Analysis of immunogenicity of different protein groups from malaria parasite *Plasmodium falciparum*

Katarzyna Krzyczmonik (k.krzyczmonik@gmail.com)

Michał Świtnicki (indykpol@gmail.com)

Szymon Kaczanowski* (szymon@ibb.waw.pl)

*Corresponding author

Bioinformatics Department, Institute of Biochemistry and Biophysics, Polish Academy of Sciences, ul. Pawinskiego 5a, 02-106, Warszawa, Poland. Tel: +48 22 592 5761

Abstract

It was observed that pressure of host immune system leads to diversifying selection (which can be measured in terms of pN/pS ratio). In this research we checked whether *Plasmodium falciparum* proteins containing experimentally evident epitopes from the IEDB database are subject to diversifying selection. We also investigated which life stage of this parasite and which proteins are subject to the strongest immune pressure. To answer these questions we used information about experimentally evident epitopes from *Plasmodium*, that interact with human immune system and sequences of different isolates of *P. falciparum.*

We confirmed the expectations that proteins containing IEDB epitopes are subject to stronger diversifying selection which is evidenced by higher pN/pS ratio. A stage characterized by the highest average pN/pS ratio is that of the sporozoite. The greatest fraction of putative antigens is also present at this stage. We also found that the sporozoite stage is particularly interesting for further analysis as it potentially contains the highest number of unidentified epitopes.

1. Introduction

Malaria is one of the most serious parasitic infections affecting humanity. It causes about one million deaths and over 200 million infections globally each year (WHO World Malaria Report, 2010). There are four major species belonging to *Plasmodium* genus that were found to be responsible for causing malaria in humans: *P. falciparum, P. vivax, P. malariae* and *P. ovale.* Among them, *P. falciparum* causes the most severe and lethal form of this disease (Greenwood, 2005).

The life cycle of *P. falciparum* is complex since its nature requires expression of multiple specialized proteins necessary to survive in both invertebrate (mosquito) and vertebrate (human) host organisms as well as for the invasion of different types of cells (Florens et al., 2002). Due to the complexity of the life cycle, each life stage expresses a number of antigenic proteins that can induce an immune response and which presents a real challenge for vaccine design (Vaughan et al., 2009).

Another aspect, characteristic for *Plasmodium* parasites is their high genetic variability. The important factor that determines this variability is a number of single nucleotide polymorphisms (SNPs) (Ekland and Fidock, 2007; Jeffares et al., 2007; Volkman et al., 2007), or, more precisely, the ratio of non-synonymous to synonymous SNPs (pN/pS). Research has revealed that recombination rates are variable for *P. falciparum* strains and achieve the highest values for African strains (Mu et al., 2005). It was also observed that genes related to immunity are under stronger diversifying selection when compared to other genes as they have higher pN/pS ratio (Jeffares et al., 2007; Neafsey et al., 2008). Identification and characterization of such genes may be crucial for better understanding of the disease, its different clinical manifestations and potential resistance to vaccination (Richie and Saul, 2002).

The publication of the *Plasmodium falciparum* genome and the development of novel genetic and bioinformatics tools offer new possibilities for detailed analysis. Hitherto the sequences of approximately 5300 proteins were annotated for *Plasmodium falciparum* genome (Gardner et al., 2002). Among them, there were identified about 4500 unique epitopes (present in about 1500 proteins), which are short protein motifs recognized by the host immune system (Vaughan et al., 2009). All useful data about evident epitopes from *Plasmodium* and many other species are collected by IEDB – Immune Epitope Database and Analysis Resource (www.immuneepitope.org, Vita, 2009). In another useful resource known as PlasmoDB (www.plasmodb.org, Aurrecoechea et al., 2009) those epitopes were mapped in relationship to *Plasmodium* proteins. Such proteins are regarded as antigens as they can elicit response of the host immune system.

The aim of this research was to use data collected in PlasmoDB and IEDB to investigate whether these antigens are subject to diversifying selection. We have also employed an analysis of epitopes and SNPs to compare the immunogenicity across developmental stages and different protein classes from *Plasmodium*.

Recent study of Tetteth et. al (2009), based on deep resequencing, has shown that set of best known antigens are subject of balancing and diversifying selection. That was demonstrated by using different population genetic tests (McDonald-Kreitman test, HKA test, and Tajima's D). These results shows that comparison between currently existing diversity and genome of *P. reichenovi* may reveal immune pressure acting on pathogen after the speciation.

However, these tests require very good quality sequences obtained from many isolates. Unfortunately majority of accessible genomes are only rough draft and often available sequences do not cover the full gene length. Therefore it is hard to use this type of statistics for global

genome scan. This is why we decided to analyze the ratio of non-synonymous versus synonymous SNPs and check if this ratio is elevated in antigens. Another advantage of this approach is the fact that this it tests selection occurring recently during evolution of new infection strains, which are much more closely related than different species (*P. reichenowi* and *P. falciparum*). Therefore one could find it useful in epidemiological studies. Simplicity of this approach cause its robustness to unusual situations which may lead to the false signals in other tests (for example stronger pressure on antigen and higher speed of protein evolution during speciation of *Plasmodium falciparum* versus *Plasmodium reichenowi* than immune pressure observed now).

2. Methods

In the following research we aimed to investigate whether the immunogenicity of different malaria protein classes is correlated with differences in the frequency of molecular evolutionary changes observed in the available data. The whole analysis was divided into two parts – the first was focused on the general analysis of the proteome and the second on the investigation of evolutionary changes.

Plasmodium falciparum has a complex life cycle and each of its developmental stages is characterized by a unique protein expression profile. In the first part of the research we took into consideration individual merozoite, trophozoite, gametocyte, and sporozoite proteomes. We used the set of 2384 proteins described in PlasmoDB 8.1 which expression in investigated stages was confirmed by mass spectroscopy (Florens et al. 2002).

All the data used in the second part of the research were also obtained from the PlasmoDB project website. For this analysis the ratio of the number of non-synonymous single nucleotide polymorphisms (pN) to those of synonymous ones (pS) was necessary (See the table in

supplementary materials). Several different methods for investigation of natural selection have been developed. In this analysis we decided to use the pN/pS ratio for this purpose. For comparison we conducted also K_a/K_s ratio test (see supplementary materials). This method considers the correction for the number of synonymous and non-synonymous sites within the gene structure (Nei and Gojobori, 1986). We did no decide to use it as the main method in this research since such a correction is justified only in analysis of sequences obtained from different species since in case of different genes the probability of occurrence of synonymous mutation due to random chance is different. This is caused by the gene structure and codons composition. However, in the analysis of polymorphism there is another fundamental problem caused by the fact that often number of synonymous SNPs is equal to 0. We approximated such cases by number 1. Although such a correction is far from being perfect, it is hard to add justified correction for gene structure in such cases.

McDonald-Kreitman test (McDonald and Kreitman, 1991) and similar methods such as HKA test (Hudson et al., 1987) shows if non-synonymous protein mutations present as polymorphisms tend to be fixed in the longer period of time. When MK test indicates that such polymorphisms tend to be removed by selection it may indicate the action of diversifying selection or accumulation of deleterious polymorphisms, which are gradually removed by purifying selection. In case of *Plasmodium* it has been demonstrated by MK and HKA tests that well known antigens are subject of diversifying/purifying selection (Polley and Conway, 2001, Baum et al., 2002). In contrast it is expected that in case of good vaccine candidates polymorphisms are usually fixed. Immune pressure of host organism leads to elimination of non-mutated version of antigens and mutated version becomes more frequent. The pN/pS test is more general as it shows whether non-synonymous polymorphisms tend to occur in a given gene or sets of genes. It is

expected that elevated number of non-synonymous mutations manifested as polymorphisms in antigens is the adaptation to the pressure of host immune system and does not result only from relaxation of selection constraint.

To investigate different protein classes we divided all expressed proteins according to two different criteria. First of them was cellular location and the proteins were classified as membrane (containing at least one transmembrane domain or predicted signal peptide) and non-membrane. Prediction of signal peptides in PlasmoDB is provided by SignalIP program (Petersen et al., 2011) and transmembrane domains are predicted by TMHMM2 (Krogh et al., 2001), for both of them we used the PlasmoDB default settings. We decided to classify proteins containing signal peptides as membrane-associated because they have similar properties to those containing transmembrane domains (i.e. in both cases proteins can be expressed on the surface of the cell). The second classification (into antigenic and non-antigenic) was based on the antigenicity of proteins. Proteins are considered as antigenic if they contain experimentally evident epitopes (T cells, B cells or MHC binding) provided by IEDB website. In PlasmoDB there is an option ('Epitope Presence') which allows for identification of proteins containing epitopes from IEDB and we took advantage of this tool.

All necessary calculations and statistical analysis were done in R-project. In order to analyze the data we have applied the χ^2 test, the bootstrap confidence intervals, Wilcoxon test (known also as Mann-Whitney U test) and Kruskal-Wallis test.

3. Results and Discussion

3.1. Proteome analysis

We have found a set of 161 proteins common to all investigated parasite lifecycle stages (which is 6.75 % of all proteins). Some other are shared between two or three stages (respectively 560

and 310 proteins) and the rest 1353 (56.75 %) are proteins unique only to one form (stage-specific proteins).The results are presented by the Venn diagram in Fig. 1A.

Additional analysis was performed according to the cellular location of proteins. We have distinguished membrane-associated and non-membrane proteins. The same division we did for the whole proteome was done for those two groups and presented in the Fig. 1B and 1C.

Another division important for this research took into consideration the presence of epitopes. The PlasmoDB website provides proteins containing experimentally evident epitopes mapped on the basis of the results collected in the IEDB database. Fig. 2. presents the percentage distribution of antigenic and non-antigenic proteins in the main four developmental stages taking into consideration all the proteins expressed in each stage. The plot in the Fig. 3. shows the same distribution but considers only proteins unique to the given form.

From the above plots (Fig 2 and 3) it can be noticed that sporozoites are characterized by the highest number of proteins containing epitopes in both the case of membrane-associated and non-membrane proteins. However this tendency is stronger if we consider only stage-specific proteins. We used the χ^2 test to assess the statistical significance of this distribution. The results revealed that the type of expressed proteins is strongly dependent on the developmental stage of *P. falciparum.* Significantly higher number of antigenic proteins expressed in sporozoite may be explained by the fact that this form of parasite is able to infect both mosquito and human. It means that the parasite must be able to attack different target cells and for this purpose more specialized proteins may be required. (Garcia et al., 2006).

3.2. Analysis of diversifying selection

For this analysis we took into consideration only stage specific proteins i.e. proteins that are expressed exclusively in one developmental stage. We started from the estimation of an average

ratio of the number of non-synonymous single nucleotide polymorphisms (pN) to that of synonymous ones (pS) for each developmental stage of *P. falciparum*.

We calculated the average pN/pS ratio values for all stage specific proteins expressed in all investigated stages and we got the following results: 2.10 (merozoite), 2.26 (trophozoite), 1.92 (gametocyte) and 2.91 (sporozoite). The results are presented in the Fig. 4. We used Kruskal-Wallis test to check whether the distribution in all four developmental stages is the same or not. We obtained p-value equal to $5.09 \cdot 10^{-6}$, which means that there are statistically important differences between these groups. Further investigation by pairwise comparison of all stages between each other revealed that sporozoite is significantly different from other stages and that there are no significant differences between merozoite, trophozoite and gametocyte To further analyze the estimates of pN/pS values, we investigated how they differ across the 4 considered stages regarding the distinction of proteins with epitopes. The average values of pN/pS ratio for antigenic and non-antigenic proteins in all developmental stages are presented in the Fig. 5. Confidence intervals were calculated by bootstrap method with 10000 random permutations of the data. We used Wilcoxon rank sum test to evaluate the differences in means between antigenic and non-antigenic proteins. The results indicate that in all four stages there is a statistically significant difference between pN/pS values for antigenic and non-antigenic proteins. We obtained the following p-values: $1.96 \cdot 10^{-6}$ (merozoite), $7.85 \cdot 10^{-5}$ (trophozoite), $4.85 \cdot 10^{-14}$ (gametocyte), $6.89 \cdot 10^{-8}$ (sporozoite). These observations are consistent with the theory that in the case of antigenic proteins, positive natural selection is expected in order to maintain genetic diversity (Escalante et al., 2004).

In the next step of the research, we divided the same set of proteins into different classes. This time we analyzed the membrane-associated and non-membrane proteins in all developmental

stages (Fig. 6). Again the Wilcoxon test was used to assess whether the difference in means is significant. In this case only we obtained significant p-value only for trophozoite ($P=9.04 \cdot 10^{-3}$). The shape of the plot suggests that the dependencies are similar to those presented by the previous plot although not significantly so. The observation that genes coding for membrane-associated proteins in *Plasmodium* are characterized by higher pN/pS ratio has been previously demonstrated by Volkman et al. (2002). The fact that membrane-associated proteins can be antigenic is not surprising. However, some non-membrane proteins are also able to induce an immune response for example in case of autoimmune diseases (Dumont, Payseur, 2011 and Worman, Courvalin, 2003)We have also noticed that there is a certain problem with searching for *Plasmodium* membrane proteins in PlasmoDB. The database's tools allow only for identification of proteins containing at least one transmembrane domain and it does not take into consideration GPI-anchored proteins, which should also be classified as membrane-associated. The research done by Gilson et al. (2006) shows that there exist some proteins that do not possess transmembrane domains but are GPI-anchored. This fact can slightly influence the values that we obtained in our analysis. Moreover, it may be expected that at least some proteins that we have identified as non-membrane antigens are in fact GPI-anchored proteins. For example merozoite surface protein 2 (MSP-2) coded by gene PFB0300c (Gilson et al., 2006). From each developmental stage taken into account, we selected antigenic membrane-associated and non-membrane proteins that are characterized by the highest pN/pS value. The results are presented in tables 1 and 2 (full list of pN/pS scores is provided as supplementary material). Such proteins contain both experimentally evident epitopes and elevated pN/pS ratio indicated that they are antigens which are subject of the strongest diversifying selection in a given stage.

Table. 1. Membrane-associated proteins with epitopes found to be under strong diversifying selection

Gene ID	Stage	Product name	GO function	GO process	pN/pS ratio
PFD0230c	Merozoite	dipeptidyl peptidase 3			11.0
PFB0105c	Merozoite	Plasmodium exported protein (PHISTc), unknown function	null	null	9.00
PFL2565w	Merozoite	Plasmodium exported protein (PHISTa), unknown function, pseudogene	metal ion binding, transferase activity	null	7.00
PF13_0338	Merozoite	cysteine-rich surface protein (Pf92)	null	attachment of GPI anchor to protein	7.00
PFL1330c	Merozoite	cyclin (CYC2)	null	negative regulation of cyclin-dependent protein kinase activity, regulation of cell cycle	6.33
PFL0050c	Trophozoite	Plasmodium exported protein (PHISTb), unknown function	ATP binding	null	15.00
PFB0675w	Trophozoite	conserved Plasmodium membrane protein, unknown function	null	null	15.00
PFL0035c	Trophozoite	acyl-CoA synthetase (ACS7)	catalytic activity	metabolic process	14.00
PFA0620c	Trophozoite	glutamic acid-rich protein (GARP)	null	null	9.00
PF10_0160	Trophozoite	serine/threonine protein kinase, FIKK family (FIKK10.1)	ATP binding, protein serine/threonine kinase activity	protein phosphorylation	9.00
PFA0340w	Gametocyte	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase, putative	catalytic activity	translation, translational initiation	16.00
PF13_0248	Gametocyte	6-cysteine protein (P47)	null	attachment of GPI anchor to protein	14.00
PFB0190c	Gametocyte	conserved Plasmodium protein, unknown function	binding, transferase activity	null	10.00
PF14_0313	Gametocyte	conserved Plasmodium protein, unknown function	ATP binding, DNA binding	cell cycle, cell division, chromosome segregation	6.00
PF14_0732	Gametocyte	Plasmodium exported protein (PHISTb), unknown function	nucleotide binding	null	6.00
PF10_0095	Gametocyte	conserved Plasmodium membrane protein, unknown function	null	null	6.00
PFE1180c	Gametocyte	conserved Plasmodium membrane protein, unknown function	null	null	6.00
PF13_0201	Sporozoite	sporozoite surface protein 2	null	cellular component movement	50.00
PFL0440c	Sporozoite	zinc finger protein, putative	protein binding, zinc ion binding	null	25.00

PFL0800c	Sporozoite	cell traversal protein for ookinetes and sporozoites	null	null	22.00
PFB0920w	Sporozoite	DnaJ protein, putative	heat shock protein binding, unfolded protein binding	protein folding	14.00
PF14_0736	Sporozoite	Plasmodium exported protein, unknown function	ATP binding, ATPase activity, coupled to transmembrane movement of ions, phosphorylative mechanism	ATP biosynthetic process	14.00

Table. 2. Non-membrane proteins with epitopes found to be under strong diversifying selection

Gene ID	Stage	Product name	GO function	GO process	pN/pS ratio
PFD0180c	Merozoite	CGI-201 protein, short form	RNA binding	RNA processing, spliceosome assembly	24.00
PFB0300c	Merozoite	merozoite surface protein 2	null	attachment of GPI anchor to protein, cell adhesion	19.00
PF14_0379	Merozoite	conserved Plasmodium protein, unknown function	ATP binding, actin binding, motor activity	null	16.00
PF11_0362	Merozoite	protein phosphatase, putative	catalytic activity	null	10.00
PF14_0537	Merozoite	conserved Plasmodium protein, unknown function	ATP binding, arginine-tRNA ligase activity, glycine-tRNA ligase activity	arginyl-tRNA aminoacylation, glycyl-tRNA aminoacylation, translation	8.00
PF14_0374	Trophozoite	CCAAT-binding transcription factor, putative	sequence-specific DNA binding	null	14.00
PFE1320w	Trophozoite	conserved Plasmodium protein, unknown function	binding	RNA splicing, cell cycle, mRNA processing	13.50
PFL2010c	Trophozoite	DEAD/DEAH box ATP-dependent RNA helicase, putative	ATP binding, ATP-dependent RNA helicase activity, nucleic acid binding	RNA metabolic process	13.00
PF10_0278	Trophozoite	nucleolar preribosomal assembly protein, putative	null	null	11.00
PFF0480w	Trophozoite	conserved Plasmodium protein, unknown function	null	null	9.67
PF07_0069	Gametocyte	conserved Plasmodium protein, unknown function	null	null	16.00
PFE0780w	Gametocyte	conserved Plasmodium protein, unknown function	null	phosphate transport	13.00
MAL13P1.260	Gametocyte	alveolin, putative	transferase activity	null	12.00
PF10_0078	Gametocyte	histone deacetylase, putative	inositol trisphosphate 3-kinase activity	null	11.50
PFF1065c	Gametocyte	conserved Plasmodium protein, unknown function	null	null	11.00
PF14_0135	Gametocyte	conserved Plasmodium protein, unknown function	null	null	11.00

PF14_0727	Sporozoite	conserved Plasmodium protein, unknown function	null	null	46.00
PF14_0073	Sporozoite	conserved Plasmodium protein, unknown function	ATP binding, protein binding	null	18.00
PFL1565c	Sporozoite	CG2-related protein, putative	null	null	17.00
MAL7P1.76	Sporozoite	conserved Plasmodium protein, unknown function	null	null	17.00
PF13_0120	Sporozoite	conserved Plasmodium protein, unknown function	ATP binding, actin binding, calmodulin binding, motor activity	null	15.00
MAL13P1.296	Sporozoite	conserved Plasmodium protein, unknown function	hydrolase activity	null	15.00

4. Conclusions

Our aim was to analyze immune pressure acting on different stages of malaria life cycle. We identified proteins which are under putative immune pressure taking into consideration two criteria: presence of epitopes from the IEDB and pN/pS ratio, which is an approximate measure of host immune pressure. We confirmed the expectations that proteins which contain epitopes have elevated pN/pS ratio. This observation suggests that it is likely that both methods detect the pressure of host immune system. Using this approach we asked the question, what was the ecology of interactions between host immune system and parasite. Comparison of plots showing pN/pS ratios for different protein classes (Fig. 5. and 6.) confirmed that antigenicity has stronger impact on diversifying selection than cellular location. This observation confirms that hypothesis that elevated pN/pS ratio of membrane proteins is a result of immune pressure (Volkman 2002) is true. The same analysis repeated with K_a/K_s ratio test gave mostly the same dependencies which suggests that both the method and the results are correct.

As examples of proteins which could be interesting for further investigation, we have identified five to seven proteins from each developmental stage with the highest pN/pS values. Membrane-

associated and non-membrane proteins were distinguished for each stage. The data are presented in the tables 1 and 2.

The results obtained also indicate that among all investigated developmental stages, that of the sporozoite contains the highest fraction of antigenic proteins (42.77 % for all proteins and 41.83 % for stage-specific ones). This form is also under particularly strong immune pressure which is indicated by the highest pN/pS values. Since the proteins expressed in this stage - currently believed to be non-antigenic - are characterized by higher pN/pS value than those from other stages, we suspect that a high number of sporozoite-specific epitopes remains undiscovered.

Acknowledgments

Funding: The work was supported by COST action BM0802 and grant 772/N-COST/2010 from the Polish Ministry of Science. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Figures captions

Fig. 1. Distribution of *Plasmodium falciparum* proteins across four developmental stages. A. All proteins. B. Non-membrane proteins. C. Membrane-associated proteins. Numbers represent the amount of genes expressed in the particular life stages. Data were obtained from PlasmoDB website (Florens et al., 2002).

Fig. 2. Distribution of all proteins in four main stages, categorized into different groups: non-membrane without epitopes, non-membrane with epitopes, membrane without epitopes and membrane with epitopes.

Numbers in the boxes represent the amount of protein coding genes expressed in the particular life stages. Different classes of proteins encoded by the respective genes are marked with different colors. Data were obtained from PlasmoDB website (Florens et al., 2002).

Fig. 3. Distribution of stage-specific proteins in four main stages, categorized into different groups: non-membrane without epitopes, non-membrane with epitopes, membrane without epitopes and membrane with epitopes.

Numbers in the boxes represent the amount of protein coding genes expressed in the particular life stages. Different classes of proteins encoded by the respective genes are marked with different colors. Data were obtained from PlasmoDB website (Florens et al., 2002).

Fig. 4. Estimates of pN/pS of stage- specific proteins across the four main *P. falciparum* stages.

Mean values and 95 % confidence intervals of the pN/pS ratio were calculated for each developmental stage.

Fig. 5. Comparison between estimates of pN/pS across all considered forms with divisions into antigenic and non-antigenic protein classes.

Fig. 6. Comparison between estimates of pN/pS across all considered forms with divisions into non-membrane and non-membrane protein classes.

References

Aurrecoechea, C., Brestelli, J., Brunk, B. P., Dommer, J., Fischer, S., Gajria, B., Gao, X., Gingle, A., Grant, G., Harb, O. S., Heiges, M., Innamorato, F., Iodice, J., Kissinger, J. C., Kraemer, E., Li, W., Miller, J. A., Nayak, V., Pennington, C., Pinney, D. F., Roos, D. S., Ross, C., Stoeckert, C. J., Treatman, Wang, H., 2009. PlasmoDB: a functional genomic database for malaria parasites. Nucleic Acids Res. 37(Database issue), D539-543.

Baum, J., Thomas, A. W., Conway, D. J., 2003. Evidence for diversifying selection on erythrocyte-binding antigens of Plasmodium falciparum and P. vivax. Genetics. 163(4), 1327-1336.

Dumont, B.L., and Payseur, B.A. ,2011. Genetic analysis of genome-scale recombination rate evolution in house mice. PLoS Genet 7, e1002116.

Ekland, E. H., Fidock, D. A., 2007. Advances in understanding the genetic basis of antimalarial drug resistance. Curr Opin Microbiol. 10(4), 363-370.

Escalante, A. A., Cornejo, O. E., Rojas, A., Udhayakumar, V., Lal, A. A., 2004. Assessing the effect of natural selection in malaria parasites. Trends Parasitol. 20(8), 388-395.

Florens, L., Washburn, M. P., Raine, J. D., Anthony, R. M., Grainger, M., Haynes, J. D., Moch, J. K., Muster, N., Sacci, J. B., Tabb, D. L., Witney, A. A., Wolters, D., Wu, Y., Gardner, M. J., Holder, A. A., Sinden, R. E., Yates, J. R., Carucci, D. J., 2002. A proteomic view of the Plasmodium falciparum life cycle. Nature. 419(6906), 520-526.

Garcia J. E., Puentes A., Patarroyo M. E.,2006, Developmental biology of sporozoite-host interactions in plasmodium falciparum malaria: Implications for vaccine design. Clin Microbiol Rev 19(4), 686-707.

Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, K. E., Bowman, S., Paulsen, I. T., James, K., Eisen, J. A., Rutherford, K., Salzberg, S. L., Craig, A., Kyes, S., Chan, M. S., Nene, V., Shallom, S. J., Suh, B., Peterson, J., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., Haft, D., Mather, M. W., Vaidya, A. B., Martin, D. M., Fairlamb, A. H., Fraunholz, M. J., Roos, D. S., Ralph, S. A., McFadden, G. I., Cummings, L. M., Subramanian, G. M., Mungall, C., Venter, J. C., Carucci, D. J., Hoffman, S. L., Newbold, C., Davis, R. W., Fraser, C. M., Barrell, B., 2002. Genome sequence of the human malaria parasite Plasmodium falciparum. Nature. 419(6906), 498-511.

Gilson P. R., Nebl T., Vukcevic D., Moritz R. L., Sargeant T., Speed T.P., Schofield L., Crabb B. S., 2006, Identification and stoichiometry of glycosylphosphatidylinositol-anchored membrane proteins of the human malaria parasite plasmodium falciparum. Mol Cell Proteomics 5(7),1286-99.

Greenwood, B. M., Bojang, K., Whitty, C. J., Targett, G. A., 2005. Malaria. Lancet. 365(9469), 1487-1498.

Hudson, R. R., Kreitman, M., Aguadé, M., 1987. A test of neutral molecular evolution based on nucleotide data. Genetics. 116(1), 153-159.

Jeffares, D. C., Pain, A., Berry, A., Cox, A. V., Stalker, J., Ingle, C. E., Thomas, A., Quail, M. A., Siebenthall, K., Uhlemann, A. C., Kyes, S., Krishna, S., Newbold, C., Dermitzakis,

E. T., Berriman, M., 2007. Genome variation and evolution of the malaria parasite Plasmodium falciparum. Nat Genet. 39(1), 120-125.

Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E. L., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 305(3), 567-580.

McDonald, J. H., Kreitman, M., 1991. Adaptive protein evolution at the Adh locus in Drosophila. Nature. 351(6328), 652-654.

Mu, J., Awadalla, P., Duan, J., McGee, K. M., Joy, D. A., McVean, G. A., Su, X. Z., 2005. Recombination hotspots and population structure in Plasmodium falciparum. PLoS Biol. 3(10), e335.

Neafsey, D. E., Schaffner, S. F., Volkman, S. K., Park, D., Montgomery, P., Milner, D. A., Lukens, A., Rosen, D., Daniels, R., Houde, N., Cortese, J. F., Tyndall, E., Gates, C., Stange-Thomann, N., Sarr, O., Ndiaye, D., Ndir, O., Mboup, S., Ferreira, M. U., Moraes, S. o. L., Dash, A. P., Chitnis, C. E., Wiegand, R. C., Hartl, D. L., Birren, B. W., Lander, E. S., Sabeti, P. C., Wirth, D. F., 2008. Genome-wide SNP genotyping highlights the role of natural selection in Plasmodium falciparum population divergence. Genome Biol. 9(12), R171.

Nei, M., Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol. 3(5), 418-426.

Petersen, T. N., Brunak, S., von Heijne, G., Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 8(10), 785-786.

Polley, S. D., Conway, D. J., 2001. Strong diversifying selection on domains of the Plasmodium falciparum apical membrane antigen 1 gene. Genetics. 158(4), 1505-1512.

Richie, T. L., Saul, A., 2002. Progress and challenges for malaria vaccines. Nature. 415(6872), 694-701.

Tetteh, K. K., Stewart, L. B., Ochola, L. I., Amambua-Ngwa, A., Thomas, A. W., Marsh, K., Weedall, G. D., Conway, D. J., 2009. Prospective identification of malaria parasite genes under balancing selection. PLoS One. 4(5), e5568.

Vaughan, K., Blythe, M., Greenbaum, J., Zhang, Q., Peters, B., Doolan, D. L., Sette, A., 2009. Meta-analysis of immune epitope data for all Plasmodia: overview and applications for malarial immunobiology and vaccine-related issues. Parasite Immunol. 31(2), 78-97.

Vita, R., Zarebski, L., Greenbaum, J. A., Emami, H., Hoof, I., Salimi, N., Damle, R., Sette, A., Peters, B., 2010. The immune epitope database 2.0. Nucleic Acids Res. 38(Database issue), D854-862.

Volkman, S. K., Hartl, D. L., Wirth, D. F., Nielsen, K. M., Choi, M., Batalov, S., Zhou, Y., Plouffe, D., Le Roch, K. G., Abagyan, R., Winzeler, E. A., 2002. Excess polymorphisms in genes for membrane proteins in Plasmodium falciparum. Science. 298(5591), 216-218.

Volkman, S. K., Sabeti, P. C., DeCaprio, D., Neafsey, D. E., Schaffner, S. F., Milner, D. A., Daily, J. P., Sarr, O., Ndiaye, D., Ndir, O., Mboup, S., Duraisingh, M. T., Lukens, A., Derr, A., Stange-Thomann, N., Waggoner, S., Onofrio, R., Ziaugra, L., Mauceli, E., Gnerre, S., Jaffe, D. B., Zainoun, J., Wiegand, R. C., Birren, B. W., Hartl, D. L., Galagan, J. E., Lander, E. S., Wirth, D. F., 2007. A genome-wide map of diversity in Plasmodium falciparum. Nat Genet. 39(1), 113-119.

Worman, H.J., and Courvalin, J.C., 2003. Antinuclear antibodies specific for primary biliary cirrhosis. Autoimmun Rev 2, 211-217.