COMPUTATIONAL DESIGN OF ANCESTRAL AND CONSENSUS
SEQUENCE OF APICAL MEMBRANE ANTIGEN 1 (AMA1) OF
Plasmodium spp.

R Nurdiansyah1*, RA Kemal2

1Department of Bioinformatics, School of Life Sciences, Indonesia International Institute for Life Sciences, Jakarta, Indonesia
2Department of Medical Biology, Faculty of Medicine, University of Riau, Pekanbaru, Indonesia

*Corresponding author: rizky.nurdiansyah@i3l.ac.id

ABSTRACT

Background: It is important to design a malaria vaccine targeting all human malaria parasites as well as non-human primate parasites to eradicate malaria and prevent zoonotic malaria. Apical membrane antigen 1 (AMA1) protein is shared by human-infecting Plasmodium species. Ancestral sequence reconstruction (ASR) and consensus sequence construction on AMA1 might be able to overcome the antigenic distinction between those species.

Objective: We aimed to computationally design the ancestral and consensus sequence of Plasmodium AMA1 protein and analyze the sequences for its putative immunogenicity.

Methods: We utilized bioinformatics software to computationally design ancestral and consensus sequences of AMA1 protein. AMA1 protein sequences of human-infecting Plasmodium and non-human primate Plasmodium were retrieved from PlasmoDB. ASR was designed using MEGA X while consensus was inferred using UGENE. Phylogenetic tree consisting of existing Plasmodium sequences and the ancestral sequence was constructed using IQTREE webserver and visualized with FigTree.

Results: Phylogenetic analysis showed that Plasmodium spp. were divided into 2 major groups, P. falciparum (Clade F) and non-falciparum (Clade NF) thus three ancestral and consensus sequences were designed based on each clade and both clades at once. Reconstructed ancestral sequences were located as sister branch for naturally occurring strains. On the contrary, consensus sequences are located within the branch of corresponding naturally occurring strains. Sequence analysis showed the presence of CD8+ T cell epitope in all computationally-designed sequences.

Conclusion: Ancestral and consensus AMA1 sequences are potential for further studies as a malaria vaccine candidate.

Keywords: AMA1, Ancestral sequence reconstruction, Consensus sequence, Plasmodium, Vaccine

Received Dec 12, 2019 ; Revised Jan 27, 2020 ; Accepted Jan 28, 2020
INTRODUCTION

Malaria is a persistent disease transmitted by parasite *Plasmodium spp.*[1] WHO reported there were 219 million cases with 435 thousand mortalities in 2017.[2] One way to cure malaria infection is through artemisinin combination therapy (ACT) which attacks the parasite in the blood-stage.[3] Unfortunately, the cases of ACT resistance are becoming prominent every year. Some studies reported ACT resistance in the South East Asia region.[4,5] Even though still partially resistant, the parasite will eventually become fully resistant if left unchecked. Moreover, there are also reports in the cross-species transmission from non-human primates to humans. *P. knowlesi* was once known for its infectious nature to macaque, but now it actually can infect the human in the Southeast Asia Region.[6,7] The potential zoonosis of malaria could be caused by the human habitation and also the adaptive nature of the parasite and vector.[8,9]

Those reports indicated that the parasite is evolving to gain an edge in infecting humans. To prevent that, a novel method needs to be devised in creating preventive or curative measures for malaria. In this sense, an evolutionary biology approach could be an alternative way. One of the approaches is using the ancestral and consensus sequence. Ancestral sequence reconstruction (ASR) is a tool to infer the primordial sequence from the contemporary sequences and represents the common ancestor for those sequences.[10] While the consensus sequence looked for the residues with the highest frequency at a certain position after multiple sequence alignment (MSA) of the extant sequences. Those residues at a given position reflect the relative importance for the whole sequence, such as common function or domain.[11] Several studies used this approach to design vaccines for viruses.[12,13] For example, ancestral and consensus sequences of HIV-1 envelope protein can be utilized to recognize the broader natural variant spectrum.[14]

A previous study in looking for a new target candidate found several proteins were shared in the *Plasmodium* species, one of them is apical membrane antigen 1 (AMA1).[15] This protein was found in the human infecting ones, including the newly zoonotic *Plasmodium* species, *P. knowlesi*, and several non-human infecting ones. AMA1 is expressed in the form 83-kD precursor and then cleaved to create a 66-kDa as an integral membrane protein with an ectoplasmic domain, a transmembrane domain, and a C-terminal cytoplasmic domain.[16] Interestingly, this protein is also one of the prime candidates for the new malaria vaccine in several malaria species, such as *P. falciparum* and *P. vivax.*[16–20] This is due to its location on the surface of malaria and one of the crucial protein for the infection properties of the parasite to red blood cells.[18,21] Additionally, this 622 amino acid (AA) long protein is expressed on both the liver and blood-stage, make it suitable for both anti-infection and anti-disease vaccine.[22,23]

It is also reported that the AMA1 has high antigenic diversity due to its sequence polymorphism[24]; a longitudinal study comparing the data from Mali with the published sequences in the database found about 200 unique haplotypes with some key changes of the amino acid residue in the putative invasion machinery binding site.[25] This could pose a challenge in creating the vaccine, even though most of the published studies only focus on *P. falciparum.*[23–25]
Interestingly, a study reported that the multi-allele AMA1 vaccine could give broad coverage against the diversity of AMA1, highlighting the need for a vaccine with a broad coverage.[24] To this end, the broad coverage vaccine could be achieved by targeting the conserved region in the protein.[25]

Based on those arguments, this study is trying to utilize the ancestral and consensus sequence on AMA1 protein to determine the potential vaccine candidate for several Plasmodium species at once. This approach mainly uses the phylogenetic analysis of AMA1 proteins from several species. In the end, the result could be valuable information in supporting the creation of the universal malaria vaccine.

**MATERIAL AND METHODS**

**Data mining**

AMA1 protein sequences from eight Plasmodium species were retrieved from the PlasmoDB database (https://plasmodb.org/) based on the previous data mining analysis.[15] Five plasmodia were known to infect humans (P. falciparum, P. vivax, P. knowlesi, P. ovale, and P. malariae) and the rest could infect non-human primates (P. coatneyi, and P. cynomolgi). One species infect murine (P. berghei) and served as outgroup. From those eight species, a total of 24 protein sequences were retrieved from the database (Table 1).

**Phylogenetic tree reconstruction**

The phylogenetic tree reconstruction was done twice in this study. The first one was to establish the relationship between the retrieved AMA1 sequences and to help in inferring the ancestral and consensus sequences. The first phylogenetic tree was reconstructed based on Hall’s protocol.[26] Multiple sequence alignment (MSA) was conducted using the MUSCLE algorithm[27] and then the model selection was conducted using the IQTREE server (http://iqtree.cibiv.univie.ac.at/)[28]. The maximum likelihood tree was reconstructed using the Jones, Taylor, and Thornton with gamma distribution (JTT+G) model based on the best model selector and 1000 bootstraps to check the tree robustness and validity. MEGA X software was used to reconstruct the first tree.[29] The second tree was made after the ancestral and consensus sequence of AMA1 was inferred. Different from the first one, the tree was made using the IQTREE server even though the model selection was using the same method as before.[28] JTTDCmut+F+G4 and 1000 bootstraps were used to reconstruct the second tree with the ancestral and consensus sequence. FIGTREE software was used to modify all of the trees for publication purposes.

**Ancestral and Consensus sequence inference and analysis**

The ancestral sequence of retrieved AMA1 was inferred using MEGA X based on the first phylogenetic tree and the default parameter from MEGA X.[29] After that, the ancestral sequence from the falciparum and non-falciparum were retrieved. Consensus sequences were inferred using the consensus function in the UGENE software with a strict 50% cutoff consensus.[30] The ancestral sequence and the consensus sequence for each of the clade and both clades were analyzed and retrieved to create the final tree. Ancestral and consensus sequences were aligned to find the conserved region.


<table>
<thead>
<tr>
<th>Sequence code</th>
<th>Accession Number</th>
<th>Sequence code</th>
<th>Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>P. berghei</em> ANKA</td>
<td>PBANKA_0915000</td>
<td><em>P. falciparum</em> IT</td>
<td>PHTT_110038000</td>
</tr>
<tr>
<td><em>P. knowlesi</em> strain H</td>
<td>PKNH_0931500</td>
<td><em>P. falciparum</em> KE01</td>
<td>PKE01_110038000</td>
</tr>
<tr>
<td><em>P. knowlesi</em> Malayan Strain Pki</td>
<td>PKNOH_S120150200</td>
<td><em>P. falciparum</em> KH01</td>
<td>PKH01_110037800</td>
</tr>
<tr>
<td><em>P. vivax</em> P01</td>
<td>PVP01_0934200</td>
<td><em>P. falciparum</em> VH02</td>
<td>PKH02_110038700</td>
</tr>
<tr>
<td><em>P. vivax</em> Sal-1</td>
<td>PVX_092275</td>
<td><em>P. falciparum</em> MD01</td>
<td>PKH02_110038000</td>
</tr>
<tr>
<td><em>P. falciparum</em> 7G8</td>
<td>Pf7G8_110037300</td>
<td><em>P. falciparum</em> SD01</td>
<td>PISD01_110036100</td>
</tr>
<tr>
<td><em>P. falciparum</em> CD01</td>
<td>PfCD01_110038900</td>
<td><em>P. falciparum</em> SN01</td>
<td>PFSN01_110036600</td>
</tr>
<tr>
<td><em>P. falciparum</em> Dd2</td>
<td>PfDd2_110036700</td>
<td><em>P. falciparum</em> TG01</td>
<td>PFTG01_110037900</td>
</tr>
<tr>
<td><em>P. falciparum</em> GA01</td>
<td>PFGA01_110037700</td>
<td><em>P. malariae</em> UG01</td>
<td>PnUG01_09042600</td>
</tr>
<tr>
<td><em>P. falciparum</em> GB4</td>
<td>PFGB4_110040000</td>
<td><em>P. ovale</em> curtisi GH01</td>
<td>PocGH01_09039800</td>
</tr>
<tr>
<td><em>P. falciparum</em> GN01</td>
<td>PfGN01_110038000</td>
<td><em>P. coatneyi</em> Hackeri</td>
<td>PCOAH_00026700</td>
</tr>
<tr>
<td><em>P. falciparum</em> HB3</td>
<td>PfHB3_110036900</td>
<td><em>P. cynomolgi</em> strain M</td>
<td>PcyM_0938200</td>
</tr>
</tbody>
</table>

The observed conserved region was analyzed for epitope presence available in the literature. Additionally, the sequences were analyzed using VaxiJen (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) for immunoprotective protein prediction with 0.5 thresholds.[31]

RESULTS

Phylogenetic trees

The first phylogenetic tree (Figure 1A) consisted of only natural sequences (retrieved from the PlasmoDB). It showed that the AMA1 sequences were clustered into the *P. falciparum* group (Clade F) and the non-falciparum one (Clade NF). *P. berghei* was used as the outgroup and therefore was not included in the ancestral and consensus inference (Figure 1A). The clustering served as the basis of the ancestral and consensus sequences inference. When ancestral and consensus sequences were included in phylogenetic tree construction, the same cluster pattern as observed (Figure 1B).

All of the ancestral sequences were located in the sister branch of the extant sequences, while the consensus sequences were located within the sequences. Interestingly, both of the ancestral and consensus sequences from every AMA1 clade were located in the middle of the phylogenetic tree, near the outgroup. The consensus sequence of all species resided in the falciparum cluster while the ancestral resided in the non-falciparum cluster (Figure 1B).

Ancestral and Consensus sequence epitope analysis

Ancestral and consensus sequences were analyzed for epitopes that have been previously characterized. Compared to CD8+ T cell epitopes TLDEMRHFY and NEVVVKEEY from *P. falciparum* AMA1, ancestral and consensus sequences have the 520NEVV(V/I)K(E/D)EY peptide (Figure 2).[23] Analysis with PROVEAN (provean.jcvi.org) showed that the V524I and E526D substitutions were neutral.

Computationally designed sequences were also analyzed for residues required for binding of the invasion-inhibitory monoclonal antibody, mAb 4G2, to *P. falciparum* AMA1.[32]
Figure 1. Phylogenetic tree of AMA1 sequences. A. Natural sequences. B. Natural sequences with its ancestral and consensus sequences. Red colored sequences: Consensus sequences. Blue colored sequences: Ancestral sequences.
Figure 2. The alignment of ancestral and consensus sequences showed a relatively conserved CD8+ epitope.

Figure 3. The alignment of ancestral and consensus sequences showed a relatively B-cell conserved epitope.

All of the sequences have conserved residues of Q352, F385, and D388. Consensus and ancestral sequences of the non-falciparum clade as well as the ancestral sequence for all clade had K351R and R389N substitutions. Analysis with PROVEAN showed that these substitutions were neutral. B-cell epitope characterized by *P. vivax* AMA1, SASDQPQYEQQHLTDYEK[33] was analyzed on the ancestral and consensus sequences. The epitope was present in all six sequences (Figure 3) with several substitutions. The epitope observed in the sequences was 345SASDQP(K/R)QYE(Q/E)(H/E)LTDYEK. PROVEAN analysis showed that the substitutions were neutral. Finally, analysis by VaxiJen showed that all computational sequences were considered as probable antigens with ancestral sequences that had a higher probability that consensus sequences (Table 2).

### DISCUSSION

The phylogenetic tree construction positioned the consensus sequence of all species in the falciparum cluster.

Table 2. AMA1 Sequences retrieved from the PlasmoDB database

<table>
<thead>
<tr>
<th>Sequence</th>
<th>VaxiJen Antigen probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consensus Clade Falciparum</td>
<td>0.5798</td>
</tr>
<tr>
<td>Consensus Clade Non-Falciparum</td>
<td>0.5957</td>
</tr>
<tr>
<td>Consensus All Clades</td>
<td>0.5386</td>
</tr>
<tr>
<td>Ancestral Clade Falciparum</td>
<td>0.6402</td>
</tr>
<tr>
<td>Ancestral Clade Non-Falciparum</td>
<td>0.6566</td>
</tr>
<tr>
<td>Ancestral All Clades</td>
<td>0.6520</td>
</tr>
</tbody>
</table>

The position of consensus sequence might due to the abundance of *P. falciparum* sequences in the database. However, even though the data mostly came from the *P. falciparum*, the ancestral sequence resides in the non-falciparum cluster. The ancestral AMA1 sequence might hint the evolutionary history of the *Plasmodium* species. This result is following the hypothesis of the evolution that the *Plasmodium* initially infected the non-human primates and then underwent zoonosis to humans.[8,34] The molecular pathway of this evolution was supported by an analysis of the ancestral sequence of *Plasmodium* RH5 protein.[34]
As one of the big three communicable diseases in the world, a lot of efforts have been done to combat malaria yet many challenges persist. The complexity of the *Plasmodium* spp. and its host-parasite interactions hinders the development in eradicating this parasite.[35] Interestingly, out of many proposed ideas, vaccine development has been considered to be the most feasible.[36–38] Some *Plasmodium* vaccine development has reached the trial version, even though the performance could be improved.[23,39,40] This, in turn, highlights the importance of the strong and long-lasting *Plasmodium* vaccine via the response of CD8+ T cells.[23] Besides the large size of the *Plasmodium* nuclear genome, the complex life cycle and the gene expression pattern of this species make it hard and challenging to do so.[41] In this regard, our target, AMA1 protein is expressed in both of life cycle during the human host period, the pre-erythrocytic which infects the liver and the blood-stage which infects the red blood cell, making it an interesting target in vaccine design.[42]

A putative AMA1 vaccine study detected CD8+ T cell response at epitopes TLDEMRHYF and NEVVVKEEY with the response frequency of 66.7% and 100%, respectively.[43,44] While we did not find the TLDEMRHY epitopes in any of computationally-designed sequences, our result using the human-infecting and non-human infecting species found the second CD8+ epitope, NEVV(V/I)K(E/D)EY, in domain III.[45] The presence of B-cell epitope in domain II[33] and recognition residues of mAb 4G2[32], as well as VaxiJen prediction for immunogenic protein, supported the hypothesis that all computationally-designed sequences to be immunogenic. However, this hypothesis needs to be further tested to develop a universal vaccine candidate against many human-infecting plasmodia.

**CONCLUSION**

This study provided the initial phase of the vaccine development of *Plasmodium* spp. based on the ancestral and consensus of AMA1 protein sequences. The clustering of the AMA1 sequences correlates with the current understanding of the host-parasite dynamics of *Plasmodium* spp. and it also revealed a relatively conserved epitope that could be recognized by the CD8+ cell, B-cell, and invasion-inhibitory antibody. Future studies should be focused on the potency of the conserved region as a vaccine candidate that could target many *Plasmodium* species at once.

**Acknowledgment**

The authors sincerely thank the I3L Department of Research and Community Service for the administrative support and DRPM DIKTI for the study funding. The research is funded by DRPM DIKTI funding with the scheme “Penelitian Dosen Pemula” in the 2019 funding year (Contract No: 48/AKM/MONOPNT/2019).

**REFERENCES**


