Molecular Characterization of Genetic Diversity of
Plasmodium falciparum Associated with Imported
Cases of Malaria in Southeastern Region of
Saudi Arabia

Lorina I Badger-Emeka1,*, Abdulrahman A Al Sultan1, Ahlam A Al Qatari2, Abdullatif S Al Radhed3, Ahmed Al Sayyah2

1Department of Biomedical Sciences, College of Medicine, King Faisal University, Al-Ahsa, SAUDI ARABIA.
2Department of Pathology, College of Medicine, Imam Abdulrahman Bin Faisal University, Dammam, SAUDI ARABIA.
3Department of Microbiology, College of Medicine, Imam Abdulrahman Bin Faisal University, Dammam, SAUDI ARABIA.

ABSTRACT
Growing global international travel has reflected significantly on the number of imported cases of malaria into non endemic regions of the world inclusive of Saudi Arabia where transmission of the disease is significantly controlled. This success in control could be challenged by the influx of imported cases of the disease. This research looks at the genetic diversity in merozoite proteins 1 and 2 (msp 1 and msp 2) of Plasmodium falciparum gene in imported malaria cases. Blood samples collected by pathology department were preserved as dried blood spot on WhatmanTM 903TM Protein saver card. Extraction of genomic DNA was with Qiagen QIAamp DNA kit. Using primary and secondary nested PCR, Plasmodium species were classified with targeted primers. The nucleotides of K1, MAD20, RO33 msp 1 and 3D7, FC27 msp 2 of P. falciparum gene were sequenced by Macrogen (South Korea) and analyzed with BioEdit version 7 software. Frequencies of distribution of msp 1 and msp 2 allelic families varied, with statistically significant differences (p-value 0.001). Mono and poly clonal infections were also encountered. Clonal fragment sizes ranged between low (110 bp) to high (625 bp). Sequence analysis showed genetic diversity in imported malaria cases in the present study, indicating that current control measures might not be adequate to maintain the status quo.

Key words: Malaria, Imported, Plasmodium falciparum, Merozoite proteins, msp 1, msp 2, Genetic, Diversity.

INTRODUCTION
Malaria is described as the “undisputed queen of parasitic diseases”, and this was attributed to a number of reasons. At the eighth World health congress in 1955, malaria was the first human parasitic disease scheduled for global eradication. Today, the disease remains one of the most important parasitic diseases of man that has led to more deaths than those of all other combined parasitic diseases. The World malaria report placed estimated global cases at 228 million in 2018 with about 405,000 recorded deaths. In the same year, the World malaria report indicated that 61 indigenous cases of malaria were reported in Saudi Arabia while there were 2517 other reported cases of imported malaria. This shows the vulnerability of the Kingdom to imported cases malaria. For malaria, travel remains a risk factor with a recent literature citing that the increase in cases of malaria could be attributed to continuous long distance travels as well as the huge global migration movement. All these factors, are components that could be a lead in characterizing the epidemiology of imported malaria in non-endemic countries. Cross border importation between neighboring
countries and Saudi Arabia remains of great public health importance with sources of imported malaria grouped into three main types. The first of these sources are those of expatriate work force, while others include yearly visits from all geographical regions of the world for religious rites and national borders with disease endemic countries such as the national border with the republic of Yemen. All these sources constitute a public health problem in the control of malaria a view shared by other researchers.

Generally, the topography of Saudi Arabia is divided into groups based on incidences of transmitted cases of malaria. There is the non-malarious based, made up of the central, the north and eastern regions of the Kingdom that have terminated the transmission of the disease through effective vector control measures. There is also the low transmission areas in the Western parts while the high/medium transmission zones are in the southern and south-western parts of the Kingdom.

Thus, these regions in Saudi Arabia could be grouped from high risk (Jeddah, Mecca, Taif, Najran), to low risk or no risk. Dammam, Dhahran and Hofuf the region of the present study, are in the low or no risk region. However, due to the high yearly migration into the Kingdom, there will be the need for close monitoring. On the other hand, there are mosquitoes in these localities and they could probably, be potential vectors in the region. Subsequently, imported cases of malaria in the Eastern Province of Saudi Arabia should be of a major concern and be under continuous monitoring. Generally, molecular epidemiological studies of malaria can be used to study the genetic diversity of infections in relation to disease phenotypes. Generally, the identification of parasite molecular markers involved in resistance to antimalarial compounds are known to be of great interest in the monitoring of development and spread of resistance. Also, individuals infected with *Plasmodium falciparum* often consist of genetically distinct parasite populations, such as clones of same parasite species. Malaria parasite genotyping is an important tool in determining the multiplicity of infections [MOI] of *P. falciparum*. It had been documented that *P. falciparum* infections in patients could present as multiple parasite genotype variants, which could play a role in parasite evolution. However, the relationship between the numbers of *Plasmodium falciparum* genotypes, multiplicity of infection (MOI) continues to receive attention from researchers. Literature is silent on the genetic diversity of *Plasmodium* species associated with imported cases of malaria in this southeastern region of Saudi Arabia either for indigenous malaria infections or of visiting/migrant workers. The present investigation therefore aims at looking into the genetic diversity in imported cases of malaria in this southeastern region of Saudi Arabia. It is expected that the obtained results will be important information for the effective management and in keeping the disease under control in the kingdom in general. This is in view of the fact that Saudi Arabia has a dynamic workforce migration.

**MATERIALS AND METHODS**

**Study area and Ethical consideration**

The study was carried out in the south-eastern region of Saudi Arabia. No patients were directly involved in the study. Samples were those that had tested positive to *Plasmodium* parasites and stored in the pathology department of the college of medicine. The samples had originated from visitors to the Kingdom and formed part of routine healthcare checks, which is a pre-requisite for employment. Samples were collected between from January 2018 to January 2020.

**Molecular diagnosis of malaria parasites, collection and preparation of dried blood spot (DBS) samples**

Conventional microscopic examination of thick and thin blood smears as recommended by WHO were used in the hospital laboratories to confirm the presence of malaria parasite. Dried blood spots (DBS) samples were prepared according to the standard operating procedures on WhatmanTM 903TM Protein saver card [GE Healthcare UK] with 40–50 µL of blood introduced into each circle at the center point. One card was used per sample and given a code, date of sample collection, gender and age of patient. Collected blood spots were allowed to air-dry at room temperature and then kept in sealed labelled bags with silica desiccant and stored at -80°C until when they were examined. Earlier described protocol was used for the amplification of 18S ribosomal RNA through primary and secondary nested PCR using specific primers probes for *P. falciparum*, *P. malariae* and *P. vivax* were used shown in Table 1.

**Extraction of genomic DNA from DBS**

Isolation of genomic DNA was carried out with Qiagen QIAamp 96 Blood DNA extraction kit following manufacturers’ guidelines. With a paper puncher, 3mm in diameter punches were cut out from each DBS card and placed in 1.5 ml micro centrifuge tubes. To each tube was added 180 µL of ATL buffer and incubated at 85°C for 10 min, followed by the addition of 20 µL stock solution of proteinase K. The resultant mixture was mixed thoroughly by vortexing, incubated for 10 min at 70°C. To the resulting mixture 200 µL of 100% ethanol was added, mixed thoroughly by vortexing.
### Table 1: Primers used for nested PCR 18S ribosomal RNA amplification.

<table>
<thead>
<tr>
<th>PCR</th>
<th>Name of Primer</th>
<th>Primer sequence (5'-3')</th>
<th>PCR bp</th>
<th>Plasmodium species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary</td>
<td>rPLUf</td>
<td>TTTAAAATGGTTGGACATTTAAACGC</td>
<td>1.6 - 1.7</td>
<td>Plasmodium genus</td>
</tr>
<tr>
<td>Primary</td>
<td>rPLUr</td>
<td>CCTGTTGTGCTTAAACTTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd Nested</td>
<td>rFALf</td>
<td>TTTAAAATTGTGCTTTGAGTTAAAATATT</td>
<td>206</td>
<td>P. falciparum</td>
</tr>
<tr>
<td>2nd Nested</td>
<td>rFALr</td>
<td>ACAAAATGGAACCTCAATTGACTGACTGTC</td>
<td>145</td>
<td>P. malariae</td>
</tr>
<tr>
<td>2nd Nested</td>
<td>rMALf</td>
<td>AACTACATGTGTTGTAAGTTAACACTGCG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd Nested</td>
<td>rMALr</td>
<td>AAAATTCCCATGAACTAAAAATTTATCAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd Nested</td>
<td>rVIVf</td>
<td>CTTTCTAGTTAATCCACATATGATAC</td>
<td>121</td>
<td>P. vivax</td>
</tr>
<tr>
<td>2nd Nested</td>
<td>rVIVr</td>
<td>ACTTCAAGCCGAGCAAAGACTCTTGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd Nested</td>
<td>rOVAf</td>
<td>ATCTTATTTGCTATTGTTGAAGA</td>
<td>226</td>
<td>P. ovale</td>
</tr>
<tr>
<td>2nd Nested</td>
<td>rOVAR</td>
<td>ATCTTATTTGCTATTGTTGAAGA</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 2: Primers sequences used for genotyping by PCR for msp 1 and msp 2

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Round of PCR</th>
<th>Sequences (5'-3')</th>
<th>Allelic family</th>
<th>Annealing temp.</th>
<th>control</th>
<th>PCR Cycling parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>msp 1</td>
<td>External</td>
<td>CTAGAAGCTTTAGAAGATGCAGATTG-F</td>
<td>61°C</td>
<td>K1</td>
<td>3D7</td>
<td>Initial denaturation 5 minutes at 94 °C; PCR: Denaturation 94 °C for 1 min, annealing 61 °C for 45 secs; Extension 72 °C for 90 secs. (40 circles each). Final extension 72 °C for 5 min. Hold 4°C.</td>
</tr>
<tr>
<td>msp 1</td>
<td>Internal</td>
<td>CTTAAATAGTATTCTAATTCAAGTTGGAATCA-R</td>
<td></td>
<td>AD20</td>
<td>HB3</td>
<td>Initial denaturation 2 minutes at 94 °C; PCR: Denaturation 94 °C for 1 min, annealing 61 °C for 45 secs; Extension 72 °C for 90 secs. (30 circles each). Final extension 72 °C for 10 min. Hold 4°C.</td>
</tr>
<tr>
<td>msp 2</td>
<td>External</td>
<td>ATGAAGGTAAATTAAAACATTGTCTATTATA-F</td>
<td>55°C</td>
<td>IC/3D7</td>
<td>3D7</td>
<td>Initial denaturation: 94 °C for 5 min; PCR: 40 cycles of 94 °C for 1.5 min, 55 °C for 45 s, 72 °C for 1.5 min; final elongation: 72 °C for 10 min. Hold 4°C.</td>
</tr>
<tr>
<td>msp 2</td>
<td>Internal</td>
<td>ATATGGCAAAGATAAAAAACAGTGTGCTG-R</td>
<td></td>
<td>FC27</td>
<td>HB3</td>
<td>Initial denaturation: 94 °C for 2 min; PCR: 30 cycles of 94 °C for 1.5 min, 55 °C for 45 s, 72 °C for 1.5 min; final elongation: 72 °C for 10 min. Hold 4°C.</td>
</tr>
</tbody>
</table>
before centrifuging briefly. The complete process of parasite DNA extraction was carried out according to the detail protocol in the handbook of Qiagen-DNA purification from DBS.

**Plasmodium parasite genotyping**

Distinct subpopulations present in *P. falciparum* parasites are generally characterised by genotyping and this was carried out through nested PCR of extracted DNA. Thus, for their extensive polymorphism in size and sequences, genetic markers of merozoite surface protein (*msp*) 1 and 2 were chosen for the investigation. The allelic types for *msp* 1, K1, MAD3, and RO33, those of *msp* 2, FC27 and IC/3D7 were investigated with specific primers in the secondary nested PCR. A combination of methods described by Biodefence and emerging infections research resources repository (BEI) methods for Malaria research as well as the protocol for *msp* 2 genotyping by University of Maryland School of Medicine, malaria group16 version were used for parasite genotyping. With the first amplification reaction, oligonucleotide primers shown in Table 2 were used for the first nested PCR. The resultant product from the first amplified reaction served as DNA template for the second nested PCR process using specific-sequenced primers for *msp* 1 and *msp* 2 as described previously.15,17 The final volume of the primary and secondary PCR constituted products was 25 µL each. This was made up of the following, sterile water 17.4 µL; 2.5 µl of 10x buffer; 0.2 µL of dNTP (25mM); 0.8 µL of MgCl₂ (50mM); 0.5 µL each (10 µM) of forward and reverse primers; 0.2 µL of Taq polymerase and 3.0 template DNA. Thus, tubes targeting the allelic families of both *msp* 1 and *msp* 2 were set up. Bio-Rad thermal cycler was used for PCR amplifications with 40 and 30 cycles for primary and secondary reactions respectively as shown in Table 2.

**Detection of *msp* 1 and *msp* 2 alleles by Gel electrophoresis**

The resultant secondary PCR product were analysed on 1% agarose gel by electrophoresis. 4 µL of amplified DNA products were stained with Ethidium bromide and documented with an electronic photographic documentation system. Amplicon sizes were detected as previously described17 using a 100bp DNA ladder (Promega USA). DNA positive controls from BEI resources (www.beiresource.org) were used for the comparison of the DNA fragments of *Plasmodium* parasite species with molecular weights applied in characterising *msp* 1 and *msp* 2 alleles.

**Determination of multiplicity of infection (MOI)**

MOI or the complexity of infection for *msp* 1 and *msp* 2 was determined for positive samples according to set guidelines.17,18 The number of genotypes detected per infection was determined by dividing the total number of alleles individually for *msp* 1 and *msp* 2 by the total number of samples that were positive to the markers of the genes. Isolates were grouped as either monoclonal or polyclonal infections if they carried a single allele or more than one genotype respectively.19

**Sequencing of *msp* 1 and *msp* 2 *P. falciparum* gene and sequence analysis**

DNA PCR products of K1, MAD3, RO33 *msp* 1 alleles and 3D7 and FC27 *msp* 2 allele of *P. falciparum* were constituted with the specific primers and submitted to Macrogen (South Korea) in individual 1.5 µL micro-centrifuge tubes for PCR-clean up and bidirectional sequencing. Pre-sequencing sample preparation was according to the Macrogen-sample guide. Resultant sequences were analysed using available online tools at National Center for Biotechnology information (NCBI) https://www.ncbi.nlm.nih.gov/nucleotide/. Also, BLAST (Basic local alignment search tool) https://blast.ncbi.nlm.nih.gov/Blast.cgi was used in finding regions of similarities in nucleotide/protein sequences as well as statistical analysis related to the sequences. Also, BioEdit version 7 was use for entropy Hx plot of amino acids sequence analysis.

**Statistical analysis**

GraphPad Prism version 8 was used for statistical analysis. Data is presented as percentages while the one sample t-test was used to compare statistical differences between monoclonal and polyclonal infections. Also, two sample t-test was used to compare differences in number of clones of *msp* 1 and *msp* 2 malaria infections with significant difference taken at *p*<0.05. Sequences analyses were carried out with BioEdit version 7. Also, entropy plots were obtained using BioEdit version 7 software. Phylogenetic trees for obtained sequences were inferred using the Neighbor-Joining method with all trees drawn to scale. Positions containing gaps and missing data were eliminated. While evolutionary analyses were conducted in MEGA5.

**RESULTS**

**Demographic description of collected samples**

Fifty-four (54) positive malaria samples were collected for over a period of two years (Figure 1A). Samples were
from males (65%) and females (35%). There was 9% of cases imported from the South-western part of the Kingdom, to the region of the study. The highest number of samples (35.2%) were from the east African country of Sudan. Others included travelers from the West African country of Nigeria (30%). While the remaining were from Pakistan (18.5%), India and Uganda, which constituted of 3% each of the total number of samples (Figure 1B).

Results showed age groups ranging from 3 to 70 years of age (Figure 1C). The highest samples were from the 30 – 39 age group. Others within age groups of 20-29 and 3 – 9 years age groups were more than 40 years old and above.

**Encountered species of Plasmodium parasite**

Using nested PCR (nPCR), with specific primers targeting 18S rRNA gene, malaria parasite was encountered in 51 samples and were grouped species of *P. falciparum* (*n* = 47; 92.2%) and *P. vivax* (*n* = 4; 7.8%). Other *Plasmodium* species, *P. ovale* and *P. malariae* were not detected. In terms of species by country of importation, results showed *P. falciparum* to be the most and was encountered in samples from all the countries with the exception of India which had only *P. vivax*. Samples of *P. falciparum* parasite were genotyped by nested PCR for *msp 1* and *msp 2* for the investigation of molecular analysis of clonal distribution of the infections.

**Encountered *msp 1* and *msp 2* allelic families of imported cases of malaria**

PCR fragments amplification obtained were used in grouping the allelic families of *msp 1* and *msp 2*. Results showed alleles for *msp 1* to be more frequently encountered (82.35%) in the infections than those of *msp 2* (60.8%). The allelic families of *msp 1* included MAD20, KI, and RO33 all of which had been detected at varying degrees (Figure 1D) with RO33 being predominantly more as compared to MAD20 and KI. For *msp 2*, the detected allelic families were 3D7 and FC27 of the PCR amplified samples. In this group, the FC27 allelic family was predominant (74.2% *n* = 23/31) while 3D7 allelic family was the least (58.1% *n* = 18/31) (Figure 1E). *P. falciparum* infections were grouped either as monoclonal or polyclonal. There were more monoclonal than polyclonal infections involving both allelic families of *msp 1* and *msp 2* (Table 3). A t-test comparison showed differences in percentages to be statistically significant with *p*-value of 0.001. In *msp 1*, more than half (52.38%) of the samples were monoclonal infections (Table 3). However, the differences between them and the polyclonal infections was not statistically significant (*p*-value = 0.74 at 41d.f). Also, in the *msp 2*, alleles, there were more (70.97%) monoclonal than polyclonal infections involving both allelic families of *msp 1* and *msp 2* (Table 3). A t-test comparison showed differences in percentages to be statistically significant with *p*-value of 0.015 at 30 d.f (Table 3). Trimorphic infections which was classified as samples that had 3 allelic families in *msp 1* alleles constituting of MAD20/KI/RO33 was encountered in some samples (7.14%). The polyclonal infections constituting of MAD20/RO33 were more

<table>
<thead>
<tr>
<th>Table 3: Multiplicity of infection associated with <em>msp 1</em> and <em>msp 2</em> genes of <em>P. falciparum</em> isolates of imported malaria.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gene</strong></td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td><em>Msp 1</em></td>
</tr>
<tr>
<td><em>Msp 2</em></td>
</tr>
</tbody>
</table>

MOI = Multiplicity of infection. * = significant *p* < 0.05
Badger-Emeka, et al.: Molecular Diversity in Imported Cases of Malaria

(23.81%) than those of MAD20/KI and KI/RO33 (Figure 1D). However, in the msp 2 alleles, polyclonal infections constituting of 3D7/FC27 were few (22.6%) in number (Figure 1E). The multiplicity of infection (MOI) for msp 1 was seen to be higher (1.6) than that of msp 2 (1.32) as shown in Table 3.

Associated frequencies of alleles for msp 1 and msp 2 P. falciparum gene

There were nine observed alleles in msp 1 based on the results from PCR fragment amplification. For MAD20, fragment sizes ranged between 200 - 225 bp, in KI fragment sizes ranged between 115 - 225 bp while only one fragment size of 150 bp was associated with RO33.
Details of percentage distribution of clone sizes among the samples is shown in Figure 1F. The majority of \( msp \, 1 \) were the 150 bp RO33 allele and the 225 bp clone size of MAD20 allele. For \( msp \, 2 \), fifteen clones were encountered eleven of which were detected in the 3D7 allelic family with fragment sizes ranging between 175 – 375 bp (Figure 1G). Majority of the \( msp \, 2 \) alleles were 3D7 of which the dominate fragment size was 175 bp. There were 5 clones seen in FC27 allele with fragment sizes that ranged between 325 – 625 bp. In this group, clones with 405 bp fragment sizes were dominant (Figure 1G).

### Diversity in \( P. \, falciparum \) nucleotide sequences of imported Malaria cases

The genetic similarities of alleles of \( msp \, 1 \) and 2 with those earlier reported in malaria endemic countries globally, were retrieved from NCBI databases. The GeneBank ascensions numbers https://www.ncbi.nlm.nih.gov/nucleotide of earlier reported sequences and region of isolation are presented in Table 4. For \( msp \, 1 \) allelic families, K1, MAD20, RO33 haplotypes were found to be similar to those reported from other regions in countries that included Ghana, Central sub-Saharan African, Tanzania, Senegal, Malawi, Solomon...
Badger-Emeka, et al.: Molecular Diversity in Imported Cases of Malaria

Islands, Thailand Kanchanaburi Province, Vietnam, Philippines, Western Brazilian Amazon, Papua New Guinea and various regions of India (Table 4). For 3D7 and FC27 alleles of \textit{msp} 2 \textit{P. falciparum}, some of the haplotypes are similar to those retrieved from NCBI as follows: AJ318754.2 (Papua New Guinea); JX885948.1 (Thailand); AE217036.1 (Ghana); JX885948.1, JX885940.1; JX885946.1 (Thailand); JX283523.1 (India); KC887547.1 (Gabon). Others included LN999943.1 of 3D7 chromosome 2, DQ166542.1 isolate wos58 (\textit{msp} 2) gene, and AY378310.1 \textit{msp} 2-O allele, partial cds (Table 4).

Multiple sequence alignment (MSA) and phylogenetic analysis for \textit{msp} 1 alleles of \textit{P. falciparum} gene

Three alleles, KI, MAD$_{20}$ and RO33 were sequenced and analysed for \textit{msp} 1 \textit{P. falciparum} gene (SS 3 A – C). In KI allelic family, four DNA nucleotide sequences of 114 positions, which constituted 15\% of the original 736 positions, were analysed. No gap positions were allowed with the obtained flanked positions of 8 selected blocks, namely 75-84, 137-146, 264-279, 381-394, 533-561, 576-587, 609-618 and 697-709. In addition, the obtained nucleotides frequencies for the KI alleles were as follows: Adenine (A) 0.36842, Cytosine (C) 0.16667, Guanine (G) 0.17763 and Thymine (T) 0.28728. For MAD$_{20}$ allele of \textit{msp} 1 \textit{P. falciparum} gene, six nucleotide sequences were analysed of which positions with gaps and missing data were eliminated (Figure 2A). A final data set of 96 positions were obtained with flank positions of three selected blocks 45-103, 123-140 and 153-191.

Obtained nucleotide frequencies were as follows: Adenine (A) 0.20690, Cytosine (C) 0.17672, Guanine (G) 0.28233 and Thymine (T) 0.33405. An optimal tree with 0.00313072 branch length is shown in Figure 2B with associated samples clustered together in bootstrap test indicated while evolutionary analysis was carried out using MEGA version 4.

Multiple sequence alignment analysis for RO33 allelic family for \textit{msp} 1 \textit{Plasmodium falciparum} gene (Figure 2C) was of fourteen nucleotide sequences analysed using MEGA.

Figure 4: Amino acid sequence alignments of \textit{Plasmodium falciparum}. MAD$_{20}$ (A) Distinct variants were found in the sequence analysis of \textit{msp} 1 MAD$_{20}$ alleles as follows: 28 from BJ MAD$_{20}$; 27 from BA MAD$_{20}$; 19 from BM MAD$_{20}$ and 3 from BV MAD$_{20}$. In RO33 of \textit{msp} 1 amino acid sequence (B), distinct variants were found by sequence analysis as follows: 15 from CC RO33; 10 from CW RO33; 9 from CM RO33; 5 from CR RO33 msp 1; 2 from CV RO33 and 1 each from CV RO33; CS RO33 and CT RO33. 7C - D: shows amino acid sequence alignment for \textit{msp} 2 3D7 and FC27. Distinct variants were found by sequence analysis of \textit{msp} 2 3D7 (C). The highest was 71 from DA 3D7; followed with 42 from DY 3D7; 11 and 10 from DK 3D7 and DJ 3D7 respectively. However, sequence alignments of amino acid of \textit{msp} 2 FC27 distinct allelic variants of \textit{Plasmodium falciparum} isolates of imported cases of malaria (D) exhibited minor variants. Variations in sequence analysis were as a result of deletions represented as dashes. Identical residues have been highlighted in colors.

Supplementary Figure 1: Entropy (HX) plot for \textit{Plasmodium falciparum} nucleotides (A-D). MAD20 (A) and RO33 (B) msp 1 amino acid nucleotide of \textit{Plasmodium falciparum} of imported malaria case. 688 Also, Entropy [Hx] plot of amino acid sequence alignment for \textit{msp} 2 3D7 and FC27. Distinct variants were found by sequence analysis of \textit{msp} 2 3D7 (C). The highest was 71 from DA 3D7; followed with 42 from DY 3D7; 11 and 10 from DK 3D7 and DJ 3D7 respectively. However, sequence alignments of amino acid of \textit{msp} 2 FC27 distinct allelic variants of \textit{Plasmodium falciparum} isolates of imported cases of malaria (D) exhibited minor variants. Variations in sequence analysis were as a result of deletions represented as dashes. Identical residues have been highlighted in colors.
version 4 with a total of 114 positions in the final data set.\textsuperscript{37} The obtained nucleotide frequencies for this allelic family were A (0.31923), C (0.17179), G (0.24744) and T (0.26154). The optimal tree with the sum branch length of 0.01764909 was inferred using the neighbour-joining method previously described.\textsuperscript{38} Associated taxa clustered together of replicate trees are indicated according to the confidence limits on phylogenies by Felsenstein.\textsuperscript{39} The tree drawn to scale is shown in Figure 2D.

**Multiple sequence alignment (MSA) and phylogenetic analysis for msp 2 allelic families of *P. falciparum* gene**

MSA for *msp 2* *P. falciparum* gene involved the 3D7 and FC27 allelic families. For 3D7 allele of *msp 2* analysis involved six nucleotide sequences (Figure 3A) with all gaps or missing data eliminated. Flanked positions of eight selected blocks 1-24, 30-98, 107-119, 127-148, 151-187, 225-241, 272-285, 288-304 were obtained. There were a total of 222 positions of which nucleotide frequencies were as follows: (A) 0.25822, (C) 0.21596, (G) 0.24178 and (T) 0.28404. The neighbouring joining method\textsuperscript{40} was used for inferring the phylogenetic trees of 3D7 and FC27 allelic families of the *msp 2* gene (Figures 3B, 3C). Three nucleotide sequences were analysed with 320 positions in the final data set for the FC27 allele of *msp 2* *P. falciparum* gene (SS 4B). Flanked positions of four selected blocks 79-95, 102-120, 126-148, and 160-207 were obtained with nucleotide frequencies of A (0.31923), C (0.17179), G (0.24744), and T (0.26154).

**Entropy plots for allelic families of *msp 1* and *msp 2* *P. falciparum* gene**

Entropy plots used to measure the amount of variability in the amino acids for merozoite proteins 1 and 2 are shown in Supplementary Figure 1. Entropy Hx plots were obtained with BioEdit version 7 with values defined as described earlier.\textsuperscript{40} For *msp 1*, entropy plots showed minimal entropy with most positions barely getting to the scale of one (Figure S1 A, B). Similar observation was seen with *msp 2* 3D7 and FC27 amino acids entropy plot. However, the entropy [Hx] for FC27-*msp 2*, is seen to be below the scale of one being signs of high similarities with lower entropy implying the most conserved regions (Figure S1 C, D).

**Genetic variations in amino acid of alleles of *msp 1* and *msp 2* *P. falciparum***

Tripeptide repeat units varied in the sequences of *msp 1* and *msp 2* as shown in Figure 4 A-D. In the MAD\textsubscript{30} allelic type, the tripeptide alleles starts with SGG, and ends with two tripeptide units (SGGAV/DNSRRT). The different tripeptide repeat units are, SGG, SGT, SVA, SGN, SRR, and TNP. The allelic diversity as seen in this group (Figure 4A), can be attributed to substitutions and deletions in these motifs. Allelic diversity in RO33-*msp 1* amino acid sequence showed that for this allelic type, the tripeptide region started with NTQ and ended with AAN. Diversity in this allelic type can therefore be attributed to both deletions and substitution in repeat motif in the (NTQVVKAPADAVSTQSAK-NPP/GATVPSGTASTKGAIRSPGAAN) sequence as shown (Figure 4B). For *msp 2* 3D7 allelic variants of imported *P. falciparum* malaria cases, tripeptide repeat units starts with four tripeptide units, AES, KPS, TGA and GAS (Figure 4C). Genetic variations are attributed to duplications, deletions as well as substitutions in the repeat motifs. Sequence alignments of amino acid sequences of *msp 2*-FC27 distinct allelic variants among *Plasmodium falciparum* isolates showed no distinct variants by sequence analysis (Figure 4D).

**DISCUSSION**

The genetic diversity of imported malaria parasites impacts malaria control strategies and hence portends danger of recrudescence.\textsuperscript{41} From the foregoing, the identification of population structure of migrants bringing in malarial parasites associated endemity, should be a public health objective. In the present study, we examined the genetic diversity of imported *P. falciparum*. We also studied their clonal assortments in terms of associated population structure of these imported cases. Our study comes on the wake of the fact that Saudi Arabia has a large dynamic migrant workforce therefore, imported malaria could underscore their control measures. Demographic results showed infections were from children and adults with history of travel into and within the Kingdom, calls for the need for a continuation in stringent control measures if there is to be a complete eradication of imported malaria cases. However the small sample size collected over a period of two years could be possible pointers to attained success so far in this region of study that will need to be maintained if Saudi Arabia is to retain its status as a “country on the verge of malaria elimination”.\textsuperscript{42,43} On the other hand, based on WHO (2008) recommendation\textsuperscript{44} of *msp 1* and *msp 2* being standard markers in distinguishing recrudescence from newly infecting parasite of malaria, the present investigation evaluated genetic diversity in these allelic families of *P. falciparum* gene. The obtained results which showed genotypes of *msp 1
predominating those of msp 2 with differing genetic polymorphism, are similar to those reported from disease endemic countries.\textsuperscript{17,45}

The study also showed that RO33 were encountered in all the samples thus predominating MAD\textsubscript{20} and KI in the allelic families of the msp 1 P. falciparum gene. These findings are similar to those of earlier reports.\textsuperscript{46-48}

However, contrary to this are other reports in which KI and MAD\textsubscript{20} msp 1 were found to predominate RO33.\textsuperscript{17}

Thus, there are differences in the predominance of alleles which could be attributed to a number of suggested reasons. One of such is that KI allele dominance was associated with severe malaria,\textsuperscript{17} while others associated this dominance with asymptomatic malaria.\textsuperscript{49,50} The present study did not look into the severity or otherwise of symptoms of malaria in the patients from where the samples had been taken. Neither could the infections be said to be asymptomatic. Predominance of RO33 could possible simply be a reflection of that obtainable from the origin of travel into Saudi Arabia. In the msp 2 alleles, FC27 predominated 3D7 in this study with the findings similar to those of earlier reports in disease endemic regions.\textsuperscript{51-54} However, while 3D7 allele was predominant in countries like Pakistan, Papua New Guinea, Thailand,\textsuperscript{55-57} FC27 was predominant in reports from countries such as Nigeria, Cameroon and Gabon.\textsuperscript{57,59}

Thus predominance of FC27 in this study could again be a reflection of the origin of travel into the Kingdom. Considering the fact that majority of the imported cases here were from African countries of Sudan and Nigeria, there is therefore the possibility that predominance in alleles of either msp 1 or msp 2, could be attributed to geographical regions of travel origin. On the other hand, FC27-msp 2 allele had been generally associated with severity of malaria while it is thought that the 3D7-msp 2 is common in areas where transmission of malaria is high. Thus, while there is no such transmission in this region of Southeast Saudi Arabia, results in this report again could be reflecting region of travel of the imported cases. Some authors\textsuperscript{60} did however suggest that dominance of some alleles could either be accidental or could be due to a stable life cycle of P. falciparum parasite. There is the possibility that low P. falciparum number of clones seen here are as a result of a restricted pool of genes as earlier suggested.\textsuperscript{40,61} However, there were diversities generally in the genotypes of msp 1 and msp 2 P. falciparum gene in this investigation. By PCR agarose visualisation (supplementary data Figure 1, 2 and 3) diversity seen among the clones of alleles of msp 1 (9) and msp 2 (11) cannot be described as high when compared with those seen in other reports.\textsuperscript{45,49}

High diversity those reports had attributed to the intensity of malaria transmission as well as exposures to multiple parasite infections carrying different genotypes. This could explain the reason for the results here of the imported malaria samples with similar findings having been reported previously.\textsuperscript{62-64} There is also the possibility of treatment with antimalarial before travelling into the Kingdom and this could reflect in the results. Also, MOI of the imported cases of malaria in this region of study was low when compared with those of other findings.\textsuperscript{65} This is a positive direction with a possibility of reduced risk for parasite transmission. On the other hand, similar to our findings are those of Somé\textit{et al.,}\textsuperscript{17} in which low MOI had been attributed to possible success of malaria control measures. Others had associated high values of MOI with hyper endemic regions where there was high transmission of the disease.\textsuperscript{55,66} However, the samples here are imported into a southeast region of the Kingdom with no active parasite transmission. Also worthy of note are the polyclonal samples seen in this study. Such infections are reportedly common in high endemic regions.\textsuperscript{64,67} The results here, are therefore not unexpected when considering the origins of the samples. Similarly, nucleotide sequences of msp 1 and msp 2 allelic families found here are similar to those found in other regions of the world. Phylogenetic tree analysis of sequenced KI msp 1 grouped this allele into four categories. Three haplotypes of which two (AB502545, AB502526), were 95 and 97 % in similarity as those reported in Ghana\textsuperscript{20} while the other (AB827755) was 98 % similar to that reported in Thailand.\textsuperscript{21} Also, the KI-msp 1 tripeptide motifs (SAQ, SGA, SG, SG) which were seen in combination with tripeptides (SGASAQ, SGTSAG, SAQSGA) reported in this study are similar to other documented findings.\textsuperscript{68} Thus, this pattern of tripeptide motifs and their combinations here could be associated with African allele trend as earlier suggested.\textsuperscript{68} For MAD\textsubscript{20}-msp 1, genetic diversity was detected in the number of variants obtained from analysed sequences of this allele. Four variants (HM568580, KR063228, and HM568568) seen here, are between 94 – 99% similar to those in India.\textsuperscript{25,27} One variant (AB502745) with 100 % similarity had been reported in Solomon Island,\textsuperscript{29} while others variants with 99 % similarity, in Tanzania (AB502487) Central Africa (AB502597), Senegal (EU032224), and Ghana (AB502528, AB502626).\textsuperscript{20,24} There were also other variants from Vietnam (AF509654, AF509696),\textsuperscript{42} Thailand (AB8227752),\textsuperscript{21} Western Brazilian Amazon (XJ412319)\textsuperscript{27} and Papua New Guinea (AB502678, AB502637).\textsuperscript{20} Worthy of note is that some of the listed countries are not among the listed for imported cases of malaria to the region of the present study and could
only be explained to be due to human movement to different regions of the world. Also, the phylogenetic of analysed sequence of RO33 showed four categories variants of which had been reported by other researchers in India (KR063231), JF460903, DQ485450, JX283499 and Tanzania (AB502443). For 3D7-ms, two, there were nine variants in imported malaria case in this study which are similar to those from Thailand (JX885948, JX885946, JX885940), Papua New Guinea (AJ318754) and Ghana (AF217036) had been reported. There were other variants (LN999944, DQ166542, and AY378310) that were reportedly deposited with NCBI by Welcome Trust UK. On the other hand however, only one haplotype (KC887547) with 99% similarity to that reported in Gabon was detected in FC27-ms.  

CONCLUSION

This study described the genetic diversity of Plasmodium falciparum and their population structure of imported cases of malaria into South East region of Saudi Arabia during 2018–2020. Genotyping of msp1 and msp2 of Plasmodium falciparum revealed varying levels of genetic and clonal diversity. Furthermore, our findings show that the population structure of imported Plasmodium falciparum parasites appeared to exhibit high diversity as seen from the allelic families of sequenced data. The findings highlight the possible threat posed by cases of imported malaria to the success achieved by Saudi Arabia in the elimination of the disease as had earlier been highlighted. The yet to be attained malaria free status planned for 2020 will need continuous monitoring.

Funding

The research was funded by a grant from Deanship of Scientific Research, King Faisal University, Al-Ahsa, Saudi Arabia. Grant number 160107.

ACKNOWLEDGEMENT

Authors acknowledge Deanship for Scientific research, King Faisal University, Al-Ahsa for financing of the research. We would like to thank Dr. Promise Emeka and Dr. Krishnaraj Thirugnanasambantham for their help with data analysis. We thank Hani Al-Farhan, Hani Al-Rasasi and Fatimah Al-Najja for their technical support. We would also like to acknowledge BEI Resources, 10801 University Boulevard Manassas, VA 20110-2209 for providing us with MRA-102G Genomic DNA 3D7 Plasmodium falciparum and MRA-155G Genomic DNA from HB3 Plasmodium falciparum.

ABBREVIATIONS

MOI: Multiplicity of infection; nPCR: nested PCR; PCR: Polymerase chain reaction; DBS: Dried blood spots; msp: merozoite protein.

CONFLICT OF INTEREST

The authors declare there are no conflict of interest.

REFERENCES


28. Sharma YD. Genetic diversity of Plasmodium falciparum vaccine candidate antigens from Madhya Pradesh, India; unpublished.


35. Enderes C, Bruske EI, Tschun S, Kombila D, Krensmen P, Frank M. Plasmodium falciparum erythrocyte membrane protein 1 ( PfEMP1) is not the main variant surface antigen of field isolates from a chronic asymptomatic infection; unpublished.


The 21st century has witnessed an increase in global travel, and this has reflected in cases of malaria parasite of different species of *Plasmodium* being imported into non-endemic regions as well as those where there is no active transmission. The study looked at genetic diversity of merozoites proteins 1 and 2 of *Plasmodium falciparum* gene. This investigation has shown the genetic diversity in stains of imported case of *P. falciparum* in the region of this study. Encountered genes reflected those seen in different tropical regions of the world. Hence if the area of this study is to achieve a malaria free status, there will be need for continuous monitoring.