

Genetic diversity and effect of natural selection at apical membrane antigen-1 (AMA-1) among Iranian *Plasmodium vivax* isolates

Ahmad Reza Esmaceli Rastaghi, Fatemeh Nedaeei, Hossein Nahrevanian and Nazanin Hoseinkhan

Department of Parasitology, Pasteur Institute of Iran, Tehran, Iran

Abstract: Apical membrane antigen-1 (AMA-1) of *Plasmodium vivax* Grassi et Feletti, 1890 is a promising malaria vaccine candidate. However, antigenic variation is a major problem to design a universal malaria vaccine. Hence, detailed understanding of the *pvama-1* gene polymorphism can provide conductive information on this potential vaccine component. Therefore, this study investigated the extent of genetic polymorphisms at domain I (DI), DII and partial DIII of AMA-1 among Iranian *P. vivax* isolates. Out of 107 blood samples, 92 were analysed based on the quality of the sequencing data. The sequences were classified into 53 haplotypes. Amino acid changes were observed at 31 positions that 17 were located at DI, 11 were at DII and the rest of them (3 positions) were at DIII. Thus, codon polymorphisms at DI were found to be higher than DII. Also, five of these polymorphic codons (D242E, T374P, S389R, Y391F, I395F) were novel and have not been reported yet. Neutrality analysis by using the dN-dS difference (the difference between the rate of non-synonymous and synonymous mutations) showed a negative diversifying selection at DI, DII and across the length of both domains. The potential B-cell epitopes were found in 5 regions of the PvAMA-1 with 10 mutation sites (E145A, K188N, E189N/K/D, K190Q/E, P210S, E227V, D242E, R249H, G253E, K352E), whereas only one mutation (G288E) has been detected in intrinsically unstructured/disordered regions. Fixation index (Fst) estimation between Iranian and Indian isolates (0.0131) indicated a significant low genetic differentiation. Distribution of the polymorphic sites and IURs mapped on a three dimensional structure of PvAMA-1 showed that these regions were located at two opposite faces of the molecule. In conclusion, the results have significant value in the design and development of a malaria vaccine based on this antigen.

Keywords: malaria parasite, vaccine candidate, diversifying selection, Nested-PCR, DNA sequencing

This article contains supporting information (Table. S1) online at <http://folia.paru.cas.cz/suppl/2014-5-385.pdf>

Malaria is one of the most important public health problems. Human malaria parasites are widely distributed throughout the tropical and subtropical regions of the world (WHO 2009) with approximately 300–500 million clinical cases each year leading 1–3 million death (Cortés et al. 2003). *Plasmodium vivax* Grassi et Feletti, 1890 is the most prevalent of the four malaria parasite species that infect humans (Brutus et al. 2013) causing an estimated 70–80 million cases per year. Vivax malaria infection can be an important contributor to morbidity and social economic loss (Cui et al. 2003). In addition, the emergence of chloroquine-resistant *P. vivax* is a major global health problem (Nomura et al. 2001).

Adequate knowledge of parasite population structure, usually showing remarkable geographical differences, is a prerequisite to improve the control strategies such as vaccines and drugs. Knowledge of the parasite's geographical distribution permits us to determine if a particular vaccine or drug formulation in a malaria endemic region

is efficient (Cui et al. 2003). Several antigens including Duffy-Binding Protein (DBP), Merozoite Surface Protein-1 (MSP-1) and Apical Membrane Antigen-1 (AMA-1) are immunogenic proteins involved in the process of erythrocyte invasion. Among them, AMA-1 is a promising blood-stage malaria vaccine candidate (Remarque et al. 2008). Antibodies against AMA-1 have been detected in populations exposed to natural human infections. The introduced antibodies, blocking parasite invasion *in vitro*, can be protective against rodent and simian malaria parasites (Cortés et al. 2003). The failure of 'knockout' technology suggests that AMA-1 is involved in host cell invasion (Remarque et al. 2008).

The complete AMA-1 protein contains an ectodomain divided in three subdomains (DI, DII and DIII) by eight disulfide bonds (Pizarro et al. 2005). In contrast to other merozoite antigens, AMA-1 is almost conserved among different *Plasmodium* species and its limited polymorphism results from single nucleotide mutations in certain

regions of the gene. In addition, its position as a leading vaccine candidate is associated with conserved and functionally crucial parts of this antigen including DII loop, whereas expression of DII alone did not induce antibodies inhibiting the parasite growth *in vitro*.

DI and DII covering the polymorphic regions have shown to be major targets to elicit the inhibitory responses (Remarque et al. 2008). Therefore, understanding of antigenic variants of the parasite is important to design an effective vaccine. Single Nucleotide Polymorphisms (SNPs) are identified in certain regions of the gene that may serve as a useful molecular marker for typing parasite populations including microsatellites and repetitive regions (Cui et al. 2003). These markers can also provide the basic information for strategic development of mono- or polyvalent vaccine (Hwang et al. 2009).

A recent study from Iran has analysed a small number of *P. vivax* isolates (29 haplotypes in 37 isolates – Zakeri et al. 2013). In the present study, those 37 isolates were also included to provide more comprehensive information on the population structure of *P. vivax* isolates in Iran (37 from Zakeri et al. 2013 and 92 from the present study).

MATERIALS AND METHODS

Sample collection

A total of 107 isolates were collected from two different areas of tropical malaria endemic regions in southern and southeastern Iran (Hormozgan, Sistan and Baluchistan provinces) during 2009–2011. The most prevalent species in these regions from May to August is *Plasmodium vivax* and then from October to November *P. vivax* and *P. falciparum* Welch, 1897. Totally, *P. vivax* is responsible for more than 90% of the recorded malaria cases. In this investigation, 107 blood samples were obtained from individuals with symptomatic *P. vivax* infections. All samples were diagnosed as *P. vivax* by light microscopic examination after Giemsa staining. The characteristics of the patients and their travel history were recorded before sampling. The majority of the patients (60%) were between 10–30 years old (ranging from 2–65 years old) and 70% of them were male. 2–5 ml venous blood was collected in an EDTA-containing tube from the *P. vivax* positive patients after all studied individuals or their parents/legal guardians gave informed consent according to institutional ethical guidelines. Blood samples were stored at -20°C and then transported to the Department of Parasitology at the Pasteur Institute of Iran.

Isolation of genomic DNA and PCR amplification

Genomic DNA was extracted from the infected blood samples using the QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. An aliquot of eluted DNA was kept at 4°C for PCR amplification. Nested PCR was performed to amplify a fragment of the *pvama-1* gene containing DI, DII and partial DIII. Initially, a 1139 bp fragment of the gene was amplified using primary set of primers PvAMAF1 (5'-GTTGAGAGAAGCACACGAATG-3') and PvAMA2 (5'-GAGATAAATATCCTCGGCAGG-3'). The primary amplified DNA fragments were diluted and 2 µl were used as template in second round of PCR. The second round of amplification was performed using primers PvAMAF2 (5'-TGCA-

GAAGTGGAAAATGCAAAG-3') and PvAMA4 (5'-CCATCAACACTGTACAGATTC-3'). The quality of PCR products was analysed on 1.2% agarose gel. All PCR amplifications were performed in a 25 µl PCR reaction volume containing 0.2 µM of each primer pair, 1.5 mM MgCl₂, 1× PCR buffer (50 mM KCl, 10 mM Tris-Cl, pH 8.3), 0.2 mM dNTP and 0.5 unit of Taq DNA polymerase (Bioron, Ludwigshafen, Germany).

The cycling parameters to amplify the primary fragments were as follows: initial denaturation at 94°C for 10 min, 35 cycle of denaturation at 94°C for 1 min, annealing at 57°C for 1 min, extension at 72°C for 1 min and a final extension at 72°C for 10 min. The cycling conditions to amplify the internal 1000 bp fragment were: 10 min initial denaturation at 94°C followed by 35 cycle with 30 s denaturation at 94°C, 30 s annealing at 57°C, 30 s extension at 72°C and final extension was done at 72°C for 10 min. PCR fragments were electrophoresed on a 1.2% agarose gel.

Nucleotide sequencing and data analysis

PCR products in second round of amplification were purified using Qiagen gel extraction kit (Qiagen) from agarose gel. The DNA fragments were sequenced using ABI Big Dye Terminator Reaction Ready Kit Version 3.1 (PE Applied Biosystems, Foster City, CA, USA) and an automated DNA sequencer (ABI 310 Genetic Analyzer; PE Applied Biosystems). The 1000 bp amplicons were sequenced on both strands using PvAMAF2 and PvAMA4 primers. Nucleotide sequences were edited using the Chromas version 2.33 program. Nucleotide and amino acid sequences were aligned using MEGA4 program (Tamura et al. 2007) with the Salvador-1 (NCBI GenBank ID AF063138) sequence as reference strain. The nucleotide sequence data in this study are available in the GenBank database under accession numbers KF181626–KF181642 and KF422636–KF422681.

Out of 107 sequences, 92 were analysed with a high quality in sequencing data. In order to classify the sequences into haplotypes (H), the 92 PvAMA-1 data from the present study were compared with the previously reported PvAMA-1 data from Iran (n = 37, Zakeri et al. 2013); the haplotypes were identified using DnaSP ver. 5. 10. 01 (Librado and Rozas 2009). The haplotype diversity (Hd), nucleotide diversity (π), segregating or polymorphic sites (S) and average number of pairwise nucleotide differences within the population (K) were measured using the options available in DnaSP program. Evidence of recombination in terms of recombination per gene (Ra), recombination between adjacent polymorphic sites (Rb) and minimum number of recombination events (Rm) was also estimated using DnaSP program. The linkage disequilibrium between nucleotide variants at different polymorphic sites was estimated by R² index.

The LD plot observed for 129 Iranian isolates was compared with the LD plot for 60 isolates reported from India (accession Nos. EF025187–EF025197, see Rajesh et al. 2007 and accession Nos. EU282774–EU282822, see Thakur et al. 2008). To analyse the hypothesis that all mutations are neutral, the difference between the rate of the average number of synonymous substitutions (dS) and the average number of non-synonymous substitutions (dN) was calculated using the Nei and Gojobori method (Nei and Gojobori 1986) with the Jukes and Cantor (JC) correction by the Z-test of selection (P < 0.05) in MEGA software package. In addition, Tajima's test (Tajima 1989) and other analysis to detect the natural selection in a population including Fu and Li's D* and F* (Fu and Li 1993) tests were performed by

DnaSP software. To test the neutral hypothesis based on a comparison of synonymous and non-synonymous variation within and between species, the McDonald and Kreitman (McDonald and Kreitman 1991) test was performed with a single *P. falciparum* sequence as a second species (accession No. EU586503, see Duan et al. 2008). To determine the significant departure from neutrality, the two-tailed Fisher's exact test was performed.

Phylogenetic analysis was performed by the neighbour joining method in MEGA program to evaluate the genetic relationships among 3 predominant haplotypes found in the present study and 22 *pvama-1* sequences reported from 13 different geographic locations as well as one *P. falciparum ama-1* sequence from Iran (accession No. HM776820, see Mardani et al. 2012). The *pvama-1* sequences were chosen randomly from Africa, Brazil (Cuiaba, the capital of Mato Grosso state and Belem), Honduras, India, Indonesia, Myanmar, North Korea, Solomon, South Korea, Sri Lanka, Thailand and Venezuela.

To calculate the extent of genetic differentiation among populations, Fixation-statistics (F_{st}) was measured among isolates from Iran, India, Sri Lanka, Thailand and Indonesia using DnaSP. The quantity of F_{st} was defined in four categories as follows: a value between 0–0.05 indicates low genetic differentiation; 0.05–0.15, moderate differentiation; 0.15–0.25, high differentiation and above 0.25 shows very high genetic differentiation (Balloux and Lugon-Moulin 2002). The ABCpred server (<http://www.imtech.res.in/raghava/abcp> – Saha and Raghava 2006) was used to predict the potential B-cell epitopes. Sixteen residues were predicted by the threshold of 0.8 to show a given region as a reliable linear B-cell epitope. Intrinsically unstructured/disordered regions (IURs) are important to understand the protein structure and its affection on antigenicity. Thus, the program regional order neural network (RONN) server (<http://www.strubi.ox.ac.uk/RONN> – see Yang et al. 2005) was used to detect the potential IURs in the PvAMA-1. Furthermore, we have compared the variations of AMA-1 domain I sequences between *P. vivax* and *P. falciparum* species among Iranian isolates (Mardani et al. 2012) (only DI sequencing data of Iranian *P. falciparum* isolates were available). Finally, a three-dimensional structure of PvAMA-1 (PDB file: 1w8k; Pizarro et al. 2005) was generated to show the distribution of polymorphic sites, B-cell epitopes and IURs using Weblab Viewer Lite 4.2 (<http://www.scalacs.org/TecherResources>).

RESULTS

Genetic variation and allelic forms of *pvama-1* gene

A 1000 bp region of the *pvama-1* gene was amplified and sequenced from 107 blood samples, with 92 of them having been chosen for further analysis with the best DNA sequencing results. However, the sequencing data was readable only for 911 bp region (covering predominant region of DI, complete DII and partial DIII). The 92 sequences were grouped into 53 haplotypes and compared with the previously reported *pvama-1* sequences ($n = 37$, Zakeri et al. 2013) from Iran (Table S1). Based on the results, 70 haplotypes have so far been reported among 129 Iranian isolates. Analysis of 129 sequences at the nucleotide level showed that out of 911 bp, 861 nucleotide sites were conserved whereas 50 of them were

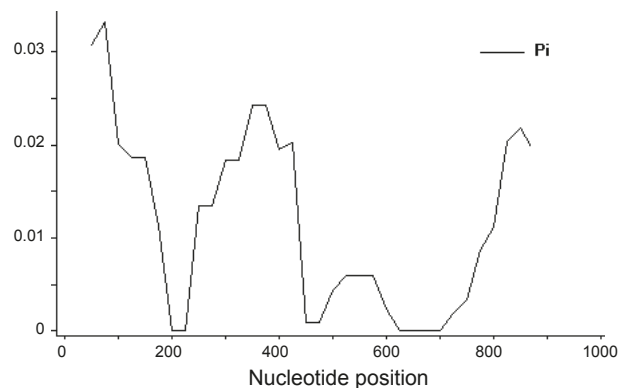


Fig. 1. Sliding window plot of nucleotide diversity per site (π) at the PvAMA-1 in 129 Iranian isolates with the window length 100 bp and step size of 25 bp. The analysis was performed using DnaSP ver. 5. 10. 01. The level of nucleotide diversity was high between the nucleotide positions 1 to 200 and then 225 to 425 at domain I also nucleotide positions 725 to 868 at domain II.

variable (7 singleton variables and 43 parsimony informative). Amino acid changes were found at 31 positions where 17 polymorphic positions were located at DI, 11 positions at DII and the rest of them (3 positions) at DIII. Of these 31 polymorphic sites, 24 sites were dimorphic and four were trimorphic (K190Q/E, V382E/I, L384P/R, E385G/D), whereas only three of them showed four alternative amino acid residues, thus being tetramorphic (R112T/N/K, E189N/K/D, S228D/N/G) (Table S1). Also, codon polymorphisms was found in this study (D242E, T374P, S389R, Y391F, I395F).

Nucleotide sequence polymorphism, natural selection and recombination analysis

Analysis of the nucleotide sequence polymorphism was performed for DI, DII and across the length of both domains of *ama-1* gene within all 129 Iranian *Plasmodium vivax* isolates (Table 1). Total number of mutation sites for these regions were 27, 21 and 48 positions, respectively. The overall nucleotide diversity (π) and haplotype diversity (H_d) for the entire length of DI and DII were 0.013 ± 0.002 (mean \pm standard deviation) and 0.976 ± 0.005 (mean \pm standard deviation) respectively. Also, the average number of pair wise nucleotide differences (K) was 11.61 for all 129 isolates from Iran. Computing the π values (window length of 100 bp and step size of 25 bp) revealed that the values of nucleotide diversity ranged 0–0.03320. Maximum diversity was seen between nucleotide positions 1–200 and 22–425 at DI of PvAMA-1. The nucleotide positions 725 to 868 have also a high level of diversity at DII (Fig. 1).

Comparison of the genetic diversity extent between DI and DII showed that DI is more variable than DII (Table 1, Fig. 1). For recombination analysis, all 129 Iranian isolates were also included. Minimum number of recombination events between adjacent variable sites (R_m) for

Table 1. Estimation of DNA sequence polymorphisms and neutrality tests at AMA-1 among Iranian of *Plasmodium vivax* isolates.

	S	Singleton variable sites	Parsimony informative sites	Total no. of mutation	K	$\pi \pm SD$	Hd $\pm SD$	dN - dS $\pm SE$	Tajima's D	D*(F&L)	F*(F&L)
DI	25	3	22	27	8.74	0.019 \pm 0.004	0.954 \pm 0.008	-0.60 \pm 0.01	2.6073	0.7679	1.6108
DII	19	4	15	21	2.86	0.007 \pm 0.002	0.918 \pm 0.012	-0.34 \pm 0.02	-0.5026	-0.0476	-0.3725
Total	44	7	37	48	11.60	0.013 \pm 0.002	0.976 \pm 0.005	-0.54 \pm 0.01	1.3302	0.4740	0.8184

92 isolates are from present study and 37 reported by Zakeri et al. (2013); S – number of polymorphic (segregating) sites; K – average number of pairwise nucleotide differences; π – observed average pairwise nucleotide diversity; Hd – haplotype diversity; dN - dS – the difference between the rate of non-synonymous and synonymous mutations; D*(F&L) – Fu and Li's D* value; F*(F&L) – Fu and Li's F* value; the values of dN-dS were not statistically significant based on the Z-test of selection ($P > 0.05$); the values for Fu and Li's D* and F* were considered as nonsignificant difference ($P > 0.05$).

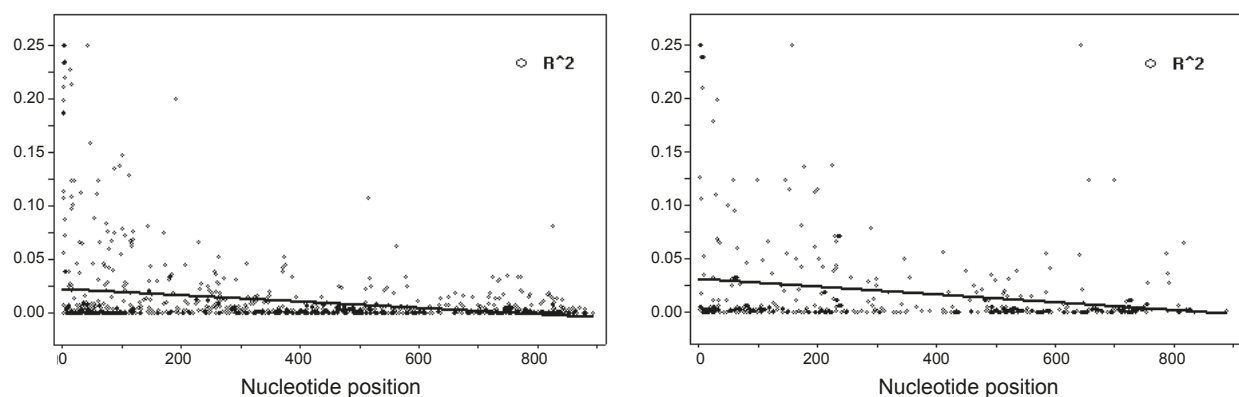


Fig. 2. The R2 values are plotted against the nucleotide distances with two-tailed Fisher's exact test of significance using DnaSP ver. 5.10.01. **A** – plot is constructed for 129 Iranian *pvama-1* sequences; **B** – plot is based on 60 Indian isolates. These plots show non-random association between nucleotide variants at different polymorphic sites.

Table 2. Results of the null hypothesis with the McDonald and Kreitman test.

PvAMA-1	Polymorphic changes within <i>P. vivax</i>		Fixed differences between species		NI	Fisher's exact test
	Ps	Pn	Ds	Dn		
DI	4	17	76	89	3.629	0.019882*
DII	4	11	72	64	3.094	0.06156
Total	8	28	148	153	3.386	0.002365*

Ps – the number of synonymous substitutions within *pvama-1* gene; Pn – the number of non-synonymous substitutions within *pvama-1* gene; Ds – the number of synonymous changes between species; Dn – the number of non-synonymous changes between species; NI – neutrality index; * significant values with $P < 0.05$.

Table 3. Estimation of genetic differentiation at *pvama-1* gene among populations from various geographic regions.

country (n)	Iran (129)	India (60)	Sri Lanka (23)	Thailand (60)	Venezuela (40)
Iran	-	0.0001*	0.0011*	0.0067*	0.0023*
India ^a	0.0131	-	0.0017*	0.0054*	0.0018*
Sri Lanka ^b	0.0875	0.1266	-	0.0077*	0.0049*
Thailand ^c	0.4139	0.3647	0.4823	-	0.0048*
Venezuela ^d	0.1699	0.1391	0.3291	0.3968	-

Values of Fst are indicated in the lower left and P-values are indicated in the upper right of the table; n – number of samples analysed; ^a Thakur et al.(2008) (EU282774–EU282822), Rajesh et al.(2007) (EF025187–EF025197); ^b Gunasekera et al.(2007) (EF218679–EF218701); ^c Putaporntip et al.(2009) (FJ784891–FJ784950); ^d Ord et al.(2008) (EU346015–EU346054); * the Fst values were considered as significant difference ($P < 0.05$).

the entire 911 bp region was 18.0, whereas recombination between adjacent sites and recombination per gene were 0.0622 and 56.6 respectively. The linkage disequilibrium (LD) plot shows non-random association between nucleotide variations in 129 Iranian isolates at different poly-

morphic sites (Fig. 2). A similar result was found when the LD plot was constructed for *pvama-1* sequences from India (Rajesh et al. 2007, Thakur et al. 2008). The difference between dN and dS (dN - dS) for DI (-0.6 \pm 0.01 SE), DII (-0.34 \pm 0.02 SE) and across the length of both do-

Table 4. Comparison of the amino acid polymorphisms at AMA-1 domain I in *Plasmodium vivax* and *P. falciparum* isolates from Iran.

PvAMA-1 (n = 129) ^a			PfAMA-1 (n = 48) ^b		
codons	mutations	f	codons	mutations	f
107	A/D	41/88	162	N/K	29/19
112	R/T/N/K	11/44/1/73	167	T/K	37/11
117	G	-	172	E/G	32/16
118	D/N	128/1	173	N	-
120	K/R	7/122	175	D/N	47/1
123	D	-	178	D/N	43/5
130	N/K	109/20	185	P	-
132	N/D	79/50	187	K/N	33/15
134	H	-	189	L/P	47/1
135	I	-	190	M/I	35/13
140	L/I	71/58	195	L	-
141	A/E	91/38	196	D/N/Y	4/41/3
142	N	-	197	H/G/D	17/7/24
145	A/E	114/15	200	L/R/D/H	2/2/15/29
146	R	-	201	F/L	46/2
149	D	-	204	D/N	18/30
151	V	-	206	K/N	21/27
152	E	-	207	Y/N	46/2
170	V	-	225	I/N	28/20
175	Q	-	230	K/E/Q	31/12/5
187	E	-	242	Y/A/D	19/3/26
188	K/N	127/2	243	K/N	45/3
189	E/N/K/D	75/20/33/1	244	D	-
190	K/Q/E	107/12/10	245	K	-
193	H/Y	112/17	248	H	-
210	P/S	71/58	265	K	-
212	A	-	267	E/Q	20/28
214	N	-	269	K/I	36/12
227	E/V	77/52	282	I/K	25/23
228	S/D/N/G	74/46/8/1	283	S/L	45/3
230	E	-	285	Q/E	45/3
241	N	-	296	D/H	41/7
242	D/E	128/1	297	N	-

f – frequency; n, number of isolates; the common polymorphic amino acids between PvAMA-1 and PfAMA-1 are shown in bold; dashes show invariant sites; ^aPvAMA-1 data for 92 sequences from this study and 37 sequences reported by Zakeri et al. 2013 (Iran); ^bPfAMA-1 data from Mardani et al. 2012 (Iran).

mains (-0.54 ± 0.01 SE) was negative and statistically nonsignificant (Z-test, $P > 0.3$) indicating the effect of purifying or negative selection (Table 1).

The values of Tajima's D, $D^*(F\&L)$ and $F^*(F\&L)$ showed an insignificant departure from neutral substitution changes leading to the tendency of stronger diversifying selection at DI, DII and the entire length of both domains (Table 1). Neutrality analysis estimated by the McDonald Kreitman test showed an excess of synonymous mutations between species and non-synonymous mutations within species compared to the neutral substitution patterns ($PnDs/PsDn > 1$) and Neutrality Index was found to be 3.629 ($NI > 1$) for DI, 3.094 for DII and 3.386 for the entire length of both domains suggesting a negative departure from neutrality at these regions due to purifying natural selection (Table 2). The Fisher exact test

indicated a significant diversifying selection for DI and throughout the length of DI and DII.

Comparison of Iranian *P. vivax* isolates with the previously reported sequences

Phylogenetic analysis separated the *ama-1* gene of *P. vivax* and *P. falciparum* into two major phylogenetic groups. Similarly, *pvama-1* sequences from different geographic regions of the world were found in different subdivisions of phylogenetic tree (Fig. 3). The Fst estimation showed a low genetic differentiation between Iranian isolates and Indian population ($Fst = 0.0131$), whereas the fixation index between isolates from Iran and Thailand was found indicates high genetic differentiation ($Fst = 0.4139$) (Table 3). Comparison of the polymorphic residues in DI of AMA-1 between the *P. vivax* and *P. falciparum* (accession Nos. HM776795–HM776803, HM776805–HM776820, HM776822–HM776834 and HQ658150–HQ658159; see Mardani et al. 2012) species showed that 9 of these residues were common (Table 4).

Molecular structure and antigenicity analysis of PvAMA-1

Location of polymorphic sites, IURs and B-cell epitopes was mapped to examine the distribution of these regions at the entire DNA fragment amplified in this study. The common regions between the SNPs and the potential B-cell epitopes were observed in 10 positions including E145A, K188N, E189N/K/D, K190Q/E, P210S, E227V, D242E, R249H, G253E, K352E. Moreover, the IURs were observed in two regions with one mutation site (G288E) (Fig. 4). Also, the crystal structure of PvAMA-1 has been demonstrated to visualise the surface distribution of these regions (Fig. 5). Interestingly, almost all of the polymorphic residues and IURs were located in two opposite sides of the molecule.

DISCUSSION

Elimination of malaria disease from various regions of the world requires new strategies, which are quite different from those applied to control disease. However, available tools cannot be effective for global eradication of malaria (Targett and Greenwood 2008). Therefore, comprehensive research is needed to develop vaccine strategy. Although antigenic variation is one of the major problems to develop an effective anti-malaria vaccine in different geographic regions (Genton and Reed 2007), the detailed knowledge of the parasite antigenic variants is a prerequisite to design a vaccine that might be efficient in various endemic areas. This study analysed the genetic diversity at PvAMA-1 within a large number of Iranian isolates of *Plasmodium vivax* from two different malaria endemic regions. We also evaluated the phylogenetic relationships of *ama-1* gene between *P. vivax* and *P. falciparum* from the same geographical regions of Iran as well as among

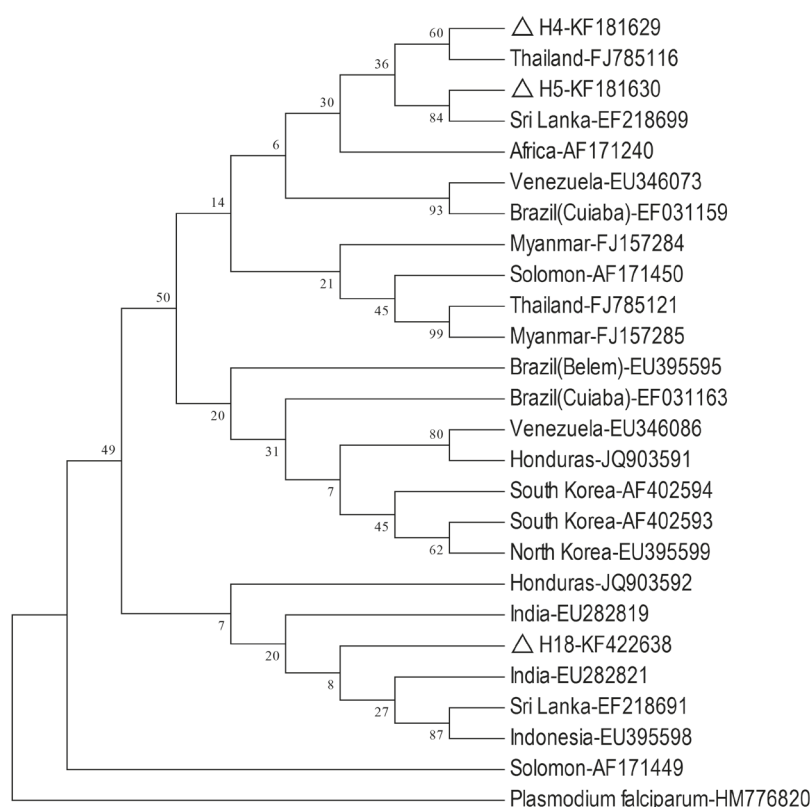


Fig. 3. Phylogenetic analyses was performed among three predominant haplotypes (labelled with the triangle) found in the present study and 22 *pvama-1* sequences reported from 13 different geographic locations as well as one *Plasmodium falciparum ama-1* sequence. The evolutionary history was reconstructed using the Neighbor-Joining method in MEGA4. The bootstrap consensus tree was inferred from 1 000 replicates and values for the nodes are shown in the tree. Abbreviation: H – haplotype

97 FGKGIVENS^DVSFLKPVATGDQRLKDGGFAPNANDHISPMTIANLKARYKDNVEMMKLNDIALCRTHAASFVMAGDQNSSYRHPAVYDEK 188
 EKTCHMLYLSAQENMGPRYCSDAQNRDAVFCFKPDKNESFENLVYLSKNVRNDWDKKCPKRNKLNAGFLWVDGNCEEIPYVKEVEAKDL 279
 RECNRIVF^GASASDQPTQYEEEMTDYQKIQGFRQNNREMIKSAFLPVGAFNSDNFSKSGRGNWANFDSVKKCYIFNTKPTCLINDINFIAT 373
 TALSHPQEVDPFPCSIYKDEIERE 399

Fig. 4. Polymorphic amino acid residues reported in the present study, the potential B-cell epitopes and intrinsically unstructured/disordered regions (IURs) at the entire 911 bp fragment are mapped using ABCpred server and RONN server. Ten mutation sites were found in the B-cell epitopes, whereas only one mutation site (G288E) is in IURs. Residues 43–248 constitute DI, 249–385 constitute DII and 386–487 make DIII. The predicted B-cell epitopes are in boldfaced; grey shadow shows polymorphic amino acid residues and underlined shows IURs.

P. vivax isolates reported from various locations of the world.

Analysis of *pvama-1* sequences using the DnaSP program showed that the 92 sequences were classified into 53 haplotypes (Table S1). Also the GenBank BLAST search for each haplotype revealed that 39 haplotypes were novel and have not been reported so far, whereas 11 haplotypes were previously identified among isolates from Iran (Zakeri et al. 2013) and 3 of them were common with Sri Lankan population (accession Nos. EF218699, EF218682, EF218697; see Gunasekera et al. 2007). Some haplotypes were found among isolates obtained from individuals who had a travel history to Pakistan (H12) and Afghanistan (H25, H27, H28, H30, H33, H34) one month

before sampling. It is suggested that migration of human hosts from/to neighbouring countries is responsible for the existence of the similar variants among these regions.

The result of the nucleotide diversity analysis showed more diversity at DI compared to DII, which is similar to results of previous studies (Ord et al. 2008, Thakur et al. 2008, Moon et al. 2010, Zakeri et al. 2013). In contrast Gunasekera et al. (2007) reported from Sri Lanka slightly more variation at DII ($\pi = 0.0097 \pm 0.0007$ SD; $H = 13$) than DI ($\pi = 0.0092 \pm 0.013$ SD; $H = 11$). It may be due to the different number of isolates analysed in the present study ($n = 129$) and Sri Lanka study ($n = 23$).

Furthermore, we have found a high level of diversity at PvAMA-1 among Iranian isolates that may be related

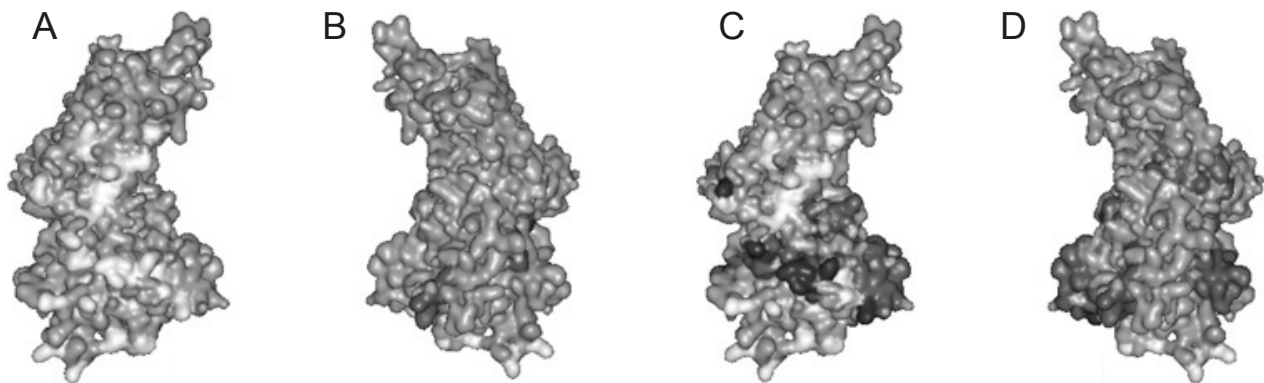


Fig. 5. Three-dimensional structure of PvAMA-1 and the surface representation of polymorphic amino acid residues, B-cell epitopes and IURs. **A, B** – two opposite sides of the same structure; white colour residues are polymorphic sites; dark grey, IURs; light grey, all other positions. The variable sites reported in this study are located predominantly on one molecular face without IURs (**A**), whereas IURs are on the opposite face with one variable site showed in black colour (**B**). **C, D** – two opposite sides of the molecule; white colour, polymorphic regions; dark grey, B-cell epitopes; black colour shows the overlap between B-cell epitope regions and polymorphic sites found in ten variable amino acid positions.

to the migration and population movement between Iran with low level of endemicity (<http://emro.who.int/dsaf/dsa741.pdf>) and eastern neighbouring countries (Afghanistan and Pakistan) that are hyperendemic for malaria disease (WHO 2007). However, diversity of *pvama-1* gene is often influenced by the natural selection and recombination events (Gunasekera et al. 2007, Thakur et al. 2008, Moon et al. 2010). The high values of the recombination parameters found in this study show that the meiotic recombination occurring between adjacent variable sites may be contributing to the high genetic diversity at *pvama-1* as confirmed by the decline of LD index R^2 with increasing nucleotide distances (Fig. 2).

Phylogenetic analysis showed 100% similarity between haplotype H5 and the Sri Lankan isolate whereas haplotypes H4 and H18 were specific to Iran (Fig. 3). Separation of the *pvama-1* and *pfama-1* genes into two major heterogeneous groups indicates DNA divergence between the two orthologues. Distribution of *P. vivax* isolates from various geographic regions throughout the different subdivisions of phylogenetic tree suggests that these different populations might be diverged from an original population. The geographic origin of *P. vivax* is supposed to be southeast Asia (Escalante et al. 2005) and regular movement of human hosts differentiates the populations through isolation by distance. Therefore, similar allelic forms of the AMA-1 may exist in populations from different geographic areas of the world that should be considered to provide a worldwide repertoire of common genotypes of the AMA-1 for the malaria vaccine development.

The estimation of the genetic differentiation at *pvama-1* gene among isolates from Iran, India, Sri Lanka, Thailand and Venezuela showed minimum genetic differentiation (0.0131) between Iran and India that may be explained by the migration between these regions due to the close

proximity. However, the F_{st} estimation between Iran and Thailand indicated very high geographical genetic differentiation (0.4139) because the populations are genetically differentiated by distance (Table 3).

The neutrality analysis by computing the difference between dN and dS showed an insignificant negative deviation from neutrality at DI, DII and the entire length of both domains indicating the tendency of stronger purifying selection across these regions (Table 1). A similar result was observed when the McDonald Kreitman test was computed to analyse the null hypothesis. In addition, this test showed a significant departure from neutral substitution patterns for DI and the entire length of DI and DII. The purifying selection at DI is supported by the results from Indian (Thakur et al. 2008) and Sri Lankan (Gunasekera et al. 2007) studies and the result of the neutrality analysis for DII reported in Venezuela (Ord et al. 2008) is in agreement with what we have found for this domain. However, it is in contrast to the report by Thakur et al. (2008) in India and Gunasekera et al. (2007) in Sri Lanka who have detected positive selection at DII.

The population size analysed, host immune selection pressure and different levels of the transmission intensity in various part of the world may be responsible for such a difference in the results of natural selection analysis. On the other hand, the exchange of genes among isolates from different regions homogenises allele frequencies between populations that may be associated with the relative effects of selection (Balloux and Lugon-Moulin 2002). The results of the Tajima's D , F_u and F^* (D^* and F^*) showed insignificant departure from neutrality for the sequences analysed in this investigation and the previous Iranian study (Zakeri et al. 2013). Knowing the location of antigenic epitopes, induced antibody response could be a conducive strategy to develop a useful vaccine against malaria parasites. In addition, mapping the IURs sup-

posed to be exposed to the immune system (Bueno et al. 2011) and effective on the protein structure (Dunker et al. 2002) provides significant information for experimental designs including the improvement of the protein expression with a correct folding. Indeed, the co-occurrence of B-cell epitopes and the polymorphic sites represents the regions which are more likely to be exposed to the immune system and 10 such residues have been found in the present study (Figs. 4, 5).

Interestingly, we have found only one mutation site in IURs (G288E) that may be related to the crucial role of IURs in many fundamental cell functions. In contrast, the IURs in the PvAMA-1 DII indicate a high level of antigenicity in the populations exposed to the natural human infections (Bueno et al. 2011). Therefore, the pressure of the host immune system may be responsible for G288E mutation in this particular IUR. Examination of the 3D structure of PvAMA-1 showed that variable sites reported in this study are located predominantly in one molecular face as shown by Thakur et al. (2008), whereas IURs are in the opposite face of the molecule (Fig. 5).

Finally, neutrality analysis indicated that DI and DII of PvAMA-1 are dominant targets of purifying selection suggesting the tendency of natural selection to conserve

the function and structure of this antigen. Sequence conservation and high immunogenicity of DI and DII show that these regions can be used as the important components of a PvAMA-1-based vaccine. In addition, to design an effective vaccine against malaria, it is momentous not only to consider the region showing strongest natural immune target but also to include alleles that elicit broadly reactive antibodies to cover the antigenic variation of *P. vivax* isolates (Arnott et al. 2013). Antibody response raised against parasites expressing different allelic forms is more effective against one form (Thakur et al. 2008). These findings suggest that AMA-1 is still a viable vaccine target in spite of genetic variability revealed by comparative sequence analyses.

Acknowledgements. We thank Zahedan and Hormozgan Universities of Medical Sciences staffs and patients for their participation in this study. We are grateful to Dr. Ali Ehsan Heidari from Alborz University of Medical Sciences for providing some blood samples and Dr. Nahid Borhani for molecular assistance. The authors also acknowledge the co-operation of Center for Diseases Management and Control (CDMC), particularly Dr. Mohammad Mehdi Gouya and Dr. Ahmad Raeisi. This work was supported by the Pasteur Institute of Iran (grant No. 434 to A.R. Esmaili Rastaghi).

REFERENCES

- ARNOTT A., MUELLER I., RAMSLAND P.A., SIBA P.M., REEDER J.C., BARRY A.E. 2013: Global population structure of the genes encoding the malaria vaccine candidate, *Plasmodium vivax* apical membrane antigen 1 (PvAMA1). *PLoS Negl. Trop. Dis.* 7: e2506.
- BALLOUX F., LUGON-MOULIN N. 2002: The estimation of population differentiation with microsatellite markers. *Mol. Ecol.* 11: 155–165.
- BRUTUS L., SANTALLA J., SCHNEIDER D., AVILA JC., DELORON P. 2013: *Plasmodium vivax* malaria during pregnancy, Bolivia. *Emerg. Infect. Dis.* 19: 1605–1611.
- BUENO L.L., LOBO F.P., MORAIS C.G., MOURAO L.C., DE AVILA R.A., SOARES I.S., FONTES C.J., LACERDA M.V., CHAVEZ OLORTEGUI C., BARTHOLOMEU D.C., FUJIWARA R.T., BRAGA E.M. 2011: Identification of a highly antigenic linear B cell epitope within *Plasmodium vivax* apical membrane antigen 1 (AMA-1). *PLoS ONE* 6: e21289.
- CORTÉS A., MELLOMBO M., MULLER I., BENET A., REEDER J.C., ANDERS R.F. 2003: Geographical structure of diversity and differences between symptomatic and asymptomatic infections for *Plasmodium falciparum* vaccine candidate AMA-1. *Infect. Immun.* 71: 1416–1426.
- CUI L., ESCALANTE A.A., IMWONG M., SNOUNOU G. 2003: The genetic diversity of *Plasmodium vivax* populations. *Parasitology* 19: 220–226.
- DUAN J., MU J., SU X.-Z., MILLER L. 2008: Population structure of the genes encoding the polymorphic *P. falciparum* apical membrane antigen 1: implications for vaccine design. *Proc. Natl. Acad. Sci. USA* 105: 7857–7862.
- DUNKER A.K., BROWN C.J., LAWSON J.D., IAKOUCHEVA L.M., OBRADOVIC Z. 2002: Intrinsic disorder and protein function. *Biochemistry* 41: 6573–6582.
- ESCALANTE A.A., CORNEJO O.E., FREELAND D.E., POE A.C., DURREGO E., COLLINS W.E., LAL A.A. 2005: A monkey's tale: the origin of *Plasmodium vivax* as a human malaria parasite. *Proc. Natl. Acad. Sci. USA* 102: 1980–1985.
- FU Y.-X., LI W.-H. 1993: Statistical tests of neutrality of mutations. *Genetics* 133: 693–709.
- GENTON B., REED Z.H. 2007: Asexual blood-stage malaria vaccine development: facing the challenges. *Curr. Opin. Infect. Dis.* 20: 467–475.
- GUNASEKERA A.M., WICKRAMARACHICHI T., NEAFSEY D.E., GANGULI I., PERERA L., PREMARATNE P.H., HARTL D., HANDUNNETTI S.M., UDAGAMA-RANDENIYA P.V., WIRTH D.F. 2007: Genetic diversity and selection at the *Plasmodium vivax* apical membrane antigen-1 (PvAMA-1) locus in a Sri Lanka population. *Mol. Biol. Evol.* 24: 939–947.
- HWANG S.Y., KIM S.H., KHO W.G. 2009: Genetic characteristic of polymorphic antigenic markers among Korean isolates of *Plasmodium vivax*. *Kor. J. Parasitol.* 47: 51–58.
- LIBRADO P., ROZAS J. 2009: DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.
- MARDANI H., KESHAVARZ H., HEIDARI A., HAJJARAN H., RAEISI A., KHORRAMIZADEH M.R. 2012: Genetic diversity and natural selection at the domain I of apical membrane antigen-1 (AMA-1) of *Plasmodium falciparum* in isolates from Iran. *Exp. Parasitol.* 130: 456–462.
- MCDONALD J.H., KREITMAN M. 1991: Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* 351: 652–654.
- MOON S.U., NA B.K., KANG J.M., KIM J.Y., CHO S.H., PARK Y.K., SOHN W.M., LIN K., KIM T.S. 2010: Genetic polymorphism and effect of natural selection at domain I of apical mem-

- brane antigen-1 (AMA-1) in *Plasmodium vivax* isolates from Myanmar. *Acta Trop.* 114: 71–75.
- NEI M., GOJOBORI T. 1986: Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* 3: 418–426.
- NOMURA T., CARLTON J.M., BAIRD J.K., DEL PORTILLO H.A., FRYAUFF D.J., RATHORE D., FIDOCK D.A., SU X., COLLINS W.E., MCCUTCHAN T.F., WOOTTON J.C., WELLEMS T.E. 2001: Evidence for different mechanisms of chloroquine resistance in two *Plasmodium* species that cause human malaria. *J. Infect. Dis.* 183: 1653–1661.
- ORD R.L., TAMI A., SUTHERLAND C.J. 2008: AMA-1 genes of sympatric *Plasmodium vivax* and *P. falciparum* from Venezuela differ significantly in genetic diversity and recombination frequency. *PLoS ONE* 3: e3366.
- PIZARRO J.C., VULLIEZ-LE NORMAND B., CHESNE-SECK M.L., COLLINS C.R., WITHERS-MARTINEZ C., HACKETT F., BLACKMAN M.J., FABER B.W., REMARQUE E.J., KOCKEN C.H., THOMAS A.W., BENTLEY G.A. 2005: Crystal structure of the malaria vaccine candidate apical membrane antigen 1. *Science* 308: 408–411.
- PUTAPORN TIP C., JONGWUTIWES S., GRYNBERG P., CUI L., HUGHES A.L. 2009: Nucleotide sequence polymorphism at the apical membrane antigen-1 locus reveals population history of *Plasmodium vivax* in Thailand. *Infect. Genet. Evol.* 9: 1295–1300.
- RAJESH V., ELAMARAN M., VIDYA S., GOWRISHANKAR M., KOCHAR D., DAS A. 2007: *Plasmodium vivax*: genetic diversity of the apical membrane antigen-1 (AMA-1) in isolates from India. *Exp. Parasitol.* 116: 252–256.
- REMARQUE E.J., FABER B.W., KOCKEN C.H., THOMAS A.W. 2008: Apical membrane antigen 1: a malaria vaccine candidate in review. *Trends Parasitol.* 24: 74–84.
- SAHA S., RAGHAVA G.P.S. 2006: Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins* 65: 40–48.
- TAJIMA F. 1989: Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585–595.
- TAMURA K., DUDLEY J., NEI M., KUMAR S. 2007: MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24: 1596–1599.
- TARGETT G.A., GREENWOOD B.M. 2008: Malaria vaccines and their potential role in the elimination of malaria. *Malar. J.* 1: S10.
- THAKUR A., ALAM M.T., BORA H., KAUR P., SHARMA Y.D. 2008: *Plasmodium vivax* sequence polymorphism and effect of natural selection at apical membrane antigen-1 (PvAMA-1) among Indian population. *Gene* 419: 35–42.
- WHO (WORLD HEALTH ORGANIZATION) 2007: Strategic plan for malaria control and elimination in the WHO eastern Mediterranean region. World Wide Web electronic publication, www.emro.who.int/dsaf/dsa741.pdf, 3/2010.
- WHO (WORLD HEALTH ORGANIZATION) 2009: World malaria report 2009. World Wide Web electronic publication, www.who.int/malaria/world_malaria_report_2009/en/index.html, 4/2011.
- YANG Z.R., THOMSON R., MCNEIL P., ESNOUF R.M. 2005: RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics* 21: 3369–3376.
- ZAKERI S., SADEGHI H., ABOUIE MEHRIZI A., DINPARAST DJADID N. 2013: Population genetic structure and polymorphism analysis of gene encoding apical membrane antigen-1 (AMA-1) of Iranian *Plasmodium vivax* wild isolates. *Acta Trop.* 126: 269–279.

Received 30 September 2013

Accepted 29 March 2014