ABSTRACT

Malaria is a disease that causes more than half a million deaths each year. Compounding the disease burden is high level of genetic variability of the causative parasites which are evasive to control efforts. Genetic diversity studies provide evidence of *Plasmodium falciparum* differentiation that affects fitness and adaptation to drugs and target antigens for vaccine development. This investigation describes the genome-wide analysis and diversity of *P. falciparum* in two southwestern states in Nigeria. Blood samples were collected from participants aged two years and above presenting with symptoms of malaria at health facilities in Lekki (LEK) and Badagry (BDG) in Lagos State and Aramoko (AMK) in Ekiti State, southwestern Nigeria. Microscopically confirmed P. falciparum-positive samples were subjected to further analysis. Following established protocols for parasite deoxyribonucleic acid (DNA) extraction and Polymerase Chain Reaction (PCR), molecular genotyping of Merozoite Surface Proteins (MSP) – 1 and 2 was carried out to assess the clonality and multiplicity of *P. falciparum* infections. Hemi-nested PCR of ten neutral microsatellite loci (Poly α, TA42, TA81, TA109, TA40, 2490, ARAII, PfG377, PfPk2, and TA60) was carried out to inquire into the extent of genetic diversity, linkage disequilibrium, population structure and inter-population differentiation. Genome-wide analysis of the parasite isolates was undertaken to survey for genes under signatures of directional and balancing selection. Bioinformatic analysis was done to call variants, filter and annotate raw sequence reads. Allele-frequency based neutrality test, Tajima's D; integrated haplotype score (iHS) and the ratio of synonymous to non-synonymous polymorphisms (π N /π S) were calculated to identify genes under selection. Of the 2,657 participants screened for uncomplicated malaria, 834 (31.4%) were confirmed *P. falciparum*- positive by microscopy. Three hundred isolates were randomly analysed for MSP1 and MSP2 (100 from each study area). The individual size ranges for the amplified DNA fragments were 160-225bp, 130-220bp, 160bp, 290-420bp, and 470-620bp for K1, MAD20, RO33, FC27 and 3D7 respectively. All the three families of MSP-1 (K1, MAD20 and RO33) and two of MSP-2 (FC27 and 3D7) were observed among the isolates in each population. The frequency of isolates with K1 family was 66.7% in the overall population (i.e., 70/100 in LEK, 50/100 in BDG and 80/100 in AMK). The frequency of isolates with MAD20 was 56.3% in the overall population. RO33 proportions were 50/100 in LEK, 40/100 in BDG and 84/100 in AMK. Dimorphic infections with both 3D7 and FC27 allele types were detected among the isolates. The frequency of samples possessing FC27 type (66.3%) was found to be higher than the samples with only 3D7 family (55%). The MSP-2 infections with both allelic types were identified in 40% of the parasite isolates. Highest and lowest mean Multiplicity of Infections (MOIs) with varying clones of *P. falciparum* was recorded in Aramoko and Badagry respectively although the difference across the populations was not significant ($P > 0.01$). Mean MOI was highest among participants between 2 and 4 years old across the three populations. Genetic diversity values were similar across all populations, with mean HE values across all loci between 0.65 (for LEK) and 0.79 (for AMK). Mann-Whitney U-test result showed no significant difference in the mean expected heterozygosity (H_E) values between LEK and BDG as well as BDG and AMK at $P > 0.01$. However, the difference in the H_E values between LEK and AMK was significant ($P < 0.01$). Although the mean number of genotypes detected per isolate was highest in AMK, Kruskal-Wallis test ($P > 0.01$) showed no substantial difference in the mean number of genotypes in the three parasite populations. Analysis of molecular variance (AMOVA) showed that genetic differentiation was low with $\Phi PT = 0.017$ ($P > 0.01$). Pairwise genetic distances between LEK and BDG, LEK and AMK and BDG and AMK parasite populations, calculated as Nei unbiased genetic distance (uD), were 0.164, 0.175 and 0.074 respectively. There was no observable relationship between genetic distance and the natural log of the geographical distance for each pair of parasite populations studied. Principal coordinates analysis (PCoA) showed two clusters of parasites not defined by the origins of individual population. Analysis of multilocus LD showed no significant index of association in all the parasite populations. Fourteen shared iHS regions that had at least 2 SNPs with a score > 2.5 were identified. These regions contained genes that were likely to have been under strong directional

selection. Two of such genes were chloroquine resistance transporter (CRT) located on chromosome 7 and multidrug resistance 1 (MDR1) located on chromosome 5. Although there was no evidence of recent directional selection in dihydropteroate synthase (DHPS) gene on chromosome 8, there was a weak signature of selection in dihydrofolate reductase (DHFR) on chromosome 4 and MDR5 genes on chromosome 13 with only 2 and 3 SNPs respectively identified within the iHS window. There was also a major selective sweep on chromosome 6 which had 32 SNPs within the shared iHS region. Tajimas's D values were mostly negative with a mean value of -0.86. One hundred and twelve genes (3.59%) had positive values. Tajima's D values of Circumsporozoite Protein (CSP), Erythrocyte-Binding Antigen (EBA-175), Merozoite Surface Proteins - MSP3 and MSP7, Merozoite Surface Protein Duffy Binding-Like (MSPDBL2) and Serine Repeat Antigen (SERA- 5) were 1.38, 1.29, 0.73, 0.84 and 0.21 respectively. This investigation has revealed a low-level of population sub-structuring of the parasites between Lagos and Ekiti States. The low clonal diversity and MOIs observed across all the parasite populations suggest the need to review the existing approach for categorizing recrudescent and re-infections with *P. falciparum* based on MSP 1 and 2 antigenic markers. The strong selection pressure observed in the markers of drug resistance may be attributable to continued chloroquine and sulfadoxine-pyrimethamine use despite their official proscription in the treatment of uncomplicated malaria; hence the re- introduction of the drug as was done in other endemic countries after a period of withdrawal may yet be inappropriate. High values of Tajima's D in some target genes of host immunity may suggest their potential as vaccine candidates.