Genomic from *Plasmodium falciparum* field isolates from Benin allows the identification of PfEMP1 variants by LC-MS/MS at the patient's level

Claire Kamaliddin[a,b], David Rombaut[b,c], Emilie Guillochon[d], Jade Royo[e], Sem Ezinmegnon[a,f], Stéphanie Huguet [g,h], Sayeh Guemouri[a,b], Céline Peirera[a], Cédric Broussard[b,c], Jules M Alao[i], Agnès Aubouy[e], François Guillonneau[b,c], Philippe Deloron[a,b] and Gwladys Bertin[a,b]

[a]UMR 216 – MERIT, IRD, Paris Descartes University, Paris, France

[b]COMUE Sorbonne Paris Cité, Paris, France

[c]3p5 Proteomic Facility, Paris Descartes University, Paris, France

[d]Inovarion, Paris, France

[e]UMR 152 – PHARMADEV, IRD, Paul Sabatier Toulouse III University, Toulouse, France

[f]Centre pour la Recherche et l'Etude du paludisme associé à la grossesse et à l'enfance, Cotonou, Bénin

[g]Institute of Plant Sciences Paris-Saclay (IPS2), CNRS, INRA, Université Paris-Sud, Université d'Evry, Université Paris-Saclay, Gif sur Yvette, France

[h]Institute of Plant Sciences Paris-Saclay (IPS2), CNRS, INRA Université Paris-Diderot, Sorbonne Paris-Cité, Gif sur Yvette, France

[i]CHU-MEL, Pediatric department, Cotonou, Bénin

#Address correspondence to Gwladys I Bertin, gwladys.bertin@ird.fr

Abstract word count: 160 words

Text word count

Abstract

PfEMP1 are the major protein family from parasitic origin involves in the pathophysiology of severe malaria, and PfEMP1 domain subtypes are associated with the infection outcome. In addition, PfEMP1 variability in endless and current protein repository do not reflect the immense diversity of the sequences of PfEMP1 proteins. The aim of our study was to identify the different PfEMP1 variants expressed within a patient sample by mass spectrometry. We performed a proteogenomic approach to decipher at the patient's level PfEMP1 expression in different clinical settings: cerebral malaria, severe anemia and uncomplicated malaria. The combination of whole genome sequencing approach, RNAsequencing, and mass spectrometry proteomic analysis allowed to attribute PfEMP1 sequences to each sample and classify the relative expression level of PfEMP1 proteins within each sample. We predicted PfEMP1 structures using the newly identified protein sequences. We confirmed the involvement of DBLβ in malaria pathogenesis and observed that CIDRα domains linked to ICAM-1 binding DBLβ domains displayed EPCR binding structure.

Introduction

Plasmodium falciparum is one of the causal agents of malaria in humans and leads to the severe form of the disease. In endemic areas, children under 5 years old are victims of cerebral malaria (CM) and severe anemia (SA).Despite the introduction of intra venous artesunate therapy, severe malaria mortality remains significant - 0.5% of malaria outbreak according to World Health Organization (WHO) (1).

Through its development in human erythrocytes, *P. falciparum* grows inside the erythrocyte and the infected erythrocyte's (iE) surface changes in shape and rigidity. Parasite proteins exported at the host cell surface mediate infected erythrocyte's adhesion to the host's endothelium. This phenomenon, called cytoadhesion, is the main virulence factor of the parasite and contributes to the pathophysiology of severe malaria. The sequestration of infected erythrocytes in the capillaries leads to hypoxia, occlusion and endothelial activation. In CM pathophysiology, the sequestration of iE in the brain capillaries is believed to trigger coma and brain swelling (2).

Among the proteins exported at the erythrocyte's surface, three variant surface antigens (VSA) families have been described: the repetitive interspersed family (RIFIN) (3, 4), subtelomeric variable open reading frame (STEVOR) (5, 6) and *Plasmodium falciparum* Erythrocyte Membrane Protein 1 (PfEMP1) (7), which is the most described VSA family. They are encoded by the multigenic *var* gene family (8–10), consisting in ~ 60 copies per parasite genome. The diversity among *var* sequences is almost endless (11, 12) thus participating to the infected erythrocyte ability to evade the immune system.

PfEMP1 proteins are high molecular weight transmembrane proteins (200-350 kDa), and are composed of an intra-erythrocytic segment, which is conserved, and a highly variable extracellular segment (13). The extra-erythrocytic segment is composed of 4 to 9 alternated Duffy Binding Like (DBL) or Cystein Inter Domain Rich (CIDR) domains. The nature and the arrangement of these domains determine the binding phenotype of the infected erythrocyte (13, 14).

Among the PfEMP1 receptor in human endothelium, the most common is the broadly expressed in human cells CD36, but it is not related to any specific form of malaria. In the context of severe malaria, two human host receptors for PfEMP1 binding have been identified: the InterCellular Adhesion Molecule-1 receptor (ICAM-1) (15) and the Endothelial Protein C Receptor (EPCR) (16), both expressed by brain endothelial cells (17), and co-localized with the sequestered iEs (15). Specific domains from PfEMP1 are known to

interact with ICAM-1, including the Domain Cassette 4 (DC4) (DBLα1.4/1.1-CIDRα1.6-DBLβ3) (18). The binding domain is located in the C-terminal third of the DBLβ3 (18). A study conducted on lab-adapted patient's isolates after selecting on ICAM-1 highlighted that the residues involved in PfEMP1 binding to ICAM-1 are highly variable with a limited binding pattern (19) (20).

EPCR role in PfEMP1 binding has been more recently discovered (16). EPCR binding is mediated by highly variable but structurally conserved CIDRα1 PfEMP1 domains (more precisely CIDRα1.1 and CIDRα1.4-1.8) (21, 22). Importantly, the level of PfEMP1transcript associated with EPCR binding is higher in samples from patients suffering from severe malaria, and increases with the disease severity (21).

A dual-binding with EPCR and ICAM-1 has been suggested, since not all CM isolates present an increase in binding-EPCR PfEMP1 coding transcript (23). In addition, expression levels of *var* coding transcripts are increased in parasites able to bind to EPCR, ICAM-1 and CD36 *in vitro* enforcing the idea of a mutual binding (24). The expression of DBL involved in ICAM-1 binding is associated with dual ICAM-1 and EPCR binding (20).

Despite their high diversity, PfEMP1 proteins are a target of naturally acquired immunity against *P. falciparum* (25), and the parasite express VSA corresponding to gap in the antibodies repertoire from an individual (26, 27). Therefore, identification and characterization of severe malaria associated PfEMP1 proteins is crucial. It is believed that despite the endless variability of PfEMP1 proteins (28), a cross-reactivity between expressed epitopes exists (29), as children in endemic area acquire protective immunity to the high binding phenotypes after a limited number of infection compared to PfEMP1 variability (26). PfEMP1 proteins are a proposed target for vaccine to protect people against severe forms of malaria (25, 30).

Most field studies looking for *P. falciparum* binding phenotypes are based on molecular biology analysis, and have shown that transcript coding for specific PfEMP1 domains expression level are associated with disease outcome (21, 24, 31, 32). However, this strategy is currently limited to the already identified PfEMP1 domains and does not give proficiency of the expressed proteins.

To complement this deficiency, mass-spectrometry-based proteomics is a powerful and sensitive tool for bottom-up protein identification after trypsin digestion. The application of mass spectrometry for VSA identification, and more especially PfEMP1 identification remains challenging for the following reasons.

First, PfEMP1 are transmembrane proteins, and therefore present hydrophobic segments which decrease its solubility and its accessibility to proteases such as trypsin. Second, PfEMP1 are high molecular weight proteins (250-350 kDa), and finally, PfEMP1 are highly variable in sequences, yet database repository is usually simplified by eliminating redundancy. Thus, they do not reflect the natural sequence diversity which may occur in such a context.

To date, PfEMP1 identification with mass-spectrometry remains limited. In fact, the identification of the semi-conserved variant VAR2CSA in the context of Pregnancy Associated Malaria has been performed (33), but the identification of PfEMP1 variant in a context of severe malaria (CM or SA) has not been reported yet. The association of PfEMP1 domains with clinical outcome (UM, CM and Pregnancy Associated Malaria) was performed, with relaxed analysis parameters to allow the identification of PfEMP1 domains (34).

To identify PfEMP1 associated with *P. falciparum* clinical outcome in endemic settings, we used a "proteogenomic" strategy on field sample from children from Benin, West Africa, retrieved from May 2016 to August 2016. Specific PfEMP1 sequences from each isolate were reconstructed, using whole genome sequencing (WGS) data to enrich the protein database (Figure 1). We analyzed the whole proteome of samples from patients presenting either CM, SA or UM, and attribute PfEMP1 sequences within these samples. Corresponding samples were analyzed in RNAsequencing for PfEMP1 expression analysis, in relation to proteomic results.

We managed to identify the PfEMP1 sequence associated with 4 CM samples, 9 SA and 9 UM samples using mass spectrometry (MS). For two samples, the relative expression levels from the PfEMP1 identified in MS was assessed using RT-qPCR using the corresponding mRNA. We confirmed the expression of several PfEMP1 within a single field isolates and provided the first identification at the patient's level of PfEMP1 expressed by the parasite in the context of acute *P. falciparum* infection. We were also able to predict PfEMP1 structure within each newly identified sequence.

Results

Included Samples

We included 95 patients, covering 31 SA, 18 CM and 46 UM. Average patient's age was similar among all inclusion groups. Parasite density geometric mean was 8,055 p/µL for UM group, 34,191 p/µL for CM and 24,313 p/µL for SA. Parasite density was only significantly different between SA and UM samples (p = 0.0158 with Bonferroni's Multiple Comparison Test). Hemoglobin level was measured for 20 UM, 17 CM and 31 SA, respectively 11.28 [10.26; 12.75], 5.51 [4.10; 6.56] and 4.393 [3.90; 5.00] g/dL. Hemoglobin level was statistically different for UM samples (*vs.* CM and *vs.* SA) ($p < 0.05$), but no difference was observed between SA and CM samples. No difference in erythrocyte count (p = 0.1274) and temperature (p = 0.9125) was retrieved between SA and CM. All CM patients presented a coma (average BS 2 [2;2]), while SA patients did not (average BS 4.6 [4;5]) ($p < 0.0001$). 14/18 CM patients presented convulsions, and 7/38 SA patients ($p < 0.0001$).

For mass spectrometry analysis, we selected samples among those showing successful maturation. Four CM, 9 SA and 9 UM samples have been further investigated. RNAsequencing was successfully performed on 10 samples corresponding to the one studied in mass spectrometry (1 UM, 5 SA (3 patients) and 4 CM (2 patients)). (Supplementary Table 1)

Var genes transcripts identification with RNAsequencing

We then focused on a set of 120 proteins from PlasmoDB reference, corresponding either to PfEMP1, or to other proteins associated with knobs formation (Supplementary data Table 3). The presence of skeleton binding protein 1 (Pf3D7_0501300) and EMP1 (Pf3D7_0730900) was assessed in all samples except SA09. Focusing on PfEMP1, we observed that the conserved ATS coding region was the only site for PfEMP1 mRNA identification using reference sequences from PlasmoDB. We then performed a selective mapping of RNAsequencing reads using the sequences from *var* genes reconstruction, which allowed to identify the expressed *var* transcript for 3 samples (Supplementary Table 4).

Protein identification

Protein identification was performed using a home-made database (reference sequences from human and *P. falciparum* repositories, and the assembled *var* from field samples) containing 295,601 protein sequences,

among which 87,489 were *P. falciparum*-associated sequences, while the other referred to the human proteome. Overall, we identified 3,300 proteins. After applying contaminant depletion and identification criteria, 3 214 proteins remained. A total of 1,302 were associated to the human proteome, and 1,912 to *P. falciparum'*s. Among those later, 460 proteins were identified as *P. falciparum* membrane-associated proteins, including 60.4% of hypothetical or putative, 12 % of PfEMP1s, 3.5% of RIFINs, 0.9% of STEVORs, 1.5% of PHISTs and 21.7% belong to other protein families.

A total of 155 proteins associated with PfEMP1 were identified, among which 96 were unique proteins, and clustered in 55 protein groups. Only 10 of the identified PfEMP1 (as part of protein group, or majority protein) were known sequences from public database repository (Uniprot and PlasmoDB). All other identified PfEMP1 sequences resulted from the translation of the reconstructed *var* genes from our samples.

PfEMP1 identification and composition

We then focused on the 55 protein groups associated to PfEMP1 proteins. The minimum peptide number for protein identification was 2. Sequence coverage ranged from 0.6 to 24.6 % maximum. Average molecular weight of the identified PfEMP1 was 228.3 kDa. Using the VarDom online server, we reconstructed the domain architecture from the identified proteins (Figure 2). Fifty-three sequences displayed at least one hit with the repository sequences. NTS domain was found in 73.6 % of the sequences (39/53 sequences presenting at least one hit with VarDom reference sequences) and 97 % (38/39) of the sequences presenting NTS displayed the following combination: NTS-DBLα-CIDRα. The three-major head-terminal domain organizations were the following: NTS-DBLα-CIDRα-DBLβ (58.9%; n = 23/39), NTS-DBLα-CIDRα-DBLδ (30.8%; n = 12/39) and NTS-DBLα-CIDRα-DBLγ (2.6%; n = 1/39). To attribute more precisely subdomains to the identified PfEMP1 proteins, we performed a local blast from the 110 peptides associated to the PfEMP1 proteins against the VarDom sequences repository (Figure 3 and Figure 4). Fifty four percent (n = 60/110) of the peptides displayed at least one hit with the reference sequences from VarDom (median hit number per peptide was 5 (interquartile range [2; 23])), among which 40/110 peptides displayed hits with more than 90% identity with a reference sequence (among these, median number of hits per peptide was 2 (interquartile range [1; 5])). Most of the obtained hits from peptides with reference sequences corresponded to ATS sequences (171 hits), and 71 hits covering DBL domains (among them, 25 for DBLα, 20 for DBLβ, 14 for DBLδ, 1 for DBLε and 11 for DBLγ). One hit corresponded to NTS domain. Subdomain attribution

was performed successfully for the corresponding peptides. No peptide was issued from CIDRα and seven peptides where issued from DBLβ.

The specific search of the binding pattern for ICAM-1 retrieved 6 identifications within the *Pf*EMP1 sequences identified in mass spectrometry.

PfEMP1 associated with clinical outcome of infection

Among the identified PfEMP1 proteins in LC-MS/MS, 11 were shared between all patient's group, 2 were specifically associated with CM, 5 with UM and 18 with SA. SA and CM shared 8 proteins, UM and CM 14 proteins, and SA and UM 6 proteins (Figure 5). The corresponding PfEMP1 are listed in Table 1.

The PfEMP1 containing ICAM-1 binding pattern were associated to several protein groups, and matching with 5/9 of SA samples, 3/4 CM, and 4/9 UM samples (Figure 3). However, no peptide was identified corresponding directly to the pattern. We performed a 3D structure prediction (using 3D7 template available in PDB) from all DLBβ identified within the sequences obtained from WGS and identified in LC-MS/MS. Average quality score was 0.554 (± 0.047). RMSD calculation showed that the obtained DBLβ structure are highly similar (Figure 6 A).

Focusing on CIDRα, we were able to produce acceptable (quality score 0.470 ± 0.092) folding for all sequences except two which were excluded from the clustering. As expected, we observed that CIDRα1-like sequences and CIDRα "non α1" were clustered respectively (Figure 6 B). Using PyMol visualization of predicted 3D structure, we focused on the two α helix involved respectively in CD36 binding and interaction with EPCR (35). Among the CIDRα sequences, 14 presented a CIDRα "non α1" helix fold, among which 12 were folded correctly.

We then predicted the structure from CIDRα linked to DBLβ presenting ICAM-1 binding predicted pattern – 6 PfEMP1 proteins containing DBLβ-ICAM-1 binding pattern (Figure 6C). These CIDRα presented a phenylalanine (5/6 sequences) or tyrosine (1/6 sequences) in position 187 (1/6 sequences), 188 (3/6 sequences), 184 (1/6 sequences), 190 (1/6 sequences) from the corresponding CIDRα (Figure 6B). These amino acids were juxtaposed and exposed similarly to the phenylalanine involved in EPCR binding from Hb3var3 CIDRα 3D structure.

Genomic and transcript data for specific protein identification

Since we had access to preserved RNA for 2 samples in addition to RNAsequencing data (CM04 and SA06 at 3 additional sampling time each: H0 (diagnosis of malaria) and post diagnosis sample, H16, H24 for CM04 and H0, H8, H16, for SA06), we selectively designed primers that would specifically amplify the PfEMP1 transcripts from the mass spectrometry matching sequences (Figure 1). Our goal was to assess which ones of these proteins are expressed, and their corresponding relative expression level using the Tu measurement method. Using alignment tools, we managed to find discriminant primers between the 6 identified sequences for sample SA06 (g45, g949, g1030, g1025, g150 and g62) and the 3 sequences (g276, g299 and g522) for sample CM04. Identified primers are presented in Supplementary data table 1 and quantification on Figure 7 for sample SA06. The amplification profile of reference genes was not satisfying for CM04-H0 and SA06–H08 samples, which were excluded from the analysis. For patient CM04, at H16, transcript expression was at the detection level threshold and no major transcript was identified. For patient CM04, at H24, the major transcript is g276 followed by g299 and g522 (Tu respectively: 43.26 – 4.24 – 1.32). For sequences that displayed different Melting Temperature (Tm), selective sequencing did not reveal differences between sequences, except for CM04 g276 sequence amplified with the designed primers.

Discussion

The evolution of *P. falciparum* infection from uncomplicated forms of the disease to cerebral malaria, the most fatal, is a complex phenomenon which entails host and parasite factors, in addition to socio-economics settings (36). There are strong evidence that the PfEMP1 proteins are involved in the disease progression since they allow the parasite to bind to host endothelium (14). It is believed that a distinct subset of PfEMP1 proteins is involved in severe malaria (23, 37), most likely by providing to the parasite the ability to sequester to a given receptor. However, PfEMP1 identification in natural infection remained challenging, due to the large size of PfEMP1 and their high sequences diversity. Recently, Jespersen et *al* (23) provided a new insight towards *var* genes sequences expression analysis in patient's sample using transcript reconstruction after DBLα barcoding. They confirmed the preferential expression of CIDRα associated with EPCR binding in severe malaria patients.

We used a mass spectrometry–based proteomic approach to analyse the *P. falciparum* proteome in the context of severe malaria (SA and CM) compared to UM. We aimed to accurately identify, at the protein level, the PfEMP1 sequence variants associated with diseases severity. To this end, we initiated a "proteogenomic" study of field samples. In collaboration with the WT Sanger institute, we performed whole genome sequencing of the samples and had access to the assembled *var* gene sequences of the samples. Second, we performed RNAsequencing of these samples and analysed by LC-MS/MS the corresponding matured parasites. RNAsequencing reads were aligned for *P. falciparum* transcripts expression both on reference transcriptome and sequences issued from WGS of our samples. Reference database allowed the identification of *var* transcripts; however, all mapped reads were in semi conserved ATS coding region. Protein identification was performed by matching mass-spectrometry data against a custom database containing the translated PfEMP1 sequences (Figure 1). As anticipated, most of the identified PfEMP1 (considering all PfEMP1 from all protein group) came from the newly added sequences to the database (10/96 were known sequences), confirming the validity of our approach.

The technical improvement allowed us to identify more proteins than previously published studies (34, 38), with higher sequence coverage. We identified a set of 55 PfEMP1 associated in protein group. We investigated the structure of theses sequences and found that the two main domain organisations were NTS-DBLα-CIDRα-DBLβ and NTS-DBLα-CIDRα-DBLδ. The high proportion of NTS-DBLα-CIDRα-DBLβ

among expressed PfEMP1 identified in our samples compared to genomic sequences within the same sample pool reflects the preferential expression of the PfEMP1 containing this pattern compared with the other PfEMP1 protein sequences in the studied context. The CIDRα-DBLβ tandem is associated with the potential "double binding" PfEMP1 (20, 24), targeting both ICAM-1 (through DBLβ (20)) and EPCR (through CIDRα (22)) human endothelial receptor. PfEMP1 binding pattern to ICAM-1 was not identified with specific peptides; however, proteins containing the binding pattern were identified in our samples (all malaria clinical presentation). We were able to predict domains structure for the CIDRα from PfEMP1 identified within our sequences pool. The proportion of "non α1" CIDR sequences identified is lower than the corresponding sequences in parasite genomes, (CIDRα2-6 represent 85% of PfEMP1 in parasite genomes) (37), enforcing the idea of preferential expression from PfEMP1 encoding EPCR binding CIDRα within malaria clinical presentation. In addition, all sequences presenting CIDRα-DBLβ- [ICAM-1 binder] harboured EPCR binding like CIDRα, enforced by 3D representation of domain structure.

For two isolates (SA09 and CM04), the identified PfEMP1 proteins were directly and unambiguously associated to transcript from the same isolate. To ensure whether several PfEMP1 proteins were expressed in the sample or if this was due to the identification method (peptides mapping sequences), we performed selective sequencing of the corresponding mRNA within the same sample. We managed to identify the most expressed transcript for one (CM04-H24), and the two main expressed transcripts for the other (SA09-H16). However, transcript corresponding to the other identified proteins existed at low expression level. It seems legitimate that along multiplication cycles, parasites express different PfEMP1. Furthermore, natural infection are often polyclonal and patients in endemic areas harbour usually 2 to 3 clones (39). However, we identified more sequences per sample than clones, strengthening the idea of simultaneous expression of several sets of PfEMP1 proteins by the parasite population within an individual. In addition, there are less PfEMP1 proteins identified in mass spectrometry compared to RNA sequencing/RT-qPCR approach. It seems plausible that, for technical reasons, PfEMP1 identification in LC-MS/MS provides a subset of the major proteins associated with each sample. In addition, all mRNA might not occur translation, as pseudogenes from *var* family are described. mRNA analysis reflects both coding and non-coding RNA expressed by a sample.

Focusing on patient's group (CM, SA and UM), no consensus or highly similar sequence was obtained for a given clinical outcome of *P. falciparum* infection. Most sequences were shared between patient groups, since their identification relies mostly on reads mapping in the semi-conserved region ATS. We could not specifically identify sequences coding for the ICAM-1 binding pattern or the CIDRα involved in EPCR binding within severe malaria group. However, for sample SA06, one of the major proteins and transcript identified corresponds to a sequence expressing the binding pattern for ICAM-1.

We were not able to attribute domains subtype to the obtained sequences. The high diversity of PfEMP1 protein sequences compared with the repository database prevented us from assigning a subdomain to every domain. This limitation did not allow us to specifically identify subtypes of domains, in relation to reference sequences, associated with *P. falciparum* infection severity. However, selective alignment from obtained peptides to reference sequences repository allowed to identify domain subtypes for 36% (40/110) of identified peptides. This difficulty to attribute a domain may be the consequence of either: the peptide being shared between two or more domains, or the peptide has never been previously identified thus absent from the database (Figure 2). Variant protein identification in mass spectrometry is challenging, since protein identification is the reflection of the quality of the database available. Restrictive use of reference sequences from public repositories reduces the "chance" of protein identification, but we ride out this challenge with our home-custom sequence database obtained through whole genome sequencing of our samples. The variability of PfEMP1 primary sequence was overcome using protein structure modelling, which allowed us to classify the CIDRα identified within predicted binding target. We confirmed that PfEMP1 protein sequences which display ICAM-1 binding pattern present CIDRα domains with EPCR binding folding.

This study highlights the perspective offered by proteogenomic approaches for VSA studies. Using a home-made database containing PfEMP1 sequences obtained from WGS of the corresponding samples, we managed to identify specific PfEMP1 pattern associated with each sample, where RNAsequencing alone could not discriminate the expressed PfEMP1 within a sample. However, the sequence coverage of the identified proteins remained limited. Further attempts should focus on increasing the number of identified peptides per protein, using for example selective enrichment in membrane proteins. The identification of new PfEMP1 variants using whole genome sequencing and the confirmation of their expression at the protein level is, to our knowledge, the first report using field isolates.

In conclusion, we identified PfEMP1 proteins expressed by parasite in patients presenting several forms of malaria. This is the first report of PfEMP1 direct identification and is providing insight towards malaria pathogenesis understanding. PfEMP1 are adhesin which mediates iE adhesion to the host endothelium. The high proportion of CIDRα among the identified sequences enforce the idea that iE sequestration occurs either through CD36 binding, or EPCR binding, pending of clinical presentation (22, 40). We also preferentially identified PfEMP1 protein harbouring DBLβ, among which 20% (6/30 identified DBLβ) displayed the binding pattern for ICAM-1. This strengthen the hypothesis that DBLβ is involved in the disease development, as demonstrated with antibodies against DBLβ in Tanzania (41) and Papua New Guinea (42). Antibodies against full length DBLβ3$_{PF11_0521}$ are associated with reduction risk of severe malaria and protection against clinical infection (42). We demonstrated that predicted 3D structure from all DBLβ identified in the studied followed closely DBLβ3$_{PF11_0521}$ structure, enforcing the idea of DBLβ3$_{PF11_0521}$ as a vaccination target to prevent severe malaria. Further studies to identify immune response to DBLβ3$_{PF11_0521,}$ in combination with different CIDRα subtype are needed. There is little doubt that upcoming strategies to prevent severe malaria will target DBLβ3$_{PF11_0521}$ in combination with EPCR binder CIDRα in a multi epitope vaccination strategy.

Our study opens opportunities to identify PfEMP1 variants and later implement these newly identified sequences in PfEMP1 based vaccine development strategies. However, our approach was restricted to one geographic area (Benin, West Africa) and included a limited number of patients. Further studies should include patients from various *P. falciparum* endemic areas to better represent PfEMP1 associated within *P. falciparum* disease in general and specifically to severe malaria.

Material and Methods

Ethic statement

Ethical clearance was obtained from the Institutional Ethics Committee of the faculty of health science at the Abomey-Calavi University in Benin (clearance n°90, 06/06/2016). Before inclusion, written informed consent was obtained from children' guardians. Patients were treated in accordance to the national malaria program policy. The methods were carried out in accordance with the relevant guidelines and regulations.

Sample collection

Patients under age 5, presenting severe malaria were included in the Lagune Mother and Child Hospital in Cotonou, Benin. Two distinct clinical groups were constituted, as following. CM was defined as a *Plasmodium falciparum* infection associated with a coma (Blantyre score ≤ 2) and the absence of meningitis detected by CSF count and culture. SA was defined as a *P. falciparum* infection associated with Hb < 5g/dL, measured using Hemocue® device (Radiometer). UM patients were included in Saint-Joseph Hospital, in Sô-Ava, Benin (16 km from Cotonou). UM was defined as a *P. falciparum* infection with fever, in the absence of any other complication. The study was conducted at the rainy season (May – August) 2016.

Five mL of peripheral whole blood were collected on EDTA, (with additional sampling at 8h, 16h and 24h post treatment initiation for CM and SA patients). Parasite density was evaluated with Giemsa-stained thick blood smear. Only pure *P. falciparum* infection were retained for the study. Samples were depleted from white blood cells using a gradient based separation technique Ficoll (GE Healthcare Life Science).

Whole genome sequencing

Fifty µL of erythrocyte's pellet was extracted using DNEasy Blood kit (Qiagen). WGS was performed by the Malaria Genomic Epidemiology Network (MalariaGEN) at the Welcome Trust Sanger Institute (Hinxton, UK).

Transcriptome studies

Ring staged parasite were preserved in 5 volumes of pre-warmed (37°C) TriZol (Life Technology), vortexed then immediately frozen at -80°C until further utilization. RNA was extracted as described (43), then digested with DNAse I (Qiagen) and purified using RNEasy MinElute Cleanup kit column (Qiagen). Absence of genomic DNA was assessed by the absence of amplification of p90 gene in qPCR from the

direct use of the RNA extract, using Syber Green reagent (Life Technology) on Rotorgene 6000 (Corbett Technology). RNA quality was assessed using PicoChip Agilent 2100TM Bioanalyzeur profile (Agilent). Only RNA presenting a RNA Integrity Number (RIN) > 7 were retained for downstream analysis (44). Sequencing technology used was an Illumina NexSeq500 (IPS2 POPS platform). RNA-seq libraries were performed by TruSeq Stranded mRNA protocol (Illumina®, California, U.S.A.). The RNA-seq samples have been sequenced in paired-end (PE) with a sizing of 260 base pairs and a read length of 150 bases. 54 samples by lane of NextSeq500 using individual bar-coded adapters and giving approximately 5 million of PE reads by sample are generated.

RNA-sequencing data bioinformatics treatment and analysis

RNA-Seq preprocessing includes trimming library adapters and performing quality controls. The raw data (fastq) were trimmed with Trimmomatic (45) tool for Phred Quality Score Qscore >20, read length >30 bases, and ribosome sequences were removed with tool sortMeRNA (46). Reads were mapped against the *Plasmodium falciparum* 3D7 reference genome (PlasmoDB v35) and against the reconstructed sequences from WGS, using STAR (version 2.5.4b) with default parameters. A sequence was considered as successfully sequenced if more than 70% of the transcript was covered with at least 5 reads.

Obtention of mature forms from *P. falciparum*

Blood samples were matured *in vitro* for 18 to 32 hours in RPMI medium supplemented with human serum and Albumax (Gibco) and preserved after MACS^{TM} (Myltenyi Biotech) enrichment as described (33).

Protein extraction and sample preparation for mass spectrometry analysis

Whole cell infected erythrocyte lysates were solubilized and digested in solution using trypsin (Promega, sequencing Grade). Briefly, 50 µg of proteins from whole cell lysates were diluted to 25 µl in solubilization buffer (1% sodium desoxycholate, 100 mM Tris/HCl, pH8.5, 10mM TCEP, 40 mM chloroacetamide), heated for 5 min at 95°C and sonicated three times for 30 s. Once at room temperature, extracts were diluted with 25 µl Tris-ACN buffer (50mM Tris/HCl pH 8.5, 10%ACN). Collected peptides were fractionated in 5 fractions per sample by strong cationic exchange (SCX) StageTips (47).

Mass spectrometry analysis

Mass Spectrometry (MS) analysis were performed on a Dionex U3000 RSLC nano-LC-system coupled to an Orbitrap-fusion mass spectrometer (Thermo Fisher Scientific) as described (48). Peptides were separated on a C18 reverse-phase resin (75-µm inner diameter and 15-cm length) with a 3-hr gradient. The mass spectrometer acquired data throughout the elution process and operated in a data-dependent scheme.

Accession number

The raw reads were available under the accession number listed in supplemental data 5. VarDOM server for *var* genes is available at the following address www.cbs.dtu.dk/services/VarDom/. PlasmoDB reference sequences are available on EupathDB portal dedicated to *Plasmodium* (http://plasmodb.org/plasmo/).

Database construction

To get an accurate identification of proteins, a database was created as follows: Since our samples have been sequenced by the MalariaGen consortium (UK), whole genome sequences were available. Reconstructed *var* genes were kindly provided by Thomas Otto, Matt Berriman and Chris Newbold from the Welcome Trust Sanger Institute. We concatenated *P. falciparum* proteins sequences from PlasmoDB (v35) Uniprot and NCBI. We added our own PfEMP1 sequences, obtained after *in silico* translation from *var* genes reconstruction. Duplicate sequences were removed. FASTA sequence headers were uniformed to facilitate MaxQuant sequences analysis.

Raw data processing

The mass spectrometry data were analyzed using Maxquant version 1.5.2.8 (49). The database used was our homemade database (see above section) and the list of contaminant sequences from Maxquant. The precursor mass tolerance was set to 4.5 ppm and the fragment mass tolerance to 0.5 Da. Carbamidomethylation of cysteins was set as constant modification and acetylation of protein N-terminus and oxidation of methionine were set as variable modifications. Second peptide search was allowed, and minimal length of peptides was set at 7 amino acids. False discovery rate (FDR) was kept below 1% on both peptides and proteins. Label-free protein quantification (LFQ) was done using both unique and razor peptides. At least 2 such peptides were required for LFQ. The “match between runs” (MBR) option was allowed with a match time window of 1 min and an alignment time window of 30 min.

For analysis, LFQ results from MaxQuant were imported into the Perseus software (version 1.5.1.6). Reverse and contaminant proteins were excluded. Only proteins from *P. falciparum* were selected for further analysis. We then focused on the membrane associated and putative proteins from *P. falciparum*.

Protein sequence analysis and structure prediction

PfEMP1 sequences were aligned using the VarDom server (available at http://www.cbs.dtu.dk/services/VarDom/) for domain type identification (11). We specifically searched the pattern for ICAM-1 binding I[V/L]x_9N[E]GG[P/A]xYx_{27}GPPx_3H (20) in the identified PfEMP1 sequences using the ProSite online interface (50).

Using domain limitation performed with VarDOM server, we predicted the structure of the CIDRα and DBLβ domains identified within our sequences. Structure were generated using Modeller (v9.20.) (51). We selected templates from the Protein Data Bank (52) using BLAST algorithm. Template selection was performed manually using BLAST results, sequence coverage within each template and template quality. DBLβ structure prediction was performed using PDB 5mza (PF11_0521). For the CIDRα domain, the following templates were used: 4v3d (from HB3var03), 4v3e (from IT4var07), 3c64 (MC179) and 5lgd (from MCvar1). Template and query quality scores were calculated using ProQ3 online tool. For 5mza the quality score was 0.560. For 4v3d, 4v3e, 3c64 and 5lgd, quality scores were respectively 0.484, 0.585, 0.396 and 0.540. For each domain, 5 models have been generated, and 4 additional models were generated after loop refining. Template and query mapping has been performed using align2D tool from Modeller script. For structure modelling using several templates, we used salign script (part of Modeller script).

The choice of the best model was made based on the Discrete Optimized Protein Energy (DOPE) score values. In order to analyze and compare the modeled structures were compared using the Bio3D v 3.2 R package (53). Structure comparison was based on Root-Mean-Square Deviation (RMSD) calculated using Cα distance from predicted structures. Obtained RMSD values were clustered using Bio3D R package (53).

Statistical analysis

Patient's samples information's were compared between the 3 patient's groups (UM, CM and SA) using one-way ANOVA. Bonferroni's Multiple Comparison Test was applied for individual group comparison.

We considered a *p* value < 0.05 as significant. Qualitative data were compared with Chi Squared test using contingency table. All analyses were performed using Prism v5 (Graphpad).

Primer design

We tried to specifically assign PfEMP1 expression within the samples. In addition to RNAsequencing data, we disposed of preserved RNA from 2 samples. Our strategy was as follows: first, we compared the sequences of the identified proteins using Multalin (54). Second, we designed specific primers targeting the sequences using Primer3Plus web interface (55). Finally, we assessed primers specificity using BLAST tools. (Primers are listed in Supplementary data 2).

RT-qPCR and PCR

RNA was preserved using Trizol reagents (Life Technology) and RNA extraction was performed as described (43). cDNA was obtained using Superscript II (Life Technology) retrotranscriptase following the manufacturer's instructions.

qPCR was performed using Power SYBR Green PCR Master Mix (Applied Biosystems) on Rotorgen (Corbett technology) as described (34). Melting stage was set up as default on the Rotorgen software. PCR specificity was assessed by the melting temperature (Tm). We considered that a sequence is unique for Tm included in the range Tm +/- 0.5°C. Transcript expression was quantified using the Transcript Unit (Tu) method, as described by Lavsten et *al* (31). GAPDH and Seryl-tRNA synthetase transcripts were used as reference genes.

For transcripts which displayed a Tm difference upper than 0.5°C for a same primer set, we performed a conventional PCR using AmpliTaGold DNA polymerase (Applied Biosystem) with the following conditions: initial denaturation at 95°C for 10 min, followed by 52°C hybridization for 30s – 68°C elongation for 15 s, repeated 35 times. PCR products length was assessed using 2% agarose gel containing Syber Safe (Life Technology) at 1/10,000.

Amplicons were purified using the QIAquick PCR Purification Kit (Qiagen). The purified fragments were sequenced with BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) using g949, g150 (for patient SA06), and g299 primers (for patient CM04) (Supplementary data 1). The sequence reaction products were purified using the BigDye XTerminator® Purification Kit (Applied Biosystems), in

455 accordance with the manufacturer's instructions. The purified products were sequenced using an ABI Prism
456 3100 analyzer (Applied Biosystems), and the sequences were analyzed using DNA Baser v4 (Heracle
457 BioSoft).

458

Acknowledgments

The authors would like to thank patients who participate in the study, the clinicians and nurses who were involved in patient's inclusion. We especially acknowledge Dr Nadine Fievet help and counselling during the field study. We would like to thank Matt Berriman, Chris Newbold and Thomas Otto from the Welcome Trust Sanger Institute for the access to the unpublished PfEMP1 sequences. We also would like to thank Emilie-Fleur Gautier for her helpful advices in mass spectrometry, and Evangeline Bennana for her technical support in sample preparation. The authors thank Romain Coppée for his critical comments on the manuscript, and Antoine Claessens for his help in WGS. We also thank Feder for the opportunity to use Orbitrap Fusion mass spectrometer.

Author's contribution

CK, PD, FG and GB designed the study. CK, SE, JR, JMA and GB conducted the field study. CK, GB, CB, SG, CP and SH performed the experiments. CK, DR and EG analyzed the data. CK, GB and PD wrote the paper.

Conflict of interest

The authors declare no conflict of interest.

Funding

This work was funded by Merieux Research Grant (awarded to Dr Gwladys Bertin for CIVIC project), by the Laboratoire d'Excellence GR-Ex, Paris, France, reference ANR-11-LABX-0051, that is funded by the program Investissements d'avenir of the French National Research Agency, reference ANR-11-IDEX-0005-02 and by NeuroCM project, that is funded by ANR-17-CE17-0001

The POPS platform benefits from the support of the LabEx Saclay Plant Sciences-SPS (ANR-10-LABX-0040-SPS).

CK is awarded a PhD Scholarship from French Minister of Research through MTCI Doctoral School (ED 563, Paris Descartes University).

References

1. WORLD HEALTH ORGANIZATION. 2018. WORLD MALARIA REPORT 2017. WORLD HEALTH ORGANIZATION, S.l.

2. Kessler A, Dankwa S, Bernabeu M, Harawa V, Danziger SA, Duffy F, Kampondeni SD, Potchen MJ, Dambrauskas N, Vigdorovich V, Oliver BG, Hochman SE, Mowrey WB, MacCormick IJC, Mandala WL, Rogerson SJ, Sather DN, Aitchison JD, Taylor TE, Seydel KB, Smith JD, Kim K. 2017. Linking EPCR-Binding PfEMP1 to Brain Swelling in Pediatric Cerebral Malaria. Cell Host Microbe 22:601-614.e5.

3. Kyes SA, Rowe JA, Kriek N, Newbold CI. 1999. Rifins: a second family of clonally variant proteins expressed on the surface of red cells infected with Plasmodium falciparum. Proc Natl Acad Sci U S A 96:9333–9338.

4. Goel S, Palmkvist M, Moll K, Joannin N, Lara P, Akhouri RR, Moradi N, Öjemalm K, Westman M, Angeletti D, Kjellin H, Lehtiö J, Blixt O, Ideström L, Gahmberg CG, Storry JR, Hult AK, Olsson ML, von Heijne G, Nilsson I, Wahlgren M. 2015. RIFINs are adhesins implicated in severe Plasmodium falciparum malaria. Nat Med 21:314–317.

5. Cheng Q, Cloonan N, Fischer K, Thompson J, Waine G, Lanzer M, Saul A. 1998. stevor and rif are Plasmodium falciparum multicopy gene families which potentially encode variant antigens. Mol Biochem Parasitol 97:161–176.

6. Niang M, Bei AK, Madnani KG, Pelly S, Dankwa S, Kanjee U, Gunalan K, Amaladoss A, Yeo KP, Bob NS, Malleret B, Duraisingh MT, Preiser PR. 2014. STEVOR is a Plasmodium falciparum erythrocyte binding protein that mediates merozoite invasion and rosetting. Cell Host Microbe 16:81–93.

7. Leech JH, Barnwell JW, Miller LH, Howard RJ. 1984. Identification of a strain-specific malarial antigen exposed on the surface of Plasmodium falciparum-infected erythrocytes. J Exp Med 159:1567–1575.

8. Su XZ, Heatwole VM, Wertheimer SP, Guinet F, Herrfeldt JA, Peterson DS, Ravetch JA, Wellems TE. 1995. The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of Plasmodium falciparum-infected erythrocytes. Cell 82:89–100.

9. Baruch DI, Pasloske BL, Singh HB, Bi X, Ma XC, Feldman M, Taraschi TF, Howard RJ. 1995. Cloning the P. falciparum gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitized human erythrocytes. Cell 82:77–87.

10. Smith JD, Chitnis CE, Craig AG, Roberts DJ, Hudson-Taylor DE, Peterson DS, Pinches R, Newbold CI, Miller LH. 1995. Switches in expression of Plasmodium falciparum var genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. Cell 82:101–110.

11. Rask TS, Hansen DA, Theander TG, Gorm Pedersen A, Lavstsen T. 2010. Plasmodium falciparum erythrocyte membrane protein 1 diversity in seven genomes--divide and conquer. PLoS Comput Biol 6.

12. Claessens A, Hamilton WL, Kekre M, Otto TD, Faizullabhoy A, Rayner JC, Kwiatkowski D. 2014. Generation of antigenic diversity in Plasmodium falciparum by structured rearrangement of Var genes during mitosis. PLoS Genet 10:e1004812.

13. Smith JD. 2014. The role of PfEMP1 adhesion domain classification in Plasmodium falciparum pathogenesis research. Mol Biochem Parasitol 195:82–87.

14. Wahlgren M, Goel S, Akhouri RR. 2017. Variant surface antigens of Plasmodium falciparum and their roles in severe malaria. Nat Rev Microbiol.

15. Smith JD, Craig AG, Kriek N, Hudson-Taylor D, Kyes S, Fagan T, Fagen T, Pinches R, Baruch DI, Newbold CI, Miller LH. 2000. Identification of a Plasmodium falciparum intercellular adhesion molecule-1 binding domain: a parasite adhesion trait implicated in cerebral malaria. Proc Natl Acad Sci U S A 97:1766–1771.

16. Turner L, Lavstsen T, Berger SS, Wang CW, Petersen JEV, Avril M, Brazier AJ, Freeth J, Jespersen JS, Nielsen MA, Magistrado P, Lusingu J, Smith JD, Higgins MK, Theander TG. 2013. Severe malaria is associated with parasite binding to endothelial protein C receptor. Nature 498:502–505.

17. Hess DC, Bhutwala T, Sheppard JC, Zhao W, Smith J. 1994. ICAM-1 expression on human brain microvascular endothelial cells. Neurosci Lett 168:201–204.

18. Bengtsson A, Joergensen L, Rask TS, Olsen RW, Andersen MA, Turner L, Theander TG, Hviid L, Higgins MK, Craig A, Brown A, Jensen ATR. 2013. A novel domain cassette identifies Plasmodium falciparum PfEMP1 proteins binding ICAM-1 and is a target of cross-reactive, adhesion-inhibitory antibodies. J Immunol Baltim Md 1950 190:240–249.

19. Madkhali AM, Alkurbi MO, Szestak T, Bengtsson A, Patil PR, Wu Y, Al-Harthi SA, Alharthi S, Jensen ATR, Pleass R, Craig AG. 2014. An analysis of the binding characteristics of a panel of recently selected ICAM-1 binding Plasmodium falciparum patient isolates. PloS One 9:e111518.

20. Lennartz F, Adams Y, Bengtsson A, Olsen RW, Turner L, Ndam NT, Ecklu-Mensah G, Moussiliou A, Ofori MF, Gamain B, Lusingu JP, Petersen JEV, Wang CW, Nunes-Silva S, Jespersen JS, Lau CKY, Theander TG, Lavstsen T, Hviid L, Higgins MK, Jensen ATR. 2017. Structure-Guided Identification of a Family of Dual Receptor-Binding PfEMP1 that Is Associated with Cerebral Malaria. Cell Host Microbe 21:403–414.

21. Mkumbaye SI, Wang CW, Lyimo E, Jespersen JS, Manjurano A, Mosha J, Kavishe RA, Mwakalinga SB, Minja DTR, Lusingu JP, Theander TG, Lavstsen T. 2017. The Severity of Plasmodium falciparum Infection Is Associated with Transcript Levels of var Genes Encoding Endothelial Protein C Receptor-Binding P. falciparum Erythrocyte Membrane Protein 1. Infect Immun 85:e00841-16.

22. Lau CKY, Turner L, Jespersen JS, Lowe ED, Petersen B, Wang CW, Petersen JEV, Lusingu J, Theander TG, Lavstsen T, Higgins MK. 2015. Structural conservation despite huge sequence diversity allows EPCR binding by the PfEMP1 family implicated in severe childhood malaria. Cell Host Microbe 17:118–129.

555 23. Jespersen JS, Wang CW, Mkumbaye SI, Minja DT, Petersen B, Turner L, Petersen JE, Lusingu JP,
556 Theander TG, Lavstsen T. 2016. Plasmodium falciparum var genes expressed in children with severe
557 malaria encode CIDRα1 domains. EMBO Mol Med 8:839–850.

558 24. Tuikue Ndam N, Moussiliou A, Lavstsen T, Kamaliddin C, Jensen ATR, Mama A, Tahar R, Wang
559 CW, Jespersen JS, Alao JM, Gamain B, Theander TG, Deloron P. 2017. Parasites Causing Cerebral
560 Falciparum Malaria Bind Multiple Endothelial Receptors and Express EPCR and ICAM-1-Binding
561 PfEMP1. J Infect Dis 215:1918–1925.

562 25. Bull PC, Abdi AI. 2016. The role of PfEMP1 as targets of naturally acquired immunity to childhood
563 malaria: prospects for a vaccine. Parasitology 143:171–186.

564 26. Bull PC, Lowe BS, Kortok M, Molyneux CS, Newbold CI, Marsh K. 1998. Parasite antigens on the
565 infected red cell surface are targets for naturally acquired immunity to malaria. Nat Med 4:358–360.

566 27. Nielsen MA, Staalsoe T, Kurtzhals JAL, Goka BQ, Dodoo D, Alifrangis M, Theander TG, Akanmori
567 BD, Hviid L. 2002. Plasmodium falciparum variant surface antigen expression varies between isolates
568 causing severe and nonsevere malaria and is modified by acquired immunity. J Immunol Baltim Md
569 1950 168:3444–3450.

570 28. Barry AE, Leliwa-Sytek A, Tavul L, Imrie H, Migot-Nabias F, Brown SM, McVean GAV, Day KP.
571 2007. Population genomics of the immune evasion (var) genes of Plasmodium falciparum. PLoS
572 Pathog 3:e34.

573 29. Nielsen MA, Vestergaard LS, Lusingu J, Kurtzhals JAL, Giha HA, Grevstad B, Goka BQ, Lemnge
574 MM, Jensen JB, Akanmori BD, Theander TG, Staalsoe T, Hviid L. 2004. Geographical and temporal
575 conservation of antibody recognition of Plasmodium falciparum variant surface antigens. Infect Immun
576 72:3531–3535.

577 30. Hviid L, Lavstsen T, Jensen AT. 2018. A vaccine targeted specifically to prevent cerebral malaria - is
578 there hope? Expert Rev Vaccines.

31. Lavstsen T, Turner L, Saguti F, Magistrado P, Rask TS, Jespersen JS, Wang CW, Berger SS, Baraka V, Marquard AM, Seguin-Orlando A, Willerslev E, Gilbert MTP, Lusingu J, Theander TG. 2012. Plasmodium falciparum erythrocyte membrane protein 1 domain cassettes 8 and 13 are associated with severe malaria in children. Proc Natl Acad Sci U S A 109:E1791-1800.

32. Argy N, Bertin GI, Milet J, Hubert V, Clain J, Cojean S, Houzé P, Tuikue-Ndam N, Kendjo E, Deloron P, Houzé S, CNRP Study Group. 2017. Preferential expression of domain cassettes 4, 8 and 13 of Plasmodium falciparum erythrocyte membrane protein 1 in severe malaria imported in France. Clin Microbiol Infect Off Publ Eur Soc Clin Microbiol Infect Dis 23:211.e1-211.e4.

33. Bertin GI, Sabbagh A, Guillonneau F, Jafari-Guemouri S, Ezinmegnon S, Federici C, Hounkpatin B, Fievet N, Deloron P. 2013. Differential protein expression profiles between Plasmodium falciparum parasites isolated from subjects presenting with pregnancy-associated malaria and uncomplicated malaria in Benin. J Infect Dis 208:1987–1997.

34. Bertin GI, Lavstsen T, Guillonneau F, Doritchamou J, Wang CW, Jespersen JS, Ezimegnon S, Fievet N, Alao MJ, Lalya F, Massougbodji A, Ndam NT, Theander TG, Deloron P. 2013. Expression of the domain cassette 8 Plasmodium falciparum erythrocyte membrane protein 1 is associated with cerebral malaria in Benin. PloS One 8:e68368.

35. Hsieh F-L, Turner L, Bolla JR, Robinson CV, Lavstsen T, Higgins MK. 2016. The structural basis for CD36 binding by the malaria parasite. Nat Commun 7:12837.

36. Miller LH, Baruch DI, Marsh K, Doumbo OK. 2002. The pathogenic basis of malaria. Nature 415:673–679.

37. Smith JD, Rowe JA, Higgins MK, Lavstsen T. 2013. Malaria's deadly grip: cytoadhesion of Plasmodium falciparum-infected erythrocytes. Cell Microbiol 15:1976–1983.

38. Bertin GI, Sabbagh A, Argy N, Salnot V, Ezinmegnon S, Agbota G, Ladipo Y, Alao JM, Sagbo G, Guillonneau F, Deloron P. 2016. Proteomic analysis of Plasmodium falciparum parasites from patients with cerebral and uncomplicated malaria. Sci Rep 6:26773.

604 39. Jafari S, Le Bras J, Bouchaud O, Durand R. 2004. Plasmodium falciparum clonal population dynamics
605 during malaria treatment. J Infect Dis 189:195–203.

606 40. Avril M, Bernabeu M, Benjamin M, Brazier AJ, Smith JD. 2016. Interaction between Endothelial
607 Protein C Receptor and Intercellular Adhesion Molecule 1 to Mediate Binding of Plasmodium
608 falciparum-Infected Erythrocytes to Endothelial Cells. mBio 7:e00615-16.

609 41. Cham GKK, Turner L, Kurtis JD, Mutabingwa T, Fried M, Jensen ATR, Lavstsen T, Hviid L, Duffy
610 PE, Theander TG. 2010. Hierarchical, domain type-specific acquisition of antibodies to Plasmodium
611 falciparum erythrocyte membrane protein 1 in Tanzanian children. Infect Immun 78:4653–4659.

612 42. Tessema SK, Utama D, Chesnokov O, Hodder AN, Lin CS, Harrison GLA, Jespersen JS, Petersen B,
613 Tavul L, Siba P, Kwiatkowski D, Lavstsen T, Hansen DS, Oleinikov AV, Mueller I, Barry AE. 2018.
614 Antibodies to Intercellular Adhesion Molecule 1-Binding Plasmodium falciparum Erythrocyte
615 Membrane Protein 1-DBLβ Are Biomarkers of Protective Immunity to Malaria in a Cohort of Young
616 Children from Papua New Guinea. Infect Immun 86.

617 43. Ponts N, Chung D-WD, Le Roch KG. 2012. Strand-specific RNA-seq applied to malaria samples.
618 Methods Mol Biol Clifton NJ 883:59–73.

619 44. Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, Lightfoot S, Menzel W,
620 Granzow M, Ragg T. 2006. The RIN: an RNA integrity number for assigning integrity values to RNA
621 measurements. BMC Mol Biol 7:3.

622 45. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data.
623 Bioinforma Oxf Engl 30:2114–2120.

624 46. Kopylova E, Noé L, Touzet H. 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in
625 metatranscriptomic data. Bioinforma Oxf Engl 28:3211–3217.

626 47. Kulak NA, Pichler G, Paron I, Nagaraj N, Mann M. 2014. Minimal, encapsulated proteomic-sample
627 processing applied to copy-number estimation in eukaryotic cells. Nat Methods 11:319–324.

48. Gautier E-F, Ducamp S, Leduc M, Salnot V, Guillonneau F, Dussiot M, Hale J, Giarratana M-C, Raimbault A, Douay L, Lacombe C, Mohandas N, Verdier F, Zermati Y, Mayeux P. 2016. Comprehensive Proteomic Analysis of Human Erythropoiesis. Cell Rep 16:1470–1484.

49. Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M. 2014. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol Cell Proteomics MCP 13:2513–2526.

50. Sigrist CJA, Cerutti L, Hulo N, Gattiker A, Falquet L, Pagni M, Bairoch A, Bucher P. 2002. PROSITE: a documented database using patterns and profiles as motif descriptors. Brief Bioinform 3:265–274.

51. Sali A, Blundell TL. 1993. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234:779–815.

52. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. 2000. The Protein Data Bank. Nucleic Acids Res 28:235–242.

53. Grant BJ, Rodrigues APC, ElSawy KM, McCammon JA, Caves LSD. 2006. Bio3d: an R package for the comparative analysis of protein structures. Bioinforma Oxf Engl 22:2695–2696.

54. Corpet F. 1988. Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res 16:10881–10890.

55. Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JAM. 2007. Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res 35:W71–W74.

Figure's legends

Figure 1: Proteogenomic approach on field samples for PfEMP1 identification.

Whole blood sample from patients are collected. DNA and RNA are extracted from parasite's ring forms. From the same collected tube, parasites are matured and the corresponding proteins from mature form (within one maturation cycle) are extracted and analyzed in LC-MS/MS. Whole genome sequencing data enrich the protein database used for protein identification with LC-MS/MS data. Selective mRNA

sequencing provides additional information's to elucidate proteins identification within the home-made database.

Figure 2: Schematic representation of domain attribution to PfEMP1 sequences and peptides

The PfEMP1 « x » has been identified in mass spectrometry with the peptides « A », « B », « C », « D », and « E ». The general domains organisation of PfEMP1 « x » can be attributed using the VarDom online servers.

However, subdomains attribution is not over possible with confident score. Local blast of the peptides used for protein identification retrieves different types of results. First option (peptides « A », and « E ») is that the peptide local blast allows for a « general » domain attribution, without subtype. Second option (peptide « C ») is that the peptide is known within the precise sequence of a subdomain and allows identification of a subdomain. Unattributed peptides within the reference sequences has two origins. 1/The peptide is shared between two domains and the amino acids belonging to each domain are not sufficient for confident attribution. 2/The peptide is « unknown » in the reference sequence repository (a new PfEMP1 related peptide has been implemented).

Figure 3: Domain description of the identified PfEMP1

PfEMP1 domain assignment for the one identified through mass spectrometry experiment. Proteins sequences have been mapped on VarDom server for domain attribution. Each color correspond to a domain type, corresponding to the VarDom server color code. DBLβ containing the binding pattern to ICAM-1 are highlighted with bold red rectangle.

Figure 4: PfEMP1 domain count in the corresponding samples

Number of each domain attribute to the PfEMP1 peptides identified in CM, SA and UM samples. Color code is proportional to the number of identified domains (blue: low number and red: high number). NTS associated peptide is found once in each sample, except for 1 CM isolate and 4 SA isolates. CIDRα matching peptides have been observed in 2/4 CM, 5/9 SA and 2/9 UM. DBLδ matching domains have been

677 found in 3/4 CM, 9/9 SA and 9/9 UM. In all samples, peptides corresponding to the protein sequence did
678 not match with a known domain.

679 **Figure 5: Venne Diagram representation of PfEMP1 identification**

680 Representation of sequence identified in accordance with clinical presentation. 18 sequences were associated
681 with SA only, 2 with CM and 5 with UM. SA and CM shared 8 sequences, CM and UM 14 sequences, and
682 SA and UM 6 sequences. All three groups shared 11 sequences as well. CIDRα domain in found in ½ CM
683 specific sequence; 13/18 SA sequences; 3/8 sequences shared between CM and SA and 3/5 UM sequences.
684 Concerning DBLβ, the domain is contained in ½ CM sequences, 11/18 SA sequences, 4/8 sequences shared
685 between CM and SA, and 4/5 UM sequences

686 **Figure 6: 3D representation of the CIDRα associated to DBLβ containing ICAM-1 binding pattern**

687 A) PCA representation of DBLβ RMSD values obtained after structure modelling. B) PCA representation of
688 CIDRα RMSD values obtained after structure modelling. Green circle highlight sequences presenting
689 CIDRα binding EPCR C) Juxtaposition of the structure predicted from the 6 CIDRα associated to DBLβ
690 containing ICAM-1 binding pattern. Amino acid known for binding to EPCR is highlighted in blue. The two
691 front α helix are folded accordingly to the structure from Hb3var3 and ITvar7, which bind EPCR. Structure
692 have been drawn using Pymol.

693 **Figure 7 : Transcript proportion for sample SA06, identified by selective RT-qPCR**

694 Expression of *var* transcript within sample SA06 at diagnosis and 16h post treatment initiation. The relative
695 percentage of each transcript is presented. Each transcript expression has been evaluated using dedicated
696 primers specific to the newly identified *var* sequences.

697 Using two reference genes for quantification, we identified the transcript g45 as the major expressed
698 transcript for SA06 (at H16) samples, followed by transcript g949, g1030, g1025, g150, g62 (Tu
699 respectively 3114.96 – 75.85 – 18.19 – 17.81 – 13.98 – 1.90). In the H0 sample, the following transcript
700 were identified g1025, g1030, g45, g150, g949, g62 (Tu respectively: 23.59 – 6.77 – 2.64 – 2.41- 2.25 –
701 1.66)

702 **Table 1**

703 PfEMP1 sequences identified within each clinical group

704 # Supplemental material legends

705 **Supplemental material 1**

706 Control transcript evaluation for RNAsequencing

707 **Supplemental material 2**

708 Primers used for selective PfEMP1 sequencing

709 **Supplemental material 3**

710 Identification of PfEMP1

711 **Supplemental material 4**

712 Selective sequencing of *var* genes within field samples

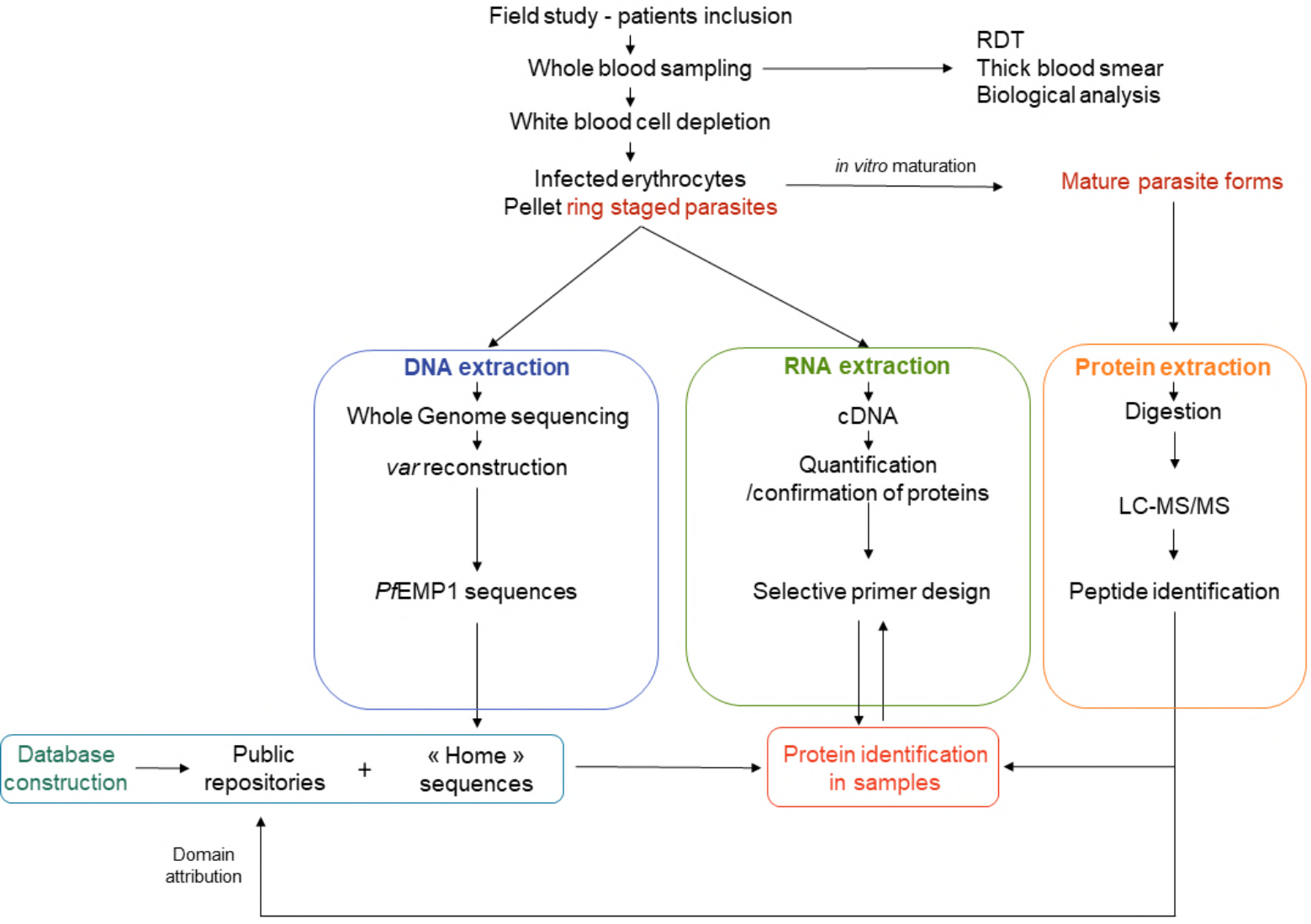
713 **Supplemental material 5**

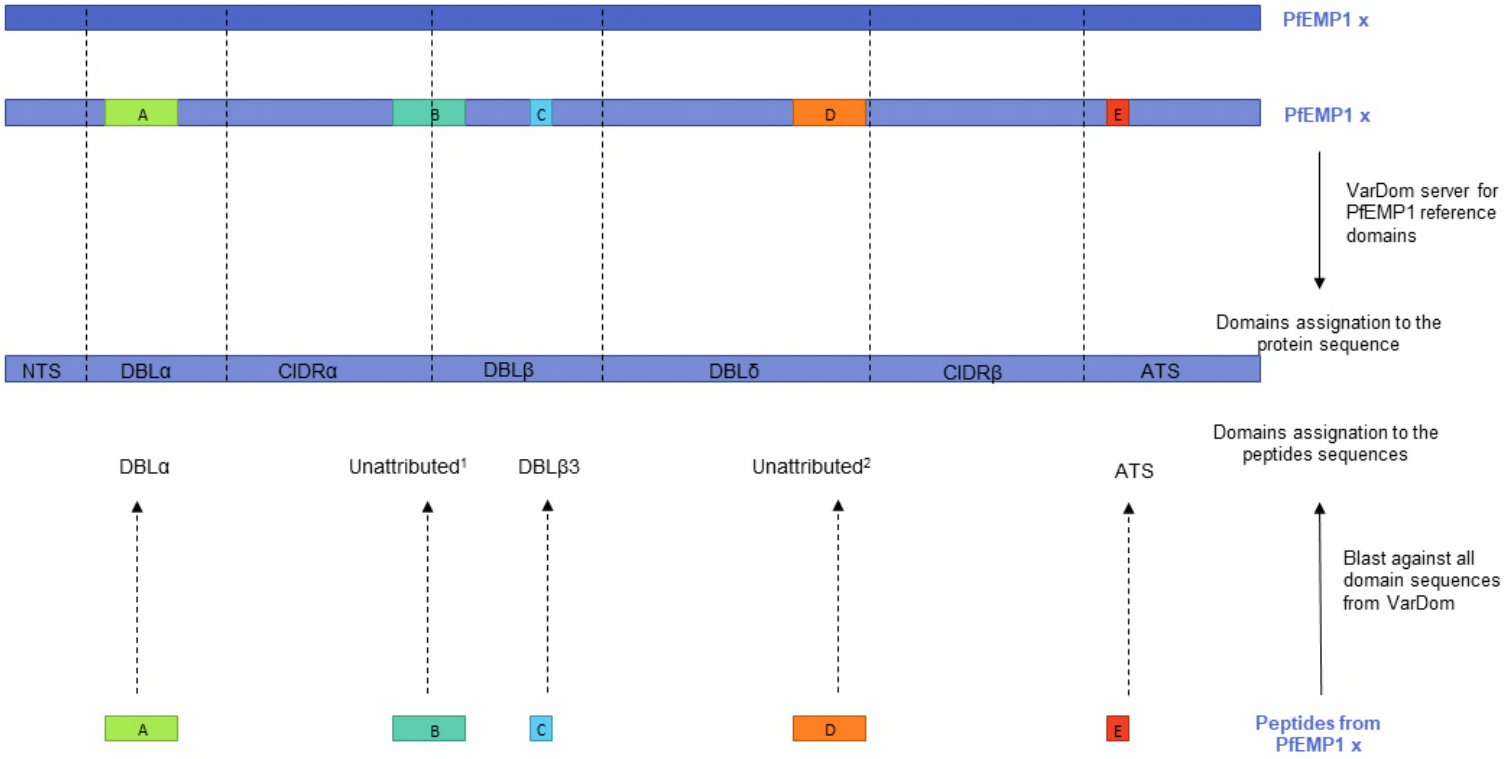
714 ENA accession numbers from genomic datafiles

715

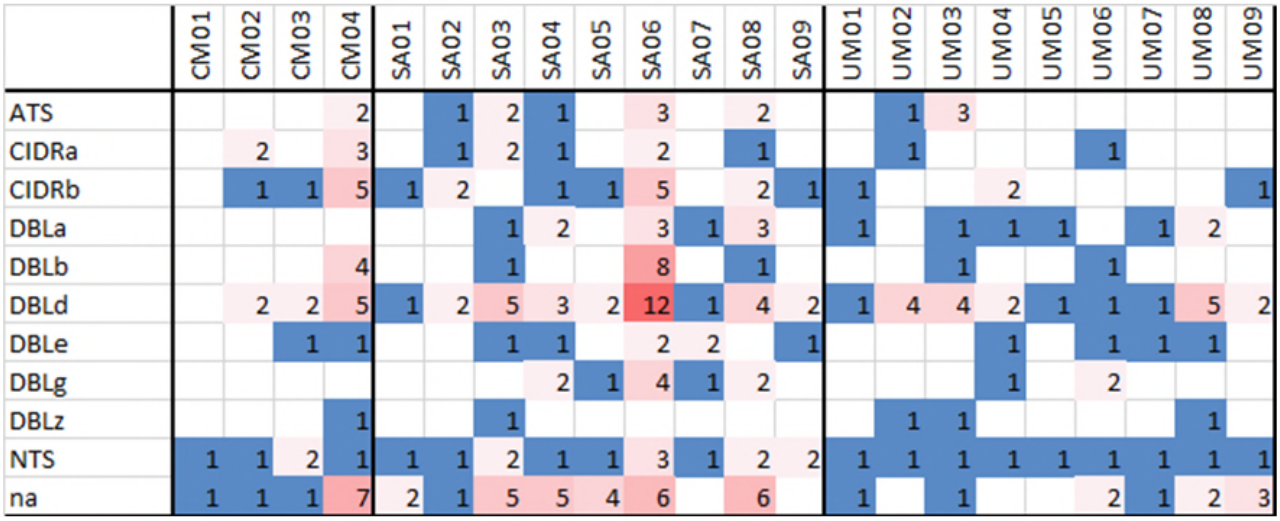
716

717


Field study - patients inclusion
Whole blood sampling
RDT
Thick blood smear
Biological analysis
White blood cell depletion
Infected erythrocytes
Pellet ring staged parasites
in vitro maturation
Mature parasite forms
DNA extraction
Whole Genome sequencing
var reconstruction
PfEMP1 sequences
RNA extraction
cDNA
Quantification
/confirmation of proteins
Selective primer design
Protein extraction
Digestion
LC-MS/MS
Peptide identification
Database construction
Public repositories
+
« Home » sequences
Protein identification in samples
Domain attribution

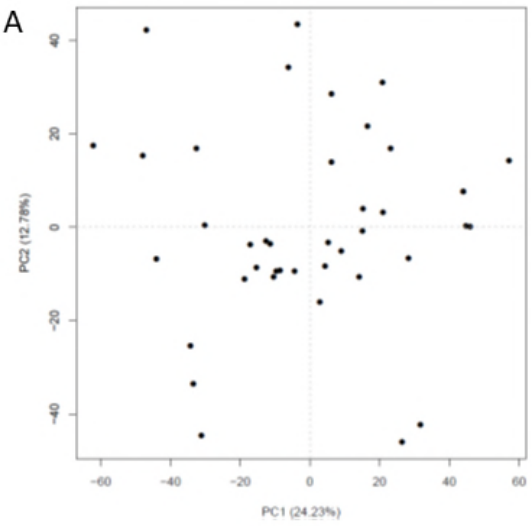

PfEMP1 x
PfEMP1 x
A
B
C
D
E
VarDom server for PfEMP1 reference domains
Domains assignation to the protein sequence
NTS
DBLα
CIDRα
DBLβ
DBLδ
CIDRβ
ATS
Domains assignation to the peptides sequences
DBLα
Unattributed¹
DBLβ3
Unattributed²
ATS
Blast against all domain sequences from VarDom
A
B
C
D
E
Peptides from PfEMP1 x

sequence id	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	D16	D17	D18	D19	D20	D21
QT0009-CW4.g3425	NTS	DBLα	CIDRα	DBLδ	CIDRβ																
574984302	NTS	DBLα	CIDRα	DBLβ	DBLδ	CIDRγ															
069H_H00-C.g949	DBLβ	DBLβ	DBLδ																		
069H_H00-C.g150	CIDRα	DBLβ	DBLγ	DBLδ	CIDRγ																
061H_H16-C.g486	NTS	DBLα	CIDRα	DBLβ	DBLγ	DBLδ	CIDRγ														
069H_H08-C.g1025	DBLα	CIDRγ	DBLβ	DBLγ	DBLγ	DBLγ	DBLγ														
910271409	NTS	DBLα	CIDRα	DBLδ	CIDRγ	ATS															
QT0032-CM.g2324	ATS																				
QT0017-C.g13	NTS	DBLα	CIDRα	DBLβ	DBLγ	DBLz	DBLε														
RAJ116var08	NTS	DBLα	CIDRα	DBLβ	DBLγ	DBLε	DBLε	ATS													
074H_H24-C.g522	CIDRα																				
064H_H08-C.g700	NTS	DBLα	CIDRα	DBLδ	CIDRβ																
QT0015-CW2.g2888	ATS																				
033H_H08-C.g192	NTS	DBLα	CIDRα	DBLβ	DBLδ	CIDRβ															
572SJ-C.g456	NTS	DBLα	CIDRα	DBLβ	DBLγ	DBLγ	DBLδ	CIDRβ													
049H_H24-C.g515	NTS	DBLα	CIDRα	DBLβ	DBLβ	DBLγ	DBLγ	DBLδ	CIDRβ	ATS											
QT0036-CW.g1117	NTS	DBLα	CIDRα	DBLβ	DBLβ	DBLγ	DBLδ	CIDRβ													
IT4var15	NTS	DBLα	CIDRα	DBLβ	DBLδ	CIDRβ	ATS														
056H_H24-C.g258	NTS	DBLα	CIDRα	DBLβ	DBLγ	DBLγ															
069H_H00-C.g45	DBLγ	DBLδ	CIDRβ																		
025H_H24-C.g225	ATS																				
QT0027-CM.g6369	DBLα	CIDRα	DBLβ	DBLβ	DBLγ	DBLδ	CIDRβ														
024H_H00-C.g52	NTS	DBLα	CIDRα	DBLβ	DBLγ	DBLδ	CIDRβ														
QT0031-CW2.g3676	NTS	DBLα	CIDRα	DBLβ	DBLγ	DBLδ	CIDRβ	DBLβ	DBLγ												
PF07_0048	NTS	DBLα	CIDRα	DBLδ	CIDRβ	ATS															
QT0001-C.g391	NTS	DBLα	CIDRα	DBLβ	DBLδ	CIDRγ	DBLε	DBLz	DBLε	DBLβ	DBLδ	DBLε	DBLγ	DBLz	DBLε	ATS					
QT0023-C.g24	NTS	DBLα	CIDRα	DBLδ	CIDRβ	NTS	DBLα	CIDRα	DBLβ	DBLδ	CIDRβ	NTS	DBLα	CIDRα	DBLδ	DBLγ	DBLε	DBLγ	DBLz	DBLε	ATS
QT0012-CW2.g612	NTS	DBLα	CIDRα																		
069H_H00-C.g1030	NTS	DBLα	CIDRα	DBLβ	DBLγ	DBLδ	CIDRβ	DBLβ	DBLγ												
QT0032-CM.g2637	NTS	DBLα	CIDRα	DBLβ	DBLδ	CIDRγ															
DD2var28	NTS	DBLα	CIDRα	DBLβ	DBLγ	DBLε	ATS														
060H_H08-C.g192	NTS	DBLα	CIDRα	DBLβ	DBLγ	DBLz	DBLε														
574984185	NTS	DBLα	CIDRα	DBLβ	DBLγ	DBLz	DBLε	ATS													
029H_H00-C.g187	NTS	DBLα	CIDRα	DBLδ	CIDRβ																
712SJ-C.g117	NTS	DBLα	CIDRα	DBLβ	DBLγ	DBLz															
024H_H08-C.g133	NTS	DBLα	CIDRα	DBLδ	CIDRβ	DBLz	DBLε														
074H_H08-C.g299	CIDRβ																				
QT0002-C.g12	NTS	DBLα	CIDRα	DBLδ	CIDRβ																
PF07_0050	NTS	DBLα	CIDRα	DBLβ	DBLγ	ATS															
DD2var39	NTS	DBLα	CIDRα	DBLδ	CIDRβ	ATS															
047H_H08-C.g672	NTS	DBLα	CIDRα	DBLδ	CIDRβ																
610SJ-C.g1	ATS																				
DD2var29	NTS	DBLα	CIDRα	DBLδ	CIDRβ	DBLγ	ATS														
073H_H08-C.g576	NTS	DBLα	CIDRα	DBLγ	DBLδ	CIDRβ	DBLε	DBLz	ATS												
579318932	NTS	DBLα	CIDRα	DBLδ	CIDRβ	ATS															
QT0033-CW3.g12	DBLβ	DBLβ	DBLδ	CIDRγ																	
IGHvar26	NTS	DBLα	CIDRα	DBLβ	DBLγ	DBLγ	DBLδ	CIDRβ	ATS												
PFCLINvar21	CIDRβ	ATS																			
044H_H24-C.g315	ATS																				
069H_H00-C.g62	NTS	DBLα	CIDRα																		
021H_H00-C.g675	NTS	DBLα	CIDRδ	DBLβ	DBLβ	DBLγ	DBLδ	CIDRβ													
QT0002-C.g9	NTS	DBLα	CIDRα	DBLβ	DBLγ	DBLδ	CIDRβ	DBLz	DBLε												
065H_H16-C.g1122	NTS	DBLα	CIDRα	DBLβ	DBLδ	CIDRγ	DBLγ	DBLz	ATS												

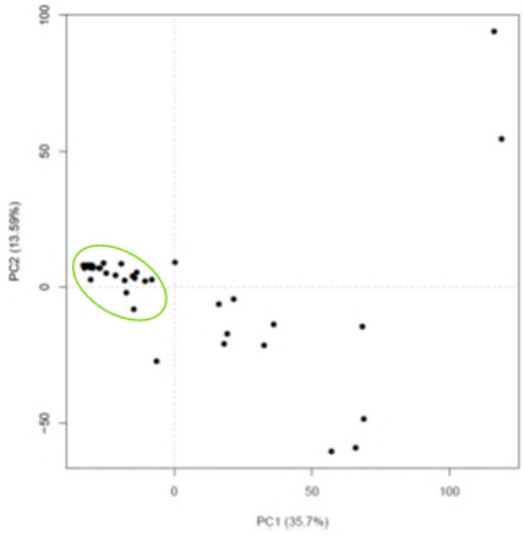



CM
2
8
14
11
SA
6
UM
18
5

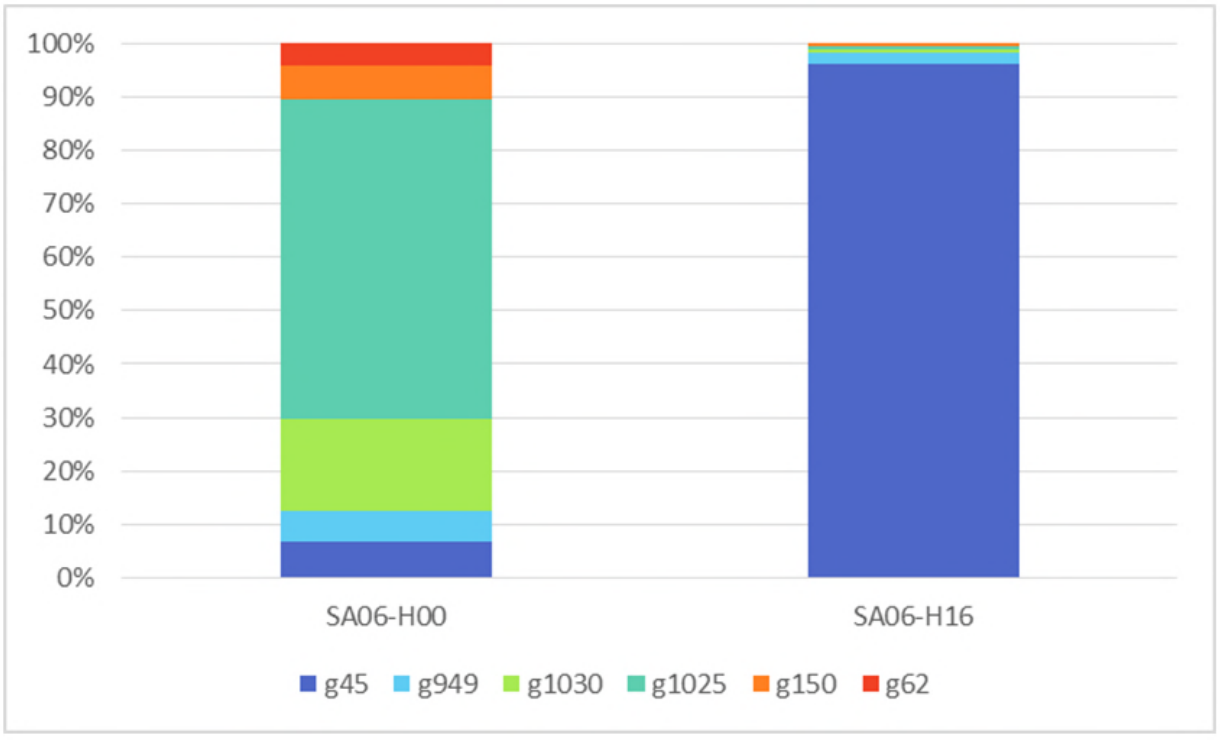
A


PC2 (12.78%)
PC1 (24.23%)
40
20
0
−20
−40
−60
−40
−20
0
20
40
60

B


PC2 (13.59%)
PC1 (35.7%)
100
50
0
−50
0
50
100

C


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
SA06-H00
SA06-H16
g45
g949
g1030
g1025
g150
g62

Sequence ID	Origin	Phenotype	Domains combination																		
UM																					
049H_H24-C.g515	Field	na	NTS	DBLα	CIDRα	DBLβ	DBLβ	DBLγ	DBLγ	DBLδ	CIDRβ										
QT0036-CW.g1117	Field	na	NTS	DBLα	CIDRα	DBLβ	DBLβ	DBLγ	DBLδ	CIDRβ											
024H_H08-C.g133	Field	na	NTS	DBLα	CIDRα	DBLδ	CIDRβ	DBLz	DBLε												
044H_H24-C.g315	Field	na	ATS																		
065H_H16-C.g1122	Field	na	NTS	DBLα	CIDRα	DBLβ	DBLδ	CIDRγ	DBLγ	DBLz	ATS										
CM																					
074H_H08-C.g299	Field	na	CIDRβ																		
PF07_0050	PlasmoDB	na	NTS	DBLα0.18	CIDRα6	DBLβ6	DBLγ9	ATS													
SA																					
QT0009-CW4.g3425	Field	na	NTS	DBLα	CIDRα	DBLδ	CIDRβ														
574984302	NCBI	na	NTS	DBLα	CIDRα	DBLβ	DBLδ	CIDRγ													
069H_H00-C.g949	Field	na	DBLβ	DBLβ	DBLδ																
069H_H00-C.g150	Field	na	CIDRα	DBLβ3	DBLγ	DBLδ	CIDRγ														
061H_H16-C.g486	Field	na	NTS	DBLα	CIDRα	DBLβ	DBLγ	DBLδ	CIDRγ												
QT0017-C.g13	Field	na	NTS	DBLα	CIDRα	DBLβ	DBLγ	DBLz	DBLε												
572SJ-C.g456	Field	na	NTS	DBLα	CIDRα	DBLβ	DBLγ	DBLγ	DBLδ	CIDRβ											
IT4var15	VarDom	CD36	NTS	DBLα0.8	CIDRα3.5	DBLβ3	DBLδ1	CIDRβ1	ATS												
069H_H00-C.g45	Field	na	DBLγ	DBLδ	CIDRβ																
025H_H24-C.g225	Field	na	ATS																		
QT0001-C.g391	Field	na	NTS	DBLα	CIDRα	DBLβ	DBLδ	CIDRγ	DBLε	DBLz	DBLε	DBLβ	DBLδ	DBLε	DBLγ	DBLz	DBLε	ATS			
069H_H00-C.g1030	Field	na	NTS	DBLα	CIDRα	DBLβ	DBLγ	DBLδ	CIDRβ	DBLβ	DBLγ										
QT0032-CM.g2637	Field	na	NTS	DBLα	CIDRα	DBLβ	DBLδ	CIDRγ													
574984185	NCBI	na	NTS	DBLα	CIDRα	DBLβ	DBLγ	DBLz	DBLε	ATS											
610SJ-C.g1	Field	na	ATS																		
DD2var29	VarDom	CD36	NTS	DBLα0.16	CIDRα3.4	DBLδ1	CIDRβ1	DBLγ10	ATS												
073H_H08-C.g576	Field	na	NTS	DBLα	CIDRα	DBLγ	DBLδ	CIDRβ	DBLε	DBLz	ATS										
PFCLINvar21	VarDom		DBLδ1	CIDRβ1	ATS																
CM + SA																					
069H_H08-C.g1025	Field	na	DBLα	CIDRγ	DBLβ	DBLγ	DBLγ	DBLγ	DBLγ												
DD2var28	VarDom	?	NTS	DBLα0.4	CIDRα6	DBLβ5	DBLγ18	DBLε8	ATS												
029H_H00-C.g187	Field	na	NTS	DBLα	CIDRα	DBLδ	CIDRβ														
712SJ-C.g117	Field	na	NTS	DBLα	CIDRα	DBLβ	DBLγ	DBLz													
579318932	NCBI	na	NTS	DBLα	CIDRα	DBLδ	CIDRβ	ATS													
QT0033-CW3.g12	Field	na	DBLβ	DBLβ	DBLδ	CIDRγ															

069H_H00-C.g62	Field	na	NTS	DBLα	CIDRα					
021H_H00-C.g675	Field	na	NTS	DBLα	CIDRδ	DBLβ	DBLβ	DBLγ	DBLδ	CIDRβ

Legend

ICAM 1 binding +

reference sequence

Table 1 : Sequences identification and origin from the PfEMP1 sequences identified using mass spectrometry in field isolates from Benin.