malaria parasites in a national malaria surveillance study in Haiti, 2011. *Malar J.* 2014;13:462.

34. Lima GFM, de C, Lucchi, Silva-Flannery NW, Macedo-de-Oliveira L, Hristov A, Inoue AD. J, et al. Still searching for a suitable molecular test to detect hidden *Plasmodium* infection: a proposal for blood donor screening in Brazil. PLoS One. 2016;11:e0150391.
35. Snounou G, Zhu X, Siripoon N, Jarra W, Thaithong S, Brown KN, et al. Biased distribution of msp1 and msp2 allelic variants in *Plasmodium falciparum* populations in Thailand. Trans R Soc Trop Med Hyg. 1999;93:369–74.
36. Matranga CB, Andersen KG, Winnicki S, Busby M, Gladden AD, Tewhey R, et al. Enhanced methods for unbiased deep sequencing of Lassa and Ebola RNA viruses from clinical and biological samples. Genome Biol. 2014;15:519.
37. Park D, Tomkins-Tinch C, Ye S, et al. broadinstitute/viral-ngs: v1.19.2. Geneva: Zenodo, February 6, 2018 (<https://zenodo.org/record/1167849>).
38. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29:644–52.
39. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser. 1999;41:95–8.
40. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25:1451–2.
41. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989;123:585–95.
42. Weir BS, Cockerham CC. Estimating f-statistics for the analysis of population structure. Evolution. 1984;38:1358–70.
43. Agyeman-Budu A, Brown C, Adjei G, Adams M, Dosoo D, Dery D, et al. Trends in multiplicity of *Plasmodium falciparum* infections among asymptomatic residents in the middle belt of Ghana. Malar J. 2013;12:22.
44. Lê HG, Kang JM, Jun H, Lee J, Thái TL, Myint MK, et al. Changing pattern of the genetic diversities of *Plasmodium falciparum* merozoite surface protein–1 and merozoite surface protein–2 in Myanmar isolates. Malar J. 2019;18:241.
45. Aspeling–Jones H, Conway DJ. An expanded global inventory of allelic variation in the most extremely polymorphic region of *Plasmodium falciparum* merozoite surface protein 1 provided by short read sequence data. Malar J. 2018;17:345.
46. Ferreira M. Sequence diversity and evolution of the malaria vaccine candidate merozoite surface protein-1 (MSP-1) of *Plasmodium falciparum*. Gene. 2003;304:65–75.
47. Kang JM, Moon SU, Kim JY, Cho SH, Lin K, Sohn WM, et al. Genetic polymorphism of merozoite surface protein-1 and merozoite surface protein-2 in *Plasmodium falciparum* field isolates from Myanmar. Malar J. 2010;9:131.
48. Conway DJ, Roper C, Oduola AM, Arnot DE, Kremsner PG, Grobusch MP, et al. High recombination rate in natural populations of *Plasmodium falciparum*. Proc Natl Acad Sci USA. 1999;96:4506–11.

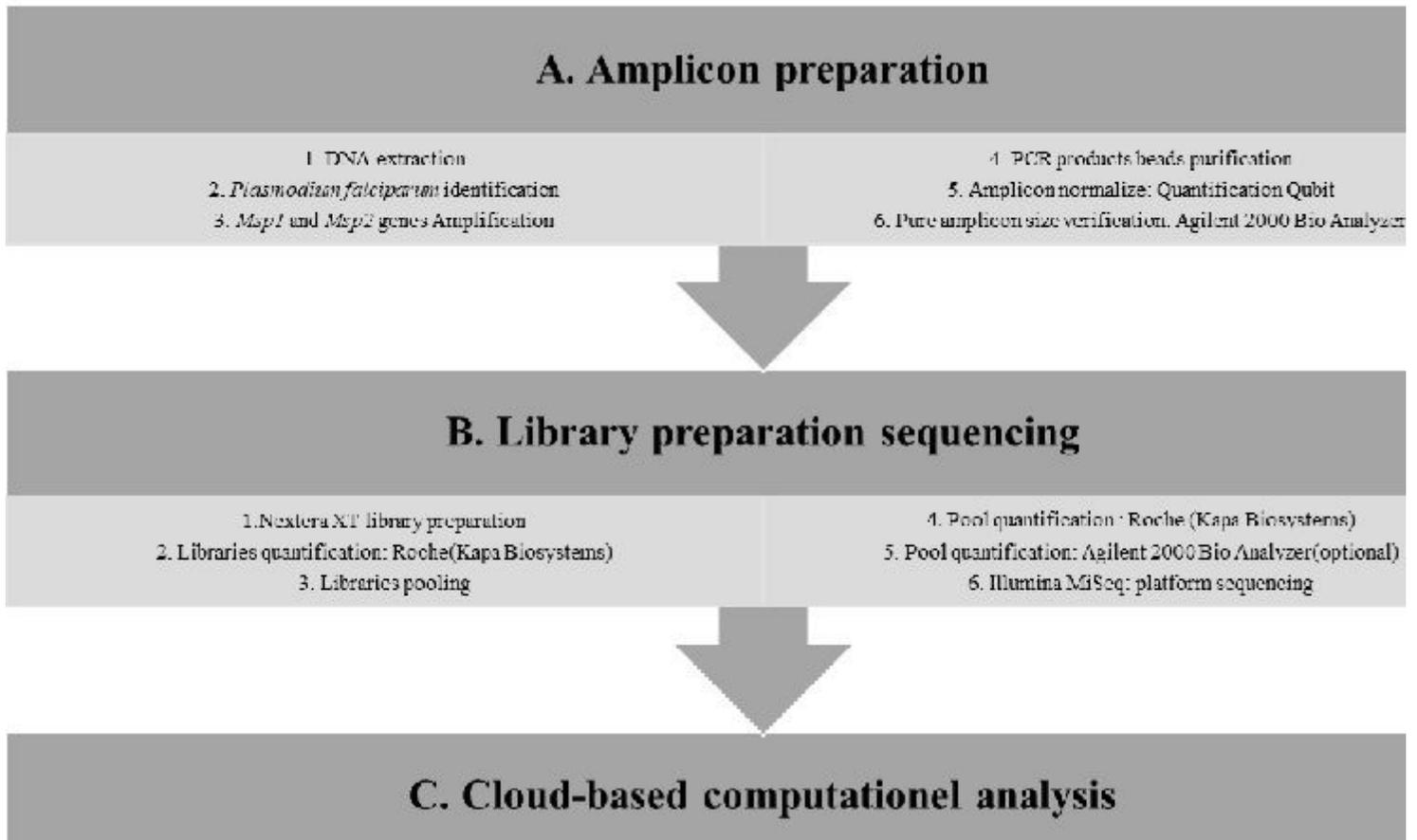
49. Ghansah A, Amenga-Etego L, Amambua-Ngwa A, Andagalu B, Apinjoh T, Bouyou-Akotet M, et al. Monitoring parasite diversity for malaria elimination in sub-Saharan Africa. *Science*. 2014;345:1297–8.
50. Osier FHA, Weedall GD, Verra F, Murungi L, Tetteh KKA, Bull P, et al. Allelic diversity and naturally acquired allele-specific antibody responses to *Plasmodium falciparum* apical membrane antigen 1 in Kenya. *Infect Immun*. 2010;78:4625–33.
51. 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*. *Parasit Vectors*. 2017;10:320.
52. Molina-Cruz A, Barillas-Mury C. The remarkable journey of adaptation of the *Plasmodium falciparum* malaria parasite to New World anopheline mosquitoes. *Mem Inst Oswaldo Cruz*. 2014;109:662–7.
53. Rono MK, Nyonda MA, Simam JJ, Ngoi JM, Mok S, Kortok MM, et al. Adaptation of *Plasmodium falciparum* to its transmission environment. *Nat Ecol Evol*. 2018;2:377–87.
54. Chenet SM, Schneider KA, Villegas L, Escalante AA. Local population structure of *Plasmodium*: impact on malaria control and elimination. *Malar J*. 2012;11:412.
55. Contamin H, Fandeur T, Rogier C, Bonnefoy S, Konate L, Trape JF, et al. Different genetic characteristics of *Plasmodium falciparum* isolates collected during successive clinical Malaria episodes in Senegalese children. *Am J Trop Med Hyg*. 1996;54:632–43.
56. Hastings IM, D'Alessandro U. Modelling a predictable disaster: the rise and spread of drug-resistant malaria. *Parasitol Today*. 2000;16:340–7.
57. Atroosh WM, Al-Mekhlafi HM, Mahdy MA, Saif-Ali R, Al-Mekhlafi AM, Surin J. Genetic diversity of *Plasmodium falciparum* isolates from Pahang, Malaysia based on MSP-1 and MSP-2 genes. *Parasit Vectors*. 2011;4:233.
58. Wanji S, Kengne-Ouafo AJ, Eyong EEJ, Kimbi HK, Tendongfor N, Ndamukong-Nyanga JL, et al. Genetic diversity of *Plasmodium falciparum* merozoite surface protein-1 block 2 in sites of contrasting altitudes and malaria endemicities in the Mount Cameroon region. *Am J Trop Med Hyg*. 2012;86:64–774.
59. Nkhoma SC, Nair S, Al-Saai S, Ashley E, McGready R, Phyo AP, et al. Population genetic correlates of declining transmission in a human pathogen. *Mol Ecol*. 2013;22:273–85.
60. Pumpaibool T, Arnathau C, Durand P, Kanchanakhan N, Siripoon N, Suegorn A, et al. Genetic diversity and population structure of *Plasmodium falciparum* in Thailand, a low transmission country. *Malar J*. 2009;8:155.
61. Noviyanti R, Coutrier F, Utami RAS, Trimarsanto H, Tirta YK, Trianty L, et al. Contrasting transmission dynamics of co-endemic *Plasmodium vivax* and *P. falciparum*: implications for malaria control and elimination. *PLoS Negl Trop Dis*. 2015;9:e0003739.

## Figures



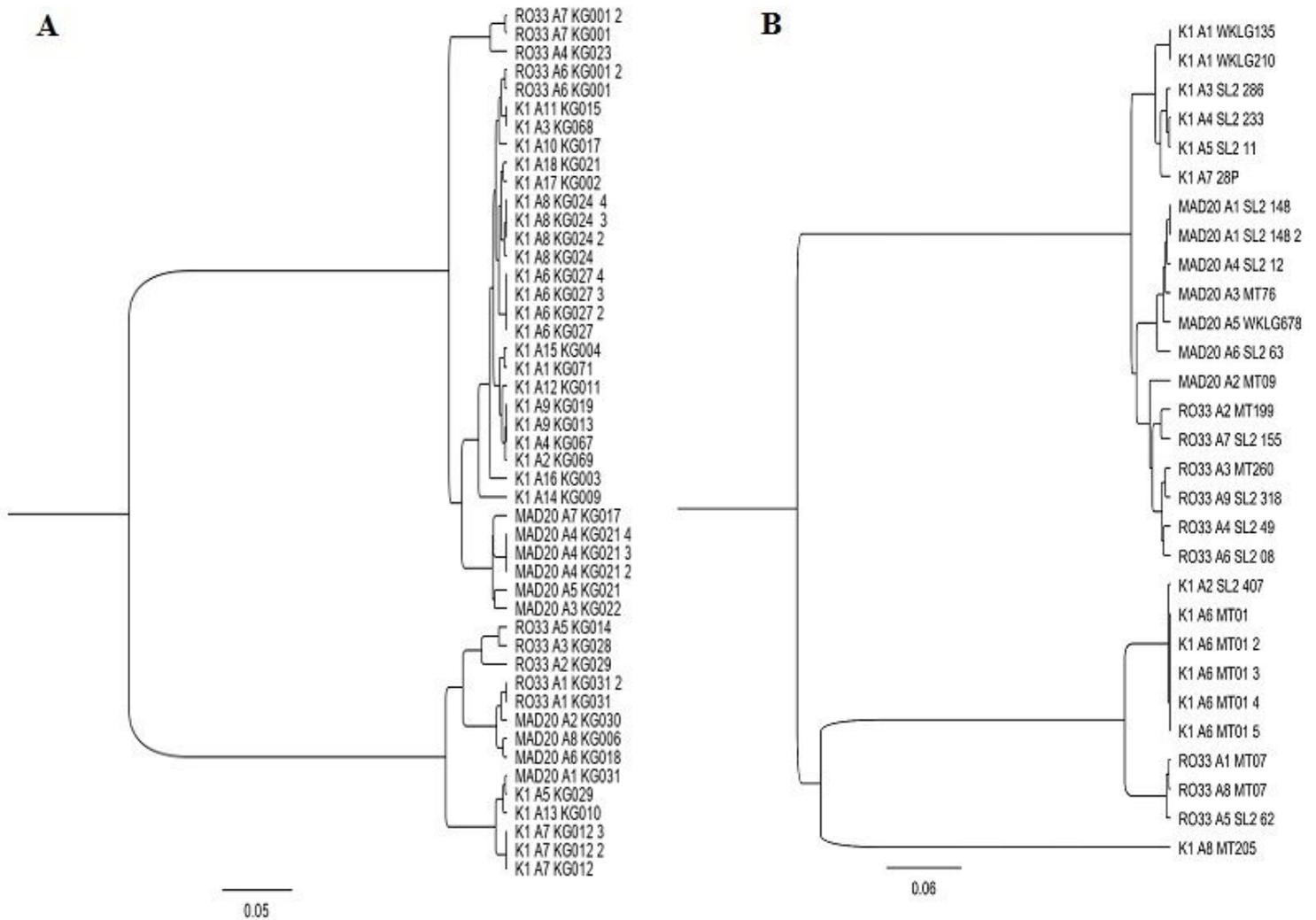
**Figure 1**

Map of Senegal showing the two malaria studies areas, Kedougou in orange (Southern Senegal) and Podor and Matam in blue (Northern Senegal). This map was generated using online website (<http://www.d-maps.com>)



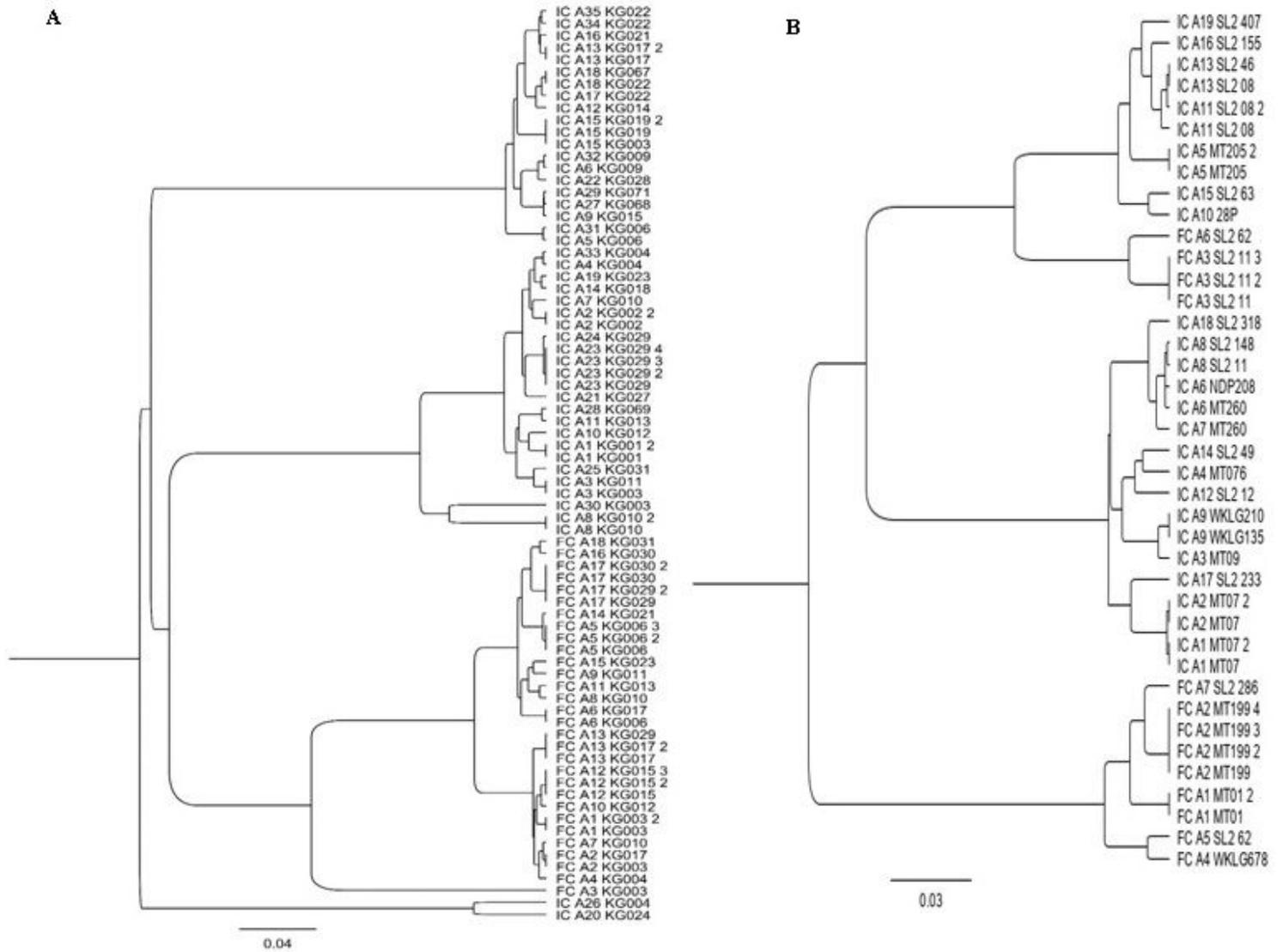
**Figure 2**

Experimental workflow from amplicon preparation to data analysis with three steps: amplicon preparation (A), library preparation and sequencing (B) and sequences data analysis (C).



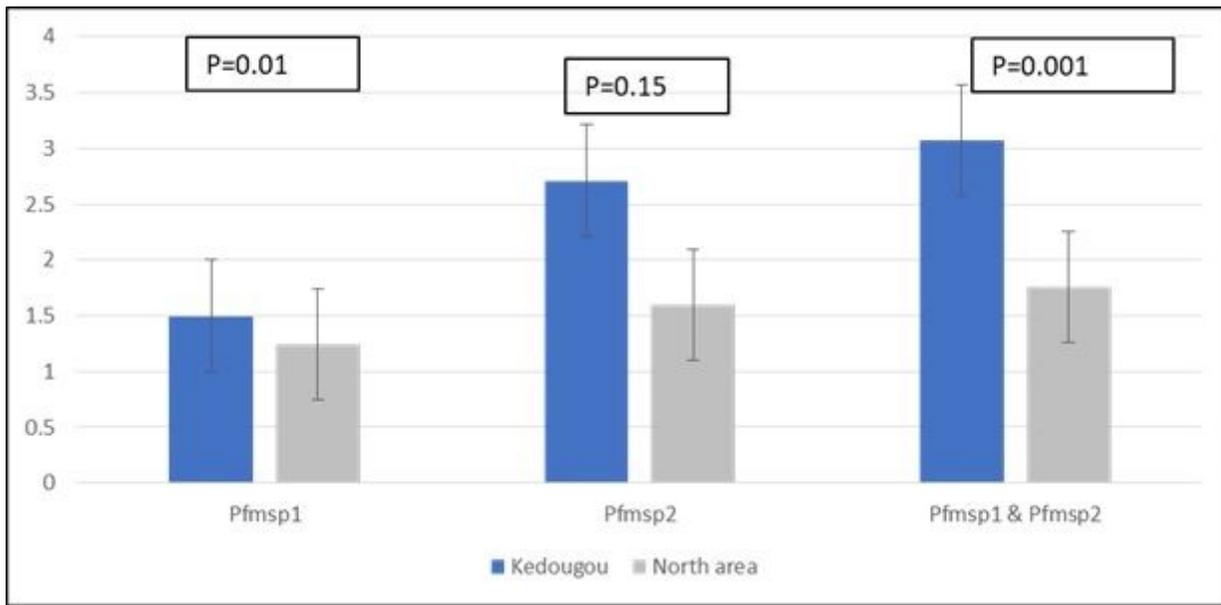
**Figure 3**

Neighbor-joining tree of K1, MAD20 and RO33 allele types of *msp1* gene of 53 *P.falciparum* isolates A: Neighbor-joining tree showing the genetic relatedness between *msp1* alleles in samples from the South. B: Neighbor-joining tree showing the genetic relatedness between *msp1* alleles in samples from the North.



**Figure 4**

Neighbor-joining tree of IC3D7 and FC27 allele types of *msp2* gene of 53 *P.falciparum* isolates A: Neighbor-joining tree showing the genetic relatedness between *msp2* alleles in samples from the South. B: Neighbor-joining tree showing the genetic relatedness between *msp2* alleles in samples from the North.



**Figure 5**

Plasmodium falciparum multiplicity of infections (MOI) mean for Pfmosp1 and Pfmosp2 genes in the South area and North area. The MOI mean of Pfmosp1&Pfmosp2was statistically different between the South and North (p=0.001).

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionnalfile2.xlsx](#)
- [Additionalfile1.xlsx](#)