Population Genetic Analyses Inferred a Limited Genetic Diversity in the *Pvama-1* of *Plasmodium Vivax* Isolates from Khyber Pakhtunkhwa Regions of Pakistan

Ibrar Ullah
Abdul Wali Khan University Mardan
Sahib Gul Afridi
Abdul Wali Khan University Mardan
Muhammad Israr
University of Swat
Hizbullah Khan
Chinese Academy of Sciences
Sulaiman Shams
Abdul Wali Khan University Mardan
Komal Zaib
Abdul Wali Khan University Mardan
Huong Giang Le
Gyeongsang National University
Jung-Mi Kang
Gyeongsang National University
Byoung-Kuk Na
Gyeongsang National University
Asifullah Khan (asifullah111@gmail.com)
Abdul Wali Khan University Mardan https://orcid.org/0000-0002-1444-4249

Research Article

Keywords: Plasmodium vivax, Apical membrane antigen-1, Genetic diversity, Khyber Pakhtunkhwa, Pakistan

Posted Date: January 19th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1229126/v1

Abstract

Background: *Plasmodium vivax* apical membrane antigen-1 (*pvama-1*) is an important vaccine candidate. Assessment of the genetic composition of *pvama-1* is preliminary important to better plan the vaccine designing strategies based on the antigen.

Methods: Blood samples of 84 vivax malaria patients from Khyber Pakhtunkhwa (KP) province of Pakistan were collected. The *pvama-1* domain 1 (DI) region was amplified and sequenced. The QC based sequences raw data filtration was done using DNASTAR package. The downstream population genetic analyses were performed using MEGA4, DnaSP, Arlequin v3.5 and network.5.

Results: The data analyses unveiled total 57 haplotypes of *pvama-1* DI among 84 KP *P. vivax* samples with majorly prevalent H-14 and H-5haplotypes. Limited to moderate pairwise genetic distinction was observed among the samples collected from different districts of KP. Likewise, no geography-specific genetic correlation was inferred among KP samples. In context of worldwide available data, the KP samples showed major genetic differentiation against Korean samples with *Fst* = 0.40915 (*P*-value =0.0001), while low distinction observed against India and Iranian samples. An excess of low frequency polymorphism and negative Tajima's D indicate purifying selection signatures across the *pvama-1* in KP *P.vivax* samples. Comparison of KP *pvama-1* DI with the reference *pvama-1* SalI (AF063138) unveiled total 09 KP samples-specific novel non-synonymous single nucleotide polymorphisms (nsSNPs) including several trimorphic and tetramorphic substitutions. Few of these nsSNPs were mapped within the B-cell predicted epitopic motifs of the *pvama-1* and, suggesting to modulate the immune response mechanisms.

Conclusion: The genetic composition of the *pvama-1* appeared uniform across the KP regions of Pakistan. However, the KP samples exhibited marked genetic distinction in context of worldwide samples. The information may worthy to understand genetic nature of Pakistani *P. vivax* and implicate in future in *pvama-1* vaccine designing.

Background

Malaria is an acute febrile infectious disease caused by vector-borne apicomplexan parasites of the genus *Plasmodium*. *P. vivax* and *P. falciparum* are predominant species responsible for malaria [1], among which *P. vivax* is the most widely distributed human malaria parasite endemic in tropical and subtropical countries of Asia, South Pacific, Central and South America, Middle East, and North Africa [2]. According to the latest WHO report, about 229 million cases and 40,900 deaths occur due to malaria in 2019 [3].

Treatment and control of *P. vivax* and *P. falciparum* have become a serious challenge due to drug resistance and lack of a proper vaccine. The wide-range distribution, antigenic variation, relapsing and co-infection led to a collective interest towards the development of effective vaccine against *P. vivax* [4]. The implementation of RTS,s/AS01 in three African countries started in 2019 and this considered as the most effective vaccine against malaria to date. Furthermore, the researchers reported the trial results of R21/Matrixs-M vaccine tested among children in Bukina Faso and reported it as the first malaria vaccine to meet the WHO's goal up to 77% [5]. Several surface antigens of *Plasmodium* species such as Apical Membrane Antigen 1 (AMA-1), Circumsporozoite proteins (CSP), Merozoite surface proteins (MSP) and Duffy binding protein (DBP) are reported as potent malaria vaccine candidates and many studies have been tried for the candidate antigens to develop productive vaccine [6].

The genetic composition assessments of vaccine candidates' loci are indispensable in modern-age to develop a malaria vaccine. Ample of evidences suggested that AMA-1 of *Plasmodium* species *pama-1* is one of the promising malaria vaccine candidate antigens [7]. AMA-1 is a type I integral membrane protein with molecular size of 83 kDa and is mainly expressed in the merozoite and sporozoite stages of *Plasmodium* parasites [8, 9]. The main biological function of AMA-1 is not clearly understood yet but its stage-specific expression and localization suggest its potent crucial role during invasion of erythrocytes and hepatocytes by malaria parasites [10–12]. The protein consists of cysteine rich ectodomain having three separated domains (i.e. Domain I, II, and III), a conserved cytoplasmic region and a transmembrane region [13]. The ectodomain of the protein is highly immunogenic and evokes natural immune responses in patients naturally exposed to *P. falciparum* and *P.vivax* [14.15.16.17]. The protein ama-1 is also reported to elicit the antibody production that effectively halt the invasion of erythrocytes by malaria parasite and hence confers protective immune responses [16, 18], suggesting a leading malaria vaccine candidate. The domain I of AMA-1 exhibits high level of genetic polymorphism and this region appears to be a major target of anti-AMA-1 protective antibodies [19–22]. It is therefore noteworthy to monitor genetic variation and polymorphism of the vaccine candidate antigen among global malaria isolates circulating in endemic areas, in order to design effective malaria vaccine [23]. Several studies about antigenic variation of *pvama-1* have been conducted in malaria endemic countries of the World [24–28]. However, limited studies are reported about *pvama-1* genetic features from Pakistan. Especially, no study till date is reported from remote malaria endemic regions in KP province of Pakistan. The current study was therefore pursued to evaluate the genetic composition of *pvama-1* in *P. vivax* isolates from KP regions of Pakistan (Figure 1).

Methods

Ethical approval, blood sample collections and DNA extraction

The current study was approved from Ethical committee of Abdul Wali Khan University Mardan (AWKUM/Biochem/Dept/Commit/eth/18). Blood samples were obtained from 100 consented patients tested positive for *P. vivax* who presented to different hospitals and private laboratories including Mardan, Swat, Buner, Hangu, Swabi, Kohat, Bannu, Timergara and Peshawar to cover the broad area of KP province, Pakistan (Figure 1). The region has the average annual rain fall of 384 mm during the two seasons from March to May and from August to November, during the time malaria incidences are peaked. The mean temperature in the region ranges from 20°C to 40°C. The blood samples from the patients were collected prior to treatment, spotted on filters, air-dried, and kept in individual sealed plastic bags at ambient temperature until use. The genomic DNA was extracted from the spotted blood samples using a QIAmp blood kit (Qiagen, CA, USA) according to manufacturer's instructions. The DNA samples were stored at -20°C.

Amplification and sequencing of pvama-1 DI

A DNA fragment flanking the DI of *pvama-1* was amplified by polymerase chain reaction (PCR) using the specific primers and amplification conditions described previously [21, 29]. The resulted PCR products were analyzed on 1.5% agarose gel, purified, and cloned into the T&A vector (Real Biotech Corporation, Banqiao City, Taiwan). Ligation mixture was transformed into *Escherichia coli* DH5α competent cells, and positive clones were selected by colony PCR. The nucleotide sequence of cloned insert was analyzed by automatic DNA sequencing with M13 forward and M13 reverse primers (Genotech Inc., Daejoen, Korea).The raw data was filtered for quality assessment using DNASTAR Lasergene package.

Population Genetic Analyses

The DnaSP v6.12 software package was used to estimate parsimony informative sites and haplotypes composition from the KP *pvama-1* sequences [30]. The population genetics statistics including pairwise fixation index (*Fst*), analysis of molecular variance (AMOVA), haplotype frequencies, and nucleotide diversity based on Nei's net distance (*DA)* were computed using Arlequin v3.5 [31]. The haplotype paradigm was generated by the median-joining method implemented in NETWORK 5.0 [32]. The data generated from the median-joining calculation was subjected to a refine network plot.

Dna Sequence Polymorphisms Analysis

The sequence generated in study was further analyzed in comparison with the Genbank- deposited *pvama-1* sequences from China Myanmar Boarder (KX495505- KX495577), Iran (KF422636.1- KF422681.1), Korea (KM230319.1 - KM230384.1), Myanmar (FJ157248.1- FJ157285.1), Papua New Guinea (PNG) (KC702402.1- KC702501.1), Sri Lanka (EF218679.1 - EF218701.1), Venezuela (EU346015.1 - EU346087.1), Thailand (FJ784891.1 - FJ784990.1), and India (EU282774.1- EF025196.1). The DnaSP v6.12 [30] was employed to estimate the parsimony informative sites, total number of mutations, pairwise nucleotide diversity (π), segregating sites (S), haplotypes diversity (Hd), haplotypes (H) composition of the sequences. The Tajima's D, Fu and Li's *D** and *F** indices were calculated via a sliding window method [33].

Recombination And Linkage Disequilibrium

The recombination parameter (R) between adjacent nucleotides per generation and the minimum number of recombination events (Rm) were calculated by using the DnaSPv6.12 [30]. Likewise, the linkage disequilibrium (LD) was estimated between the various polymorphic sites based on the R^2 index via v6.12 [33].

Functional Prediction Of Nssnps

The BepiPred - 2.0 [34] servers was used for prediction of Linear B-cell epitopes of *Pvama-1*with a threshold value (0.5) score. The higher score shows the higher binding affinity. The nsSNPs mapping within the top predicted epitopes of *pvama-1* was checked. The IURs regions and RBC binding sites within the *Pvama-1* have previously been predicted [35, 25] and their annotations were utilized to check the novel nsSNPs, being identified in current study, mapping within these motifs of *pvama-1*.

Results

Genetic polymorphic features of KP pvama-1

The 416 bp sequences of *pvama-1* flanking the DI (322-737 nucleotide positions) were successfully amplified in 84 *P. vivax* KP samples. Comparison of the sequences to the reference sequence, Sal I (AF063138), revealed that the 84 Pakistani *pvama-1* classified into 57 haplotypes (Figure 2). Analysis of the KP *pvama-1* sequences compared to reference Sal I identified a large numbers of single nucleotide polymorphisms (SNPs) in KP *pvama-1* sequences. Among these, 68 were non-synonymous SNPs (nsSNPs) causing amino acid substitutions including 53 dimorphic, 10 trimorphic, 3 tetramorphic, and 2 pentamorphic. The two pentamorphic amino acid changes were R112K/T/E/S and S228D/N/R/K, The ten trimorphic amino acid substitution included N132D/G, A141E/G, E145A/G, K190E/Q, T191K/P, A199T/V, S209G/C, P210S/L, P223L/S, and V233L/P. While the three tetramorphic amino acid changes are K120R/S/G, E189N/K/G, and E227V/K/G. These amino acid substitutions were observed at varied frequencies in the KP samples. Among the 68 nsSNPS, the 59 have previously been reported in literature for *P. vivax* isolates from different geographical origins. However, the rest of 9 nsSNPs were specific to KP samples set (Table 1). These nsSNPs were observed at low frequencies (1.19%). Few nsSNPs such as K120R, N132D, L140I, A141E, K190E, E227V, and S228D were commonly observed with high frequency in KP, as well as some other continental *pvama-1* sequences (Table 1). The KP *pvama-1* showed overall haplotype diversity (Hd) of 0.978±0.008. A total of 62 segregating sites (S) and 67 mutations were identified for the samples. The Fu and Li D's test inferred the effect of natural selection on genetic composition. The negative values of Tajima's D implied an excess of low frequent polymorphism, suggesting the population size expansion (Table 2).

Table 1
The sixty eights SNPs identified in KP *P. vivax* samples in comparison to the reference *pvama1* Sal I (AF063138) sequence

1	R112K/T/E/S	15	L140I*	29	V170A	43	E201G	57	N226D
2	P113S	16	A141E*	30	M171T	44	M203T	58	E227V*
3	G117R#	17	N142D	31	A172T	45	G204D	59	S228D*
4	D118N	18	K144T	32	V184A	46	R206G	60	N231D#
5	Q119H	19	E145A/G	33	K188N	47	S209G#	61	V233L/P
6	K120R*	20	K148Q	34	E189N/K/G	48	P210S/L*	62	Y234H#
7	F126S	21	D149N	35	K190E*	49	A212V#	63	L235S
8	N130K	22	M153T	36	T191K/P	50	N214S	64	N238D#
9	A131T	23	I159T	37	C192R	51	R215T	65	R240C
10	N132D/G*	24	A160T	38	H193Y	525	V218L	66	N241D
11	D133N	25	L161V	39	M194V	53	F221L	67	D242E
12	H134R	26	C162R#	40	Y196H#	54	K222N	68	W234R
13	S136T	27	A166P	41	S198P	55	P223L#		
14	T139A	28	A167P	42	A199T/V	56	K225E		

* The common nsSNPs identified in KP and other*P. vivax* samples deposited in Genbank, NCBI.

#Novel amino acid polymorphism in KP sample acquired from different region and high and low frequency observed.

Table 2
The neutrality test and genetic polymorphism estimation for *Pvama-1* domain-1 DNA sequences of KP-Pakistan and global samples

Countries	Total isolates(n)	Segregating sites (S)	Singleton variable sites	Parsimony informative sites	Total no of Mutation	K	H	Hd± SD	π ± SD	Tajimas D	D*(F&L)	F*(F&L)
KP (Pakistan)	84	62	41	21	67	7.335	57	0.978 ±0.008	0.01763 ±0.00084	-1.490 *P* > 0.10	-5.167 *P*< 0.02	(-4.418) *P*< 0.02
China-Myanmar boarder	73	22	3	19	24	6.249	25	0.914 ±0.021	0.01502 ±0.00077	0.81996 *P*> 0.10	0.00217 *P*> 0.10	0.3622 *P*> 0.10
Iran	80	19	4	17	23	7.101	30	0.975 ±0.010	0.01707 ±0.00072	1.17202 *P*> 0.10	0.43585 *P*> 0.10	0.8179 *P*> 0.10
Korea	66	23	4	19	23	4.205	15	0.782 ±0.047	0.01011 ±0.00111	-0.40425 *P*> 0.10	0.32761 *P*> 0.10	0.0739 *P*> 0.10
Myanmar	38	45	23	22	47	8.580	37	0.999 ±0.006	0.02063 ±0.00091	-0.83341 *P*> 0.10	-1.88701 *P*> 0.10	-1.8082 *P*> 0.10
PNG	100	21	2	19	22	6.10667	28	0.941 ±0.009	0.01468 ±0.00057	1.28215 *P*> 0.10	0.95445 *P*> 0.10	1.2894 *P*> 0.10
Sri lanka	23	15	3	12	15	0.01024	9	0.858 ±0.047	0.01024 ±0.00158	0.17272 *P*> 0.10	0.44859 *P*> 0.10	0.4266 *P*> 0.10
Venezuela	73	12	0	12	13	4.718	12	0.847 ±0.019	0.01134 ±0.00044	2.15646 *P*< 0.05	1.43284 *P*< 0.05	2.0629 *P*< 0.02
Thailand	100	18	1	17	23	6.6000	34	0.919 ±0.015	0.01587 ±0.00057	1.43285 *P*> 0.10	0.59998 *P*> 0.10	1.0963 *P*> 0.10
India	59	23	4	19	26	7.10286	41	0.980 ±0.008	0.01707 ±0.000	0.86316 *P*> 0.10	0.21989 *P*> 0.10	0.5381 *P*> 0.10

S: number of polymorphic sites (Segregating sites), K: average number of pair-wise nucleotide differences, H: haplotype, Hd: haplotype diversity, π: observed average pair-wise nucleotide diversity, D^* (F&L): Fu and Li's D^* value, F^* (F&L): Fu and Li's F^* value. P value < 0.05 is considered as significant difference.

Haplotype Networking Analysis

Total of 57 KP *pvama-1* haplotypes were identified for the 84 isolates sequences with the haplotype diversity (Hd) of 0.978 (±0.008). The haplotype (H14) was identified with high frequency and shared among samples collected from six different KP districts including, Kohat, Hungo, Buner, Swat, Timergara and Bannu. The haplotype (H5) was identified as second predominant haplotype shared among samples collected from five different KP districts (i.e. Mardan, Swat, Hungo, Bannu and Kohat). The haplotype (H3) was also found with the highest frequency in samples collected from Swat, Mardan, Peshawar and Bannu. The pairwise AMOVA (Analysis of Molecular Variance) inferred the pairwise distances among haplotypes. The haplotype-53, i.e. predominant in Peshawar samples, was identified as distinct and showed significant genetic differentiation against the haplotype-6 and haplotype-55. The H-6 and H-55 were identified with high frequency in samples collected from Mardan and Peshawar regions respectively. The size of each node in haplotype network plot indicates the frequency of a particular haplotype. The length of the line between nodes is proportion to the number of nucleotide substitutions, composing the haplotypes. The majorly shared haplotypes of KP samples, collected from different districts appeared on shared nodes, however, some haplotypes for samples collected from Timergara, Peshawar, Kohat and Hungo districts occupied distinct nodes in the network plot which inferred their distinctive features (Figure 3).

The functional impact of the novel nsSNPs was assessed with respect to amino acids substitution in the IURs motifs of *pvama-1*. This region considered important in vaccine designing and diagnosis based on *pvama-1* [30]. None of the residue substituted due to KP samples specific novel SNPs are mapped within the disordered regions of the *pvama-1*. The result showed that two SNPs i.e. M171T, V172T mapped within the IURs motifs and four SNPs i.e. R240C, N241D, D242E and W243R were detected in RBC binding region, while most of the amino acid changes caused by nsSNPs were mapped within the predicted B-cell epitopes of *pvama-1*. Among novel nsSNPs, the G117R, S209G, A212V, and P223L, being identified in current study, are mapped within the epitopic region of *pvama-1* and predicted to modulate the possible host immune response. The top lead epitopes were predicted based on BepiPred - 2.0 threshold score of > 0.5. The region comprises of 240-254 residues have four SNPs (i.e. R240C, N241D, D242E, and W243R) that mapped within the B-cell epitopes as well as RBC binding sites.

Recombination and linkage disequilibrium analysis

The recombination events across *pvama-1* and decline of LD index R^2 with the increase of nucleotide distance was identified for the KP samples. This speculate high meiotic recombination events across the *pvama-1* in the KP samples (Figure 4). The R value for KP samples were observed higher compare to those of East Asian (i.e. China Myanmar boarder, and Korea), South Asia (Sri lanka) samples, while lower than those of the Myanmar samples. The lowest R value for Myanmar samples depicts opportunity of high multiclonal infections, cross fertilization and recombination. The higher values of recombination and rapid LD decay observed in KP and some other geographical samples indicate high meiotic recombination in *pvama-1*, supporting the recombination as a possible factor to provoke genetic diversity (Table 3).

Table 3
Comparison of different estimates of recombination in *Pvama-1*(Domain-1) among KP and global *P. vivax* samples

	R^a	R^b	Rm
KP, Pakistan	0.0892	37	5
China Myanmar	0.0506	21	5
Iran	0.0846	35.1	6
Korea	0.0022	0.9	2
Myanmar	0.1940	8.6	6
PNG	0.0513	21.3	5
Sri lanka	0.0154	6.4	4
Venezuela	0.0381	15.8	4
Thailand	0.0879	36.5	5
India	0.1210	50.2	6
Abbreviations:			
R^a recombinant parameter between adjacent sites, R^b recombinant parameter for the whole genes, Rm minimum number of recombinant events.			

Nucleotide diversity across pvama-1 in context of global isolates

The sequences of KP isolates (*n* = 84) were compared to the global *pvama-1* sequences deposited in Genbank. The values of *K* and π observed for KP sequences were more or less similar to previously reported sequences from Iran and India, while differentiated from rest global sequences (Table 2). The fixation index *Fst* statistic was used to assess the genetic differentiation across *pvama-1* gene among KP samples collected from different regions as well as in context of global samples. The pairwise analysis inferred genetic distinction of samples collected from Swabi district compare to rest of the KP regions. The top *Fst* differentiation was detected between the Bannu and Swabi isolates (*Fst* = 0.16258, *P*-value = 0.00977), followed by Swabi and Kohat (*Fst* = 0.12932; *P*-value = 0.04199) samples. The lowest *Fst* depicted between Swat and Bannu groups (*Fst* = -0.07427; *P*- value = 0.96973), followed by Swat and Hungo groups (*Fst*= -0.06635; *P*-value = 0.89551) (Figure 5A). In context of global samples, marked genetic distinction inferred for KP samples compare to India, Iran, Thailand, Sri-Lanka, Korea, Venezuela, Myanmar, PNG, and China-Myanmar. High genetic differentiation was observed between KP and Korean samples (*Fst* = 0.40915). The Korean samples showed significant genetic distinction in pairwise comparison to rest of the global samples. Meanwhile, least genetic differentiation was observed among KP, Iranian, and Indian samples (Figure 5B). The highest pairwise net number of nucleotide variation (DA) and mean pairwise differences (π_{xy}) was observed between Bannu and Swabi samples (Figure 5a), i.e. congruent to *Fst* analysis. The highest within population genetic differentiation (π) was found for Korean samples followed by South East Asian samples (Figure 5b). Pearson correlation plot showed relationship among KP, Sri lanka, Iran, India and Myanmar samples, congruent to pairwise *Fst* (Figure S1). The plot showed correlation among the populations in hierarchical order. However, the Korean samples showed high genetic distinction in term of *Fst* value, probably due to geographical separation. Distinction for Korean samples also depicted in correlation plot. Likewise, the PCA plot also unveiled the samples clustered with more or less same fashion (Figure S2).

The AMOVA test was performed to determine genetic differentiation at single and multiple loci because of variation within a population group as well as between population groups. The AMOVA analysis depicted that genetic diversity in KP samples set mainly arose due to within population differentiation i.e. 100.24%, instead of among groups differentiation (-0.24%). This indicates limited or no genetic differentiation in KP samples despite their geographical distinction across the KP region of Pakistan (Table 4).

Table 4
AMOVA-based genetic differentiation analysis across *Pvama-1* (domain-1) in samples acquired from different districts of KP, Pakistan

Source of variation	d.f	Sum of squares deviation	estimates of Variance components	Percentage of variation	P-value
Among populations	8	28.742	-0.00888 Va	-0.24%	P= 0.50556
Within populations	75	275.650	3.67534 Vb	100.24%	P= 0.50556
Total	83	304.393	3.66646		
Local F*st* value = -0.00242					

Discussion

The comprehensive knowledge about the antigenic variants in *Plasmodium* parasites is perquisite to design effective vaccine strategies workable in different endemic regions [36]. The current study aimed to analyze genetic composition of *pvama-1*, a leading malaria vaccine candidate antigen, in *P. vivax* isolates from different districts of KP, Pakistan.

The southern and northern regions of KP province of Pakistan are distinct with respect to geographical and environmental perspectives. However, limited genetic diversity of *P. vivax pvama-1* was identified in the current study, suggesting no significant genetic heterogeneity between the isolates from southern or northern KP. The low genetic diversity across the DI domain of *pvama-1* in KP region might be due to low endemicity of *Plasmodium* genotypes [37], as the low endemic region is generally characterized with limited parasitic genetic diversity [38]. The low transmission and endemicity of *P. vivax* in the KP might have been provoked due to active malaria control program in these regions from last several years. Additionally, the limited genetic diversity speculates that malaria infection in KP region might be monoclonal and may combat with a single type of *pvama-1* based immune vaccine. The pair wise genetic analyses indicate the close genetic feature of KP samples to South/Central Asian samples from India and Iran regions. This might be due to close geographical contacts among these countries. The negative values of Tajima's D imply an excess of low frequency polymorphism and indicate the population size expansion. Besides, this clue toward stronger diversifying selection and host immune selection signature across the *pvama-1* in KP samples. The Tajima's D indicates the balancing selection event across the *pvama-1* in global samples however it was not statistically significant for the samples set in the current study.

The analyses of KP samples in context of global samples inferred unique genetic features and 9 KP specific nsSNPs were identified in the samples. The genetic polymorphisms (nsSNPs) identified in the current study were further analyzed with respect to their possible functional consequences in the predicted RBC binding sites, B cell epitopes, and IURs regions of *pvama-1*. Several nsSNPs were found to be located at the predicted RBC- binding sites, B-cell epitopes and IURs region of *pvama-1*. However, most of these nsSNPs were mapped at the B-cell predicted epitopic motifs, indicating a high degree of balancing natural selection across the B-cell epitopes region of PvAMA-1. The protein structure affected by amino acid changes due to these nsSNPs may affect the physicochemical perspectives of the PvAMA-1 that might help the parasite to escape from host protective immunity. The IURs play an important role in molecular recognition, assembling and protein modification [39]. The *pvama-1* IURs are indispensable for attachment and invasion of the parasite into red blood cell [40]. Several *pvama-1* SNPs were detected in PvAMA-1 IURs region. However, none of KP samples-specific nsSNP, identified in current study, mapped within the IURs region of *pvama-1.*

Conclusion

The *pvama-1* is considered as the promising candidates for malaria vaccine which targeting the blood stage of *P. vivax* isolates. The partial DNA sequencing and analysis unveiled limited genetic diversity of *pvama-1* across the KP regions. This inferred that *pvama-1* based vaccine might be promising to effectively combat and contribute in malaria eradication throughout the KP province of Pakistan.

Declarations

Ethics approval and consent to participate

The study approved from ethical review committee of Abdul Wali Khan University Mardan, Pakistan. Besides, the study pursued with informed written consent from all the study participants.

Consent for publication

All authors read and agreed to publish the study.

Availability of supplementary data

All the relevant data is provided in supplementary figures and tables. The newly generated sequences data deposited to Genbank, NCBI.

Competing interests

The authors declare that they have no competing interests.

Funding

The study pursued without any proper grant allotment.

Authors' contributions

IU, S.G.A and A.K conceived the basic research idea. IU, M.I S.S contributed in samples collection. IU, H.K, K.Z, A.K performed data analyses. H.G.L, J-M.K and B-K.N contributed in PCR amplification and DNA sequencing experiments. I.U, S.G.A and A.K prepared the initial manuscript draft. S.G.A, B-K.N and A.K finalized and reviewed the draft. All authors have read and agreed to the published version of the manuscript.

Acknowledgement

The authors acknowledge the supportive laboratory personnel, who contributed in samples collection and diagnosis.

References

1. McFadden GI. Plasmodium: more don'ts. Trends Parasitol. 2019;35:4–6.
2. Mendis K, Sina BJ, Marchesini P, Carter R, et al. The neglected burden of *Plasmodium vivax* malaria. T Am J Trop Med Hyg. 2001;64(1_suppl):97–106.
3. WHO. World Malaria Report 2020: 20 years of global progress and challenges. Geneva: World Health Organization; 2020. p. 2020.
4. Herrera S, Corradin G, Arévalo-Herrera M, et al. An update on the search for a *Plasmodium vivax* vaccine. Trends Parasitol. 2007;23:122–8.
5. Datoo MS, Natama MH, Somé A, Traoré O, Rouamba T, Bellamy D, Yameogo P, Valia D, Tegneri M, Ouedraogo F, Soma R, et al. Efficacy of a low-dose candidate malaria vaccine, R21 in adjuvant Matrix-M, with seasonal administration to children in Burkina Faso: a randomised controlled trial. The Lancet. 2021;397:1809–18.
6. Dobaño C, Sedegah M, Rogers WO, Kumar S, Zheng H, Hoffman SL, Doolan DL, et al. Plasmodium: mammalian codon optimization of malaria plasmid DNA vaccines enhances antibody responses but not T cell responses nor protective immunity. Exp Parasitol. 2009;122:112–23.
7. Thera MA, Coulibaly D, Kone AK, Guindo AB, Traore K, Sall AH, et al. Phase 1 randomized controlled trial to evaluate the safety and immunogenicity of recombinant Pichia pastoris-expressed *Plasmodium falciparum* apical membrane antigen 1 (PfAMA1-FVO [25-545]) in healthy Malian adults in Bandiagara. Malar J. 2016;15:1–11.
8. Healer J, Murphy V, Hodder AN, Masciantonio R, Gemmill AW, Anders RF, Cowman AF, Batchelor A, et al. Allelic polymorphisms in apical membrane antigen-1 are responsible for evasion of antibody-mediated inhibition in *Plasmodium falciparum*. Mol Microbiol. 2004;52:159–68.
9. Silvie O, Franetich JF, Charrin S, Mueller MS, Siau A, Bodescot M, et al. A role for apical membrane antigen 1 during invasion of hepatocytes by *Plasmodium falciparum* sporozoites. J Biol Chem. 2004;279:9490–6.
10. Peterson MG, Marshall VM, Smythe JA, Crewther PE, Lew A, Silva AN, Anders RF, Kemp DJ. Integral membrane protein located in the apical complex of *Plasmodium falciparum*. Mol Cell Biol. 1989;9:3151–4.
11. Mitchell GH, Thomas AW, Margos G, Dluzewski AR, Bannister LH. Apical membrane antigen 1, a major malaria vaccine candidate, mediates the close attachment of invasive merozoites to host red blood cells. Infect Immun. 2004;72:154–8.
12. Yap A, Azevedo MF, Gilson PR, Weiss GE, O'Neill MT, Wilson DW, et al. Conditional expression of apical membrane antigen 1 in *Plasmodium falciparum* shows it is required for erythrocyte invasion by merozoites. Cell Microbiol. 2014;16:642–56.
13. Hodder AN, Crewther PE, Anders RF, et al. Specificity of the protective antibody response to apical membrane antigen 1. Infect Immun. 2001;69:3286–94.
14. Udhayakumar V, Kariuki S, Kolczack M, Girma M, Roberts JM, Oloo AJ, et al. Longitudinal study of natural immune responses to the *Plasmodium falciparum* apical membrane antigen (AMA-1) in a holoendemic region of malaria in western Kenya: Asembo Bay Cohort Project VIII. Am J Trop Med Hyg. 2001;65:100–7.
15. Cortés A, Mellombo M, Masciantonio R, Murphy VJ, Reeder JC, Anders RF. Allele specificity of naturally acquired antibody responses against *Plasmodium falciparum* apical membrane antigen 1. Infect Immun. 2005;73:422–30.
16. Rodrigues MH, Rodrigues KM, Oliveira TR, Cômodo AN, Rodrigues MM, Kocken CH, et al. Antibody response of naturally infected individuals to recombinant *Plasmodium vivax* apical membrane antigen-1. Int J Parasitol. 2005;35:185–92.
17. Moncunill G, Aponte JJ, Nhabomba AJ, Dobano C, et al. Performance of multiplex commercial kits to quantify cytokine and chemokine responses in culture supernatants from *Plasmodium falciparum* stimulations. PloS one. 2013;8:e52587.
18. Gentil EC, Damgaard A, Hauschild M, Finnveden G, Eriksson O, Thorneloe S, Kaplan PO, Barlaz M, Muller O, Matsui Y, Ii R, et al. Models for waste life cycle assessment: Review of technical assumptions. Waste Manag. 2010;30:2636–48.
19. Escalante AA, Grebert HM, Chaiyaroj SC, Magris M, Biswas S, Nahlen BL, et al. Polymorphism in the gene encoding the apical membrane antigen-1 (AMA-1) of *Plasmodium falciparum*. X. Asembo Bay Cohort Project. Mol Biochem Parasitol. 2001;113:279–87.
20. Polley SD, Conway DJ. Strong diversifying selection on domains of the *Plasmodium falciparum* apical membrane antigen 1 gene. Genetics. 2001;158:1505–12.
21. Moon SU, Na BK, Kang JM, Kim JY, Cho SH, Park YK, et al. Genetic polymorphism and effect of natural selection at domain I of apical membrane antigen-1 (AMA-1) in *Plasmodium vivax* isolates from Myanmar. Acta Trop. 2010;114:71–5.
22. Arnott A, Mueller I, Ramsland PA, Siba PM, Reeder JC, Barry AE. Global population structure of the genes encoding the malaria vaccine candidate, *Plasmodium vivax* apical membrane antigen-1 (Pvama1). PLoS Negl Trop Dis. 2013;7:e2506.

23. Volkman SK, Neafsey DE, Schaffner SF, Park DJ, Wirth DF, et al. Harnessing genomics and genome biology to understand malaria biology. Nat Rev Genet. 2012;13:315–28.
24. Gunasekera AM, Wickramarachchi T, Neafsey DE, Ganguli I, Perera L, Premaratne PH, Hartl D, Handunnetti SM, Udagama-Randeniya PV, Wirth DF, et al. Genetic diversity and selection at the *Plasmodium vivax* apical membrane antigen-1 (Pvama-1) locus in a Sri Lankan population. Mol Biol Evol. 2007;24:939–47.
25. Zhu X, Zhao P, Wang S, Liu F, Liu J, Wang J, Yang Z, Yan G, Fan Q, Cao Y, Cui L, et al. Analysis of Pvama1 genes from China-Myanmar border reveals little regional genetic differentiation of *Plasmodium vivax* populations. Parasites vectors. 2016;9:1–2.
26. Heidari A, Keshavarz H, Hajjaran H, Ebrahimi SM, Kabir K, Naseri MH, et al. Genetic variation and selection of domain I of the *Plasmodium vivax* apical membrane antigen-1 (Pvama1) gene in clinical isolates from Iran. Iranian J Parasitol. 2013;8:536.
27. Rajesh V, Elamaran M, Vidya S, Gowrishankar M, Kochar D, Das A, et al. *Plasmodium vivax*: genetic diversity of the apical membrane antigen-1 (AMA-1) in isolates from India. Exp Parasitol. 2007;116:252–6.
28. Dias S, Somarathna M, Manamperi A, Escalante AA, Gunasekera AM, Udagama PV, et al. Evaluation of the genetic diversity of domain II of *Plasmodium vivax* Apical Membrane Antigen 1 (PvAMA-1) and the ensuing strain-specific immune responses in patients from Sri Lanka. Vaccine. 2011;29:7491–504.
29. Figtree M, Pasay CJ, Slade R, Cheng Q, Cloonan N, Walker J, Saul A, et al. *Plasmodium vivax* synonymous substitution frequencies, evolution and population structure deduced from diversity in AMA 1 and MSP 1 genes. Mole Biochem Parasitol. 2000;108:53–66.
30. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A, et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol. 2017;34:3299–302.
31. Excoffier L, Laval G, Schneider S, et al. Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinform. 2005;1:117693430500100003.
32. Bandelt HJ, Forster P, Röhl A, et al. Median-joining networks for inferring intra-specific phylogenies. Mol Biol Evol. 1999;16:37–48.
33. Librado P, Rozas J. DnaSP v5: software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;23:1451–2.
34. Larsen JE, Lund O, Nielsen M, et al. Improved method for predicting linear B-cell epitopes. Immunome research. 2006;2:1–7.
35. Zakeri S, Sadeghi H, Mehrizi AA, Djadid ND, et al. Population genetic structure and polymorphism analysis of gene encoding apical membrane antigen-1 (AMA-1) of Iranian *Plasmodium vivax* wild isolates. Acta Trop. 2013;126:269–79.
36. Genton B, Reed ZH, et al. Asexual blood-stage malaria vaccine development: facing the challenges. Curr Opin Infect Dis. 2007;20:467–75.
37. Heidari A, Keshavarz H, Rokni MB, Jelinek T, et al. Genetic diversity in merozoite surface protein (MSP)-1 and MSP-2 genes of *Plasmodium falciparum* in a major endemic region of Iran. Korean J Parasitol. 2007;45:59.
38. Ghanchi NK, Mårtensson A, Ursing J, Jafri S, Bereczky S, Hussain R, Beg MA, et al. Genetic diversity among *Plasmodium falciparum* field isolates in Pakistan measured with PCR genotyping of the merozoite surface protein 1 and 2. Malar J. 2010;9:1–6.
39. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradović Z, et al. Intrinsic disorder and protein function. Biochemistry. 2002;41:6573–82.
40. Feng ZP, Zhang X, Han P, Arora N, Anders RF, Norton RS, et al. Abundance of intrinsically unstructured proteins in *P. falciparum* and other apicomplexan parasite proteomes. Mol Biochem Parasitol. 2006;150:256–67.

Figures

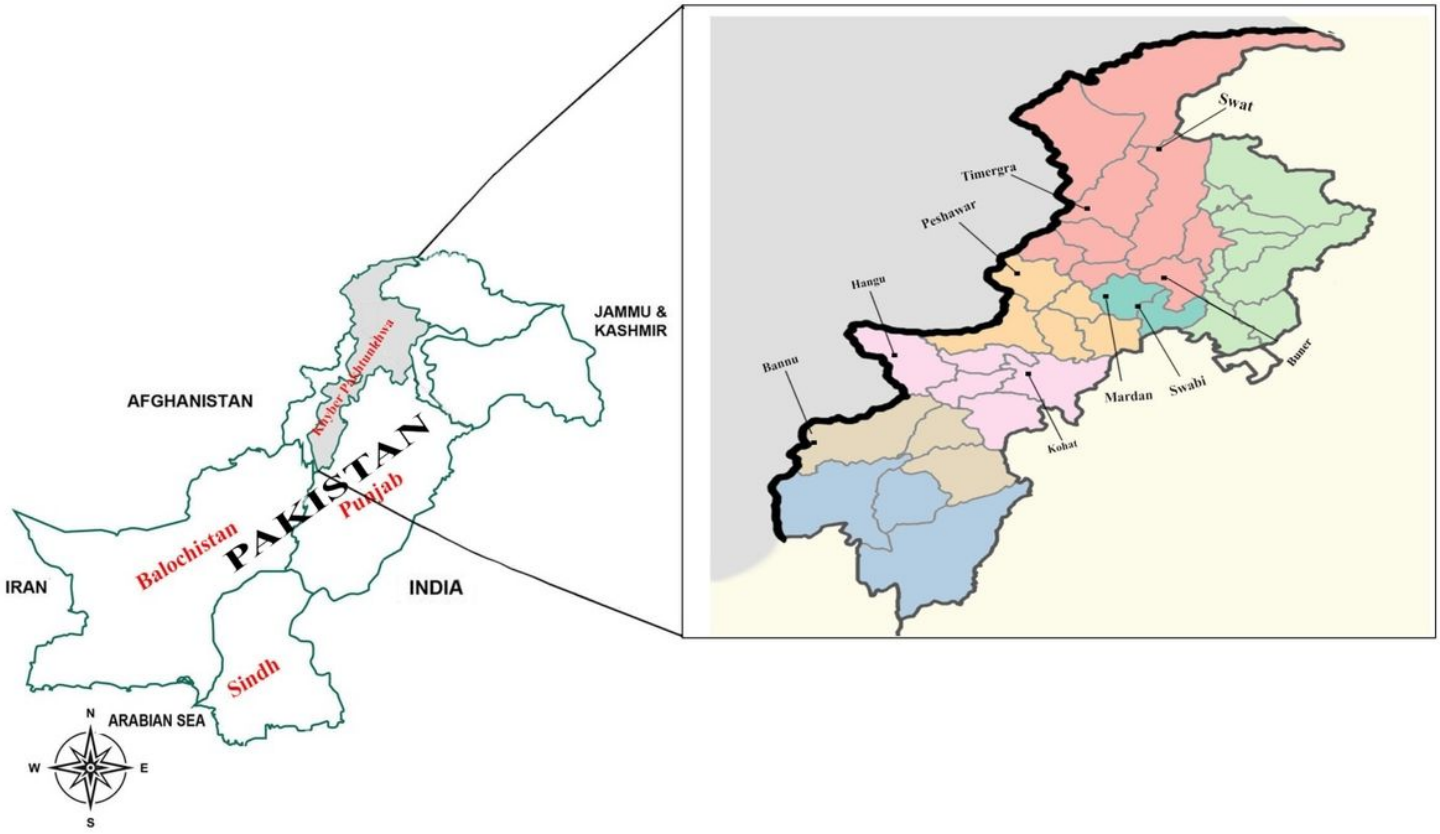

Figure 1

Map of different districts of Khyber Pakhtunkhwa (KP), Pakistan.

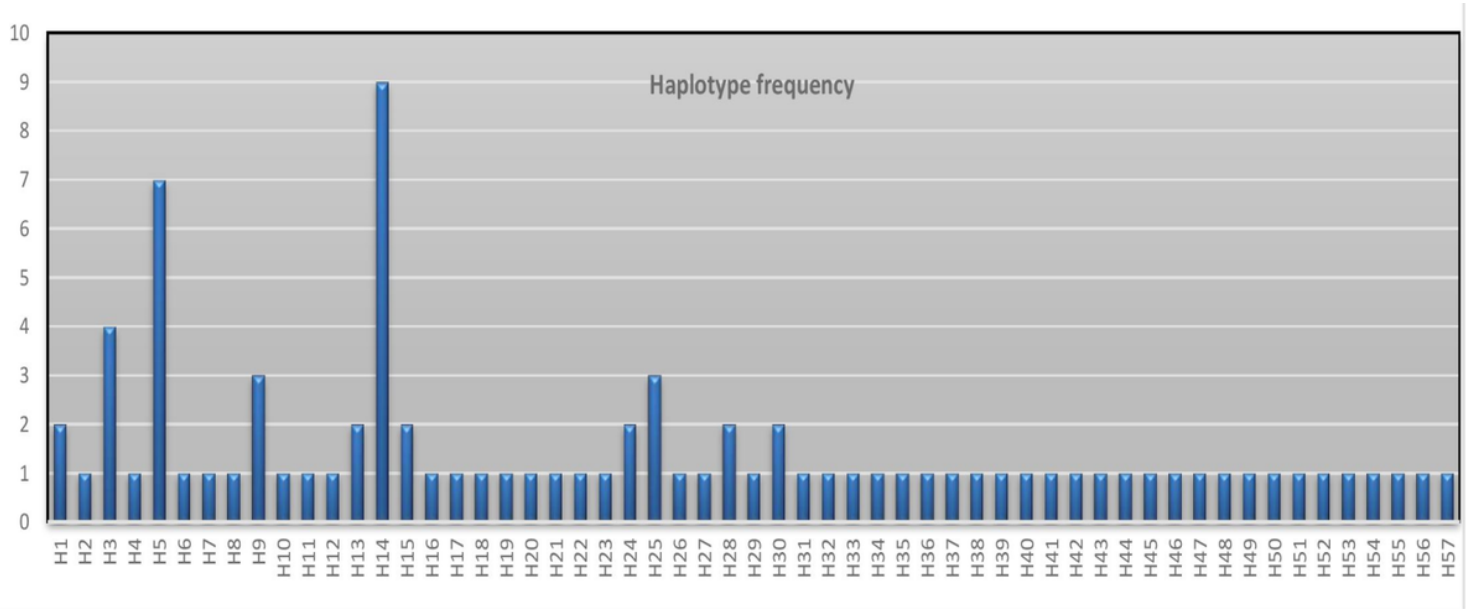

Figure 2

Frequency of different haplotypes of *Pvama-1* in KP region, Pakistan

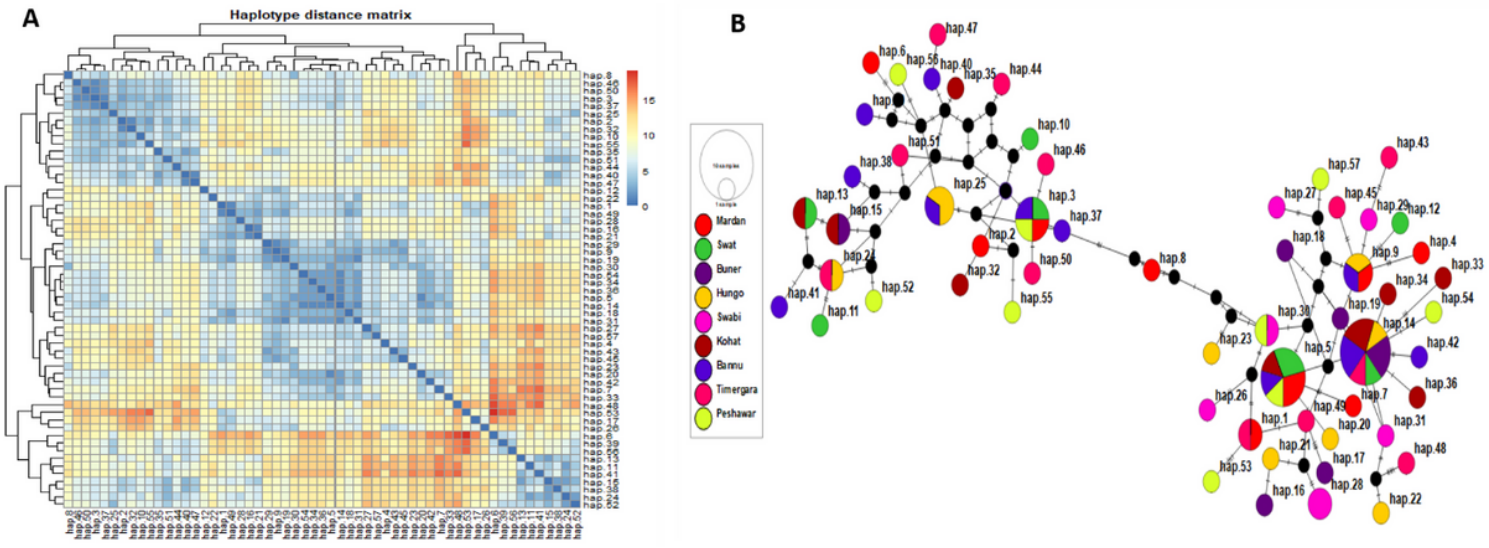

Figure 3

[A] The pairwise haplotypes difference between the Pakistani pvama-1populations. [B] The Network plots was generated by PopArt to detect haplotypes, in which a circle represents each haplotype, and the size of each circle is proportional to the number of individual samples of haplotype. The lines connecting them reflected the distance between n haplotypes. While the colored shows distinction between population groups.

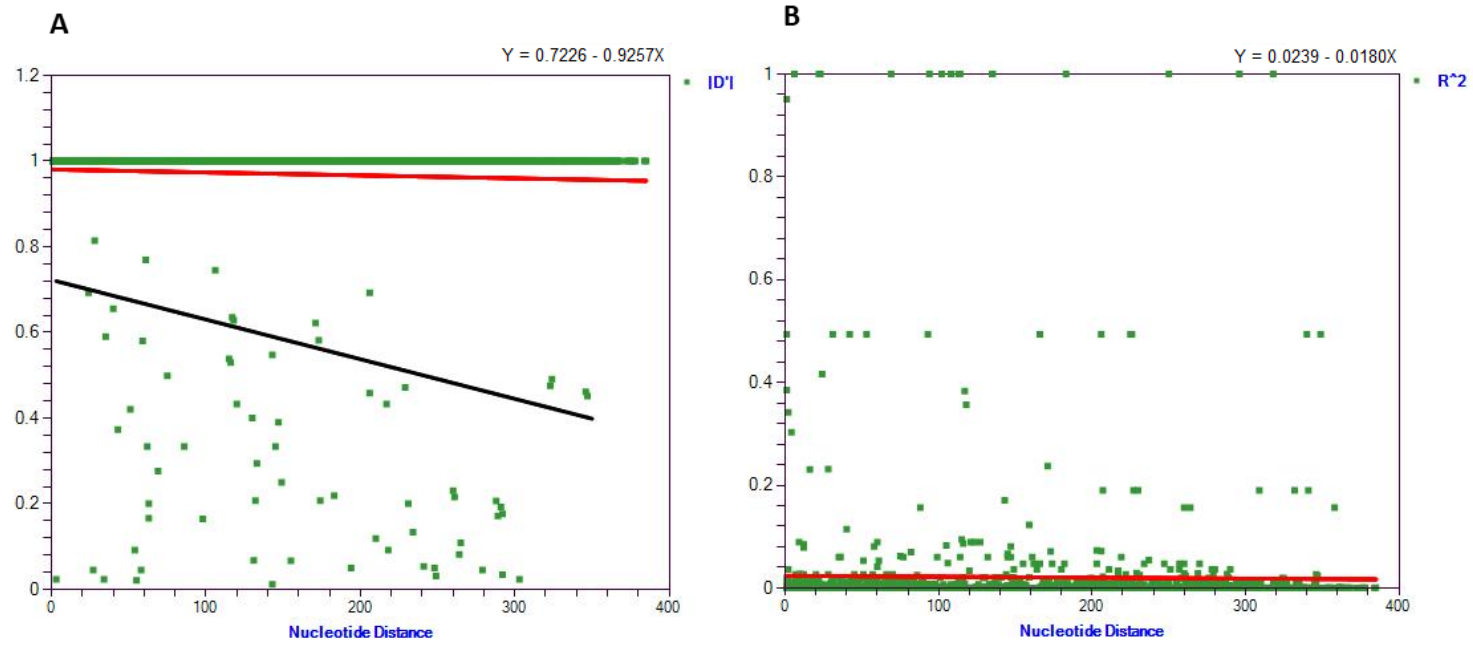

Figure 4

Patterns of linkage disequilibrium (LD) based on the linear regression line: (A, B) illustrate the relationships between the distance between loci (expressed in nucleotides) and |D'| and r^2, respectively.

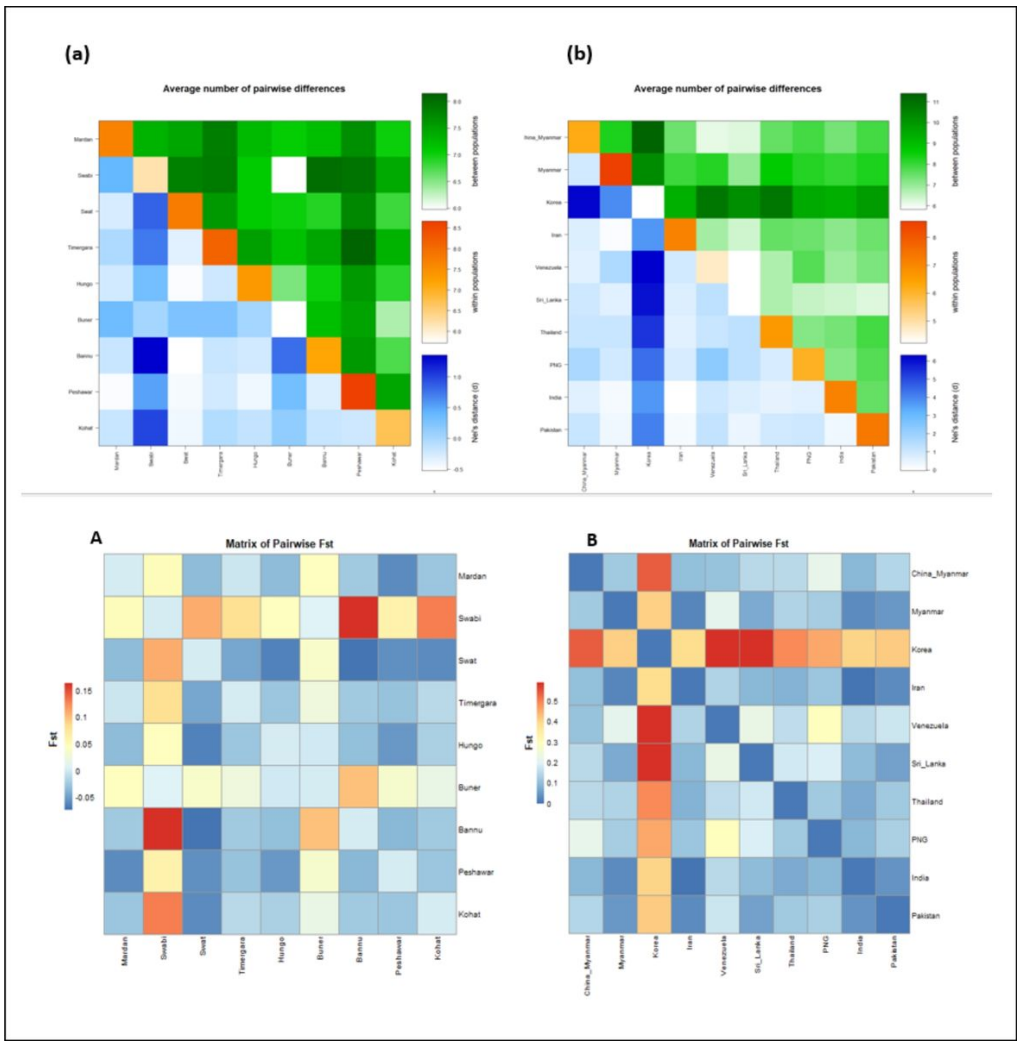

Figure 5

(a) The graph represents the average number of pairwise differences (πxy), between sampled population groups (Green above diagonal); within-population πxx(orange diagonal) and the net number of nucleotide differences among population's groups (Nei distance DA) (blue below diagonal) based on Pvama-1 gene variants among Pakistani groups. (b) The graph represents the average number of pairwise differences (πxy), between sampled population groups (Green above diagonal); within-population πxx (orange diagonal) and the net number of nucleotide differences among population's groups (Nei distance DA) (blue below diagonal) based on *Pvama-1* gene variants in context of other world population. (A) Heat-map plot of pair wise *Fst* between Pakistani populations based on *Pvama-1* gene sequences. (B) Based on *Pvama-1* gene sequences, heat map of pairwise *Fst* among Pakistani population and other world population group.

Supplementary Files

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

- ibraretal2021SupplyMaterial.docx