MAJOR ARTICLE

Running title: Expression of malaria *clag3* genes

Expression of the *Plasmodium falciparum* clonally variant *clag3* genes in human infections

Sofía Mira-Martínez[1,2], Evi van Schuppen[2], Alfred Amambua-Ngwa[3], Emmanuel Bottieau[1], Muna Affara[3], Marjan Van Esbroeck[1], Erika Vlieghe[1], Pieter Guetens[1], Núria Rovira-Graells[2], Gloria P. Gómez-Pérez[2], Pedro L. Alonso[2], Umberto D'Alessandro[1,3,4], Anna Rosanas-Urgell[1,a] and Alfred Cortés[2,5, a]

[1] Institute of Tropical Medicine (ITM), 2000 Antwerp, Belgium

[2] ISGlobal, Barcelona Ctr. Int. Health Res. (CRESIB), Hospital Clínic - Universitat de Barcelona, 08036 Barcelona, Catalonia, Spain

[3] Medical Research Council Unit (MRC), Fajara, The Gambia

[4] London School of Hygiene and Tropical Medicine, WC1E 7HT London, UK

[5] ICREA, 08010 Barcelona, Catalonia, Spain

P.L.A., present address: WHO, CH-1211 Geneva, Switzerland

[a] A.R.-U. and A.C. contributed equally to this work

Correspondence: arosanas@itg.be (A.R.-U.), alfred.cortes@isglobal.org (A.C.).

Abstract word count: 200. **Manuscript word count** (Background to Discussion): 3495.

Summary: Many important *Plasmodium falciparum* genes show clonally variant expression, but little is known about how these genes are used during human malaria infections. This article reports the expression in human infections of the clonally variant *clag3* genes linked to solute uptake.

FOOTNOTE PAGE

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by Spanish Ministry of Economy and Competitiveness (MINECO) [SAF2013-43601-R to A.C.], co-funded by the European Regional Development Fund (ERDF, European Union); the Secretary for Universities and Research under the Department of Economy and Knowledge of the Government of Catalonia [2014 SGR 485 to A.C.]; and ITM funding to A.R.U. ISGlobal receives support from the CERCA Programme (Government of Catalonia). Funding for manufacture, quality control release and stability studies of Sanaria's PfSPZ Challenge product (CHMI) was provided in part by the National Institute of Allergy and Infectious Disease [R44AI058375] 'Universal Attenuated Malaria Sporozoite Vaccine and Challenge System'. S.M.M. holds a TransGlobalHealth - Erasmus Mundus Joint Doctorate Programme scholarship (European Community).

Presented at meetings

Presented in part at the 11th BioMalPar Conference, Heidelberg, Germany, 11–13 May 2015 (abstract 111) and 13th MEEGID Conference, Institute of Tropical Medicine, Antwerp, Belgium, 10–13 May 2016.

Corresponding authors contact information

A. Cortés, CEK building, c./ Rosselló 153, 08036 Barcelona, Catalonia, Spain (+34 932275400, alfred.cortes@isglobal.org); A. Rosanas-Urgell, Nationalestraat 155, 2000 Antwerp, Belgium (+32 32476354, arosanas@itg.be).

ABSTRACT

Background. Many genes of the malaria parasite *Plasmodium falciparum* show clonally variant expression regulated at the epigenetic level. These genes participate in fundamental host-parasite interactions and contribute to adaptive processes. However, little is known about their expression patterns during human infections. A peculiar case of clonally variant genes are the two nearly identical *clag3* genes, *clag3.1* and *clag3.2*, which mediate nutrient uptake and are linked to resistance to some toxic compounds.

Methods. We developed a procedure to characterize the expression of *clag3* genes in naturally-infected malaria patients and in experimentally-infected human volunteers.

Results. We provide the first description of *clag3* expression during human infections, which revealed mutually exclusive expression and identified the gene predominantly expressed. Adaptation to culture conditions or selection with a toxic compound resulted in isolate-dependent changes in *clag3* expression. We also found that *clag3* expression patterns are reset during transmission stages.

Conclusions. Different environment conditions select for parasites with different *clag3* expression patterns, implying functional differences between the proteins encoded. The epigenetic memory is likely erased before parasites start infection of a new human host. Altogether, our findings support the idea that clonally variant genes facilitate the adaptation of parasite populations to changing conditions through bet-hedging strategies.

KEYWORDS

Malaria; *Plasmodium falciparum*; transcription; epigenetics; adaptation; bet-hedging; controlled human malaria infection (CHMI); mutually exclusive gene expression; transcriptional variation; *clag3*

BACKGROUND

Plasmodium falciparum is responsible for the most severe forms of malaria. Asexual growth of the parasites in the human blood is responsible for all clinical symptoms and also for chronic infection. During the ~48 h asexual multiplication cycle parasites live inside of human erythrocytes except for the short time between bursting of parasites at the schizont stage and invasion of new erythrocytes [1]. While the human blood is a relatively stable environment, parasites need to adapt to fluctuating conditions such as nutrient concentrations, presence of drugs, occurrence of fever episodes, or immune responses. Recent studies have demonstrated that populations of genetically identical parasites show extensive transcriptional heterogeneity [2], which potentially allows adaptation by dynamic natural selection of parasites with transcriptional patterns associated with increased fitness as the environment changes. This type of adaptive strategies, known as bet-hedging, is used by many microbial organisms [3-5]. Genes that can be found in either an active or a silenced state in genetically identical parasites at the same stage of cycle progression, known as clonally variant genes, participate in multiple biological pathways involved in fundamental host-parasite interactions [2, 6, 7]. The silenced or active state of these genes is transmitted from one generation to the next by epigenetic mechanisms [8-10]. Switches between the two alternative states of these genes occur spontaneously, albeit with low frequency, allowing for the constant generation of transcriptional diversity within parasite populations.

In spite of the large number of families of clonally variant genes identified in *P. falciparum*, there are few for which an adaptive role or an association between the transcriptional state of specific genes and the resulting phenotypes has been clearly established. The best characterized family of clonally variant genes is *var*, which consists of about 60 genes per genome encoding PfEMP-1, a major virulence factor linked to cytoadherence and antigenic

variation. *var* genes show mutually exclusive expression, such that a single parasite typically expresses only one *var* gene at a time and keeps all the others silenced [11]. Spontaneous switches in the expression of *var* genes play an adaptive role, mediating immune evasion and altering the sequestration tropism of infected erythrocytes [12].

A second case of *P. falciparum* clonally variant genes for which transcriptional switches have been associated with specific phenotypes and adaptation to changes in the environment is *clag3*. The two *clag3* genes, *clag3.1* (PF3D7_0302500) and *clag3.2* (PF3D7_0302200), are separated by only 10 kb and show 95% sequence coincidence. The *clag3.1* or *clag3.2* identity is determined by the relative position of each gene in the chromosome and by the conserved *clag3.1*- or *clag3.2*-specific flanking regions [13]. These genes are part of the five-member *clag* family, which encodes the CLAG/RhopH1 component of the RhopH complex. While early reports linked CLAG proteins with erythrocyte invasion or cytoadherence [14], more recent research has provided strong genetic and biochemical evidence for a key role of CLAG3 proteins in the formation of the Plasmodial Surface Anion Channel (PSAC), a broad selectivity channel that mediates the uptake of nutrients and several other solutes at the infected erythrocyte membrane [14-17]. CLAG3 proteins are validated drug targets [16]. Epigenetic silencing of *clag3* genes is mediated by formation of heterochromatin, similar to other clonally variant genes [10, 18]. Together with the *var* family, *clag3* is the only known example of mutually exclusive expression in *P. falciparum* [19]. However, while under normal conditions the vast majority of parasites express only one of the two *clag3* genes at a time, mutually exclusive expression is not strict, which allows for the occurrence of small selectable subpopulations of parasites with alternative expression patterns that enable additional phenotypic plasticity [20]. We and others have recently demonstrated that a compound that is toxic for the parasite, blasticidin S (BS), can select for low-abundance

subpopulations of parasites with both *clag3* genes simultaneously silenced [21, 22]. However, lower concentrations of the drug select for parasites that express a specific paralog, which is suggestive of phenotypic differences associated with expression of one or the other *clag3* gene [22]. Altogether, these results indicate that *clag3* expression patterns determine the permeability phenotype of infected erythrocytes and can mediate drug resistance at the epigenetic level. Variant expression of these genes needs to fulfill two competing requirements: efficiently acquiring nutrients, and restricting the entrance of harmful compounds.

While the expression patterns of *var* genes in field isolates have been the subject of intensive investigation [12, 23-26], very little is known about the expression of *clag3* genes during human infections. Previous studies of *clag3* expression were conducted with culture-adapted parasites. Genome-wide transcriptomic analysis of *P. falciparum* field isolates [27-29] could not reliably characterize the expression patterns of these genes because the sequences of *clag3.1* and *clag3.2* are almost identical and the regions that are more distinct between the two genes are highly polymorphic between parasite isolates [13]. Here we developed a procedure to analyze *clag3* expression in natural human infections and applied it to study the expression of these genes in clinical malaria cases and after parasite adaptation to culture conditions or to BS pressure. We also studied *clag3* expression in experimental human malaria infections.

METHODS

Ethics approval

This study was approved by the Institute of Tropical Medicine review board (ITM IRB) (ITG913/13), University Hospital of Antwerp review board (UZA IRB) (B300201319284),

and The Gambia Government/MRC Joint Ethics Committee (SCC1392). Approval for the controlled human malaria infection (CHMI) trial has been previously described [30]. All participants provided written informed consent before enrolment. The study was conducted according to the principles stated in the Declaration of Helsinki.

Sample collection

P. falciparum clinical malaria blood samples were obtained from returning travelers (≥18 years old) attending clinics in Antwerp (Belgium), and from children (≤12 years old) attending Health Centers in The Gambia. We also analyzed blood samples collected from volunteers participating in a CHMI study [30]. Additional details of sample collection and processing are provided in the Supplemental Methods.

Genetic and transcriptional analysis

Multiplicity of infection was estimated by genotyping the *msp1* and *msp2* loci [31]. To assess for recombination events between the two *clag3* genes, the *clag3* loci were analyzed by long PCR [13]. For each isolate, we sequenced the hypervariable region (HVR) [13] of each *clag3* gene from the long PCR products to design gene- and isolate-specific primers.

To prepare RNA for *clag* transcriptional analysis, cultures were harvested when the majority of parasites were at the schizont stage. For natural infections, parasites were cultured only until they reached the schizont stage, with the exception of two samples (see Supplemental Methods). Day 9 and day of malaria samples from the CHMI were cryopreserved and, after thawing, parasites were cultured for 2-3 weeks or ~1 week, respectively, until they reached a ≥0.02% parasitaemia.

RNA was purified approximately as described [32]. The protocol was validated for use at very low parasitaemias (Supplemental Fig. S1). Quantitative PCR analysis was performed using the standard curve method approximately as described [10] with primers listed in Supplemental Table S1.

All new sequences obtained in this study have been deposited to GenBank with accession numbers KY092485-KY092488 (full sequences) and KY364642-KY364689 (HVR sequences). Additional details of the methods for the genetic and transcriptional analysis and also for the CLAG3 sequence analysis can be found in the Supplemental Methods.

RESULTS

Parasites predominantly express *clag3.2* in natural malaria infections

We analyzed *clag3* expression patterns in blood from *P. falciparum*-infected symptomatic patients. Parasites were cultured *ex vivo* until they reached the schizont stage, when *clag3* genes are expressed, and harvested for genomic DNA and RNA extraction (Fig. 1A). To reduce the complexity of the analysis, only samples with a single infection or a clearly predominant clone were retained for *clag3* expression characterization. Isolates presenting a single *clag3* gene in their genome as a consequence of recombination [13] were excluded. For the twenty remaining samples, we sequenced the HVR of the two genes to design isolate-specific primers for *clag3.1* and for *clag3.2*, and used them to analyze *clag3* expression by reverse transcription followed by quantitative PCR. All isolates show predominant expression of one of the two *clag3* paralogs (Fig. 1B), consistent with the mutually exclusive expression observed in culture-adapted parasite lines [15, 16, 18, 19]. However, in contrast to the majority of culture-adapted lines, which predominantly express *clag3.1* [15, 18], we observed predominant *clag3.2* expression in all isolates (Fig. 1B).

Adaptation of field isolates to culture conditions or to low drug pressure is associated with isolate-dependent changes in *clag3* expression

Two of the field isolates, P04 and P12, were maintained under culture conditions for 17 or 13 weeks, respectively, with regular monitoring of *clag3* expression (Fig. 2A). In the P12 isolate, parasites expressing *clag3.2* were progressively replaced by parasites expressing *clag3.1*. In contrast, in the P04 isolate the majority of parasites maintained *clag3.2* expression throughout the experiment. Next we selected the same isolates with a sub-lethal concentration of BS that in the 3D7 genetic background selects for parasites that express *clag3.1* [22]. In the P04 isolate, parasites expressing *clag3.1* were quickly selected and by 10 weeks of selection they had almost completely displaced *clag3.2*-expressing parasites (Fig. 2B). However, BS had the opposite effect on the P12 isolate and rather favored survival of parasites expressing *clag3.2*: while normal culturing of this isolate resulted in selection of *clag3.1*-expressing parasites (Fig. 2A), this did not occur in the presence of BS (Fig. 2B). These results indicate that the selective advantage conferred by expression of one or the other *clag3* paralog under different conditions is isolate-specific and likely depends on the *clag3* genes sequences.

Analysis of *clag3* sequences identifies paralog-specific and promiscuous sequence features

To gain insight into the sequence determinants of the fitness advantages conferred by expression of one or the other protein, we analyzed publicly available CLAG3 sequences and the newly obtained sequences from the P04 and P12 isolates. We identified a sequence feature at the N-terminal end of the protein (positions 10-23) that is different between CLAG3.1 and CLAG3.2 but is conserved for each protein among all isolates analyzed (Fig.

3A). This hydrophobic region, which likely corresponds to either a signal peptide or a transmembrane domain, is a candidate for determining general CLAG3 properties that under normal human blood conditions confer a selective advantage to parasites expressing CLAG3.2. At other positions, polymorphisms are not paralog-specific, such that the same sequence can be found in CLAG3.1 or in CLAG3.2 in different isolates. This is probably attributable to frequent recombination and gene conversion events between the two paralogs (Supplemental Fig. S2A). However, clustering analysis of the full CLAG3 sequences confirmed that CLAG3.1 and CLAG3.2 separate into discrete clades (Fig. 3B), as previously reported [21]. This is a consequence of several polymorphisms occurring at very different frequencies between the two paralogs. In contrast, clustering analysis of the HVR does not show separation of CLAG3.1 and CLAG3.2 sequences (Fig. 3C), which indicates that a CLAG3.1 HVR can be as similar to a CLAG3.2 HVR as to another CLAG3.1 HVR, and vice versa. HVR sequences separate in two distinct clades, both including CLAG3.1 and CLAG3.2 sequences (Fig. 3C and Supplemental Fig. S2B). Phenotypic traits such as resistance to BS that are associated with *clag3.1* expression in some isolates and with *clag3.2* expression in others likely depend on polymorphism at the HVR or other positions where polymorphism is not paralog-specific.

clag3 expression in parasites obtained from experimentally infected humans

We analyzed *clag3* expression patterns in parasites collected from human volunteers participating in a CHMI trial [30] in which cryopreserved sporozoites of the culture-adapted line NF54 (from which 3D7 was derived) were used for infection (Fig. 4A). In samples collected from six different volunteers when parasites were first detected by microscopy (day of malaria, 12-15 days post-injection), we observed higher levels of *clag3.2* transcripts compared with *clag3.1*, in contrast to the parental NF54 line used to infect the mosquitoes

that almost exclusively expresses *clag3.1* (Fig. 4B). Next we re-adapted to culture conditions the parasites obtained from two of the volunteers, and in both cases we observed a progressive increase in the ratio of *clag3.1* to *clag3.2* transcripts, reflecting selection of parasites expressing *clag3.1* (Fig. 4C). This result indicates that *clag3.1* expression confers a fitness advantage under culture conditions in the NF54 genetic background. Similar levels of transcripts for both *clag3* genes in volunteer samples (Fig. 4B) imply that they likely consist of a mixture of parasites expressing *clag3.2* and parasites expressing *clag3.1*, rather than a homogeneous population of individual parasites expressing the two genes simultaneously. This idea is based on previous results with culture-adapted parasite lines with similar transcript levels of *clag3.1* and *clag3.2* [19, 20] and is also supported by the analysis of *clag3* expression in subclones of one of the volunteer samples (Supplemental Fig. S3).

The expression patterns of *clag3* genes are reset during transmission stages

The *clag3* expression patterns observed in parasites obtained from experimentally infected volunteers can be explained by two non-exclusive scenarios: *clag3.2*-expressing parasites are selected under the conditions of the human circulation, or there is a reset of *clag3* expression patterns during transmission stages. To distinguish between the two possibilities, we analyzed *clag3* expression in blood samples from two of the volunteers collected at day 9 post sporozoite injection. Considering that parasite liver stage development takes 6-7 days [33], parasites collected at day 9 had been in the peripheral blood for only about one multiplication cycle. In spite of this, we observed similar levels of transcripts for both *clag3* genes at day 9 (Fig. 4D), a pattern similar to that observed at the day of malaria. This result is inconsistent with a 'blood stages selection only' scenario and supports the idea that *clag3* expression patterns are reset when parasites go through transmission stages. Because parasite densities were extremely low at day 9 post injection, parasites had to be cultured for 2-3 weeks before

we could obtain sufficient material for transcriptional analysis. However, this is unlikely to be a confounder for these results because in parasites of the NF54 genetic background, culture conditions progressively select for parasites that express *clag3.1* (Fig. 4C). Hence, at day 9 post-injection the parasites population contains a large proportion of parasites expressing *clag3.2*, a composition that is clearly distinct from the parental NF54 line.

Transcript levels of *clag2, clag8* and *clag9* did not show major differences among isolates or between different growth conditions

We also analyzed the expression of *clag2* (PF3D7_0220800), *clag8* (PF3D7_0831600) and *clag9* (PF3D7_0935800) in all samples described in this study. There was little variation in the transcript levels of these genes (Supplemental Fig. S4-8). Even *clag2*, which shows clonally variant expression in culture-adapted lines [2, 19], was expressed at similar levels in all samples.

DISCUSSION

Variantly expressed genes play key roles in malaria host-parasite interactions and contribute to parasite adaptation to changes in its environment, but little is known about the expression patterns of these genes during human infections. This is an important limitation because the expression of clonally variant genes in a population of parasites is shaped by the environment, and the environment is different between culture conditions and the natural conditions of the human blood circulation. Here we characterized the expression in human infections of the clonally variant *P. falciparum clag3* genes, which provide one of the best models in malaria to study functional variation linked to epigenetic switches. We observed differences from the *clag3* expression patterns commonly observed under culture conditions, but expression conformed with the mutual exclusion principle previously described in

cultured parasites. By comparing expression patterns in the same parasite lines between human circulation and culture conditions, or challenging parasites with a toxic compound, we observed that different environments dynamically select for parasites with different patterns of *clag3* expression in an isolate-dependent manner. These results support the idea that transcriptional variation and bet-hedging strategies play an important role in malarial adaptation.

In twenty clinical malaria infections, parasites predominantly expressed *clag3.2*, in contrast to most culture-adapted parasite lines that predominantly express *clag3.1* [15, 18]. This result suggests that under the conditions of the human circulation, with lower concentrations of most nutrients than in the regular parasite culture medium [16], expression of *clag3.2* confers a growth advantage. Whether different clinical presentation (e.g. asymptomatic or severe malaria), host malnutrition, exposure to drugs, or other conditions are associated with different *clag3* expression patterns remains to be determined. Together with previous observations showing that in the 3D7 genetic background expression of *clag3.1* appears to restrict the entrance of the toxic compound BS [22], this result may suggest that the PSAC resulting from *clag3.1* expression mediates less efficient solute uptake than the PSAC resulting from *clag3.2* expression. However, we found that BS pressure or growth under culture conditions select for parasites expressing a different *clag3* gene in isolates of different genetic background, revealing a more complex scenario. Predominant expression of *clag3.2* during clinical malaria infections was the only observation common to all isolates, which suggests that the advantage conferred by *clag3.2* expression in this type of infections depends on characteristics unique to the CLAG3.2 protein, such as the conserved sequence feature identified at its N-terminus. On the other hand, phenotypic characteristics that in different isolates are associated with the expression of a different *clag3* paralog may depend on non-

conserved CLAG3 sequence features that in some isolates occur in CLAG3.1 and in others in CLAG3.2. In any case, considering that natural selection only operates on phenotypes, our culture adaptation and BS selection experiments (together with our previous studies using BS selection of a culture-adapted line [22]) clearly demonstrate that expression of one or the other *clag3* paralog results in phenotypic differences. This is remarkable considering that the two CLAG3 proteins have nearly identical sequences. These phenotypic differences are likely linked to infected erythrocyte permeability, although we cannot exclude the possibility that they also involve processes such as cytoadherence or erythrocyte invasion in which CLAG3 proteins may also play a role [14].

Mutually exclusive expression is a phenomenon that affects gene families of utmost importance in several pathogens [34]. In *P. falciparum*, it has been observed for *var* [11] and *clag3* [15, 16, 18, 19] genes in culture-adapted parasites, although it was found not to be strict: for both gene families, single-cell analysis or strong selection applied to cultures revealed the existence of small parasite subpopulations that do not conform with mutually exclusive expression patterns [20, 22, 35-37]. Considering that the selective pressures operating on parasites in the human blood circulation are different from those under culture conditions, this raises the formal possibility that mutual exclusion may not be the most common pattern in human infections. By focusing only on single infections and using isolate-specific *clag3.1* and *clag3.2* primers, here we provide evidence for predominant mutually exclusive expression in *P. falciparum* genes during natural infections. In the majority of isolates, *clag3.2* transcript levels were more than 10-fold higher than *clag3.1* levels, although residual expression of the latter was observed in all cases. Residual expression of the silenced paralog is also observed in clonal culture-adapted parasite lines and likely corresponds to small subpopulations of parasites that spontaneously switch the active *clag3* at each cycle of

multiplication. The existence of these subpopulations of parasites with alternative expression patterns is essential to enable natural selection when changes in host conditions occur.

We compared *clag3* expression between blood stage parasites obtained from infected volunteers and the parental cultured parasite line used for the infections. An analogous approach has been previously used to study the expression of *var* genes in the context of a human infection, which revealed a reset of the expression patterns of this gene family during transmission stages [38-40]. Here we show that the expression of *clag3* genes is also reset. This result strongly suggests that the epigenetic memory for the expression of *clag3* genes is erased during gametocyte, mosquito or liver stages and stochastically re-established before the onset of a new blood infection, thus providing support to the idea that the epigenetic memory for the expression of clonally variant genes in general is erased during transmission stages, rather than only the epigenetic memory for the peculiar *var* family. An alternative explanation would be that the reset of *clag3* expression depends on selection of parasites expressing a specific *clag3* gene during one or more of the transmission stages, but we consider this an unlikely possibility because such selection seems incompatible with the relatively small parasite population sizes and few multiplications cycles occurring during transmission stages [41]. Furthermore, no function has been described for CLAG3 proteins outside the asexual cycle. The idea that mosquito passage resets the epigenetic patterns for virulence genes has also been proposed for *P. chabaudi* [42]. Erasing the epigenetic memory and releasing a transcriptionally diverse population of parasites at the onset of a blood infection is an intuitively advantageous strategy for the parasite to ensure the survival of the population in a new human host with unpredictable conditions.

Altogether, our results support the idea that variant expression of *clag3* genes plays an important adaptive role, and provide the first insight into how these genes are used under the natural conditions of a human infection.

FUNDING

This work was supported by Spanish Ministry of Economy and Competitiveness (MINECO) [SAF2013-43601-R to A.C.], co-funded by the European Regional Development Fund (ERDF, European Union); the Secretary for Universities and Research under the Department of Economy and Knowledge of the Government of Catalonia [2014 SGR 485 to A.C.]; and ITM funding to A.R.U. ISGlobal receives support from the CERCA Programme (Government of Catalonia). Funding for manufacture, quality control release and stability studies of Sanaria's PfSPZ Challenge product (CHMI) was provided in part by the National Institute of Allergy and Infectious Disease [R44AI058375] 'Universal Attenuated Malaria Sporozoite Vaccine and Challenge System'. S.M.M. holds a TransGlobalHealth - Erasmus Mundus Joint Doctorate Programme scholarship (European Community).

ACKNOWLEDGMENTS

We thank the patients and clinical staff at ITM and UZA (Belgium) and MRC Unit The Gambia for their cooperation, and the volunteers and clinical staff who participated in the CHMI in Barcelona. We are also grateful to José Muñoz (ISGlobal) for clinical support in the CHMI trial, to Alfredo Mayor and Ariel Magallón-Tejada (ISGlobal) for their contribution to the collection of samples for transcriptional analysis in the CHMI study, to Stephen L. Hoffman and Kim Lee Sim (Sanaria) for providing cryopreserved *P. falciparum* sporozoites (PfSPZ Challenge), and to Conor Meehan (ITM) for assistance with phylogenetic analysis.

We thank Jacqueline E. Broerse (VU Amsterdam) for useful discussion and co-supervision of SMM.

REFERENCES

1. Miller LH, Baruch DI, Marsh K, Doumbo OK. The pathogenic basis of malaria. Nature **2002**; 415:673-9.

2. Rovira-Graells N, Gupta AP, Planet E, et al. Transcriptional variation in the malaria parasite *Plasmodium falciparum*. Genome Res **2012**; 22:925-38.

3. Veening JW, Smits WK, Kuipers OP. Bistability, epigenetics, and bet-hedging in bacteria. Annu Rev Microbiol **2008**; 62:193-210.

4. Starrfelt J, Kokko H. Bet-hedging--a triple trade-off between means, variances and correlations. Biol Rev Camb Philos Soc **2012**; 87:742-55.

5. Levy SF, Ziv N, Siegal ML. Bet hedging in yeast by heterogeneous, age-correlated expression of a stress protectant. PLoS Biol **2012**; 10:e1001325.

6. Voss TS, Bozdech Z, Bartfai R. Epigenetic memory takes center stage in the survival strategy of malaria parasites. Curr Opin Microbiol **2014**; 20:88-95.

7. Cortés A, Deitsch K. Malaria Epigenetics. Cold Spring Harb Perspect Med **2017**; in press, doi:10. 1101/CSHPERSPECT.a025528.

8. Lopez-Rubio JJ, Gontijo AM, Nunes MC, Issar N, Hernandez Rivas R, Scherf A. 5' flanking region of *var* genes nucleate histone modification patterns linked to phenotypic inheritance of virulence traits in malaria parasites. Mol Microbiol **2007**; 66:1296-305.

9. Jiang L, Lopez-Barragan MJ, Jiang H, et al. Epigenetic control of the variable expression of a *Plasmodium falciparum* receptor protein for erythrocyte invasion. Proc Natl Acad Sci USA **2010**; 107:2224-9.

10. Crowley VM, Rovira-Graells N, de Pouplana LR, Cortés A. Heterochromatin formation in bistable chromatin domains controls the epigenetic repression of clonally variant *Plasmodium falciparum* genes linked to erythrocyte invasion. Mol Microbiol **2011**; 80:391-406.

11. Scherf A, Lopez-Rubio JJ, Riviere L. Antigenic variation in *Plasmodium falciparum*. Annu Rev Microbiol **2008**; 62:445-70.

12. Smith JD, Rowe JA, Higgins MK, Lavstsen T. Malaria's deadly grip: cytoadhesion of *Plasmodium falciparum*-infected erythrocytes. Cell Microbiol **2013**; 15:1976-83.

13. Iriko H, Kaneko O, Otsuki H, et al. Diversity and evolution of the *rhoph1/clag* multigene family of *Plasmodium falciparum*. Mol Biochem Parasitol **2008**; 158:11-21.

14. Gupta A, Thiruvengadam G, Desai SA. The conserved *clag* multigene family of malaria parasites: Essential roles in host-pathogen interaction. Drug Resist Updat **2015**; 18:47-54.

15. Nguitragool W, Bokhari AA, Pillai AD, et al. Malaria Parasite *clag3* Genes Determine Channel-Mediated Nutrient Uptake by Infected Red Blood Cells. Cell **2011**; 145:665-77.

16. Pillai AD, Nguitragool W, Lyko B, et al. Solute Restriction Reveals an Essential Role for *clag3*-Associated Channels in Malaria Parasite Nutrient Acquisition. Mol Pharmacol **2012**; 82:1104-14.

17. Nguitragool W, Rayavara K, Desai SA. Proteolysis at a Specific Extracellular Residue Implicates Integral Membrane CLAG3 in Malaria Parasite Nutrient Channels. PLoS ONE **2014**; 9:e93759.

18. Comeaux CA, Coleman BI, Bei AK, Whitehurst N, Duraisingh MT. Functional analysis of epigenetic regulation of tandem *RhopH1/clag* genes reveals a role in *Plasmodium falciparum* growth. Mol Microbiol **2011**; 80:378-90.

19. Cortés A, Carret C, Kaneko O, Yim Lim BY, Ivens A, Holder AA. Epigenetic silencing of *Plasmodium falciparum* genes linked to erythrocyte invasion. PLoS Pathog **2007**; 3:e107.

20. Rovira-Graells N, Crowley VM, Bancells C, Mira-Martínez S, Ribas de Pouplana L, Cortés A. Deciphering the principles that govern mutually exclusive expression of *Plasmodium falciparum clag3* genes. Nucleic Acids Res **2015**; 43:8243-57.

21. Sharma P, Wollenberg K, Sellers M, et al. An epigenetic antimalarial resistance mechanism involving parasite genes linked to nutrient uptake. J Biol Chem **2013**; 288:19429-40.

22. Mira-Martínez S, Rovira-Graells N, Crowley VM, Altenhofen LM, Llinás M, Cortés A. Epigenetic switches in *clag3* genes mediate blasticidin S resistance in malaria parasites. Cell Microbiol **2013**; 15:1913-23.

23. Kaestli M, Cortes A, Lagog M, Ott M, Beck HP. Longitudinal assessment of *Plasmodium falciparum var* gene transcription in naturally infected asymptomatic children in Papua New Guinea. J Infect Dis **2004**; 189:1942-51.

24. Rottmann M, Lavstsen T, Mugasa JP, et al. Differential expression of *var* gene groups is associated with morbidity caused by *Plasmodium falciparum* infection in Tanzanian children. Infect Immun **2006**; 74:3904-11.

25. Lavstsen T, Turner L, Saguti F, et al. *Plasmodium falciparum* erythrocyte membrane protein 1 domain cassettes 8 and 13 are associated with severe malaria in children. Proc Natl Acad Sci USA **2012**; 109:E1791-800.

26. Abdi AI, Warimwe GM, Muthui MK, et al. Global selection of *Plasmodium falciparum* virulence antigen expression by host antibodies. Sci Rep **2016**; 6:19882.

27. Daily JP, Scanfeld D, Pochet N, et al. Distinct physiological states of *Plasmodium falciparum* in malaria-infected patients. Nature **2007**; 450:1091-5.

28. Mackinnon MJ, Li J, Mok S, et al. Comparative transcriptional and genomic analysis of *Plasmodium falciparum* field isolates. PLoS Pathog **2009**; 5:e1000644.

29. Vignali M, Armour CD, Chen J, et al. NSR-seq transcriptional profiling enables identification of a gene signature of *Plasmodium falciparum* parasites infecting children. J Clin Invest **2011**; 121:1119-29.

30. Gomez-Perez GP, Legarda A, Munoz J, et al. Controlled human malaria infection by intramuscular and direct venous inoculation of cryopreserved *Plasmodium falciparum* sporozoites in malaria-naive volunteers: effect of injection volume and dose on infectivity rates. Malar J **2015**; 14:306.

31. Schoepflin S, Valsangiacomo F, Lin E, Kiniboro B, Mueller I, Felger I. Comparison of *Plasmodium falciparum* allelic frequency distribution in different endemic settings by high-resolution genotyping. Malar J **2009**; 8:250.

32. Wampfler R, Mwingira F, Javati S, et al. Strategies for detection of *Plasmodium* species gametocytes. PLoS ONE **2013**; 8:e76316.

33. Sauerwein RW, Roestenberg M, Moorthy VS. Experimental human challenge infections can accelerate clinical malaria vaccine development. Nat Rev Immunol **2011**; 11:57-64.

34. Deitsch KW, Lukehart SA, Stringer JR. Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens. Nat Rev Microbiol **2009**; 7:493-503.

35. Duffy MF, Brown GV, Basuki W, et al. Transcription of multiple *var* genes by individual, trophozoite-stage *Plasmodium falciparum* cells expressing a chondroitin sulphate A binding phenotype. Mol Microbiol **2002**; 43:1285-93.

36. Joergensen L, Bengtsson DC, Bengtsson A, et al. Surface co-expression of two different PfEMP1 antigens on single *Plasmodium falciparum*-infected erythrocytes facilitates binding to ICAM1 and PECAM1. PLoS Pathog **2010**; 6:e1001083.

37. Merrick CJ, Jiang RH, Skillman KM, et al. Functional Analysis of Sirtuin Genes in Multiple *Plasmodium falciparum* Strains. PLoS ONE **2015**; 10:e0118865.

38. Peters J, Fowler E, Gatton M, Chen N, Saul A, Cheng Q. High diversity and rapid changeover of expressed *var* genes during the acute phase of *Plasmodium falciparum* infections in human volunteers. Proc Natl Acad Sci USA **2002**; 99:10689-94.

39. Wang CW, Hermsen CC, Sauerwein RW, Arnot DE, Theander TG, Lavstsen T. The *Plasmodium falciparum var* gene transcription strategy at the onset of blood stage infection in a human volunteer. Parasitol Int **2009**; 58:478-80.

40. Bachmann A, Petter M, Krumkamp R, et al. Mosquito Passage Dramatically Changes var Gene Expression in Controlled Human *Plasmodium falciparum* Infections. PLoS Pathog **2016**; 12:e1005538.

41. Sinden RE. A biologist's perspective on malaria vaccine development. Hum Vaccin **2010**; 6:3-11.

42. Spence PJ, Brugat T, Langhorne J. Mosquitoes Reset Malaria Parasites. PLoS Pathog **2015**; 11:e1004987.

FIGURE LEGENDS

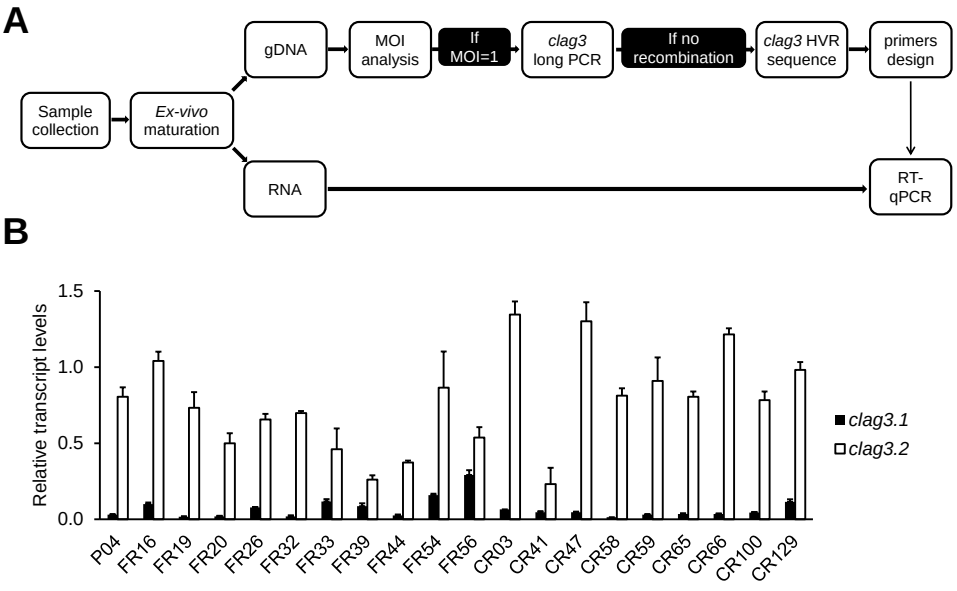
Fig. 1. Transcriptional analysis of *clag3.1* and *clag3.2* in *P. falciparum* natural human infections. (**A**) Schematic of the workflow. Forty field isolates were cultured *ex vivo* until they reached the schizont stage and harvested for genomic DNA and RNA extraction. In samples with multiplicity of infection (MOI) =1 (single infections) or a clear predominant clone and without recombination between the two *clag3* genes, the hypervariable region (HVR) of *clag3.1* and *clag3.2* was sequenced to design paralog- and isolate-specific primers for the analysis of *clag3* expression by reverse transcription followed by quantitative PCR (RT-qPCR). (**B**) Expression of *clag3.1* and *clag3.2* in twenty field isolates. Pxx are samples collected in Belgium, FRxx are samples collected in The Gambia and analyzed directly without freezing, and CRxx are cryopreserved isolates collected in The Gambia. Of note, *clag3.1* has a stronger promoter than *clag3.2* [10, 20], implying that the actual proportion of *clag3.1*-expressing parasites in each isolate is even lower than the proportion of *clag3.1* transcripts relative to total *clag3* transcripts. The significance of variation in total *clag3* transcript levels (*clag3.1* + *clag3.2* transcripts) among these samples is unclear (see Supplemental Methods); the focus of this analysis is on the relative transcript levels between the two *clag3* genes. Transcript levels are normalized against *rhoph2*. Error bars are SD.

Fig. 2. Expression of *clag3.1* and *clag3.2* in field isolates maintained under culture conditions or under drug pressure. (**A**) Transcript levels of *clag3.1* and *clag3.2* in the P04 and P12 isolates maintained under culture conditions. (**B**) Transcript levels of *clag3.1* and *clag3.2* in the same isolates selected with a low concentration of BS (0.3 µg/ml). Transcript levels are normalized against *rhoph2*. Error bars are SD.

Fig. 3. Analysis of CLAG3 sequences. (**A**) Schematic of CLAG3 sequences showing the single paralog-specific conserved sequence feature identified (positions 10-23) and the hypervariable region located near the C-terminal end of the protein (HVR, starting at position ~1110). The analysis is based on publicly available full CLAG3 sequences and the newly obtained sequences from the P04 and P12 isolates (ten sequences for each gene in total). (**B**) Cladogram of full CLAG3 sequences. (**C**) Cladogram of HVR sequences including publicly available sequences and new sequences from 24 patient isolates. In panels B and C, CLAG3.1 and CLAG3.2 sequences are represented as black and red branches, respectively (black and dark grey in the black and white printed version of the manuscript). Asterisks indicate bootstrap values >70%.

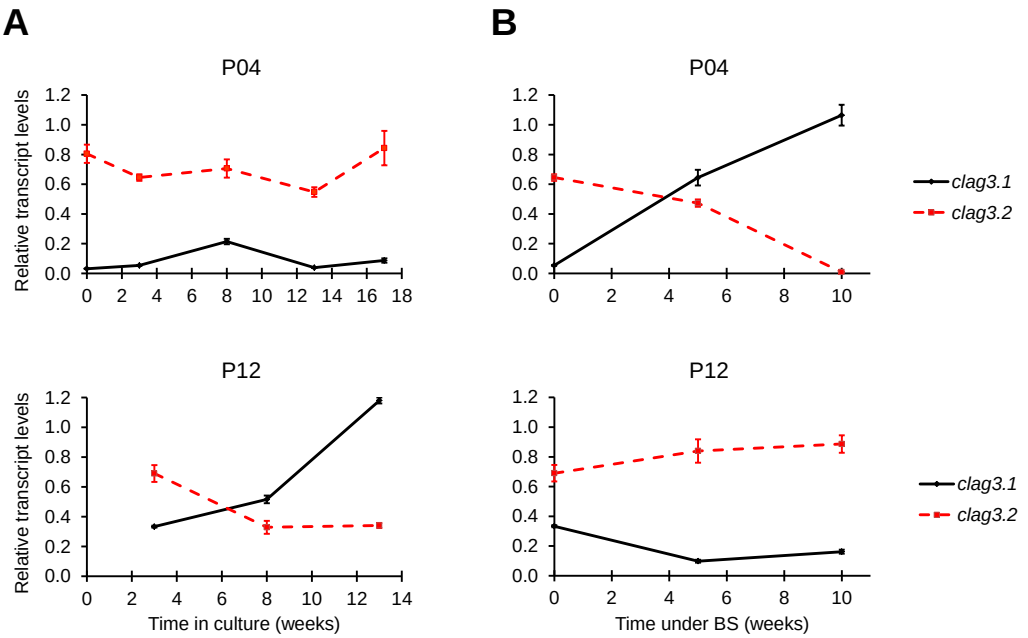
Fig. 4. Expression of *clag3.1* and *clag3.2* in controlled human malaria infections. (**A**) Schematic of the controlled human malaria infection trial. Sporozoites obtained from mosquitoes fed with NF54-infected blood were cryopreserved and injected into healthy volunteers. (**B**) Expression of *clag3.1* and *clag3.2* in the NF54 parental line used to infect the mosquitoes and in parasites collected from volunteers at the day of malaria diagnosis by light microscopy (day of malaria). The significance of variation in total *clag3* transcript levels (*clag3.1* + *clag3.2* transcripts) among these samples is unclear (see Supplemental Methods); the focus of this analysis is on the relative transcript levels between the two *clag3* genes. (**C**) Relative transcript levels of *clag3.1* and *clag3.2,* expressed as the $\log_2$ of the ratio of *clag3.1*/*clag3.2* transcript levels, in parasites from two volunteers maintained under culture conditions for 9 weeks. (**D**) Expression of *clag3.1* and *clag3.2* in parasites collected from two volunteers at day 9 after sporozoite injection (~1 cycle in the blood circulation). In all panels, transcript levels are normalized against *rhoph2*. Error bars are SD.

559 **Figure 1.**

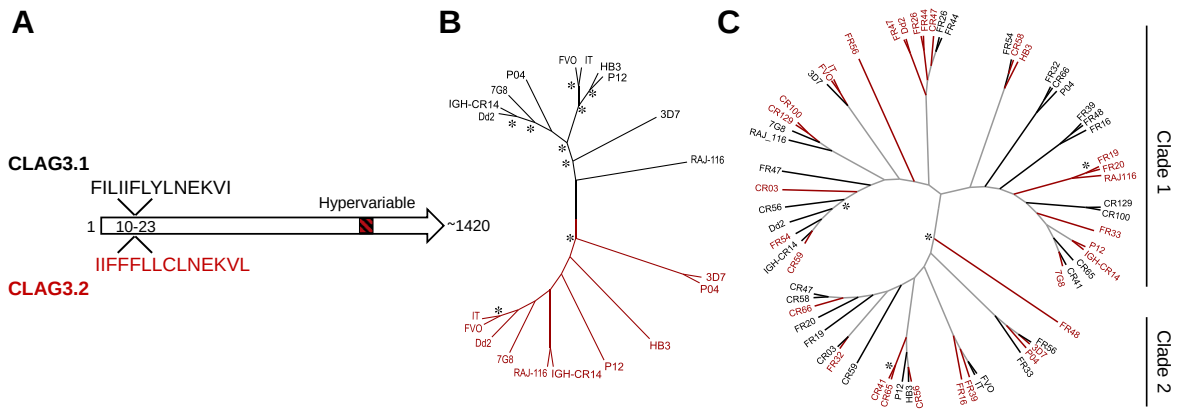


560

Figure 2.

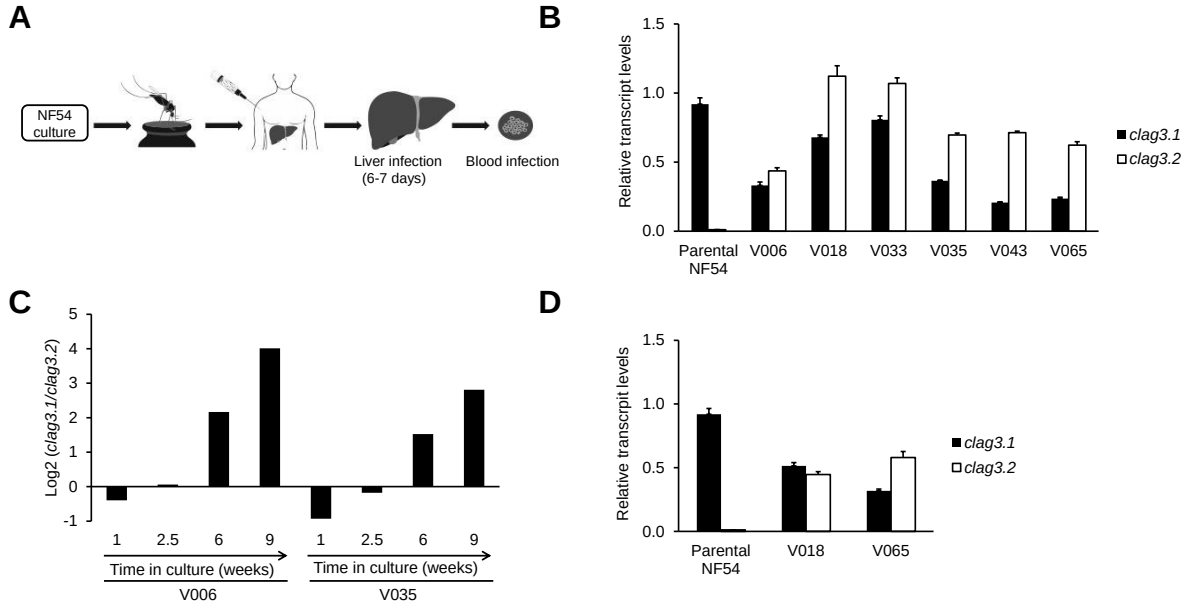

A
B
P04
P12
Relative transcript levels
Time in culture (weeks)
Time under BS (weeks)
clag3.1
clag3.2

563 **Figure 3.**



564

Figure 4.


A
NF54 culture
Liver infection (6-7 days)
Blood infection
B
Relative transcript levels
1.5
1.0
0.5
0.0
Parental NF54
V006
V018
V033
V035
V043
V065
clag3.1
clag3.2
C
Log2 (clag3.1/clag3.2)
5
4
3
2
1
0
-1
1
2.5
6
9
Time in culture (weeks)
V006
V035
D
Relative transcript levels
Parental NF54
V018
V065
clag3.1
clag3.2

SUPPLEMENTAL METHODS

Sample collection

P. falciparum-infected blood samples were obtained from returning travelers (≥18 years old) attending the clinics at the Institute of Tropical Medicine or University Hospital in Antwerp (Belgium), and from children (≤12 years old) attending the Brikama Health Centre (The Gambia) as part of an ongoing drug efficacy study. 1-3.5 ml of venous-blood were collected in citrate phosphate dextrose adenine (CPDA) or EDTA vacutainers. Parasites were placed in culture within 6 h of collection. In addition, cryopreserved parasite isolates obtained from children participating in previous studies at MRC Unit The Gambia (MRC biobank) were included in the study. We also analyzed blood samples collected from volunteers participating in a controlled human malaria infection (CHMI) study [1]. In brief, *P. falciparum* sporozoites of the NF54 parasite line at Sanaria were injected into malaria-naïve volunteers and venous blood samples were collected and cryopreserved at day 9 after sporozoites injection and at the day of malaria (when parasites are first detected by microscopy, 12-15 days post-injection).

Parasite cultures

Parasites were cultured at 3% hematocrit under standard conditions, with Albumax II or inactivated human serum according to regular procedures at each of the laboratories involved in this study. To verify that supplementing with Albumax II or inactivated serum does not affect *clag3* expression, transcript levels of these genes were compared between 3D7 cultures maintained in parallel under both conditions, which confirmed similar *clag3* expression patterns (Supplemental Fig. S1A). To prepare RNA for *clag* transcriptional analysis, cultures were harvested when the majority of parasites were at the schizont stage.

In the case of natural infections, parasites were cultured only until they reached the schizont stage of the first cycle, with the exception of two of the samples (FR39 and CR100), which had to be cultured for one additional cycle due to very low parasitaemia. Day 9 and day of malaria samples from the CHMI were cryopreserved and, after thawing, parasites were cultured for 2-3 weeks or ~1 week, respectively, until they reached a ≥0.02% parasitaemia. Cultures were then sorbitol-synchronized and harvested at the schizont stage. Selection with BS (ThermoFisher) was performed with a sub-lethal concentration of the drug (0.3 µg/ml) as previously described [2].

Genetic analysis

Genomic DNA (gDNA) was extracted using the QIAamp DNA Blood Mini Kit (Qiagen) following manufacturer's instructions. The multiplicity of infection of field samples was estimated as the highest number of alleles observed by genotyping the *msp1* and *msp2* loci. Amplification of the *msp1* and *msp2* target sequences was performed using fluorescently-labeled primers and analyzed by capillary electrophoresis approximately as previously described [3]. To assess for recombination events between the two *clag3* genes, the *clag3* loci were analyzed by long PCR with primers against the conserved paralog-specific regions flanking the genes, as previously described [4].

We also used *msp1* and *msp2* genotyping to confirm that samples P04 and P12 contained the same haplotype at different times along the selection experiments reported in Fig. 2 in the main manuscript, thus excluding the possibility that changes in *clag3* expression were associated with selection of genetically different parasites present in the population. We also excluded recombination between the two *clag3* genes [4] during adaptation.

For each isolate, we sequenced the hypervariable region (HVR) of each *clag3* gene from the long PCR products, using primers against the conserved regions around the HVR (Supplemental Table S1). The sequences obtained were used to design new gene- and isolate-specific primers or to determine which existing primers are suitable for the analysis of *clag3.1* and *clag3.2* transcript levels in each isolate. The full *clag3.1* and *clag3.2* genes from isolates P04 and P12 were sequenced from long PCR products using conventional Sanger sequencing.

Preparation of RNA, reverse transcription and quantitative real time PCR

RNA was purified approximately as described [5]. In brief, parasites were collected in Trizol (Invitrogen) and, after phase separation, RNA was purified using the RNeasy Mini Kit (Qiagen), following the supplier instructions except for DNase digestion that was performed on-column. RNA was reverse transcribed using the High Capacity RNA-to-cDNA Kit (Applied Biosystems). To exclude the possibility of gDNA contamination, parallel reactions were performed in the absence of reverse transcriptase (-RT controls). The optimized protocol was validated for use at very low parasitaemias by comparing *clag3* expression in two parasite lines between 2% and 0.02% parasitaemia (Supplemental Fig. S1B). Quantitative PCR analysis of cDNAs and -RT controls was performed in triplicate wells using the Power SYBR Green Master Mix (Applied Biosystems) or SensiMix SYBR No-ROX Kit (Bioline). Each 96-well plate contained a standard curve made with serial dilutions of 3D7 gDNA, except in the case of *clag3* analysis of field isolates for which the standard curve was prepared with gDNA from the same isolate. Expression values, in arbitrary units, were calculated using the standard curve method as previously described [6]. Expression values of *clag* genes were normalized against expression of *rhoph2* (PF3D7_0929400), which

shows a similar temporal expression pattern. We also measured expression of the constitutive gene *seryl tRNA synthetase* (*seryl*, PF3D7_0717700) to estimate the proportion of parasites at the schizont stage. Samples with a *rhoph2/seryl* ratio <2 were excluded from further analysis based on our previous observations that well-synchronized cultures with a predominance of schizonts show higher values of this ratio. In spite of using *rhoph2* for normalization and always harvesting cultures for transcriptional analysis at a similar stage of parasite development, total normalized *clag3* transcript levels (*clag3.1* + *clag3.2* transcript levels) show intrinsic variability that may not reflect actual biological differences, especially among samples from patients or volunteers that had to be analyzed at very low parasitaemia after minimal in vitro culture and could not be tightly synchronized. Variability in total normalized *clag3* transcript levels, which sometimes is observed even between biological replicates, is probably attributable to small differences in the temporal expression pattern between *rhoph2* and *clag3* genes and to small differences in life cycle progression between samples. Hence, we drew our conclusions from the analysis of the relative abundance of *clag3.1* and *clag3.2* transcripts, which can be measured accurately and is highly reproducible.

For *clag3* genes we used paralog- and isolate-specific primers, whereas for other *clag* genes we designed primers against conserved regions in which no polymorphism is reported at www.malariagen.net [7] (Supplemental Table S1). Absence of polymorphism was also confirmed by sequencing these regions in some of the isolates.

CLAG3 sequences analysis

Publicly available CLAG3.1 and CLAG3.2 full length or HVR sequences were obtained from PlasmoDB (www.plasmodb.org) and GenBank. Alignments were performed with ClustalW and manual refinement. Pairwise sequence distances were computed using the Jones-Taylor-

Thorton substitution model with a gamma distribution (shape parameter = 1) as implemented in MEGA6 [8]. A phylogenetic tree was then generated in MEGA6 using the Neighbor-Joining method with pairwise removal of gaps and 500 bootstrap replicates to assess topology reliability. An unrooted radial cladogram was generated using FigTree v1.4.2 (http://tree.bio.ed.ac.uk), whereas split networks of CLAG3 sequences were inferred using SplitsTree4 (www.splitstree.org) [9].

SUPPLEMENTAL REFERENCES

1. Gomez-Perez GP, Legarda A, Munoz J, et al. Controlled human malaria infection by intramuscular and direct venous inoculation of cryopreserved *Plasmodium falciparum* sporozoites in malaria-naive volunteers: effect of injection volume and dose on infectivity rates. Malar J **2015**; 14:306.

2. Mira-Martínez S, Rovira-Graells N, Crowley VM, Altenhofen LM, Llinás M, Cortés A. Epigenetic switches in *clag3* genes mediate blasticidin S resistance in malaria parasites. Cell Microbiol **2013**; 15:1913-23.

3. Schoepflin S, Valsangiacomo F, Lin E, Kiniboro B, Mueller I, Felger I. Comparison of *Plasmodium falciparum* allelic frequency distribution in different endemic settings by high-resolution genotyping. Malar J **2009**; 8:250.

4. Iriko H, Kaneko O, Otsuki H, et al. Diversity and evolution of the *rhoph1/clag* multigene family of *Plasmodium falciparum*. Mol Biochem Parasitol **2008**; 158:11-21.

5. Wampfler R, Mwingira F, Javati S, et al. Strategies for detection of *Plasmodium* species gametocytes. PLoS ONE **2013**; 8:e76316.

6. Crowley VM, Rovira-Graells N, de Pouplana LR, Cortés A. Heterochromatin formation in bistable chromatin domains controls the epigenetic repression of clonally variant *Plasmodium falciparum* genes linked to erythrocyte invasion. Mol Microbiol **2011**; 80:391-406.

7. Manske M, Miotto O, Campino S, et al. Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. Nature **2012**; 487:375-9.

8. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol **2013**; 30:2725-9.

9. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol **2006**; 23:254-67.

SUPPLEMENTAL DATA

Expression of the *Plasmodium falciparum* clonally variant *clag3* genes in human infections

Sofía Mira-Martínez, Evi van Schuppen, Alfred Amambua-Ngwa, Emmanuel Bottieau, Muna Affara, Marjan Van Esbroeck, Erika Vlieghe, Pieter Guetens, Núria Rovira-Graells, Gloria P. Gómez-Pérez, Pedro L. Alonso, Umberto D'Alessandro, Anna Rosanas-Urgell and Alfred Cortés

Supplemental Table S1. Primers used in this study.

Name	Sequence 5'--3'	Target gene	Samples
long PCR			
clag3.1_long_F*	TGTGCAATATATCAAAGTGTACATGC	*clag3.1*	All
clag3.1_long_R*	TAGAAAATATTAGAATTGCTATTATGTAC	*clag3.1*	All
clag3.2_long_F*	AATAGTTGAGTACGCACTAATATGTC	*clag3.2*	All
clag3.2_long_R*	ACACAAATTCTTAATAATTATATAAAACC	*clag3.2*	All
HVR Sequencing			
seq3.1&3.2B F	ATCAGGCTGTGGCAAGTGC	*clag3.1 & clag3.2*	Field isolates
seq3.1&3.2B R	CTGACGGTTTGCGTTCGTAGG	*clag3.1 & clag3.2*	Field isolates
RT-qPCR			
PF07_0073 F*	AAGTAGCAGGTCATCGTGGTT	*seryl*	All
PF07_0073 R*	TTCGGCACATTCTTCCATAA	*seryl*	All
rhopH2_qRT_P2F*	TGTTGCTGTCCATATTTAGTTTT	*rhoph2*	All
rhopH2_qRT_P2R*	AATATATCGCTACATAACTTCGT	*rhoph2*	All
clag3.1 clag3.2 – 6F*	TAGTAATGAGAATTAGTTGGACA	*clag3.1 & clag3.2*	NF54/3D7, some field isolates
clag3.1 –6R*	ATAAATATTTGGATGCTTCAGCA	*clag3.1*	NF54/3D7, some field isolates
clag3.2 –6R*	ACAAATATGTTTCTGAACTAGGA	*clag3.2*	NF54/3D7, some field isolates
BR054 _3.1F	TCTGCTAAAAAGGTAGGTCAGT	*clag3.1*	Some field isolates
BR054 _3.1R	GGATTTATATTTACCACTTGCAC	*clag3.1*	Some field isolates
BR054 _3.2R	CGCACTAGGACAAGTTTTCTG	*clag3.2*	Some field isolates
CRYO947 _3.2F	ACACATGCTTTTACAACTGGAC	*clag3.2*	Some field isolates
CRYO1000_3.1R	GCTTTTATAGAATTAGTATTAGCG	*clag3.1*	Some field isolates
CRYO1000_3.2R	TTACAATTAGTAGTAGAACTTGAG	*clag3.2*	Some field isolates
CRYO1029 _3.1R	CTTTTATAGAATTAGTATTAGCGG	*clag3.1*	Some field isolates
CRYO1029 _3.2R	ACAATTAGTAGTAGAACTTGAGG	*clag3.2*	Some field isolates
CRYO1029 _3.1&3.2F	TATCAGGCTGTGGCAAGTGC	*clag3.1 & clag3.2*	Some field isolates
BR044 & BR026_3.1F	GATCCCAAAAGTTGTACTAGTAG	*clag3.1*	Some field isolates
BR044 & BR026_3.1R	CAGCAAGTGCGTGAGTGAAG	*clag3.1*	Some field isolates
BR044 & BR026_3.2F	CCTCAAGTTCTACTACTAATTGT	*clag3.2*	Some field isolates
BR044 & BR026_3.2R	CAGCAAGTCCGTGAGTAAAGA	*clag3.2*	Some field isolates
clag9_qRT_F2	AATCACTTACCTGAAGAATTGAG	*clag9*	Field isolates, Selection
clag9_qRT_R2	ACGAAAGGGACAAACCATGAC	*clag9*	Field isolates, Selection
clag9_qRT_F1*	GTAAATCAATGGCAAATACTTGG	*clag9*	CHMI
clag9_qRT_R1*	CTGGTTGTTGTAATTCTACACC	*clag9*	CHMI
clag8_qRT_F2	ACGGAAGATACGGATTTCGAC	*clag8*	Field isolates, Selection
clag8_qRT_R2	TCGAAAGTATCTTCCTCATCCT	*clag8*	Field isolates, Selection
clag8_qRT_F1*	CATCGGTTCATGGTTTTACACA	*clag8*	CHMI
clag8_qRT_R1*	AAGCATATATTTGTGAAAGGCTC	*clag8*	CHMI
clag2_qRT_F3	GTCATAAGGAAGAAACCAACAC	*clag2*	Field isolates
clag2_qRT_R3	TGTTTGAATATAGAAATGCCCCT	*clag2*	Field isolates
clag2_qRT_F1*	TTCGTGCATCATATGGTTGGG	*clag2*	CHMI
clag2_qRT_R1&2*	TATATAGGTGCATCAGATTTCCA	*clag2*	CHMI, Selection
clag2_qRT_F2	ATCTATCTCTCAGAATTCGTGC	*clag2*	Selection
Others			
clag9_long_F1	GATTCCAATAATGAAAGGTGATCTTG	*clag9*	Field isolates
clag9_long_R1	TCATAACGTTTCGTTCTATCTATACC	*clag9*	Field isolates
clag2_seqF1	GTCTTTTGTGTGAATACCAAGC	*clag2*	Field isolates
clag2_seqR1	ATGAAGTAGAAAATCCTCCAGG	*clag2*	Field isolates
clag9_seqF1	AATATCAAGCCATTGGAAGTGC	*clag9*	Field isolates
clag9_seqR1	TTGCCAATCAATACTATGAACAG	*clag9*	Field isolates
seqR3.2_BR044 & BR026	TGACGGTTTGCGTTCATACGT	*clag3.2*	Some field isolates
seqC_R	ATACTTAAATAAGCCTCACGTTC	*clag3.1 & clag3.2*	Some field isolates

The primers marked with an asterisk have been previously described [1-3]. "Selection" refers to the samples from the selection experiments shown in Figs. 2 and S4. Primers for *clag3* genes were used to analyze *clag3.1* expression in some isolates and to analyze *clag3.2* expression in other isolates. The name of the primer only indicates the isolate and the paralog against which the primer was originally designed. "Others" corresponds to primers used to obtain sequences for the genes indicated to confirm the absence of polymorphism in the regions where the primers for transcriptional analysis were designed.

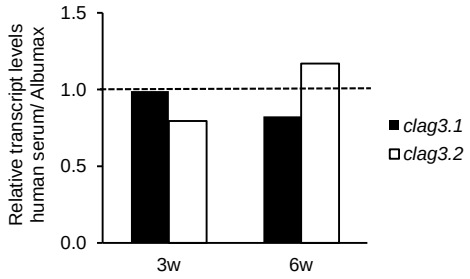
References.

1. Iriko H, Kaneko O, Otsuki H, et al. Diversity and evolution of the *rhoph1/clag* multigene family of *Plasmodium falciparum*. Mol Biochem Parasitol **2008**; 158:11-21.

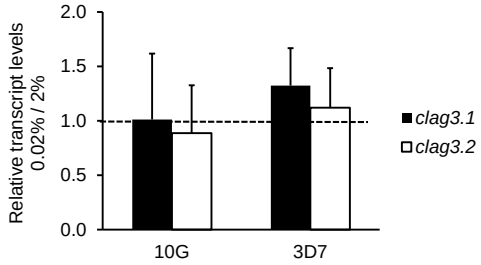
2. Crowley VM, Rovira-Graells N, de Pouplana LR, Cortés A. Heterochromatin formation in bistable chromatin domains controls the epigenetic repression of clonally variant *Plasmodium falciparum* genes linked to erythrocyte invasion. Mol Microbiol **2011**; 80:391-406.

3. Mira-Martínez S, Rovira-Graells N, Crowley VM, Altenhofen LM, Llinás M, Cortés A. Epigenetic switches in *clag3* genes mediate blasticidin S resistance in malaria parasites. Cell Microbiol **2013**; 15:1913-23.

A

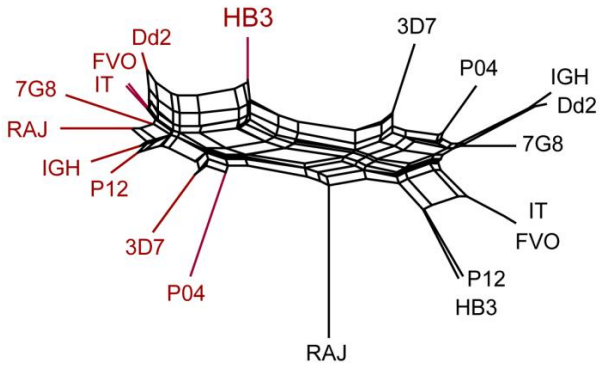


B

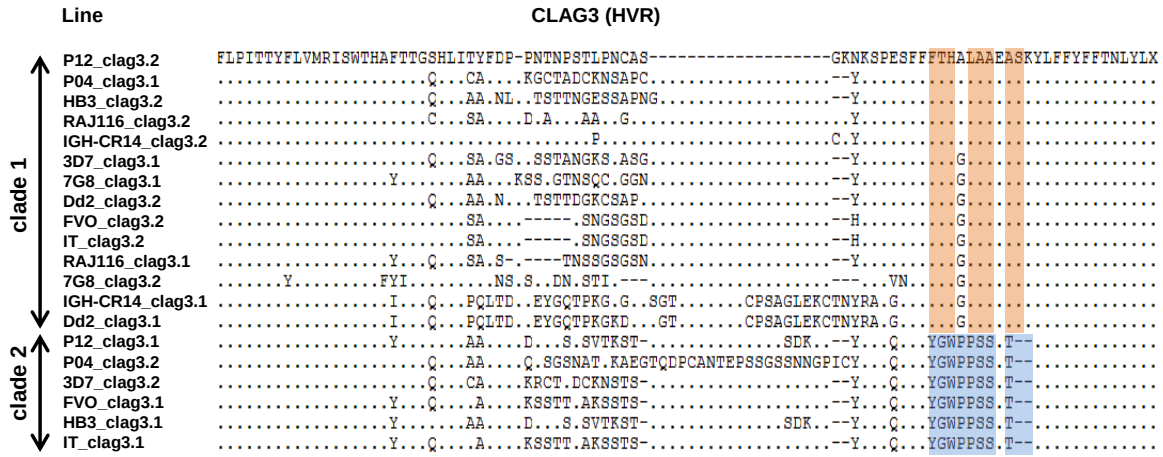


Supplemental Fig. S1. Validation of the methods used. (A) Expression of *clag3.1* and *clag3.2* in the P04 isolate maintained in culture media supplemented with human serum relative to expression in media supplemented with Albumax II. Cultures were maintained under the different conditions for up to 6 weeks. **(B)** Validation of the RT-qPCR protocols for the analysis of samples with low parasite densities. We compared *clag3* expression in the 10G and 3D7-ITM parasite lines (expressing predominantly *clag3.2* and *clag3.1*, respectively) between 2% and 0.02% parasitaemia. Results are expressed as the transcript levels (normalized against *rhoph2*) at 0.02% parasitaemia relative to transcript levels at 2% parasitaemia. Values are the average of three independent experiments, with SD.

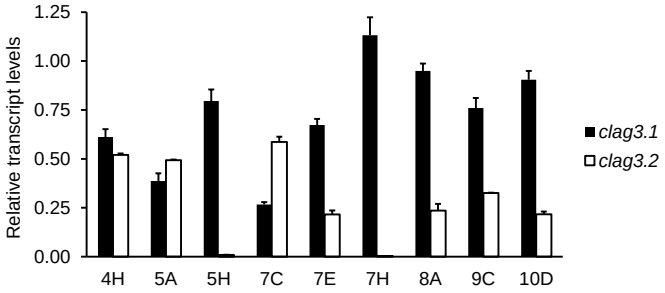
A



B



Supplemental Fig. S2. CLAG3 sequence analysis. (A) Split network of full CLAG3 sequences (publicly available sequences and new sequences from the P04 and P12 isolates). Connections between branches (parallel edges) represent recombination events. The distribution of these branches indicates that recombination events between the two paralog genes have likely occurred. CLAG3.1 and CLAG3.2 sequences are represented as black and red branches, respectively. **(B)** Alignment of a subset of CLAG3 hypervariable region (HVR) sequences (from the same parasite lines as in panel A) ordered according to the two separate clades observed in Fig. 3C. Dots indicate identity with the first sequence. This subset of sequences is representative of the full set of sequences analyzed (22 additional new sequences for each gene). Conserved sequence features specific for clades 1 or 2 are highlighted. These features are also conserved in all other clade 1 and clade 2 sequences analyzed. Each isolate had at least one *clag3* gene belonging to clade 1, which is suggestive of a non-random distribution of HVR sequences.

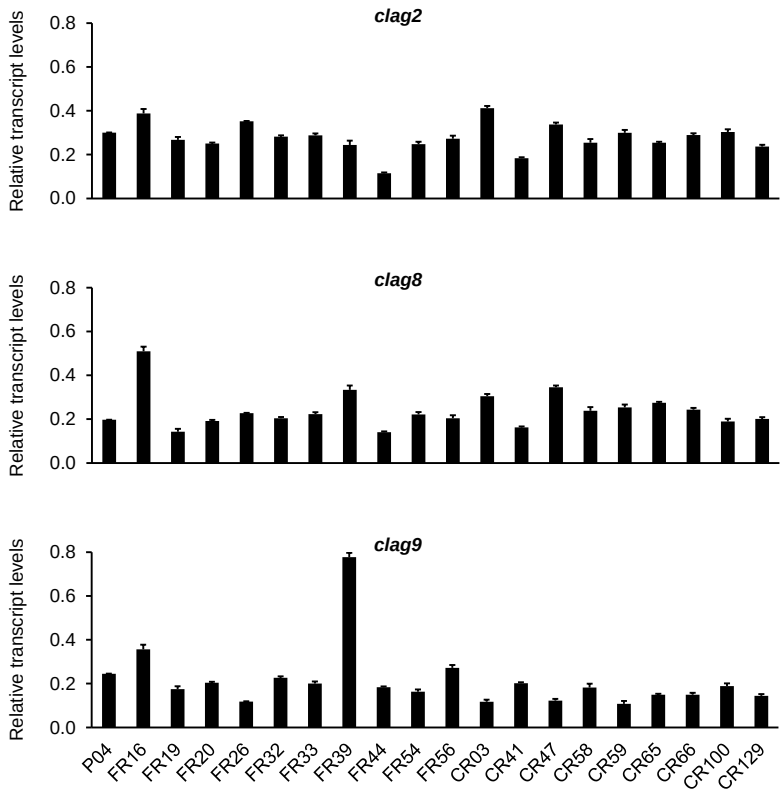


Supplemental Fig. S3. Expression of *clag3.1* and *clag3.2* in subclones of isolate V033 (obtained from a volunteer infected with parasites of the NF54 line). Subclones were obtained by limiting dilution. Two out of nine subclones (5H and 7H) expressed *clag3.1* almost exclusively, whereas all other subclones likely originated from *clag3.2*-expressing single parasites. In the latter subclones, during the 3-4 weeks in culture elapsed between limiting dilution and transcriptional analysis parasites that spontaneously switched to *clag3.1* expression were progressively selected. This interpretation is based on the observation that in the NF54 genetic background parasites that express *clag3.1* are selected under culture conditions (Fig. 4C). The similar transcript levels of *clag3.1* and *clag3.2* observed in parasites from volunteer V033 (Fig. 4B) are consistent with the number of *clag3.1*-expressing subclones obtained (two out of nine) because *clag3.1* has a stronger promoter than *clag3.2* [1, 2], implying that similar transcript levels of *clag3.1* and *clag3.2* reflect a higher proportion of *clag3.2*-expressing parasites. Altogether, the transcriptional analysis of V033 subclones revealing two clearly distinct types of subclones suggests that the V033 isolate predominantly consists of a mixture of parasites expressing *clag3.1* and parasites expressing *clag3.2*, rather than a homogeneous population of parasites expressing the two genes simultaneously, but mutual exclusion is directly demonstrated only for parasites expressing *clag3.1*, and inferred for parasites expressing *clag3.2*. Transcript levels are normalized against *rhoph2*. Error bars are SD.

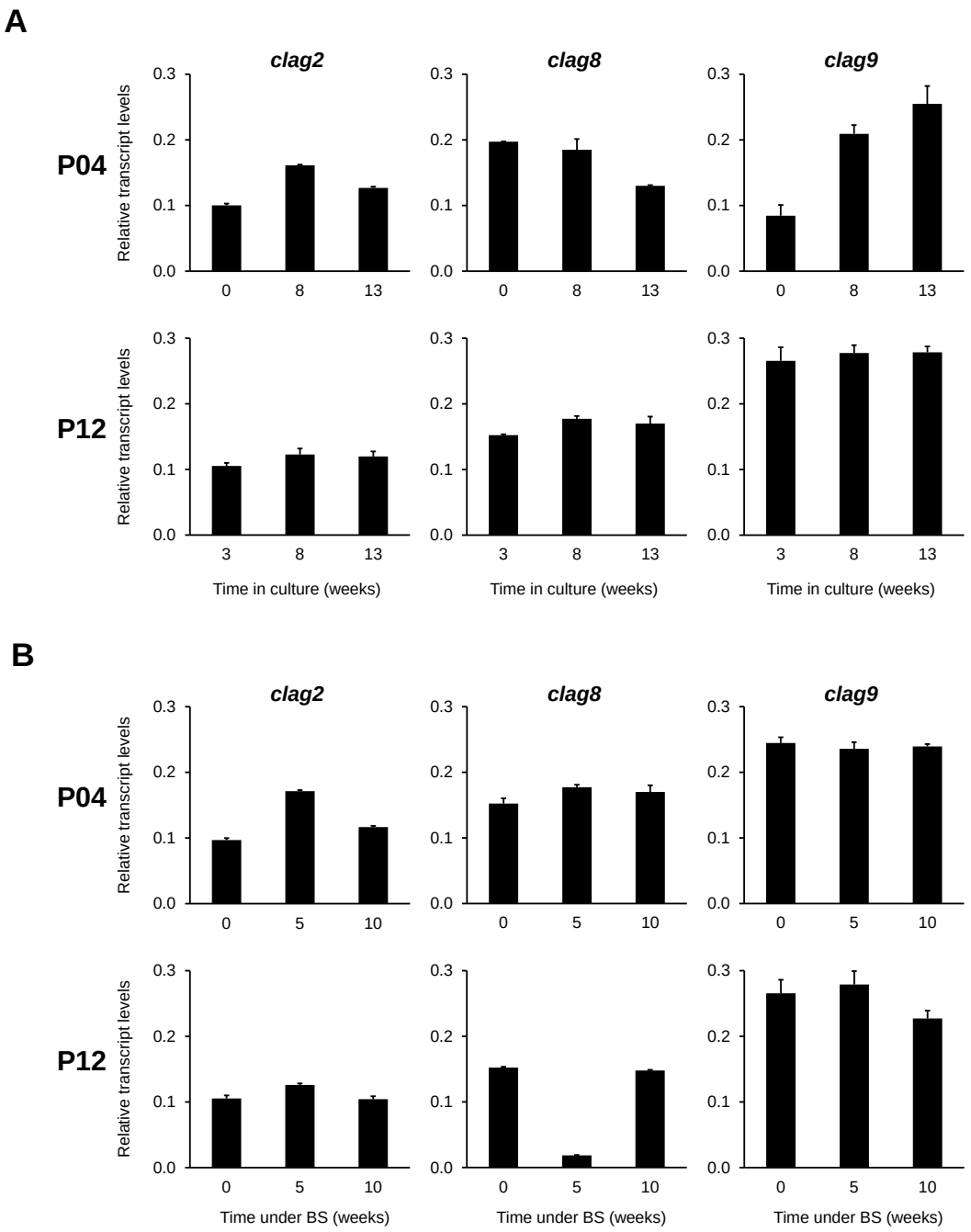
References.

1. Crowley VM, Rovira-Graells N, de Pouplana LR, Cortés A. Heterochromatin formation in bistable chromatin domains controls the epigenetic repression of clonally variant *Plasmodium falciparum* genes linked to erythrocyte invasion. Mol Microbiol **2011**; 80:391-406.

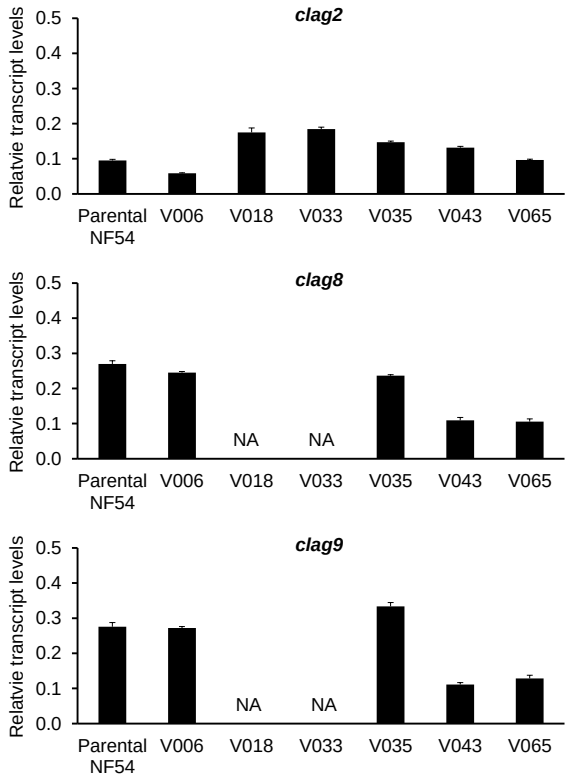
2. Rovira-Graells N, Crowley VM, Bancells C, Mira-Martínez S, Ribas de Pouplana L, Cortés A. Deciphering the principles that govern mutually exclusive expression of *Plasmodium falciparum clag3* genes. Nucleic Acids Res **2015**; 43:8243-57.



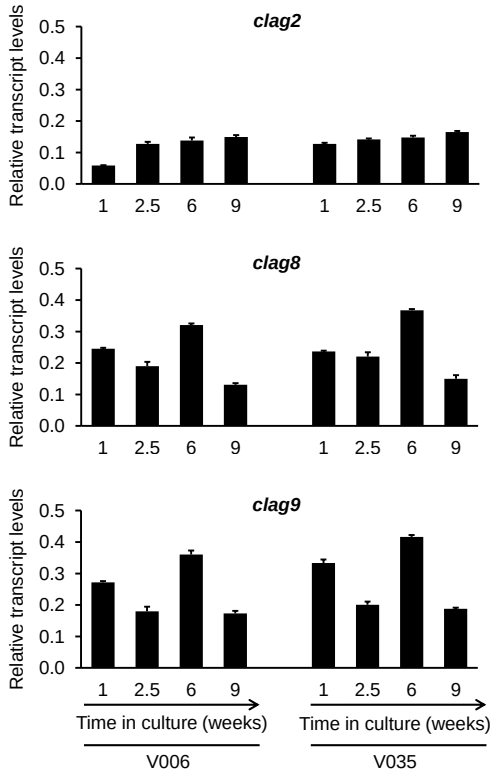
Supplemental Fig. S4. Expression of *clag2, clag8* and *clag9* in *P. falciparum* natural infections. RNA from field isolates cultured *ex vivo* to the schizont stage was used for this analysis, as in Fig. 1B. Transcript levels are normalized against *rhoph2*. Error bars are SD. Only isolate FR39 presented substantially higher expression of *clag9* than other isolates. Analysis of FR39 genomic DNA by qPCR ruled out altered *clag9* copy number in this isolate (data not shown).



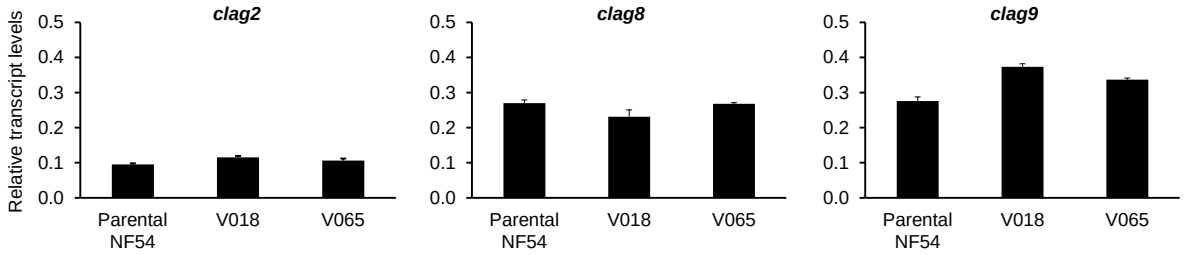
Supplemental Fig. S5. Expression of *clag2, clag8* and *clag9* in field isolates maintained under culture conditions or under drug pressure. Transcript levels were measured in the P04 and P12 isolates maintained under culture conditions for up to 13 weeks **(A)** or selected with a low concentration (0.3 µg/mL) of blasticidin S (BS) for up to 10 weeks **(B)**. Transcript levels are normalized against *rhoph2*. Error bars are SD. Only one of the samples, P12 after 5 weeks of BS pressure, presented a substantially lower expression of *clag8*, but after 10 weeks of selection *clag8* expression was back to normal levels.



Supplemental Fig. S6. Expression of *clag2, clag8* and *clag9* in parasites collected from *P. falciparum*-infected volunteers at the day of malaria and in cultures of the parental NF54 line used to produce the sporozoites. Parasites were cultured until parasitaemia was high enough for transcriptional analysis (3-4 cycles). Transcript levels are normalized against *rhoph2*. Error bars are SD. NA: not analyzed.



Supplemental Fig. S7. Expression of *clag2, clag8* and *clag9* in parasites collected from two infected human volunteers at the day of malaria and maintained under culture conditions for up to 9 weeks. Transcript levels are normalized against *rhoph2*. Error bars are SD.



Supplemental Fig. S8. Expression of *clag2, clag8* and *clag9* in parasites collected from infected human volunteers at day 9 and in the parental NF54 line. Parasites were cultured until parasitaemia was high enough for transcriptional analysis (2-3 weeks). Transcript levels are normalized against *rhoph2*. Error bars are SD.