Infection, Genetics and Evolution 18 (2013) 213–219

Contents lists available at SciVerse ScienceDirect

Infection, Genetics and Evolution

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

Differential sequence diversity at merozoite surface protein-1 locus of *Plasmodium knowlesi* from humans and macaques in Thailand

Chaturong Putaporntip [a,*], Siriporn Thongaree [b], Somchai Jongwutiwes [a]

[a] Molecular Biology of Malaria and Opportunistic Parasites Research Unit, Department of Parasitology, Faculty of Medicine, Chulalongkorn University, Bangkok 10330, Thailand
[b] Hala-Bala Wildlife Research Station, National Park and Wildlife Research Division, Department of National Parks, Wildlife and Plant Conservation, Bangkok 10900, Thailand

ARTICLE INFO

Article history:
Received 10 January 2013
Received in revised form 16 May 2013
Accepted 20 May 2013
Available online 30 May 2013

Keywords:
Human-to-human transmission
Merozoite surface protein 1
Plasmodium knowlesi
Recombination
Mosaic organization
Phylogenetic networks

ABSTRACT

To determine the genetic diversity and potential transmission routes of *Plasmodium knowlesi*, we analyzed the complete nucleotide sequence of the gene encoding the merozoite surface protein-1 of this simian malaria (*Pkmsp-1*), an asexual blood-stage vaccine candidate, from naturally infected humans and macaques in Thailand. Analysis of *Pkmsp-1* sequences from humans ($n = 12$) and monkeys ($n = 12$) reveals five conserved and four variable domains. Most nucleotide substitutions in conserved domains were dimorphic whereas three of four variable domains contained complex repeats with extensive sequence and size variation. Besides purifying selection in conserved domains, evidence of intragenic recombination scattering across *Pkmsp-1* was detected. The number of haplotypes, haplotype diversity, nucleotide diversity and recombination sites of human-derived sequences exceeded that of monkey-derived sequences. Phylogenetic networks based on concatenated conserved sequences of *Pkmsp-1* displayed a character pattern that could have arisen from sampling process or the presence of two independent routes of *P. knowlesi* transmission, i.e. from macaques to human and from human to humans in Thailand.

1. Introduction

The fifth human malaria parasite, *Plasmodium knowlesi*, has been identified in both humans and macaques inhabiting Southeast Asian region (Jongwutiwes et al., 2004, 2011; Singh et al., 2004; Singh and Daneshvar 2013). Although the prevalence of human infection with *P. knowlesi* is generally less than those caused by *P. falciparum* and *P. vivax* (Putaporntip et al., 2009), recent reports on severe and fatal consequences of knowlesi malaria in humans have highlighted the public health importance of this simian parasite (Cox-Singh et al., 2008; William et al., 2011; Rajahram et al., 2013). Several species of mosquitoes can transmit *P. knowlesi* under experimental or natural conditions. Some of them such as *Anopheles dirus* and other '*leucosphyrus*' group also serve as the vectors for other human malaria parasites (Coatney et al., 1971; Vythilingam et al., 2006; Nakazawa et al., 2009); thereby it is not uncommon to detect co-infections of *P. knowlesi* and other human malaria species in several endemic areas (Putaporntip et al., 2009; Jongwutiwes et al., 2011). The natural hosts of *P. knowlesi* are long-tailed macaques (*Macaca fascicularis*), pig-tailed macaques (*Macaca nemestrina)* and leaf monkeys (*Presbytis* spp) that are populated in Southeast Asia (Coatney et al., 1971). Although *P. knowlesi* infections in humans are known to acquire from macaque natural hosts, experimental evidences indicate possible human-to-human transmission via mosquito vectors (Chin et al., 1968).

To better understand the transmission routes of *P. knowlesi*, analysis of polymorphic genetic markers may provide important information on this issue. Among the polymorphic loci of *Plasmodium*, the merozoite surface protein-1 genes of *P. falciparum* (*Pfmsp-1*) and *P. vivax* (*Pvmsp-1*) have been extensively studied in terms of structural, functional and immunological relevance for asexual blood stage vaccine development (Holder, 1988). The precursor of malarial msp-1 is synthesized during schizogony and undergoes two successive processing events. Msp-1 is encoded by a single copy gene whose sequence diversity is mainly dimorphic. Polymorphism in *msp-1* of *P. vivax*, the human malaria closely related genetically to *P. knowlesi*, is partly attributable to frequent meiotic intragenic recombination and balancing selection that probably due to host immune selective pressure (Putaporntip et al., 2002, 2006). Herein, we explored the extent of sequence variation in *Pkmsp-1* from human and macaque isolates in Thailand. Possible genetic mechanisms underlying the observed polymorphism on this gene locus has been analyzed in this study.

2. Materials and methods

2.1. Human and monkey populations

Blood samples (~1 ml) were obtained at malaria clinics from febrile individuals residing in major endemic provinces (Tak,

* Corresponding author. Tel.: +66 2 2564000x4761; fax: +66 2 2525944.
E-mail address: p.chaturong@gmail.com (C. Putaporntip).

1567-1348/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.meegid.2013.05.019

Prachuab Khirikhan, Chantaburi, Yala and Narathiwat) of Thailand in 1996 (n = 210), from 2006 to 2007 (n = 1874) and from 2008 to 2009 (n = 3770) as previously reported (Putaporntip et al., 2009; Jongwutiwes et al., 2011). From 2007 to 2009, blood samples (~1 ml) were collected from 754 monkeys (449 *M. nemestrina*, 294 *M. fascicularis*, seven *Semnopithecus obscurus* and four *M. arctoides*) inhabiting Prachuab Khirikhan, Ranong, Pattalung, Pattani, Yala and Narathiwat Provinces as described (Seethamchai et al., 2008; Putaporntip et al., 2010). These studies were approved by the Institutional Review Board of Faculty of Medicine, Chulalongkorn University. The monkey studies were conducted in compliance with standard of animal care and use established under ethical principles and guidelines for the use of animals for scientific purposes and policies of Chulalongkorn University.

2.2. Nested PCR diagnosis of human malaria and P. knowlesi

DNA samples from all human subjects were prepared for molecular analysis including nested PCR assays targeting the *18S rRNA* of all five human malaria species as described (Putaporntip et al., 2009; Jongwutiwes et al., 2011). For monkey blood samples, only primers specific to *18S rRNA* of *P. knowlesi* were used for nested PCR analysis.

2.3. PCR amplification and sequencing of the Pkmsp-1 gene

The complete nucleotide sequences of *Pkmsp-1* were determined directly from the PCR-amplified products as described (Putaporntip et al., 2006). Sequences have been deposited in the GenBank™ database under the accession numbers JF837339–JF837353 and JX046791–JX046798.

2.4. Data analysis

Alignment of the *Pkmsp-1* sequences was performed using the default option of the MUSCLE program (Edgar, 2004) and manually edited. Insertions/deletions (indels) in coding regions were determined from multiple alignments of amino acid sequences to maintain the reading frame. Sequences from the first naturally acquired human infection with *P. knowlesi* in Thailand (Jongwutiwes et al., 2004) and the H strain isolated from a human who had naturally acquired the infection while working in Peninsular Malaysia (Chin et al., 1965) were used for comparative analysis (accession numbers DQ220743 and XM002258546.1, respectively). Tandem repeats were detected by using the Tandem Repeats Finder version 4.0 program (Benson, 1999). Prediction whether amino acid substitutions could potentially affect protein function was done by using the Sorting Intolerant from Tolerant (SIFT) program (Ng and Henikoff, 2003).

Nucleotide diversity (π) was computed from the average number of pairwise sequence differences in the sample and the square root of the variance defines its standard deviation (Nei, 1987). Haplotype diversity (h) and its sampling variance were calculated according to equations 8.4 and 8.12 but n was replaced by $2n$ (Nei, 1987). The number of synonymous substitutions per synonymous site (d_S) and the number of nonsynonymous substitutions per nonsynonymous site (d_N) were computed using modified Nei and Gojobori's method (Nei and Gojobori, 1986) with Juke and Cantor correction (Jukes and Cantor, 1969). Significant differences in d_S and d_N (p value less than 0.05) indicate selective pressure on tested regions. Standard errors of these parameters were estimated by the bootstrap method with 1000 pseudoreplicates using the MEGA 5.05 program (Tamura et al., 2011).

The relationship between the r^2 measure of linkage disequilibrium and molecular distance was determined under random permutation of the physical position of the single nucleotide polymorphism (SNP) using 1000 permutations as implemented in the LDhat 2.1 package (McVean et al., 2002). Recombination parameter (R_m) or the minimum number of recombination events in the history of the sample (Hudson and Kaplan, 1985) was analyzed by using the DnaSP version 5.10.01 software (Librado and Rozas, 2009). Detection of recombination by phylogenetic approach was performed by using Genetic Algorithm Recombination Detection (GARD) and goodness of fit was assessed by Akaike Information Criterion derived from a maximum likelihood model fit to each segment (AICc) as implemented in the software package HyPhy (Pond et al., 2005, 2006). Parsimony informative sites were used in all analyses pertaining to recombination and linkage disequilibrium analysis.

Phylogenetic analysis was performed by the neighbor-net method (a distance-based method based on displaying the data in two dimensions) that generates an optimized result between the median networks (a character-based method based on displaying all differences among the taxa as separate branches) and split decomposition (a transformation-based approach by decomposing the data to a sum of weakly compatible splits); thereby, the preponderance of false positives (evolutionarily irrelevant reticulations) and false negatives (uninformative multifurcations) can be minimized (Hudson and Bryant, 2006). The neighbor-net networks were constructed based on various distance models (uncorrected-*p*, Juke-Cantor, Kimura-2-parameter, LogDet and F84) and the reliability of reticulation patterns was evaluated by bootstrapping using 1000 iterations as implemented in SplitsTree4 program (Hudson and Bryant, 2006). Likewise, a phylogenetic network was also constructed by the split decomposition method to reduce the visual complexity of the networks using least square optimization of branch lengths (Winkworth et al., 2005). Both sum of differences goodness of fit and least squares goodness of fit were calculated to examine how well the graph represents the distance of the data set (Winkworth et al., 2005).

3. Results

3.1. Amplification and sequencing of Pkmsp-1

Infections with *P. knowlesi* were diagnosed in 35 Thai patients (0.67% of all malaria positive cases) by nested *18S rRNA* PCR (Jongwutiwes et al., 2011). Of these, nine patients had *P. knowlesi* monoinfections and the remaining had co-infections with one or more malaria species (11 co-infected with *P. falciparum*, nine co-infected with *P. vivax* and five co-infected with *P. falciparum* and *P. vivax*) (Putaporntip et al., 2009; Jongwutiwes et al., 2011). *P. knowlesi* was diagnosed by *18S rRNA* PCR in 24 monkeys (3.18%) (three *M. fascicularis*, 20 *M. nemestrina*, and one *Semnopithecus obscurus*) (Putaporntip et al., 2010). *Pkmsp-1* was successfully amplified by PCR in 12 patients and 14 monkeys. However, sequencing revealed that two isolates from monkeys contained clonal mixture of different *Pkmsp-1* alleles due to the presence of superimposed signals in some segments intervening regions with clear electropherogram. Therefore, 24 *Pkmsp-1* sequences from field isolates in Thailand (Table 1) and a previously reported sequence XM002258546.1 were available for comparative analysis. The complete nucleotide sequences of *Pkmsp-1* varied from 5430 to 5613 bp.

3.2. Conserved domains in Pkmsp-1

Sliding window analysis of nucleotide diversity (window length = 100 sites and step size = 15 sites) along the entire 25 *Pkmsp-1* sequences reveals five conserved domains spanning ~4563 bp (Fig. 1) with an overall average π value of 0.0125 ± 0.0011 (Table 2). The nucleotide diversity in the conserved