

# Genetic Characterization of *Plasmodium vivax* Isolates from Pakistan Using Circumsporozoite Protein (*PvCSP*) and Merozoite Surface Protein-1 (*PvMSP-1*) genes as genetic markers

**Zainab Bibi**

Virtual University of Pakistan

**Anam Fatima**

Polyclinic Hospital Islamabad

**Rehana Rani**

Abasyn University

**Ayesha Maqbool**

Virtual University of Pakistan

**Samea Khan**

Virtual University of Pakistan

**Shumaila Naz** (✉ [naz.pmasaaur@gmail.com](mailto:naz.pmasaaur@gmail.com))

National University of Medical Sciences <https://orcid.org/0000-0001-8088-0084>

**Shahid Waseem**

Alpha Genomics (Pvt) Ltd

---

## Research

**Keywords:** Plasmodium vivax, Malaria, Genetic diversity, CSP, MSP-1

**Posted Date:** January 4th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-32064/v2>

**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 at Malaria Journal on February 25th, 2021. See the published version at <https://doi.org/10.1186/s12936-021-03654-w>.

# Abstract

**Background:** *Plasmodium vivax* contributes to over 70% malaria burden in Pakistan, but limited data exists on various aspects including genetic diversity of the parasite as compared to other parts of the world. Since the information about the genetic diversity of *P. vivax* assists to understand the population dynamics of the parasite, the current study was designed to understand population divergence of *Plasmodium vivax* in Pakistan using circumsporozoite protein (*PvCSP*) and merozoite surface protein-1 (*PvMSP-1*) genes as molecular markers.

**Methods:** *PvCSP* and *PvMSP-1* specific PCR and DNA sequencing were carried out for 150 blood samples collected from Islamabad and Rawalpindi, Pakistan. Genetic diversity and polymorphism was analyzed using ChromasPro, ClustalW, MEGA7, DnaSP v.5 and WebLogo programs.

**Results:** The PCR for *PvCSP* and *PvMSP-1* genes was carried out for 150 *P. vivax* isolates and resulting the PCR products ranging from 900 to 1100 bp for *PvCSP* and ~400bp for *PvMSP-1* genes respectively. Majority (93%; 141/150) of the *P. vivax* isolates were of VK210 variant and only 9 isolates were found to be of VK247 variant based on *PvCSP* gene. Out of the numerous peptide repeat motifs (PRMs) detected, GDRADGQPA (PRM1) and GDRAAGQPA (PRM2) were more extensively dispersed among the *P. vivax* isolates. Partial sequences (~400bp) at the N-terminal of *PvMSP-1* gene depicted high level of diversity.

**Conclusion:** High levels of genetic diversity based on *PvCSP* and *PvMSP-1* genes was observed in the isolated samples from the study area. Parasite typing is essential in predicting pattern of antigenic variations and drug resistance and for effective vaccine designing and development which can further assist in evaluating measures for malaria control at individual and community level. The base-line data presented here warrants future studies to investigate more into the genetic diversity of *P. vivax* with large sample size from across the country for better understanding of the transmission patterns of vivax malaria.

## Background

Malaria remains a serious infectious disease of public health importance and one of the leading causes of morbidity and mortality worldwide [1]. Globally, around 42% population is at risk with almost 228 million cases of malaria with 0.4 million reported deaths in 2018 based on the data collected from 91 countries [2, 3]. Among the five *Plasmodium* species, *Plasmodium vivax* (*P. vivax*) contributes around 70% malaria cases in Pakistan [4] with variable severity [5-7]. To circumvent the parasite load, there is a need to investigate the population structure, genetic diversity [8], parasite typing of the local *P. vivax* species, pattern of antigenic variations and drug resistance [9, 10]. It would lead to interruption of transmission cycle of the parasite in human host in endemic areas. Despite multiple academic research ventures, scanty data is available regarding the diverse genetic make-up of *P. vivax* in Pakistan [5, 6]. Genetic sequences of *PvCSP* and *PvMSP-1* are being used to understand the genetic diversity [11-13]. *PvCSP*, a highly immunogenic sporozoite surface protein, hence a good vaccine candidate, is encoded by single copy gene [8] and comprises of a central repeat domain that varies across *Plasmodium* species having two non-

repetitive domains at N- and C-terminals [14, 15]. The varying degree of number of peptides in the central repeat region reveals three variants of *P. vivax* namely VK210, VK247, *P. vivax*-like [16, 17]. Across the globe, these variants exhibit certain spatial predispositions. With a GDRA(A/D)GQPA amino acid repetition, VK210 strain dominates in the endemic region [18, 19]. Originating in Thailand, the VK247 strain is mostly reported from the areas where mixed infections are prominent [20]. VK247 depicts ANGAGNQPQ amino acid repeat in the central region [20, 21]. In *PvCSP*, polymorphism is reported to be limited in central tandem repeat among the isolates from the regions of Sub-Saharan Africa [22]. *PvMSP-1* is expressed on the surface of the blood-stage parasite [23]. *PvMSP-1*, a large gene, covers conserved and polymorphic regions [24] and has mosaic organization with 13 regions of variable blocks [25]. The three main regions of sequence divergence are block 2 (F1 region), 6–8 (F2 region) and 10 (F3 region) [26]. In the representative blocks, the genetically distinct *PvMSP-1* populations within the regions and polymorphism can be detected through PCR [5, 27]. Selective pressure of the host immune maintains the diversity of *PvMSP-1* gene, however, immunogenic properties can be affected by single-point mutation [28, 29]. Despite *P. vivax* contributing to 88% of malaria burden in Pakistan [5- 7], data regarding genetic diversity of this key circulating species is lacking. The present study was designed to investigate the genetic diversity of *P. vivax* in Potohar region of Pakistan exploiting *PvCSP* and *PvMSP-1* genes as molecular markers. The understanding of sequence diversity in *PvCSP* would contribute to comprehend the population dynamics and transmission pattern of the parasite in this region.

## Methods

### Study area

The study was conducted in Islamabad and Rawalpindi districts of the Punjab with longitudes 72°45' and 73°30' E and latitudes 33°30' and 33°50' N [30] (Figure 1). The climate in the study region ranges from showery warm to chilly dry wintry with the attributes of the semi-arid region of Pakistan. The monsoon rains typically start in June, get peak in August, and finish by September. The rainfall is between 620 and 1,200 mm per year. The weather and geographical settings of this region are favorable for the mosquito breeding with highest frequency reported in Rawalpindi (25.5%) and lowest in Chakwal (15.9%) [31].

### Sample collection

Blood samples (n=150) were collected during malaria transmission season (April and October, 2019) from Pakistan Institute of Medical Sciences (PIMS), Islamabad and Rawalpindi General Hospital, Rawalpindi, Pakistan. Venous blood (5 mL) was collected in ethylene diamine tetra-acetic acid (EDTA) (BD, USA) vacutainers from malaria patients. The samples were collected from the patients fulfilling the inclusion criteria of clinical signs and symptoms (fever, chills, headache, sweats, fatigue, nausea and vomiting). Initially malaria patients were screened by microscopic examination of Giemsa-stained thin and thick blood smears to ensure the *P. vivax* parasites and exclude samples with presence of other infections. Blood samples were stored at -20°C for a month until further analysis was carried out.

### DNA extraction and PCR amplification of *PvCSP* and *PvMSP-1* genes

Genomic DNA from 150 malarial blood samples were extracted by using standard phenol-chloroform method [32]. PCR primers for *PvCSP* and *PvMSP-1* genes were designed at Geneious software by using reference sequence of *P. vivax* (AB539044 and GQ890906). The primers for *PvCSP* gene were F: 5'-GGCCATAAATTTAAATGGAG-3' and R: 5'-ATGCTAGGACTAACAAATATG-3'. The PCR conditions were as follows: pre-denaturation at 94°C for 10 min followed by 35 cycles of annealing at 52°C 1 min and extension at 72°C 1 min, followed by final extension at 72°C for 10 min. The PCR of *PvMSP-1* gene was performed with primers F: 5'-ACATCATTAAAGGACCCATACAAG 3' and R: 5' GCAATTTCTTTACAGTGATCTCG-3' with similar PCR cycling conditions except that the annealing was at 56°C for 1 min. The PCR reactions were carried out in a 25 µL reaction mixture comprising of 2µL DNA template, 0.5mM dNTPs, 1X PCR reaction buffer (SolarBio Life Sciences, China), 0.2mM of each primer (BGI, China), 2.5mM MgCl<sub>2</sub> and 1 unit of *Taq* DNA polymerase (BLIRT, Poland). The PCR products were visualized using 1% agarose gel (ThermoFisher, USA) stained with ethidium bromide (SolarBio Life Sciences, China) and visualized under UV-transilluminator (ThermoFisher, USA).

### Sanger sequencing and analysis

Out of 150 malaria amplified genomic DNA samples, only 35 and 30 amplified PCR products of *PvCSP* and *PvMSP-1* genes were purified respectively by QIAquick PCR Purification Kit (Qiagen, Germany according to manufacturer's protocol) and sent to Beijing Genomics Institute (BGI), China for Sanger sequencing. DNA sequences of both *PvCSP* and *PvMSP-1* genes were read and assembled on both upstream and downstream ends for 35 sequences of *PvCSP* gene and 30 sequences of *PvMSP-1* gene. The sequences were then analyzed by using ChromasPro (version 1.5) software (<http://technelysium.com.au/wp/chromaspro/>) and Bio Edit alignment editor (version 7.2) (<https://bioedit.software.informer.com/7.2/>). The sequenced samples were validated by BlastN (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) and alignment of top hit resulted sequences were done by ClustalW (<https://www.genome.jp/tools-bin/clustalw>). The number of haplotypes (H), nucleotide diversity ( $\pi$ ) and haplotype diversity (Hd) were calculated by DNAsp v5 [33]. The evolutionary relationships of the both genes were established, and evolutionary tree were constructed by the neighbor-joining method using Molecular Evolutionary Genetics Analysis (MEGA 7.0) software [34]. The neutral theory of natural selection was also measured by Tajima's D, Fu and Li's D\*, and Fu and Li's F\* tests using DNAsp v5 [33]. The negative value of Tajima's D, Fu and Li's D indicates an excess of rare alleles that might results from selective sweep. A plot was constructed to look for polymorphic patterns of the N- and C-terminal in Pakistani *PvMSP-1* gene by using the WebLogo program (<https://weblogo.berkeley.edu/logo.cgi>). The nucleotide sequences of *PvCSP* and *PvMSP-1* genes obtained from this study were submitted in NCBI database under accession number MT222296 to MT222330 and MT303819 to MT303848, respectively.

## Results

In the present study, a total of 150 blood samples were collected from *P. vivax* malaria infected patients from the hospitals of twin cities of Islamabad and Rawalpindi. The blood samples were verified by microscopic examination to ensure the *P. vivax* parasites and exclude samples with any other infections.

The PCR for *PvCSP* and *PvMSP-1* genes were carried out of 150 *P. vivax* isolates out of which 35 sequences of *PvCSP* gene and 30 sequences of *PvMSP-1* gene were sequenced with PCR products ranging from 900 to 1100 bp for *PvCSP* gene and ~400 bp for partial sequence of *PvMSP-1* gene.

### Sequence analysis of *PvCSP* gene

Multiple sequence alignment of the translated nucleotide sequences was carried out for the analysis of polymorphisms in the pre-, post- and central repeats of the *PvCSP* gene. The top hits for *PvCSP* gene were extracted from GenBank protein database using Blastp and one of the sequences of *PvCSP* gene of Iranian isolate was retrieved and used as reference sequence (KT588208.1). The multiple sequence alignment of extracted amino acid sequences were performed using ClustalW. When compared with the reference sequence (KT588208.1), the sequence analysis of *PvCSP* gene showed the VK210 and VK247 variant types infection. *PvCSP* gene sequence analysis revealed that majority (94%; 141/150) of the *P. vivax* isolates were of VK210 variant and only 9 isolates were found to be VK247 type. All *PvCSP* gene-based *P. vivax* variants started with the same pre-repeat sequence (KLLKQP region). In the central-repeat region (CRR), the VK210 sequences comprised of variable repeats of PRMs, GDRADGQPA (PRM1), GDRAAGQPA (PRM2) which were found in all the isolates. It was followed through two conserved post-repeat sequence GNGAGGQAA (PRM3) and GGNAANK (PRM4) and one post-repeat insert i.e., KAEDA region. The one-copy repeat region of GGNA was found after the CRR in all the analyzed sequences. The frequency of peptide repeat motifs (PRMs) in the central repeat region (CRR) of *PvCSP* has been summarized in Figure 2. The observed non-synonymous substitution based on diverse types of repetition in allotypes (RATs), which leads to different PRM,s are mentioned in Table 1.

### *PvCSP* CRR based genetic population structure

The population genetic structure based on the *PvCSP* CRR of the *P. vivax* isolates was analyzed and compared with *PvCSP* isolates of neighboring countries Iran, India and Myanmar. The haplotype (gene) diversity of *PvCSP* was categorized into fifteen distinct haplotypes with an estimated Hd of 0.547 and ten distinct haplotypes with an estimated Hd of 0.345 in Pakistani and Iran *PvCSP* samples respectively. Adding to this, Tajima's D, Fu and Li's D\* and F\* tests also accepted occurrence of a neutral model of polymorphism with values for these tests are given in Table 2 for the *PvCSP* variants from Pakistan, Iran, India and Myanmar. The overall nucleotide and haplotype diversity were  $0.02371 \pm 0.00056$  and  $0.084 \pm 0.00701$ , respectively. These results suggested that the CRR region of *PvCSP* population of Pakistan was under positive natural selection. Further, the effect of natural selection was explicated by the Tajima's D which was 0.54276 ( $P > 0.10$ ). The Fu and Li's D and F values for CRR region was also positive. The nucleotide diversity and natural selection were also analyzed in *PvCSP* population of Iran, India and Myanmar. *PvCSP* population from India and Iran showed high nucleotide diversity but values from Myanmar of *PvCSP* population were negative, suggesting negative selection. The values of the Tajima's D, Fu and Li's D and F values for CRR region was also positive for *PvCSP* population of Iran and India as shown in Table 2.

### Phylogenetic analysis of *PvCSP* gene

A phylogenetic tree drawn from the sequence findings of *PvCSP* gene is presented in Figure 3. Two separate clades can be inferred from the tree; one having VK210 variant type while the other has VK247 variant type of *PvCSP* isolates. Four sub-clusters of VK210 and VK247 can be distinguished in the leading clade. The associated taxa were clustered together and shown after the branches with the branch length as of the evolutionary distances used to calculate the phylogenetic tree. The evolutionary distance was computed by the p-distance method and analyzed using 55 nucleotide sequences. All the positions that had gaps or missing data were discarded. VK210 strain sequences from Pakistan showed 48- 100% identity with *PvCSP* sequences from countries such as Iran, Greece, India, USA, Sri Lanka, Australia, Vanuatu and Myanmar, whereas the sequences of VK247 strains from Pakistan showed 100% identity with sequences from Iran, Columbia, Vanuatu, USA, India, Latin America and Korea.

### **Sequence analysis of *PvMSP-1* gene**

The top hits for *PvMSP-1* gene were extracted from GenBank protein database using Blastp and one of the sequences of *PvMSP-1* gene of Iranian isolate was retrieved and used as reference sequence (KX697612.1). Sequence of *PvMSP-1* gene was compared with reference sequence KX697612.1 of *P. vivax* strain. It revealed that *PvMSP-1* gene sequences of 30 isolates were corresponding to partial sequence of other *PvMSP-1* gene sequence at N-terminal. Overall, 13 single nucleotide polymorphisms (SNPs) were found amongst 30 samples with an average  $\pi$  value of 0.00143 in *PvMSP-1* gene. The average conserved sequence between Pakistani and reference Iranian *PvMSP-1* gene was C: 0.835 indicating that sequences have remained relatively unchanged with close evolutionary relationship. Overall genetic polymorphisms of the *PvMSP-1* population were analyzed as shown in Figure 4. The N-terminal non-repeat region of the *PvMSP-1* was well-conserved, although low frequencies of uneven amino acid changes were identified. The significant variation was observed from amino acid position K55N to M78T/N showing uneven and low frequencies with less conserved sequence of *PvMSP-1*.

### ***PvMSP-1* N-terminal based genetic population structure**

Population genetic structure based on the N-terminal of *PvMSP-1* gene of the *P. vivax* isolates was analyzed and compared with isolates of neighboring country Iran as shown in Table 2. The haplotype diversity of *PvMSP-1* gene was comparable between two countries ranging from 0.962 to 0.954. Adding to this, Tajima's D, Fu and Li's D\* and F\* tests also accepted occurrence of a neutral model of polymorphism with values for Fu and Li's D and Fu and Li's F are given in Table 2 for the *PvMSP-1* variants from Pakistan and Iran. The results of *PvMSP-1* population of Pakistan also indicated that positive natural selection may occur in the region. The overall nucleotide and haplotype diversity were  $0.00162 \pm 0.0000026$  and  $0.012 \pm 0.00014$  respectively. The effect of natural selection was estimated by the Tajima's D which was 1.67790 ( $P > 0.10$ ). The Fu and Li's D and F values for *PvMSP-1* population were also positive. The nucleotide diversity and natural selection were also analyzed in *PvMSP-1* population of Iran and India which showed high nucleotide diversity with positive values of the Tajima's D, Fu and Li's D and Fu and Li's F as shown in Table 2.

### **Phylogenetic analysis of *PvMSP-1* gene**

Based on sequence of *PvMSP-1* gene, a phylogenetic tree was constructed (Figure 5). Two distinct clades can be inferred from the tree. The first is divided further into three sub-clades and contains Pakistani isolates and isolates belong to East Africa, Thailand, Mexico, India and USA with these isolates having 16 to 95% identity with sequences from Pakistani *PvMSP-1* population. Second clade is further divided into two sub-clades having Turkey, Iran, Korea and Southern Mexico isolates in addition to sequences of Pakistani isolates with 62- 94% sequence identity.

## Discussion

Malaria is one of the major public health concerns with limited data in Pakistan [35]. As Pakistan shares border with malaria endemic countries like Iran, India and Afghanistan, the human migration across the border is inevitable possibly facilitating substantial cross-border transmission of malaria. The resultant recombination leads to genetic diversity and affects the frequency of new alleles in the parasite population [36, 37]. There are limited studies from Pakistan which have analyzed the diversity of local *P. vivax* in detail [5, 6, 24, 36]. *PvCSP* and *PvMSP-1* are among the other important genetic markers used by the researchers to understand population structure and evolutionary dynamics from different geographical regions [5, 26]. In the present study, the genetic polymorphism in the *PvCSP* and *PvMSP-1* genes were studied in detail. The results of the study support the findings of the previously published data showing that VK210 strain is predominant type with the prevalence rate ranging between 56–100% in Iran, Myanmar, Brazil, India, Thailand, Azerbaijan and Mexico [15, 19, 20, 24, 43, 39-43]. There are only few malaria endemic areas where VK247 isolates are commonly present [44].

The analysis of translated nucleotide sequences suggest that GDRADGQPA (PRM1) and GDRAAGQPA (PRM2) are two major PRMs. Earlier studies also reported the dominance of the two prime PRMs in the clinical isolates [8, 9, 14]. All these isolates were composed of similar pre-repeat sequence (KLKQP) region and conserved post-repeat sequence GGNAANK (PRM4) present as a last section in all of the VK210 isolates as aforementioned in studies from India, Iran and Sri Lanka [9, 14, 20]. Another peptide repeat motif GNGAGGQAA (PRM3) was found at lower frequency (0.6%) in the isolates. The variations exist in the amino acid and nucleotide sequences of the *Plasmodium* antigens due to variations in the repeat unit numbers which is indicative of the natural selection pressure by the host immune system [14, 44]. The arrangement of the main PRM1 and PRM2 factors leads to 15 different haplotypes of *PvCSP*. The analysis of CRR region of Pakistani *PvCSP* isolates is indicative of positive selection when compared with *PvCSP* isolates of Iran and India. The negative Tajima's D values of Myanmar *PvCSP* population imply purifying negative selection [15]. The evidence from the previous studies has reported that the arrangements and numbers of PRMs in CRR are indicative of occurrence of phenomenon of natural selection on the *PvCSP* isolates [45, 46].

*PvMSP-1* is one of the most promising vaccine candidates and is available for antigenic and genetic variation studies of *P. vivax* populations [47]. In this study, partial sequence (~400bp) at N-terminal of *PvMSP-1* gene has depicted a high-level of diversity, which is in concordance with what has already been observed in neighboring country Iran [24] and also in previous study from Pakistan [5]. In the northwestern

region of Thailand, a high degree of mutational variety was observed in *PvMSP-1* genes of *P. vivax* isolates [29, 46]. The *PvMSP-1* population of Pakistan has also indicated positive natural selection when compared with *PvMSP-1* isolates of Iran and India. The significantly positive values may be the result of balancing the selection and population bottlenecks. The sequence diversity of the population is best studied by the intragenic recombination of *PvMSP-1* gene where the allelic recombination frequencies may aid as a character reference for understanding the parasitic population structure [28]. Kibria et al. [26] also indicated a high genetic diversity of *PvMSP-1* gene which undergoes selective pressure for the existence and spread of the parasite. Similar pattern of genetic diversity was observed in *PfMSP-1* gene of *P. falciparum* populations in Pakistan [36]. The *PvMSP-1* gene sequences were useful in distinguishing the two central localities of origin in terms of geography as well as helping them to group in two different clades. These biological groups are further sub divided into different clusters according to their geographic origin [47]. The polymorphic nature of MSP-1 markers can be used to differentiate between reemergence and reinfections and to determine ultimate shifts in population dynamics of parasites [48].

Other studies have suggested that the mode of evolution in *PvCSP* gene can lead to cohort of variants that can elude the host immune response under the effect of both mitotic recombination and positive selection of new variants of *P. vivax* [17]. Therefore, it is safe to assume that the wide variety of *P. vivax* may be interrelated with multiple other variables including, but not limited to, genetic and biological characteristics, immunity of the host and the displacement of individuals within the boundaries of the endemic areas. Furthermore, the spread of *P. vivax* infections is also re-inforced by relapse and early gametocytaemia, which in turn sustains local diversity, paving way for a more efficient transmission to the vector mosquitoes [13, 49]. The results revealed a broad range of genetic variety of *PvCSP* and *PvMSP-1* genes in *P. vivax* population. The localities of the study areas from where samples were collected are inhabited by various ethnicities which suggested that the migration of people may carry diverse parasite entities that increase the variety of the gene-pool.

## Conclusion

High-level genetic diversity based on *PvCSP* and *PvMSP-1* genes for *P. vivax* in clinical isolates was observed in selected areas of Pakistan. A better understanding of genetic variability and stable as well as unstable reemergence of the infection due to population dynamics of parasite will help to control malaria at individual and community level. In order to design and put new and more effective vaccinations into effect, these studies can provide substantial help. The base-line data presented here warrants further studies to investigate more into the genetic diversity of *P. vivax* with large sample size from across the country for better understanding of the vivax malaria transmission patterns.

## Declarations

### Ethics approval and consent to participate

The study was approved by institutional review board (IRB), Virtual University of Pakistan, (letter No. VU/ASRB/131-7). Blood samples were collected from diagnosed malaria patients after informed consent.

## Consent for publication

Not applicable

## Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

## Competing interests

The authors declare no competing interests

## Funding

There was no financial support for this research.

## Authors' contributions

ZB conducted laboratory experiments and data analysis, AF helped in sample collection, clinical inputs and malaria identification, RR helped in primer designing and molecular data input, AM provided scientific inputs and critical review of manuscript, SK helped specifically in CSP gene analysis, SN supervised the wet labs, critically reviewed manuscript, helped in manuscript drafting and data analysis, SW conceived the idea and study design, provided all wet lab reagents/lab facilities and contributed to the critical revision of the manuscript, . All authors read and approved the final manuscript.

## Acknowledgements

We would like to thank medical staff of hospital and all the study participants to participate in the study

## Authors' information

<sup>1</sup>Department of Molecular Biology, Virtual University of Pakistan, Lahore, Pakistan. <sup>2</sup>Department of Medicine, Polyclinic Hospital, Islamabad, Pakistan. <sup>3</sup>Department of Life Sciences, Abasyn University, Islamabad, Pakistan. <sup>4</sup>Department of Biological Sciences, National University of Medical Sciences, Rawalpindi, Pakistan. <sup>5</sup>Alpha Genomics (Pvt) Ltd., Islamabad, Pakistan.

## References

1. Ghanchi NK, Khan MH, Arain MA, Zubairi M, Raheem A, Khan MA, Beg MA. Hematological Profile and Gametocyte Carriage in Malaria Patients from Southern Pakistan. *Cureus*. 2019;11(3), e4256.
2. World Health Organization. World malaria report 2020. <https://www.who.int/news-room/fact-sheets/detail/malaria>

3. Bettencourt P. Current Challenges in the Identification of Pre-Erythrocytic Malaria Vaccine Candidate Antigens. *Frontiers in Immunology*. 2020;11:190.
4. Pakistan Malaria Annual Report, 2019. Malaria Control Program, Directorate of Malaria Control, Pakistan.
5. Raza A, Ghanchi NK, Thaver AM, Jafri S, Beg MA. Genetic diversity of *Plasmodium vivax* clinical isolates from southern Pakistan using PvCSP and pvmSP1 genetic markers. *Malaria journal*. 2013;12(1):16.
6. Khan SN, Khan A, Khan S, Ayaz S, Attaullah S, Khan J, Khan MA, Ali I, Shah AH. PCR/RFLP-based analysis of genetically distinct *Plasmodium vivax* population of PvmSP-3 $\alpha$  and PvmSP-3 $\beta$  genes in Pakistan. *Malaria journal*. 2014;13(1):355.
7. Ahmad T. *Plasmodium vivax* and *Plasmodium falciparum* are Common Malaria Species in Pakistan. *Biomedical Research and Therapy*. 2016;3(6):1-7.
8. Zhang LL, Yao LN, Chen HL, Lu QY, Ruan W. Genetic diversity analysis of PvCSP and its application in tracking of *Plasmodium vivax*. *Experimental parasitology*. 2018;188:26-35.
9. Dias S, Wickramarachchi T, Sahabandu I, Escalante AA, Udagama PV. Population genetic structure of the *Plasmodium vivax* circumsporozoite protein (PvCSP) in Sri Lanka. *Gene*. 2013;518(2):381-7.
10. Talha AA, Pirahmadi S, Mehrizi AA, Djadid ND, Nour BY, Zakeri S. Molecular genetic analysis of *Plasmodium vivax* isolates from Eastern and Central Sudan using PvCSP and pvmSP-3 $\alpha$  genes as molecular markers. *Infection, Genetics and Evolution*. 2015;32:12-22.
11. Kaur H, Sehgal R, Goyal K, Makkar N, Yadav R, Bharti PK, Singh N, Sarmah NP, Mohapatra PK, Mahanta J, Bansal D. Genetic diversity of *Plasmodium falciparum* merozoite surface protein-1 (block 2), glutamate-rich protein and sexual stage antigen Pfs25 from Chandigarh, North India. *Tropical Medicine & International Health*. 2017;22(12):1590-8.
12. Simon B, Sow F, Al Mukhaini SK, Al-Abri S, Ali OA, Bonnot G, Bienvenu AL, Petersen E, Picot S. An outbreak of locally acquired *Plasmodium vivax* malaria among migrant workers in Oman. *Parasite*. 2017;24.
13. Liu Y, Zhou RM, Zhang YL, Wang DQ, Li SH, Yang CY, Qian D, Zhao YL, Zhang HW, Xu BL. Analysis of polymorphisms in the circumsporozoite protein gene of *Plasmodium vivax* isolates from Henan Province, China. *Malaria journal*. 2018;17(1):103.
14. Kaur H, Sehgal R, Kumar A, Sehgal A, Bharti PK, Bansal D, Mohapatra PK, Mahanta J, Sultan AA. Exploration of genetic diversity of *Plasmodium vivax* circumsporozoite protein (PvCSP) and *Plasmodium vivax* sexual stage antigen (Pvs25) among North Indian isolates. *Malaria journal*. 2019;18(1):308.
15. Vö TC, Lê HG, Kang JM, Moe M, Naw H, Myint MK, Lee J, Sohn WM, Kim TS, Na BK. Genetic polymorphism and natural selection of circumsporozoite protein in Myanmar *Plasmodium vivax*. *Malaria journal*. 2020; 19(303).
16. Alves RT, Póvoa MM, Goldman IF, Cavasini CE, Rossit AR, Machado RL. A new polymerase chain reaction/restriction fragment length polymorphism protocol for *Plasmodium vivax* circumsporozoite

- protein genotype (VK210, VK247, and *vivax*-like) determination. Diagnostic microbiology and infectious disease. 2007;59(4):415-9.
17. Patil A, Orjuela-Sánchez P, da Silva-Nunes M, Ferreira MU. Evolutionary dynamics of the immunodominant repeats of the *Plasmodium vivax* malaria-vaccine candidate circumsporozoite protein (CSP). Infection, Genetics and Evolution. 2010;10(2):298-303.
  18. Rosenberg R, Wirtz RA, Lanar DE, Sattabongkot J, Hall T, Waters AP, Prasittisuk C. Circumsporozoite protein heterogeneity in the human malaria parasite *Plasmodium vivax*. Science. 1989;245(4921):973-6.
  19. Leclerc MC, Durand P, Gauthier C, Patot S, Billotte N, Menegon M, Severini C, Ayala FJ, Renaud F. Meager genetic variability of the human malaria agent *Plasmodium vivax*. Proceedings of the National Academy of Sciences. 2004;101(40):14455-60.
  20. Zakeri S, Akram AM, Navid DD, Georges S. Circumsporozoite protein gene diversity among temperate and tropical *Plasmodium vivax* isolates from Iran. Tropical Medicine & International Health. 2006;11(5): 729-737.
  21. Qari SH, Shi YP, Goldman IF, Udhaykumar V, Collins WE, Lal AA, Alpers MP. Identification of *Plasmodium vivax*-like human malaria parasite. The Lancet. 1993;341(8848):780-3.
  22. Tahar R, Ringwald P, Basco LK. Heterogeneity in the circumsporozoite protein gene of *Plasmodium malariae* isolates from sub-Saharan Africa. Molecular and Biochemical Parasitology. 1998;92(1):71-8.
  23. Birkenmeyer L, Muerhoff AS, Dawson GJ, Desai SM. Isolation and characterization of the MSP1 genes from *Plasmodium malariae* and *Plasmodium ovale*. The American journal of tropical medicine and hygiene. 2010;82(6):996-1003.
  24. Zakeri S, Raeisi A, Afsharpad M, Kakar Q, Ghasemi F, Atta H, Zamani G, Memon MS, Salehi M, Djadid ND. Molecular characterization of *Plasmodium vivax* clinical isolates in Pakistan and Iran using pvmsp-1, pvmsp-3a and PvCSP genes as molecular markers. Parasitology International. 2010;59(1):15-21.
  25. Bonilla JA, Validum L, Cummings R, Palmer CJ. Genetic diversity of *Plasmodium vivax* PvCSP and pvmsp1 in Guyana, South America. The American Journal of Tropical Medicine and Hygiene. 2006;75(5):830-5.
  26. Kibria MG, Elahi R, Mohon AN, Khan WA, Haque R, Alam MS. Genetic diversity of *Plasmodium vivax* in clinical isolates from Bangladesh. Malaria journal. 2015;14(1):267.
  27. Imwong M, Pukrittayakamee S, Grüner AC, Rénia L, Letourneur F, Looareesuwan S, White NJ, Snounou G. Practical PCR genotyping protocols for *Plasmodium vivax* using Pvcs and Pvmsp1. Malaria journal. 2005;4(1):20.
  28. Putaporntip C, Jongwutiwes S, Sakihama N, Ferreira MU, Kho WG, Kaneko A, Kanbara H, Hattori T, Tanabe K. Mosaic organization and heterogeneity in frequency of allelic recombination of the *Plasmodium vivax* merozoite surface protein-1 locus. Proceedings of the National Academy of Sciences. 2002;99(25):16348-53.

29. Thanapongpichat S, Khammanee T, Sawangjaroen N, Buncherd H, Tun AW. Genetic Diversity of *Plasmodium vivax* in Clinical Isolates from Southern Thailand using PvMSP1, PvMSP3 (PvMSP3 $\alpha$ , PvMSP3 $\beta$ ) Genes and Eight Microsatellite Markers. The Korean journal of parasitology. 2019;57(5):469.
30. Sheikh IM, Pasha MK, Williams VS, Raza SQ, Khan KS. Environmental geology of the Islamabad-Rawalpindi area, northern Pakistan. Regional Studies of the Potwar-Plateau Area, Northern Pakistan. Bull. G. 2008;2078.
31. Qureshi NA, Fatima H, Afzal M, Khattak AA, Nawaz MA. Occurrence and seasonal variation of human Plasmodium infection in Punjab Province, Pakistan. BMC infectious diseases. 2019;19(1):935.
32. Sambrook J, Russell DW, Sambrook J. The condensed protocols: from molecular cloning: a laboratory manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 2006.
33. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25:1451–2
34. Kumar S, Stecher G, Tamura, K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular biology and evolution. 2016;33(7):1870-1874.
35. Khatoon L, Baliraine FN, Bonizzoni M, Malik SA, Yan G. Prevalence of antimalarial drug resistance mutations in *Plasmodium vivax* and *falciparum* from a malaria-endemic area of Pakistan. The American journal of tropical medicine and hygiene. 2009;81(3):525-8.
36. Khatoon L, Baliraine FN, Bonizzoni M, Malik SA, Yan G. Genetic structure of *Plasmodium vivax* and *Plasmodium falciparum* in the Bannu district of Pakistan. Malaria journal. 2010;9(1):112.
37. Ord R, Polley S, Tami A, Sutherland CJ. High sequence diversity and evidence of balancing selection in the Pvmsp3 $\alpha$  gene of *Plasmodium vivax* in the Venezuelan Amazon. Molecular and biochemical parasitology. 2005;144(1):86-93.
38. Shabani SH, Zakeri S, Mehrizi AA, Mortazavi Y, Djadid ND. Population genetics structure of *Plasmodium vivax* circumsporozoite protein during the elimination process in low and unstable malaria transmission areas, southeast of Iran. Acta tropica. 2016;160:23-34.
39. González-Cerón L, Martínez-Barnette J, Montero-Solís C, Santillán F, Soto AM, Rodríguez MH, Espinosa BJ, Chávez OA. Molecular epidemiology of *Plasmodium vivax* in Latin America: polymorphism and evolutionary relationships of the circumsporozoite gene. Malaria journal. 2013;12(1):243.
40. Moon SU, Lee HW, Kim JY, Na BK, Cho SH, Lin K, et al. High frequency of genetic diversity of Plasmodium vivax field isolates in Myanmar. Acta Trop. 2009;109:30–6.
41. Cui LW, Mascorro CN, Fan Q, Rzomp KA, Khuntirat B, Zhou G, et al. Genetic diversity and multiple infections of Plasmodium vivax malaria in western Am J Trop Med Hyg. 2003;685:613–9.
42. Chenet SM, Tapia LL, Escalante AA, Durand S, Lucas C, Bacon DJ. Genetic diversity and population structure of genes encoding vaccine candidate antigens of *Plasmodium vivax*. Malaria journal. 2012

Dec 1;11(1):68.

43. Kim SH, Hwang SY, Shin JH, Moon CS, Kim DW, Kho WG. Molecular genetic characterization of the merozoite surface protein 1 Gene of *Plasmodium vivax* from reemerging Korean isolates. *Clinical and vaccine immunology*. 2009;16(5):733-8.
44. Lopez AC, Ortiz A, Coello J, Sosa-Ochoa W, Torres RE, Banegas EI, Jovel I, Fontecha GA. Genetic diversity of *Plasmodium vivax* and *Plasmodium falciparum* in Honduras. *Malaria journal*. 2012;11(1):391.
45. Hughes AL. The evolution of amino acid repeat arrays in Plasmodium and other organisms. *Journal of molecular evolution*. 2004;59(4):528-35.
46. Parobek CM, Bailey JA, Hathaway NJ, Socheat D, Rogers WO, Juliano JJ. Differing patterns of selection and geospatial genetic diversity within two leading *Plasmodium vivax* candidate vaccine antigens. *PLoS Neglected Tropical Diseases*. 2014;8:e2796.
47. Putapornpipit C, Hongsrimuang T, Seethamchai S, Kobasa T, Limkittikul K, Cui L, Jongwutiwes S. Differential prevalence of Plasmodium infections and cryptic *Plasmodium knowlesi* malaria in humans in Thailand. *The Journal of infectious diseases*. 2009;199(8):1143-50.
48. Lopez AC, Ortiz A, Coello J, Sosa-Ochoa W, Torres RE, Banegas EI, Jovel I, Fontecha GA. Genetic diversity of *Plasmodium vivax* and *Plasmodium falciparum* in Honduras. *Malaria journal*. 2012;11(1):391.
49. Barry AE, Waltmann A, Koepfli C, Barnadas C, Mueller I. Uncovering the transmission dynamics of *Plasmodium vivax* using population genetics. *Pathogens and global health*. 2015;109.

## Tables

**Table 1.** Nucleotide sequence of four repeated allotypes (RATs) and the peptide repeat motif (PRMs) in the central-repeat region of *PvCSP*

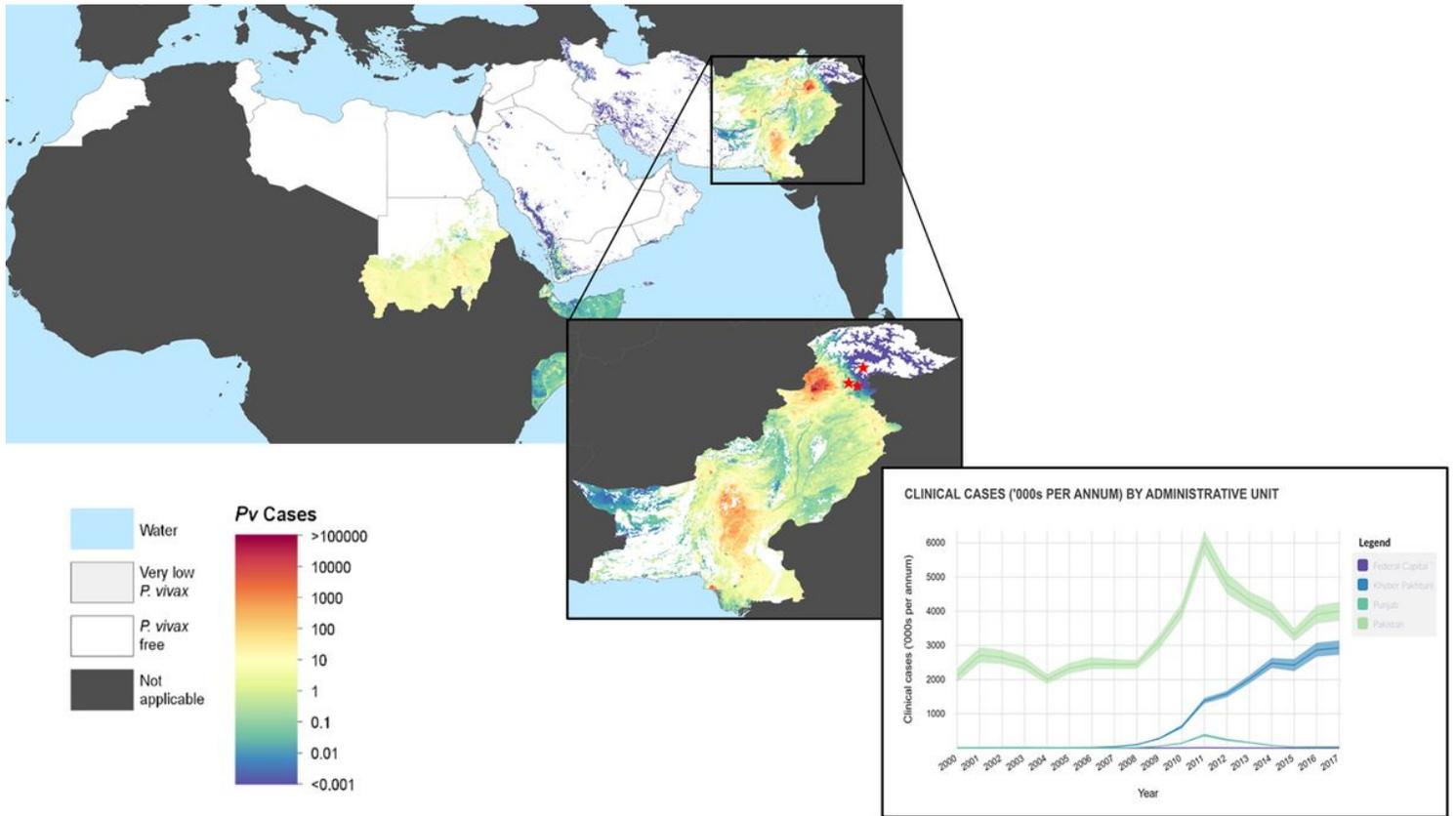
PRMs	Nucleotide sequence of the repeat allotypes (RATs)
GDRADGQPA (PRM1)	GGAGACAGAGCAGATGGACAGCCAGCA GGAGACAGAGCAGATGGACAGCCAGCA GGTGATAGAGCAGCTGGACAACCAGCA GGTGATAGAGCAGATGGACAGCCAGCA GGCGATAGAGCAGCTGGACAGCCAGCA GGCGATAGAGCAGATGGACAGCCAGCA GGAGATAGAGCAGCTGGACAGCCAGCA GGCGATAGAGCAGATGGACAGCCAGCA
GDRAAGQPA (PRM2)	GGAGATAGAGCAGCTGGACAGCCAGCA GGCGATAGAGCAGATGGACAGCCAGCA GGAGATAGAGCAGCTGGACAACCAGCA GGTGATAGAGCAGCTGGACAACCAGCA GGAGATAGAGCAGATGGACAACCAGCA GGAGATAGAGCAGCTGGACAGCCAGCA GGAGATAGAGCAGCTGGACAGCCAGCA GGAGATAGAGCAGCTGGACAGCCAGCA
GNGAGGQAA (PRM3)	GGAGATAGAGCAGCTGGACAGCCAGCA
GGNAANK (PRM4)	GGAAATGGTGCAGGTGGACAGGCAGCA GGAGGAAATGCGGCAAACAAG

**Table 2.** Estimates of nucleotide-, haplotype-diversity and DNA sequence polymorphisms of *P. vivax PvCSP* and *PvMSP-1* genes in Pakistan

\* $P > 0.10$

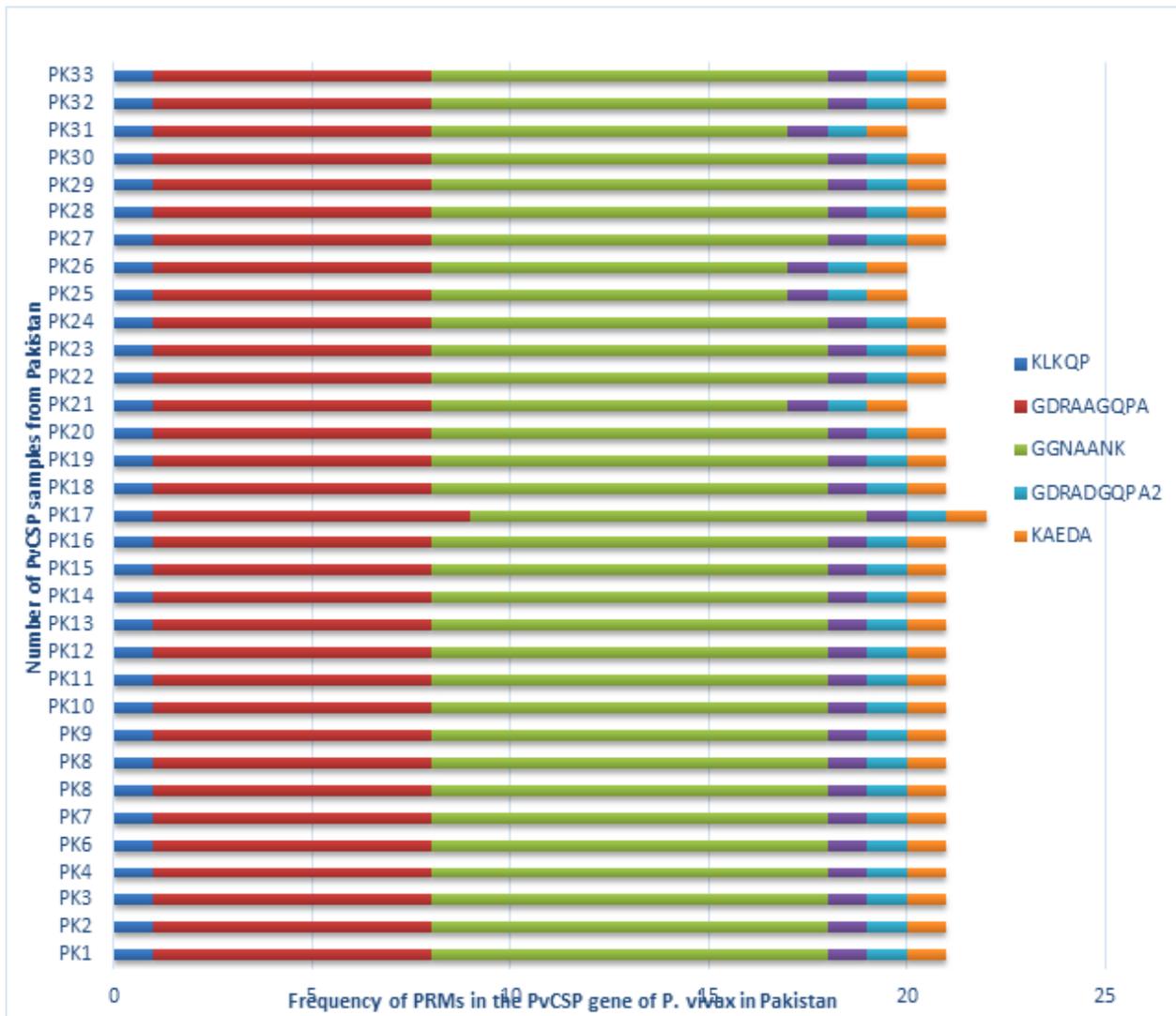
## Figures

Population	Type <i>P. vivax</i>	No. of samples	No. of Haplotypes	Haplotype diversity	Diversity ± SD	Fu & Li's <i>D</i> *	Fu & Li's <i>F</i> *	Tajima's <i>D</i> *
					Nucleotide			
					Haplotype			
Pakistan	<i>PvCSP</i>	35	15	0.547	0.02371± 0.00056	1.17870	1.12083	0.54276
Iran	<i>PvCSP</i>	28	10	0.345	0.084±0.00701	1.01980	1.80767	0.43556
Myanmar	<i>PvCSP</i>	15	7	0.175	0.02001± 0.00031	-1.20965	-1.06781	-0.78645
India	<i>PvCSP</i>	25	17	0.556	0.057± 0.00478 0.01781 ±0.0008 0.056±0.015 0.0370±0.0064 0.681±0.076	1.05422	1.02237	0.35674
Pakistan	<i>PvMSP-1</i>	30	10	0.962	0.00162±0.0000026 0.012±0.00014	1.86276	2.13897	1.67790
Iran		32	10	0.954		1.78902	1.90878	1.56845
India	<i>PvMSP-1</i>	25	07	0.576	0.00159±0.0000023 0.012±0.00014	1.55433	1.744009	1.66792
	<i>PvMSP-1</i>				0.0212 ± 0.0005 0.989 ± 0.010			



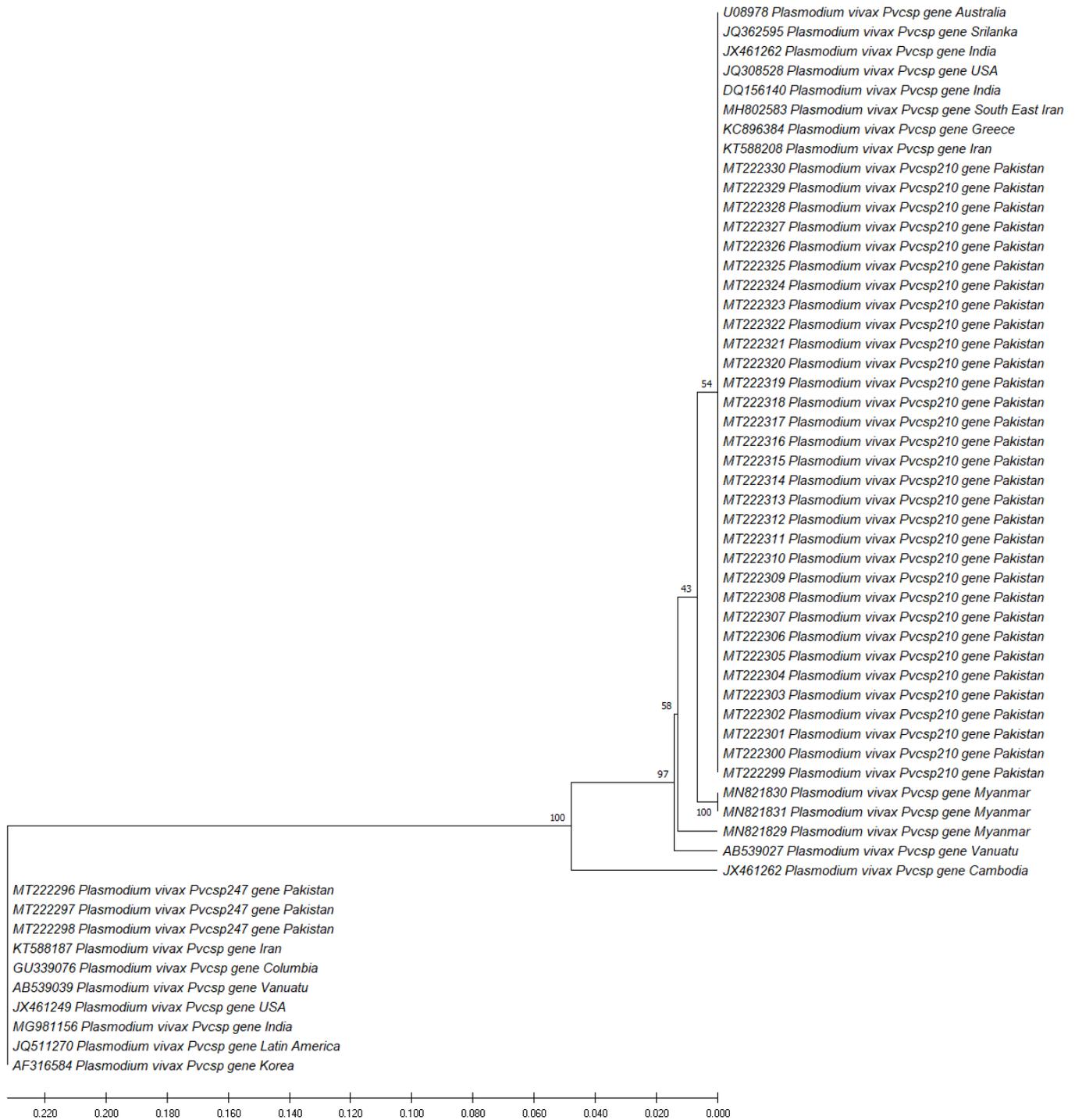
**Figure 1**

Map of *P. vivax* prevalence in EMRO region and Pakistan. The map showing all age clinical cases of *P. vivax* from year 2000-2017 of EMRO region and Pakistan. These maps were originally created by the Malaria Atlas Project, University of Oxford available at <https://malariaatlas.org/>. The estimated sites of the study sites described here are showed with red star: Rawalpindi and Islamabad



**Figure 2**

Frequency of peptide repeat motifs (PRMs) in the central repeat region (CRR) of Pakistani PvCSP variants



**Figure 3**

Phylogenetic relationships of VK210 and VK247 (based on PvCSP gene analysis) isolates from Pakistan. The Neighbor-Joining method was used to deduce evolutionary history with the branch length of =0.50423062.

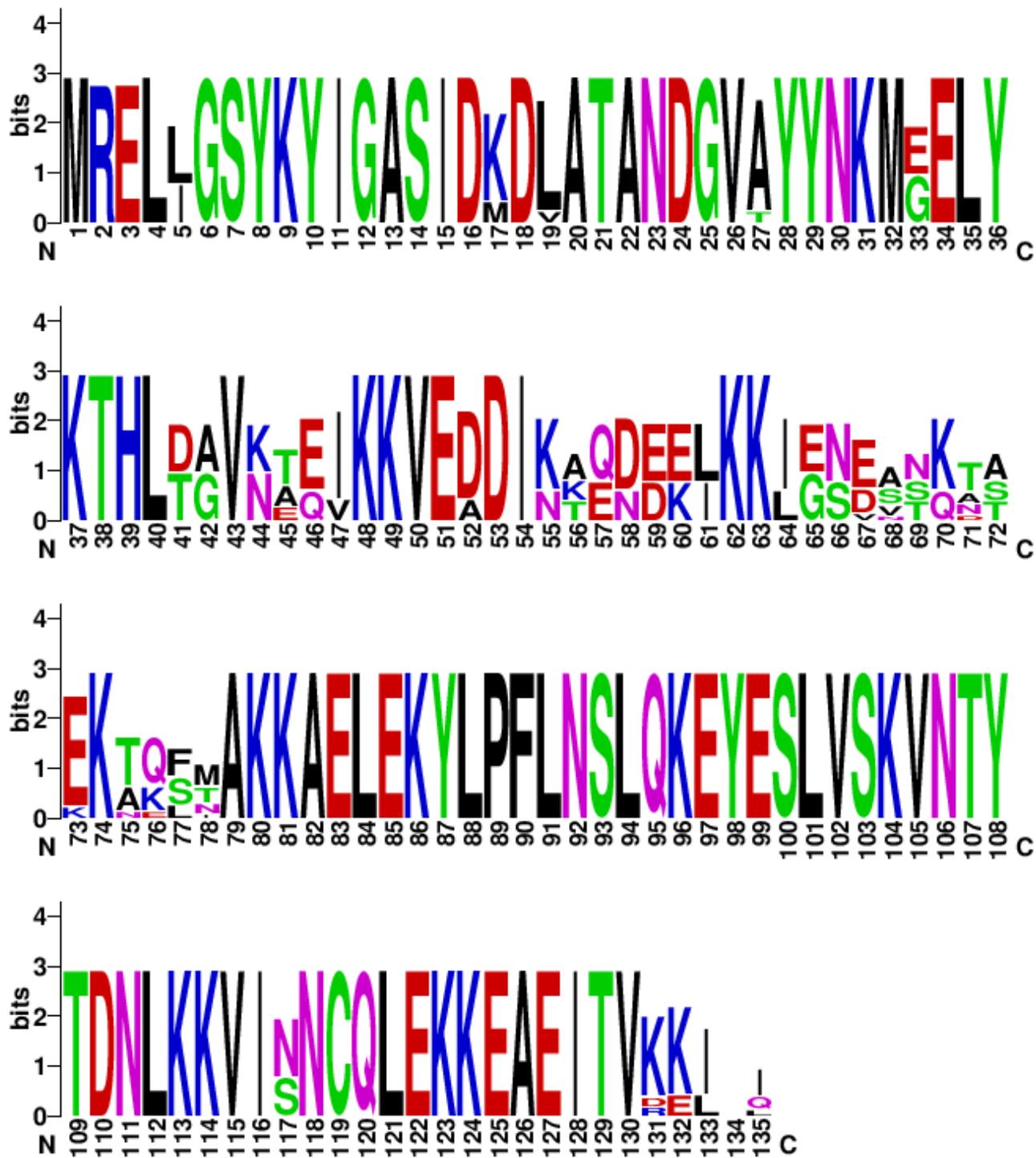
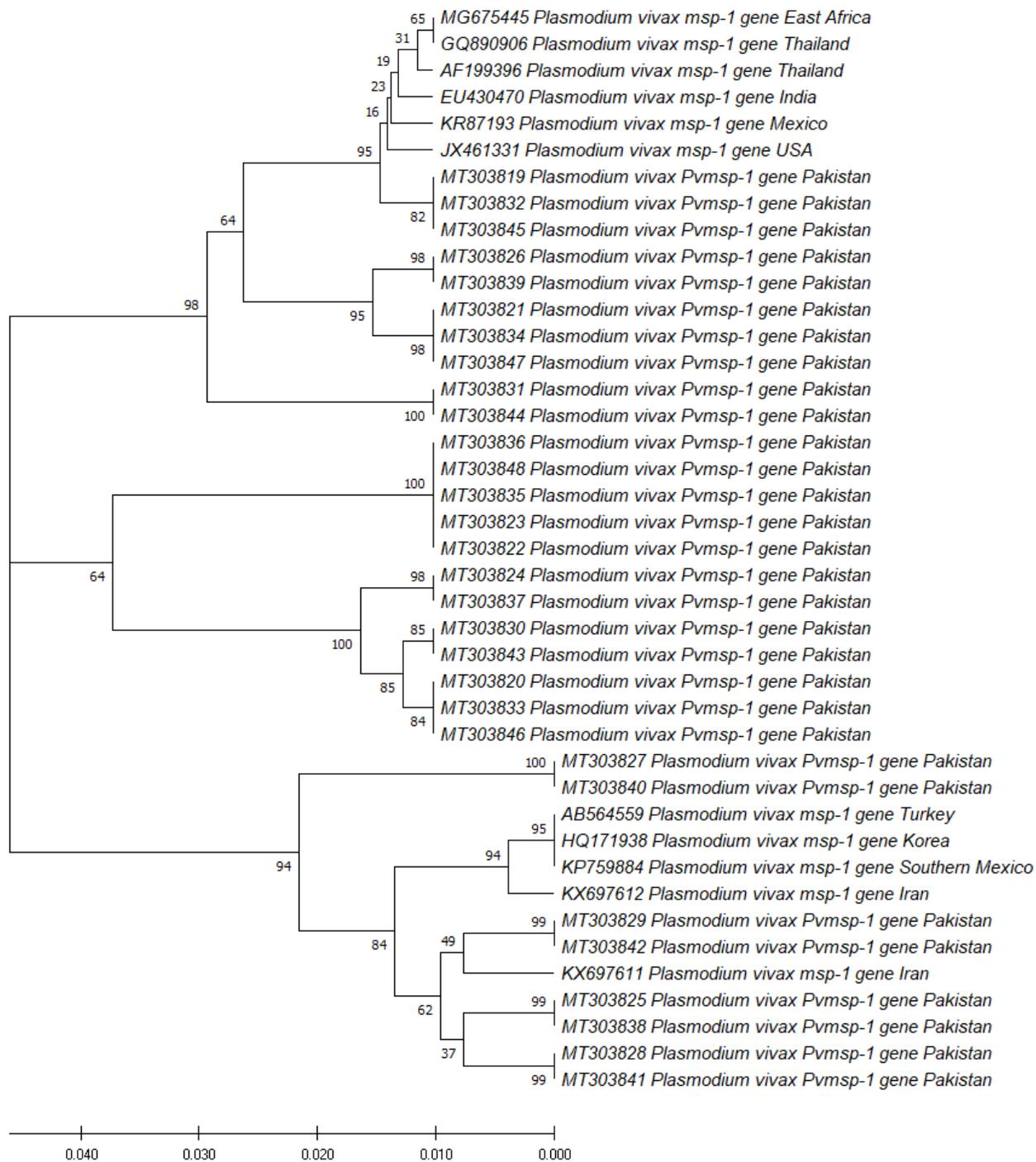


Figure 4

Analysis of polymorphic patterns of Pakistani PvMSP-1 gene



**Figure 5**

Phylogenetic relationships (based on PvMSP-1 gene) of *P. vivax* isolates from Pakistan. The Neighbor-Joining method was used to deduce evolutionary history with the branch length of =0.21993778.