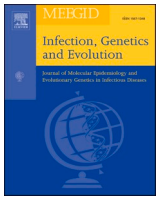
Infection, Genetics and Evolution 107 (2023) 105397

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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



Detection of positive selection acting on protein surfaces at the whole-genome scale in the human malaria parasite *Plasmodium falciparum*

Szymon Kaczanowski

Department of Bioinformatics, Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warsaw, Poland

ARTICLE INFO

Keywords:
Positive selection
AlphaFold
Malaria parasites
dN/dS
Evolutionary arms race

ABSTRACT

The host–parasite evolutionary arms race is a fundamental process with medical implications. During this process, the host develops parasite resistance, and the parasite develops host immune evasion strategies. Thus, this process accelerates relevant protein evolution. This study test hypothesizes that proteins subject to sequence evolution structural constraints play a crucial role and that these constraints hinder the modification of such proteins in this process. These hypotheses were tested using *Plasmodium falciparum* model and evaluated protein structures predicted for the entire proteome by the AlphaFold method. Based on *dN/dS* test results and *P. falciparum* and *P. reichenowi* comparisons, the presented approach identified proteins subject to purifying selection acting on the whole sequence and buried residues ($dN < dS$) and positive selection on nonburied residues. Of the 26 proteins, some known antigens (ring-exported protein 3, RAP protein, erythrocyte binding antigen-140, and protein P47) targeted by the host immune system are promising vaccine candidates. The set also contained 11 enzymes, including FIKK kinase, which modifies host proteins. This set was compared with genes for which the *dN*/dS test suggested that positive selection acts on the whole gene (i.e., $dN > dS$). The present study found that such genes encode enzymes and antigenic vaccine candidates less frequently than genes for which evolution is not subject to selection constraints and positive selection acts on only exposed residues. The analysis was repeated comparing *P. falciparum* with *P. alderi*, which is more distantly related. The study discusses the potential implications of the presented methodology for rational vaccine design and the parasitology and evolutionary biology fields.

1. Introduction

The evolutionary arms race between hosts and parasites is a fundamental phenomenon in evolutionary ecology. In the case of human pathogens, this process has medical implications.

During this continuous process, the host acquires resistance through the evolution of host genes. In the case of vertebrates such as humans, the host may even gain immunity due to the activity of the adaptive immune system. However, the evolution of parasites leads to the origination of host resistance evasion mechanisms. This mode of evolution is described metaphorically as the Red Queen phenomenon(Van Valen, 1973).

According to this metaphor, the evolutionary race is similar to the race between the Red Queen and Alice described in Lewis Carroll's book Through the Looking-Glass. During this race, both the Red Queen and Alice had to run as fast as they could just to stay in the same place.

I hypothesize that genes encoding proteins that are subject to structural constraints on evolution, such as enzymes or factors involved in parasitic invasion, play essential roles in this process. Here, the structural constraints on evolution and positive selection acting on constrained genes were identified.

The detection of positive selection requires statistical methods, among which one of the most fundamental and widely used is the analysis of the *dN/dS* ratio (Yang and Bielawski, 2000). This method takes advantage of the fact that there are two types of mutations in protein-encoding genes: synonymous (S) and nonsynonymous (N). Assuming that synonymous mutations are neutral, the type of selection (neutral, positive, or purifying) acting on the protein can be determined from the ratio of the two classes of mutations observed in that protein.

The accumulation of both types of mutations is described by the ratio of the nonsynonymous substitution rate (*dN*) to the synonymous substitution rate (*dS*), *dN/dS*; the two rates are determined via the alignment of coding regions of homologous genes. When it is evolutionarily beneficial for a protein sequence to rapidly change, nonsynonymous mutations are observed more frequently than synonymous mutations (i.e., $dN/dS > 1$). Conversely, when a change in the protein sequence is

E-mail address: szymon@ibb.waw.pl.

https://doi.org/10.1016/j.meegid.2022.105397
Received 27 April 2022; Received in revised form 20 September 2022; Accepted 21 December 2022
Available online 23 December 2022

deleterious, synonymous mutations accumulate more rapidly than nonsynonymous mutations (i.e., $dN/dS < 1$). Numerous studies have revealed that the majority of proteins are subject to purifying selection, while surface antigens of viruses and parasites are frequently subject to putative positive selection (Endo et al., 1996), as could be expected.

However, the *dN/dS* ratio is rarely higher than 1, even in proteins that are likely subject to adaptive evolution. The types of selection acting on different sites or regions of a protein often differ (Endo et al., 1996). This phenomenon is called mosaic selection. Some specific sites within proteins, such as ligand binding sites or catalytic residues in enzymes, are evolutionarily more conserved than others. As a result, the *dN/dS* ratios differ among individual sites, and averaging these ratios across the entire protein length obliterates the information they provide and is counterproductive. In the case of mosaic selection, when different selection types act on different residues, an elevated *dN/dS* (particularly above 1) indicates that there are residues that are subject to positive selection. The power of *dN/dS* analysis can be enhanced by the alignment of sequences of genes from many species (as in the branch-site test (PAML (Zhang et al., 2005)) (HyPhy (Kosakovsky Pond et al., 2011))). It is generally accepted that an absolute minimum requirement for such analyses is four or five sequences of interest (Yang, 2005).

Multiple alignments are particularly useful for the detection of selection acting on particular residues. The most popular method in this regard is the Bayes Empirical Bayes (BEB) inference of amino acid sites under positive selection (Yang et al., 2005). Notably, the statistical power of these tests cannot be established using real sequence data, as the selection acting on genes and different residues is not known. Therefore, this question has been investigated via the simulation of sequence evolution. An analysis of such simulated evolution with mosaic selection showed that this test was able to detect approximately 25% cases of positive selection in a comparison of five taxonomic groups and approximately 50% of cases of positive selection in a comparison of 30 taxonomic groups (Yang et al., 2005).

It is worth mentioning, that besides the *dN/dS* –based analyses two other approaches have been used to uncover non-neutral evolution.

One group of methods comprises neutrality tests using population genetic data. Examples include analyses of allelic frequency at individual loci (Ewens–Watterson test) (Slatkin, 1994) and of the frequency distribution of segregation sites at multiple loci (Tajima's D test) (Tajima, 1989).

Second approach is based on the genetic data of a population with sequences from different taxa. The McDonald–Kreitman (MK) test (McDonald and Kreitman, 1991) compares the ratio of nonsynonymous and synonymous mutations between and within species. The Hudson–Kreitman–Aguade test (Hudson et al., 1987) is similar and compares patterns of polymorphisms and mutations between species in several loci.

As with *dN/dS* approach statistical power of these tests cannot be determined using real sequence data, but based on simulated sequence evolution, the *dN/dS* method almost always shows a better statistical power to detect selection than alternative approaches based on population genetic data. Notably, the dN/dS approach is particularly strong in the case of mosaic selection(Zhai et al., 2009).

Different tests can detect selection affecting different mutations. Tests based on population genetics data detect selection acting on polymorphic mutations observed in extant populations. Tests based on the *dN/dS* ratio identify selection in all positions, including those in which mutations are lethal and cannot manifest as polymorphisms.

This difference has consequences in the case of rapid male evolution i.e. rapid evolution of genes with a male-biased expression (Parsch and Ellegren, 2013). It has been shown in model organism *D. melanogaster* that such genes have relatively high *dN/dS* ratio in comparison with other genes, albeit it is still usually well below 1 (Zhang et al., 2004). This latter observation indicates that the majority of mutations in these genes are deleterious. In contrast, tests based on population genetics indicate that the majority of polymorphic mutations in these genes are beneficial (Pröschel et al., 2006; Sawyer et al., 2007).

In this paper, the hypothesis that proteins that are subject to structural constraints on sequence evolution play a crucial role in the evolutionary arms race between pathogens and hosts was tested.

The study used the evolution of the human malaria parasite *P. falciparum* as a model. The genomes of malaria parasites are known. The first identified closely related sibling species of *P.falciparum* was *P. reichenowi*—a chimpanzee parasite. Phylogenetic analyses indicate that all extant *P. falciparum* populations originated from *P. reichenowi*, likely by single host transfer (Rich et al., 2009).

The recent progress in the taxonomic knowledge has revealed more such species: *P. praefalciparum*, *P. billcollinsi*, *P. gaboni* and *P. adleri* (Rayner et al., 2011).

The analysis of selection based on genomic data still is challenging in the case of the human malaria parasite, as methods based on protein alignment (such as BEB or side-branch models) are difficult to apply. They require correct gene phylogenetic trees, which in turn cannot be with high confidence because of ample recent gene flow between sibling species (Otto et al., 2018). As a consequence gene trees are often different than species trees. Indeed, gene flow indicates recombination events, and detection of positive selection in the presence of recombination events is highly challenging and requires to be analyzed numerous species (Anisimova et al., 2003) (Wilson and McVean, 2006). Initially, I tried to detect selection using PAML, but frequently (in about 20% of cases), values of likelihood ratio tests were unrealistically high (not shown).

On the other hand, due to the problems caused by codon saturation more distantly related species cannot be used in such analysis.

The genomes of *Plasmodium* parasites, such as *P. falciparum* usually have a very high AT content (Gardner et al., 2002; Videvall, 2018; Weber, 1987), which causes the codon saturation (as synonymous positions are usually occupied only A or T only). This means it is highly likely that, when comparing more distant species, synonymous mutations will occur more than once in one site (eg. A to T and then T to A) resulting in lack of any mutation and thereby making the *dN/dS* approach unreliable.

This limitation may be circumvented by studying closely related species, in which the occurrence of such back-mutations is less likely due to the short time since their evolutionary separation.

Sites were classified according to the tertiary structure into buried and nonburied classes, which were then analyzed separately for signs of adaptive evolution. Several studies revealed that buried protein residues evolve very slowly and are subject to purifying selection(Huang, 2021; Yang and Swanson, 2002; Zhou et al., 2008).

I took advantage of the AlphaFold method shown earlier to predict protein structures with high accuracy (Jumper et al., 2021) and the database containing whole-proteome prediction for the human malaria parasite *P. falciparum*. Unfortunately, the AlphaFold method usually provides good models only for some parts of proteins. Therefore, this method provides information only on the structural constraints on the evolution of buried sites.

Since predicting the ancestral state is difficult in the presence of gene flow, a comparison of extant gene sequences with ancestral sequences was not feasible. Therefore, I aligned only extant sequences and compared these sequences between *P. falciparum* and *P. reichenowi* and between *P. falciparum* and *P. adleri*.

2. Materials and methods

The study applies the method introduced by Yang and Swanson. The sites experiencing mutation were grouped into different classes, using AlphaFold to detect positive selection. In *dN/dS* analyses the sites undergoing mutations are often grouped into different classes. For example, surface residues of proteins evolve more rapidly than buried ones and antigenic sequences tend to change more rapidly.

Thus, adaptive evolution can be detected more easily in a more

rapidly evolving class (Yang and Swanson, 2002).

As mentioned in the introduction, I applied the method of Yang and Swanson. I grouped the sites experiencing mutation into different classes using AlphaFold to detect positive selection.

Protein structures were obtained from the AlphaFold protein structure database (https://www.alphafold.ebi.ac.uk/) (Jumper et al., 2021). Sites were classified according to their buried or exposed status using relative solvent accessibility (RSA), as described previously (Serafimova et al., 2020).

Confidence in the protein structure was predicted using a per-residue confidence score (referred to as pLDDT in AlphaFold) produced by AlphaFold. Confidence in the structure is high when pLDDT>70. Some regions with a score below 50 pLDDT may be unstructured in isolation. A protein will usually contain fragments with both high and low confidence (see Fig. 1).

The calculation of RSA is important because amino acids differ in their surface area. To account for these differences, RSA is calculated by normalizing the solvent-accessible surface area of a residue in a structure of interest based on its accessible surface area in a reference state (e.g., residue X in an extended tripeptide, such as Gly-X-Gly). Although nonnormalized surface areas can also be used to classify residues as buried or exposed, normalization reduces the bias toward classifying smaller residues as more buried and larger ones as more exposed. A residue was considered buried when its RSA < 0.2 and pLDDT>70.

In my analysis, I grouped sites into the following two classes: buried and nonburied. Additionally, I detected positive selection using the second classification at sites that were probably incorrectly modeled by AlphaFold (pLDDT<50) and other sites ('confident' and 'non-confident' classes in the 'non-buried' class).

The *dN/dS* ratios were calculated separately for different classes of residues in entire genes.

It is worth mentioning that the software also produces *dN* and *dS* likelihood estimates based on the more complex model of Goldman and Yang (Goldman and Yang, 1994). This algorithm is not very useful in the case of the analyzed genomes, as it underestimates the dS/dN ratio in AT-rich genomes and, thus, is not able to detect positive selection in the malaria genome (Yap et al., 2010).

Orthology tables based on OrthoMCL (Chen et al., 2006) were downloaded from PlasmoDB (Aurrecoechea et al., 2009). The *dN/dS* ratios were calculated separately for different classes of residues in entire genes using ClustalW (Thompson et al., 2002), PAML (Yang and Bielawski, 2000) and the author's own pipeline written in UNIX C shell (see SOM 2). ClustalW was used for the pairwise alignment of *Plasmodium falciparum* and *P. reichenowi* (or *P. adleri*) sequences. It was mathematically proven that pairwise global alignment was optimal (Altschul et al., 1990; Needleman and Wunsch, 1970).

I estimated *dN* and *dS* using the method of Nei and Gojobori (Nei and Gojobori, 1986), calculated by PAML.

I also applied the Fisher one-sided test to test the results obtained via the Nei and Gojobori method. (Zhang et al., 1997). Here, the numbers of synonymous and nonsynonymous changes and positions were estimated using PAML. PAML provides estimates of both *dS/dN* and the numbers of synonymous (S)/nonsynonymous (N) sites.

The numbers of synonymous and nonsynonymous changes can easily be estimated by multiplying their rates by the number of sites of a given type: *dN**N or *dS**S. The resultant values were rounded to obtain natural numbers.

3. Results

The main aim of this paper was to detect *P. falciparum* genes that are subject to structural constraints on sequence evolution and in which only a fraction of sites are subject to positive selection. I hypothesize that such proteins play important roles in the parasite-host arms race.

As previously mentioned, based on protein structures and the per-residue confidence score (pLDDT), the evaluated protein sites were

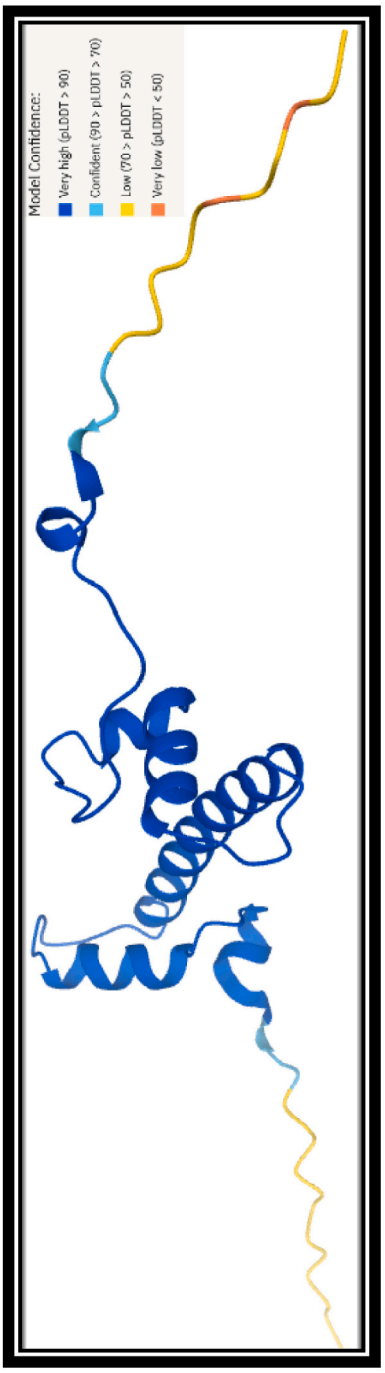


Fig. 1. An example of an AlphaFold model of a *P. falciparum* protein (model of histone h2A).

divided into the following two classes: buried and nonburied (actually, 'confidently' buried sites and other sites).

According to this classification, 18% of sites were 'confidently buried'. Additionally, based on pLDDT, sites were classified as 'confident' or 'nonconfident'.

According to the second classification, 45% of sites were members of the confident class, and the other sites belonged to the 'nonconfident' class.

As expected, buried sites were observed to evolve significantly more slowly than other sites. Additionally, the poor quality of the protein model is a good predictor of rapid evolution. It should be noted that the differences in the speed of neutral evolution among different classes, estimated using the *dS* parameter, were very small.

See Fig. 2.

In this paper, I focus on detecting signs of positive selection among these genes, and the analysis of complete sequences identified purifying selection ($dN < dS$). In the comparison of whole *P. falciparum* vs. *P. reichenowi* genes, 105 genes were indicated to be subject to positive selection ($dN > dS$) and 4382 to purifying selection.The results are presented in Table 1.

However, when *dN* was calculated only for the 'nonburied' class, the *dN* was higher than the *dS* calculated using whole genes in the case of an additional 26 protein-coding genes (see Table 1) and was higher than the *dS* calculated using only 'nonburied' sites in the case of an additional 63 protein-coding genes. In a set of 15 additional genes, the *dN* of the nonburied class was higher than the *dS* estimated using both methods.

In the above-described set of 63 protein-coding genes in which the *dN* calculated for the buried class was higher than the *dS* estimated for the buried class, the dS estimated for the exclusively buried class was smaller than the *dS* estimated for the whole gene in the majority of cases (58). This observation suggests that it is likely that *dS* is underestimated in these 63 genes. In the signal results indicating positive selection, $dN > dS$ was observed due to the underestimation of *dS*. Therefore, I concluded that the detection of positive selection using *dS* estimated based on whole genes provides more accurate results.

I tested the statistical significance of the obtained results using the one-sided Fisher exact test. None of the 26 cases of putative selection acting on "nonburied" sites were found to be statistically significant ($dN_{Buried} > dS$)(see Table 1). As explained in the introduction however, if $dN/dS = 1$ (or is even slightly <1), since there are amino acid restrictions, it can be concluded with a high probability that there is positive selection on some protein sites.

Therefore, I checked the assumptions of this reasoning.

I used another site classification approach to detect signals of putative positive selection. Sites were divided into 'confident' and 'not-confident' sites. The 'nonconfident' sites are actually a subclass of the 'nonburied' class, as I included 'nonconfident' sites in this class. The signals of positive selection were statistically significant for three genes according to the Fisher test. Moreover, the *P* values were very low (in the range of 0.00059—10^{-8}), indicating that it is unlikely that the signal of positive selection is false. This observation confirms that some sites are subject to positive selection.

I also checked the expectation that the evolution of the 26 proteins would be subject to selection constraints on exposed residues. For this purpose, I applied *dN* and *dS* analysis to compare these 26 genes of *P. falciparum* with syntenic orthologs of the remotely related *Plasmodium* species *P. knowlesi*. Fourteen of these genes had appropriate orthologs. The analysis of *dS* indicated that during the evolution of each synonymous position, many mutations have appeared. In most cases (ten cases in a set of fourteen genes); the rate was too high to be estimated, and the program showed an error. The smallest *dS* was 1.8. The nonsynonymous substitution rates at nonburied positions ranged from 0.19 to 1.14,

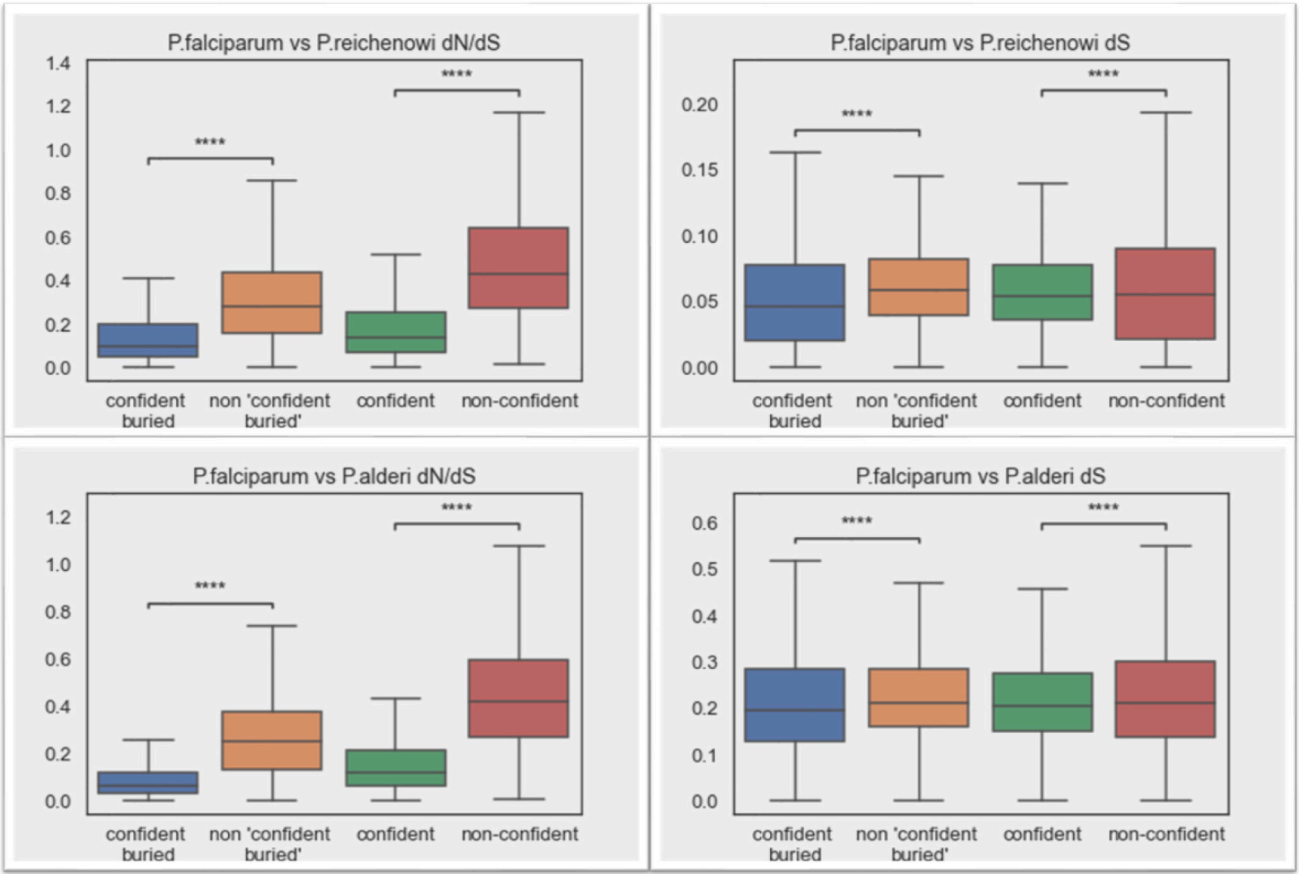


Fig. 2. Orthologous genes were compared in a pairwise manner between *P. falciparum* and *P. reichenowi* (top) and between *P. falciparum* and *P. adleri* (bottom).

Table 1
Genes under putative positive selection according to analysis of *P. falciparum* vs. *P. reichenowi* alignments.

	Genes under positive selection and structural constraints dN < dS but $dN_{non-buried}$ > dS for nonburied residues	Genes under positive selection dN > dS
Total number of genes	**26**	**105**
Putative antigenes	**11**	**40**
Promising vaccine candidates	6 (merozoite surface protein 1, erythrocyte binding antigen-140, RAP,P47, ring-infected erythrocyte surface antigen, ring-exported protein 3)	13 (MSP-2 MSP-4, MSP-7 like, merozoites-associated armadillo repeats protein, erythrocyte binding antigens-181 and 175,sporozoite threonine and asparagine-rich protein, SAS-6, gametocyte exported protein 2, sporozoite invasion-associated protein 2, thrombospondin-related anonymous protein, EMP1-trafficking protein, exported protein 1)
Clonally variant genes	5 (Hyp *1, PHIST *2, Surfin *1 Pfmc-2TM *1)	16 (Hyp *7, PHIST *6, Pfmc-2TM *1,DnaJ *1, early transcribed membrane protein *1)
Exported proteins unknown function	0	11
Other proteins than putative antigens	**15**	**65**
Translation and DNA/RNA metabolism	0	3
Enzymes and mitochondrial carriers	10 (protein phosphatase PPM8, FIKK9.3 kinase, lysophospholipase LPL20, haloacid dehalogenase-like hydrolase, putative esterases (2), 4-diphosphocytidyl-2-*C*-methyl-D-erythritol kinase, S-adenosylmethionine, phosphatidylinositol *N*-acetylglucosaminyltransferase, mitochondrial carrier protein 6)	3 (shikimate dehydrogenase, 1-acyl-sn-glycerol-3-phosphate acyltransferase, ATP synthase subunit O, bifunctional methylenetetrahydrofolate dehydrogenase/cyclohydrolase,12)
Calodulin	1	0
Tetratricopeptide repeat protein	1	0
Not known function	3	45
Other genes not mentioned above	0	14
Reanalysis of dN/dS results presented above. Detection of statistically significant cases of positive selection was based on the comparison of nonsynonymous substitution rates at sites belonging to the 'nonconfident' class $dN_{non-confident}$ > dS ($P < 0.05$)		
Total number of genes	**3**	**7**
promising vaccine candidate	1 (merozoite surface protein)	2 (erythrocyte-binding antigen 181, TRAP)

suggesting that some nonsynonymous 'nonburied' sites have remained unchanged during the evolution of these remotely related *Plasmodium* species, while many mutations have occurred at the majority of synonymous sites. This observation confirms that it is likely that selection constraints acting on 'nonburied sites' have removed some nonsynonymous mutations changing the residues at these sites. Additionally, I confirmed that buried positions were subject to structural constraints acting on sequence evolution. The nonsynonymous rates at buried positions were much lower (0.05–0.42) than those at exposed positions, suggesting that these sites are subject to selection constraints.

In summary, the above method identified a set of 26 proteins subject to putative positive selection acting on the protein surface and purifying selection acting on buried sites. Fig. 3 shows two examples of models of such proteins. Accelerated evolution occurs across the whole protein surface, as mutations appear on the surfaces of different parts of a protein.

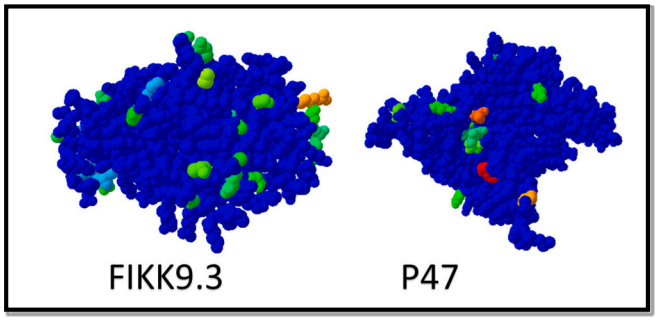


Fig. 3. Models of proteins encoded by genes that are subject to putative positive selection acting on 'nonburied' residues according to the alignment of *P. falciparum* vs. *P. reichenowi* ($dN_{buried} > dS$) and for which there is a signal of purifying selection. Mutations are indicated by different colors. Very long loops with low confidence scores are not shown.

I used this set of 26 genes to test the hypothesis that the evolution of structurally constrained proteins plays a crucial role in a host-parasite evolutionary arms race. According to this hypothesis, structurally constrained proteins are promising vaccine candidates.

I compared proteins subject to putative positive selection to test this hypothesis: 1. structurally constrained proteins were compared with 2. other proteins whose speed of evolution is not significantly constrained by the protein structure (positive selection indicated by the conventional *dN/dS* test).

Similar fractions of the genes in the two sets encoded putative antigens. However, the group of proteins whose speed of evolution was not structurally constrained contained 11 exported proteins with unknown functions. The predictions of the antigenic properties of these proteins are very uncertain, as they are based only on the fact that they are exported.

A significant fraction of antigens from both sets consisted of clonally variant genes (and putative clonally variant genes) belonging to multigene families such as the *pfmc-2tm, phist*, hyp, and surfin families (Guizetti and Scherf, 2013; Rovira-Graells et al., 2012). "Clonally variant gene expression" refers to the occurrence of genes in different states (active or silent) in different individual parasites with identical genomes in the same stage of life cycle progression. The parasite can thereby evade host immunity by switching the expression of clonally variant genes from genes recognized by the host immune system to genes that are not yet recognized. This property means that such genes are not promising targets for vaccination. The assumption that multigene families are not good vaccine candidates was applied in a very recent study describing discovery of novel vaccine antigens against *Trypanosoma vivax* (Autheman et al., 2021).

Therefore, I hypothesized that antigens whose function is described in the PlasmoDB database and are not encoded by clonally variant genes may be promising vaccine candidates. The currently available evidence supports this assumption. Some of these proteins are already accepted vaccine candidates and are being subjected to different types of studies,

including clinical tests, as reported for merozoite surface protein-1 (MSP1) [73, 74], MSP2 (Tuju et al., 2017), and erythrocyte-binding antigens 140 [75], 175 (Koram et al., 2016), RAP-1 (Tuju et al., 2017), and P47 (Molina-Cruz et al., 2017; van Dijk et al., 2010). In fact, a higher fraction of promising vaccine candidates were identified in my analysis of the set of genes whose evolution is structurally constrained and subject to positive selection on nonconstrained sites (although this difference was not statistically significant).

Among the identified protein-coding genes subject to putative positive selection, the set whose evolution is structurally constrained contained substantially more enzymes. This result indicates that enzymes are subject to structural constraints and that positive selection acts only on "nonburied" residues. As a result, it is difficult to detect positive selection acting on the whole sequence.

Relevant published experimental studies suggest that the prediction of enzymes under positive selection performed in this study was correct. The obtained results revealed molecular mechanisms that may lead to the accelerated evolution of FIKK9.3 kinase and dehalogenase-like hydrolase. FIKK is a member of the FIKK kinase family. The parasite exports FIKK kinases, which phosphorylate host erythrocyte proteins (for example adducin S726 phosphorylation) (Davies et al., 2020). Therefore, it is likely that mutations altering the substrates of FIKK kinases are frequently beneficial for the host. (particularly if FIKK kinases cannot phosphorylate modified substrates). It is also likely that the evolution of FIKK kinases may change their activity in response to changes in their targets. Furthermore, it has been shown that dehalogenase-like hydrolases participate in the metabolic plasticity of parasites. Mutations in this enzyme confer drug resistance (Frasse and Odom John, 2019). Therefore, the experimental results presented above indicate that it is likely that my predictions of enzymes under positive selection are correct.

I repeated the whole analysis comparing *P. falciparum* and *P. alderi*. The results are presented in Table 2. The number of identified genes with signals of positive selection was smaller than that in the comparison of *P. falciparum* vs. *P. reichenowi*. There were only 40 genes with dN > dS among 4244 genes aligned with *P. falciparum*. Methodology based on the detection of positive selection acting only on nonburied sites again identified antigens more frequently. However, this method did not identify enzymes in this case, and only 6 cases of positive selection were identified.

Table 2
Genes under putative positive selection according to analysis of *P. falciparum* vs. *P. alderi* alignments.

	Genes under positive selection and structural constraints dN < dS but $dN_{non\text{-}buried}$ > dS for nonburied residues	Genes under positive selection dN > dS
Total number of genes	**6**	**40**
Putative antigenes	5	16
Promising vaccine candidates	3 (2 genes encoding membrane associated erythrocyte binding-like protein, Reticulocyte binding protein homolog 5)	5 (MSP-2,MSP4, MSP7, MSP-11 S-antigen)
Clonally variant genes	1 (Hyp)	3 (Hyp, PHIST, Parasite-infected erythrocyte surface protein)
Exported proteins unknown function	1	8
Other proteins than putative antigens	**1**	**24**
Translation and DNA/RNA metabolism	0	3
Enzymes and mitochondrial carriers	0	1
Not known function	1	11
Other genes not mentioned above	0	9
Reanalysis of dN/dS results presented above. Detection of statistically significant cases of positive selection was based on the comparison of nonsynonymous substitution rates at sites belonging to the 'nonconfident' class$dN_{non\text{-}confident}$ > dS ($P < 0.05$)		
Total number of genes	**2**	**7**
promising vaccine candidate	1 (membrane associated erythrocyte binding-like protein)	4 (MSP-2, MSP-4, MSP-7, S-antigen)

4. Discussion

This paper tests the hypothesis that proteins that are subject to structural constraints on sequence evolution play a crucial role in the evolutionary arms race between pathogens and hosts.

I tested this hypothesis using *P. falciparum* as a model. Malaria is one of the world's major medical problems, and despite of years of medical trials, an effective malaria vaccine is still lacking (Kaslow, 2020; Maxmen, 2021). However, the evolutionary arms race in which this organism is engaged is relatively well studied in the context of vaccination.

Early studies showed promising results, and an effective malaria vaccine was believed to be just around the corner (Miller et al., 1986). However, immune escape has been one of the main challenges to producing an effective vaccine (Gomes et al., 2016; Good et al., 1988; Neafsey et al., 2015). Modern attempts aimed at malaria vaccine development began with a 1960s study of mice immunized with irradiated sporozoites (Nussenzweig et al., 1967). Subsequent now-classical studies described different malarial antigenic proteins that could be potential targets for vaccination (Bull et al., 1998). Both host-protecting vaccines (Kaslow et al., 1988; Miller et al., 1986) and transmission-blocking vaccines (Carter, 2001), which induce immunity against the stages of parasites that infect mosquitos appeared to be effective. However, progress in the field has been very slow. Despite decades of efforts, only one vaccine candidate, the pre-erythrocytic circumsporozoite (CSP)-based RTS,S/AS01E, has advanced through licensure and pilot implementation (Kaslow, 2020; Maxmen, 2021).

The main difficulty in producing an effective vaccine is immune escape (i.e., selection of mutations in antigens that are not recognized by the host immune system) (Good et al., 1988; Krzyczmonik et al., 2012).

Another mechanism of immune evasion is host immune suppression. *Plasmodium* parasites have been shown to modulate the maturation of dendritic cells, which subsequently suppresses the activity of T-cells (Urban et al., 1999; Wykes and Good, 2008). Malaria infection has also been shown to induce regulatory T-cells and suppress experimental autoimmune encephalomyelitis(Farias et al., 2011). The plasmodial factors involved in host immune suppression are not well described. However, homology searches indicate that there are two protein homologues of animal proteins that may be involved in this process: a Macrophage Migration Inhibitory Factor (MIF) (Miller et al., 2012) and a T-cell Immunomodulatory Protein of *Plasmodium* (TIP) (Fiscella et al., 2003; Kaczanowski and Zielenkiewicz, 2003).

As it was already mentioned, another important mechanism of host immune evasion is the mono-allelic expression of multi-gene families by malaria parasites. The best-described example is the group of *var* genes, which encode PfEMP1 surface proteins. There are approximately 60 *var* genes encoded by the *P. falciparum* parasite (Guizetti and Scherf, 2013). These proteins are highly antigenic, and each individual of *P. falciparum* expresses only a single type of this gene. Other copies are transcriptionally silent. As the antibody response against a given type of PfEMP1 develops, a subgroup of the parasites switches the expression to an alternative form of PfEMP1 and re-establishes the infection.

To sum up, the good vaccine should be evolution prove and should induce long lasting memory prove to parasitic immunosuppression mechanisms.

As it was already mentioned, evolution of the human malaria parasites is relatively well studied.

The analysis of the action of selection on genes of these parasites has a long history. After the *P. falciparum* genome was published (Gardner et al., 2002), Plotkin developed and applied a method for detecting positive selection based on a single nucleotide sequence (the volatility method)(Plotkin et al., 2004). However, the strong codon bias in *P. falciparum* genomes makes this ineffective (Hahn et al., 2005).

The first identified closely related sibling species of *P.falciparum* was *P. reichenowi* a chimpanzee parasite. Phylogenetic analyses indicate that all extant *P. falciparum* populations originated from *P. reichenowi*, likely by single host transfer (Rich et al., 2009).When compared with *P. reichenowi*, almost all *P. falciparum* genes had *dN/dS* below 1 in (Jeffares et al., 2007), indicating that most residues in proteins are subject to putative purifying selection. However, some genes had a significantly higher *dN/dS* ratio. Authors hypothesized that putative positive selection acts on part of the residues of such genes. A similar conclusion was drawn from an earlier analysis of the rodent malaria parasites: *P. yoelii* and *P. berghei* (Carlton et al., 2002). A whole-genome scan of *P. falciparum* based on the MK test revealed genes in which polymorphic mutations are subject of putative positive or purifying selection. For example, indications of putative positive selection were found for the MSP-8 gene (Jeffares et al., 2007). Analysis of whole-genome sequences of different sibling species revealed that 136 genes are subject to putative episodic positive selection in *P. falciparum* according to the branch-site test (Otto et al., 2018).

Earlier we used the BEB method to study the evolution of three rodent species *P. berghei*, *P. yoelii* and *P. chabaudi*, and identified residues in the p47 and p230 proteins, subject to adaptive evolution (van Dijk et al., 2010). Using the MK test we also showed rapid male evolution phenomenon in *Plasmodium* species: polymorphisms in genes with male-biased expression (in contrast to genes with female-biased expression), showed signs of positive selection. The male-biased genes in both rodent and human Plasmodium parasites had elevated *dN/dS* ratios in comparison with the female-biased genes Khan et al., 2013).

Notably we showed that the antigenic proteins of *P. falciparum* were likely to undergo positive selection, as they showed elevated levels of non-synonymous to synonymous polymorphisms (Krzyczmonik et al., 2012; Khan et al., 2013).

This paper presents a novel approach to the study of evolution of parasitic protein. It identifies constrained protein sites (confidently predicted buried sides) to answer this question, using AlphaFold models. Based on this information, the cases of positive selection acting only on nonconstrained sites were identified.

The above results show that the developed methodology correctly identifies structural constraints and detects cases of positive selection that are not detected by the conventional dN/dS test considering the entire sequence.

The identified cases of positive selection are promising vaccine candidates.

Therefore, this method could be useful in the process of deriving vaccines from the starting point of genomic sequence data named "reverse vaccinology," which involves the following steps: bioinformatic software to screen genomes for surface-expressed proteins; high-throughput expression of these proteins and in vitro confirmation of their surface location; animal-based immunogenicity testing; and finally, conventional human vaccine trials (see as a review (Kelly and Rappuoli, 2005; Masignani et al., 2002). Using bioinformatic methods, for instance, one can predict the protein epitopes recognized by the immune system, secreted proteins, and proteins exposed on the surface of the pathogen (see as a review (Dalsass et al., 2019)).

Additionally, the presented methodology detects enzymes under putative positive selection that are not detected by the conventional *dN/dS* test. It is well established that enzymes are subject to structural constraints on sequence evolution. There are two different nonexclusive potential explanations for the rapid evolution of parasite enzymes. The first is the pressure imposed by the host immune system targeting them. Such enzymes are promising vaccine candidates. Parasites require the activity of these enzymes for life, and they are subject to structural constraints. Therefore, they cannot readily undergo changes to achieve immune escape (i.e., modification in a form that is not recognized by the host immune system). The second explanation for the rapid evolution of enzymes is the rapid evolution of their enzymatic activity, which is likely the case for FIKK9.3 kinase and dehalogenase-like hydrolase.

It is worth mentioning that I identified signals of positive selection acting on PPM8 protein phosphatase. This represents a case of guilt by association with FIKK9.3 kinase, as kinases show the opposite activity to phosphatases.

5. Conclusions

In summary, this paper shows that proteins whose evolution is subject to structural constraints play fundamental roles in the host-parasite arms race. The presented methodology based on the analysis of AlphaFold models identifies such proteins, which are promising vaccine candidates. Progress in protein modeling may improve the ability of this method to predict promising vaccines, as the AlphaFold method has its own limitations and cannot predict the structures of many protein fragments.

Funding

This work was supported by grant 2017/27/B/NZ8/02502 from the Polish National Science Centre.

Credit author statement

Szymon Kaczanowski wrote the paper, developed the presented methodology, and made a bioinformatic analysis.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

I have shared links to my data: 1. Dataset 1. https://osf.io/vyh4s/ and 2. Dataset 2. https://osf.io/7m43p/

Acknowledgments

Many useful comments were provided by my colleague Piotr Zielenkiewicz.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2022.105397.

References

Altschul, S., Gish, W., Miller, W., Myers, E., Lipman, D., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Anisimova, M., Nielsen, R., Yang, Z., 2003. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. Genetics 164, 1229–1236. https://doi.org/10.1093/genetics/164.3.1229.

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, C., Wang, H., 2009. PlasmoDB: a functional genomic database for malaria parasites. Nucleic Acids Res. 37, D539–D543. https://doi.org/10.1093/nar/gkn814.

Autheman, D., Crosnier, C., Clare, S., Goulding, D.A., Brandt, C., Harcourt, K., Tolley, C., Galaway, F., Khushu, M., Ong, H., Romero-Ramirez, A., Duffy, C.W., Jackson, A.P., Wright, G.J., 2021. An invariant Trypanosoma vivax vaccine antigen induces

protective immunity. Nature 595, 96–100. https://doi.org/10.1038/s41586-021-03597-x.
Bull, P.C., Lowe, B.S., Kortok, M., Molyneux, C.S., Newbold, C.I., Marsh, K., 1998. Parasite antigens on the infected red cell surface are targets for naturally acquired immunity to malaria. Nat. Med. 4, 358–360.
Carlton, J.M., Angiuoli, S.V., Suh, B.B., Kooij, T.W., Pertea, M., Silva, J.C., Ermolaeva, M. D., Allen, J.E., Selengut, J.D., Koo, H.L., Peterson, J.D., Pop, M., Kosack, D.S., Shumway, M.F., Bidwell, S.L., Shallom, S.J., van Aken, S.E., Riedmuller, S.B., Feldblyum, T.V., Cho, J.K., Quackenbush, J., Sedegah, M., Shoaibi, A., Cummings, L. M., Florens, L., Yates, J.R., Raine, J.D., Sinden, R.E., Harris, M.A., Cunningham, D.A., Preiser, P.R., Bergman, L.W., Vaidya, A.B., van Lin, L.H., Janse, C.J., Waters, A.P., Smith, H.O., White, O.R., Salzberg, S.L., Venter, J.C., Fraser, C.M., Hoffman, S.L., Gardner, M.J., Carucci, D.J., 2002. Genome sequence and comparative analysis of the model rodent malaria parasite plasmodium yoelii yoelii. Nature 419, 512–519. https://doi.org/10.1038/nature01099.
Carter, R., 2001. Transmission blocking malaria vaccines. Vaccine 19, 2309–2314.
Chen, F., Mackey, A.J., Stoeckert, C.J., Roos, D.S., 2006. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. Nucleic Acids Res. 34, D363–D368. https://doi.org/10.1093/nar/gkj123.
Dalsass, M., Brozzi, A., Medini, D., Rappuoli, R., 2019. Comparison of Open-Source Reverse Vaccinology Programs for Bacterial Vaccine Antigen Discovery. Front. Immunol. 10, 113. https://doi.org/10.3389/fimmu.2019.00113.
Davies, H., Belda, H., Broncel, M., Ye, X., Bisson, C., Introini, V., Dorin-Semblat, D., Semblat, J.P., Tibúrcio, M., Gamain, B., Kaforou, M., Treeck, M., 2020. An exported kinase family mediates species-specific erythrocyte remodelling and virulence in human malaria. Nat. Microbiol. 5, 848–863. https://doi.org/10.1038/s41564-020-0702-4.
Endo, T., Ikeo, K., Gojobori, T., 1996. Large-scale search for genes on which positive selection may operate. Mol. Biol. Evol. 13, 685–690. https://doi.org/10.1093/oxfordjournals.molbev.a025629.
Farias, A.S., Talaisys, R.L., Blanco, Y.C., Lopes, S.C., Longhini, A.L., Pradella, F., Santos, L.M., Costa, F.T., 2011. Regulatory T cell induction during plasmodium chabaudi infection modifies the clinical course of experimental autoimmune encephalomyelitis. PLoS One 6, e17849. https://doi.org/10.1371/journal.pone.0017849.
Fiscella, M., Perry, J.W., Teng, B., Bloom, M., Zhang, C., Leung, K., Pukac, L., Florence, K., Concepcion, A., Liu, B., Meng, Y., Chen, C., Elgin, E.C., Kanakaraj, P., Kaufmann, T.E., Porter, J., Cibotti, R., Mei, Y., Zhou, J., Chen, G., Roschke, V., Komatsoulis, G., Mansfield, B., Ruben, S., Sanyal, I., Migone, T.S., 2003. TIP, a T-cell factor identified using high-throughput screening increases survival in a graft-versus-host disease model. Nat. Biotechnol. 21, 302–307. https://doi.org/10.1038/nbt797.
Frasse, P.M., Odom John, A.R., 2019. Haloacid dehalogenase proteins: novel mediators of metabolic plasticity in. Microbiol. Insights 12. https://doi.org/10.1177/1178636119848435, 1178636119848435.
Gardner, M., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R., Carlton, J., Pain, A., Nelson, K., Bowman, S., Paulsen, I., James, K., Eisen, J., Rutherford, K., Salzberg, S., Craig, A., Kyes, S., Chan, M., Nene, V., Shallom, S., Suh, B., Peterson, J., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., Haft, D., Mather, M., Vaidya, A., Martin, D., Fairlamb, A., Fraunholz, M., Roos, D., Ralph, S., McFadden, G., Cummings, L., Subramanian, G., Mungall, C., Venter, J., Carucci, D., Hoffman, S., Newbold, C., Davis, R., Fraser, C., Barrell, B., 2002. Genome sequence of the human malaria parasite plasmodium falciparum. Nature 419, 498–511.
Goldman, N., Yang, Z., 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol. Biol. Evol. 11, 725–736. https://doi.org/10.1093/oxfordjournals.molbev.a040153.
Gomes, P.S., Bhardwaj, J., Rivera-Correa, J., Freire-De-Lima, C.G., Morrot, A., 2016. Immune escape strategies of malaria parasites. Front. Microbiol. 7, 1617. https://doi.org/10.3389/fmicb.2016.01617.
Good, M.F., Miller, L.H., Kumar, S., Quakyi, I.A., Keister, D., Adams, J.H., Moss, B., Berzofsky, J.A., Carter, R., 1988. Limited immunological recognition of critical malaria vaccine candidate antigens. Science 242, 574–577.
Guizetti, J., Scherf, A., 2013. Silence, activate, poise and switch! Mechanisms of antigenic variation in plasmodium falciparum. Cell. Microbiol. 15, 718–726. https://doi.org/10.1111/cmi.12115.
Hahn, M.W., Mezey, J.G., Begun, D.J., Gillespie, J.H., Kern, A.D., Langley, C.H., Moyle, L.C., 2005. Evolutionary genomics: codon bias and selection on single genomes. Nature 433, E5–E6 discussion E7-8. https://doi.org/10.1038/nature03221.
Huang, Y.F., 2021. Dissecting genomic determinants of positive selection with an evolution-guided regression model. Mol. Biol. Evol. https://doi.org/10.1093/molbev/msab291.
Hudson, R.R., Kreitman, M., Aguadé, M., 1987. A test of neutral molecular evolution based on nucleotide data. Genetics 116, 153–159.
Jeffares, D., Pain, A., Berry, A., Cox, A., Stalker, J., Ingle, C., Thomas, A., Quail, M., Siebenthall, K., Uhlemann, A., Kyes, S., Krishna, S., Newbold, C., Dermitzakis, E., Berriman, M., 2007. Genome variation and evolution of the malaria parasite plasmodium falciparum. Nat. Genet. 39, 120–125.
Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D., 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589. https://doi.org/10.1038/s41586-021-03819-2.
Kaczanowski, S., Zielenkiewicz, P., 2003. A TIP on malaria (genomics). Nat. Biotechnol. 21, 733. https://doi.org/10.1038/nbt0703-733.
Kaslow, D.C., 2020. Malaria vaccine research & innovation: the intersection of IA2030 and zero malaria. NPJ Vaccines 5, 109. https://doi.org/10.1038/s41541-020-00259-3.
Kaslow, D.C., Quakyi, I.A., Syin, C., Raum, M.G., Keister, D.B., Coligan, J.E., McCutchan, T.F., Miller, L.H., 1988. A vaccine candidate from the sexual stage of human malaria that contains EGF-like domains. Nature 333, 74–76. https://doi.org/10.1038/333074a0.
Kelly, D.F., Rappuoli, R., 2005. Reverse vaccinology and vaccines for serogroup B Neisseria meningitidis. Adv. Exp. Med. Biol. 568, 217–223. https://doi.org/10.1007/0-387-25342-4_15.
Khan, Shahid M., Andrew, S.E.R., Waters, P., Janse, Chris J., Kaczanowski, Szymon, 2013. Why are male malaria parasites in such a rush?: sex-specific evolution and host–parasite interactions. Evol. Med. Public Health 2013 (3), 13.
Koram, K.A., Adu, B., Ocran, J., Karikari, Y.S., Adu-Amankwah, S., Ntiri, M., Abuaku, B., Dodoo, D., Gyan, B., Kronmann, K.C., Nkrumah, F., 2016. Safety and immunogenicity of EBA-175 RII-NG malaria vaccine administered intramuscularly in semi-immune adults: a phase 1, double-blinded placebo controlled dosage escalation study. PLoS One 11, e0163066. https://doi.org/10.1371/journal.pone.0163066.
Kosakovsky Pond, S.L., Murrell, B., Fourment, M., Frost, S.D., Delport, W., Scheffler, K., 2011. A random effects branch-site model for detecting episodic diversifying selection. Mol. Biol. Evol. 28, 3033–3043. https://doi.org/10.1093/molbev/msr125.
Krzyczmonik, K., Świtnicki, M., Kaczanowski, S., 2012. Analysis of immunogenicity of different protein groups from malaria parasite plasmodium falciparum. Infect. Genet. Evol. 12, 1911–1916. https://doi.org/10.1016/j.meegid.2012.07.023.
Masignani, V., Rappuoli, R., Pizza, M., 2002. Reverse vaccinology: a genome-based approach for vaccine development. Expert. Opin. Biol. Ther. 2, 895–905. https://doi.org/10.1517/14712598.2.8.895.
Maxmen, A., 2021. Scientists hail historic malaria vaccine approval - but point to challenges ahead. Nature. https://doi.org/10.1038/d41586-021-02755-5.
McDonald, J., Kreitman, M., 1991. Adaptive protein evolution at the Adh locus in Drosophila. Nature 351, 652–654.
Miller, L.H., Howard, R.J., Carter, R., Good, M.F., Nussenzweig, V., Nussenzweig, R.S., 1986. Research toward malaria vaccines. Science 234, 1349–1356.
Miller, J.L., Harupa, A., Kappe, S.H., Mikolajczak, S.A., 2012. Plasmodium yoelii macrophage migration inhibitory factor is necessary for efficient liver-stage development. Infect. Immun. 80, 1399–1407. https://doi.org/10.1128/IAI.05861-11.
Molina-Cruz, A., Canepa, G.E., Barillas-Mury, C., 2017. Plasmodium P47: a key gene for malaria transmission by mosquito vectors. Curr. Opin. Microbiol. 40, 168–174. https://doi.org/10.1016/j.mib.2017.11.029.
Neafsey, D.E., Juraska, M., Bedford, T., Benkeser, D., Valim, C., Griggs, A., Lievens, M., Abdulla, S., Adjei, S., Agbenyega, T., Agnandji, S.T., Aide, P., Anderson, S., Ansong, D., Aponte, J.J., Asante, K.P., Bejon, P., Birkett, A.J., Bruls, M., Connolly, K. M., D'Alessandro, U., Dobaño, C., Gesase, S., Greenwood, B., Grimsby, J., Tinto, H., Hamel, M.J., Hoffman, I., Kamthunzi, P., Kariuki, S., Kremsner, P.G., Leach, A., Lell, B., Lennon, N.J., Lusingu, J., Marsh, K., Martinson, F., Molel, J.T., Moss, E.L., Njuguna, P., Ockenhouse, C.F., Ogutu, B.R., Otieno, W., Otieno, L., Otieno, K., Owusu-Agyei, S., Park, D.J., Pellé, K., Robbins, D., Russ, C., Ryan, E.M., Sacarlal, J., Sogoloff, B., Sorgho, H., Tanner, M., Theander, T., Valea, I., Volkman, S.K., Yu, Q., Lapierre, D., Birren, B.W., Gilbert, P.B., Wirth, D.F., 2015. Genetic diversity and protective efficacy of the RTS,S/AS01 malaria vaccine. N. Engl. J. Med. 373, 2025–2037. https://doi.org/10.1056/NEJMoa1505819.
Needleman, S., Wunsch, C., 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. 48, 443–453.
Nei, M., Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3, 418–426.
Nussenzweig, R.S., Vanderberg, J., Most, H., Orton, C., 1967. Protective immunity produced by the injection of x-irradiated sporozoites of plasmodium berghei. Nature 216, 160–162.
Otto, T.D., Gilabert, A., Crellen, T., Böhme, U., Arnathau, C., Sanders, M., Oyola, S.O., Okouga, A.P., Boundenga, L., Willaume, E., Ngoubangoye, B., Moukodoum, N.D., Paupy, C., Durand, P., Rougeron, V., Ollomo, B., Renaud, F., Newbold, C., Berriman, M., Prugnolle, F., 2018. Genomes of all known members of a plasmodium subgenus reveal paths to virulent human malaria. Nat. Microbiol. 3, 687–697. https://doi.org/10.1038/s41564-018-0162-2.
Parsch, J., Ellegren, H., 2013. The evolutionary causes and consequences of sex-biased gene expression. Nat. Rev. Genet. 14, 83–87. https://doi.org/10.1038/nrg3376.
Plotkin, J.B., Dushoff, J., Fraser, H.B., 2004. Detecting selection using a single genome sequence of M. tuberculosis and P. falciparum. Nature 428, 942–945. https://doi.org/10.1038/nature02458.
Pröschel, M., Zhang, Z., Parsch, J., 2006. Widespread adaptive evolution of Drosophila genes with sex-biased expression. Genetics 174, 893–900.
Rayner, J.C., Liu, W., Peeters, M., Sharp, P.M., Hahn, B.H., 2011. A plethora of plasmodium species in wild apes: a source of human infection? Trends Parasitol. 27, 222–229. https://doi.org/10.1016/j.pt.2011.01.006.
Rich, S.M., Leendertz, F.H., Xu, G., LeBreton, M., Djoko, C.F., Aminake, M.N., Takang, E. E., Diffo, J.L., Pike, B.L., Rosenthal, B.M., Formenty, P., Boesch, C., Ayala, F.J., Wolfe, N.D., 2009. The origin of malignant malaria. Proc. Natl. Acad. Sci. U. S. A. 106, 14902–14907. https://doi.org/10.1073/pnas.0907740106.
Rovira-Graells, N., Gupta, A.P., Planet, E., Crowley, V.M., Mok, S., Ribas de Pouplana, L., Preiser, P.R., Bozdech, Z., Cortés, A., 2012. Transcriptional variation in the malaria parasite plasmodium falciparum. Genome Res. 22, 925–938. https://doi.org/10.1101/gr.129692.111.

Sawyer, S., Parsch, J., Zhang, Z., Hartl, D., 2007. Prevalence of positive selection among nearly neutral amino acid replacements in Drosophila. Proc. Natl. Acad. Sci. U. S. A. 104, 6504–6510.

Serafimova, K., Mihaylov, I., Vassilev, D., Avdjieva, I., Zielenkiewicz, P., Kaczanowski, S., 2020. Using machine learning in accuracy assessment of knowledge-based energy and frequency base likelihood in protein structures. In: Proceedings of the International Conference on Computational Science. Springer, Cham, pp. 572–584.

Slatkin, M., 1994. An exact test for neutrality based on the Ewens sampling distribution. Genet. Res. 64, 71–74. https://doi.org/10.1017/s0016672300032560.

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123, 585–595.

Thompson, J., Gibson, T., Higgins, D., 2002. Multiple Sequence Alignment using ClustalW and ClustalX. Curr Protoc Bioinformatics, Chapter 2: Unit 2.3. https://doi.org/10.1002/0471250953.bi0203s00.

Tuju, J., Kamuyu, G., Murungi, L.M., Osier, F.H.A., 2017. Vaccine candidate discovery for the next generation of malaria vaccines. Immunology 152, 195–206. https://doi.org/10.1111/imm.12780.

Urban, B.C., Ferguson, D.J., Pain, A., Willcox, N., Plebanski, M., Austyn, J.M., Roberts, D. J., 1999. Plasmodium falciparum-infected erythrocytes modulate the maturation of dendritic cells. Nature 400, 73–77. https://doi.org/10.1038/21900.

van Dijk, M.R., van Schaijk, B.C., Khan, S.M., van Dooren, M.W., Ramesar, J., Kaczanowski, S., van Gemert, G.J., Kroeze, H., Stunnenberg, H.G., Eling, W.M., Sauerwein, R.W., Waters, A.P., Janse, C.J., 2010. Three members of the 6-cys protein family of plasmodium play a role in gamete fertility. PLoS Pathog. 6, e1000853 https://doi.org/10.1371/journal.ppat.1000853.

Van Valen, L., 1973. A new evolutionary law. Evol. Theory 1, 1–30.

Videvall, E., 2018. Plasmodium parasites of birds have the most AT-rich genes of eukaryotes. Microb. Genom. 4 https://doi.org/10.1099/mgen.0.000150.

Weber, J.L., 1987. Analysis of sequences from the extremely a + T-rich genome of plasmodium falciparum. Gene 52, 103–109. https://doi.org/10.1016/0378-1119(87)90399-4.

Wilson, D.J., McVean, G., 2006. Estimating diversifying selection and functional constraint in the presence of recombination. Genetics 172, 1411–1425. https://doi.org/10.1534/genetics.105.044917.

Wykes, M.N., Good, M.F., 2008. What really happens to dendritic cells during malaria? Nat. Rev. Microbiol. 6, 864–870. https://doi.org/10.1038/nrmicro1988.

Yang, Z., 2005. PAML FAQ. pp. http://abacus.gene.ucl.ac.uk/software/pamlFAQs.pdf.

Yang, Z., Bielawski, J.P., 2000. Statistical methods for detecting molecular adaptation. Trends Ecol. Evol. 15, 496–503.

Yang, Z., Swanson, W.J., 2002. Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. Mol. Biol. Evol. 19, 49–57. https://doi.org/10.1093/oxfordjournals.molbev.a003981.

Yang, Z., Wong, W.S., Nielsen, R., 2005. Bayes empirical bayes inference of amino acid sites under positive selection. Mol. Biol. Evol. 22, 1107–1118. https://doi.org/10.1093/molbev/msi097.

Yap, V.B., Lindsay, H., Easteal, S., Huttley, G., 2010. Estimates of the effect of natural selection on protein-coding content. Mol. Biol. Evol. 27, 726–734. https://doi.org/10.1093/molbev/msp232.

Zhai, W., Nielsen, R., Slatkin, M., 2009. An investigation of the statistical power of neutrality tests based on comparative and population genetic data. Mol. Biol. Evol. 26, 273–283. https://doi.org/10.1093/molbev/msn231.

Zhang, J., Kumar, S., Nei, M., 1997. Small-sample tests of episodic adaptive evolution: a case study of primate lysozymes. Mol. Biol. Evol. 14, 1335–1338. https://doi.org/10.1093/oxfordjournals.molbev.a025743.

Zhang, Z., Hambuch, T.M., Parsch, J., 2004. Molecular evolution of sex-biased genes in Drosophila. Mol. Biol. Evol. 21, 2130–2139. https://doi.org/10.1093/molbev/msh223.

Zhang, J., Nielsen, R., Yang, Z., 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol. Biol. Evol. 22, 2472–2479. https://doi.org/10.1093/molbev/msi237.

Zhou, T., Enyeart, P.J., Wilke, C.O., 2008. Detecting clusters of mutations. PLoS One 3, e3765. https://doi.org/10.1371/journal.pone.0003765.