Genetic Diversity of the *Plasmodium vivax* Circumsporozoite Protein (*Pvcsp*) in Sri Lanka

Sajani Dias[1], Thilan Wickramarachchi[1,*], Imeshi Sahabandu[1], Ananias A Escalante[2], Preethi V Udagama[1]

[1]*Department of Zoology, Faculty of Science, University of Colombo*

[2]*Center for Evolutionary Medicine & Informatics, The Biodesing Institute, Arizona State University, Tempe, AZ, USA*

**Current Affiliation - Microtech Biological (Pvt) Ltd, Pannipitiya, Sri Lanka*

Introduction

The circumsporozoite protein (CSP), sporozoite's major surface protein comprises a central repeat (CR) region, flanked with two conserved domains (Coppi et al., 2011). The protein is diverse across all *Plasmodium* species and consists of tandem arrays of relatively short amino acid motifs (Brito and Ferreira, 2011). *P. vivax* CSP (PvCSP) displays two major types of peptide repeat motifs (PRMs), each consisting of nine amino acids, GDRA[D/A]GQPA and ANGAGNQPG, defining variants VK210 and VK247, respectively (Rosenberg et al., 1989). Both variants are globally distributed, but geographic biases have been described (Leclerc et al., 2004).

We investigated the genetic diversity of *Pvcsp* in Sri Lanka, where low transmission and unstable malaria prevails. Local and global *P. vivax* isolates were analyzed for patterns of sequence variation in the *csp* gene by examining the polymorphism of the PRMs and the evolutionary relationships of the *Pvcsp* worldwide isolates were traced.

Materials and Methods

Single clone infections from two malaria endemic areas, Anuradhapura (N=17) and Kataragama (N=29) and from a malaria non-endemic area, Colombo (N=14) confirmed by genotyping at the polymorphic *Pvmsp-3α* locus using a combined nested polymerase chain reaction (PCR) and restriction fragment length polymorphism (RFLP) (Wickramarachchi et al., 2010), were used for this study.

The CR domain of *Pvcsp* was amplified by nested PCR as described by Imwong et al., (2005). Sequencing of the PCR products were done on both strands using VCS-NF and VCS-NR primers at Macrogen, Korea.

Measure of genetic polymorphism by nucleotide diversity (π) was calculated using DnaSP version 5.1 (Librado and Rozas, 2009). Natural selection was determined by the difference between the non-synonymous and synonymous substitutions (Dn − Ds) estimated by Nei and Gojobori's method (MEGA 5, Tamura et al., 2011) and recombination and linkage disequilibrium (LD) were tested by DnaSP 5.1.

The evolutionary relatedness of the available global sequences including Sri Lanka (VK210) was determined by i) the construction of a phylogenetic tree using the Maximum Likelihood (ML) method by MEGA 5, and ii) calculating the genetic differentiation between geographic regions via F_{ST} values using DnaSP 5.1.

Results

All of the 60 *P. vivax* isolates were of the VK210 variant consisting of variable repeats of 4 different PRMs, GDRA(A/D)GQPA, GNRAAGQPA and GNGAGGQAA. Three varying lengths sequences were observed where the 128 a.a. sequence (N=57; C2- C17) was significantly more common than both the 137 (N=1; C1) and the 119 (N=2; C18 & C19) a.a. sequences. Although the Sri Lankan isolates corroborated the variations in the peptide repeats motifs GDRA(A/D)GQPA of other global isolates, with different alternations of non-synonymous codons GCT or GAT, respectively, coding for Alanine (A) or Aspartic acid (D) (Leclerc et al., 2004), the PRM, GNRAAGQPA, observed in this study is unique to the island and is reported for the first time.

Four each from the eight and five RATs (different nucleotide sequences encoding the same PRM) identified by Patil et al (2010) for GDRADGQPA and GDRAAGQPA, respectively were detected from our study, where GGAGACAGAGCAGATGGACA**G**CCAGCA of the former a.a. sequence differed only by a single nucleotide polymorphism. Except for a single isolate, where the nucleotide sequence of GNGAGGQAA comprised a single nucleotide polymorphism **(**GGAAATGGTGCAGGTGGACA**T**GCAGCA), rest of the Sri Lanka isolates coincided with the nucleotide sequence observed by Patil et al., (2010).

The 19 amino acid haplotypes defined from the Sri Lankan population were exclusive to the island. Of these, 52% (N=31) was of the C4 a.a. haplotype. The 194 world wide isolates of VK210 variant obtained for this study from Azerbaijan, Brazil, China, Gabon, Iran, Korea, the Philippines, and the Solomon Islands, defined 57 a.a. haplotypes.

Elevated nucleotide diversity (π) was recorded from the Korean isolates followed by isolates from Brazil, Sri Lanka and Iran. The Sri Lankan population demonstrated a significant purifying selection at the CR region of PvCSP (Z test $P<0.05$), where excess of synonymous substitutions per site (Ds = 0.059 ± 0.014 S.D.) were observed compared to non-synonymous substitutions (Dn = 0.013 ± 0.004 S.D.). A linkage disequilibrium (LD) was maintained across the *Pvcsp* for the entire local population corroborating Patil et al., (2010).

The Sri Lankan population showed the highest degree of genetic differentiation with Iran (F_{ST} = 0.521) followed by Brazil (F_{ST} = 0.509) and Korea (F_{ST} = 0.434).

Phylogenetic tree drawn using all available VK210 worldwide isolates, defined 19 distinct groups. Most isolates were geographically clustered into groups. Of the Sri Lankan isolates, 2 from Kataragama were exclusive to group 5 while the rest depicted no clustering indicating origin from a common ancestor.

Conclusion

This study for the first time demonstrated the genetic diversity of the central repeat domain of the *P. vivax* Circumsporozite protein in Sri Lanka, where the polymorphism was due to point mutations, insertions and intragenic recombination. It has been

hypothesized that the presence of repeat arrays serve as a ‘‘smoke-screen’’ that elicits a strong but ineffective immune response on the part of the host (Kemp et al., 1987). The study further adds insights to the origin of *P. vivax* isolates from a common ancestor, where the RATs in the CR region showed a similar arrangement and some of the isolates in the phylogenetic tree revealed no clustering. The scarcity of non-synonymous polymorphisms at the CR region may not indicate a recent bottleneck in the evolutionary history of the *P. vivax* VK210 variant, rather the strain may be maintained in the population for a relatively long time.

References

Brito, C.F.A. *et al.,* 2011. Mem. Inst. Oswaldo. Cruz. 106(1), 12-26.

Coppi, A. *et al.,* 2011. J. Exp. Med. 208(2), 341-356.

Imwong, M. *et al.,* 2005. Malar. J. 4, 20.

Kemp, D.J. *et al.,* 1987. Annu. Rev. Microbiol. 41, 181–208.

Leclerc, M.C. *et al.,* 2004. Proc. Natl. Acad. Sci. USA. 101, 14455-14460.

Librado, P. *et al.,* 2009. Bioinformatics. 25, 1451-1452.

Patil, A. *et al.,* 2010. Infect. Genet. Evol. 10, 298-303.

Rosenberg, R. *et al.,* 1989. Science. 245, 973-976.

Tamura, K. *et al.,* 2011. Mol. Biol. Evol. doi: 10.1093/molbev/msr121.

Wickramarachchi, T. *et al.,* 2010. Ann. Trop. Med. Parasit. 104(2), 95–108.