Vaccine 31 (2013) 1334–1339

Contents lists available at SciVerse ScienceDirect

Vaccine

journal homepage: www.elsevier.com/locate/vaccine

Within-population genetic diversity of *Plasmodium falciparum* vaccine candidate antigens reveals geographic distance from a Central sub-Saharan African origin

Kazuyuki Tanabe[a,b,*], Toshihiro Mita[c], Nirianne M.Q. Palacpac[b], Nobuko Arisue[b], Takahiro Tougan[b], Satoru Kawai[d], Thibaut Jombart[e], Fumie Kobayashi[f], Toshihiro Horii[b,*]

[a] *Laboratory of Malariology, Research Institute for Microbial Diseases, Osaka University, Osaka 565-0871, Japan*
[b] *Department of Molecular Protozoology, Research Institute for Microbial Diseases, Osaka University, Osaka 565-0871, Japan*
[c] *Department of Molecular and Cellular Parasitology, Juntendo University School of Medicine, Tokyo 113-8421, Japan*
[d] *Department of Tropical Medicine and Parasitology, Dokkyo Medical University, Tochigi 321-0293, Japan*
[e] *MRC Centre for Outbreak Analysis and Modelling, Department of Infectious Disease Epidemiology, School of Public Health, Faculty of Medicine, Imperial College, London W2 1PG, UK*
[f] *Department of Infectious Diseases, Kyorin University School of Medicine, Mitaka, Tokyo 181-8611, Japan*

ARTICLE INFO

Article history:
Received 20 September 2012
Received in revised form
15 November 2012
Accepted 14 December 2012
Available online 4 January 2013

Keywords:
Malaria
Plasmodium falciparum
Vaccine
AMA1
CSP
MSP1

ABSTRACT

Populations of *Plasmodium falciparum*, the most virulent human malaria parasite, are diverse owing to wide levels of transmission and endemicity of infection. Genetic diversity of *P. falciparum* antigens, within and between parasite populations, remains a confounding factor in malaria pathogenesis as well as clinical trials of vaccine candidates. Variation of target antigens in parasite populations may arise from immune pressure depending on the levels of acquired immunity. Alternatively, similar to our study in housekeeping genes [Tanabe et al. Curr Biol 2010;70:1–7], within-population genetic diversity of vaccine candidate antigens may also be determined by geographical distance from a postulated origin in Central sub-Saharan Africa. To address this question, we obtained full-length sequences of *P. falciparum* genes, apical membrane antigen 1 (*ama1*) ($n = 459$), circumsporozoite protein (*csp*) ($n = 472$) and merozoite surface protein 1 (*msp1*) ($n = 389$) from seven geographically diverse parasite populations in Africa, Southeast Asia and Oceania; and, together with previously determined sequences ($n = 13$ and 15 for *csp* and *msp1*, respectively) analyzed within-population single nucleotide polymorphism (SNP) diversity. The three antigen genes showed SNP diversity that supports a model of isolation-by-distance. The standardized number of polymorphic sites per site, expressed as θ_S, indicates that 77–83% can be attributed by geographic distance from the African origin, suggesting that geographic distance plays a significant role in variation in target vaccine candidate antigens. Furthermore, we observed that a large proportion of SNPs in the antigen genes were shared between African and non-African parasite populations, demonstrating long term persistence of those SNPs. Our results provide important implications for developing effective malaria vaccines and better understanding of acquired immunity against falciparum malaria.

1. Introduction

Malaria is a serious public health problem in the tropics with estimated 216 million cases and 655,000 malaria deaths in 2010 [1], due mostly to infections of *Plasmodium falciparum*, the most virulent human malaria parasite. To effectively contain this deadly disease, malaria vaccines are decisively needed. A few parasite developmental stages have been targeted for malaria vaccine development. Currently, various vaccine constructs based on surface proteins of sporozoites and merozoites are being developed [2]. Success in interventions using any of the developmental stages will require better understanding in genetic variation within and between parasite populations.

Recently we showed a strong negative correlation between within-population genetic diversity of the parasite housekeeping genes and geographic distance from a postulated origin in Central sub-Saharan Africa over Africa, Asia and Oceania [3]. Neither regional variation in transmission intensity nor malaria interventions using antimalarial drugs or insecticides appeared to be tightly associated with the geographic distribution of within-population parasite genetic diversity. Together with age estimation of parasite populations, the study lends credence to a Central sub-Saharan African origin of *P. falciparum* and colonization along with modern

* Corresponding authors at: Department of Molecular Protozoology, Research Institute for Microbial Diseases, Osaka University, 3-1 Yamada-Oka, Suita, Osaka 565-0871, Japan. Tel.: +81 6 6879 8279; fax: +81 6 6879 8281.
E-mail addresses: kztanabe@biken.osaka-u.ac.jp (K. Tanabe), tmita@juntendo.ac.jp (T. Mita), nirian@biken.osaka-u.ac.jp (N.M.Q. Palacpac), arisue@biken.osaka-u.ac.jp (N. Arisue), ttougan@biken.osaka-u.ac.jp (T. Tougan), skawai@dokkyomed.ac.jp (S. Kawai), t.jombart@imperial.ac.uk (T. Jombart), fumfum@ks.kyorin-u.ac.jp (F. Kobayashi), horii@biken.osaka-u.ac.jp (T. Horii).

0264-410X/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.vaccine.2012.12.039

human migrations out-of-Africa about 60,000 years ago [4,5]. These resulted in smooth patterns of reduction in parasite genetic diversity along with geographic distance from the African origin, akin to the distribution of human genetic diversity. However, *P. falciparum* possesses a number of antigens that are targets of host immunity. Immune target antigen genes are under natural selection [6,7]. In highly endemic areas where repeated infections are common, parasites are frequently exposed to host immunity, whereas in low endemic areas where infections are limited, parasites are less likely to be exposed to strong protective immunity. It is therefore likely that immune target antigen genes and housekeeping genes show unique geographic distribution of within-population genetic diversity. So far very few studies have compared geographic distribution of parasite genetic diversity between antigen genes and housekeeping genes in parasite populations from wide geographic regions.

With the seven previously collected geographically wide parasite populations from Africa, Asia and Oceania [3], we determined within-population genetic diversity of *P. falciparum* vaccine candidate genes and examined its correlation with geographic distance from a Central sub-Saharan African origin. We specifically analyze SNP diversity of three vaccine candidate genes coding for the apical membrane antigen 1 (AMA1), the circumsporozoite protein (CSP) and the merozoite surface protein 1 (MSP1). CSP is expressed during the sporozoite and early liver stage [8]. It is involved in the invasion of the liver cells. Antibodies against this immunodominant surface antigen inhibit parasite invasion and was associated with reduced risk of clinical malaria [9,10]. A CSP-based vaccine, RTS,S, showed 30–50% protection against clinical episodes of malaria [11]. AMA1 and MSP1, implicated in host erythrocyte invasion, are leading blood-stage vaccine candidates. AMA1 is localized in the micronemes of the merozoite [12], an apical organelle that plays an important role in parasite invasion [13]. Anti-AMA1 antibodies seem to be associated with protection from clinical malaria [14,15]. Various constructs of AMA1 polypeptides are currently being pursued for vaccine development/trials [2]. MSP1 is initially synthesized as a 200-kDa precursor that undergoes processing during schizont rupture to produce four fragments (83-, 30-, 38- and 42-kDa) [16]. Further cleavage of the C-terminal 42-kDa fragment, $MSP1_{42}$, yields a 19-kDa fragment, $MSP1_{19}$ [17]. Naturally exposed individuals in malaria-endemic areas have antibodies specific for $MSP1_{19}$ that correlated with lower risk to clinical malaria [18]. Intense research efforts have focused on both $MSP1_{42}$ and $MSP1_{19}$ as vaccine candidates [2]. In this study, we characterized the geographic distribution of SNPs in these three antigen genes and compared them to those obtained from housekeeping genes. Results clearly show a strong negative correlation between within-population genetic diversity of the three vaccine candidates and geographic distance from the Central sub-Saharan African origin over Africa, Asia and Oceania.

2. Materials and methods

2.1. Parasite isolates and DNA collection

A total of 531 *P. falciparum* isolates were collected from seven countries: Tanzania, Ghana, Thailand, Philippines, Papua New Guinea (PNG), the Solomon Islands and Vanuatu (Table 1). These seven parasite populations were also used for the polymorphism study in two housekeeping genes, sarcoplasmic/endoplasmic reticulum Ca^{2+}-transporting ATPase gene (*serca*) and adenylosuccinate lyase gene (*adsl*) [3]. Details of the parasite isolates and procedures for extraction of the parasite DNA have been described previously [3,19–23]. In all cases, ethical clearance for sampling was obtained from relevant ethical committees.

Table 1
Seven *P. falciparum* populations analyzed in this study.

Country	Local area	Sampling year	References
Tanzania	Rufiji River Delta, Eastern coast	1993, 1998, 2003	[23]
Ghana	Winneba, Western coast	2004	[3]
Thailand	Mae Sot, Northeastern Thailand	1995	[19]
Philippines	Palawan Island	1997	[22]
Papua New Guinea	Wewak, East Sepik, Northeast coast	2001, 2002	[3]
Solomon Islands	Northeastern Guadalcanal Island	1996	[21]
Vanuatu	Four islands (Malakula, Gaua, Santo, Pentecost)	1996–1998	[20]

2.2. DNA sequencing

To obtain full-length sequences of the *P. falciparum* AMA1 (*ama1*), CSP (*csp*) and MSP1 genes (*msp1*), genomic DNA was amplified by PCR using Takara LA *Taq* polymerase (Takara Bio, Japan). Procedures and conditions for PCR amplification have been described [24]. Two sets of primers were used in two successive reactions. For *ama1* amplification, the first PCR used primers ama1-F2 and ama1-R1 (Supplementary Table 1), followed by nested PCR with primers ama1-F3 and ama1-R1. *csp* was amplified with primers CS.F1 and CS.R0, followed by nested PCR with primers CS.F1 and CS.1332R. *msp1* was amplified with primers UPF1 and DWR1, followed by nested PCR with primers UPF3 and DWR3. DNA sequencing was performed as described elsewhere [24]. Sequences showing two or more peaks at the same position in an electropherogram were considered to be mixed infections and were excluded from further analysis. Sequences of *msp1* are known to be grouped into two major allelic types, MAD20- and K1-types in blocks 6–16, a central 3.5 kb region [25,26]. We obtained very limited numbers of K1 type in Africa ($n = 4$); and this scarcity made geographic comparison for K1 type sequences infeasible, and, thus, was excluded from further analysis.

2.3. Statistical analysis

Obtained in this study were 459 *ama1*, 472 *csp* and 389 *msp1* sequences. We included additional 13 *csp* and 15 *msp1* sequences from our previous study (Supplementary Table 2) (thus, total *n* for *csp* = 485; *msp1* = 404). DNA sequences were aligned as described [24] and sequence regions that do not allow unambiguous SNP identification such as tandem repeats and insertions/deletions (indel) were excluded. In *csp*, the central NANP/NVDP repeat region and a region containing a 57 bp indel upstream to the central region were excluded. In *msp1*, excised regions included blocks 2 and 4, and short regions with degenerated tandem repeats and indels in blocks 8 and 16 [26]. Sequences of *serca* and *adsl* [3] were used for comparison.

Nucleotide diversity was estimated as θ_S, the standardized number of polymorphic sites per site, and $\theta\pi$, the average pairwise nucleotide difference, as previously described [24]. The mean numbers of synonymous substitutions per synonymous site (dS) and non-synonymous substitutions per non-synonymous site (dN) were estimated [24]. If dN is significantly higher than dS, diversifying selection is inferred. Frequency of SNPs was computed using Arlequin v3.1 [27] and singleton SNPs as well as SNPs with minor allele frequency (MAF) <5% were calculated. SNPs were classified into either geographic area-specific SNP or those shared by multiple populations. Difference in frequency of SNPs among various groups was assessed by the Student's *t*-test. Correlation between within-population nucleotide diversity and geographic distance was tested