UPPSALA
UNIVERSITET

Sequence Diversity and Antibody Response to Autologous and Heterologous MSP2 Antigens in a Prospective Malaria Immunology Cohort

Julia Zerebinski

Master Degree Project in Infection Biology, 45 credits. Spring 2021
Karolinska Institutet Department of Medicine, Department of Infectious Disease
Supervisor: David Plaza

Contents

Abstract

Malaria, caused by the *Plasmodium* parasite and transmitted by mosquitoes, kills almost half a million people each year. Drug resistance in both the parasite and its vector make preventative measures increasingly important, and a fully protective vaccine is absolutely necessary to eradicate the disease. However, genetic diversity of the parasite makes vaccine development difficult. One of the best vaccine candidates is MSP2, a surface protein present during the blood stage of *P. falciparum* infection. Antibodies, which are important for natural immunity, have been shown to bind MSP2 and prevent parasite infection of blood cells. The purpose of this study was to analyze MSP2 sequence diversity in a cohort of patients infected while traveling or living in sub-Saharan Africa, and to investigate patient antibody responses to MSP2 variants infecting other individuals. Parasite isolates from our cohort were made up of 47% 3D7 alleles and 53% FC27 alleles. Protein sequences showed similar levels of conservation within allelic families, and blocks of conserved amino acids between different variants suggest there may be epitopes that can induce antibody production targeting multiple variants. Antibody reactivity tests suggest the variable region of MSP2 is important for antibody binding to variants of the same allelic type, while the conserved region is important for reactivity to different allelic types. This thesis gives evidence to the importance of including epitopes from conserved and variable regions of both MSP2 allelic families in order to induce strain-transcending immunity against *P. falciparum* malaria.

Popular science summary

Malaria is a global threat. It causes 229 million cases and 500,000 deaths yearly and half of the globe lives where humans can be infected. The disease is caused by the complex parasite *Plasmodium falciparum*, which is transmitted through mosquito bites. Because of its complexity and the fact that there are many strains, individuals must be infected many times to become immune. Drug resistance in both the parasite and its vector make development of a fully protective vaccine an urgent priority. The *P. falciparum* life cycle include several stages, each infecting different cells of the human body. One of the best vaccine targets is a protein called MSP2, which is produced by the parasite in the stage before it infects red blood cells (RBC). After infecting our RBC, the parasite breaks open the cells and releases more parasites, and this cycle causes malaria symptoms.

Antibodies help our immune system to stop pathogens from infecting us, and they are the main goal of vaccination. In immune people, antibodies can prevent the parasite from invading RBC, which means we can theoretically use anti-MSP2 antibodies to block RBC invasion as well. However, the challenge with MSP2 lies in its almost unlimited number of genetically distinct variants, a phenomenon called polymorphism. The many variants can be divided into two families called FC27 and 3D7. Previous studies have shown that antibodies against MSP2 are produced by the human immune system during malaria infection, making it an excellent candidate to use in a vaccine. All we have to do, is find out how to put together MSP2 pieces in a way that will prime the immune response against *P. falciparum*.

The aim of this thesis was first to investigate the diversity of parasites taken from patients infected in sub-Saharan Africa. Second, we aimed to determine the extent of patient antibody binding to MSP2 clones from other patients, using an ELISA assay which detects antibodies present in a given sample. In addition, we wanted to evaluate if patient antibodies had different levels of reactivity to two different parts of the MSP2 protein called variable and conserved regions.

Parasite isolates from our cohort were made up of 47% 3D7 type MSP2 and 53% FC27 type MSP2. The protein sequences showed similar levels of conservation within allelic families, meaning they share sections of amino acids, or protein building blocks, that are exactly the same. Antibody reactivity tests suggest the variable region of MSP2 is important for antibody binding to variants of the same type, while the conserved region is important for reactivity to different allelic types. Pieces from both the conserved region and the central variable region of MSP2 should be used to induce antibody responses. The findings of this thesis can be used to direct development of a more effective vaccine capable of protecting against parasite strains across sub-Saharan Africa.

Key words

Malaria, sub-Saharan Africa, merozoite surface protein, antibody, immune response, antigen

Introduction

Malaria is an acute illness which can become severe and even fatal, totaling about 229 million cases and close to half a million deaths yearly [1,2]. Half of the world's population is at risk, with the majority of deaths carried by countries in sub-Saharan Africa [1,3]. People living in malaria-endemic areas can develop partial immunity, which means they may have asymptomatic infections [4]. Clinical malaria is caused by four species of the *Plasmodium* parasite, and the most dangerous species is *P. falciparum* [3]. Diversity of the parasite is large and its life cycle is complex [5]. During feeding of female *Anopheles* mosquitos, the parasite is passed into the bloodstream where it reaches the liver and replicates (pre-erythrocytic stage), releasing daughter merozoites into the blood. These invade red blood cells where an acute, eventually chronic infection is established (erythrocytic stage) [6]. For the purposes of this thesis the rest of the life cycle is not described here. The most vulnerable groups are children under the age of 5, pregnant women, individuals infected with HIV, and travelers or migrants with no previous exposure to malaria [2,7]. In general, non-immune adults develop fever, headache, chills, muscle pain, vomiting and lethargy 7 to 15 days post-infection, while immune individuals develop asymptomatic infection and show low parasitemia. Untreated symptomatic malaria can rapidly turn severe and lead to multi-organ failure [8].

Growing antimalarial drug resistance as well as mosquito insecticide resistance are cause for concern [4]. Many attempts have been made to produce a vaccine against *P. falciparum* malaria; a steady rate of around 10 clinical vaccine trials have been registered per year since 2000, targeting various stages of the parasite life cycle [9,10]. The most promising candidate has been the RTS,S/AS01E vaccine which targets the circumsporozoite protein present in the pre-erythrocytic stage [11]. In its phase III trial, the vaccine reduced clinical malaria by over 30% of the 15,000 children involved [9]. However, the efficacy can be significantly lower depending on the individual's age and vaccine regimen, or parasite strain [11,12]. The most recent vaccine, R21/MM, also targets the circumsporozoite protein and showed a higher protective efficacy than RTS,S/AS01E in its most recent trial, between 74%-77% [13]. Even so, eradication of malaria requires a robust, highly effective vaccine that prevents both erythrocyte invasion and transmission, and protects against many strains.

Immune response to malaria

The innate response is the first line of defense against invasion and acts rapidly but non-specifically [14]. The adaptive immune response allows us to "remember" a pathogen we have previously encountered, producing cells with unique specificity towards target antigens and

thus a more effective response [14,15]. The adaptive response includes B cells, which express and secrete both membrane-anchored and soluble antibodies in order to bind antigens. Antibodies reach a peak soon after antigen stimulation and are maintained long-term at a lower level. Maintenance of pathogen-specific antibodies providing immediate protection against pathogen attack is called 'serological memory' and is the goal of vaccination [16]. Long-term protection is usually achieved by natural infection or live vaccines, which induce the generation of memory B cells (MBC) at a high frequency and constant antibody levels.

Natural immunity against malaria develops slowly, after many infections, and in malaria-endemic areas clinical malaria episodes decline with age [17,18]. Infection with multiple genetically diverse clones is more common in high-transmission regions, such as countries in Africa [19]. These regions also show higher parasite diversity than low-transmission areas [19]. Antibodies are important for reduced infectiousness of malaria parasites. The innate response is triggered by sporozoites[20 21 8] but neutralizing antibody titres are not efficiently induced [22]. In contrast, antibody responses against blood-stage parasites [23], liver stage parasites [24] and gametocytes [25] are protective.

Studying antibody responses allows us to learn how efficiently our immune systems recognize and attack invading pathogens, and how to design vaccines that will confer protection. Several methods have been used over the years to study antibody responses. The enzyme-linked immunosorbent assay (ELISA) uses the principle of antibodies binding the antigen they have specificity for [26]. ELISAs are highly sensitive and allow to test a large number of samples and antigens at a time [27]. The method can be used to detect specific antigens [28] or antibodies, whether for research and development use or for clinical diagnostics [29 30 31]. Individual wells in an ELISA plate are coated with antigens, and specific antibodies (for example from a patient or test animal plasma) are allowed to bind. The antigen-specific antibodies are then bound by a secondary antibody coupled to an enzyme, most frequently horseradish peroxidase or alkaline phosphatase. Addition of the substrate catalyzed by the enzyme results in the formation of a colored product which absorbs light at a specific wavelength. Absorption is then directly proportional to the amount of antigen-specific antibodies present in the sample.

MSP2 protein structure and antigenicity

There are two main antigenic targets for blood-stage *P. falciparum*: merozoite stage surface proteins (MSP) or merozoite secreted proteins, and variant surface antigens which are found on the surface of infected erythrocytes [32]. One of the merozoite surface proteins is the highly polymorphic MSP2 [33]. MSP2 is covalently bound to the merozoite membrane by a glycosylphosphatidylinositol (GPI) anchor and has three structural domains: a central domain made up of repeats that are flanked by variable sequences, and the conserved N- and C-termini (Figure 1) [34,35]. Many forms of the antigen circulate in the population but they can be divided into two allelic families called FC27 and 3D7 [33,36]. Since MSP2 is a surface protein, polymorphism may allow the parasite to escape immune attack [37]. Sequence polymorphism also slows the development of memory B cell and T cell responses[38], which is an important part of developing immune response. The MSP2 polymorphic region is highly informative for genotyping of parasite populations, as well as genotyping of clinical infections to determine type, number of clones present, and to understand if infections are recrudescent or caused by a new parasite clone [39]. Studying these factors is important for malaria epidemiology and understanding the interplay of allelic variants, transmission, and severity of disease.

Natural infection by *P. falciparum* induces protective, anti-MSP2 antibodies towards epitopes within the central variable region [40], and vaccine studies have shown responses against both the central repeats and the conserved termini [7]. In mouse studies, antibodies against the terminal regions have also been elicited [17,41,42]. Studies have already shown that antibodies

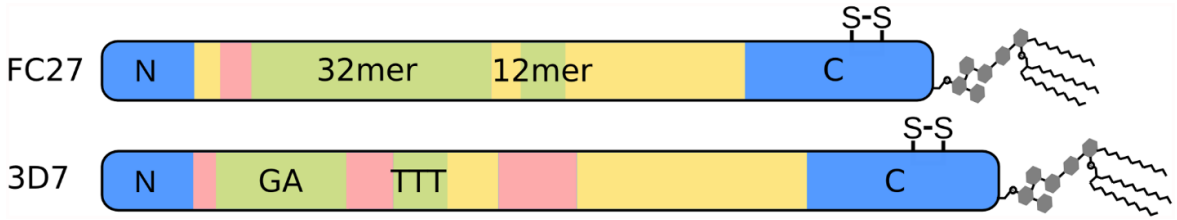


Figure 1: Structure of MSP2 FC27 and 3D7. Blue regions show the conserved N- and C-termini, where the C terminus also includes a disulfide bond and the GPI-anchor. In green, repeat sections which allow distinguishing the two allelic families. In pink and yellow, variable regions. Figure adapted from MacRaild et al 2015 [43].

against MSPs, which are acquired with exposure and boosted with re-infection, are cross-reactive and associated with reduced risk of disease or reduced morbidity [44–47]. In addition, MSP2's direct contact with and accessibility to the immune system contributes to it as a promising vaccine candidate [48]. Thus, major efforts have been underway to determine which epitopes are the most immunogenic and can be combined into a highly efficient vaccine.

Taking into account that immune responses against multiple antigens results in a cumulative, protective effect, antibodies against a single antigenic variant are not sufficient for protection [49,50]. Natural infection also shows signs of allele-specific immunity, which correlates with the broad exposure needed to produce protective immunity [17,51,52]. Within MSP2 families the immune response appears variant-specific to an extent [53,54], however studies have shown cross-reactivity [55] and some cross-protection within serogroups of MSP2 [54,56,57]. This suggests that even with extensive polymorphism and allele-specific responses, a vaccine may only need to include a subset of variants to induce cross-reactive antibodies.

Aim

With increasing drug resistance and the absence of an effective malaria vaccine, research into highly immunogenic *P. falciparum* antigens and broadly protective immune responses is absolutely necessary. Using a cloud-based analysis approach, we wanted to analyze MSP2 diversity and individual antibody responses in a cohort of malaria-infected individuals, according to the following aims:

Specific aims

1. To survey the sequence diversity of *P. falciparum* MSP2 in parasite isolates from a cohort of individuals who were infected in sub-Saharan Africa.
2. To evaluate the existence and extent of autologous and heterologous specificity of patient plasma 1 month post-treatment towards MSP2 variants cloned and expressed recombinantly in a mammalian expression system.
3. To evaluate any differences in antibody binding to conserved termini and central variable regions of MSP2 variants.

Methods

Ethical approval

Patient samples have been previously approved for study by the Regional Ethical Committee in Stockholm [58] in addition to written informed consent from participants. For cohort demographics, clinical and epidemiological data, episodes reported in the National Surveillance Database at the Public Health Agency of Sweden were linked to hospital data

through unique national identity number of Swedish residents or temporary numbers given to newly arrived migrants given that Malaria is a notifiable disease in Sweden according to the Communicable Diseases Act [59].

Cohort Info

Our Prospective Malaria Immunology Cohort was made up of 61 adults treated for malaria at Karolinska University Hospital in Stockholm, Sweden, between 2011 and 2018 and followed over one year after treatment. Blood samples were collected at diagnosis. Patients belong to two groups: Swedish individuals who travelled to countries in sub-Saharan Africa (sSA) and were treated after returning to Sweden, called "primary infected", or individuals who immigrated to Sweden from countries in sSA, called "previously exposed". 9 female (15%) and 50 male (82%) patients were included. 46 individuals (75%) had lived for an extended period in sSA. At time of sampling, there were 22 patients (36%) between 20 and 35 years old, 17 patients (28%) between 36-45 years old, 13 patients (21%) between 46-55 years old, and 7 patients (11%) above the age of 56. Demographic data was not received for 2 individuals. The most common symptoms were fever, headache, chills and body pain. 14 patients developed severe malaria. Countries with the highest number of samples were Kenya (7), Ghana (6), Tanzania (5), and Nigeria (5). Isolate origins are summarized in Supplementary Figure S1. 6 individuals from the cohort reported travel to more than one country.

Bioinformatics

Sequencing results were obtained courtesy of David Plaza and Ioanna Broumou (Karolinska Institutet). Circular consensus sequencing (CCS) was performed on *msp2* from the malaria immunology cohort by SciLifeLab (Uppsala Sweden). After demultiplexing, *in silico* PCR was run on the sequence data, adapted from capillary electrophoresis as developed by Anne Liljander et al (2009) [60] and previously a nested PCR as described by Georges Snounou et al. (1999, 2002) [61,62]. To do so, a BLAST database was constructed from CCS long reads and the capillary electrophoresis primers (shown below) were used as search queries [63–65].

>ICF
AGAAGTATGGCAGAAAGTAAKCCTYCTACT
>ICR
GATTGTAATTCGGGGGATTCAGTTTGTTCG
>FC27F
AATACTAAGAGTGTAGGTGCARATGCTCCA
>FC27R
TTTTATTTGGTGCATTGCCAGAACTTGAAC

A blastn-short program was run optimized for sequences shorter than 50 bases, with expectation value cutoff 0.001, maximum hits 1000000, percent identity cutoff 50.0, and minimum query coverage per alignment of 50.0. Size variants were calculated using the following (from D. Plaza, unpublished data):

$$size\ variant\ (bp) = \frac{Rv\ strand\ x\ (seqstartFw - seqstartRv)}{|Rv\ strand|} + 1$$

Where "Rv strand" is the strand (negative or positive) of oligo annealing, and "seqstartFw" or "seqstartRv" are the nucleotide positions at which annealing occurs.

PCR or sequencing artifacts needed to be distinguished from correct size variants calls. To define a coverage threshold, read coverages for synthetic controls were pooled and the mean coverage for the most common size variant miscall plus approximately two times the standard

deviation, rounded to ≥0.20 or 20% of reads, was calculated (Figure 2). The nested PCR product for the reference FC27 strain is 292bp long while the product for reference strain 3D7 is 473bp long as determined previously [66–68].

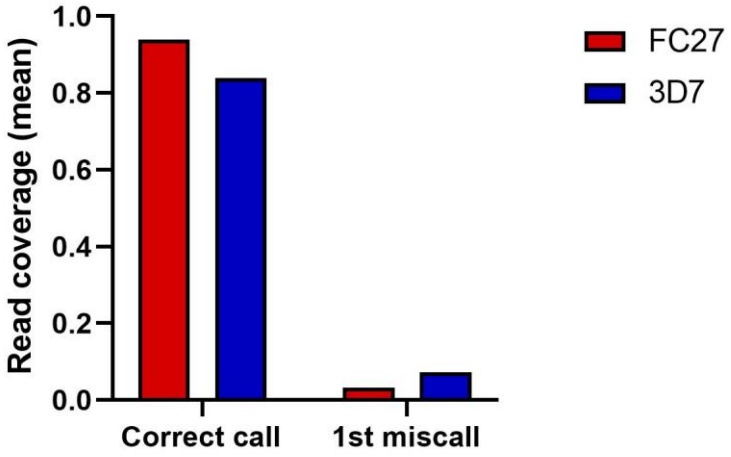


Figure 2: Size variant ranking. PCR artifacts and miscalls were filtered from sequencing results by a 20% sample coverage threshold, determined by pooling coverage for each size variant per sample, and calculating the mean coverage of the 1st miscall plus two times the standard deviation.

Before translating CCS reads passing the threshold into amino acid sequences, PCR primers from the first amplification reaction were added to the 5' and 3' ends of every read. Sequences were then translated using the Galaxy transeq tool [69]. Miscalls in the resulting amino acid sequences were then filtered out using a 10 percent read coverage, based on a stringent coverage threshold that corresponds to the mean coverage of the most common sequence miscall plus 9 times the standard deviation.

Amino acid sequences selected for cloning matched the following criteria: samples from individuals that were previously unexposed to *P. falciparum* (primary infection), having size variants with greater than 20 percent coverage per sample, having an amino acid sequence with greater than 10 percent coverage per sample, and having a support of at least 20 reads per sample. Two single-clone infections were chosen as well as two individuals with both msp2 allelic families. Table 1 summarizes this information.

Table 1: Selected antigens for cloning and expression in mammalian cells. Size variant in base pairs and number of reads from CCS sequencing. Percent coverage of highest coverage protein sequence, per sample.

Sample	Allele	Size variant (bp)	Number supporting reads	% Coverage protein sequence	Protein Sequence
2014003	FC27	292	25	72.0	MKVIKTLSIINFFIFVTFNIKNESKYSNTFINN AYNMSIRRSMANEGSNTTSVGANAPNADT IANGSQSSTNSASTSTTNNGESQTTTPTAAD TPTATKSNSPSPPITTTKSNSPSPPITTTKSNS PSPPITTTESSSSGNAPNKTDGKGEESEKQN ELNESTEEGPKAPQEPQTAENENPAAPENK GTGQHGHMHGSRNNHPQNTSDSQKECTD GNKENCGAATSLLNNSSNIASINKFVVLISA TLVLSFAIFI
2015004	3D7	575	238	97.1	MKVIKTLSIINFFIFVTFNIKNESKYSNTFINN AYNMSIRRSMEESNPSTGAGGSGSAGGSGS AGGSGSAGGSGSAGGSGSAGGSGSAGGSG SAGGSGSAGGSGSAGGSGSAGSGDGNGAN PGADAERSPSTPATTTTTTTTNDAEASTSTS SENPNHNNAETNPKGKGEVQKPNQANKET QNNSNVQQDSQTKSNVPPTQDADTKSPTA QPEQAENSAPTAEQTESPELQSAPENKGTG QHGHMHGSRNNHPQNTSDSQKECTDGNK ENCGAATSLLSNSSNIASINKFVVLISATLV LSFAIFI
2017001	3D7	707	77	80.5	MKVIKTLSIINFFIFVTFNIKNESKYSNTFINN AYNMSIRRSMAESNPSTGAGGSGSAGGSA GGSAGGSAGGSAGGSAGGSAGGSAGGSAG GSAGGSAGGSAGGSAGGSAGGSAGGSAGG SAGGSAGGSAGGSAGGSAGGSAGGSAGGS AGGSAGGSAGGSAGSGDGNGANPGADAE GSSSTPATTTTTTTTTTTNDAEASTSTSSEN PKGKGEVQKPNQANKETQNNSNVQQDSQ TKSNVPRTQDADTKSPTAQPEQAENSAPTA EQTESPELQSAPENKGTGQHGHMHGSRNN HPQNTSDSQKECTDGNKENCGAATSLLNN SSNIASINKFVVLISATLVLSFAIFI
2012PT12	FC27	376	571	95.1	MKVIKTLSIINFFIFVTFNIKNESKYSNTFINN AYNMAIRRSMANKGSNTNSVGANAPNAD TIASGSQRSTNSASTSTTNNGESQTTTPTAA DTIASGSQRSTNSASTSTTNNGESQTTTPTA ADTIASGSQRSTNSASTSTTNNGESQTTTPT AADTPTATESSSSGNAPNKADGKGEESEKQ NELNESTEEGPKAPQEPQTAENENPAAPEN KGTGQHGHMHGSRNNHPQNTADSQKECT DGNKENCGAATSLLNNAANIASINKFVVLI SATLVLSFAIFI

Four sample were chosen from the cohort for cloning and expression in HEK293T/17 cells. This included two FC27- and two 3D7-family MSP2 sequences. Minimum sample coverage of the four was 72%, and the lowest number of reads from sequencing was 25 (Table 1). Samples were taken from patients between 2012 and 2017. The selected sequences were aligned against synthetic control sequences, which were from reference strains HB3 (FC27) and 3D7 (Figure 3).

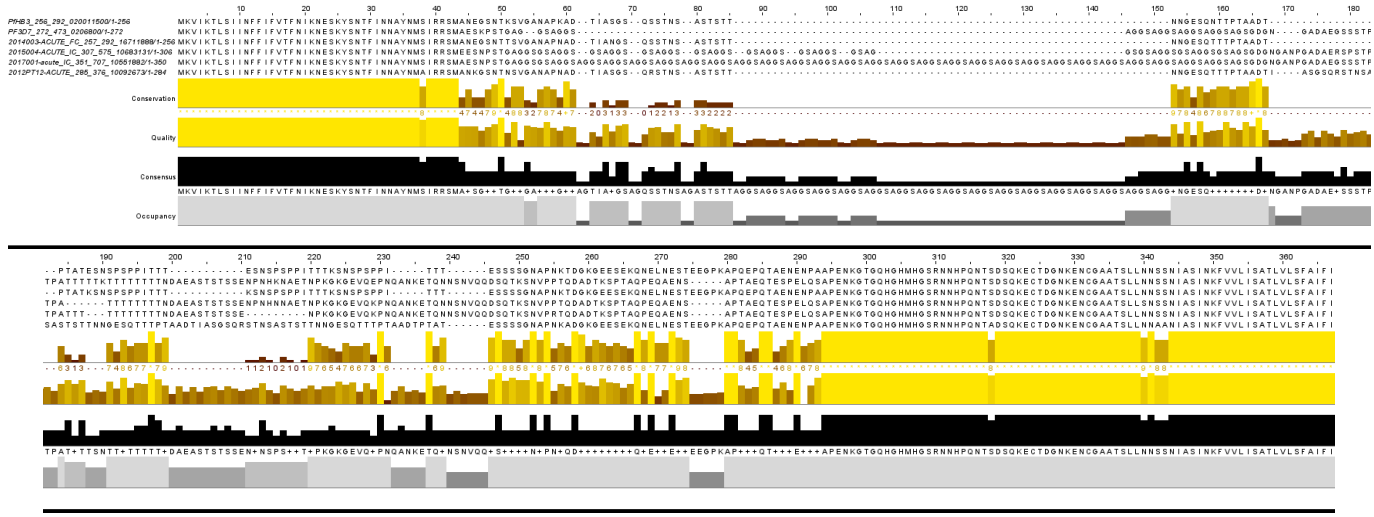


Figure 3: Selected four protein sequences aligned with reference strains HB3 (FC27), top line and 3D7, second line. Protein sequences were aligned using Jalview Software [70]. Terminal ends are highly conserved between selected four protein sequences and the reference strain MSP2 protein sequence.

N-glycosylation sites were predicted using the NetNGlyc 1.0 Server [71,72]; While controversial for many years, *Plasmodium* has been shown to use N-glycosylation although not for secretion like in eukaryotes [73]. When using mammalian cells for protein production, N-glycosylation sites should be modified so that glycosylation does not occur and synthesized proteins are folded the way they would be in their natural state. Serine or threonine at the predicted glycosylation motifs were substituted with alanine to prevent N-glycosylation during expression in human cells.

Signal peptide from mouse IgG kappa chain was added to the N-terminus of all eight constructs to drive secretion into culture supernatant to facilitate harvesting. Full sequence MSP2 constructs were also processed to lack native signal peptides [74] and GPI-anchor signals [75]. The eight sequences were then tagged (Table 2).

Table 2: Peptide tags added to the C-terminus of each optimized, glycosylation-site-mutated protein sequence, both for full version and variable-region-only.

Sample	Tag	Tag sequence
2014003	TWIN	SAWSHPQFEKGGGSGGGSGGSAWSHPQFEK
2015004	GAL	YPGQAPPGAYPGQAPPGA
2017001	WASP	PDYRPYDWASPDYRD
2012PT12	TRAP	DDFLSQQPERPRDVKLA

Tagged, codon-optimized antigen sequences were added to pcDNA3.1/Zeo(-) plasmids by GeneScript (Piscataway, NJ, USA). Transfection and recombinant protein expression in HEK293T/17 cells was performed by MabTech (Nacka Strand, Sweden) as described by Arestrom et al 2012. Transfected cells were cultured 6 days before harvesting. Recombinant proteins were purified and buffer-exchanged into PBS before storage in liquid nitrogen (MabTech, Nacka Strand, Sweden).

ELISA

ELISA was performed as previously described [50]. Briefly, 500ng/ml solutions of antigen were prepared in carbonate coating buffer (15mM Na2CO3, 35 mM NaHCO3, pH 9.3). ELISA plates were coated with 100ul/well of antigen coating solution and incubated overnight at 4°C. Wells were then washed four times in PBS-Tween (PBS-0.05% Tween 20) before blocking 5 hours at room temperature with 1% bovine serum albumin in washing buffer (200ul/well). Thereafter, plates were washed once and incubated overnight at 4°C with 100 ul test plasma, at dilutions of 1:1000, 1:2000, 1:4000, 1:8000, 1:16000, or 1:32000 in blocking buffer (schema in figure 3). After washing five times, rabbit anti-human IgG conjugated with horseradish peroxidase (HRP) was added at 1/5000 dilution in blocking buffer and incubated for 3 hours at room temperature. 100ul/well 3,3',5,5'-Tetramethylbensidine (TMB) was added for 15 minutes before stopping the reaction with 100u/well 0.18M H_2SO_4. Absorbance was measured at 450nm. Two negative controls were used: a pool of plasma from unexposed patients and secondary antibody (no plasma) only, plus a positive control of pooled exposed patient plasma ("hyperimmune"). Single-dilution plasma ELISA optical density values were used as estimation of antibody titers.

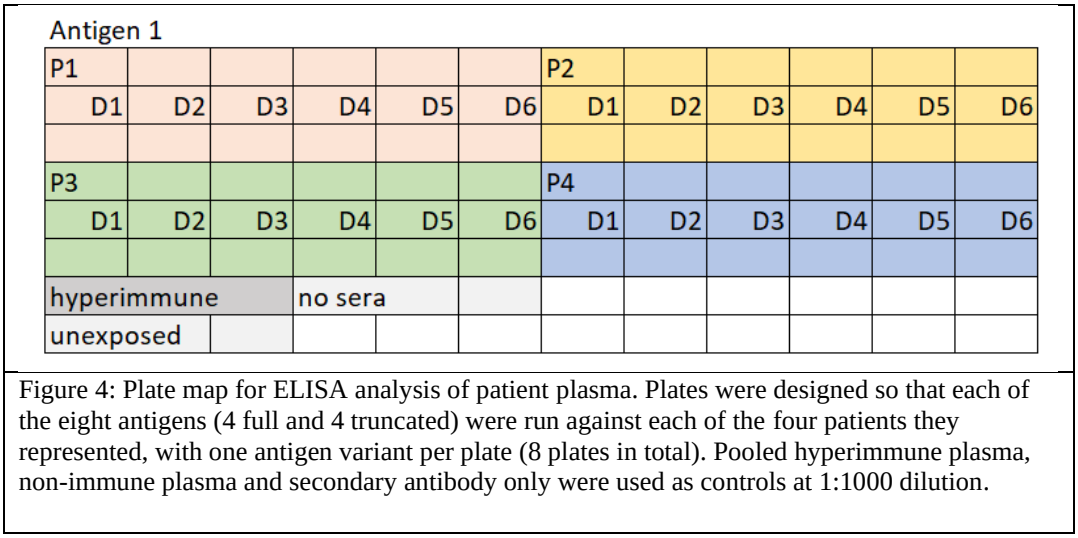


Figure 4: Plate map for ELISA analysis of patient plasma. Plates were designed so that each of the eight antigens (4 full and 4 truncated) were run against each of the four patients they represented, with one antigen variant per plate (8 plates in total). Pooled hyperimmune plasma, non-immune plasma and secondary antibody only were used as controls at 1:1000 dilution.

Results

3D7 family MSP2 are more polymorphic than FC27 MSP2

We wanted to survey the size variant polymorphism in MSP2 using *in silico* PCR. Synthetic controls were used to determine a coverage threshold to be used in calling correct variants (Figure 2), and a 20% threshold of coverage per sample was applied. Resulting nucleotide sequences were then pooled into FC27 and 3D7 type groups, and the number of sequences called for each size variant was counted (Figure 5). 29 different size variants were found for the FC27 family MSP2, and 209 nucleotide sequences passing the 20% threshold (Figure 5A). The most frequent size variants within FC27 family proteins were 292bp with 57 hits (27.2% of sequences) and 376bp long with 40 hits (19% of sequences). For 3D7-type proteins, 184 sequences passed the 20% threshold and 65 different size variants were called (Figure 5B). The most frequent size variants were 479bp with 21 hits (11.4% of sequences) and 485bp with 20 hits (10.8%).

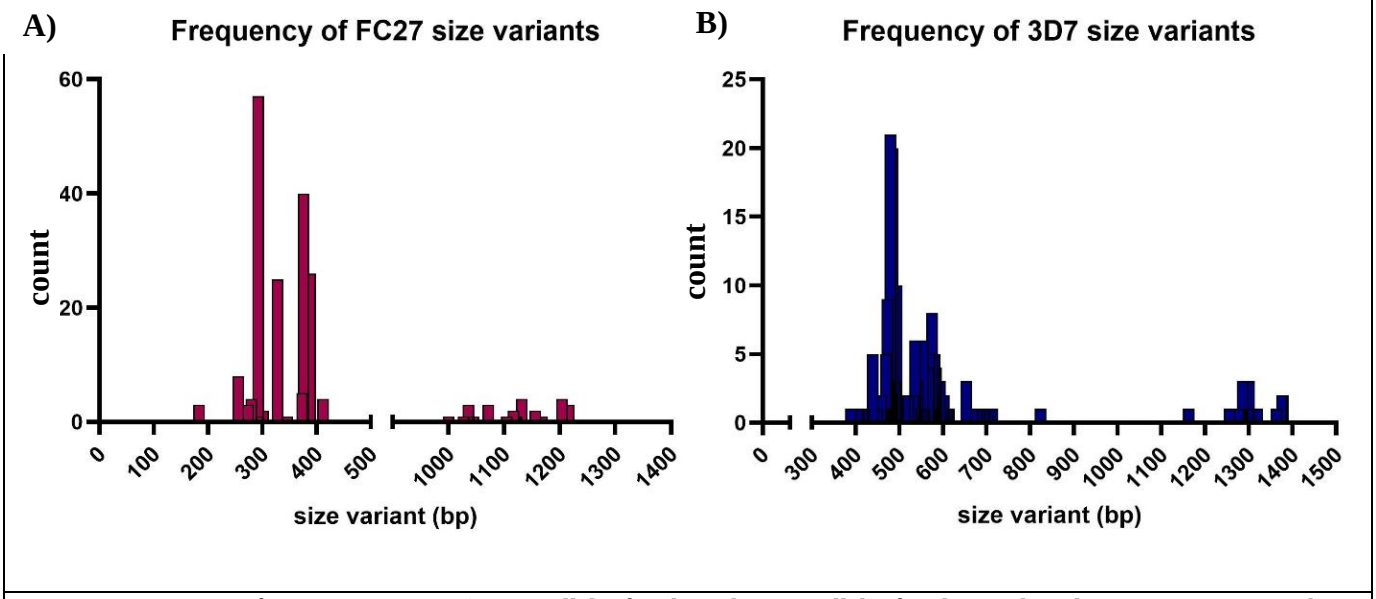


Figure 5: Frequency of size variants per A) FC27 allelic family and B 3D7 allelic family. Nucleotide sequences passing the 20% threshold, after determining size variant by *in silico* PCR, were pooled and the number of nucleotide sequences called for each unique size variant were counted. The total count of size variants for FC27 and 3D7 were 29 and 65, respectively. 209 nucleotide sequences passed the 20% threshold for FC27 and 184 sequences for 3D7.

FC27 and 3D7 nucleotide sequences passing the 20% threshold were hierarchically clustered by presence of size variant and by sample using Cluster 3.0 (Figure 6). Maximum coverage of 1 corresponded to the size variant covering 100% of sequences for that sample. According to this method, 3D7 MSP2 was more polymorphic than FC27. FC27 sequences show higher coverage for 292bp and 376bp variants (Figure 6A). 3D7 sequences show several high coverage size variants, including 485bp and 479bp (Figure 6B). Samples from the cohort showed multiple size variants within both FC27 and 3D7 families.

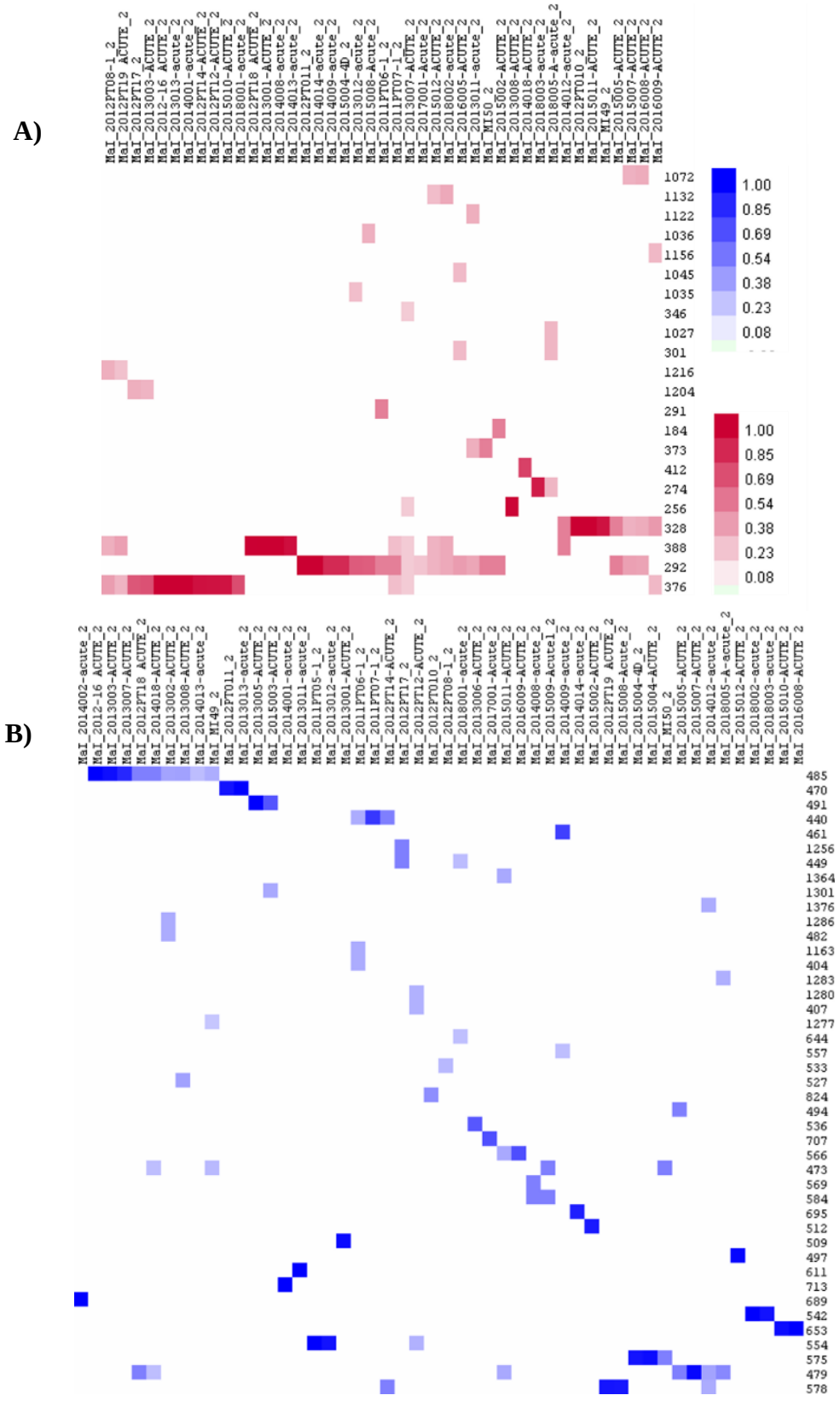


Figure 6: Hierarchical clustering of size variant and samples. A) FC27 family size variants and B) 3D7 family size variants. Coverage range shown in legend; darker color denotes higher coverage per sample, where 1 indicates a single size variant of 3D7 or FC27 per sample.

Highly conserved blocks of amino acids were shared between both allelic families

Our next aim was to evaluate sequence conservation between MSP2 for the cohort's parasite isolates. A more stringent 10% threshold was applied after nucleotide sequences were filtered by the 20% threshold described earlier, and after the resulting nucleotide sequences were translated to amino acid sequences. Protein sequences passing the 20% and 10% threshold for size and sequence variant calling were aligned using MAFFT multiple sequence alignment software [76] and visualized in Jalview [70,77,78]. Sequence conservation was determined by AAcon [79] using the Valdar consensus score [80] (Figure 7). N- and C- terminal regions were completely conserved, as expected.

Distinct blocks of conserved sequences between most of the cohort were found. FC27 sequences showed approximately double the conservation value of 3D7 sequences, with 250 residues showing conservation scores ≥ 0.5 for FC27 (out of 2269) and 259 residues (out of 4824) showing conservation scores ≥ 0.5 for 3D7. 3D7 sequences shared smaller sections of conservation than FC27, in addition to more sections with very low conservation.

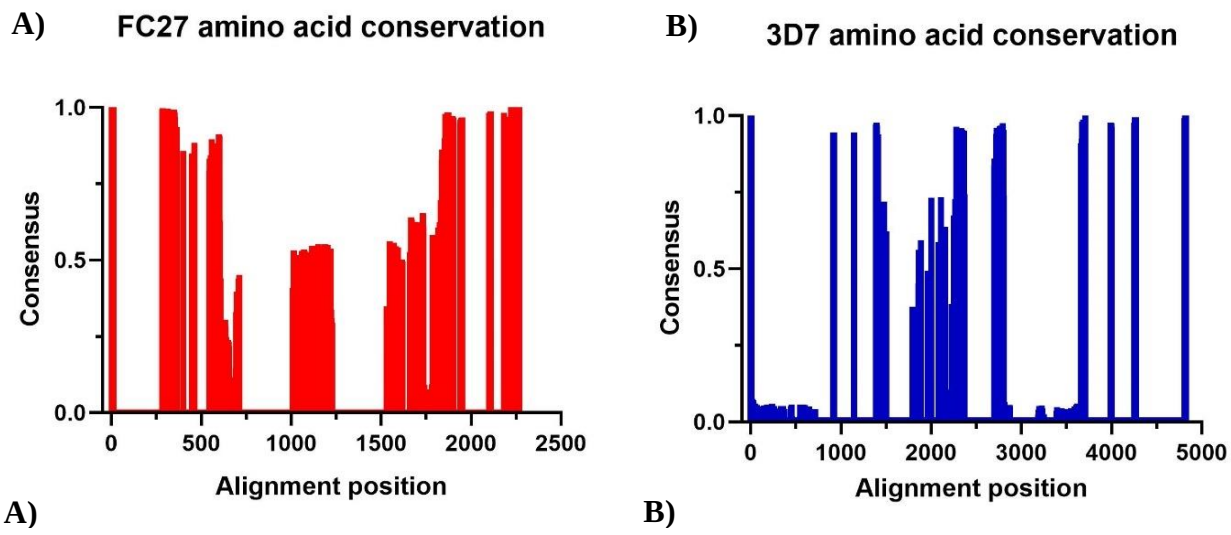
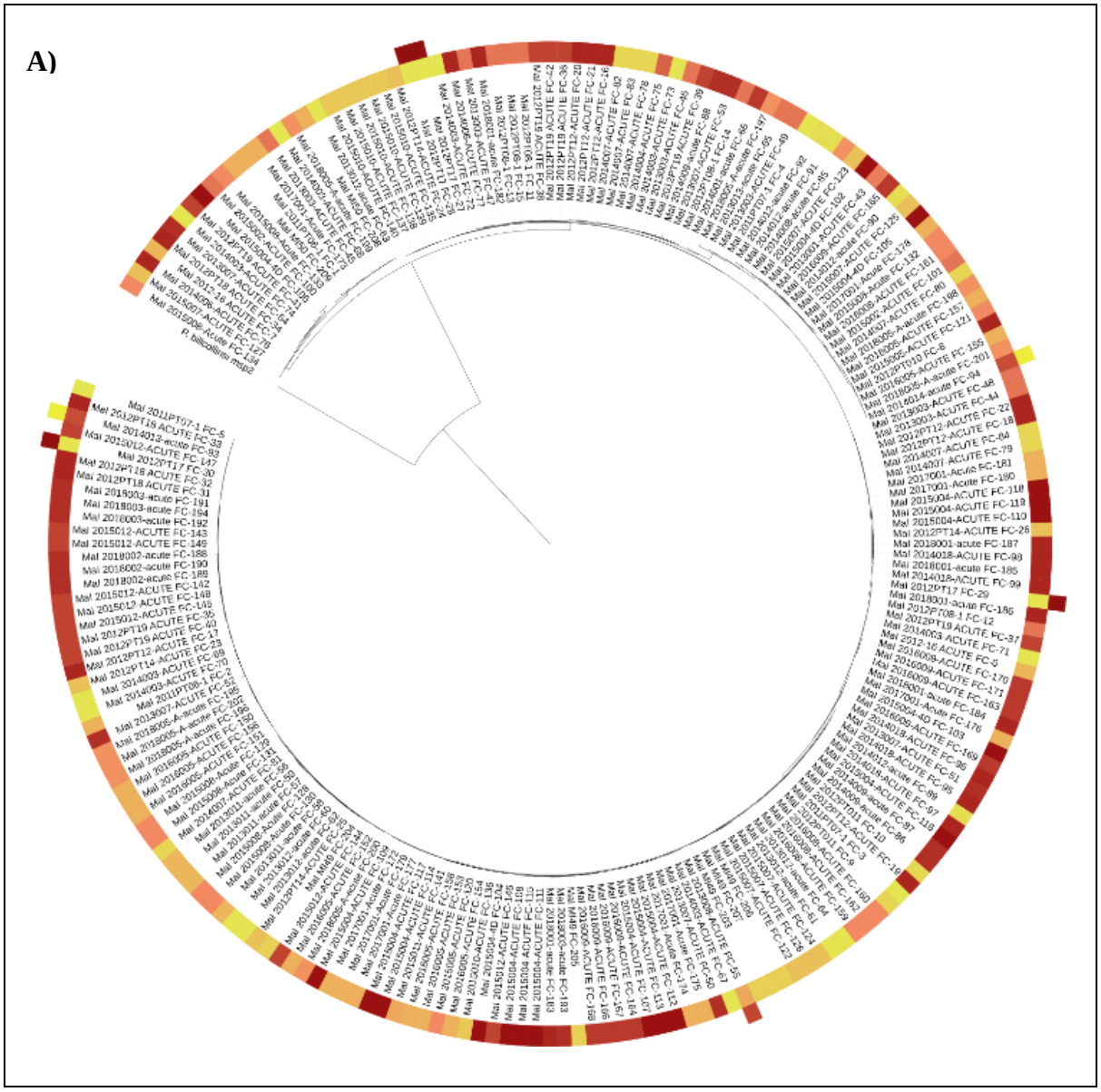


Figure 7: Amino acid consensus of A) FC27 family MSP2 and B) 3D7 family MSP2 from the cohort using normalized Valdar consensus and visualized in GraphPad Prism version 9.1.0..

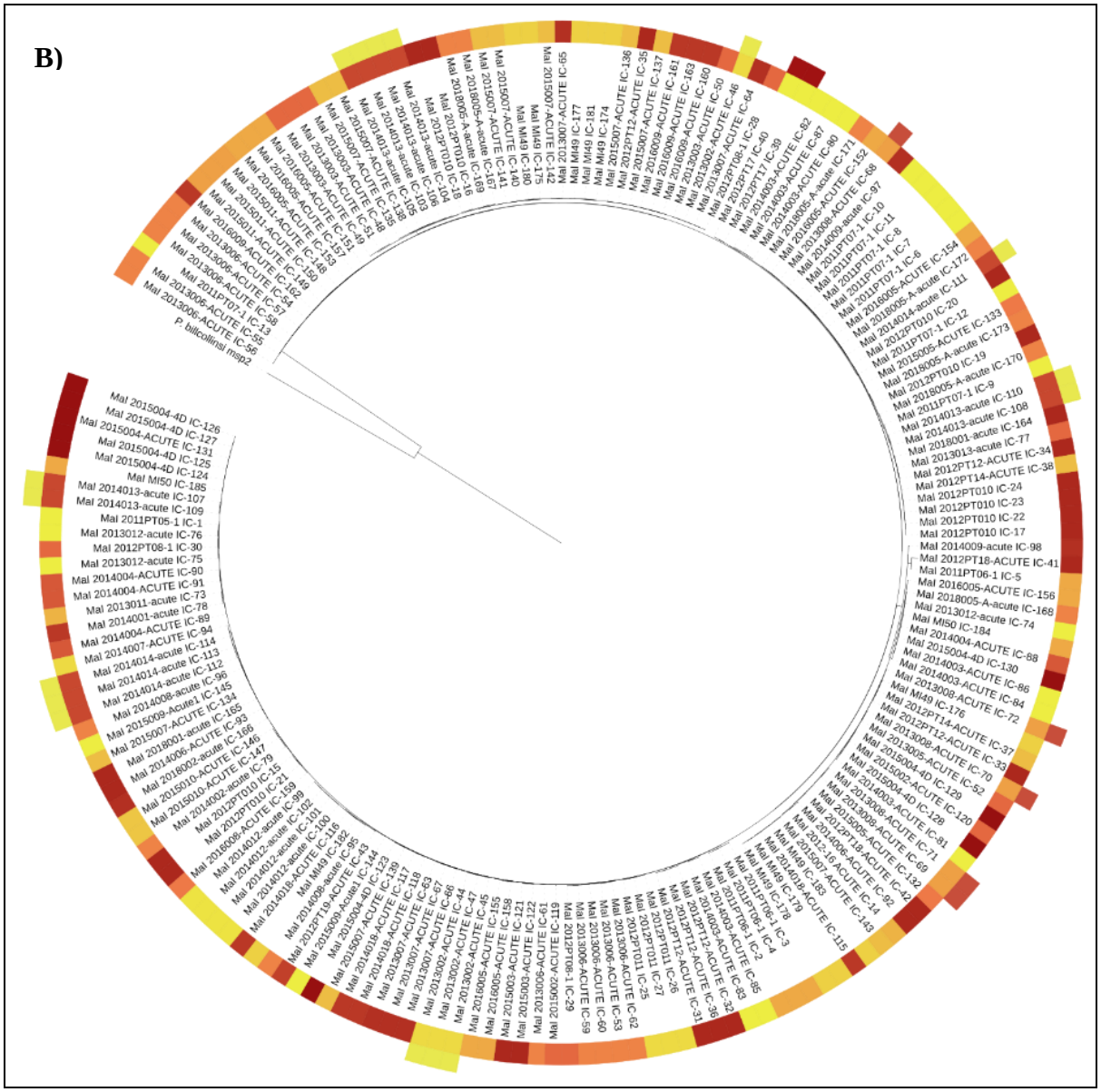
Phylogeny by MSP2 did not show a distinct geographical structure

In order to assess parasite and *msp2* relatedness, nucleotide sequences were aligned using MAFFT. For evolutionary analysis, IQ-tree [81] with Model Finder [82] was run with single branch replicates of 1000 and ultrafast bootstrapping [83] at 1000 replicates. Sequences were rooted with the ortholog of *msp2* present in *P. billcollinsi*, a species infecting chimpanzees [84]. Genetic relatedness among P. falciparum isolates was predicted using Maximum Likelihood analysis and visualized in iTOL[85] (Figure 8). The isolate phylogeny did not show a distinct geographical structure.

A)



B)


P. billcolinsi msp2
Mal 2013006-ACUTE IC-55
Mal 2013006-ACUTE IC-56
Mal 2011PT07-1 IC-13
Mal 2013006-ACUTE IC-57
Mal 2013006-ACUTE IC-58
Mal 2016009-ACUTE IC-44
Mal 2015011-ACUTE IC-150
Mal 2015011-ACUTE IC-149
Mal 2016005-ACUTE IC-153
Mal 2013003-ACUTE IC-157
Mal 2016005-ACUTE IC-154
Mal 2015007-ACUTE IC-48
Mal 2015007-ACUTE IC-49
Mal 2013003-ACUTE IC-51
Mal 2014013-acute IC-105
Mal 2014013-acute IC-106
Mal 2012PT010 IC-16
Mal 2018005-A-acute IC-167
Mal 2015007-ACUTE IC-140
Mal MI49 IC-180
Mal MI49 IC-175
Mal 2015007-ACUTE IC-142
Mal 2013007-A-ACUTE IC-95
Mal MI49 IC-177
Mal MI49 IC-181
Mal MI49 IC-174
Mal 2015007-ACUTE IC-136
Mal 2012PT12-ACUTE IC-35
Mal 2015007-ACUTE IC-137
Mal 2016009-ACUTE IC-161
Mal 2016009-ACUTE IC-163
Mal 2013003-ACUTE IC-160
Mal 2013007-ACUTE IC-50
Mal 2012PT08-1 IC-64
Mal 2012PT17 IC-40
Mal 2012PT17 IC-39
Mal 2014003-ACUTE IC-62
Mal 2014003-ACUTE IC-87
Mal 2016005-A-acute IC-171
Mal 2016005-ACUTE IC-152
Mal 2013008-ACUTE IC-68
Mal 2014009-acute IC-97
Mal 2011PT07-1 IC-10
Mal 2011PT07-1 IC-11
Mal 2011PT07-1 IC-8
Mal 2011PT07-1 IC-7
Mal 2011PT07-1 IC-6
Mal 2011PT07-1 IC-9
Mal 2018005-ACUTE IC-154
Mal 2018005-A-acute IC-172
Mal 2018005-A-acute IC-111
Mal 2014014-acute IC-20
Mal 2012PT010 IC-20
Mal 2012PT07-1 IC-12
Mal 2011PT07-1 IC-133
Mal 2015005-A-acute IC-173
Mal 2015005-A-acute IC-19
Mal 2018005-A-acute IC-170
Mal 2012PT010 IC-19
Mal 2018005-A-acute IC-9
Mal 2011PT07-1 IC-9
Mal 2014013-acute IC-110
Mal 2014013-acute IC-108
Mal 2014013-acute IC-164
Mal 2018001-acute IC-77
Mal 2013013-acute IC-34
Mal 2012PT12-ACUTE IC-38
Mal 2012PT14-ACUTE IC-24
Mal 2012PT010 IC-24
Mal 2012PT010 IC-23
Mal 2012PT010 IC-22
Mal 2012PT010 IC-17
Mal 2014009-acute IC-98
Mal 2012PT18-ACUTE IC-41
Mal 2011PT06-1 IC-5
Mal 2016005-ACUTE IC-156
Mal 2018005-A-acute IC-168
Mal 2013012-acute IC-74
Mal MI50 IC-184
Mal 2014004-ACUTE IC-88
Mal 2015004-4D IC-130
Mal 2014003-ACUTE IC-86
Mal 2013008-ACUTE IC-84
Mal MI49 IC-176
Mal 2012PT14-ACUTE IC-37
Mal 2013008-ACUTE IC-33
Mal 2013005-ACUTE IC-70
Mal 2015004-4D IC-52
Mal 2015002-ACUTE IC-129
Mal 2015004-4D IC-120
Mal 2014003-ACUTE IC-128
Mal 2013008-ACUTE IC-81
Mal 2015005-ACUTE IC-71
Mal 2012PT18-ACUTE IC-69
Mal 2014006-ACUTE IC-132
Mal 2012PT18-ACUTE IC-92
Mal 2012PT18-ACUTE IC-42
Mal 2015007-ACUTE IC-14
Mal 2014018-ACUTE IC-183
Mal MI49 IC-179
Mal 2011PT06-1 IC-3
Mal MI49 IC-178
Mal 2011PT06-1 IC-4
Mal 2011PT06-1 IC-2
Mal 2014003-ACUTE IC-85
Mal 2012PT12-ACUTE IC-32
Mal 2012PT12-ACUTE IC-36
Mal 2012PT12-ACUTE IC-31
Mal 2012PT011 IC-26
Mal 2012PT011 IC-27
Mal 2012PT011 IC-25
Mal 2013006-ACUTE IC-62
Mal 2013006-ACUTE IC-53
Mal 2013006-ACUTE IC-60
Mal 2013006-ACUTE IC-59
Mal 2012PT08-1 IC-29
Mal 2015002-ACUTE IC-19
Mal 2013006-ACUTE IC-61
Mal 2015003-ACUTE IC-122
Mal 2015003-ACUTE IC-121
Mal 2015003-ACUTE IC-158
Mal 2016005-ACUTE IC-155
Mal 2013002-ACUTE IC-45
Mal 2013002-ACUTE IC-47
Mal 2013002-ACUTE IC-44
Mal 2013002-ACUTE IC-68
Mal 2013007-ACUTE IC-67
Mal 2013007-ACUTE IC-65
Mal 2014018-ACUTE IC-117
Mal 2015007-ACUTE IC-139
Mal 2015004-4D IC-123
Mal 2015003-Acute1 IC-144
Mal 2015003-Acute1 IC-43
Mal 2014008-acute IC-95
Mal 2012PT19-ACUTE IC-182
Mal MI49 IC-116
Mal 2014008-acute IC-101
Mal 2014012-acute IC-102
Mal 2014012-acute IC-99
Mal 2014012-acute IC-100
Mal 2016008-ACUTE IC-159
Mal 2012PT010 IC-21
Mal 2014002-acute IC-79
Mal 2015010-ACUTE IC-147
Mal 2015010-ACUTE IC-146
Mal 2018002-acute IC-166
Mal 2014006-ACUTE IC-93
Mal 2018001-acute IC-165
Mal 2015007-ACUTE IC-134
Mal 2015009-Acute1 IC-145
Mal 2014008-acute IC-96
Mal 2014014-acute IC-112
Mal 2014014-acute IC-113
Mal 2014007-ACUTE IC-114
Mal 2014004-ACUTE IC-94
Mal 2014001-acute IC-89
Mal 2013011-acute IC-73
Mal 2014004-ACUTE IC-91
Mal 2014004-ACUTE IC-90
Mal 2013012-acute IC-75
Mal 2012PT08-1 IC-30
Mal 2013012-acute IC-76
Mal 2011PT05-1 IC-1
Mal 2014013-acute IC-109
Mal 2014013-acute IC-107
Mal MI50 IC-185
Mal 2015004-4D IC-124
Mal 2015004-4D IC-125
Mal 2015004-ACUTE IC-131
Mal 2015004-4D IC-127
Mal 2015004-4D IC-126

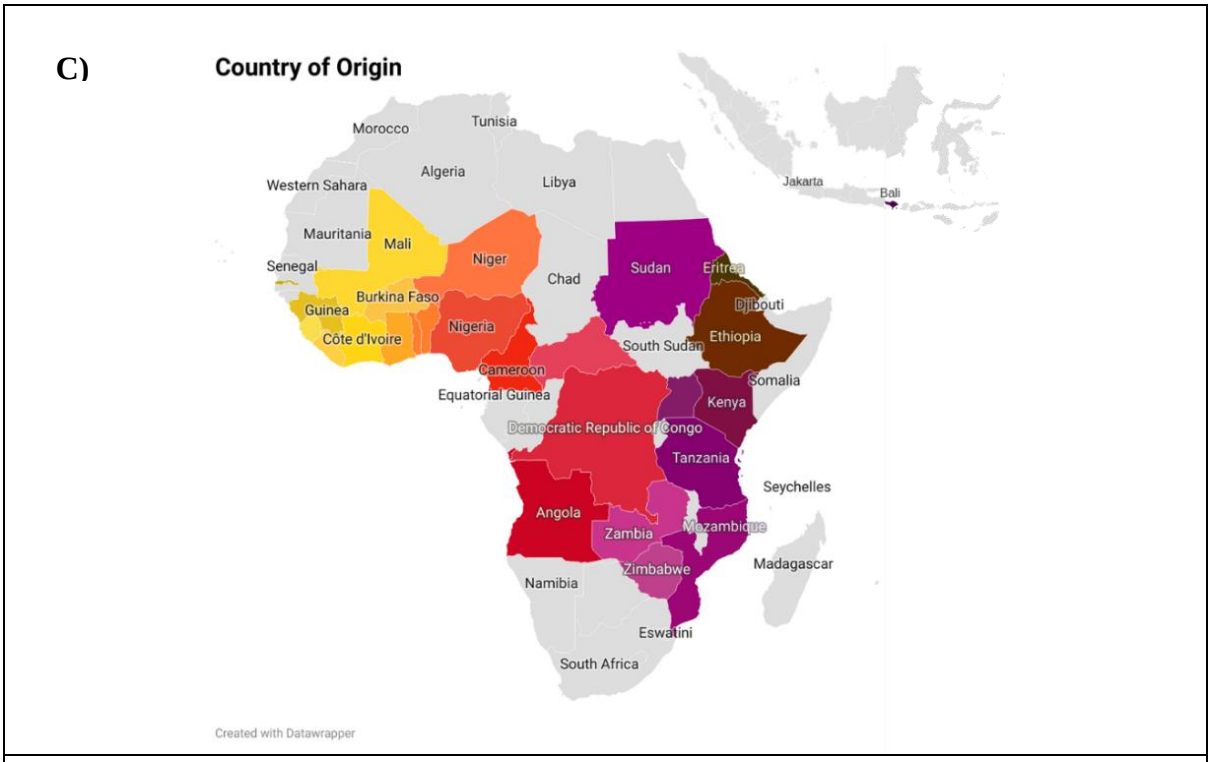


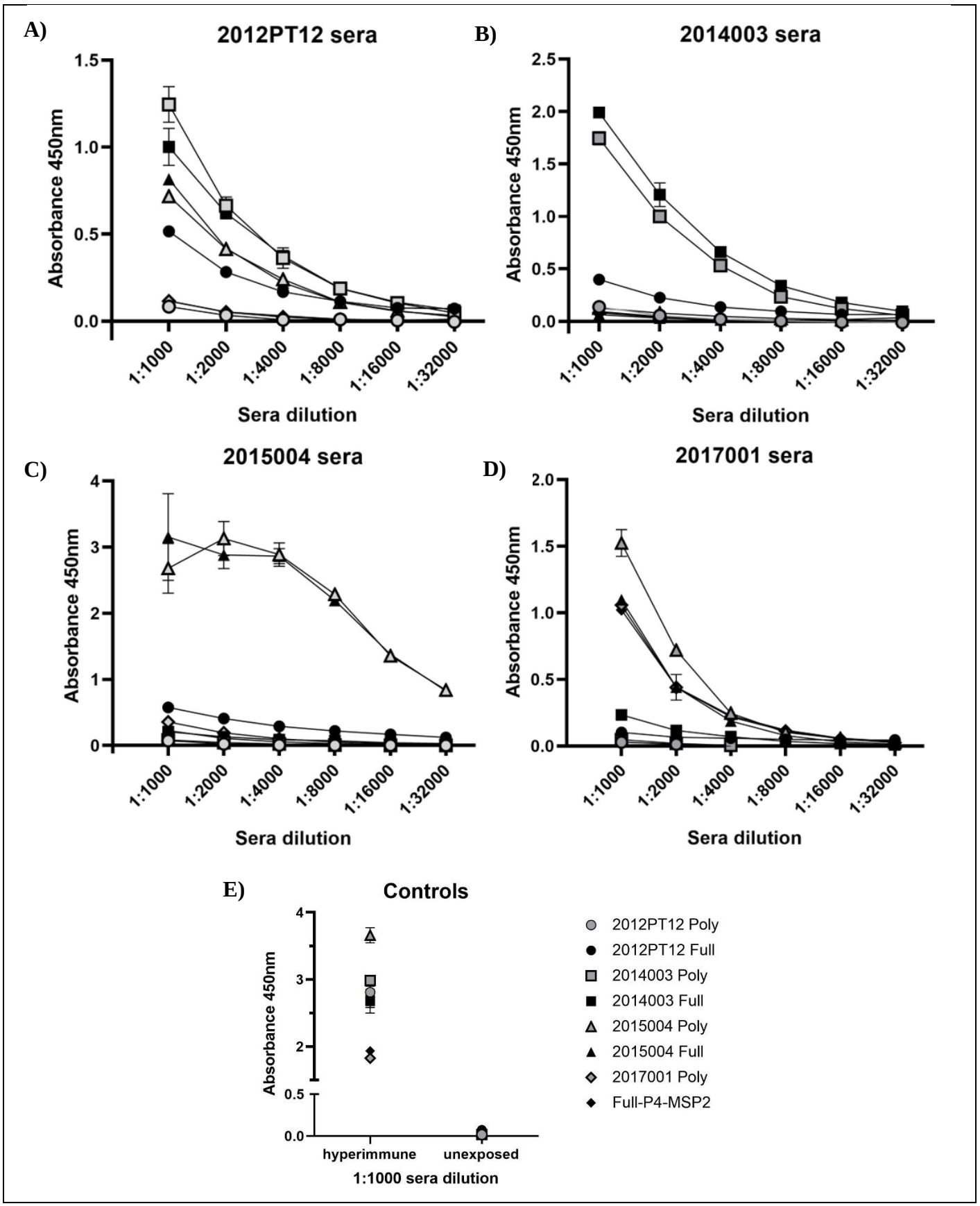
Figure 8: Phylogenetic trees using MAFFT-aligned nucleotide sequences covering both size variant and amino acid sequence thresholds. A) FC27 family B) 3D7 family. C) Map of countries of origin. Map colors match color gradient in phylogenetic trees A and B. Second row of color blocks represent samples that visited more than one country.

FC27-family MSP2 contain predicted T-cell binding sites

We were interested in potential epitopes on the MSP2 within our cohort. To do so we used the Immune Epitope Database and Analysis Resource (IEDB) to predict MHC class II binding sites [86,87], using the immunodominant HLA alleles DRB1 03:01, DRB1 07:01, DRB1 15:01, DRB3 01:01, DRB3 02:02, DRB4 01:01, and DRB5 01:01 [88]. 15-mer peptides corresponding to the average length for MHC class II epitopes were used in the predictions. MHCII epitopes were filtered by 1%, that is, the top 1% of predicted binding affinity in the IEDB database that includes peptide-HLA class II affinities measured experimentally. Over 600,000 FC27-class peptides were found of which 207 were unique and predicted to be strong binders (not shown). No MHCII binding immunodominant alleles were predicted for the 3D7 class alleles from the cohort.

Patient antibodies recognize autologous and heterologous MSP2 variants

ELISA was used to evaluate the presence and extent of antibody reactivity against full and truncated, autologous and heterologous MSP2 variants. Antibody responses from 2012PT12, 2014003 and 2017001 declined to below 0.5 at around the third dilution, 1:4000 (Figure 9A, B, D). Antibody responses from patient 2015004 were still present at the last dilution (Figure 9C). All patient samples showed autologous reactivity as expected. Hyperimmune plasma induced a strong response from all four patient samples as expected (Figure E). In general, hyperimmune plasma responded significantly more strongly to Poly-MSP2 antigens than Full-MSP2 antigens. Unexposed plasma showed negligible response as expected. The following observations are for dilution 1 (1:1000), since this showed the highest absorbance and most reliable representation of antibody level.


A)
2012PT12 sera
Absorbance 450nm
1.5
1.0
0.5
0.0
1:1000
1:2000
1:4000
1:8000
1:16000
1:32000
Sera dilution
B)
2014003 sera
Absorbance 450nm
2.5
2.0
1.5
1.0
0.5
0.0
1:1000
1:2000
1:4000
1:8000
1:16000
1:32000
Sera dilution
C)
2015004 sera
Absorbance 450nm
4
3
2
1
0
1:1000
1:2000
1:4000
1:8000
1:16000
1:32000
Sera dilution
D)
2017001 sera
Absorbance 450nm
2.0
1.5
1.0
0.5
0.0
1:1000
1:2000
1:4000
1:8000
1:16000
1:32000
Sera dilution
E)
Controls
Absorbance 450nm
4
3
2
0.5
0.0
hyperimmune
unexposed
1:1000 sera dilution
2012PT12 Poly
2012PT12 Full
2014003 Poly
2014003 Full
2015004 Poly
2015004 Full
2017001 Poly
Full-P4-MSP2

F)

Symbol	Sera	Antigen 1	Antigen 2	Significance
○	2012PT12	Poly-2012PT12	Full-2012PT12	****
		Poly-2014003	Full-2014003	****
		Poly-2015004	Full-2015004	**
		Poly-2017001	Full-2017001	ns
□	2014003	Poly-2012PT12	Full-2012PT12	****
		Poly-2014003	Full-2014003	****
		Poly-2015004	Full-2015004	***
		Poly-2017001	Full-2012PT12	ns
△	2015004	Poly-2012PT12	Full-2012PT12	****
		Poly-2014003	Full-2014003	ns
		Poly-2015004	Full-2015004	***
		Poly-2017001	Full-2017001	ns
▽	2017001	Poly-2012PT12	Full-2012PT12	**
		Poly-2014003	Full-2014003	****
		Poly-2015004	Full-2015004	****
		Poly-2017001	Full-2017001	ns

Figure 9: A-D) Autologous and heterologous binding of previously unexposed patient plasma to different variants of MSP2. Measured by absorbance at 450nm. E) Control plasma and legend F) Statistical significances as calculated by two-way ANOVA with multiple comparisons. $p > 0.0001$. Significant differences are shown only for the first dilutions.

2012PT12 plasma had higher reactivity to Poly-2014003MSP2 and Full-2012PT12MSP2, but lower reactivity to Full-2015004MSP2 (Figure 9A). 2014003 plasma showed higher reactivity to Full-2014003MSP2 and Full-2012PT12MSP2 (Figure 9B). Both of these differences were statistically significant (Figure 9F). Patient 2015004 had a stronger response to Full-2015004MSP2 at the first dilution (Figure 9C) but the opposite at dilution 2 (significance not shown). Full-2012PT12MSP2 also induced a stronger response than the Poly-variant. Patient 2017001 had a stronger antibody response towards Poly-2015004 (Figure 9D). There was no significant difference between Poly- and Full-2017001-MSP2 for any patient plasma, including patient 2017001. While plasma from patients 2017001 and 2015004 showed higher reactivity to Poly-2017001-MSP2 than the Full protein, 2012PT12 and 2014003 plasma showed no significant difference in rectivity to Full- or Var- 2017001-MSP2 (Figure 9A-B). A second ELISA set was run with highly similar results, which we have not shown here.

Discussion

Parasite isolates across sub-Saharan Africa are broadly FC27-type sequences

Malaria genetic diversity is an essential factor for developing interventions and understanding the changing endemicity of varied areas. Diversity of potential vaccine candidates is especially necessary to guide vaccine design. Malaria endemic regions such as sub-Saharan Africa show extensive diversity of *P. falciparum* infection, especially for *msp2*. The relationship between FC27 and 3D7 alleles within populations vary widely around the globe as well as between smaller regions—studies have identified populations with higher proportions of FC27 than 3D7 alleles [89], and vice versa [39,90,91].

Our Malaria Immunology Cohort was an excellent example of cross-African *msp2* diversity. While FC27 family sequences revealed 29 distinct size variants and 209 nucleotide

sequences passing the threshold, 3D7 sequences were divided into 65 distinct size variants and there were 184 sequences passing the threshold. Thus our cohort, made up of sequences originating broadly from sub-Saharan Africa, showed a higher proportion of FC27 alleles than 3D7 alleles.

3D7 MSP2 within this cohort showed more extensive polymorphism than FC27 MSP2. The two most common size variants within the FC27 allele proteins were 292bp, matching the synthetic HB3 control, and 376bp. The two most common size variants for 3D7 proteins were 479bp and 485bp, neither of which matched the 3D7 reference strain MSP2 at 473bp long. Clustering revealed that most samples were infected by multiple MSP2 clones. This should be further analyzed for clonality of infection (COI), or number of clones per individual, which is commonly used as a parasite diversity marker and can also be used to estimate exposure [92,93].

FC27 and 3D7 variants had similar levels of sequence conservation

Distinct blocks of conserved sequences between most of the cohort were found using a Valdar-scored, normalized conservation analysis. FC27 and 3D7 had similar conservation values, with 250 residues showing conservation scores ≥ 0.5 for FC27 and 259 residues showing conservation scores ≥ 0.5 for 3D7. The highest conservation was seen for terminal ends of the MSP2 proteins, as expected. This matches with our protein alignments (Figure 3) before expression and cloning, where terminal regions are the most highly conserved. Within the central variable region, highly conserved regions were found around alignment positions 250-490, 510-530, and 1800-1900 for FC27 and around positions 900, 1200, 1400-1500, 2400, 2600-2800, and 2700 for 3D7 (Figure 7). We also observed that 3D7 proteins contained more sections with low consensus (alignment positions 20-750, 3300-3700). Also, there were comparatively smaller sections of high conservation (Figure 7B). We postulate that this could be resulting from the broad geographical locations of our cohort. In other words, we hypothesize that parasite isolates from a smaller region would share more conserved sequences for both FC27 and 3D7. However, since 3D7 is more polymorphic than FC27, there may also be no obvious changes when comparing to a smaller region; This could be futher analyzed within our cohort in the future. Conserved blocks within the central variable region should be compared to known monoclonal antibody epitopes—since high conservation was observed throughout this highly diverse cohort, they are promising for cross-protective vaccine development.

Phylogeny by MSP2 did not show a distinct geographical structure

A previous study using parasite isolate SNPs for genotyping determined that sub-Saharan *P. falciparum* cluster genetically into western, central and eastern subgroups, plus a distinct Ethiopian population [94]. *P. falciparum* jumped into humans around 50,000 years ago before major migrations, according to the most recent estimates [95]. Donation of genome sections from central Africa to the west and east correlate with this origin [94]. We constructed a phylogenetic tree of genetic relatedness among the *P. falciparum* isolates, using nucleotide sequences which we split into their respective allelic families.

Although some samples from nearby countries grouped together within the tree, the isolate phylogeny did not show a distinct geographical structure. While 3D7 is highly polymorphic--this may be the reason its predicted phylogeny shows little to no relatedness within-region--this is contrary to what we expected to see for FC27: we expected to see that FC27 MSP2 in parasite isolates transmitted closer to each other would share more similar *msp2* genes. One explanation of this, at least for FC27 MSP2, is the movement of human populations mixing clones within populations and therefore contributing to the recombination of clones. An alternate explanation of this is that *msp2* as a full gene is too polymorphic to be used for wide-scale phylogenetic analysis. Genome-wide single nucleotide polymorphisms, as used in the study clustering isolates into regional groups, are the better option for such an analysis. The

observed genetic diversity and lack of phylogenetic structure may also support theories that MSP2 is "selection neutral", so that polymorphism has no phenotypic consequences [93].

Isolate MSP2 contain predicted T-cell binding sites

Over 200 strong-binding (top 1%) MHCII epitopes were predicted using IEDB and the HLA class II alleles that are most frequent across the global population. This means patients in the cohort were infected with FC27-strain parasites, where their MSP2 have epitopes that could be bound by T-cells with the selected immunodominant HLA alleles. While promising information, the patients would have to be analyzed to see if they in fact produce these T-cells able to bind the predicted sites. We cannot assume that the predicted T-cell epitopes here are functional in real life yet, but it would be an interesting research idea for the future. No MHCII-binding of immunodominant alleles were found for 3D7 class MSP2 from the cohort. This does not mean there are no epitopes present on these MSP2, rather that there are no epitopes that are predicted to be bound by the selected immunodominant HLA alleles. With more time and a more extensive search, functional T-cell epitopes could be found. This would allow us to guide vaccine design towards MSP2 epitopes recognized by human populations.

Patient antibodies recognize additional variants to their own

While antibody responses from samples 2012PT12, 2014003, and 2017001 declined around dilutions 1:4000 or 1:8000, sample 2015004 still showed reactivity even at 1:32000. This suggests an overall higher antibody response in patient 2015004 than the others.

Antibodies from patient 2012PT12 showed minimal reaction to any antigen, including Var and Full 2012PT12-MSP2. This patient developed parasitemia over 5% but no severe malaria, and showed multiple unique clones by size variant genotyping (Figure 6). It is likely that the patient may have either a stronger antibody response to another variant, or an antibody response spread between many clones. The latter explanation suggests that the level of parasitemia overwhelms the immune response, backed by low reactivity of their antibodies. Since this individual traveled to Kenya, a higher transmission area, we may also consider transmission level as a factor for case severity. Looking at the overall low absorbance for this plate, we should also consider the idea that antibody levels were too low at this time point to represent reactivity very well. Patient 2014003 developed severe malaria without hyperparasitemia, but still maintained an antibody response against their own MSP2 variant. A strong immune response during severe disease would be responsible for the reactivity maintained at 1 month post-treatment. Their antibodies also reacted to 2012PT12-MSP2 at the same level that patient 2012PT12 did, which may suggest the epitope bound on 2012PT12-MSP2 is less immunogenic than that on 2014003-MSP2.

Overall, patients displayed higher reactivity to their own recombinant MSP2 proteins, as expected. Plasma from 2015004 also recognized the 2017001 MSP2 antigens, although at a lower level. Both of these proteins are 3D7 family, so the level of cross-reactivity is most likely towards the variable region. Natural infection does induce protective, anti-MSP2 antibodies towards epitopes within the central variable region [40]. However, plasma from 2017001 did not cross-recognize the 2015004-MSP2. Previous vaccine studies have shown responses against both central repeats and the conserved termini [7], and mouse studies have shown antibodies elicited against the terminal regions[17,41,42]. Here we see significantly higher responses against full MSP2 proteins compared with the MSP2 polymorphic region only, for MSP2 variants from a different allelic family. We also see significantly higher responses against polymorphic region only, for MSP2 variants within the same allelic family.

Taken together, our data shows that, while full length proteins can induce higher antibody responses than truncated proteins, epitopes from the variable region are sufficient to induce antibody responses. An important factor to note is that in their natural conformation many epitopes may not be accessible to immune recognition [40,43]. In the absence of membrane-

anchoring and additional protein-protein interactions, monomeric MSP2 in solution may not exist in the same conformation as on the parasite surface [96] . In fact, previous vaccine trials have shown MSP2 induces weak antibody responses [97] which could mean the most protective epitopes have not yet been targeted. Even though we have seen antibody reactivity and some cross-reactivity between patients, the recombinant MSP2 proteins used in this study may not be the same as their natural forms on the parasite surface. This needs to be considered for future research into antigenicity and cross-reactivity.

Previous study from our lab evaluated antibody responses towards parasite extract and a panel of recombinant antigens over 1 year in the same cohort [98]. The study showed higher antibody levels, breadth and longevity to the panel of antigens in previously exposed patients, and that antibody response declines rapidly without re-exposure. The greatest difference between primary and exposed patients was for MSP2, which was the most polymorphic [98]. We argue that information gained in the present thesis can be used, along with antibody responses of previously exposed patients to parasite extract, to guide vaccine development using MSP2 variants. Along with optimization of adjuvant, dose and vaccine regimen [98], it should be possible to combine immunogenic epitopes from multiple variants to induce a robust antibody response in naïve or primary exposed individuals.

Conclusions

With this thesis, we wanted to analyze the sequence diversity of MSP2 present in patients infected by *P. falciparum* in sub-Saharan Africa. Results from decades of study have indicated that the immune response can control malaria infection and even prevent it, but translating this knowledge into vaccines is difficult because of the parasite's complexity [99]. In addition, MSP2's antigenic polymorphism is a significant challenge to developing an anti-merozoite stage vaccine [9]. An MSP2-based vaccine should induce significantly high antibody responses [100,101] protect against many variants prevalent in local or global populations, and include both allelic families in order to be significantly effective [56]. Furthermore, combination with other antigens, such as AMA-1 or MSP1, could produce a broader and more protective effect than MSP2 alone [44,55]. Our Malaria Immunology Cohort gives an overview of isolates from across sSA. Consensus strongly suggests that conserved sequences can be found between FC27 and 3D7 alleles, within both the terminal and variable domains, that have potential to induce cross-protective antibody responses against *P. falciparum* malaria.

Although this study has been informative, we are interested in patient immune recognition of MSP2 variants at the single-cell level. We plan to study MBC responses to autologous and heterologous MSP2, to evaluate long-lived memory responses towards the same antigens. To this end, we plan on performing a reversed B-cell FluoroSpot assay on peripheral blood mononuclear cells (PBMS) from additional previously exposed and primary infected individuals.

Acknowledgements

I would like to thank first of all Anna Färnert for the privilege of working in her lab group, and my supervisor David Plaza for teaching me all I have learned about bioinformatics and data analysis. Ioanna Broumou and David Plaza performed the circular consensus sequencing which I used to start this project. Peter Jahnmatz helped immensely with selecting *msp2* sequences to clone, as well as being our contact at MabTech for cloning and expression of our antigens. I would also like to thank my friends and family who, whether through support or through reviewing this manuscript, contributed selflessly to the completion of my thesis.

References

1. The ‘World malaria report 2019’ at a glance. https://www.who.int/news-room/feature-stories/detail/world-malaria-report-2019.
2. Greenwood, B. & Mutabingwa, T. Malaria in 2002. *Nature* **415**, 670–672 (2002).
3. Gardner, M. J. *et al.* Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature* **419**, 498–511 (2002).
4. Malaria. https://www.who.int/news-room/fact-sheets/detail/malaria.
5. Volkman, S. K. *et al.* A genome-wide map of diversity in Plasmodium falciparum. *Nat. Genet.* **39**, 113–119 (2007).
6. Venugopal, K., Hentzschel, F., Valkiūnas, G. & Marti, M. Plasmodium asexual growth and sexual development in the haematopoietic niche of the host. *Nature Reviews Microbiology* vol. 18 177–189 (2020).
7. Prevention, C.-C. for D. C. and. CDC - Malaria - Malaria Worldwide - Impact of Malaria. (2021).
8. Gazzinelli, R. T., Kalantari, P., Fitzgerald, K. A. & Golenbock, D. T. Innate sensing of malaria parasites. *Nature Reviews Immunology* vol. 14 744–757 (2014).
9. Duffy, P. E. & Patrick Gorres, J. Malaria vaccines since 2000: progress, priorities, products. *npj Vaccines* vol. 5 1–9 (2020).
10. Webster, H. K. *et al.* Development of immunity in natural Plasmodium falciparum malaria: Antibodies to the falciparum sporozoite vaccine 1 antigen (R32tet32). *J. Clin. Microbiol.* **25**, 1002–1008 (1987).
11. Efficacy and safety of RTS,S/AS01 malaria vaccine with or without a booster dose in infants and children in Africa: Final results of a phase 3, individually randomised, controlled trial. *Lancet* **386**, 31–45 (2015).
12. Neafsey, D. E. *et al.* Genetic Diversity and Protective Efficacy of the RTS,S/AS01 Malaria Vaccine. *N. Engl. J. Med.* **373**, 2025–2037 (2015).
13. Datoo, M. S. *et al.* Efficacy of a low-dose candidate malaria vaccine, R21 in adjuvant Matrix-M, with seasonal administration to children in Burkina Faso: a randomised controlled trial. (2021) doi:10.1016/S0140-6736(21)00943-0.
14. Chaplin, D. D. Overview of the immune response. *J. Allergy Clin. Immunol.* **125**, S3 (2010).
15. Tangye, S. G. & Tarlinton, D. M. Memory B cells: Effectors of long-lived immune responses. *Eur. J. Immunol.* **39**, 2065–2075 (2009).
16. Lanzavecchia, A. *et al. Understanding and making use of human memory B cells*.
17. Day, K. P. & Marsh, K. *Naturally acquired immunity to Plasmodium falciparum*.
18. Aucan, C. *et al.* High immunoglobulin G2 (IgG2) and low IgG4 levels are associated with human resistance to Plasmodium falciparum malaria. *Infect. Immun.* **68**, 1252–1258 (2000).
19. Anderson, T. J. C. *et al.* Microsatellite Markers Reveal a Spectrum of Population Structures in the Malaria Parasite Plasmodium falciparum. *Mol. Biol. Evol.* **17**, 1467–1482 (2000).
20. Liehl, P. *et al.* Host-cell sensors for Plasmodium activate innate immunity against liver-stage infection. *Nat. Med.* **20**, 47–53 (2014).
21. Miller, J. L., Sack, B. K., Baldwin, M., Vaughan, A. M. & Kappe, S. H. I. Interferon-Mediated Innate Immune Responses against Malaria Parasite Liver Stages. *Cell Rep.* **7**, 436–447 (2014).
22. Hoffman, S. L. *et al.* Naturally acquired antibodies to sporozoites do not prevent malaria: Vaccine development implications. *Science (80-.).* **237**, 639–642 (1987).
23. Stanisie, D. I. *et al.* Immunoglobulin G subclass-specific responses against

Plasmodium falciparum merozoite antigens are associated with control of parasitemia and protection from symptomatic illness. *Infection and Immunity* vol. 77 1165–1174 (2009).
24. Mitchell, G. H., Thomas, A. W., Margos, G., Dluzewski, A. R. & Bannister, L. H. Apical Membrane Antigen 1, a Major Malaria Vaccine Candidate, Mediates the Close Attachment of Invasive Merozoites to Host Red Blood Cells. *Infect. Immun.* **72**, 154–158 (2004).
25. Greenhouse, B. *et al.* Antibodies to Plasmodium falciparum antigens predict a higher risk of malaria but protection from symptoms once parasitemic. *J. Infect. Dis.* **204**, 19–26 (2011).
26. Engvall, E. & Perlmann, P. Enzyme-linked immunosorbent assay (ELISA) quantitative assay of immunoglobulin G. *Immunochemistry* **8**, 871–874 (1971).
27. Engvall, E., Jonsson, K. & Perlmann, P. Enzyme-linked immunosorbent assay. II. Quantitative assay of protein antigen, immunoglobulin g, by means of enzyme-labelled antigen and antibody-coated tubes. *BBA - Protein Struct.* **251**, 427–434 (1971).
28. Kim, J., Jang, J. W., Kim, J. Y., Oh, D. J. & Lim, C. S. Combined Use of Malaria Antigen and Antibody Enzyme-Linked Immunosorbent Assay for Blood Screening of <i>Plasmodium vivax </i>in the Republic of Korea. *Med. Princ. Pract.* **25**, 212–218 (2016).
29. Spencer, H. C., Collins, W. E., Chin, W. & Skinner, J. C. The enzyme linked immunosorbent assay (ELISA) for malaria. I. The use of in vitro-cultured Plasmodium falciparum as antigen. *Am. J. Trop. Med. Hyg.* **28**, 927–932 (1979).
30. Chiodini, P. L. *et al.* Evaluation of a Malaria Antibody ELISA and Its Value in Reducing Potential Wastage of Red Cell Donations from Blood Donors Exposed to Malaria, with a Note on a Case of Transfusion-Transmitted Malaria. *Vox Sang.* **73**, 143–148 (1997).
31. Noedl, H. *et al.* Sensitivity and specificity of an antigen detection ELISA for malaria diagnosis. *Am. J. Trop. Med. Hyg.* **75**, 1205–1208 (2006).
32. Gonzales, S. J. *et al.* Naturally Acquired Humoral Immunity Against Plasmodium falciparum Malaria. *Frontiers in Immunology* vol. 11 2809 (2020).
33. Fenton, B. *et al. Structural and Antigenic Polymorphism of the 35-to 48-Kilodalton Merozoite Surface Antigen (MSA-2) of the Malaria Parasite Plasmodium falciparum*. *MOLECULAR AND CELLULAR BIOLOGY* vol. 11 http://mcb.asm.org/ (1991).
34. Smythe, J. A. *et al.* Structural diversity in the Plasmodium falciparum merozoite surface antigen 2. *Proc. Natl. Acad. Sci.* **88**, (1991).
35. Putaporntip, C., Jongwutiwes, S. & Hughes, A. L. Differential selective pressures on the merozoite surface protein 2 locus of Plasmodium falciparum in a low endemic area. *Gene* **427**, 51–57 (2008).
36. Smythe, J. A. *et al.* Structural diversity in the Plasmodium falciparum merozoite surface antigen 2. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 1751–1755 (1991).
37. Fenton, B., Clark, J. T., Wilson, C. F., McBride, J. S. & Walliker, D. Polymorphism of a 35-48 kDa Plasmodium falciparum merozoite surface antigen. *Mol. Biochem. Parasitol.* **34**, 79–86 (1989).
38. Akpogheneta, O. J. *et al.* Duration of naturally acquired antibody responses to blood-stage Plasmodium falciparum is age dependent and antigen specific. *Infect. Immun.* **76**, 1748–1755 (2008).
39. Mohammed, H., Kassa, M., Assefa, A., Tadesse, M. & Kebede, A. Genetic polymorphism of Merozoite Surface Protein-2 (MSP-2) in Plasmodium falciparum isolates from Pawe District, North West Ethiopia. *PLoS One* **12**, e0177559 (2017).
40. Adda, C. G. *et al.* Antigenic characterization of an intrinsically unstructured protein,

Plasmodium falciparum merozoite surface protein 2. *Infect. Immun.* **80**, 4177–4185 (2012).
41. Saul, A., Lord, R., Jones, G. L. & Spencer, L. Protective immunization with invariant peptides of the Plasmodium falciparum antigen MSA2. *J. Immunol.* **148**, (1992).
42. Lawrence, N., Stowers, A., Mann, V., Taylor, D. & Saul, A. Recombinant chimeric proteins generated from conserved regions of Plasmodium falciparum merozoite surface protein 2 generate antiparasite humoral responses in mice. *Parasite Immunol.* **22**, 211–221 (2000).
43. MacRaild, C. A. *et al.* Conformational Dynamics and Antigenicity in the Disordered Malaria Antigen Merozoite Surface Protein 2. *PLoS One* **10**, e0119899 (2015).
44. Polley, S. D. *et al.* High levels of serum antibodies to merozoite surface protein 2 of Plasmodium falciparum are associated with reduced risk of clinical malaria in coastal Kenya. *Vaccine* **24**, 4233–4246 (2006).
45. Osier, F. H. A. *et al.* Breadth and magnitude of antibody responses to multiple Plasmodium falciparum merozoite antigens are associated with protection from clinical malaria. *Infect. Immun.* **76**, 2240–2248 (2008).
46. Hill, D. L. *et al.* Merozoite antigens of plasmodium falciparum elicit strain-transcending opsonizing immunity. *Infect. Immun.* **84**, 2175–2184 (2016).
47. Osier, F. H. A. *et al.* Opsonic phagocytosis of Plasmodium falciparum merozoites: Mechanism in human immunity and a correlate of protection against malaria. *BMC Med.* **12**, (2014).
48. Smythe, J. A. *et al.* Structural diversity in the Plasmodium falciparum merozoite surface antigen 2. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 1751–1755 (1991).
49. Osier, F. H. *et al.* Malaria: New antigens for a multicomponent blood-stage malaria vaccine. *Sci. Transl. Med.* **6**, 247ra102 (2014).
50. Osier, F. H. A. *et al.* Breadth and magnitude of antibody responses to multiple Plasmodium falciparum merozoite antigens are associated with protection from clinical malaria. *Infect. Immun.* **76**, 2240–2248 (2008).
51. Osier, F. H. A. *et al.* Allele-specific antibodies to Plasmodium falciparum merozoite surface protein-2 and protection against clinical malaria. *Parasite Immunol.* **32**, 193–201 (2010).
52. Stubbs, J. *et al.* Strain-Transcending Fc-Dependent Killing of Plasmodium falciparum by Merozoite Surface Protein 2 Allele-Specific Human Antibodies †. *Infect. Immun.* **79**, 1143–1152 (2011).
53. Scopel, K. K. G., Da Silva-Nunes, M., Malafronte, R. S., Braga, É. M. & Ferreira, M. U. Variant-specific antibodies to merozoite surface protein 2 and clinical expression of Plasmodium falciparum malaria in rural amazonians. *Am. J. Trop. Med. Hyg.* **76**, 1084–1091 (2007).
54. Tonon, A. P. *et al.* Plasmodium falciparum: Sequence diversity and antibody recognition of the Merozoite surface protein-2 (MSP-2) in Brazilian Amazonia. *Exp. Parasitol.* **108**, 114–125 (2004).
55. Franks, S. *et al.* Genetic diversity and antigenic polymorphism in Plasmodium falciparum: Extensive serological cross-reactivity between allelic variants of merozoite surface protein 2. *Infect. Immun.* **71**, 3485–3495 (2003).
56. McCarthy, J. S. *et al.* A phase 1 trial of MSP2-C1, a blood-stage malaria vaccine containing 2 isoforms of MSP2 formulated with montanide® ISA 720. *PLoS One* **6**, (2011).
57. Franks, S. *et al.* Genetic Diversity and Antigenic Polymorphism in Plasmodium falciparum: Extensive Serological Cross-Reactivity between Allelic Variants of Merozoite Surface Protein 2. *Infect. Immun.* **71**, 3485–3495 (2003).

58. Sundling, C. *et al.* B cell profiling in malaria reveals expansion and remodeling of CD11c+ B cell subsets. *JCI Insight* **4**, (2019).
59. Wångdahl, A. *et al.* Severity of Plasmodium falciparum and Non-falciparum Malaria in Travelers and Migrants: A Nationwide Observational Study over 2 Decades in Sweden. *J. Infect. Dis.* **220**, 1335–1345 (2019).
60. Liljander, A. *et al.* Optimization and validation of multi-coloured capillary electrophoresis for genotyping of Plasmodium falciparum merozoite surface proteins (msp1 and 2). *Malar. J.* **8**, 78 (2009).
61. Snounou, G. Genotyping of Plasmodium spp. Nested PCR. *Methods Mol. Med.* **72**, 103–116 (2002).
62. Snounou, G. *et al.* Biased distribution of msp1 and msp2 allelic variants in Plasmodium falciparum populations in Thailand. *Trans. R. Soc. Trop. Med. Hyg.* **93**, 369–374 (1999).
63. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research* vol. 25 3389–3402 (1997).
64. Camacho, C. *et al.* BLAST+: Architecture and applications. *BMC Bioinformatics* **10**, (2009).
65. Cock, P. J. A., Chilton, J. M., Grüning, B., Johnson, J. E. & Soranzo, N. NCBI BLAST+ integrated into Galaxy. *Gigascience* **4**, (2015).
66. Bahl, A. *et al.* PlasmoDB: The Plasmodium genome resource. A database integrating experimental and computational data. *Nucleic Acids Research* vol. 31 212–215 (2003).
67. Fraunholz, M. J. & Roos, D. S. PlasmoDB: Exploring genomics and post-genomics data of the malaria parasite, Plasmodium falciparum. *Redox Rep.* **8**, 317–320 (2003).
68. Aurrecoechea, C. *et al.* PlasmoDB: A functional genomic database for malaria parasites. *Nucleic Acids Res.* **37**, (2009).
69. Afgan, E. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res.* **44**, W3–W10 (2016).
70. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
71. Gupta, R., Jung, E. & Brunak, S. N-linked glycosylation sites in human proteins. *Prep.* (2004).
72. Gupta, R. & Brunak, S. Prediction of glycosylation across the human proteome and the correlation to protein function - PubMed. *Pacific Symp. Biocomput.* 310–22 (2003).
73. Berhe, S., Gerold, P., Kedees, M. H., Holder, A. A. & Schwarz, R. T. Plasmodium falciparum: Merozoite surface proteins 1 and 2 are not posttranslationally modified by classical N- or O-glycans. *Exp. Parasitol.* **94**, 194–197 (2000).
74. Almagro Armenteros, J. J. *et al.* SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* **37**, 420–423 (2019).
75. Eisenhaber, B., Bork, P. & Eisenhaber, F. Prediction of potential GPI-modification sites in proprotein sequences. *J. Mol. Biol.* **292**, 741–758 (1999).
76. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
77. MacGowan, S. A. *et al.* The Dundee Resource for Sequence Analysis and Structure Prediction. *Protein Sci.* **29**, 277–297 (2020).
78. Troshin, P. V., Procter, J. B. & Barton, G. J. Java bioinformatics analysis web services for multiple sequence alignment-JABAWS:MSA. *Bioinformatics* **27**, 2001–2002 (2011).
79. Golicz, A. *et al.* AACon: A Fast Amino Acid Conservation Calculation Service. (2018).

80. Valdar, W. S. J. Scoring residue conservation. *Proteins Struct. Funct. Genet.* **48**, 227–241 (2002).
81. Nguyen, L. T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
82. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A. & Jermiin, L. S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
83. Minh, B. Q., Nguyen, M. A. T. & Von Haeseler, A. Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* **30**, 1188–1195 (2013).
84. Krief, S. *et al.* On the Diversity of Malaria Parasites in African Apes and the Origin of Plasmodium falciparum from Bonobos. *PLoS Pathog.* **6**, e1000765 (2010).
85. Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).
86. Gonzalez-Galarza, F. F. *et al.* Allele frequency net database (AFND) 2020 update: Gold-standard data classification, open access genotype data and new query tools. *Nucleic Acids Res.* **48**, D783–D788 (2020).
87. Dhanda, S. K. *et al.* ImmunomeBrowser: A tool to aggregate and visualize complex and heterogeneous epitopes in reference proteins. *Bioinformatics* **34**, 3931–3933 (2018).
88. Paul, S. *et al.* Development and validation of a broad scheme for prediction of HLA class II restricted T cell epitopes. *J. Immunol. Methods* **422**, 28–34 (2015).
89. Zhang, C. L., Zhou, H. N., Liu, Q. & Yang, Y. M. Genetic polymorphism of merozoite surface proteins 1 and 2 of Plasmodium falciparum in the China-Myanmar border region. *Malar. J.* **18**, 367 (2019).
90. Botwe, A. K. *et al.* Dynamics in multiplicity of Plasmodium falciparum infection among children with asymptomatic malaria in central Ghana. *BMC Genet.* **18**, 1–9 (2017).
91. Kidima, W. & Nkwengulila, G. Plasmodium falciparum msp2 genotypes and multiplicity of infections among children under five years with uncomplicated malaria in Kibaha, Tanzania. *J. Parasitol. Res.* **2015**, (2015).
92. Duah, N. O., Matrevi, S. A., Quashie, N. B., Abuaku, B. & Koram, K. A. Genetic diversity of Plasmodium falciparum isolates from uncomplicated malaria cases in Ghana over a decade. *Parasit. Vectors* **9**, 416 (2016).
93. Zhong, D., Koepfli, C., Cui, L. & Yan, G. Molecular approaches to determine the multiplicity of Plasmodium infections. *Malaria Journal* vol. 17 172 (2018).
94. Amambua-Ngwa, A. *et al.* Major subpopulations of Plasmodium falciparum in sub-Saharan Africa. *Science (80-.).* **365**, 813–816 (2019).
95. Otto, T. D. *et al.* Genomes of all known members of a Plasmodium subgenus reveal paths to virulent human malaria. *Nat. Microbiol.* **3**, 687–697 (2018).
96. Adda, C. G. *et al.* Plasmodium falciparum merozoite surface protein 2 is unstructured and forms amyloid-like fibrils. *Mol. Biochem. Parasitol.* **166**, 159–171 (2009).
97. Saul, A. *et al.* Human phase I vaccine trials of 3 recombinant asexual stage malaria antigens with Montanide ISA720 adjuvant. *Vaccine* **17**, 3145–3159 (1999).
98. Yman, V. *et al.* Antibody responses to merozoite antigens after natural Plasmodium falciparum infection: Kinetics and longevity in absence of re-exposure. *BMC Med.* **17**, 1–14 (2019).
99. Long, C. A. & Zavala, F. Immune Responses in Malaria. *Cold Spring Harbor perspectives in medicine* vol. 7 (2017).
100. Murungi, L. M. *et al.* A threshold concentration of anti-merozoite antibodies is

required for protection from clinical episodes of malaria. *Vaccine* **31**, 3936–3942 (2013).
101. Stanisic, D. I. *et al.* Acquisition of antibodies against Plasmodium falciparum merozoites and malaria immunity in young children and the influence of age, force of infection, and magnitude of response. *Infect. Immun.* **83**, 646–660 (2015).

Additional Bioinformatics tools:

1. Blankenberg, D. *et al.* Manipulation of FASTQ data with galaxy. *Bioinformatics* **26**, 1783–1785 (2010).
2. Johnson, J. E. *et al.* Improve your Galaxy text life: The Query Tabular Tool. *F1000Research* **7**, (2019).
3. Bjoern A. Gruening (2014), Galaxy wrapper
4. Camacho, C., Coulouris, G., Avagyan, V., Ning, Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST\mathplus: architecture and applications. *BMC Bioinformatics*, *10*(1), 421. https://doi.org/10.1186/1471-2105-10-421
5. Cock, P. J. A., Chilton, J. M., Grüning, B., Johnson, J. E., & Soranzo, N. (2015). NCBI BLAST\mathplus integrated into Galaxy. *GigaScience*, *4*(1). https://doi.org/10.1186/s13742-015-0080-7

Supplementary Data

Countries of travel or origin for each patient in the cohort were mapped, shown in Figure S1. Although parasite isolates covered the majority of sub-Saharan Africa, the countries with the highest number of isolates were Kenya with 7 samples, Ghana, Tanzania and Nigeria with 6 samples each.

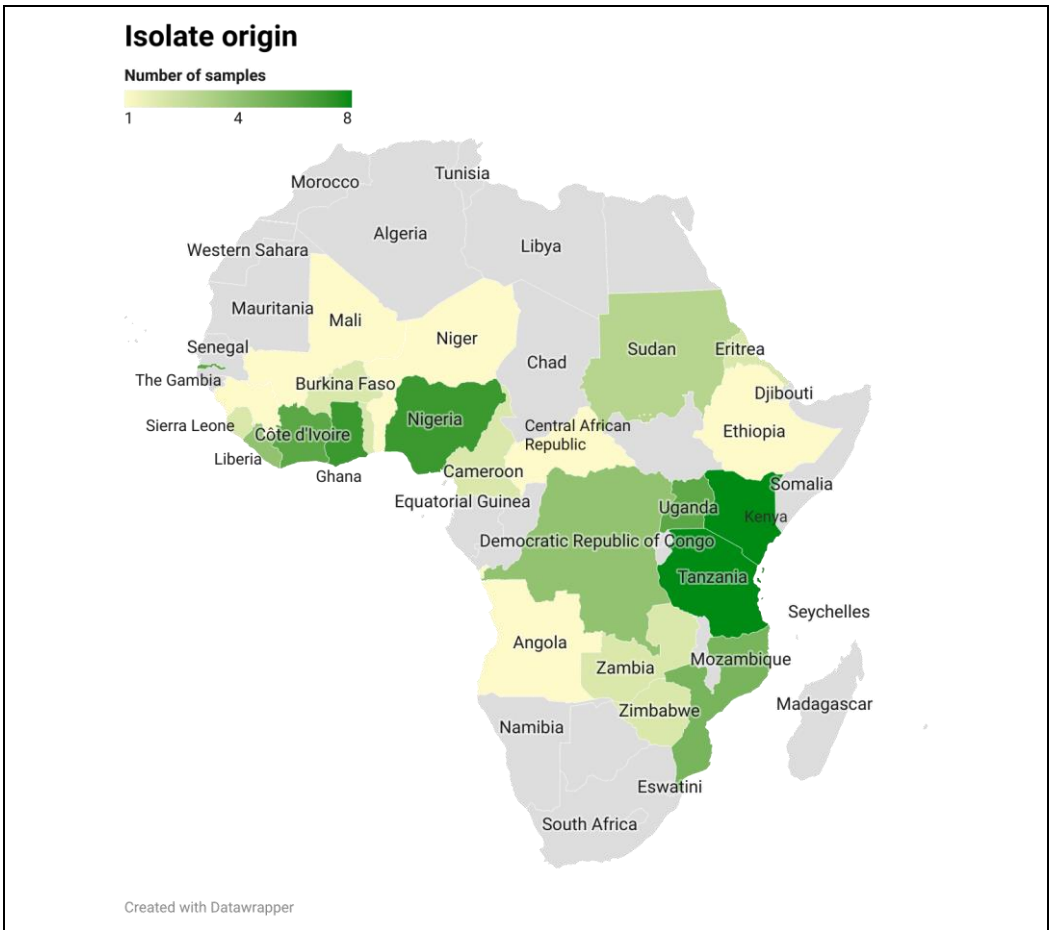


Figure S1: Map of African with number of patient isolates originating from each country. Countries with the highest number of samples were Kenya (7), Ghana (6), Tanzania (6), and Nigeria (6). One sample isolate originated in Bali, Indonesia (not shown).