UNIVERSITAT DE BARCELONA

Characterisation of the role of epigenetic variation in the adaptation of malaria parasites to changes in the conditions of the human host

Anastasia Katherine Pickford

Characterisation of the role of epigenetic variation in the adaptation of malaria parasites to changes in the conditions of the human host

Anastasia Katherine Pickford

This thesis was carried out at the Institute of Global Health of Barcelona (ISGlobal), under the supervision of **Dr Alfred Cortés Closas** and the tutelage of **Dr Gemma Marfany Nadal**

Programa de Doctorat en Biomedicina
Universitat de Barcelona

September 2021

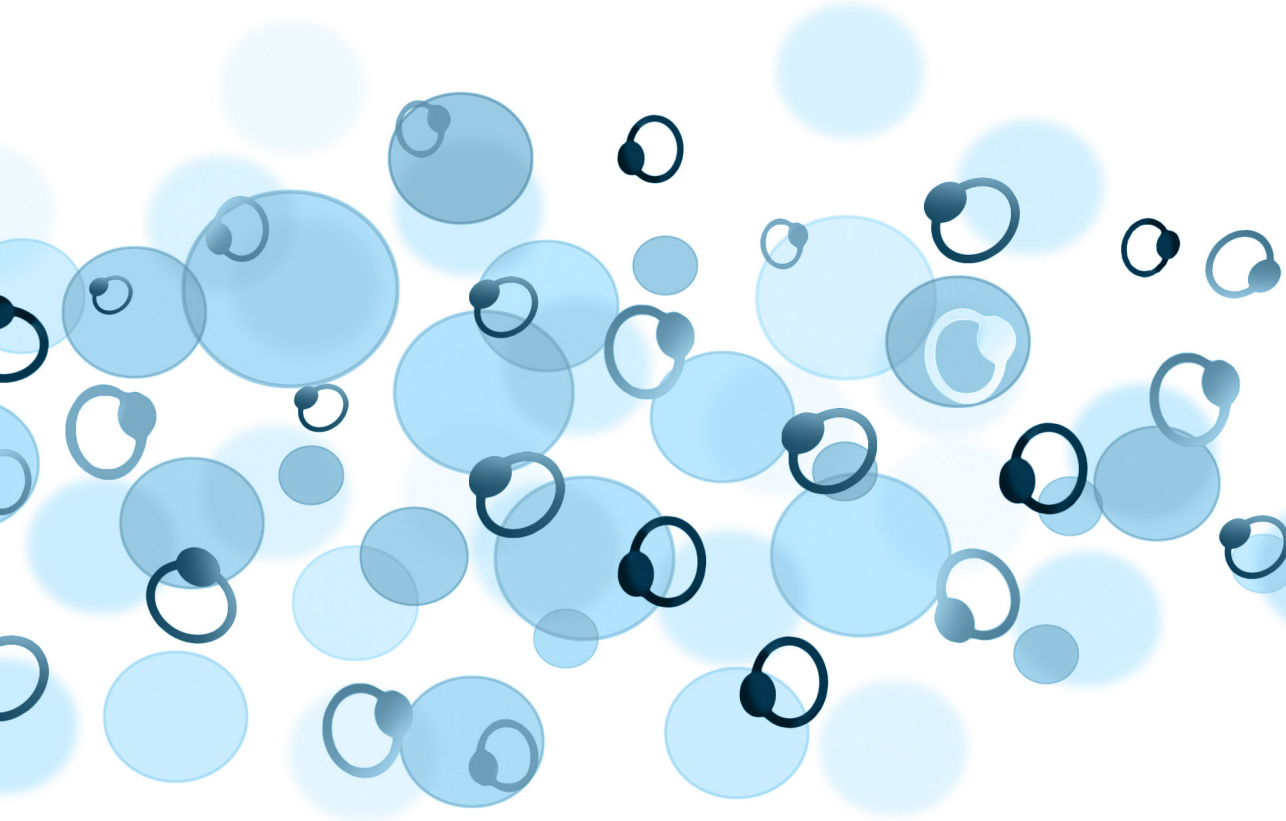
Universitat de Barcelona
ISGlobal Instituto de Salud Global Barcelona
AGAUR Agència de Gestió d'Ajuts Universitaris i de Recerca
Generalitat de Catalunya
Departament d'Empresa i Coneixement
Secretaria d'Universitats i Recerca
Unió Europea
Fons Social Europeu
L'FSE inverteix en el teu futur

“The role of the infinitely small in nature is infinitely great.”

Louis Pasteur

Contents

INTRODUCTION



1. Malaria

1.1. History and current relevance

Malaria is an ancient disease that has afflicted humans for thousands of years[1]. References to this ailment go back as far as 2,700BC. It has been described as "marsh fever" by the ancient Greeks and its current name derives from the Italian "bad air". In 1880 Alphonse Laveran discovered the *Plasmodium* parasite as the aetiological agent of the disease and almost twenty years later Ronald Ross identified that mosquitoes were responsible for its transmission[2].

Historically, malaria has been a widespread disease and was present even in Europe up until the mid-nineteenth century. Fortunately, the disease was successfully eliminated from this area thanks to the appearance of DDT insecticides in the 1940s and the Global Malaria Eradication Program in the 50s and 60s[3,4]. However, decades later the battle against malaria is still ongoing and it constitutes a huge obstacle for socioeconomic development in heavily affected regions like sub-Saharan Africa. This disease sinks these areas deeper into poverty by decreasing productivity, hindering education and costing local governments billions of euros in treatment, control and prevention[5]. Consequently, the global fight against malaria is as necessary as ever and research that enables a better understanding of the disease is essential.

1.2. Aetiology

Malaria parasites belong to the Apicomplexa phylum that together with dinoflagellates, chromerids and ciliates, form the superphylum Alveolata[6] (**Fig. 1**). Apicomplexans constitute a large and diverse taxonomic group that is composed of single-cell eukaryotic organisms that are obligate intracellular parasites and its name derives from the so-called "apical complex", a set of secretory organelles that allow the mobile invasive forms of these parasites to attach to and enter host cells[7]. Apicomplexans have complicated life cycles that take place in various vertebrate and invertebrate hosts, in different tissues and multiple cell types. Throughout their life cycle, they undergo rounds of both asexual and sexual replication in order to generate sufficient progeny to perpetuate their existence as well as to produce the specialized stages needed for transmission between hosts[8].

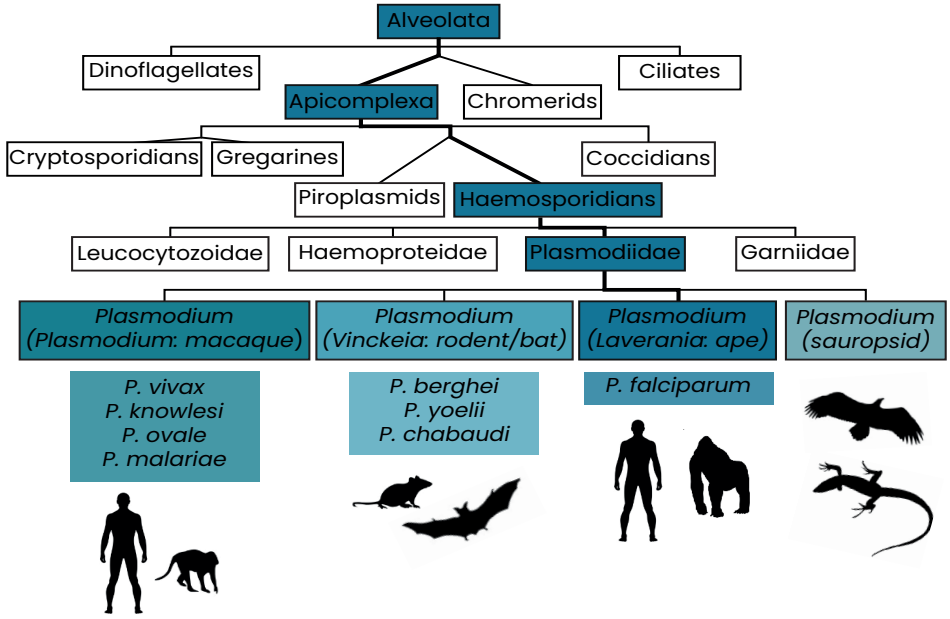


Fig. 1. Simplified phylogenetic tree of malaria parasites from superphylum to species. (Adapted from Galen *et al.*[14])

Many apicomplexan parasites are pathogens of humans and/or domestic animals and are of major public health and economic concern[9]. Taking into account traditional functional criteria rather than modern molecular data, apicomplexans can be classified into the following orders: cryptosporidians, coccidians, gregarines, piroplasmids and haemosporidians[7,8]. The first two orders include notorious zoonotic pathogens such as *Cryptosporidium* spp. that cause severe diarrheal disease[10] and *Toxoplasma gondii* which gravely jeopardizes the health of immunocompromised individuals as well as foetuses and new-borns due to its congenital transmission[11]. Next, piroplasmids and haemosporidinas are collectively known as haematozoan parasites which develop in invertebrate hosts and are injected into vertebrates during blood feeding[12]. Piroplasmids are transmitted by ticks and include *Babesia* spp. and *Theileria* spp. which cause disease in livestock resulting in grave economic losses[13]. In contrast, haemosporidians are transmitted by dipteran vectors and include over 540 species that can be classified into four families: *Plasmodidae*, *Haemoproteidae*, *Leucocytozoidae* and *Garniidae*[12]. Malaria parasites belong to the *Plasmodidae* family and more specifically to the genus *Plasmodium* which includes many species that infect birds and lizards

and are transmitted by culicine mosquitoes, as well as species that can infect mammals (including humans) which are transmitted by anopheline mosquitoes[12].

The main two groups of mammals that can be infected by *Plasmodium* parasites are rodents and primates. Rodent malaria species include *P. berghei*, *P. chabaudi* and *P. yoelii*, all of which are useful for establishing in vitro models that help study disease mechanisms and protective host immune responses[15]. On the other hand, primate malaria species infect a wide diversity of apes, monkeys and lemurs from across the globe with the distinct subgenus Laverania used to classify ape-infecting species[16]. There are six malaria species that can infect humans: *P. falciparum*, *P. vivax*, *P. malariae*, *P. ovale curtisi*, *P. ovale wallikeri* and *P. knowlesi* (**Table 1**). Mixed infections with more than one malaria species are relatively common in areas where various *Plasmodium* spp. are found[17].

1.3. Life cycle

The complex life cycle of malaria parasites (**Fig. 2**) begins when a mosquito takes a bloodmeal and injects sporozoites into the dermis of the vertebrate host. These motile forms penetrate into blood vessels and rapidly reach the liver, crossing the sinusoidal barrier and invading hepatocytes[30,31]. During the following 2-10 days, they undergo multiple rounds of asexual replication and generate a multinucleated merosome containing thousands of merozoites. At this point, some species such as *P. vivax* and *P. ovale*, produce the non-replicating latent liver stages known as hypnozoites[18].

The asexual intraerythrocytic development cycle (IDC) commences when merozoites egress from the liver, return to the bloodstream and attach themselves to red blood cells (RBCs)[32]. By inducing a vacuole derived from the cell's plasma membrane known as the parasitophorous vacuole, they gain entry into the cell and form a single nucleated stage named "ring". Over the following hours the ring undergoes remodelling and develops first into the "trophozoite" stage that accumulates hemozoin (the by-product of haemoglobin digestion) and second into the schizont stage, when the parasite replicates its DNA 16 to 32 times and then produces the same number of merozoites in a single round of replication. Once the IDC finishes 24 to 72 h later (depending on the *Plasmodium* spp.), the newly formed merozoites are explosively released from the infected red blood cell (iRBC) and are available to invade new erythrocytes[33,34].

Species (IDC duration[18])	Importance	Geographical distribution	Distinguishing characteristics
Plasmodium falciparum (48h)	Most prevalent and deadliest species[5].	Sub-Saharan Africa (99% of cases) and Southeast Asia (30% of cases)[19].	Capacity to sequester in different tissues. Easiest species to culture in vitro[20].
Plasmodium vivax (48h)	Most widespread species[21,22]. Can cause severe disease and triggers an acute inflammatory response[21].	Mainly Latin America and Southeast Asia[21,22]. Very low prevalence in Africa due to the spread of the protective Duffy-negative mutation[21,22].	Forms dormant liver stages (responsible for relapse and chronicity)[21]. Only infects reticulocytes and requires the Duffy antigen for invasion[21].
Plasmodium malariae (72h)	Most benign form of malaria. Often overlooked as it geographically overlaps with *P. falciparum*[23].	Found throughout the tropics and subtropics, especially prevalent in Sub-Saharan Africa[24].	Can produce chronic infections that last decades. Rare complications include: splenomegaly or renal damage[24].
Plasmodium ovale curtisi* & *Plasmodium ovale wallikeri (50h)	Burden of disease is unknown[25]. Death and severe illness are rare but there have been cases of splenic rupture and haematological alterations[26].	Both non-recombining sympatric species[25] are found in tropical eastern Africa, the Philippines, Indonesia and Papua New Guinea[23].	Form hypnozoites and produce round to oval shaped gametocytes[25]. Also infects reticulocytes but do not require Duffy surface antigen to invade[25].
Plasmodium knowlesi (24h)	Zoonotic species (major public health concern)[27]. Can cause severe disease leading to fatal complications with a higher frequency than other species[16,27].	The natural host of this species are macaques in Southeast Asia and its distribution is limited to this area[28]. First outbreak described in Malaysia in 2004[29].	Transmission requires exposure to mosquitos that have fed on macaques. No reported human-to-human transmission[16,28]. Morphologically indistinguishable from *P. malariae*[27].

Table 1. Characteristics of the different human infecting malaria parasite species. IDC = intraerythrocytic development cycle

Some asexual parasites undergo what is known as "sexual conversion" resulting in the formation of sexual forms termed gametocytes. This is a consequence of the activation of the transcription factor AP2-G, considered the "master regulator" of this process[35]. Before becoming fully infectious, male and female gametocytes mature during 1 to 12 days depending on the species. *P. falciparum* has the longest maturation period and its gametocytes go through five distinct morphological phases[36]. Stage I gametocytes accumulate in the bone marrow parenchyma where the iRBC membrane is completely remodelled resulting in an increase of cellular rigidity that traps gametocytes in this safe and nutrient-rich environment[34,37–39]. Once fully matured, stage V gametocytes recuperate membrane deformability and are re-released into the bloodstream. They circulate for a few days until they are ingested by a mosquito in a new bloodmeal.

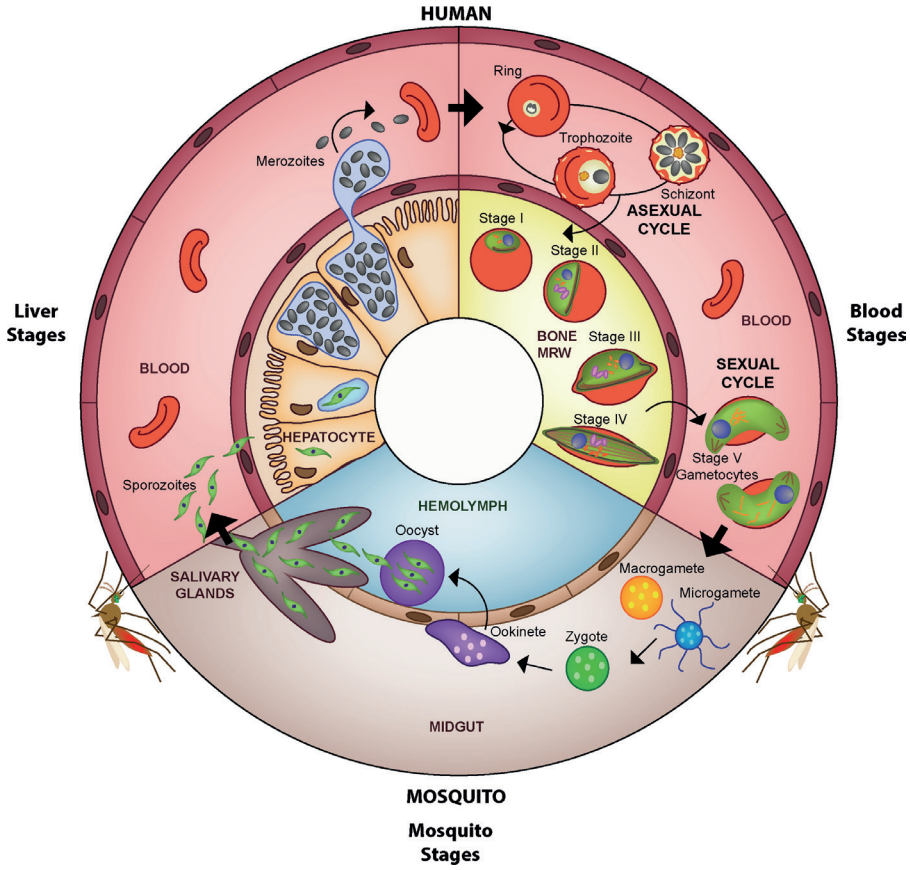


Fig. 2. Life cycle of *Plasmodium falciparum* (Adapted from Nilsson *et al.*[42]).

In the mosquito midgut, the changes in conditions between hosts triggers the rounding up of male gametocytes, followed by three rounds of mitosis with the generation of eight microgametes. These microgametes exflagellate and seek the single macrogamete that originates from each female gametocyte[40]. The two gametes fuse and form a diploid zygote, an important step that allows for the recombination of genetic material and the spread of genetic traits. Afterwards, the zygote elongates into an ookinete, a motile form that exits the lumen of the mosquito midgut across the epithelium where it establishes itself as an oocyst. A third round of asexual replication occurs producing thousands of sporozoites which are released into the haemolymph. They travel in this medium to the salivary glands, where they remain until the mosquito feeds again. On average 10-14 days pass before a mosquito can infect another vertebrate host and close the cycle[41].

1.4. Epidemiology

In 2019, the WHO reported approximately 229 million cases of malaria worldwide: 94% in Africa, 3% in Southeast Asia and the remaining percentage between the Americas, the Eastern Mediterranean and Western Pacific regions (**Fig.3**). The majority of disease burden was carried by sub-Saharan African countries and India, with the most affected countries being Nigeria, the Democratic Republic of the Congo, Uganda, Mozambique and Niger. In the same year, there were an estimated 409,000 malaria deaths, the majority in Africa and with children under the age of five accounting for 67% of mortality. The species *P. falciparum* was responsible for the majority of these cases and deaths. At present, malaria is still endemic in 87 countries and the collectives that are hit the hardest by the disease are pregnant women and children[19].

Malaria is transmitted to humans exclusively by female mosquitoes of the *Anopheles* genus. There are about 500 *Anopheles* species in the world, however, only 30 to 40 are considered important vectors of human malaria and can be found spread all over the planet[43]. In Africa, mosquitoes of the *Anopheles gambiae* complex such as *A. gambiae*, *A. arabiensis*, and *A. coluzzii* are the most efficient vectors. Other important species are *A. funestus* in Africa, *A. farauti* in Southeast Asia or *A. darlingi* in South America[44]. Additionally, two much less common routes of malaria transmission have been described: congenital and via blood transfusions in resource-poor settings[45].

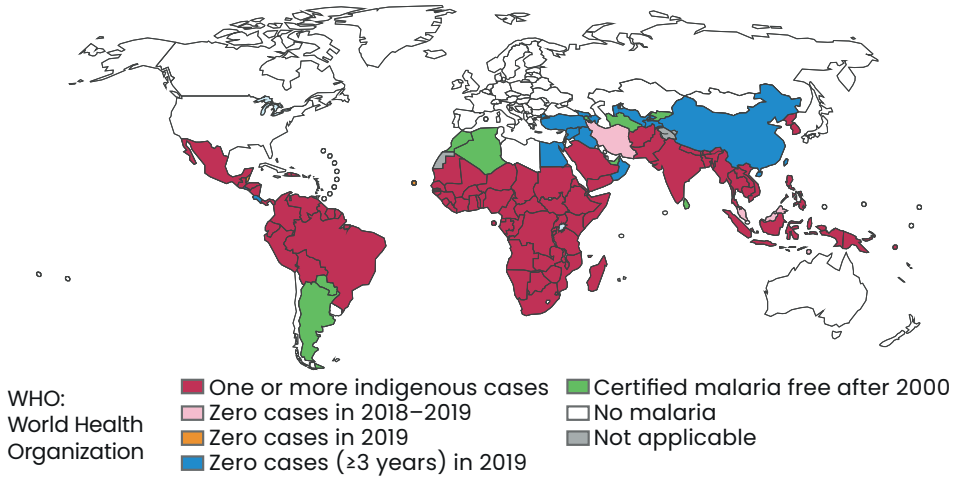


Fig. 3. Worldwide distribution of malaria cases in 2019 (Adapted from the WHO World malaria report 2020[19]).

Transmission of malaria depends on four components: the human, the vector, the parasite and the environment (**Table 2**). Each of these components plays an important role and together they have a strong impact on the distribution, intensity of transmission and seasonality of the disease[46].

Human component	Environment component	Vector component	Parasite component
Cultural beliefs Behaviour Perception of risk Political context Income levels Sanitation levels Access to healthcare Implementation of control measures Genetic or acquired immunity	Natural environment (temperature, humidity, rainfall, vegetation, natural disasters) Human-made environment (land use, deforestation, housing quality, local infrastructure, water sanitation, waste collection)	Species Biology Feeding behaviour (anthropophilic or zoophilic, nocturnal or diurnal, indoors or outdoors…) Vector density	Species Drug resistance Virulence Host flexibility Gametocyte density

Table 2. Components and subcomponents of malaria transmission.

1.5. Pathology

The main pathogenic mechanisms of malaria derive from the IDC and result in the following common symptoms of the disease. Firstly, the constant destruction of iRBC, due to either merozoite egress or spleen clearance, results in the liberation of parasites and endotoxins (combination of hemozoin and parasite DNA). This can trigger a signalling cascade that results in the appearance of fever, one of the most characteristic symptoms of malaria. Another common symptom is anaemia which can develop due to maintained haemolysis and other poorly understood mechanisms such as dyserythropoiesis. Finally, a more severe symptom is metabolic acidosis that results from the combination of the obstruction of capillaries with iRBC with reduced perfusion pressure because of anaemia, dehydration and hypovolemia. Furthermore, this can result in a decrease in the amount of oxygen reaching vital organs and, when left untreated, can lead to coma and death[20,47,48].

P. falciparum is the most virulent malaria species mainly because its trophozoite and schizont stages can cytoadhere and sequester inside small and medium-size vessels in various different organs including the lungs, adipose tissue, spleen and brain[49]. This enables parasites to avoid spleen clearance but also, depending on the location, can cause damage that results in severe disease[20]. The proteins involved in adhesion are mainly encoded by the *var* gene family and are called "*Plasmodium falciparum* erythrocyte membrane protein 1" or "PfEMP1"[50,51]. On the exterior of iRBC, PfEMP1 proteins are lifted above the dense coat of surface receptors by electron-dense structures called knobs, facilitating the interaction with different endothelial host receptors. Different PfEMP1 variants can bind to different host receptors resulting in different adhesion phenotypes and pathogenesis[52] (**Fig.4**). Lastly, iRBC can also bind to each other (rosetting) or to platelets (clumping) generating aggregates that can obstruct the microvasculature[53,54].

1.6. Clinical manifestations

Malaria has three main clinical presentations that range from asymptomatic to severe[20]. Asymptomatic patients are those that have circulating parasites yet present no symptoms, constituting a reservoir of disease that is difficult to detect. Uncomplicated malaria is characterised by very unspecific symptoms that complicate its differential diagnosis

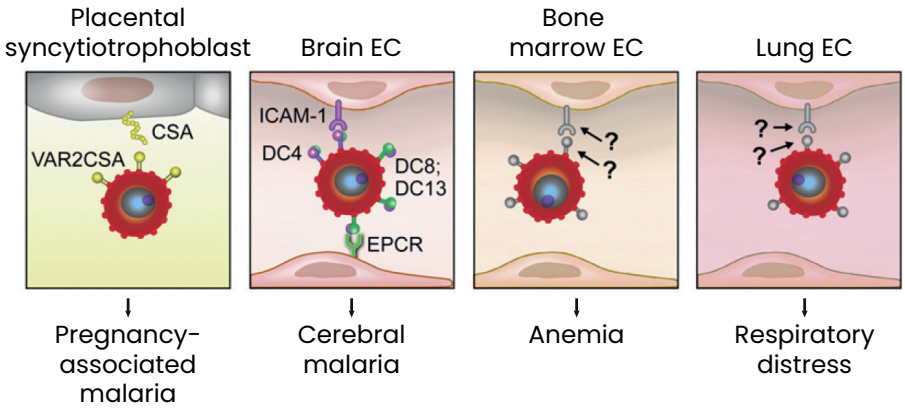


Fig. 4. Cytoadherence of *P. falciparum* to different tissues is mediated by binding specificity between different PfEMP1 variants and receptors of host endothelial cells (EC). Placental malaria is associated to the binding between the PfEMP1 variant 2-CSA (VAR2-CSA) and the chondroitin sulphate A (CSA) receptors expressed by placental syncytiotrophoblast[55]. Cerebral malaria may develop when PfEMP1 variants with specific domain cassettes (DC) bind to ICAM-1[56] or EPCR[57] receptors of brain EC. Similarly, other organ-specific malaria syndromes including respiratory distress and dyserythropoiesis anaemia, are also associated to sequestration of parasites in the lungs or bone marrow. (Adapted from Aird *et al.*[58]).

with other ailments. These symptoms include fever, chills, headaches, muscle pains, nausea, vomiting, diarrhoea and anaemia, among others[20]. Finally, severe malaria is a life-threatening condition due to the clinical deterioration of patients that can be marked by unconsciousness, respiratory distress, convulsions, prostration, shock, severe anaemia, jaundice, hypoglycaemia and metabolic acidosis[59], among others, and often it is associated with parasite sequestration in essential organs which produces more specific forms of the disease such as cerebral or placental malaria[60,61].

Regarding parasite species, *P. falciparum* is responsible for the most severe cases, though *P. vivax* and *P. knowlesi* can also cause serious complications[21,62]. Coinfections with other pathogens such as HIV, *Mycobacterium tuberculosis* and helminths (depending on the species) increase the risk of severe disease and fatality[63–65]. Furthermore, the prevalence of concurrent invasive bacterial disease in African children with severe malaria is often very high and congenital transmission of HIV has also been described to increase in malaria-infected pregnant women[20].

Host immunity is another factor that influences the clinical manifestation of malaria. In high transmission settings, morbidity and mortality are highest in young children between 6 months and 5 years of age, coinciding with the loss of maternal antibodies. Disease severity declines rapidly with age, as partial protective immunity is acquired, and most adults eventually become asymptomatic carriers of parasites, as it is considered that sterile immunity is never achieved[66] (**Fig. 5**). In low transmission areas with seasonal malaria, individuals do not acquire such high levels of protective immunity and exposure to these parasites results in more severe forms of the disease[20,66].

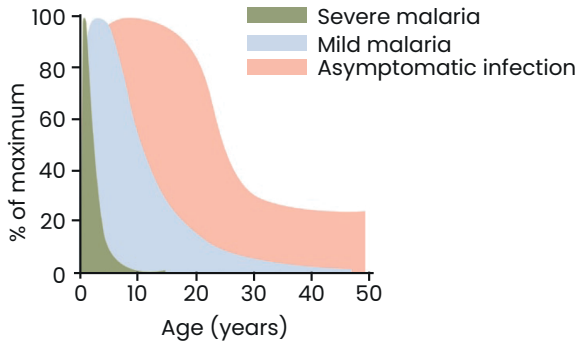


Fig. 5. Age-dependence of malaria clinical manifestations in a population from a high transmission zone. More detail in the text. (Adapted from Langhorne *et al.*[66])

1.7. Diagnosis

As malaria symptoms can be quite unspecific, sensitive diagnostic tools are necessary to detect and control the disease. At present there are three main malaria diagnostic tools[67]. The first is the examination of thick and thin blood smears using light microscopy. It is considered the classic method of detecting malaria parasites and is still used worldwide for speciation and quantification. This method is highly sensitive (it can detect as few as 10 parasites per microlitre of blood), relatively cheap and does not require huge technological investments. On the downside, it calls for trained professionals, it is time consuming, its sensitivity is reduced in field settings and it is not practical for broad-scale malaria detection[67–69].

The second type of tools are rapid diagnostic tests (RDTs) that enable the detection of parasite proteins that are released during a malaria infection,

such as the highly expressed histidine-rich protein 2 (PfHRP2) or lactate dehydrogenase (LDH). They are easy to use, do not require training and are practical for use in the field. Their main inconvenience is their sensitivity and often results need to be confirmed by microscopy, as WHO validated RDTs can detect approximately 100 parasites per microlitre of blood[68,69].

Finally, polymerase chain reaction (PCR) is the most sensitive tool that can detect very low levels of parasitaemia (as low as a single parasite per sample)[68]. However, this method requires advanced technology and it is slow, expensive and impractical for field settings. That said, considerable advances are being made for the adaptation of this technique to be used in field-based labs and for surveillance[67].

1.8. Treatment

Despite there being no universal drug with which to treat all malaria cases, there is a wide selection of antimalarials with different modes of action that can be used to fight the disease[70] (**Table 3**). When establishing a correct therapeutic regimen, it is necessary to take into account the variables of each case such as the *Plasmodium* spp. involved, the parasite stages that need to be targeted, disease severity and host factors (pregnancy, drug incompatibilities...), among others.

The initial treatment for severe falciparum malaria is intravenous artesunate and supportive management in order to stabilize the patient[59,71]. Once the patient is stable, oral artemisinin-based combination therapy (ACTs) follows, which is also the treatment of choice for uncomplicated malaria[71]. ACTs consist in the combination of a fast-acting and "short-lived" artemisinin derivative that rapidly reduces parasitaemia but is quickly eliminated due to its short in-vivo half-life, with another longer-lasting antimalarial that continues to act against "leftover" parasites once the other drug has been eliminated. The leading ACTs are: artemether + lumefantrine, artesunate + amodiaquine, dihydroartemisinin + piperaquine, artesunate + mefloquine and artesunate + sulfadoxine pyrimethamine[5,71]. Despite their high efficiency, resistance to these drug combinations in the form of delayed parasite clearance has started to appear in parts of Southeast Asia[72]. Drug resistance can be defined as "the ability of a parasite to survive or multiply despite properly administered and dosed medication"[73]. High parasite replication rates and genome mutability favour the apparition of random mutations that can lead to antimalarial resistance within only a few cycles of

replication. Subsequently, these mutations are selected in the population due to the survival advantage they confer in the presence of drug pressure[74]. The selection of resistant parasites can be also favoured by other factors such as high parasitaemias, sub-curative drug concentrations and decreased patient immunity[74,75].

Drug	Mode of action
Aryl-amino alcohols Quinine, quinidine, mefloquine, halofantrine, lumefantrine	Inhibition of parasite-mediated haeme detoxification. Mefloquine is also thought to inhibit endocytosis.
4-Aminoquinolines Chloroquine, amodiaquine, piperaquine	Prevents degradation of haemoglobin into hemozoin resulting in the toxic accumulation of haeme in the parasite's food vacuole.
Antifolates Sulfadoxine, pyrimethamine, chloroproguanil	Inhibition of dihydropteroate synthase (DHPS) or dihydrofolate reductase (DHFR), enzymes involved in folate biosynthesis required for nucleic acid synthesis.
8-Aminoquinolines Primaquine	Production of reactive oxygen species (ROS) and interference with electron transport and metabolic processes in the mitochondria.
Antibiotics Tetracycline, doxycycline, clindamycin	Inhibitors of protein synthesis by binding to different rRNAs.
Artemisinin-type compounds Artemisinin, artemether, arteether, artesunate	Production of reactive oxidative species (ROS) leading to cellular damage and parasite death. Also, prevention of the formation of hemozoin by alkylating haeme.
Naphthoquinones Atovaquone	Competitive inhibitor of ubiquinol, it inhibits the mitochondrial electron transport chain at the cytochrome bc1 complex

Table 3. Principal antimalarial drugs and their mode of action[70,76–81].

The drug of choice for other *Plasmodium* spp. is chloroquine, with the exception of *P. vivax* strains which are resistant to this drug and require ACTs[82]. Infections with species that produce hypnozoites must also be treated with primaquine, the only drug that eliminates these forms and avoids disease relapse[83]. Finally, in response to the emergence of drug resistance to nearly all current antimalarials and thanks to recent funding, there has been an increase in the research of new antimalarials[84]. Although most of the drugs under development are blood schizonticides[45], there have

also been efforts to develop alternative transmission-blocking drugs that mainly target sexual stages as an elimination tool[84].

1.9. Control and prevention

Malaria control refers to the reduction of disease burden to a level at which it is no longer a public health issue[4]. At present, three types of interventions are being put into practice: vector control, chemoprophylaxis or chemoprevention and vaccines.

The two principal vector control methods are insecticide-treated bed nets and indoor residual spraying[45]. The first protect the people sleeping beneath them from mosquito bites and the second kills mosquitos that rest on the walls after taking a bloodmeal. However, neither of these methods contemplate outdoor-feeding mosquitoes, a control gap that can only be filled by eliminating larva breeding sites or using larvicides[85]. The four classes of insecticides available for mosquito control (organochlorines such as DDT, carbamates, organophosphates and pyrethroids) all act by blocking nerve synapses, causing instant paralysis and death[19]. Due to their low toxicity, the most widely used are pyrethroids and this has led to the appearance and spread of resistance in many mosquito populations[19]. Alternative vector control methods being researched are genetically modified mosquitoes that are either refractory to *Plasmodium* spp. infection or that generate sterile progeny[86]. These genetic modifications can be spread in a population through artificial genetic engineering systems named "gene drives"[87]. Whilst very promising for malaria elimination and more environmentally friendly than mass insecticide spraying, it is very difficult to predict the ecological impact of these methods[86].

Next, chemoprophylaxis and chemoprevention are more focused on the parasite[45]. The first refers to antimalarials being given to individuals who travel from a malaria-free area to an endemic one and the second to large scale campaigns in which different drug combinations are given to a specific target group. For example, children under five who live in areas with a high risk of seasonal malaria, pregnant women (intermittent preventive treatment - IPTp) and general mass drug administration campaigns[88,89].

The third major and most appealing control intervention is the development of a malaria vaccine that provides life-long immunity and would constitute a huge step towards malaria eradication[90]. However,

natural infections only produce partial immunity and the complexity of malaria parasites makes it very difficult to establish an effective vaccine target[85]. Currently, there is a large pipeline of vaccine candidates under evaluation[91] (**Fig. 6**), yet the one with the highest chances of being approved in humans is the RTS,S vaccine[92]. This infection-blocking vaccine that targets the *P. falciparum* circumsporozoite protein started its full development in the 1990s and since then has undergone a large number of clinical trials reaching Phase III[93]. Its efficacy is between 30 and 50%, which, although far from ideal, can still have a significant impact on public health and save many children's lives, especially in combination with other interventions[93,94].

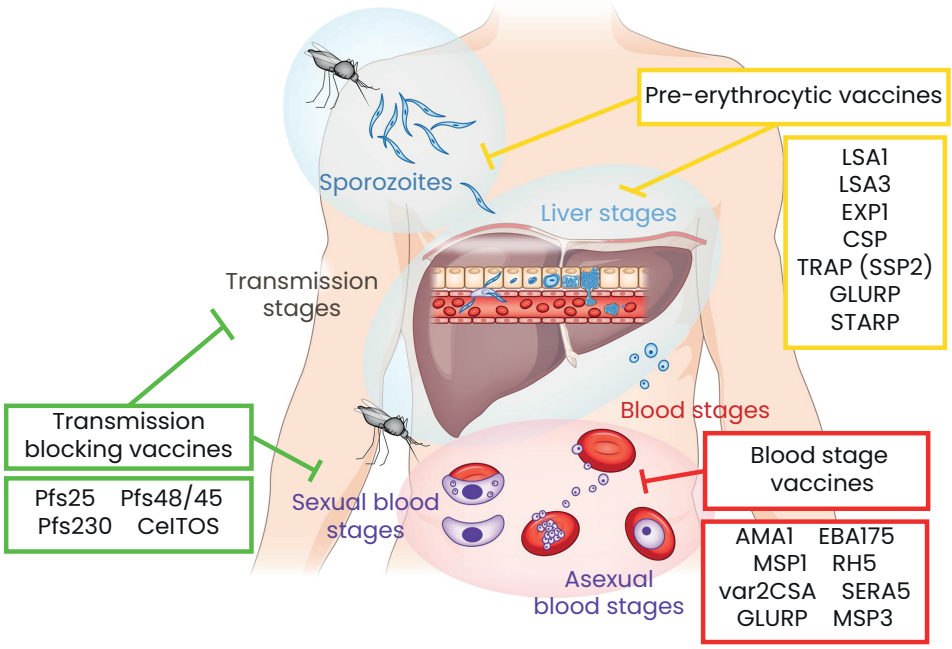


Fig. 6. Different types of malaria vaccines candidates and their target antigens currently under evaluation in clinical trials. Pre-erythrocytic vaccines target sporozoites and liver stages with the objective of blocking infection. Blood stage vaccines target the asexual blood stages and could potentially reduce disease morbidity and mortality. Transmission blocking vaccines target gametocytes and mosquito-stage parasite surface antigens or mosquito proteins involved in parasite interactions in the midgut and constitute an altruistic approach as they do not protect the vaccinated individual directly but prevent the spread of malaria in the community. (Adapted from Barry *et al.*[95] and Cockburn *et al.*[91].)

Last but not least, all the aforementioned interventions address only two of the four disease components: the parasite and the vector. However, the human and the environment component should not be forgotten. For example, involving the community in endemic areas and inducing a collective response are essential for the correct implementation of other measures[85]. Furthermore, permanently eliminating mosquito breeding habitats and improving housing are also key for stopping transmission. These interventions are effective and can have long term impact on malaria transmission. However, they unfortunately have a high short-term cost and time investment[85].

1.10. Context of malaria elimination

Malaria elimination refers to "the interruption of local, mosquito-borne malaria transmission in a defined geographical area", while malaria eradication can be defined as "the permanent reduction to zero of the worldwide incidence of infection caused by a specific malaria parasite species"[4]. The current WHO objectives for 2030 are "to reduce malaria case incidence and mortality rates by at least 90%, eliminate malaria in no less than 35 countries and prevent a resurgence of malaria in all countries that are malaria-free"[96]. Although malaria incidence has been reduced significantly over the last decade, progress is stalling and the 2030 objectives are beginning to fall out of reach[19,97]. In its current 2020 annual malaria report, the WHO lists the following challenges that must be addressed: 1) malaria funding for control and elimination efforts as well as for research must be maintained; 2) the current worldwide delivery of millions of insecticide-treat nets, rapid diagnostic tests and ACT treatment courses has to be continued; 3) various biological threats such as drug resistance, insecticide resistance, genetic mutations that alter the targets of RDTs and detection of asymptomatic individuals, among others; 4) the current COVID-19 pandemic is the newest menace to appear and has caused major disruptions in essential malaria services in endemic countries; and 5) the impact of human-related factors such as socioeconomic status, migration and behavioural issues.

In this thesis, we focus on the biological threat that arises from the adaptive plasticity of malaria parasites. A better understanding of how these pathogens adapt to different environments with fluctuating conditions could provide invaluable insight on how to continue fighting against this disease.

2. Adaptation

Before analysing how malaria parasites adapt to their environment, it is important to understand some general concepts on adaptation first. In this section we will describe the three main types of molecular mechanisms that mediate adaptive processes, followed by how these mechanisms are used in different adaptive strategies.

2.1. Molecular mechanisms of adaptation

The term "adaptation" can be described as which process that occurs when an organism is challenged by new conditions in its environment and that results in phenotypic changes that enable it to survive[98]. Adaptation can occur on a cellular level, when cells recover from a perturbation by making changes that help them regain their state of homeostasis, and at a populational level, when organisms evolve towards the maximum fitness within their environment[99].

At the cellular level, adaptive changes include receptors for sensing environmental factors, different signalling cascades, transcriptional or posttranscriptional mechanisms, translational or posttranslational changes and variations at the metabolic level, among others[98]. Depending on how these adaptive changes come about, there are three main types of adaptive molecular mechanisms: directed transcriptional responses (DTRs), epigenetic modifications and genetic mutations[100]. Each type of adaptation can be defined by a series of parameters (**Fig. 7**).

Firstly, there is the time it takes for the organism to become adapted to a new stress, which is usually shorter in DTRs and epigenetic adaptations while longer in the case of genetic mutations. Next, there is the duration of the effects of the adaptation even after the selective pressure has disappeared. Whilst DTRs generally occur while the environmental trigger is present and then gradually disappear; epigenetic and genetic adaptations can produce changes that long outlive the initial stress or can even precede it. Also, the reversibility of the adaptation follows a similar pattern with DTRs being more easily reversible than epigenetic ones and these more so than genetic mutations. Finally, there is the heritability of the process and the accuracy with which it can be transmitted to future generations, both of which are highest in genetic mutations though also possible in epigenetic adaptations[100].

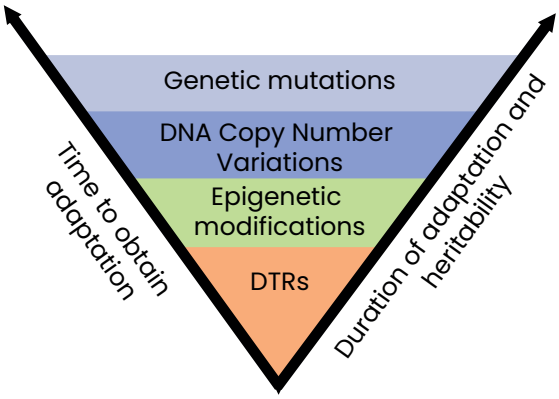


Fig. 7. Adaptation mechanisms ordered by two timescales: the time to acquire the adaptation (left axis) and the time the adaptation can be maintained in absence of the original condition that induced it (right axis).(Adapted from Yona *et al.*[100])

2.1.1. Directed transcriptional responses

DTRs are normally the first-line of defence against changes in biochemical homeostasis such as variations in temperature, pH, osmotic pressure or nutrient availability[100]. Upon sensing a change in the environment, a temporary signalling cascade resulting in changes in gene expression is put into motion and then eventually stops once the environmental challenge disappears (**Fig. 8**). Consequently, the adaptation is not actively passed on to future generations. These responses are normally characterized by the activation of "stress-genes" like chaperones or anti-oxidant enzymes[101,102].

2.1.2. Epigenetic mechanisms

2.1.2.1. Epigenetics: general concepts

"Epigenetics" refers to the heritable differences between cells that do not involve alterations in the primary DNA sequence[103]. Epigenetic mechanisms are involved in a number of fundamental biological processes that range from cellular programming during development to transcriptional changes in response to different environmental stimuli determining cellular differentiation and adaptation. Moreover, epigenetic alterations have also been associated with many types of diseases including cancer[104].

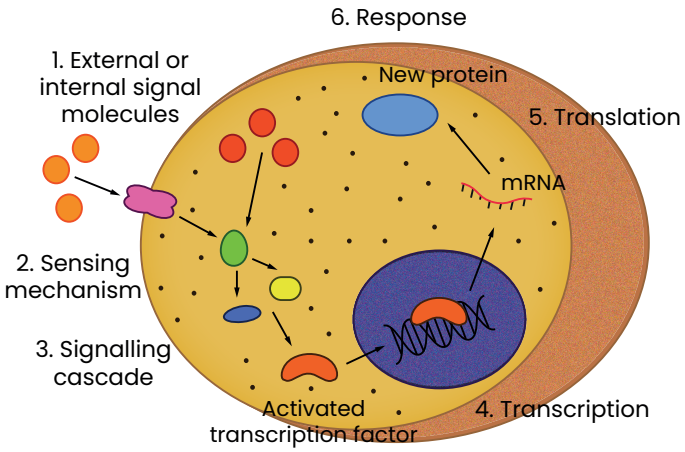


Fig. 8. DTR model. Changes in biochemical homeostasis produce different external or internal signal molecules that are detected by cells through different sensing mechanisms (e.g., membrane receptors). Next, the signal is transmitted through signalling cascades conformed of a series of molecular events that ultimately result in a cellular response such as changes in gene transcription or translation and post-translational modifications of proteins, among others.

Epigenetic characteristics can be inherited during mitosis and passed directly from one cell to the next, as well as through transgenerational inheritance in multicellular organisms. All in all, epigenetics encompasses "the study of molecules and mechanisms that can perpetuate alternative gene activity states in the context of the same DNA sequence"[104].

In eukaryotic cells, DNA is found compacted into the nucleus in the form of chromatin, a nucleoprotein complex that is condensed even further to form chromosomes. The conformation of chromatin is achieved through the interaction of DNA with proteins known as histones and together they form a basic unit called nucleosome. Each nucleosome is formed of 146 bp of DNA wrapped around a histone octamer composed of two copies of the following types of histones: H2A, H2B, H3 and H4. The linker histone H1 binds the nucleosome at the entry and exit sites of the DNA, locking it into place, allowing for the formation of a higher-order chromatin structure[105] (**Fig. 9**).

This organization of genetic material allows cells to modulate the accessibility of DNA for transcription and there are different types of chromatin depending on its grade of compactness and its consequent

transcriptional accessibility[107]. These types range from the open and transcriptionally active euchromatin to the tightly compact and principally transcriptionally inactive heterochromatin. Changes in the conformation of chromatin can lead to changes in gene expression and they can be brought about by three main mechanisms: DNA methylation, histone modification and non-coding RNA interactions.

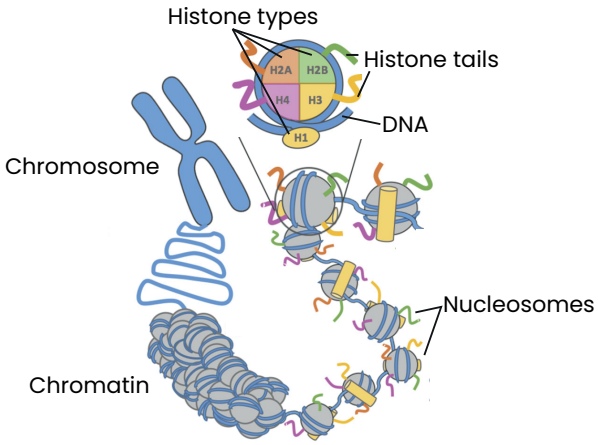


Fig. 9. DNA organization in nucleosomes, chromatin and chromosomes. More details in the text. (Adapted from Weaver *et al.*[106]).

2.1.2.2. DNA methylation

In eukaryotes, the most common form of DNA methylation consists in the addition of a methyl group onto the position 5 of a pyrimidine ring on cytosines, most frequently in cytosine-phosphate-guanine (CpG) dinucleotides[103] (**Fig. 10**). When DNA methyltransferases add methyl groups, they induce structural changes to chromatin by attracting protein complexes that favour the formation of heterochromatin[103] and therefore, this DNA modification is often transcriptionally repressive[108,109].

During cell division, methylation patterns are faithfully propagated through DNA replication as methyltransferases copy the methylation pattern of the parental strand symmetrically onto the unmethylated newly-synthesized strand[111]. However, the capacity to inherit these methylation patterns across different generations (e.g., in higher eukaryotes) is limited, given that they are mostly erased and then re-established in the early embryo[112].

Finally, there are organisms such as *Drosophila melanogaster* that lack the canonical cytosine methylation patterns[113] and possess alternative DNA methylation modifications such as 5-hydroxymethylcytosine (oxidation of methylated cytosine) and N6-methyladenine (adenine methylation) that also have epigenetic regulatory functions[111].

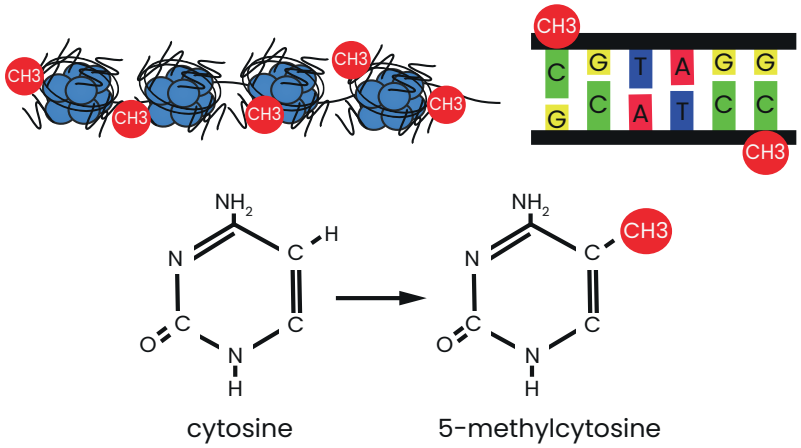


Fig. 10. DNA methylation of cytosine bases occurs when a DNA methyltransferase adds a methyl group at the 5' position on the pyrimidine ring (Adapted from Mukherjee *et al.*[110]).

2.1.2.3. Histone modifications

Histone modifications are covalent, post-translational, ATP-dependent changes to histone residues. These residues are situated on either the N-terminal histone tails that project outwards from the nucleosome or the central globular domains of histones found on the "sides" of nucleosomes. Histone post-translational modifications include acetylation, methylation, phosphorylation, ubiquitinylation, sumoylation, ADP ribosylation, and deamination, among others[114]. Each modification is catalysed by one or more protein-modifying enzymes and has varying effects on elements such as chromatin compaction and accessibility, nucleosome interactions and recruitment and blocking of transcription factors[115] (**Fig. 11**). Studies in fission yeast have shown that histone modifications can be inherited across many cell generations, independently of DNA sequence, DNA methylation, or RNA interference[116,117]. Consequently, histone marks constitute epigenetic information that can be perpetuated, even after the removal of the initiating trigger, allowing for adaptation memory.

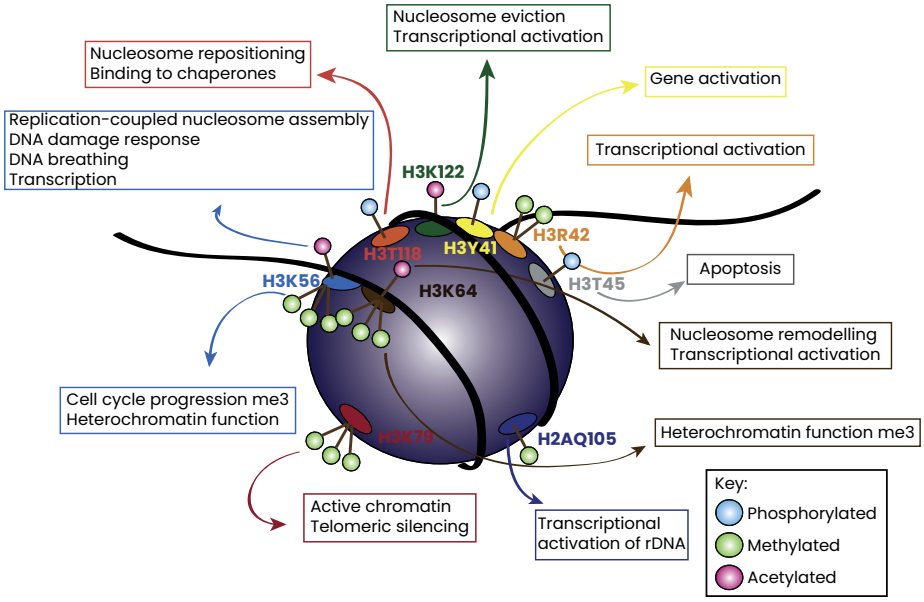


Fig. 11. Nucleosome representation detailing the location and function of some of the key modifications of histone globular domains. (Adapted from Lawrence *et al.*[115]).

2.1.2.4. Non-coding RNA interactions

In many organisms, intergenic or antisense transcription results in different classes of non-coding RNAs (ncRNAs). These RNA molecules do not encode for proteins but rather play important roles in gene expression, genome stability and defence against foreign genetic elements[118]. Multiple types of ncRNAs exist, from microRNAs (miRNAs), small interfering RNA (RNAi), small nuclear RNA (snRNA), piwi-interacting RNA (piRNA) to long ncRNA (lncRNA). There are several mechanisms through which ncRNAs can modulate gene expression: regulation of DNA methylation[119,120], participation in chromatin remodelling by recruiting chromatin-modifying complexes and consequently activating or silencing transcription[121], contributing to epiallelic interaction by modifying the expression of different alleles[122,123] and involvement in RNA editing and alternative splicing[124].

2.1.2.5. Role of epigenetics in adaptation

In addition to many other roles, epigenetic mechanisms play an important part in the adaptation of different organisms to different

environmental stimuli[100]. Changes in chromatin conformation with the consequent alterations in gene expression can result in phenotypic changes that enable organisms to overcome stressful conditions. Epigenetic mechanisms are located in the middle of the adaptive spectrum. On the one hand, they are relatively quick and can appear within the same cell generation as like DTRs[104]. On the other, as with genetic mutations, they can be maintained for longer periods of time and are heritable through mitosis and meiosis. However, unlike genetic mutations, epigenetic mechanisms are reversible and are a more flexible type of adaptation[100].

Examples of epigenetic adaptation have been well established in plants and model organisms such as *Caenorhabditis elegans*. For instance, in plants, exposure to different environmental stresses such as water deficit, viral infection, heavy metal exposure or increased saline concentrations can result in adaptation mediated by changes in DNA methylation patterns that can be passed on to following generations[125]. In *C. elegans*, histone acetylation in response to early-life heat stress promotes long-lasting defence responses and longevity[119] and RNAi has been found to help this organism cope with environmental stresses by silencing an infectious viral genome[126]. In addition to these, there are many more examples of epigenetic mediated adaptation in lots of different organisms, including malaria parasites, and constitutes the key interest of this thesis.

2.1.3. Genetic mutations

Genetic mutations are heritable variations in the DNA sequence that are most commonly associated with species evolution by the means of natural selection[100]. Mutations constitute a source of phenotypic variability that enable an organism to survive in a certain environment and which can then be transmitted to its descendants, resulting in long-term, stable adaptation. Mutations can appear spontaneously due to sporadic errors during DNA replication and recombination or be caused by external factors such as chemical damage or radiation. Subsequently, it is possible for mutations to precede a selective change in the environment, as well as for them to be the result of environmental stresses[127]. Stress-induced mutagenesis potentially accelerates adaptation by promoting random mutations when organisms are poorly adapted to their environments and is often restricted to specific genomic regions, in which mutagenesis mechanisms are associated with processes such as DNA-break repair, replication and transcription[127,128].

Mutation rates have been described to depend on the balance between the ability to adapt to new environmental conditions (adaptability) and the capacity to remain adapted to existing conditions (adaptedness)[129]. In well-adapted organisms, the majority of mutations that arise are deleterious or neutral and in a stable environment, low, constant mutations rates are considered beneficial. However, in the opposite scenario in which organisms must overcome many or frequent selective pressures, higher mutation rates can prove more advantageous[128].

There are primarily two main types of genetic mutations: DNA copy number variations (CNVs) and point mutations.

2.1.3.1. DNA copy number variation

DNA CNVs can be duplications or deletions that range from specific genes to whole chromosomes (aneuploidy). They are relatively labile and although they do not involve actual changes in the nucleotide sequence, they can originate important quantitative variations in gene expression leading to altered mRNA and protein levels[100]. For example, human populations with high-starch diets have more copies of the human salivary amylase gene AMY1, as this is positively correlated with the production of salivary amylase protein[130]. As with other genetic mutations, CNVs are often fortuitous events and therefore are frequently imprecise. For example, aneuploidy is a common by-product of chromosomal mis-segregation during meiosis and mitosis and can consequently alter the expression of many other irrelevant genes, imposing a significant burden on the cell[131]. Aneuploidies can have detrimental effects on cell growth and proliferation, yet if the initial gene imbalance is overcome, they can be advantageous for adaptation[132]. For instance, *Candida albicans* acquires chromosome five aneuploidy in response to antifungal drug exposure, yet as soon as the drug is removed the "costly" extra chromosome is eliminated from the population[133].

2.1.3.2. Point mutations

Point mutations are small variations in the DNA sequence that range from single-nucleotide polymorphisms (SNPs) with the substitution of one base for another, to small insertions and deletions (indels) that consist in the addition or elimination of one or various bases. The consequences of point mutations can vary widely depending on various factors. With regards

to location, point mutations in promoter regions can affect the regulation of gene expression, while those that fall in coding sequences may alter protein sequence and function. Depending on their impact, point mutations can be silent (no effect as the DNA sequence change does not result in an amino acid switch), missense (change in DNA sequence results in an amino acid swap that may or may not produce alterations in the final protein structure) or nonsense (change in DNA sequence that produces an early STOP codon and protein truncation)[134]. As with CNVs, examples of point mutations associated to adaptation can be found in all organisms in response to a huge variety of environmental stresses.

2.1.4. Interaction between different adaptation mechanisms

It is important to note that these three adaptation mechanisms are not totally independent from one another. In reality, they are closely linked in such a way that earlier transient adaptations can lead to or even facilitate later more durable ones[100] (**Fig. 12**). For example, DTRs can induce chromosome-based epigenetic modifications[135] and CNVs can facilitate the apparition of mutations in the DNA[136]. In addition, DNA methylation, nucleosome positioning and higher-order chromosome folding can affect mutation rates[104]. This interplay increases the effectiveness of an adaptation as it allows organisms to home towards the most efficient mechanism with which to confront certain environmental challenges.

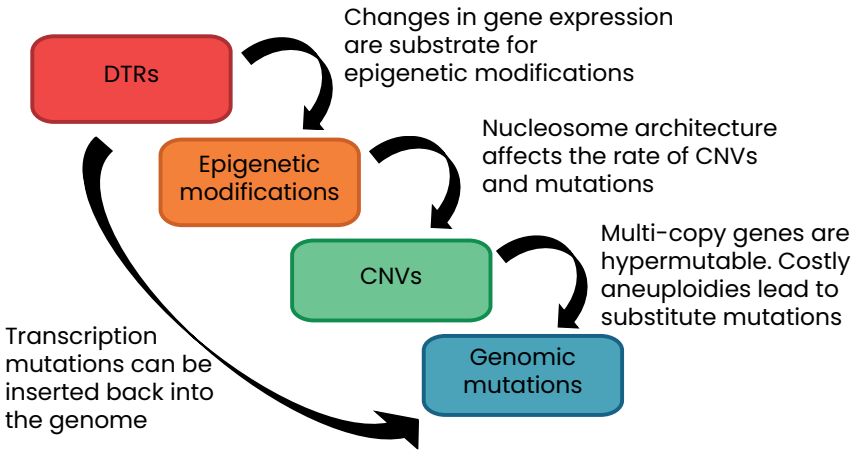


Fig. 12. Examples of how different adaptation mechanisms can lead to other more durable ones (Adapted from Yona *et al.*[100]).

2.2. Adaptive strategies

Depending on the predictability of environmental changes, organisms have developed different evolutionary strategies: adaptive plasticity, bet hedging and fixed strategies. All three have been described in many different taxa[137] and, in principle, can entail any of the aforementioned adaptive mechanisms. On occasions, organisms can use a combination of these strategies to adapt to different situations making it complicated to classify them in a single category.

2.2.1. Adaptive plasticity

Adaptive or phenotypic plasticity can be defined as the capacity of a single genome to exhibit a range of different phenotypes in response to variations in the environment[138]. For example, in vinification adaptive plasticity enables the yeast *Saccharomyces cerevisiae* to successfully carry out fermentation in a wide range of conditions (e.g., initial composition of the grape, oxygen and temperature, among others)[139]. Adaptive plasticity can be active or passive depending on whether the adaptation occurs in anticipation or as a direct consequence of the environmental challenge[138]. An example of active adaptive plasticity is environmentally-induced diapause or programmed dormancy in insects in response to cues that happen well in advance of the environmental stress (e.g., changes in photoperiod, thermoperiod or plant allelochemicals)[140]. In contrast, an example of passive adaptive plasticity could be a reduction of growth due to the concurrent decrease in resource levels[138]. Whilst active plasticity is more beneficial as it enables organisms to adapt before the environmental stress has a deleterious effect on them, it is dependent on the presence of a predictive environmental cue that does not always exist[141].

This potential for an organism to produce a range of fit phenotypes in multiple environments can be hugely advantageous but it does come at a cost. For example, this type of adaptive strategy requires the investment in sensory systems that detect environmental conditions and in mechanisms that enable organisms to adjust their phenotype in response to the captured information[142,143]. Consequently, adaptive plasticity is more advantageous in environments that vary in a predictable manner, as organisms can only present different phenotypic solutions to the range of environments they normally encounter. For instance, populations of the water flea *Daphnia magna* have evolved adaptive behavioural plasticity

in response to predators' different chemical cues, yet these responses are absent in populations that have never encountered said predators[144].

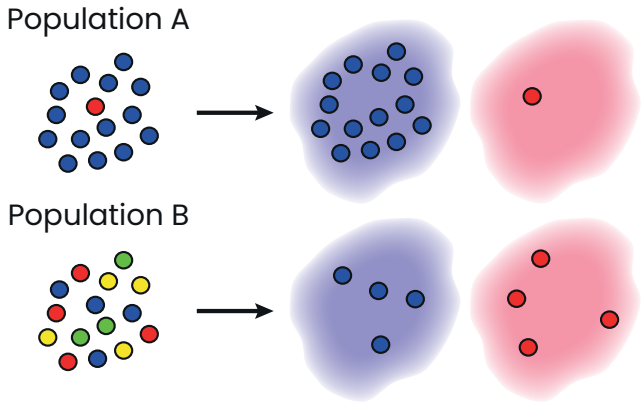
2.2.2. Bet hedging

Bet hedging is a risk-spreading strategy in which a population maximizes its geometric mean fitness at the expense of its arithmetic mean fitness across different environments[145] (**Fig.13**). In this manner, a homogenous population that is perfectly adapted to a specific environment will be fitter in this particular environment than a heterogenous population composed of individuals with different levels of adaptation. However, when these same populations come across a new environment, the heterogenous population will be fitter than the homogenous one. Therefore, bet hedging is a good adaptive strategy for unpredictable fluctuating environments in which organisms have a better chance of surviving if they diversify their phenotypes rather than invest in a single one[137,144,146–149].

Bet-hedging strategies require a proportional switching rate between phenotypes that are neither optimal nor inadequate and which is totally independent of any environmental cue. This is because they are not focused on the adaptation to the current set of conditions but rather on guaranteeing the population's survival under different potential conditions that may or may not occur[150]. Consequently, this adaptive strategy is about long-term investment as it prepares a population for future environments that the initial generations may never encounter. This makes bet hedging especially advantageous in large populations with exponential replication rates such as bacteria, yeast or unicellular parasites[151]. In these organisms the survival of individuals is not as important as that of the population. This contrasts with adaptive plasticity which is centred on the generation of an optimal phenotype in response to certain environmental cues, optimizing fitness within a single generation and allowing for the survival of the individual challenged with the environmental change[137]. Furthermore, in unicellular organisms the generation of specific responses to different unpredictable environmental cues is often more costly than the stochastic generation of phenotypic variability[146]. Subsequently, even in genetically identical cell populations there is a tendency towards stochastic phenotypic cell-to-cell variability. This is what is known as cellular decision making, "a process whereby cells spontaneously assume different, functionally important and heritable fates without an associated

genetic or environmental difference"[152]. Overall, bet-hedging strategies offer unicellular populations huge adaptive potential[146].

Examples of bet hedging in unicellular organisms include stochastic sporulation in *Bacillus subtilis*, in which some cells sporulate and become resistant to numerous environmental challenges, while others continue growing and are able to react more quickly to a positive change in conditions[153]. Similarly, in yeast different individuals of *Saccharomyces cerevisiae* populations have been shown to present different growth rhythms that enable them to survive various external stresses[147]. However, higher eukaryotes also present bet-hedging strategies such as "germ banking" or diversification in the timing of seed germination or egg hatch, multiple mating, stochastic number of ovules per flower, iteroparity or multiple reproductive cycles over the life time of a single organism, among others[144,154,155].



	Fitness blue	Fitness red	Arithmetic mean	Geometric mean
Population A	14	1	7.5	3.7
Population B	4	4	4	4

Fig. 13. Example of the bet hedging definition. Population A is better adapted and has a higher fitness in the blue environment while population B that follows a bet-hedging strategy has the same fitness level in both blue and red environments. Overall, population A has a higher arithmetic fitness due to its high level of adaptation to the blue environment, but population B has a higher geometric mean across the two environments and has a better chance at surviving in both rather than just in one.

2.2.3. Fixed strategies

Fixed strategies consist in organisms adopting the same behaviour or phenotype regardless of the environment or as a consequence of an unchanging environment. In the first case, adaptation independent of the environment can be beneficial in unpredictable environments where a lag between detection and response to external stresses may result in a high biological cost[137]. For example, periodic reproductive cycles in females that are independent of environmental stimuli and enable the propagation of a species regardless of external conditions.

In the case of unchanging environments, local adaptation enables the development of the most adequate phenotype to a certain set of invariable conditions. For example, in order to survive the dry season that is characteristic of some climates, plant species have developed mainly three different types of fixed strategies: "drought escape" in which they complete their life cycle before the decrease in water availability; "drought avoidance" that enable species to maintain their water content regardless of soil water availability; and "drought tolerance" which entails a range of adaptations that mitigate the damages produced by desiccation and enables recovery after rehydration[156].

2.2.4. Final example

The following example using plant flowering can be used to visualise the differences between adaptive plasticity, bet hedging and fixed adaptative strategies (**Fig.14**). In the case of adaptive plasticity, the plant would undergo an environment-appropriate adjustment of its flowering time dependent on a seasonality cue: for example, rice flowering occurs in response to photoperiod variations[157]. If the plant were to follow a bet-hedging strategy, different plants in the population would stochastically flower at different times and depending on the conditions of the season different individuals would thrive. For instance, in semelparous perennial species with one reproductive episode per lifetime, individuals initiate flowering at a range of different ages so that their seeds are dispersed in time[158]. Finally, a fixed adaptive strategy would consist in plants flowering independently of an environmental cue or always at the same time in an unchanging environment. This is the case of *Brassica rapa* that flowers at an earlier developmental stage in order to complete its life cycle before the onset of the dry season[159].

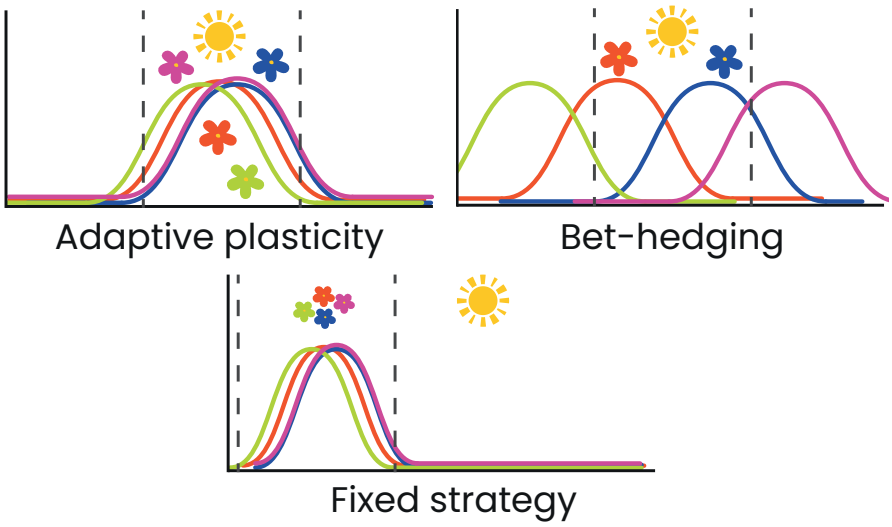


Fig. 14. Example of the three adaptive strategies for a plant's flowering time. The x axis represents time and the y axis "flowering potential". In the case of adaptive plasticity, flowering potential is adjusted to the environmental cue (represented by the sun) in order to make the most of optimal environmental conditions (represented by dotted lines). In a bet-hedging strategy, flowering occurs at different times independently of the environmental cue but is more successful when the conditions are optimal. Finally, in a fixed strategy flowering occurs always at the same time in optimal conditions in an unchanging environment and is independent of an environmental cue.

3. Fluctuating conditions of the human host circulation

Because of their complex life cycles, malaria parasites are exposed to a wide range of environments as they transition between different hosts (invertebrate and vertebrate), tissues and cells. Moreover, within each of the aforementioned environments these parasites encounter conditions that can also vary. As a result, the adaptive capacity of *Plasmodium* spp. is fundamental to ensure their viability and life cycle progression. Furthermore, another level of complexity derives from their parasitic nature as a "successful" adaptation constitutes a complex balance between parasite development and host survival[160]. This results in a constant co-evolution of parasite and host, as they adapt to each other on two different time scales. On the one hand, there is short-term adaptation of parasites within a single host and, on the other, long-term evolution reflecting natural selection and survival of entire parasite and host populations over thousands of years[161]. This thesis is focused on the short-term adaptation of malaria parasites to the fluctuating conditions of the human host circulation, which can vary both within the same individual as well as between different human hosts. Therefore, in this chapter we will describe some of the principal conditions that can fluctuate within the environment of the human bloodstream and also mention some of the adaptive strategies malaria parasites use to overcome them. Nevertheless, a more detailed account of the molecular mechanisms behind these strategies will be provided in chapter 4 of this introduction.

3.1. Changes in general environmental parameters

Before detailing specific conditions that can fluctuate in the human host circulation, it is important to mention that the majority of challenges malaria parasites encounter in this environment often lead to changes in general parameters such as temperature, pH, osmotic pressure and redox homeostasis. Hence, *Plasmodium* spp. possess various common defence mechanisms to deal with these selective pressures that are frequently shared by other protozoan parasites[162] (**Fig.15**). For instance, one common innate immune response the human host generates against infectious agents is fever[163] and symptomatic malaria infections are characterised by periodic fever episodes triggered by schizont burst at the end of each IDC[164]. Subsequently, malaria parasites can be subjected to ongoing thermal variations within the human host circulation and have consequently developed an adaptive DTR involving the activation of different heat shock proteins[165], which will be further described later on.

Another example of fluctuating conditions resulting in the alteration of a common environmental parameter is that of antimalarial drugs, such as artemisinin or primaquine, producing oxidative and nitrosative stress[166,167]. This type of stress is caused by reactive oxygen species (ROS) such as superoxide anions, hydrogen peroxide or hydroxyl radicals; and reactive nitrogen species (RNS) like nitric oxide or peroxinitrite[162]. ROS and RNS are standard products of aerobic metabolism in eukaryotic cells and malaria parasites are further exposed to both reactive species during their IDC as they are a product of haemoglobin degradation[168]. Consequently, it is hypothesized that *P. falciparum* favours anaerobic metabolism to avoid the production of ROS in the mitochondrial respiratory chain that could add to and interact with other oxidative products[169]. Nevertheless, these pathogens possess a range of low molecular weight antioxidants (e.g., tripeptide glutathione) and antioxidant enzymes (e.g., glutathione- and thioredoxin-dependent proteins, as well as superoxide dismutase) to help them face the oxidative and nitrosative stress that naturally derives from their development within iRBC as well as from additional sources, such as antimalarial treatments, to which they have to adapt to in order to survive[167].

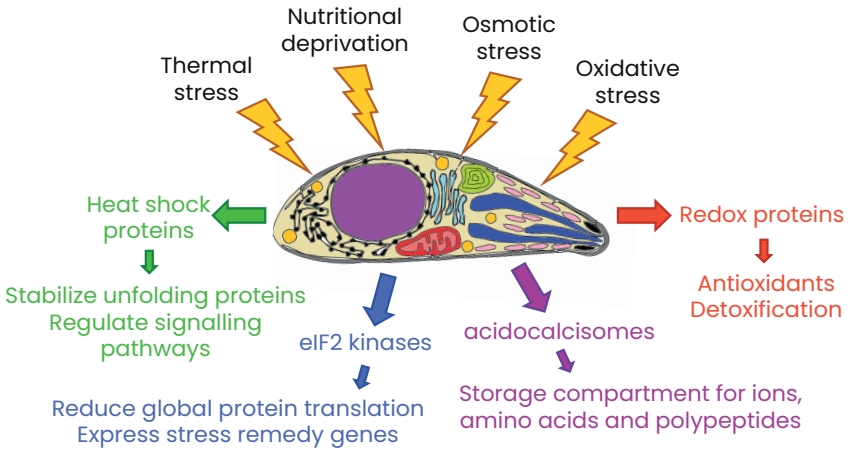


Fig. 15. Common environmental stresses and the associated protozoal responses. These responses include activation of stress remedy products such as antioxidants (red), heat shock proteins (green) or contents of organelles such as acidocalcisomes (purple); and initiation of signalling cascades like those mediated by eIF2 (eukaryotic initiation factor 2) kinases (blue) or MAPKs (mitogen activated protein kinases) (not shown). Both can result in changes in gene expression relevant to stress response and more than one mechanism can be triggered simultaneously to amend a single stress. (Adapted from Vonlaufen *et al.*[162]).

3.2. Host immunity

The human host has a wide arsenal of defence mechanisms with which to fight pathogens. These range from primary barriers to a large variety of molecules and cells capable of recognizing and destroying invasive agents. More specifically, humoral immunity mediated by antibodies targets extracellular parasitic forms, while intracellular stages are more susceptible to cell-mediated immunity[170]. The host's immune response is one of the principal conditions that can fluctuate unpredictably within the human host circulation as not only is the immune system highly variable between individuals, but also it can change over time within the same individual due to factors such as age, general health and exposure to different pathogens, including malaria parasites[171]. For instance, one study found that in malaria-naïve volunteers, the immune reaction to the same dose of *P. falciparum* parasites ranged from a strong inflammatory reaction to practically no reaction at all[172]. As a result, malaria parasites have developed a variety of sophisticated immune evasion strategies that help them to survive the complex and diverse human immune response[170], as we will proceed to describe (**Fig. 16**).

To begin, the fact that most malaria stages within the human host are intracellular already protects these organisms from direct interaction with the immune system[173]. Furthermore, RBCs offer intraerythrocytic stages an extra level of protection as they are practically inert and incapable of presenting antigens due to the lack of nucleus and other organelles[174]. Nevertheless, these parasites are capable of further modifying the cells they infect in order to avoid the immune response. They can do this by modulating iRBC deformability or by expressing different surface antigens on the iRBC surface that interact with different host cell receptors in order to sequester within different tissues such as the bone marrow, placenta or brain. This protects *Plasmodium* spp. from being cleared by the spleen as well as from inflammatory responses, given that some of the aforementioned locations are immune privileged[20,175–177]. For example, an increase in deformability helps *P. vivax* intraerythrocytic stages to navigate their way safely through the spleen[178,179], while a decrease in deformability enables sexual stages of *P. falciparum* to sequester within the bone marrow[37,180]. On the other hand, expression of different PfEMP1 variants on the iRBC surface facilitates the interaction with different host endothelial cells leading to cytoadherence and sequestration of *P. falciparum* trophozoites and schizonts within the microvasculature of different organs[50], as previously described. Similarly, iRBC can bind to other uninfected RBC in

a phenomenon known as rosetting which is also thought to protect these infected cells from the host's immune response[181].

Another strategy malaria parasites use to overcome their host's defence mechanisms is antigenic variation. This consists in these parasites periodically switching their surface antigens, which are exposed to the immune system, so that an effective long-lasting response cannot be mounted against them[181]. In *P. falciparum*, this is mediated by *var* gene expression[182]. Normally, within the parasite population there is a predominantly expressed *var* gene that encodes a PfEMP1 protein. Over time an extensive antibody response is generated against this particular PfEMP1, that can result in the elimination of the majority of the parasite population. However, as some individual parasites stochastically switch to express a different *var* gene, they are not susceptible to this specific antibody response, allowing the parasite population to clonally expand until a new antibody response is generated and so forth. In this manner, chronic infections can be established[47,183].

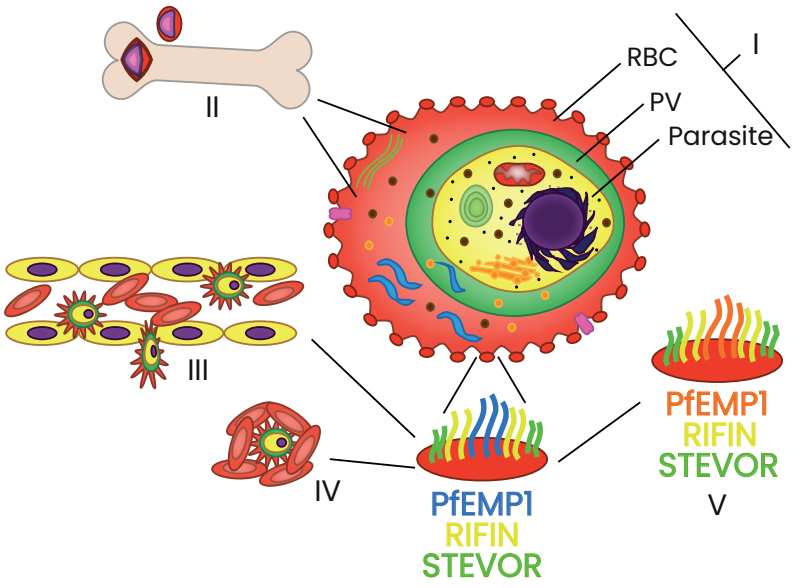


Fig. 16. Principle immune evasion mechanisms of *P. falciparum*: hiding within host cells (I), changing iRBC deformability to facilitate sequestration (II), expressing proteins on the surface of the iRBC membrane that enable cytoadherence and sequestration (III), rosetting (IV) and antigenic variation (V). RBC = red blood cell PV = parasitophorous vacuole (Adapted from Yam *et al.*[181]).

Lastly, other immune evasion mechanisms include "molecular camouflage" and modulation of the host's immune response. The first consists in parasites avoiding recognition by covering themselves with host proteins[160]. For example, merozoites bind host vitronectin protein to their surface[184], although further research is needed to establish the relevance of this with regards to avoiding the host's immune response. The second involves hindering host cell signalling in order to attenuate the immune response and generate a favourable environment for parasite growth[160]. For example, in *P. falciparum* a group of surface proteins known as RIFINs have been found to be able to mimic the natural ligand of the leukocyte immunoglobulin-like receptor subfamily B member 1 and suppress the function of natural killer cells[185].

3.3. Metabolic fluctuations

Another fundamental source of environmental stress that malaria parasites face in the human circulation are metabolic fluctuations. These fluctuations can affect the availability and overall levels of certain nutrients and they can vary unpredictably between and within different human hosts[186]. One source of metabolic fluctuations are dietary differences between hosts. For example, it has been suggested that the natural resistance to malaria of the Fulani ethnic group may be associated with their characteristic diet that is extremely rich in saturated fats derived from dairy products and cooking oils, among other factors[186–189]. Next, different physiological states can also result in substantial changes at the metabolic level. For instance, during pregnancy maternal metabolism undergoes substantial variations as early gestation is characterised by an increase in maternal fat stores and insulin sensitivity, while during late pregnancy there is a decrease in insulin sensitivity that results in an increase in maternal glucose and free fatty acid concentrations[190]. Finally, diseases and other comorbidities can also influence the availability of nutrients within the human host circulation. This is the case with malnutrition and diabetes that are both prevalent in malaria endemic areas[191,192] as well as the malaria infection itself, which in severe cases can result in hypoglycaemia[59,193,194] and in a decrease in the levels of plasma amino acids[194,] among other alterations.

Malaria parasites depend on their hosts to fulfil their metabolic requirements and therefore have evolved to make the most of what is available to them at each point of their life cycle[186,195,196]. Nevertheless, it

is highly probable that these organisms present various mechanisms with which to adapt to fluctuations in the availability of different nutrients, especially as they have been described to possess substantial metabolic plasticity[197]. The current clearest example of this is that malaria parasites have been found to be able to sense and respond to changes in host nutrient intake and adjust their multiplication levels accordingly[198]. This mechanism is mediated by a regulatory protein kinase and will be discussed in further detail later on.

3.4. Erythrocyte traits

Characteristics of human red blood cells can also vary between individuals, including both natural variations in erythrocyte surface antigens, as well as genetic abnormalities that result in structural and functional changes in these cells.

On the one hand, the surface of the erythrocyte's cytoplasmic membrane is covered in a wide range of proteins and receptors that can vary between individuals[199,200]. In this manner, a total of 34 blood groups related to these surface antigens have been described by the International Society of Blood Transfusion[201]. Various of these receptors are involved in direct host-parasite interactions and participate in erythrocyte invasion by *Plasmodium* spp. Consequently, changes in these proteins can affect this key parasite function[199,202]. For instance, individuals that do not express the Duffy antigen receptor are protected against *P. vivax* infection, which is thought to be the reason why this malaria species is rarely found in the African continent where there is a high proportion of Duffy-negative individuals in the population[21,22,202]. Similarly, structural variations in several receptors have also been associated with human resistance to *P. falciparum* infections[199,202,203]. However, this parasite species uses multiple erythrocyte receptors for invasion and expresses proteins with overlapping and redundant receptor-binding functions. As a result, this offers *P. falciparum* the potential to adapt to changes in host surface antigens[199,202,204].

On the other hand, erythrocyte disorders that are mediated by genetic mutations during erythropoiesis constitute one of the most commonly observed genetic traits in humans[202,205]. Some of these abnormalities have been found to be protective against severe malaria and therefore have been selected in populations where this disease is endemic[205]. For

example, haemoglobin disorders such as thalassemia and sickle cell disease that cause severe blood disorders in homozygous individuals, can be protective against malaria in their heterozygous form[206,207], as well as enzymatic deficiencies such as the glucose-6-phosphate dehydrogenase deficiency[207]. Subsequently, the RBCs of individuals with these abnormalities represent an important obstacle for malaria parasite development.

3.5. Antimalarial drugs

Last but not least, a fluctuating condition that malaria parasites face in the host's circulation that directly derives from human intervention are antimalarial drugs[160]. As previously described, there are a wide range of antimalarial drugs with different modes of action that can be used to treat malaria infections[70]. This can be considered a fluctuating condition because not only can *Plasmodium* spp. be subjected to different antimalarial treatments within or between different human hosts, but also because within the same individual, different factors can lead to variations in the drug concentration parasites are exposed to. These factors include the pharmacokinetic properties of each drug that affect its absorption and excretion; incomplete dosing due to prescription errors or problems with patient compliance; and also substandard medicine quality which unfortunately occurs fairly often in developing countries and tropical environments[5,73,208–211].

The exposure of *Plasmodium* parasites to intense therapeutic regimes all around the world, in addition to the aforementioned factors, has led to the inevitable selection of resistant parasites to all existing antimalarial drugs[212]. This currently constitutes one of the principal obstacles for malaria control and eradication[213,214]. Antimalarial resistance is predominantly mediated by genetic mutations[212]. However, malaria parasites also present another mechanism that is mediated at the epigenetic level and constitutes one of the main topics of this thesis. The mechanisms behind genetic and epigenetic drug resistance will be detailed in the following chapter.

4. Molecular mechanisms of adaptation in malaria parasites

In this chapter we will analyse the three main molecular mechanisms of adaptation – DTRs, genetic mutations and epigenetic mechanisms – within the context of malaria biology. Furthermore, we will describe how malaria parasites use each of these mechanisms to confront different environmental stresses, including those mentioned in the previous chapter. Finally, more emphasis will be put on the epigenetic adaptation mechanisms entailing clonally variant gene expression, as this is the main focus of this thesis.

4.1. Directed transcriptional responses in malaria parasites

The role of DTRs in the adaptation of malaria parasites to different environmental stresses has been, until quite recently, a controversial one[215]. Given the complexity of the malaria parasite's life cycle, tight control of gene expression is fundamental and transcription patterns differ greatly between each parasite stage[216–218]. This tightly regulated transcription is controlled by the members of the Apicomplexan AP2 (ApiAP2) family, which is the largest family of transcription factors present in the *Plasmodium falciparum* genome[219,220]. All 27 members of this family have marked stage-specific expression and participate in regulatory cascades that control key events in these parasites' life cycle such as invasion[221] or sexual commitment[35]. Subsequently, for a long time many authors defended that malaria parasites had a "hard-wired" transcriptional programme and were therefore incapable of mounting protective transcriptional responses to external stresses. This argument was further supported by various studies that analysed the effect of different drugs on the malaria parasite transcriptome. These studies reported either no effect or that the small transcriptional variations observed were most likely due to the damaging effect of the drug itself[222–226] or to impairment of the targeted functions[227–229], rather than to a directed protective response. Over the years different publications have presented evidence contrary to this transcriptional rigidity, suggesting that malaria parasites can indeed sense and respond to environmental changes through DTRs, albeit at a lower magnitude compared to other microorganisms such as yeast[101,102]. For instance, amino acid starvation was found to cause malaria parasites to enter into a hibernatory state[230]; hyperoxic conditions both activated an antioxidant defence system and induced metabolic adaptation in

favour of mitochondrial respiration[231]; and exposure to lethal doses of artesunate and DTT[232,233], high parasitemias[234] or high temperatures during fever[235] were all observed to induce a general stress-response involving transcriptional alteration of different genes. However, it remains unclear whether any of these transcriptional alterations are truly adaptative DTRs to the aforementioned environmental stimuli or whether they are mere indicators of parasite death or delayed cycle progression.

Moving on from these inconclusive results, in the last five years three clear examples of DTRs that enable malaria parasites to adapt to changes in their environment have been uncovered. Furthermore, these responses constitute clear examples of adaptive plasticity strategies in these pathogens (**Fig. 17**). First, in 2017, Mancio-Silva *et al.* published that *P. berghei* and possibly *P. falciparum* blood-stage parasites present a nutrient-sensing pathway that enables them to detect and cope with host dietary calorie alterations by adjusting their multiplication rate. This involves a DTR that is mediated by the regulator protein kinase "KIN", that transduces nutrient restriction signals into reduced proliferation rates[198].

Second, in the same year, Brancucci *et al.* described that intraerythrocytic *P. falciparum* stages can sense the depletion of the host serum lipid lysophosphatidylcholine (LysoPC) and react by activating a protective transcriptional response. This response not only affects the phosphatidylcholine (PC) synthesis pathway (Kennedy pathway) and other metabolism-related genes, but also induces *pfap2-g* expression and, consequently, sexual conversion of asexual parasites[236]. A year later, Filarsky *et al.* identified GDV1 (gametocyte development 1), a protein controlled by its own antisense RNA that could possibly act as an intermediate between environmental triggers of sexual conversion and the activation of *pfap2-g*[237]. However, the exact details of this sensing mechanism are still poorly understood. Over the years, several other factors have also been linked to gametocyte induction in different *Plasmodium* spp. including parasite densities, host immune factors and antimalarial drugs[238–243]. However, in some of these cases direct evidence of this gametocyte induction constituting an adaptive response is lacking. Overall, in addition to the metabolic adjustments made in response to LysoPC availability, these findings suggest that malaria parasites can change their investment in sexual commitment in response to environmental cues and this would constitute another excellent example of an adaptive plasticity strategy[137].

Finally, our team members Tintó-Font *et al.* recently found that *P. falciparum* can sense heat stress associated to fever and mount a directed transcriptional response mediated by the transcription factor PfAP2-HS, protecting parasites from cell damage and death. The heat shock induced activation of PfAP2-HS produces the upregulation of a small subset of genes, including the chaperone-encoding genes *hsp70* and *hsp90*, both of which are related to protein folding and stress response[165].

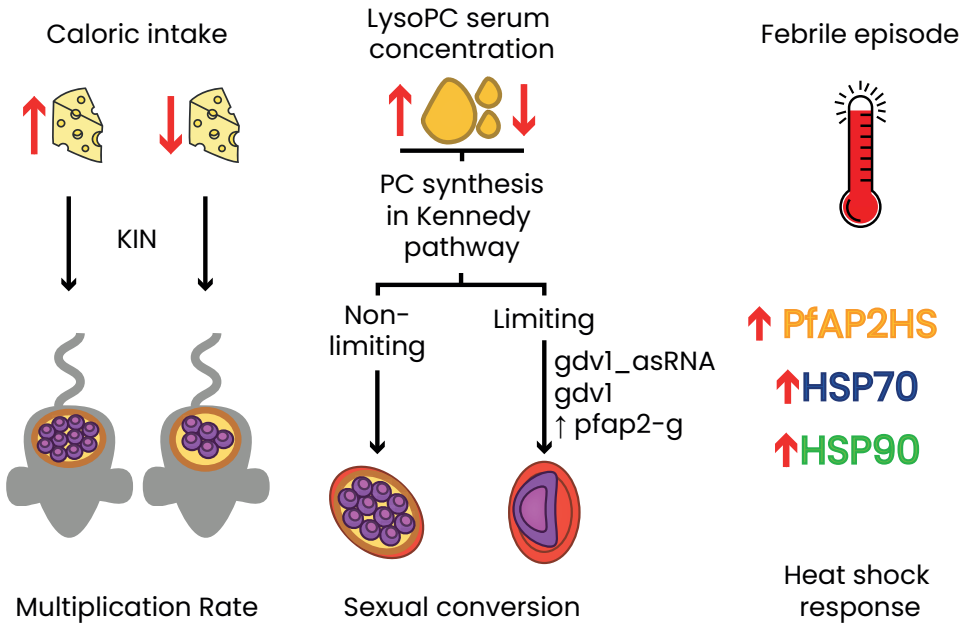


Fig. 17. Three principal examples of DTRs to environmental cues in *Plasmodium* spp. More details in the text.

4.2. Genetic adaptation in malaria parasites

Malaria parasites have relatively small haploid genomes (20-35 Mb) that are organized into 14 chromosomes and two extra-chromosomal elements: a 35 kb circular apicoplast plastid genome and a 6 kb mitochondrial DNA[244]. One characteristic that differentiates these pathogens from other eukaryotic organisms is the AT content of their genome. The *P. falciparum* genome has one of the highest AT contents known to date of around 80% and it reaches up to 90% in non-

coding regions[244]. This results in codon bias, which in addition to the presence of highly repetitive sequences in this organism's genome, can facilitate both the introduction of indels due to polymerase "slippage" during DNA replication[245] as well as unequal cross-over events during meiotic recombination[246]. These stochastic modifications have important biological consequences such as expansion and contraction of amino acid homorepeats, alteration of gene expression due to indels in AT-rich regulatory regions and increased rates of CNVs; all of which generate genomic malleability that facilitates adaptation and evolution[247].

As previously described, genetic mutations play a fundamental adaptive role in malaria parasites as they mediate resistance to nearly all existing antimalarials[212]. Although *Plasmodium* spp. can show genetic variability in the form of SNPs, indels, CNVs and chromosomal translocations[246,248,249], antimalarial drug resistance is more commonly associated with SNPs, followed by CNVs[250]. As in other organisms, CNVs often facilitate the accumulation of SNPs and are eventually substituted by these point mutations due to their lower associated fitness cost[251]. Genetic mutations can mediate various types of drug resistance mechanisms (summarised in **Table 4**). However, the two most common mechanisms are: alteration of drug transport due to mutations in the sequence of transporters and disruption of drug binding as a result of mutations that modify the antimalarial target[252]. Finally, mutations that occur in the extrachromosomal genome of malaria parasites have also been linked to drug resistance. For example, mitochondrial heteroplasmy (diversity in mitochondrial DNA within the same parasite) due to SNPs or CNVs, has been associated with atovaquone resistance[253].

In addition to drug resistance, genetic mutations have other important adaptive roles in malaria parasites. For instance, the adaptation to culture conditions has been associated to the appearance of nonsense mutations in various genes including ApiAP2 genes[255] and with the deletion of large subtelomeric regions[256–258]. Also, genetic diversity in the form of antigenic polymorphism plays an important role in immune evasion[259], with mutations often occurring in regions recognized by T-cells or antibodies[260]. Subsequently, antigenic polymorphism constitutes a huge obstacle for antimalarial vaccine development given that it limits the efficacy of vaccine candidates such as the Apical Membrane Antigen 1 (AMA-1) protein[261].

Drug		Resistance mechanism		
Family	**Specific drug**	**Genetic mutation**	**Affected gene**	**Mechanism**
4-Aminoquinolines	Chloroquine (CQ)	SNP	*pfcrt* *pfmdr1*	Both *pfcrt* and *pfmdr1* are different transporters located in the digestive vacuole membrane. Mutations result in these transporters impeding the accumulation of CQ in the digestive vacuole by incrementing the efflux of uptaken drug.
Antifolates	Sulfadoxine, pyrimethamine, chloroproguanil	SNP	*pfdhr* or *pfdhps*	Alteration of drug binding to the enzyme's (DHFR or DHPS) active site.
		CNV	Amplification of *gtp cyclo-hydrolase 1*	Amplification of a gene that encodes an enzyme upstream in the folate biosynthesis pathway, compensating the fitness cost of these drugs.
Aryl-amino alcohols	Lumefantrine, mefloquine	CNV	Amplification of *pfmdr1*	Large tandem amplification of up to 100kb which includes several genes
Antibiotics	Clindamycin	SNP	23S rRNA	Mutations in apicoplast ribosomal RNA impeding interruption of protein synthesis.
Naphtho-quinones	Atovaquone	SNP	*cyt-b*	Stops drug from interfering with electron transport chain in this cytochrome complex.
Artemisinin-type compounds	Artemisinin, artemether, DHA	SNP	*kelch13*	Lack of drug activation in early rings associated with reduced haemoglobin endocytosis[254].

Table 4: Principal genetic mutations and their associated drug resistance mechanism to the most commonly used antimalarial compounds. Adapted from Cowell *et al.*[252].

Lastly, it is important to note that genetic modifications not only result in alterations in protein products but they can also affect transcript levels and consequently protein levels, especially in the case of CNVs[262]. Moreover, genetic diversity that results in altered transcript levels can also be produced by mutations in promoter regions or in genes that encode regulatory factors[263,264]. For instance, in *P. falciparum* a deletion in the regulatory region of the *pfmrp2* gene, which encodes an ABC transporter, results in a delay in the expression of this gene, constituting another type of drug resistance mechanism[265].

4.3. Epigenetic adaptation in malaria parasites

Epigenetic mechanisms have been found to be fundamental in the regulation of gene expression in malaria parasites and are involved in essential processes such as life cycle progression, host-parasite interactions and adaptation to the host environment[217]. In *Plasmodium* spp. these epigenetic mechanisms predominantly revolve around the structure and organization of chromatin. In the following sections we will describe the characteristics of the epigenome of malaria parasites and we will discuss the main type of genes whose expression is susceptible to these epigenetic mechanisms and which are known as clonally variant genes (CVGs).

4.3.1. Epigenome of malaria parasites

The *P. falciparum* genome can be found predominantly in the form of transcriptionally active euchromatin and only presents highly condensed heterochromatin at subtelomeric regions and a few internal loci[266–268]. Various elements of the epigenome are behind this distribution of chromatin, some of which differ significantly from other eukaryotes. A visual summary of the main differential characteristics of the malaria parasite epigenome can be found at the end of this section in **Fig. 18**.

To begin, malaria parasites present a conserved nucleosome that is comprised of the four canonical histones (H2A, H2B, H3 and H4), as well as four histone variants (H2A.Z, H2Bv, H3.3 and CenH3)[269], that can also be found in other eukaryotic organisms[270]. These histone variants are associated with different types of chromatin and chromosomal locations. For example, H2A.Z replaces H2A in the promoter regions of active and inactive euchromatic genes and also in the promoter regions of active

heterochromatic genes. Similarly, CenH3 replaces H3 at the centromeres[271]. Moreover, histone variants are also important because some, like H2A.Z and H2Bv, can carry more posttranslational modifications than canonical histones, while others, such as CenH3, carry no modifications at all[272]. Finally, it has been suggested that malaria parasites lack the linker histone H1[269] and this is thought to explain why their chromatin is less tightly packed and may lack higher structural order[273] and why the malaria epigenome is mainly euchromatic[268,274].

Secondly, posttranslational modification of histone tails are one of the key epigenetic mechanisms in malaria parasites[271] as they can affect gene expression by determining chromatin compaction and accessibility, as well as by recruiting or blocking factors that can activate or repress transcription[269]. In *P. falciparum*, both constitutive and facultative heterochromatin are marked by H3K9me3 and bound by heterochromatin protein 1 (HP1)[267,268]. In euchromatin the pattern of histone modifications is much more complex. Mainly acetylation is associated with transcriptional activation in both promoter regions (H4K8ac and H3K9ac) and coding sequences (H4K16ac, H3K56ac, H3K9ac). However, methylation can also be associated to activation (H3K4me3), though for example H3K36me2 serves as a global mark of gene repression[271].

Thirdly, another element that contributes to chromatin structure and control of gene expression is the differences in nucleosome dynamics. *Plasmodium* spp. nucleosomes are identified as being less stable and, therefore, possibly favour the increased accessibility of these parasites' genome[275]. The high AT content is suggested to be partly responsible for this, as it reduces DNA flexibility and hinders it being wrapped around the nucleosome core. Nevertheless, histone variants H2A.Z and H2Bv partially solve this issue by being able to bind weakly but more effectively to AT-rich DNA[276,277].

Fourthly, chromatin organisation within the nucleus also contributes to regulating gene expression. Malaria parasites' chromosomes have a folded distribution, with both chromosomal arms folded over each other in parallel orientation so that centromeres and telomeres cluster in opposite regions of the nucleus[278,279]. Furthermore, some other subtelomeric and internal loci colocalize at the nuclear periphery and create additional chromosomal loops[268,279,280]. As a result, two compartments can be distinguished within the parasite nucleus: the repressed heterochromatic regions at the periphery and the active euchromatic regions in the interior. Moreover, the nuclear

structure also varies throughout the IDC. For example, during the most transcriptionally-active trophozoite stage, nucleus size, volume and pore number all increase, facilitating gene expression and the transport of mRNAs into the cytoplasm[279,281].

Last but not least, ncRNAs also participate in malaria parasite epigenetic regulation. *Plasmodium* spp. possess numerous ncRNAs that can be found in centromeric, subtelomeric and telomeric chromatin and are implicated in regulating housekeeping genes and virulence factors[282–286]. Also, in centromeric regions, ncRNAs contribute to the correct organizational assembly of chromatin and are bound by specific nuclear proteins[282].

Finally, despite being one of the main epigenetic regulatory mechanisms in other organisms, DNA methylation has only recently been identified in *P. falciparum* and its role in transcriptional regulation remains to be determined[287–289].

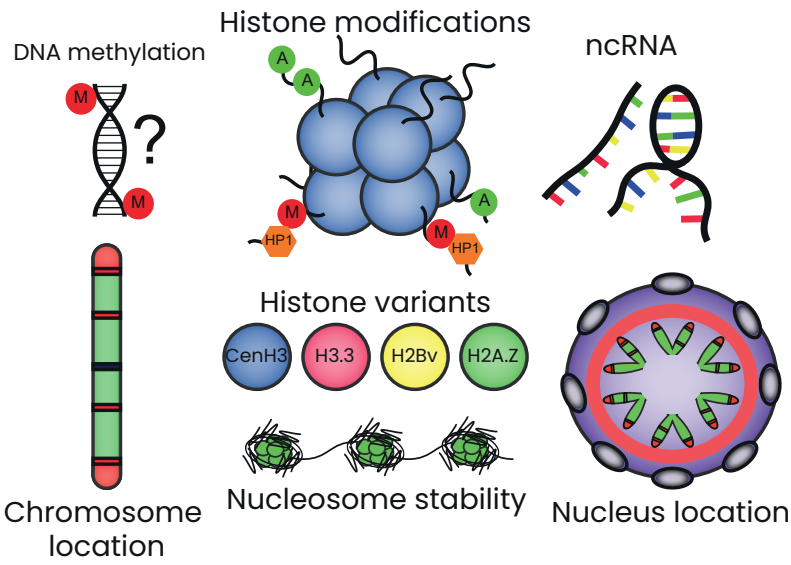


Fig. 18. Key elements of the malaria parasite epigenome. Chromatin structure determined by histone modifications (M=methylation, A=acetylation), presence of histone variants (H2A.Z, CenH3, H2Bv and H3.3), recruitment of chromatin-associated proteins (e.g., HP1), chromosomal location, nucleosome stability, organization of chromatin within the nucleus and the involvement of ncRNAs all play fundamental roles in the epigenetic regulation of the *P. falciparum* genome[279,290]. Green indicates transcriptional activation and red indicates repression.

4.3.2. Clonally variant genes

Malaria parasites possess a unique type of genes whose expression patterns have a distinct epigenetic regulation. The genes are known as clonally variant genes (CVGs) and they are characterised for being able to undergo stochastic switches between active and silent states that can generate transcriptional variability within the parasite population which can also result in phenotypic diversity[291,292]. In this manner, within a genetically homogenous population of parasites that are at the same stage of development, different individuals can have each of these genes in an active or silent transcriptional state (i.e., the same CVGs can have different expression patterns in otherwise identical parasites)[293]. Given that all *Plasmodium* spp. possess CVGs, this suggests that they are a universal characteristic of malaria parasites[291].

The transcriptional state of CVGs is regulated at the epigenetic level and is mainly determined by changes in chromatin structure[291]. Their regulatory regions are found in bi-stable chromatin domains that can shift between transcriptionally-active euchromatin and silent facultative heterochromatin[294]. Each state is characterized by a particular mark in histone H3 lysine 9, that is placed and maintained by specific enzymes and proteins. The active euchromatic state requires H3K9 histone demethylases to remove methyl groups, if present, and acetyltransferases which produce the acetylation of H3K9 (H3K9ac). Meanwhile, the silent heterochromatic state is achieved through the action of specific histone deacetylases to remove acetyl groups, if present, and methyltransferases (HMTs) which trimethylate H3K9 (H3K9me3)[268,291,295–302]. In addition, heterochromatin protein 1 (HP1) recognizes the H3K9me3 modification and plays a key role in the chromatin condensation process[267] which is fundamental in maintaining the silent state of CVGs[295]. These histone modifications are maintained throughout the full asexual IDC, allowing for the stable transmission of these transcriptional patterns over several generations and constituting a form of epigenetic memory[35,266,300–303].

The location of CVGs has also been argued to play an important role in their regulation. They are primarily found in sub-telomeric regions and various internal loci of different chromosomes that cluster at the nuclear periphery[268,279,280], where they may be influenced by the general heterochromatic environment of telomeres[279]. In some cases, clonally variant expression can also entail relocation of these genes between transcriptionally active and silent environments within

the nucleus[268,280,304–306]. This relocation, alongside other unknown and known mechanisms (involving long ncRNAs among other elements), are associated with an additional property that is "mutually-exclusive expression" which is only known to affect *var*[307–314] and *clag3* genes[315]. This type of expression means that only one *var* and *clag3* gene can be expressed in each individual parasite and its activation requires the simultaneous silencing of all the other *var* and *clag3* genes. Another consequence of being located primarily in subtelomeric regions, which contain many repetitive sequences, is that CVGs frequently undergo genetic events such as recombination[316,317], mutations[247], deletions, translocations and segmental duplications[246,318,319] during both mitotic growth[320] in the asexual life-cycle and during the fusion of gametes in the mosquito host[183]. These events result in a high genetic variability with the potential to generate of millions of different antigenic structures[183].

CVG expression is suggested to constitute the base of a bet-hedging adaptive strategy in malaria parasites[291,321,322] (**Fig. 19**). This is because the stochastic switches between active and silent transcriptional states result in transcriptional and phenotypical variability within the parasite population that precedes unpredictable fluctuations in the environment. The exact mechanism behind this stochastic switching is unknown although possibilities such as fortuitous alterations in the expression of epigenetic regulators or errors during the transmission of epigenetic memory during mitosis, have been suggested[293].

The huge adaptive potential that CVGs offers malaria parasites has so far only been demonstrated for *var* genes with regards to antigenic variation and cytoadherence[183,323,324] and for *clag3* genes in relation to the permeability of the iRBC membrane to different toxic compounds[325–327]. However, due to variability in both sequence and expression patterns, it is predicted that other CVGs may also play important adaptive roles and mediate other bet-hedging strategies[291] (**Fig. 20**). Further on we will describe the main CVG families present in the *P. falciparum* genome, their main characteristics and their roles (known or predicted) in parasite adaptation.

Finally, the expression of CVGs is influenced by the route of transmission[328]. Experiments with *P. chabaudi* revealed that parasite populations transmitted via mosquito bites express a wider repertoire of CVGs and produce a less severe infection than those transmitted by serial blood passage that express a subset of these genes associated with

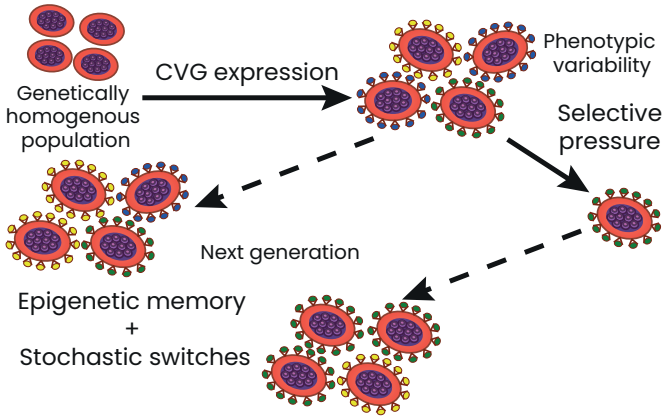


Fig. 19. CVG expression produces transcriptional and phenotypic diversity (represented by different colour surface antigens) in genetically homogenous populations. This pre-existing variability enables the parasite population to survive unpredictable selective pressures constituting an example of a bet-hedging adaptive strategy. The expression patterns of CVGs can be passed on to the following generations thanks to epigenetic memory and new variability can be introduced due to the stochastic switches between active and silent transcriptional states that these genes undergo.

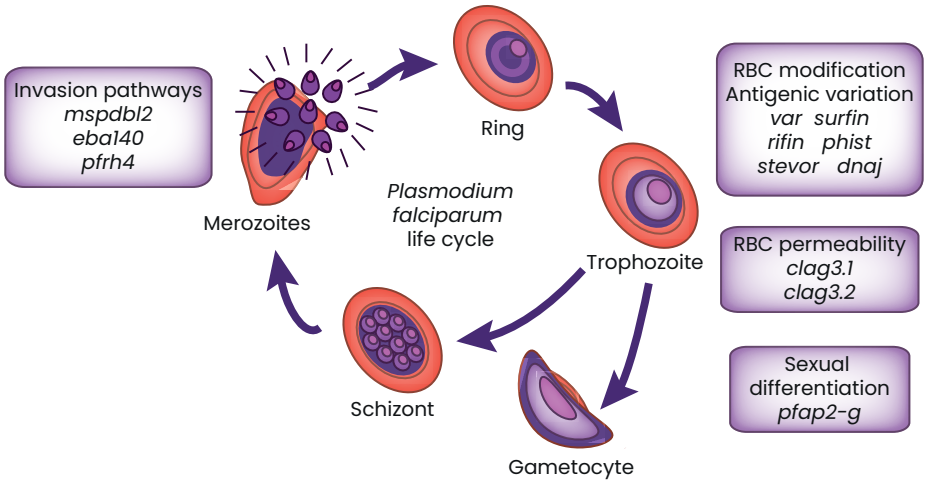


Fig. 20. CVG expression throughout the asexual IDC of *P. falciparum*. During the ring/trophozoite stage CVGs associated with antigenic variation and RBC modification are expressed, those related to RBC permeability are expressed in early schizonts and those CVGs that determine different invasion pathways are expressed in bursting schizonts / merozoites. Finally, in a small proportion of cells *pfap2-g* is stochastically activated leading to the sexual differentiation of asexual parasites and the development of gametocytes. (Adapted from Cortés *et al.*[291]).

higher virulence[329,330]. Similarly, analysis of samples from controlled human malaria infection (CHMI) trials with *P. falciparum* have unveiled that after transmission via mosquitos, parasite populations express a larger and more diverse range of *var* genes at the onset of a blood infection in a new human host[331–334]. Similarly, *clag3* expression patterns have also been found to be affected by transmission via mosquitos[335]. All in all, this data supports the hypothesis that CVGs undergo an epigenetic reset during their passage through transmission stages[328] (**Fig. 21**). In this manner, malaria parasites lose the *var* and *clag3* expression patterns selected in the previous vertebrate host and then establish new expression patterns for these genes that enable them to face the unpredictable conditions of a new vertebrate host with an increased transcriptional and phenotypical diversity. This can also be considered an example of a bet-hedging adaptive strategy. However, it remains to be determined whether in *P. falciparum* other CVGs undergo this epigenetic reset as well as *var* and *clag3* genes, which are also the only CVGs known to present mutually-exclusive expression[307–315], as previously mentioned.

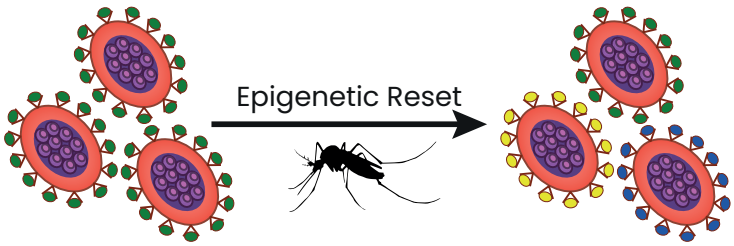


Fig. 21. Epigenetic reset of CVGs upon passing through transmission stages. More details in the text.

4.3.3. Potential adaptive roles of the main CVG families in *Plasmodium falciparum*

4.3.3.1. Antigenic variation and cytoadherence: *var* genes

The *var* gene family is comprised of approximately 60 members[244] and encodes surface antigens that are responsible for some of the most pathogenic traits of *P. falciparum*[336]. The mutually-exclusive expression of *var* genes results in a single type of PfEMP1 being expressed on the iRBC surface of each individual parasite[307–314]. This, combined with the stochastic

switching from one PfEMP1 to another in different individual parasites[314,337], constitutes the basis of antigenic variation in *P. falciparum*[183]. As previously mentioned, this adaptive strategy enables these pathogens to overcome the selective pressure of the human immune response and constitutes a clear example of bet hedging[321].

The epigenetic mechanisms behind *var* gene switching are the most extensively studied in *P. falciparum*[279] (**Fig. 22**). Silenced *var* genes are arranged in several clusters at the nuclear periphery[280,305] and covered in highly condensed heterochromatin characterised by the presence of H3K9me3, binding protein HP1 and the *var*-specific H3K36me3 (also present in active *var* genes)[267,268,274,300,309,310,338,339]. When a silent gene is switched to an active state, it dissociates from the silenced clusters and re-localizes to a different active domain, replacing the previously active gene that returns to a silenced region of the nucleus[279,280,305]. Active *var* promoters are enriched in H3K9ac and H3K4me3 and the histone variants PfH2A.Z/H2BZ dimers[268,299,302]. Further regulation of *var* gene expression is mediated by lncRNAs. Firstly, the intron that separates the two exons of these genes contains a conserved bidirectional promoter controlling two lncRNAs (a sense and antisense) that have distinct roles in controlling the transcription of different *var* genes[310,313,340,341]. Next, the members of a GC-rich ncRNA family play a key part in mutually-exclusive *var* activation and possibly in the transcription of other CVGs[342,343]. Lastly, a large lncRNA family transcribed from telomere-associated repetitive elements (TAREs) has also been suggested to participate in the regulation of *var* gene expression[279,286,344]. Overall, switching between *var* genes appears to be a non-random process and antigenic variation is a complex balance between parasite intrinsic on/off switching rates and immune-mediated selection[182].

Finally, as previously described, *var* genes are also involved in cytoadhesion and sequestration in various host tissues as different PfEMP1 variants can bind to different host receptors and enable parasites to avoid clearance by the spleen[20,50]. As with antigenic variation, the different cytoadherence phenotypes of *var* genes also constitute the basis of a bet-hedging adaptive strategy in which phenotypic diversity precedes environmental pressures. For example, *var2-csa* preferentially binds to chondroitin sulphate A (CSA) receptors of the placenta[55], therefore in pregnant hosts those parasites that stochastically express this *var* gene are selected as it confers them a fitness advantage.

All in all, this gene family plays a fundamental role in the survival of

P. falciparum within the human host and as can be expected, other *Plasmodium* spp. possess similar gene families with analogous roles. For example, in *P. knowlesi* SICA (schizont-infected cell agglutination) genes are also involved in antigenic variation, immune evasion and the establishment of chronic infections within their vertebrate hosts[345–347].

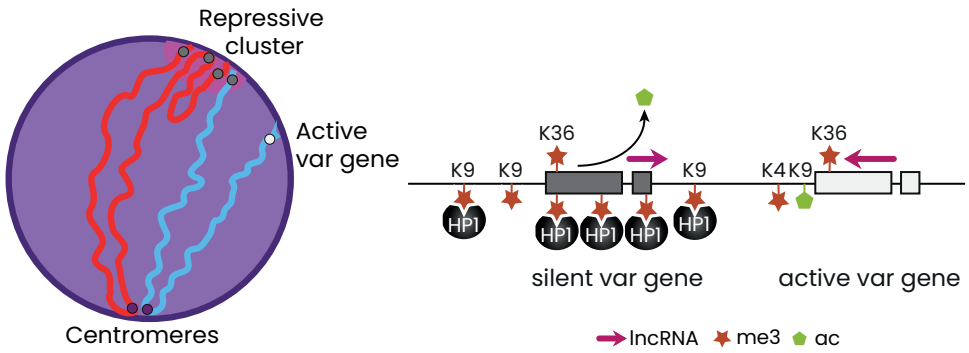


Fig. 22. Epigenetic regulation of *var* genes. Silent *var* genes are clustered in the nuclear periphery and marked with repressive H3K9me3 marks that recruit HP1 resulting in the formation of heterochromatin. The single active *var* gene is separated from the others and is marked by H3K4me3 and H3K9ac in an euchromatic environment. H3K36me3 is a mark of both active and silent *var* genes. Further regulation is achieved from lncRNAs located in bidirectional promoters of *var* introns. Transcription of the sense lncRNA is associated with silencing whilst the antisense lncRNA is required for the expression of the single active *var* gene (Adapted from Batugedara *et al.*[279]).

4.3.3.2. Other CVGs involved in antigenic variation and iRBC modification: *rif, stevor, phist* and *pfmc-2tm* genes

In addition to *var* genes, there are four other important CVG families involved in iRBC modification and antigenic variation: *rif*, *stevor*, *phist* and *pfmc-2tm*.

The repetitive interspersed (*rif*) family[348] is the largest CVG family with approximately 150 genes[244] that encode small protein adhesins that are expressed throughout the parasite's life cycle[349]. RIFIN proteins can be found in various locations such as the iRBC surface[350,351], the apex of merozoites as well as in gametocytes[349,352]. This variability in time of expression and localization suggests that this family possesses many different functions. At present, they have mainly been associated with an adaptive role in

immune evasion. For example, RIFIN proteins have been found to participate in rosetting[351] and interference with the response of host immune cells[185]. Moreover, these proteins possess a hypervariable region that in combination with the large number of *rif* genes within the parasite genome strongly hints at the involvement of these proteins in antigenic variation[299].

Next, the subtelomeric variable open reading frame (*stevor*) family is composed of 28 members in the reference genome[244], that are located subtelomerically on all 14 chromosomes[348]. Similar to *rif* genes, they are expressed at various stages of the life cycle including merozoites[353] and gametocytes[354], though they reach maximum expression at the trophozoite stage[355]. STEVOR proteins first localize to the iRBC cytosol and from there they are trafficked to the iRBC membrane via Maurer's clefts (parasite derived structures)[356]. These proteins participate in numerous functions such as erythrocyte binding, merozoite invasion and rosetting[357]. Finally, various STEVOR have been implicated in variations in iRBC membrane deformability in both asexual and sexual parasites[180]. Specifically, in gametocytes, membrane deformability is regulated by the phosphorylation of the cytoplasmic domain of these proteins and enables their sequestration and release from the bone marrow[358].

The *Plasmodium* helical interspersed subtelomeric (*phist*) family is conformed of 89 genes that encode exported proteins[359,360]. In total, only 14 *phist* genes have been found to show clonally variant expression[322], although there are likely to be many more given the total number of *phist* genes that carry heterochromatin[266]. The majority of *phist* are expressed at the beginning of the IDC[361] but the proteins they encode present an 11 h lag suggesting that they play important roles in host-cell remodelling later on in this cycle[362]. PHISTs can be found at the iRBC membrane contributing towards the formation of the knobs that expose PfEMP1, the Maurer's Clefts, the parasitophorous vacuole membrane and even in extracellular vesicles[363]. This diversity indicates that these proteins participate in a number of different roles, although their functional and physical characterization is lacking. Nevertheless, they have been implicated so far in roles such as presentation of PfEMP1 on the iRBC surface, participation in the process of gametocytogenesis[364–366], changes in cell rigidity[367], cerebral and pregnancy associated malaria[368–371] and formation of extracellular vesicles facilitating cell-to-cell communication amongst iRBC[372].

Finally, *pfmc-2tm* genes constitute a family of approximately 13 members[244]. They encode proteins with two transmembrane domains that

flank a highly polymorphic stretch of amino acids thought to be associated with antigenic variation[373]. Though a biological role for this family has yet to be elucidated[374], these proteins have been found to be associated with the membranes of the Maurer's clefts (membranous structures formed in the iRBC cytoplasm directly post-invasion[375]) and to interact with the iRBC membrane[376].

4.3.3.3. iRBC permeability: *clag3* genes

The second set of CVGs with the clearest adaptive role in *P. falciparum* are the two *clag3* genes: *clag3.1* and *clag3.2*. The proteins encoded by these genes form part of the RhopH complex[377] that is conformed of three proteins: RhopH1, RhopH2 and RhopH3[378]. In comparison to RhopH2 and RhopH3 that are encoded by single genes, RhopH1 is formed by one of the five members of the *clag* gene family: *clag2*, *clag3.1*, *clag3.2*, *clag8* and *clag9*[377,379]. In addition to the two *clag3* genes, *clag2* is the only other member of this family to show clonally-variant expression regulated at the epigenetic level[297,300,315]. The CLAG3 paralogs are almost identical and show approximately 95% sequence identity, differing mainly in a 120 nucleotide hypervariable region they both possess at their C' terminus[380]. Furthermore, they are both located in the subtelomeric region of chromosome 3, separated only by 10 kb which includes a *var* pseudogene and a ncRNA[381,382]. As a result of their sequence similarity and proximity, these genes often undergo recombination events that generate much diversity among parasite strains[380,381]. Finally, *clag3* genes present mutually-exclusive expression[315], even though exceptions have been described[383].

Over the years, various roles have been proposed for *clag* genes ranging from cytoadherence[256,384] (hence their name "cytoadherence linked asexual gene") to invasion[377,385–388]. However, it is now commonly accepted that CLAG3 proteins, in association with the other components of the RhopH complex (RhopH2 and RhopH3), participate in the formation of the plasmodial surface anion channel or PSAC[389–394]. In order for this to happen, the different members of the RhopH complex are simultaneously expressed at the schizont stage and packaged into the rhoptry organelles that are involved in merozoite invasion of new erythrocytes[393,395]. Post-invasion, the complex is deposited into the parasitophorous vacuole from where it is exported 20 h later via an unknown mechanism to the cytosol of the iRBC. There it is predicted for it to undergo large conformational changes before it is finally inserted into the iRBC membrane to form the PSAC[393,395] (**Fig. 23**).

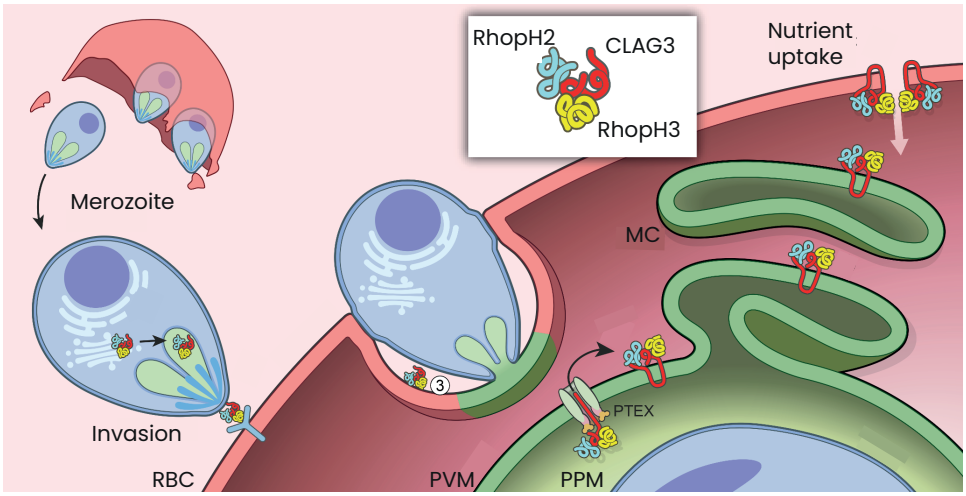


Fig. 23: Trafficking model of the RhopH complex throughout the *P. falciparum* life cycle. RBC = red blood cell, PVM = parasitophorous vacuole membrane, PPM = parasite plasma membrane, MC = Maurer's cleft. More detail in the text. (Adapted from Ito *et al.*[393]).

The PSAC is a broad selectivity channel that mediates the transport of a wide range of nutrients, anions and organic cations across the iRBC membrane[390,396]. At present, the PSAC is the only identified parasite-induced channel to form the new permeation pathways (NPPs) that are responsible for the increase in membrane permeability experimented by iRBCs[396,397]. However, there is still debate as to whether other parasite- or host-encoded channels also form part of these pathways[397–400]. What is increasingly clear thanks to ongoing research, is that CLAG3 proteins play a fundamental part in the formation of a functional PSAC and consequently in malaria parasite biology. On the one hand, these genes have been found to play an important role in nutrient acquisition[389]. Furthermore, this role is thought to be especially important in vivo, as a CLAG3-knockout parasite line was observed to be unable to grow in medium with physiological concentrations of various essential nutrients[401]. On the other hand, the selection of different *clag3* expression patterns has been associated with changes in the permeability of the iRBC membrane towards the toxic compound blasticidin S (BSD)[326,327]. While exposure to low concentrations of BSD selects for parasites expressing predominantly *clag3.1*, exposure to even higher concentrations selects for parasites with both *clag3* genes silenced constituting an exception to their characteristic mutually-exclusive expression[326] (**Fig. 24**). This constitutes a novel antimalarial drug resistance mechanism mediated at the epigenetic level as it is reversible

once drug pressure is removed due to stochastic switching at the *clag3* locus that enables the reappearance of *clag3* expressing parasites in the population[326,327]. Moreover, it also constitutes the other clear example of CVG-mediated bet-hedging in malaria parasites, as the variability in *clag3* expression patterns precedes the exposure to the drug selective pressure[291,293,321]. All in all, given the role CLAG3 proteins have in adjusting iRBC membrane permeability via the PSAC, they possess a huge adaptive potential that could help malaria parasites face environmental challenges involving nutrient availability, antimalarial drugs and other solutes present in the human host circulation.

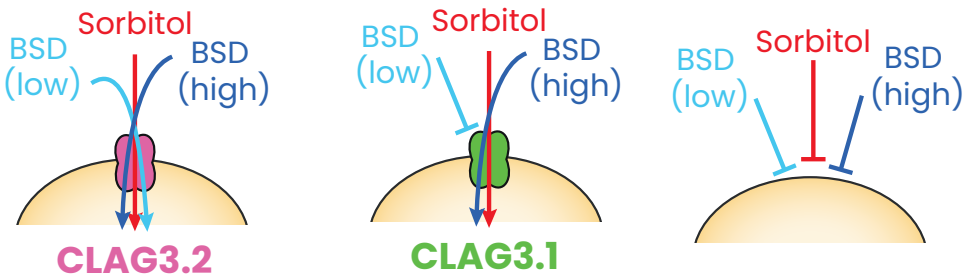


Fig. 24. *clag3* expression patterns associated with different iRBC permeability profiles. More details in the text. (Adapted from Mira-Martínez *et al.*[326]).

4.3.3.4. Sexual differentiation: *pfap2-g*

The ApiAP2 transcription factor encoded by *pfap2-g* is considered the master regulator of sexual conversion[35] and this gene shows both clonally variant expression[322] and epigenetic marks of silencing[267,268]. In addition to being stochastically activated in a small proportion of parasites[35], *pfap2-g* can also be activated in response to different environmental cues[236,238–243], as previously mentioned. PfAP2-G is conserved in all *Plasmodium* spp. highlighting its essential role in the life cycle of malaria parasites[402].

Depending on when the expression levels of this transcription factor trigger gametocytogenesis, two routes of sexual conversion have been described[403]. Same cycle conversion (SCC) occurs when *pfap2-g* is activated in very early rings that convert directly into gametocytes, while next cycle conversion (NCC) happens when *pfap2-g* is activated at a later stage so that parasites have to complete an additional asexual cycle as *pfap2-g*-expressing committed forms before converting into gametocytes

in the next cycle[403] (**Fig. 25**). These two different routes could potentially form part of an adaptive plasticity strategy because, depending on the circumstances, undergoing an additional asexual cycle or not before completing sexual conversion could be more or less beneficial, especially in response to different environmental cues.

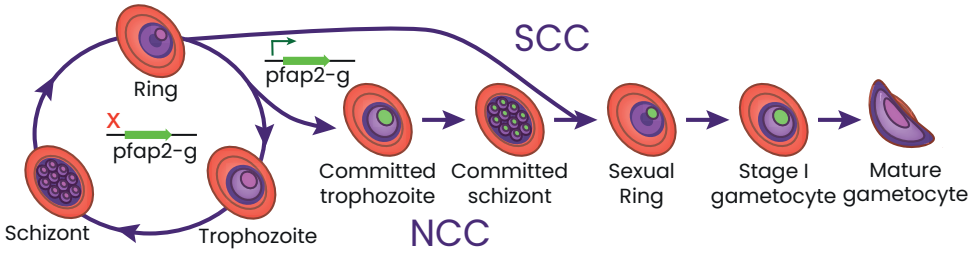


Fig. 25. Scheme of same cycle (SCC) and next cycle (NCC) conversion. Green circles indicate nuclear *pfap2-g* expression. More detail in the text. (Adapted from Bancells *et al.*[403]).

4.3.3.5. Erythrocyte invasion: *eba* and *pfrh* genes

Both *eba* (erythrocyte-binding antigen) and *pfrh* (*Plasmodium falciparum* reticulocyte-binding homologue) are small gene families that encode adhesins that are released from the apical organelles of merozoites in order to interact with different erythrocyte receptors and mediate alternative invasion pathways[404]. The non-essentiality of the majority of these genes[405–408] and their clonally-variant nature[300,301,315,408] strongly suggests that variations in their expression patterns may result in phenotypic differences that enable parasites to use different invasion routes, constituting a potential bet-hedging adaptive strategy[406,408,409]. However, this has yet to be demonstrated. That said, these proteins may play an alternative role in immune evasion, where the small number of genes in these families is supplemented by genetic polymorphism among isolates[406,409].

4.3.3.6. Other CVGs

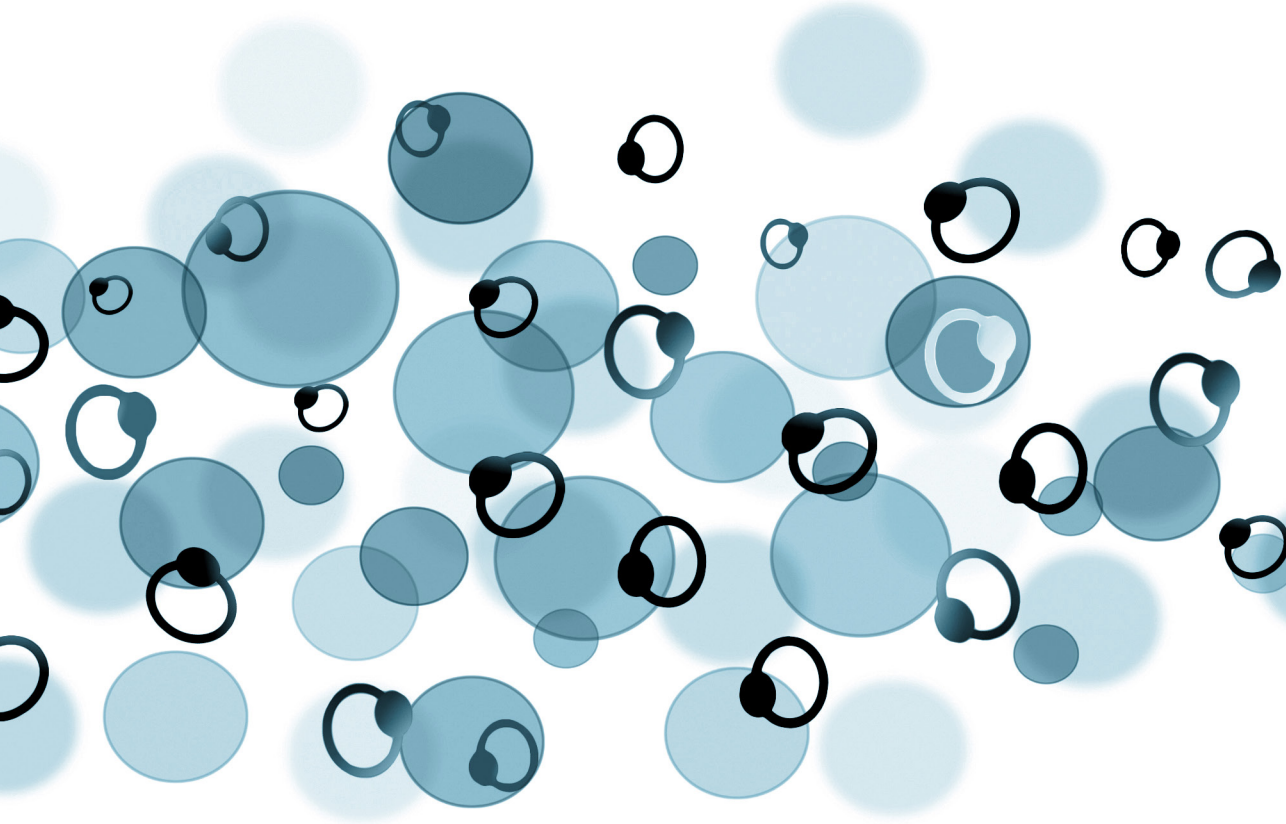
The *fikk* family, that has 21 members in *P. falciparum*[244,410], encodes exported kinases that are responsible for much of the infection-induced phosphorylation in the iRBC[411]. These kinases are unique to apicomplexan

parasites, what makes them attractive drug targets[412], and each one has an independent role with a unique phosphorylation fingerprint[411]. Most of this family is exported into the iRBC and mediates dramatic structural and functional changes related to pathogenity[413]. For instance, one particular FIKK has been found to be associated with rigidification of the erythrocyte cytoskeleton and with trafficking of PfEMP1 to the iRBC surface[411].

Next, different members of the acyl-CoA synthase (*acs*), acyl-coA binding protein (*acbp*) and *lysophospholipase* families show evidence of clonally variant expression[322]. The first two families and members of the third are involved in lipid metabolism[414–416] and it is possible that their differential expression may aid the parasite's adaptation to fluctuations in the availability of this type of nutrients[299].

Other small CVG families include the 10-member surface-associated interspersed (*surf*) multigene family that encode high molecular weight antigens known as SURFIN proteins[417]. These antigens are expressed at different stages within the intraerythrocytic cycle[418] and may play roles in immune evasion[419]. Also, the 14-member six-cysteine (*6-cys*) family encode stage-specific proteins that are often situated at the parasite surface or at the interface with the host[420] and participate in important functions such as fertilization in sexual stages, formation of the parasitophorous vacuole in liver stages and evasion against the immune response mounted in the mosquito midgut[421]. Finally, there are various other CVG families in *P. falciparum* that remain largely uncharacterized such as *etramp*, *gbp*, *hyp*, and *dnajIII* genes, among others, which could also potentially possess adaptive roles[322].

HYPOTHESIS & OBJECTIVES



It is well-established that CVGs play an important role in the adaptation of malaria parasites to their environment. However, despite CVGs constituting approximately 10% of the *Plasmodium falciparum* genome, only *var* and *clag3* genes have been demonstrated to mediate bet-hedging adaptive strategies. We hypothesise that:

- CVGs can mediate adaptation of *P. falciparum* to other fluctuating conditions of the human host circulation via bet-hedging adaptive strategies, in addition to those described for *var* and *clag3* genes.

- Other CVGs undergo an epigenetic reset upon passage through transmission stages, besides the mutually-exclusively expressed *var* and *clag3* genes.

- The *clag3*-mediated epigenetic drug-resistance mechanism is effective against other antimalarial compounds that require CLAG3 for their uptake via the PSAC, in addition to BSD.

- *clag3* expression is governed by parameters such as fitness advantages and stochastic switching rates that can be determined experimentally.

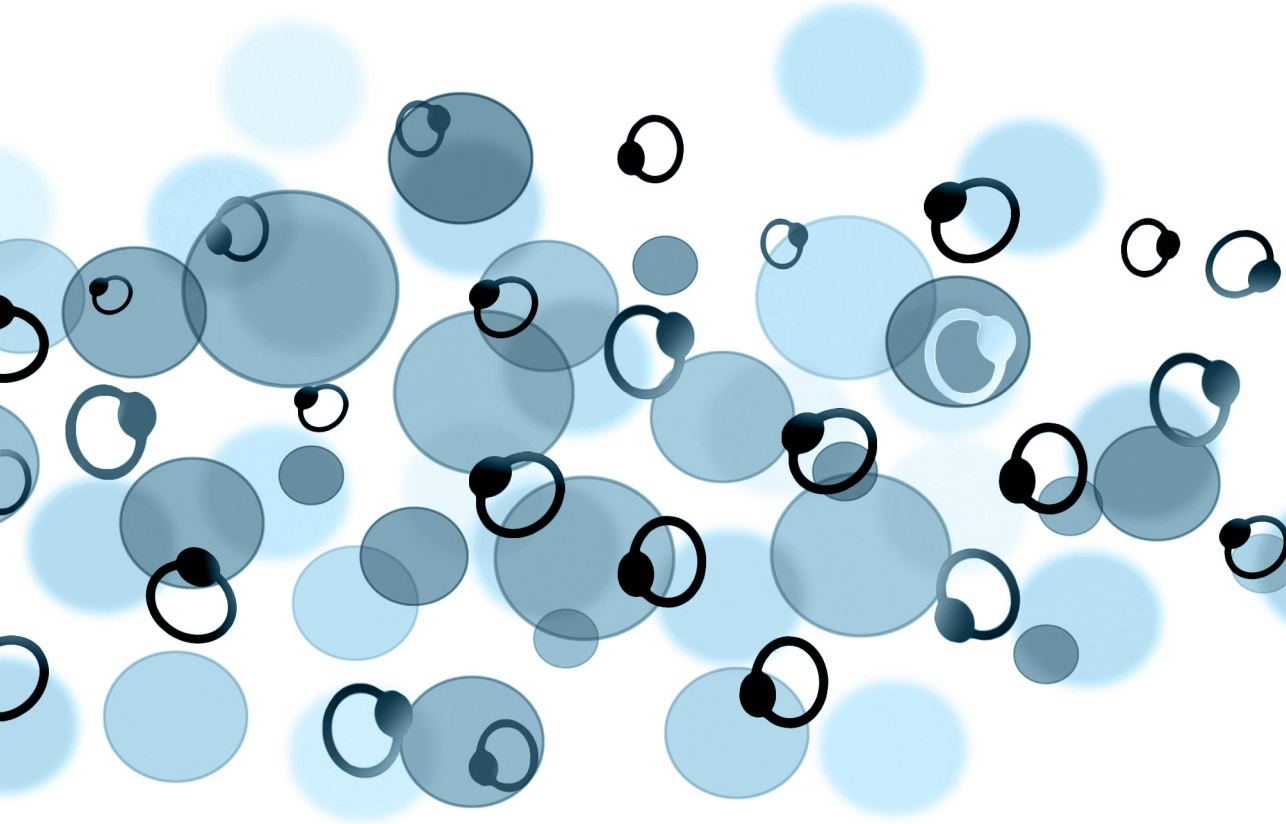
Consequently, the main goal of this thesis is to further characterise how CVGs mediate bet-hedging adaptive strategies that enable *P. falciparum* to adapt to the fluctuating conditions of the human host circulation. The specific objectives with which we aim to address our main goal are as follows:

- Further analyse the adaptive role of *clag3* genes and characterise their expression dynamics.

- Determine how CVG expression patterns are affected by the passage of *P. falciparum* through transmission stages and the exposure to the conditions of a new human host, using samples from a controlled human malaria infection trial.

- Identify new fluctuating conditions in the human host circulation to which malaria parasites adapt using bet-hedging strategies.

RESULTS



Article 1: Identification of antimalarial compounds that require CLAG3 for their uptake by *Plasmodium falciparum*-infected erythrocytes.

Mira-Martínez S, Pickford AK, Rovira-Graells N, Guetens P, Tintó-Font E, Cortés A & Rosanas-Urgell A. Identification of antimalarial compounds that require CLAG3 for their uptake by *Plasmodium falciparum*-infected erythrocytes. *Antimicrobial Agents and Chemotherapy* 63: e00052-19. (2019). https://doi.org/10.1128/AAC.00052-19.

AMERICAN SOCIETY FOR MICROBIOLOGY | Antimicrobial Agents and Chemotherapy®

MECHANISMS OF RESISTANCE

Identification of Antimalarial Compounds That Require CLAG3 for Their Uptake by *Plasmodium falciparum*-Infected Erythrocytes

Sofía Mira-Martínez,[a,b] Anastasia K. Pickford,[b] Núria Rovira-Graells,[b] Pieter Guetens,[a] Elisabet Tintó-Font,[b] Alfred Cortés,[b,c] Anna Rosanas-Urgell[a]

[a]Institute of Tropical Medicine, Antwerp, Belgium
[b]ISGlobal, Hospital Clínic—Universitat de Barcelona, Barcelona, Catalonia, Spain
[c]ICREA, Barcelona, Catalonia, Spain

ABSTRACT During the intraerythrocytic asexual cycle malaria parasites acquire nutrients and other solutes through a broad selectivity channel localized at the membrane of the infected erythrocyte termed the plasmodial surface anion channel (PSAC). The protein product of the *Plasmodium falciparum* clonally variant *clag3.1* and *clag3.2* genes determines PSAC activity. Switches in the expression of *clag3* genes, which are regulated by epigenetic mechanisms, are associated with changes in PSAC-dependent permeability that can result in resistance to compounds toxic for the parasite, such as blasticidin S. Here, we investigated whether other antimalarial drugs require CLAG3 to reach their intracellular target and consequently are prone to parasite resistance by epigenetic mechanisms. We found that the bis-thiazolium salts T3 (also known as albitiazolium) and T16 require the product of *clag3* genes to enter infected erythrocytes. *P. falciparum* populations can develop resistance to these compounds via the selection of parasites with dramatically reduced expression of both genes. However, other compounds previously demonstrated or predicted to enter infected erythrocytes through transport pathways absent from noninfected erythrocytes, such as fosmidomycin, doxycycline, azithromycin, lumefantrine, or pentamidine, do not require expression of *clag3* genes for their antimalarial activity. This suggests that they use alternative CLAG3-independent routes to access parasites. Our results demonstrate that *P. falciparum* can develop resistance to diverse antimalarial compounds by epigenetic changes in the expression of *clag3* genes. This is of concern for drug development efforts because drug resistance by epigenetic mechanisms can arise quickly, even during the course of a single infection.

KEYWORDS malaria, *Plasmodium falciparum*, drug resistance, epigenetics, plasmodium surface anion channel, *clag3*

Malaria is a major public health problem that affects half of the world's population. *Plasmodium falciparum* is the predominant species in Africa and the most deadly form of the parasite. It is responsible for half a million deaths every year, mostly among children and pregnant women (1). While chemotherapy is the main tool used for malaria control, *P. falciparum* has developed resistance to all antimalarial drugs, including artemisinin combination therapies (ACTs), which are the current frontline treatment (2, 3). Therefore, the appearance and spread of drug resistant parasites is a major obstacle to malaria control and elimination efforts and urges the discovery of new effective compounds to treat infections.

Most of the known mechanisms by which *Plasmodium falciparum* parasites develop resistance to antimalarial drugs are related to changes in the genome, such as single

Citation Mira-Martínez S, Pickford AK, Rovira-Graells N, Guetens P, Tintó-Font E, Cortés A, Rosanas-Urgell A. 2019. Identification of antimalarial compounds that require CLAG3 for their uptake by *Plasmodium falciparum*-infected erythrocytes. Antimicrob Agents Chemother 63:e00052-19. https://doi.org/10.1128/AAC.00052-19.

Address correspondence to Alfred Cortés, alfred.cortes@isglobal.org, or Anna Rosanas-Urgell, arosanas@itg.be.

S.M.-M., A.K.P., and N.R.-G. contributed equally to this work as first authors; A.C. and A.R.-U. contributed equally to this work as senior authors.

Received 9 January 2019
Accepted 1 February 2019
Accepted manuscript posted online 19 February 2019
Published 25 April 2019

nucleotide polymorphisms (SNPs) or gene amplifications (4). SNPs can occur in parasite genes encoding the enzymes targeted by the drug, reducing the drug affinity as in the case of mutations in the *P. falciparum* dihydrofolate reductase (*pfdhfr*) and dihydropteroate synthase (*pfdhps*) genes (5, 6). Mutations associated with resistance can also occur in parasite-encoded transporters causing the active extrusion of the drug out of its site of action, as in the case of mutations in the chloroquine resistance transporter (*pfcrt*) gene associated with the efflux of chloroquine out of the digestive vacuole (7). Likewise, genetic amplifications can increase the expression of the target gene, or the expression of genes encoding transporters, as in the case of amplification of the multidrug resistance protein 1 (*pfmdr1*) gene that leads to accumulation of mefloquine in the digestive vacuole away from its predicted target (8).

Transport activities present in infected erythrocytes that mediate the uptake of solutes unable to enter uninfected erythrocytes are collectively referred to as new permeation pathways (NPPs) (9–11). The plasmodial surface anion channel (PSAC) has been proposed to be the single channel responsible for NPPs. This broad selectivity channel, localized at the membrane of the infected erythrocyte, is essential for the uptake of nutrients and several other solutes (12–16). The protein product of *P. falciparum clag3.1* and *clag3.2* genes, part of the five-member *clag* family that encodes the CLAG/RhopH1 component of the RhopH complex (17), plays a key role for the activity of the PSAC (16, 18–20). Other members of the RhopH complex, RhopH2 and RhopH3, are also necessary for PSAC activity (21–23). The structure of the PSAC has not yet been determined, but protease sensitivity assays and experiments with various transgenic parasite lines suggest that CLAG3s (and possibly also RhopH2 and RhopH3) may participate directly in the formation of the channel rather than only activating a channel formed by other proteins (20, 22, 24). The RhopH complex is initially expressed at the schizont stage and localized in the rhoptries (25). About 20 h after reinvasion, it is transported to the red cell membrane, where it determines PSAC activity (22, 23, 26). The sequence of *clag3.1* and *clag3.2* genes is 95% identical. These genes display clonally variant and mutually exclusive expression, such that commonly only one of the two *clag3* genes is expressed at a time (27). The latter property was observed in culture-adapted parasite lines of different genetic backgrounds (18, 19, 28, 29) and has been later confirmed in uncomplicated human malaria infections (30), although mutual exclusion is not strict (31).

Recently, an epigenetic mechanism of drug resistance involving changes in the expression of *clag3* genes was described in *P. falciparum* (32, 33). Previous studies demonstrated that blasticidin S and leupeptin require PSAC for their transport across the membrane of infected erythrocytes and that *P. falciparum* resistance to these compounds is associated with changes in PSAC function (13, 34, 35). Later, we and others showed that changes in PSAC-mediated transport of blasticidin S were associated with switches in *clag3* genes expression regulated at the epigenetic level (32, 33). Resistance to low blasticidin S concentrations involved the selection of parasites that switched from *clag3.2* to *clag3.1* expression, whereas resistance to high concentrations of the drug was acquired by the selection of parasites with both *clag3* genes simultaneously silenced (32). In all cases, *clag3* silencing is mediated by heterochromatin (31). The pattern of *clag3* expression in the selected parasites is transmitted to the next generations by epigenetic mechanisms even when the drug is no longer present. However, simultaneous silencing of the two genes poses a fitness cost for the parasite and in the absence of selection it is progressively reverted.

Whether other antimalarial drugs require the product of *clag3* genes to reach their intracellular targets and consequently are prone to parasite resistance by this epigenetic mechanism is not known. Most antimalarials are small hydrophobic compounds that can diffuse through lipid membranes and do not require specific channels to enter infected erythrocytes (36). However, large hydrophilic compounds such as blasticidin S and leupeptin require facilitated uptake through PSAC. Drug physicochemical parameters such as molecular size and hydrophobicity indexes, e.g., the logP value, can be used to predict which antimalarial drugs require PSAC-facilitated transport (36, 37).

However, while such *in silico* predictions are informative, only experimental validation can determine which drugs are actually prone to parasite resistance by epigenetic silencing of *clag3* genes.

To address this question, we compared the 50% inhibitory concentrations (IC_{50}s) of selected antimalarial compounds between the blasticidin S-selected 10G-0.6-2 line, which has both *clag3* genes silenced and thus shows deficient PSAC transport, and the parental 10G line, which predominantly expresses *clag3.2* (32). In addition, we selected parasites with some of the drugs and monitored switches in the expression of *clag3* genes during selection. We also investigated PSAC transport in CLAG3-defficient parasites using the reporter compound 5-aminolevulinic acid (5-ALA).

RESULTS

The 10G-0.6-2 line is a valid tool to investigate transport via CLAG3-containing PSAC. We previously showed that the 10G-0.6-2 parasite line, derived from the 10G line selected with a high concentration of blasticidin S, shows dramatically reduced expression of the two *clag3* genes. The clonally variant *clag2* gene is silenced in both the 10G-0.6-2 and the parental 10G lines, whereas the not clonally variant *clag8* and *clag9* genes are expressed in both lines (32). Silencing of *clag3* genes in 10G-0.6-2 results in reduced permeability to structurally diverse compounds, such as blasticidin S, sorbitol, and the canonical amino acid L-alanine (32). To further demonstrate that the 10G-0.6-2 line is a valid tool to identify drugs that require CLAG3 to cross the membrane of the infected erythrocyte, we conducted immunofluorescence assays (IFAs) on 10G-derived lines selected with different concentrations of blasticidin S (32) using an anti-CLAG3 antibody that recognizes both CLAG3.1 and CLAG3.2 (19). By restricting the analysis only to schizonts positive for the mature schizont marker AMA1, we could unambiguously identify parasites in which absence of CLAG3 signal was attributable to epigenetic silencing rather than to parasite stage. We found that essentially all mature schizonts in the parental 10G line express CLAG3 (either CLAG3.1 or CLAG3.2), whereas the proportions of CLAG3-negative mature schizonts were 12% in 10G-0.4, 74% in 10G-0.6, and 98% in 10G-0.6-2 cultures (selected with 0.4, 0.6, and 2 μg/ml of blasticidin S, respectively) (Fig. 1). These experiments at the single cell level show that the vast majority of parasites in 10G-0.6-2 cultures do not express CLAG3 proteins at detectable levels, validating the 10G-06-2 line as an appropriate tool to identify compounds that require CLAG3 to enter the cell.

Parasites with deficient PSAC transport due to silencing of *clag3* genes are resistant to bis-thiazolium salts. To identify antimalarial compounds that require the product of *clag3* genes for efficient transport across the membrane of infected erythrocytes, we selected drugs to be tested based on two criteria. First, we included drugs of clinical relevance for which physicochemical properties suggest that they may not enter infected erythrocytes by passive membrane diffusion, i.e., doxycycline, azithromycin and lumefantrine (36). Second, we included drugs for which there is previous evidence for uptake through NPPs, i.e., fosmidomycin (38), pentamidine (39), and the bis-thiazolium salts T3 (also known as albitiazolium) and T16 (40, 41). As positive controls we included drugs for which there is already clear evidence of CLAG3-dependent uptake through PSAC, i.e., blasticidin S and leupeptin (32, 33). For the selected compounds, we compared the IC_{50} between the 10G-0.6-2 line and the parental 10G line.

Our criteria to consider that the uptake of a drug is impaired by the absence of CLAG3 was a 1.5-fold increase in the IC_{50} in 10G-0.6-2 versus 10G, plus a statistically significant difference ($P < 0.05$). The former criterion was used because we consider that differences of lower magnitude are unlikely to have a major biological significance. Using these criteria, dose-response curves revealed that the 10G-0.6-2 line shows lower sensitivity to blasticidin S, T3, T16, and leupeptin than the 10G line (Fig. 2 and Table 1). These results support the idea that T3 and T16 require the expression of *clag3* genes for their uptake, in addition to leupeptin and blasticidin S for which this was previously demonstrated (13, 32–34). The IC_{50} fold increases in 10G-0.6-2 compared to 10G were

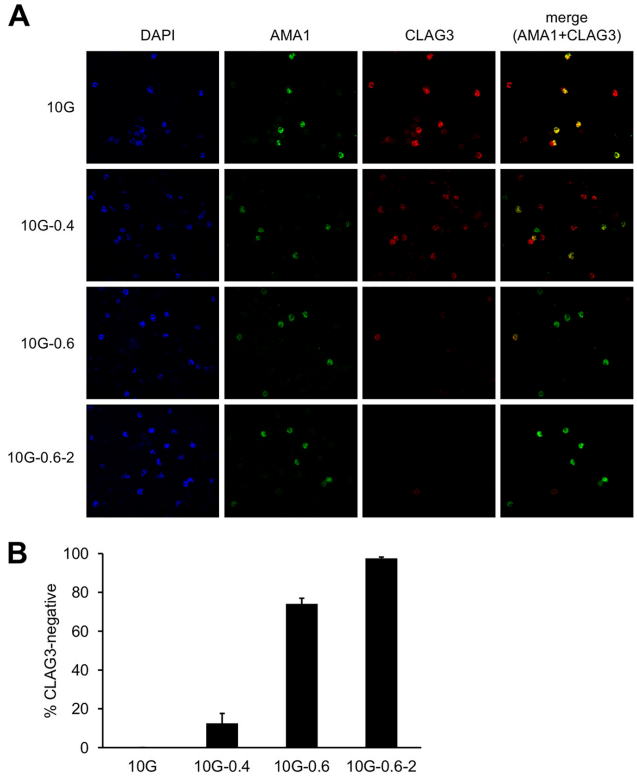


FIG 1 CLAG3 expression in parasite cultures selected with different concentrations of blasticidin S. (A) IFA analysis of mature schizont-infected erythrocytes with anti-CLAG3 and anti-AMA1 antibodies in blasticidin S-selected lines and the parental 10G line. DAPI was used to mark the parasite nuclei. Anti-AMA1 antibodies were used to identify schizonts sufficiently mature for CLAG3-expression. (B) Proportion of AMA1-positive parasites that were negative for CLAG3 fluorescence in each parasite line. The results are averages of two independent biological replicates, along with the standard deviations (SD).

3.3 for T3 and 1.8 for T16, which is lower than the 4.6- and 9.5-fold increases observed for leupeptin and blasticidin S, respectively (Table 1). On the other hand, we did not obtain evidence for *clag3* genes playing a role in the uptake of the other compounds tested, since the difference in IC_{50}s between 10G and 10G-0.6-2 cultures was not statistically significant and/or was of very low magnitude (Fig. 2 and Table 1).

***clag3* expression patterns after selection with different drugs.** We and others have previously shown that adaptation to grow in the presence of blasticidin S is associated with selection of parasites with altered *clag3* expression patterns, validating the involvement of these genes in the transport of the drug (30, 32, 33). Here, we investigated whether sublethal concentrations of other drugs can also select parasites with specific *clag3* expression patterns. In these experiments we tested drugs to which 10G-0.6-2 is less sensitive than 10G (Fig. 2). In addition, doxycycline and fosmidomycin were included in spite of not showing differences between 10G and 10G-0.6-2 because previous studies suggested that they may require NPPs for their transport into infected erythrocytes (36, 38). We considered that selection experiments may be more sensitive than IC_{50} comparisons to detect the involvement of CLAG3 in their transport.

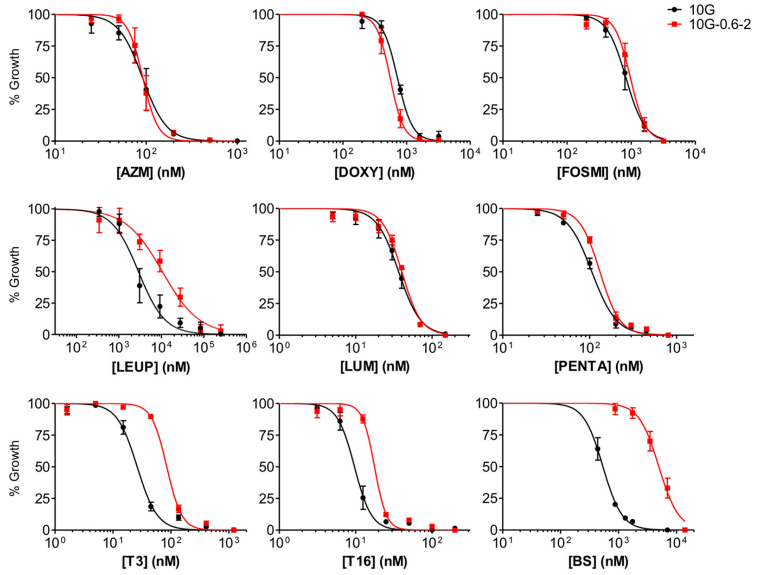


FIG 2 Drug dose-response curves for parasite lines 10G and 10G-0.6-2. The susceptibility to azithromycin (AZM), doxycycline (DOXY), fosmidomycin (FOSMI), leupeptin (LEUP), lumefantrine (LUM), pentamidine (PENTA), T3, T16, and blasticidin S (BS) were compared between the parasite lines 10G (predominantly expressing *clag3.2*) and 10G-0.6-2 (both *clag3* genes silenced). Values are averages of three independent experiments, each performed in triplicate wells, along with the SD.

We selected 10G cultures with the drugs at concentrations ranging from (approximately) the IC_{50} to the IC_{80}. We used these relatively low concentrations because we predicted that this may facilitate adaptation by selection of parasites expressing *clag3.1* or *clag3.2*, if expression of one or the other paralog was associated with differential transport of the drug as in the case of blasticidin S (32). However, for T3 and T16 we observed that toxicity is much higher in the second and subsequent cycles than in the first cycle, such that selection with this range of drug concentrations killed the cultures after only a few cycles. We determined the approximate IC_{50}s of T3 and T16 over two cycles and found that they were ~10 times lower than in a one-cycle assay; we used

TABLE 1 IC_{50}s of different drugs in the 10G and 10G-0.6-2 lines[a]

	IC_{50} (nM)			
Drug	10G	10G-0.6-2	Fold change	*P*
Azithromycin	95 (22)	93 (19)	1.0	0.899
Doxycycline	718 (11)	563 (121)	0.8	0.091
Fosmidomycin	841 (219)	985 (184)	1.2	0.434
Leupeptin	1,753 (1)	8,010 (7)	4.6	0.039*
Lumefantrine	36 (6)	40 (2)	1.1	0.394
Pentamidine	106 (8)	132 (9)	1.2	0.019*
T3	26 (5)	85 (8)	3.3	0.003*
T16	10 (2)	18 (1)	1.8	0.002*
Blasticidin S	530 (96)	5,060 (1,045)	9.5	0.002*

[a]IC_{50} values show the averages of three independent experiments (Fig. 2). Standard deviations are given in parentheses. *P* values were calculated using a two-tailed unpaired *t* test. Significant differences were determined between 10G and 10G-0.6-2 after applying the Benjamini-Hochberg correction for multiple testing, with a false discovery rate of 0.1, and are indicated by an asterisk (*).

T3 at 2.86 nM and T16 at 0.6 nM for selection. For doxycycline, which shows a delayed death effect, we also selected cultures with a concentration of the drug based on the IC_{50} in a two-cycle assay. Cultures were selected until we obtained evidence for adaptation (see Materials and Methods and Fig. S1) or for a maximum of 14 weeks. 10G cultures selected with blasticidin S (650 nM, corresponding to 0.3 μg/ml, a concentration that inhibits growth by ~60%) were used as a positive control. These cultures switched from predominant *clag3.2* to *clag3.1* expression after only 3 weeks of selection (80-fold increase in the *clag3.1/clag3.2* ratio), which is consistent with previous results (32) (Fig. 3A).

Cultures selected with T3 or T16 showed a prominent decrease in the expression levels of both *clag3* genes after only 2 weeks (Fig. 3A and Fig. S2). This was observed in selection experiments performed with the parasite lines 10G and also 3D7-A, a previously described stock of the *P. falciparum* clonal line 3D7, from which the 10G subclone was derived (27, 42). As expected from this result, in cultures selected with T3 the vast majority of parasites were resistant to sorbitol lysis (Fig. 3B) and negative for CLAG3 expression by IFA (Fig. 3C), similar to the blasticidin S-selected 10G-0.6-2 line. Together with the increased IC_{50} values observed for these drugs in the 10G-0.6-2 line, these results demonstrate that the uptake of T3 and T16 by infected erythrocytes requires CLAG3. As in the case of blasticidin S-selected cultures (32, 33), *clag3* silencing was reversible, and removal of the drug resulted in a progressive selection of parasites that express *clag3* (Fig. S3). This is probably attributable to the fitness cost associated with simultaneous silencing of the two *clag3* genes (32). We also selected cultures with lower concentrations of T3, but we did not obtain evidence for any major alteration in the *clag3.1/clag3.2* ratio that would reflect selection of parasites expressing one or the other *clag3* gene (Fig. S4).

On the other hand, 10G cultures selected with fosmidomycin, doxycycline, and leupeptin did not show any clear alteration in the expression of *clag3* genes compared to control 10G cultures maintained in parallel without drug (Fig. 3A).

Parasites with dramatically reduced levels of CLAG3 expression still acquire some compounds through the NPPs. Previous reports demonstrated that the uptake of fosmidomycin or pentamidine depends on NPPs (38, 39), and the uptake of other compounds such as doxycycline, lumefantrine, and azithromycin is strongly predicted to also require NPPs (36). Therefore, the results of the selection experiments with these drugs and especially the similar IC_{50}s between 10G and 10G-0.6-2 were somehow unexpected. To test the possibility that the transport of specific compounds via NPPs is active in the 10G-0.6-2 line in spite of its dramatically reduced levels of CLAG3, we compared the uptake of 5-ALA between the 10G-0.6-2 line and the parental 10G. Inside the infected erythrocyte, 5-ALA is processed to the fluorescent compound protoporphyrin IX (PPIX) that can be visualized by microscopy. 5-ALA has been previously observed to enter infected erythrocytes through NPPs and its uptake requires RhopH3 (a component of the RhopH complex), which indicates that it uses the PSAC (23, 43). We observed that the uptake of 5-ALA was impaired in RhopH3-defficient parasites, as previously reported (23). In contrast, essentially all mature parasite-infected erythrocytes acquired the compound in the 10G, 10G-0.6-2, and T3-selected lines (Fig. 4). This result indicates that CLAG3 is not essential for the acquisition of some specific PSAC substrates such as 5-ALA. This is in contrast to the PSAC-mediated transport of blasticidin S, leupeptin, sorbitol, L-alanine, T3, or T16, which appears to depend more strongly on CLAG3 expression.

DISCUSSION

We and others have previously shown that epigenetic changes in the expression of *clag3* genes modify the permeability of the infected erythrocyte membrane through alterations in PSAC function, and that these changes can confer resistance to compounds such as blasticidin S and leupeptin (13, 32–35, 44). Here, we investigated whether other compounds with antimalarial activity require the product of *clag3* genes to enter the infected erythrocyte and are thus prone to parasite resistance due to

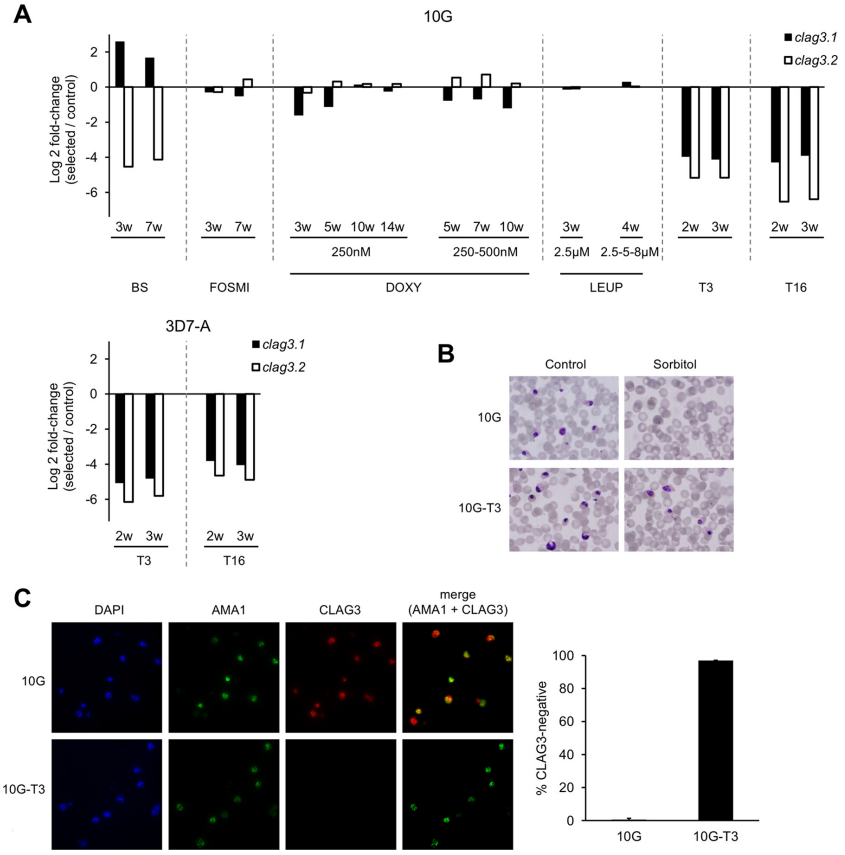


FIG 3 Changes in CLAG3 expression in cultures selected with different drugs. (A) Changes in the transcript levels of *clag3.1* and *clag3.2* in cultures selected with blasticidin S (BS), fosmidomycin (FOSMI), doxycycline (DOXY), leupeptin (LEUP), T3, and T16 compared to unselected control cultures maintained in parallel. The 10G or 3D7-A lines were used for these experiments, as indicated. Cultures selected with DOXY were maintained with the drug at 250 nM for 14 weeks or at 250 nM for 3 weeks and then at 500 nM for 7 additional weeks. Cultures selected with LEUP were first selected at 2.5 μM for 2 weeks and then maintained one additional week at the same concentration (for a total of 3 weeks at 2.5 μM) or one additional week at 5 μM and one week at 8 μM (2.5-5-8 μM bars). Transcript levels are normalized against *rhoph2*, which has temporal expression dynamics similar to *clag3* genes along the asexual cycle. Values are the $\log_2$ of the normalized expression fold change in drug-selected cultures versus cultures maintained in parallel in the absence of drug. Zero indicates the same expression in selected and control cultures, whereas positive values reflect an increase of expression in drug-selected cultures and negative values reflect reduced expression in drug-selected cultures. Individual values are the average of reactions performed in triplicate, but for each drug selection the result of independent biological samples collected at different times (indicated in weeks) is presented. (B) Resistance of late-stage parasites (pigmented trophozoites and schizonts) to treatment with sorbitol in 10G cultures selected with T3 (10G-T3) or unselected 10G cultures maintained in parallel (10G). "Control" are the same cultures before sorbitol treatment. (C) IFA analysis of mature schizont-infected erythrocytes with anti-CLAG3 and anti-AMA1 antibodies in T3-selected (10G-T3) or unselected 10G cultures. DAPI marks parasite nuclei. Anti-AMA1 antibodies were used to identify schizonts sufficiently mature for CLAG3 expression. The proportion of AMA1-positive infected erythrocytes that are negative for CLAG3 in 10G-T3 or unselected 10G cultures is shown. Values in the bar chart are averages of two independent biological replicates, along with the SD.

epigenetic changes in *clag3* genes expression. For this aim, we compared the IC_{50} for several compounds with antimalarial activity between the parental 10G line and the 10G-0.6-2 line, which expresses dramatically reduced levels of *clag3* genes. Of note, the clonally variant gene *clag2*, which has also been implicated in infected erythrocyte

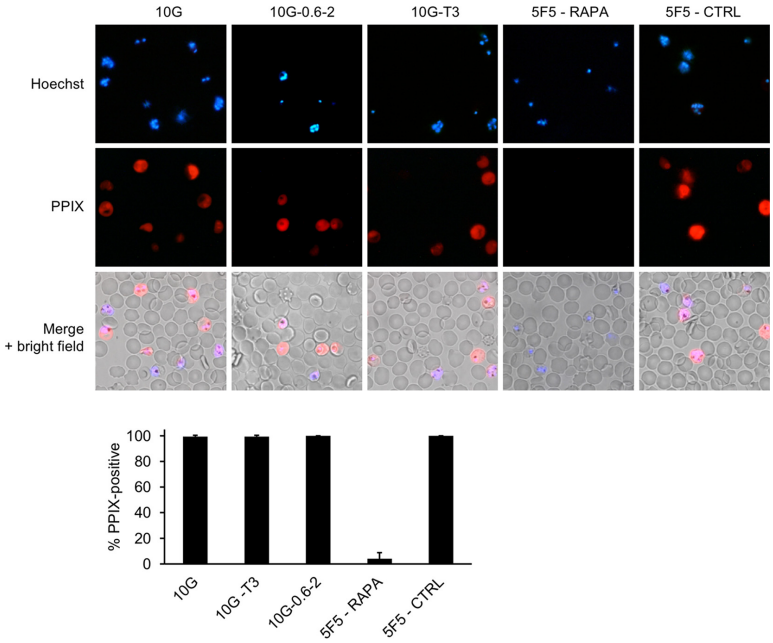


FIG 4 Uptake of 5-ALA by erythrocytes infected with CLAG3-defficient parasites. Analysis was performed with the CLAG3-defficient 10G-0.6-2 and T3-selected 10G lines, the parental 10G control line, and the previously described 5F5 transgenic line treated with rapamycin to induce the deletion of *rhoph3* exons 4 to 6 (5F5-RAPA) or treated in parallel with DMSO solvent (5F5-CTRL). The uptake of 5-ALA and its subsequent conversion to PPIX in infected erythrocytes was visualized by fluorescence microscopy. Parasite nuclei were stained with Hoechst. The bar chart shows the proportion of PPIX fluorescence-positive cells among pigmented parasite-infected erythrocytes. Values are averages of three independent biological replicates, along with the SD.

permeability, is silenced in both the 10G and the 10G-0.6-2 lines (27, 32). This excludes changes in the expression of this gene as a confounding factor for the results obtained. In addition, we selected cultures with some of the drugs for several weeks and measured *clag3.1* and *clag3.2* transcript levels to investigate whether the drugs select for parasites with altered *clag3* expression. Both approaches led to the identification of T3 and T16 as new antimalarial compounds that require CLAG3 for their uptake.

We characterized the 10G-0.6-2 line by IFA using anti-CLAG3 antibodies and found that ~98% of the parasites in the population do not express CLAG3 at detectable levels, which is consistent with the previously described high level of resistance to sorbitol and blasticidin S and dramatically reduced *clag3* transcript levels in this parasite line (32). However, we also characterized PSAC functionality in this parasite line using 5-ALA, a compound that specifically enters infected erythrocytes and is then converted to fluorescent PPIX. The uptake of 5-ALA is inhibited by furosemide (a PSAC/NPPs inhibitor) or by conditional depletion of RhopH3 (23, 43), which demonstrates that it is a PSAC substrate that can be used as a convenient reporter for PSAC activity. Surprisingly, we observed PPIX fluorescence indicative of transport through PSAC in the 10G-0.6-2 line, despite silencing of both *clag3* genes. Although we cannot exclude the possibility that undetected quantitative differences may occur in the transport of 5-ALA between control and CLAG3-defficient parasites, these results suggest that the transport of some compounds into infected erythrocytes may be mediated by CLAG3-

independent PSAC, possibly involving other CLAGs such as CLAG8 or CLAG9. These genes appear to be expressed by all parasites, as they have not been found to show clonally variant expression or to carry epigenetic marks of silencing (45–48). We hypothesize that PSAC involving CLAG8 or CLAG9 may ensure transport of some solutes even when total PSAC transport is highly reduced due to silencing of both *clag3* genes. The PSAC formed in the absence of CLAG3 appears to be able to mediate the uptake of compounds such as 5-ALA, but its ability to transport other compounds such as blasticidin S, sorbitol, L-alanine, T3, or T16 is severely impaired. Altogether, these results indicate that the 10G-0.6-2 line provides an appropriate tool to measure drug uptake in the absence of CLAG3 and CLAG2. However, NPPs likely mediated by PSAC formed in the absence of CLAG3 are active in this parasite line, resulting in the uptake of some solutes. Considering that variant expression and epigenetic silencing has only been reported for CLAG3s and CLAG2, only drugs that show lower activity on the 10G-0.6-2 line are prone to parasite resistance by epigenetic silencing of *clag* genes. Drugs for which uptake may require PSAC but independently of the presence of CLAG3 are unlikely to develop parasite resistance by this mechanism.

Our results demonstrate that in addition to blasticidin S, leupeptin, sorbitol and L-alanine (19, 32, 33), T3 and T16 also require the product of *clag3* genes to enter infected erythrocytes. The bis-thiazolium salts T3 and T16 are choline analogs that inhibit the synthesis of phosphatidylcholine. They are able to cure malaria infections *in vivo* in mice, primates, and humans, but their clinical development was discontinued because of rapid drug clearance in children (49). T3 concentrates massively in infected erythrocytes in an energy-dependent and saturable process, which underlies its antiplasmodial activity at low nanomolar concentrations (50). We observed a clear increase in the IC_{50} values for T3 and T16 in 10G-0.6-2 compared to 10G, which together with silencing of both *clag3* genes after only 2 weeks of selection with these drugs demonstrates that *P. falciparum* can develop resistance to them by changes in the expression of *clag3* genes (30, 32). The concentration of T3 and T16 that was used in the selection experiments was below the IC_{50} of the drugs, which explains the difference with previous reports concluding that choline analogs are likely not prone to drug resistance (50). Resistance to drugs at low concentrations is physiologically relevant because low drug concentrations can be encountered by parasites at some points during a treatment course. However, in spite of selecting parasites with T3 and T16 at low concentrations, we observed silencing of both *clag3* genes, a pattern that in the case of blasticidin S was only observed when using high concentrations ($\gg IC_{50}$) of the drug. Although resistance to blasticidin S at low concentration can be acquired by selection of parasites that switched from expression of one *clag3* gene to expression of the other, which does not pose a fitness cost, this was not observed for T3. Thus, our experiments did not identify additional compounds for which there are differences in sensitivity (likely reflecting differences in transport efficiency) between parasites that express *clag3.1* and parasites that express *clag3.2*.

The IC_{50} increase for T3 and T16 in the 10G-0.6-2 line relative to 10G is moderate compared to the increase observed for blasticidin S (~2- to ~3-fold versus ~10-fold). This is consistent with previous reports showing that the transport of T3/T16 into the infected erythrocytes is highly reduced, but not eliminated, by treatment with the NPP inhibitor furosemide. The authors concluded that transport of these compounds occurs mainly through PSAC/NPPs, but residual transport (~15%) is nonsaturable and continues to occur even in the presence of high concentrations of the inhibitor (40, 41). Residual transport of T3/T16 through membrane diffusion or endogenous transporters may explain the relatively modest differences in IC_{50} values between 10G and 10G-0.6-2 cultures.

A previous study analyzing the effect of T4, a bis-thiazolium compound structurally related with T3 and T16, did not detect significant transcriptional changes in 3D7 cultures after 30 min to 36 h of exposure to the drug (51). The apparent discrepancy with our results reflects the very different experimental approaches used: while the previous study analyzed changes in mRNA levels soon after drug exposure to explore

the occurrence of directed protective transcriptional responses, we studied adaptations at the transcriptional level after several cycles of selection with the drug. Our approach revealed changes in the expression of *clag3* genes linked to adaptation, indicating that selection of parasites with specific expression patterns of clonally variant genes can occur in the development of resistance to compounds to which the parasite is unable to mount a directed transcriptional response. Of note, natural selection of parasites with specific transcriptional patterns is the basis of bet-hedging adaptive strategies in *P. falciparum* (47, 48).

We did not observe major differences in the IC_{50} for doxycycline, fosmidomycin, azithromycin, lumefantrine, or pentamidine between the 10G and 10G-0.6-2 lines. However, there is robust previous data for fosmidomycin and pentamidine indicating that these drugs require NPPs for their entry into infected erythrocytes: their uptake is abrogated by NPPs inhibitors, and they concentrate massively inside the infected erythrocytes (38, 39). Hence, the most plausible explanation for the similar sensitivity to these compounds between the 10G-0.6-2 and 10G lines is that their uptake occurs via CLAG3-independent NPPs, as in the case of the PSAC substrate 5-ALA. We consider it likely that PSAC involving nonvariantly expressed CLAGs, which we predict to mediate 5-ALA uptake in CLAG3-defficient lines, also mediates the uptake of drugs such as fosmidomycin and pentamidine in these lines. However, we cannot fully exclude the possibility that the residual expression of CLAG3 in 10G-0.6-2, which is undistinguishable from background signal in IFA experiments, is sufficient to mediate transport of these compounds. In any case, given that doxycycline, fosmidomycin, lumefantrine, azithromycin, and pentamidine can access and kill parasites expressing dramatically reduced levels of *clag3* genes, it is unlikely that *P. falciparum* can develop resistance to these compounds by selection of parasite subpopulations with altered *clag3* expression as in the case of blasticidin S, T3, or T16.

Altogether, we identified T3 and T16 as antimalarials that require CLAG3-containing PSAC to enter the infected erythrocyte, and thus these drugs are prone to parasite resistance by epigenetic changes in the expression of *clag3* genes. This is an important concern because resistance acquired at the epigenetic level can arise quickly during the course of a single infection and is easily reversible, providing the parasite with a level of plasticity toward susceptible drugs that would promptly render them ineffective. On the other hand, our results suggest that other compounds known or predicted to require NPPs for their uptake by infected erythrocytes may use PSAC involving CLAG8 or CLAG9, or other channels different from PSAC. To determine which drugs require any form of PSAC to access infected erythrocytes, experiments similar to the ones presented here could be performed using conditional KO lines for *rhoph2* or *rhoph3* (21–23). However, even if these experiments identified additional drugs that require PSAC for their uptake, these drugs would not be prone to the resistance mechanism regulated at the epigenetic level studied here if they show normal transport in the 10G-0.6-2 line.

MATERIALS AND METHODS

Parasite cultures. The 3D7-A stock of the clonal *P. falciparum* line 3D7 and the 10G subclone of 3D7-A have been previously described and characterized (27, 42, 52). The 10G-0.4 and 10G-0.6 lines were generated by selection of the 10G line with blasticidin S at 0.4 and 0.6 μg/ml, respectively, whereas the 10G-0.6-2 line was generated by sequential selection with 0.6 and 2 μg/ml (32). The 5F5 inducible *rhoph3* disruption line has been previously described (23). Parasites were cultured in O erythrocytes at a 3% hematocrit with inactivated human serum under standard culture conditions, except for selection experiments with T3 and their controls that were performed with B^+ erythrocytes and media containing Albumax II instead of serum. The 10G-0.6-2 parasite line was regularly cultured under blasticidin S pressure (2 μg/ml) to maintain silencing of both *clag3* genes (32). For synchronization, we used treatment with 5% sorbitol, except for the 10G-0.6-2 line, the 10G line selected with T3 or T16, and the unselected 10G controls analyzed in parallel that were synchronized with L-proline as previously described (31). To prepare RNA for transcriptional analysis, cultures were harvested when the majority of parasites were at the schizont stage, i.e., when *clag3* genes are expressed, and a small proportion of schizonts had already burst.

IFAs. IFAs were performed on Percoll-purified parasites at the mature schizont stage because at this stage CLAG3 yields a strong and unambiguous signal in rhoptries, whereas detection at other stages in which the protein is only present in the surface of infected erythrocytes is less clear. Air-dried smears were fixed for 10 min with 1% paraformaldehyde and permeabilized for 10 min with 0.1% Triton X-100

in phosphate-buffered saline (PBS). Smears were incubated with rabbit anti-3D7 AMA1 (1:2,000; kindly provided by Robin F. Anders, La Trobe University, Australia) (53) and mouse anti-CLAG3 (1:2,000; from mouse 167#2, kindly provided by Sanjay A. Desai, NIAID-NIH) (19) polyclonal antibodies. We used secondary anti-rabbit antibodies conjugated with Alexa Fluor 488 (Life Technologies, A-11034) and anti-mouse antibodies conjugated with Alexa Fluor 546 (Life Technologies, A-10036). Nuclei were stained with DAPI (4′,6′-diamidino-2-phenylindole). Preparations were observed under a confocal Leica TCS-SP5 microscope with LAS-AF image acquisition software and processed using ImageJ. AMA1, which is expressed later than CLAG3 during intraerythrocytic development and does not show variant expression, was used to identify parasites that were mature enough for CLAG3 expression. This enabled us to distinguish between parasites that did not express CLAG3 because they were at a too early stage of development from parasites in which both *clag3* genes were simultaneously silenced at the epigenetic level. The analysis of CLAG3 expression was restricted to AMA1-positive parasites: the proportion of schizonts that have both CLAG3s silenced was determined by counting >100 AMA1-postitive mature schizonts in each replicate experiment.

Drugs. Blasticidin S, lumefantrine, doxycycline, fosmidomycin, leupeptin, and azithromycin were purchased from Sigma-Aldrich (reference numbers 15205, L5420, D9891, F8682, L8511, and PZ0007, respectively). T3, T16, and pentamidine were kindly provided by Henri J. Vial (CNRS, Montpellier, France). Stock solutions for each drug were prepared as follows: azithromycin and pentamidine in dimethyl sulfoxide (DMSO) at 10 mM, blasticidin S in H_2O at 10 mM, doxycycline in methanol at 20 mM, fosmidomycin in PBS at 1 mM, T3 and T16 in H_2O at 10 mM, and lumefantrine at 1 mM in a 1:1:1 volume mix of Tween 80, ethanol, and linoleic acid (prepared from a 5 mM stock in DMSO).

Growth inhibition assays. To determine the IC_{50} of the different compounds in 10G and 10G-0.6-2 cultures, we used a previously described SYBR Green-based assay (54) with some adjustments. We incubated parasites in triplicate wells for 96 h for experiments with blasticidin S, leupeptin, lumefantrine, fosmidomycin, pentamidine, T3, and T16, which kill the parasites in the first cycle after drug administration, or for 144 h for drugs that produce a delayed death effect (55), killing parasites only at the second cycle after adding the drug (doxycycline and azithromycin). Initial parasitemias were adjusted for each incubation time and parasite strain in order to prevent the culture from collapsing, while ensuring a sufficiently high final parasitemia (in the absence of drug) for accurate determination of the inhibition levels. The parasitemia of synchronized ring-stage cultures was determined by light microscopy and adjusted to 0.1 to 0.2 or 0.05% for 96- and 144-h experiments with the 10G line, respectively, and to 0.2 to 0.5% or 0.1% for 96- and 144-h assays with 10G-0.6-2, respectively. In experiments with 10G-0.6-2 cultures, which were regularly maintained under blasticidin S pressure, blasticidin S was removed 5 h before starting the assay. After 96 or 144 h of incubation, parasites were exposed to one freeze-thaw cycle, and 100 μl of lysis buffer (54) with SYBR Green 5× was added per well. Plates were incubated at 37°C in the dark for 3 h, and then fluorescence was measured on a plate reader (Victor X3; Perkin-Elmer) with excitation and emission wavelength bands centered at 485 and 535 nm, respectively. After log transforming the drug concentrations, the data were fit to sigmoidal dose-response curves using GraphPad Prism (version 5), setting the maximum to 100 and the minimum to 0. IC_{50} values were compared between 10G and 10G-0.6-2 using a two-tailed *t* test for unpaired data (Stata, version 12). We applied the Benjamini-Hochberg multiple testing correction of *P* values (56), with a false discovery rate of 0.1.

Drug selection experiments. To select cultures with doxycycline, fosmidomycin, T3, T16, leupeptin, or blasticidin S, the drugs were initially applied to cultures at the ring stage. Cultures were maintained under sublethal drug concentrations (determined in preliminary experiments) for a maximum of 14 weeks, or until clear evidence of adaptation to the drug was observed, i.e., an increase in the growth rate in the presence of the drug compared to the initial growth rate when the drug was first added (see Fig. S1 in the supplemental material). Unselected cultures were maintained in parallel. In all selection experiments, parasites were harvested for RNA extraction every 3 to 5 weeks.

Transcriptional analysis. For RNA purification, erythrocyte pellets were collected in TRIzol (Invitrogen), and phase separation was conducted according to the manufacturer's instructions. RNA was purified from the aqueous phase using an RNeasy minikit (Qiagen) as previously described (30) and reverse transcribed using a reverse transcription system (Promega). To exclude gDNA contamination, parallel reactions were performed in the absence of reverse transcriptase. cDNAs were analyzed by quantitative PCR in triplicate wells using Power SYBR Green master mix (Applied Biosystems) in a StepOnePlus or a 7900HT Fast real-time PCR system (Applied Biosystems). *clag3.1* and *clag3.2* expression values, in arbitrary units, were calculated using the standard curve method as previously described (30). The primers used have been previously described (30).

Analysis of erythrocyte membrane permeability by 5-ALA uptake. The 5F5 inducible *rhoph3* disruption line was treated with rapamycin at the cycle prior to testing 5-ALA uptake to induce deletion of exons 4 to 6 (23) or treated in parallel with DMSO as a control. 5-ALA uptake was determined as previously described (23, 43), with minor modifications. In brief, cultures of synchronous ring-stage parasites were incubated overnight in normal RPMI-based parasite culture medium with Albumax II supplemented with 200 μM 5-ALA (Sigma-Aldrich, catalog no. A3785). After uptake by infected erythrocytes, 5-ALA is converted to PPIX. Just prior to analysis, parasite nuclei were stained with 2 μg/ml Hoechst for 10 min at 37°C and washed with PBS. Samples were placed in an eight-well chamber slide for live cell fluorescence microscopy analysis using the Fluorescence Imaging System Leica AF6000. Images were captured with the same acquisition settings for all samples, so that signal intensities are directly comparable. Images were analyzed using ImageJ software.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/AAC.00052-19.

SUPPLEMENTAL FILE 1, PDF file, 0.1 MB.

ACKNOWLEDGMENTS

We thank Sanjay A. Desai (National Institutes of Health) for kindly providing mouse anti-CLAG3 polyclonal antibody; Robin F. Anders (La Trobe, Australia) for the rabbit anti-3D7 AMA1 polyclonal antibody; Henri J. Vial and Sharon Wein (CNRS, Montpellier, France) for providing T3, T16, and pentamidine; and Emma Sherling and Mike Blackman for the *P. falciparum* conditional *rhoph3* deletion line 5F5.

This study was supported by the Spanish Ministry of Economy and Competitiveness through the Agencia Estatal de Investigación, cofunded by the European Regional Development Fund (ERDF/FEDER), European Union (SAF2013-43601-R and SAF2016-76190-R to A.C.); the Secretary for Universities and Research, Department of Economy and Knowledge, Government of Catalonia (2014 SGR 485 to A.C.); the Institute of Tropical Medicine, Antwerp (funding to A.R.-U.). ISGlobal is a member of the CERCA Program, Government of Catalonia. ITM and ISGlobal are members of the Trans Global Health–Erasmus Mundus Joint Doctorate Program, European Union (scholarship to S.M.-M.). E.T.-F. is supported by a fellowship from the Spanish Ministry of Economy and Competitiveness (BES-2014-067901), cofunded by the European Social Fund (ESF), European Commission. A.K.P. is supported by a fellowship from the Secretary for Universities and Research, Catalan Government (FI_B 00373), cofunded by the European Social Fund (ESF), European Commission. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

REFERENCES

1. World Health Organization. 2016. Global malaria programme, World Malaria Report 2016. World Health Organization, Geneva, Switzerland.
2. Ashley EA, Dhorda M, Fairhurst RM, Amaratunga C, Lim P, Suon S, Sreng S, Anderson JM, Mao S, Sam B, Sopha C, Chuor CM, Nguon C, Sovannaroth S, Pukrittayakamee S, Jittamala P, Chotivanich K, Chutasmit K, Suchatsoonthorn C, Runcharoen R, Hien TT, Thuy-Nhien NT, Thanh NV, Phu NH, Htut Y, Han K-T, Aye KH, Mokuolu OA, Olaosebikan RR, Folaranmi OO, Mayxay M, Khanthavong M, Hongvanthong B, Newton PN, Onyamboko MA, Fanello CI, Tshefu AK, Mishra N, Valecha N, Phyo AP, Nosten F, Yi P, Tripura R, Borrmann S, Bashraheil M, Peshu J, Faiz MA, Ghose A, Hossain MA, Samad R, Rahman MR, Hasan MM, Islam A, Miotto O, Amato R, MacInnis B, Stalker J, Kwiatkowski DP, Bozdech Z, Jeeyapant A, Cheah PY, Sakulthaew T, Chalk J, Intharabut B, Silamut K, Lee SJ, Vihokhern B, Kunasol C, Imwong M, Tarning J, Taylor WJ, Yeung S, Woodrow CJ, Flegg JA, Das D, Smith J, Venkatesan M, Plowe CV, Stepniewska K, Guerin PJ, Dondorp AM, Day NP, White NJ. 2014. Spread of artemisinin resistance in *Plasmodium falciparum* malaria. N Engl J Med 371:411–423. https://doi.org/10.1056/NEJMoa1314981.
3. Amaratunga C, Lim P, Suon S, Sreng S, Mao S, Sopha C, Sam B, Dek D, Try V, Amato R, Blessborn D, Song L, Tullo GS, Fay MP, Anderson JM, Tarning J, Fairhurst RM. 2016. Dihydroartemisinin-piperaquine resistance in *Plasmodium falciparum* malaria in Cambodia: a multisite prospective cohort study. Lancet Infect Dis 16:357–365. https://doi.org/10.1016/S1473-3099(15)00487-9.
4. Petersen I, Eastman R, Lanzer M. 2011. Drug-resistant malaria: molecular mechanisms and implications for public health. FEBS Lett 585:1551–1562. https://doi.org/10.1016/j.febslet.2011.04.042.
5. Peterson DS, Walliker D, Wellems TE. 1988. Evidence that a point mutation in dihydrofolate reductase-thymidylate synthase confers resistance to pyrimethamine in falciparum malaria. Proc Natl Acad Sci U S A 85:9114–9118. https://doi.org/10.1073/pnas.85.23.9114.
6. Triglia T, Menting JG, Wilson C, Cowman AF. 1997. Mutations in dihydropteroate synthase are responsible for sulfone and sulfonamide resistance in *Plasmodium falciparum*. Proc Natl Acad Sci U S A 94:13944–13949. https://doi.org/10.1073/pnas.94.25.13944.
7. Lakshmanan V, Bray PG, Verdier-Pinard D, Johnson DJ, Horrocks P, Muhle RA, Alakpa GE, Hughes RH, Ward SA, Krogstad DJ, Sidhu AB, Fidock DA. 2005. A critical role for PfCRT K76T in *Plasmodium falciparum* verapamil-reversible chloroquine resistance. EMBO J 24:2294–2305. https://doi.org/10.1038/sj.emboj.7600681.
8. Barnes DA, Foote SJ, Galatis D, Kemp DJ, Cowman AF. 1992. Selection for high-level chloroquine resistance results in deamplification of the *pfmdr1* gene and increased sensitivity to mefloquine in *Plasmodium falciparum*. EMBO J 11:3067–3075. https://doi.org/10.1002/j.1460-2075.1992.tb05378.x.
9. Elford BC, Haynes JD, Chulay JD, Wilson RJM. 1985. Selective stage-specific changes in the permeability to small hydrophilic solutes of human erythrocytes infected with *Plasmodium falciparum*. Mol and Biochem Parasitol 16:43–60. https://doi.org/10.1016/0166-6851(85)90048-9.
10. Ginsburg H, Kutner S, Krugliak M, Cabantchik ZI. 1985. Characterization of permeation pathways appearing in the host membrane of *Plasmodium falciparum* infected red blood cells. Mol Biochem Parasitol 14:313–322. https://doi.org/10.1016/0166-6851(85)90059-3.
11. Neame KD, Homewood CA. 1975. Alterations in the permeability of mouse erythrocytes infected with the malaria parasite, *Plasmodium berghei*. Int J Parasitol 5:537–540. https://doi.org/10.1016/0020-7519(75)90046-6.
12. Desai SA, Bezrukov SM, Zimmerberg J. 2000. A voltage-dependent channel involved in nutrient uptake by red blood cells infected with the malaria parasite. Nature 406:1001–1005. https://doi.org/10.1038/35023000.
13. Hill DA, Pillai AD, Nawaz F, Hayton K, Doan L, Lisk G, Desai SA. 2007. A blasticidin S-resistant *Plasmodium falciparum* mutant with a defective plasmodial surface anion channel. Proc Natl Acad Sci U S A 104:1063–1068. https://doi.org/10.1073/pnas.0610353104.
14. Alkhalil A, Cohn JV, Wagner MA, Cabrera JS, Rajapandi T, Desai SA. 2004. *Plasmodium falciparum* likely encodes the principal anion channel on infected human erythrocytes. Blood 104:4279–4286. https://doi.org/10.1182/blood-2004-05-2047.
15. Desai SA. 2014. Why do malaria parasites increase host erythrocyte permeability? Trends Parasitol 30:151–159. https://doi.org/10.1016/j.pt.2014.01.003.
16. Desai SA. 2012. Ion and nutrient uptake by malaria parasite-infected erythrocytes. Cell Microbiol 14:1003–1009. https://doi.org/10.1111/j.1462-5822.2012.01790.x.

17. Kaneko O, Tsuboi T, Ling IT, Howell S, Shirano M, Tachibana M, Cao YM, Holder AA, Torii M. 2001. The high molecular mass rhoptry protein, RhopH1, is encoded by members of the *clag* multigene family in *Plasmodium falciparum* and *Plasmodium yoelii*. Mol Biochem Parasitol 118:223–231. https://doi.org/10.1016/S0166-6851(01)00391-7.
18. Pillai AD, Nguitragool W, Lyko B, Dolinta K, Butler MM, Nguyen ST, Peet NP, Bowlin TL, Desai SA. 2012. Solute restriction reveals an essential role for *clag3*-associated channels in malaria parasite nutrient acquisition. Mol Pharmacol 82:1104–1114. https://doi.org/10.1124/mol.112.081224.
19. Nguitragool W, Bokhari AA, Pillai AD, Rayavara K, Sharma P, Turpin B, Aravind L, Desai SA. 2011. Malaria parasite *clag3* genes determine channel-mediated nutrient uptake by infected red blood cells. Cell 145:665–677. https://doi.org/10.1016/j.cell.2011.05.002.
20. Nguitragool W, Rayavara K, Desai SA. 2014. Proteolysis at a specific extracellular residue implicates integral membrane CLAG3 in malaria parasite nutrient channels. PLoS One 9:e93759. https://doi.org/10.1371/journal.pone.0093759.
21. Counihan NA, Chisholm SA, Bullen HE, Srivastava A, Sanders PR, Jonsdottir TK, Weiss GE, Ghosh S, Crabb BS, Creek DJ, Gilson PR, de Koning-Ward TF. 2017. *Plasmodium falciparum* parasites deploy RhopH2 into the host erythrocyte to obtain nutrients, grow, and replicate. Elife 6:e23217. https://doi.org/10.7554/eLife.23217.
22. Ito D, Schureck MA, Desai SA. 2017. An essential dual-function complex mediates erythrocyte invasion and channel-mediated nutrient uptake in malaria parasites. Elife 6:e23485. https://doi.org/10.7554/eLife.23485.
23. Sherling ES, Knuepfer E, Brzostowski JA, Miller LH, Blackman MJ, van Ooij C. 2017. The *Plasmodium falciparum* rhoptry protein RhopH3 plays essential roles in host cell invasion and nutrient uptake. Elife 6:e23239. https://doi.org/10.7554/eLife.23239.
24. Gupta A, Balabaskaran-Nina P, Nguitragool W, Saggu GS, Schureck MA, Desai SA. 2018. CLAG3 self-associates in malaria parasites and quantitatively determines nutrient uptake channels at the host membrane. Mol Biol 9:e02293-17. https://doi.org/10.1128/mBio.02293-17.
25. Cooper JA, Ingram LT, Bushell GR, Fardoulys CA, Stenzel D, Schofield L, Saul AJ. 1988. The 140/130/105 kilodalton protein complex in the rhoptries of *Plasmodium falciparum* consists of discrete polypeptides. Mol Biochem Parasitol 29:251–260. https://doi.org/10.1016/0166-6851(88)90080-1.
26. Gupta A, Thiruvengadam G, Desai SA. 2015. The conserved *clag* multigene family of malaria parasites: essential roles in host-pathogen interaction. Drug Resist Updat 18:47–54. https://doi.org/10.1016/j.drup.2014.10.004.
27. Cortes A, Carret C, Kaneko O, Yim Lim BY, Ivens A, Holder AA. 2007. Epigenetic silencing of *Plasmodium falciparum* genes linked to erythrocyte invasion. PLoS Pathog 3:e107. https://doi.org/10.1371/journal.ppat.0030107.
28. Comeaux CA, Coleman BI, Bei AK, Whitehurst N, Duraisingh MT. 2011. Functional analysis of epigenetic regulation of tandem RhopH1/*clag* genes reveals a role in *Plasmodium falciparum* growth. Mol Microbiol 80:378–390. https://doi.org/10.1111/j.1365-2958.2011.07572.x.
29. Crowley VM, Rovira-Graells N, de Pouplana LR, Cortés A. 2011. Heterochromatin formation in bistable chromatin domains controls the epigenetic repression of clonally variant *Plasmodium falciparum* genes linked to erythrocyte invasion. Mol Microbiol 80:391–406. https://doi.org/10.1111/j.1365-2958.2011.07574.x.
30. Mira-Martinez S, van Schuppen E, Amambua-Ngwa A, Bottieau E, Affara M, Van Esbroeck M, Vlieghe E, Guetens P, Rovira-Graells N, Gomez-Perez GP, Alonso PL, D'Alessandro U, Rosanas-Urgell A, Cortes A. 2017. Expression of the *Plasmodium falciparum* clonally variant *clag3* genes in human infections. J Infect Dis 215:938–945. https://doi.org/10.1093/infdis/jix053.
31. Rovira-Graells N, Crowley VM, Bancells C, Mira-Martínez S, Ribas de Pouplana L, Cortés A. 2015. Deciphering the principles that govern mutually exclusive expression of *Plasmodium falciparum clag3* genes. Nucleic Acids Res 43:8243–8257. https://doi.org/10.1093/nar/gkv730.
32. Mira-Martinez S, Rovira-Graells N, Crowley VM, Altenhofen LM, Llinas M, Cortes A. 2013. Epigenetic switches in *clag3* genes mediate blasticidin S resistance in malaria parasites. Cell Microbiol 15:1913–1923. https://doi.org/10.1111/cmi.12162.
33. Sharma P, Wollenberg K, Sellers M, Zainabadi K, Galinsky K, Moss E, Nguitragool W, Neafsey D, Desai SA. 2013. An epigenetic antimalarial resistance mechanism involving parasite genes linked to nutrient uptake. J Biol Chem 288:19429–19440. https://doi.org/10.1074/jbc.M113.468371.
34. Lisk G, Pain M, Gluzman IY, Kambhampati S, Furuya T, Su XZ, Fay MP, Goldberg DE, Desai SA. 2008. Changes in the plasmodial surface anion channel reduce leupeptin uptake and can confer drug resistance in *Plasmodium falciparum*-infected erythrocytes. Antimicrob Agents Chemother 52:2346–2354. https://doi.org/10.1128/AAC.00057-08.
35. Hill DA, Desai SA. 2010. Malaria parasite mutants with altered erythrocyte permeability: a new drug resistance mechanism and important molecular tool. Future Microbiol 5:81–97. https://doi.org/10.2217/fmb.09.109.
36. Basore K, Cheng Y, Kushwaha AK, Nguyen ST, Desai SA. 2015. How do antimalarial drugs reach their intracellular targets? Front Pharmacol 6:91. https://doi.org/10.3389/fphar.2015.00091.
37. Lisk G, Pain M, Sellers M, Gurnev PA, Pillai AD, Bezrukov SM, Desai SA. 2010. Altered plasmodial surface anion channel activity and *in vitro* resistance to permeating antimalarial compounds. Biochim Biophys Acta 1798:1679–1688. https://doi.org/10.1016/j.bbamem.2010.04.013.
38. Baumeister S, Wiesner J, Reichenberg A, Hintz M, Bietz S, Harb OS, Roos DS, Kordes M, Friesen J, Matuschewski K, Lingelbach K, Jomaa H, Seeber F. 2011. Fosmidomycin uptake into *Plasmodium* and *Babesia*-infected erythrocytes is facilitated by parasite-induced new permeability pathways. PLoS One 6:e19334. https://doi.org/10.1371/journal.pone.0019334.
39. Stead AM, Bray PG, Edwards IG, DeKoning HP, Elford BC, Stocks PA, Ward SA. 2001. Diamidine compounds: selective uptake and targeting in *Plasmodium falciparum*. Mol Pharmacol 59:1298–1306. https://doi.org/10.1124/mol.59.5.1298.
40. Biagini GA, Richier E, Bray PG, Calas M, Vial H, Ward SA. 2003. Heme binding contributes to antimalarial activity of bis-quaternary ammoniums. Antimicrob Agents Chemother 47:2584–2589. https://doi.org/10.1128/AAC.47.8.2584-2589.2003.
41. Wein S, Maynadier M, Bordat Y, Perez J, Maheshwari S, Bette-Bobillo P, Tran Van Ba C, Penarete-Vargas D, Fraisse L, Cerdan R, Vial H. 2012. Transport and pharmacodynamics of albitiazolium, an antimalarial drug candidate. Br J Pharmacol 166:2263–2276. https://doi.org/10.1111/j.1476-5381.2012.01966.x.
42. Cortes A, Benet A, Cooke BM, Barnwell JW, Reeder JC. 2004. Ability of *Plasmodium falciparum* to invade Southeast Asian ovalocytes varies between parasite lines. Blood 104:2961–2966. https://doi.org/10.1182/blood-2004-06-2136.
43. Sigala PA, Crowley JR, Henderson JP, Goldberg DE. 2015. Deconvoluting heme biosynthesis to target blood-stage malaria parasites. Elife 4:e09143. https://doi.org/10.7554/eLife.09143.
44. Lisk G, Scott S, Solomon T, Pillai AD, Desai SA. 2007. Solute-inhibitor interactions in the plasmodial surface anion channel reveal complexities in the transport process. Mol Pharmacol 71:1241–1250. https://doi.org/10.1124/mol.106.030734.
45. Lopez-Rubio JJ, Mancio-Silva L, Scherf A. 2009. Genome-wide analysis of heterochromatin associates clonally variant gene regulation with perinuclear repressive centers in malaria parasites. Cell Host Microbe 5:179–190. https://doi.org/10.1016/j.chom.2008.12.012.
46. Flueck C, Bartfai R, Volz J, Niederwieser I, Salcedo-Amaya AM, Alako BT, Ehlgen F, Ralph SA, Cowman AF, Bozdech Z, Stunnenberg HG, Voss TS. 2009. *Plasmodium falciparum* heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors. PLoS Pathog 5:e1000569. https://doi.org/10.1371/journal.ppat.1000569.
47. Rovira-Graells N, Gupta AP, Planet E, Crowley VM, Mok S, Ribas de Pouplana L, Preiser PR, Bozdech Z, Cortes A. 2012. Transcriptional variation in the malaria parasite *Plasmodium falciparum*. Genome Res 22:925–938. https://doi.org/10.1101/gr.129692.111.
48. Cortes A, Deitsch KW. 2017. Malaria epigenetics. Cold Spring Harb Perspect Med 7:a025528. https://doi.org/10.1101/cshperspect.a025528.
49. Wein S, Taudon N, Maynadier M, Tran Van Ba C, Margout D, Bordat Y, Fraisse L, Wengelnik K, Cerdan R, Bressolle-Gomeni F, Vial HJ. 2017. High accumulation and *in vivo* recycling of the new antimalarial albitiazolium lead to rapid parasite death. Antimicrob Agents Chemother 61:e00352-17. https://doi.org/10.1128/AAC.00352-17.
50. Wein S, Tran Van Ba C, Maynadier M, Bordat Y, Perez J, Peyrottes S, Fraisse L, Vial HJ. 2014. New insight into the mechanism of accumulation and intraerythrocytic compartmentation of albitiazolium, a new type of antimalarial. Antimicrob Agents Chemother 58:5519–5527. https://doi.org/10.1128/AAC.00040-14.
51. Le Roch KG, Johnson JR, Ahiboh H, Chung DW, Prudhomme J, Plouffe D, Henson K, Zhou Y, Witola W, Yates JR, Mamoun CB, Winzeler EA, Vial H. 2008. A systematic approach to understand the mechanism of action of the bisthiazolium compound T4 on the human malaria parasite, *Plasmodium falciparum*. BMC Genomics 9:513. https://doi.org/10.1186/1471-2164-9-513.
52. Cortes A. 2005. A chimeric *Plasmodium falciparum* Pfnbp2b/Pfnbp2a

gene originated during asexual growth. Int J Parasitol 35:125–130. https://doi.org/10.1016/j.ijpara.2004.11.004.
53. Healer J, Murphy V, Hodder AN, Masciantonio R, Gemmill AW, Anders RF, Cowman AF, Batchelor A. 2004. Allelic polymorphisms in apical membrane antigen-1 are responsible for evasion of antibody-mediated inhibition in *Plasmodium falciparum*. Mol Microbiol 52:159–168. https://doi.org/10.1111/j.1365-2958.2003.03974.x.
54. Worldwide Antimalarial Resistance Network. 2011. *In vitro* module: *Plasmodium falciparum* drug sensitivity assay using SYBR Green I assay technique WWARN procedure. Worldwide Antimalarial Resistance Network, LOCATION.
55. Goodman CD, Pasaje CF, Kennedy K, McFadden GI, Ralph SA. 2016. Targeting protein translation in organelles of the Apicomplexa. Trends Parasitol 32:953–965. https://doi.org/10.1016/j.pt.2016.09.011.
56. Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol 57:289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

SUPPLEMENTARY FIGURES

Identification of antimalarial compounds that require CLAG3 for their uptake by *Plasmodium falciparum*-infected erythrocytes.

Sofía Mira-Martínez, Anastasia K. Pickford, Núria Rovira-Graells, Pieter Guetens, Elisabet Tintó-Font, Alfred Cortés & Anna Rosanas-Urgell

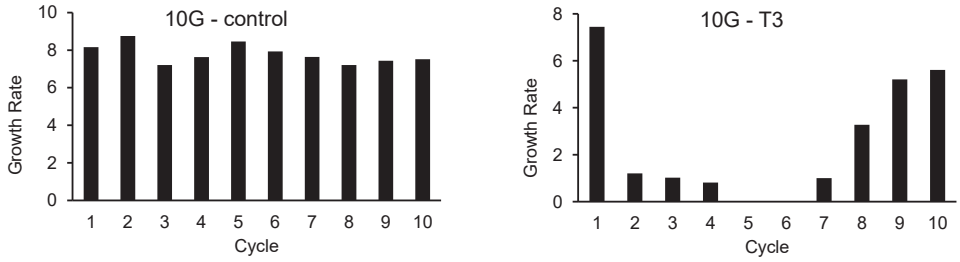


Fig. S1. Parasite growth dynamics in a representative drug selection experiment (representative of two independent selection experiments with T3). Growth rate of 10G cultures during selection with the drug T3 (2.86 nM, 10G - T3) or maintained in parallel in the absence of drug (10G - control). There is strong growth inhibition in the initial cycles after drug addition. The clear increase in the growth rate observed from cycle 8 onwards reflects adaptation of the parasite population to the drug. Cultures were regularly synchronized with sorbitol or L-Proline. At each cycle, after determining parasitaemia, cultures were diluted with fresh erythrocytes to a parasitaemia of approximately 1%. Growth rate was determined at each cycle by dividing the parasitaemia (measured at the ring stage) by the parasitaemia at the previous cycle (after diluting the culture). Parasitaemia was determined by flow cytometry after staining parasite nuclei with SYTO11.

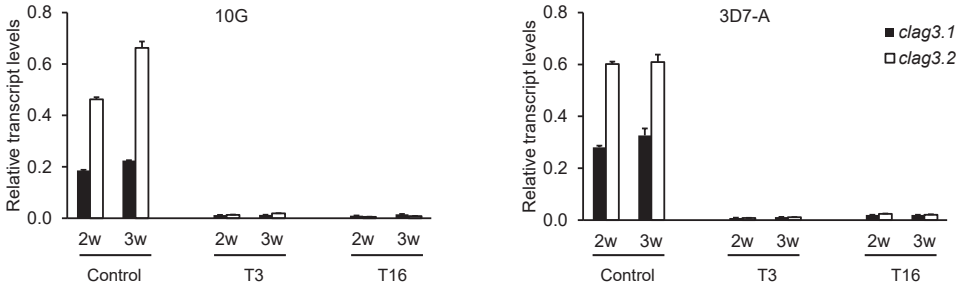


Fig. S2. *clag3* expression in cultures selected with T3 or T16. Transcript levels of *clag3.1* and *clag3.2* (normalized against *rhoph2*) were measured in 10G and 3D7-A cultures selected with T3 (2.86 nM) or T16 (0.6 nM) and unselected cultures maintained in parallel (Control), at the times indicated in weeks (w). Values are the average of reactions performed in triplicate, with SD. These results show silencing of both *clag3* genes after only 2 weeks of T3 or T16 selection.

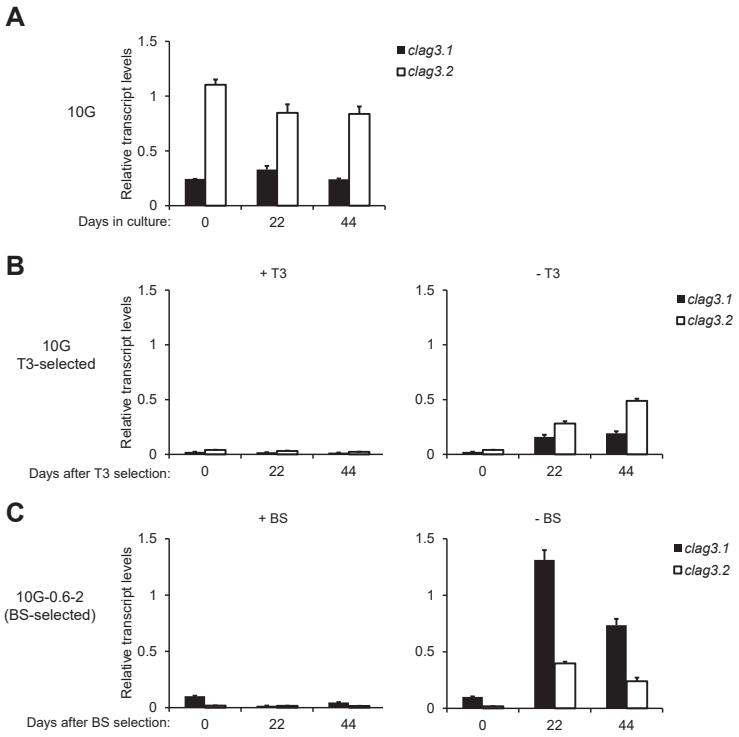


Fig. S3. *clag3* expression in cultures selected with T3 or blasticidin S and then maintained either in the presence or absence of the drug. Relative transcript levels of *clag3.1* and *clag3.2* (normalized against *rhoph2*) in control 10G cultures (A), or in cultures previously selected with T3 (B) or blasticidin S (C). Cultures were maintained in parallel either in the presence of the selecting drug (+T3 and +BS bar charts) or after removing it (- T3 and - BS bar charts). Day 0 is the day at which previously selected cultures were split and maintained under either continued presence or absence of the drug. Values are the average of three technical quantitative PCR replicates, with SD; independent biological samples for transcriptional analysis collected at different times are presented as separate data points (22 and 44 days after splitting the culture). These results show that silencing of *clag3* genes associated with resistance to the two drugs is reversible: in the absence of drug, parasites that re-activate *clag3* expression are progressively selected because simultaneous silencing of the two *clag3* genes poses a fitness cost. However, in blasticidin S-selected cultures, *clag3.1*-expressing parasites were predominant after drug removal, consistent with the previously reported partial protection against this drug associated with expression of this *clag3* paralog. This was not observed in T3-selected cultures, which maintained a *clag3.1*/*clag3.2* transcripts ratio similar to the parental 10G cultures.

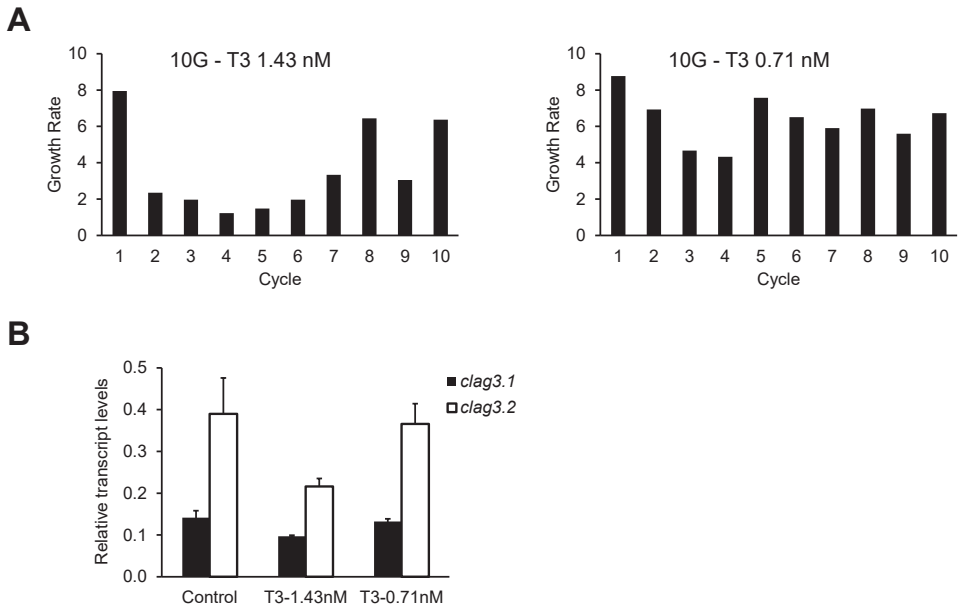


Fig. S4. Selection of parasites with low concentrations of T3. Cultures were selected with 1.43 or 0.71 nM T3 (corresponding to 1/2 or 1/4 of the concentration used in the initial selection experiments shown in Fig. 3) to determine whether adaptation to low concentrations of the drug can occur by selection of parasites that express one or the other *clag3* gene. (A) Growth dynamics during selection with lower concentrations of T3, as in Fig. S1. (B) Relative transcript levels of *clag3.1* and *clag3.2* (normalized against *rhoph2*) in control cultures and in cultures selected for eight cycles (~17 days) with T3 at the concentrations indicated. No major alteration in the relative expression of the two genes compared to control cultures maintained in parallel was observed. Values are the average of two biological replicates, with SD.

Article 2: Expression patterns of *Plasmodium falciparum* clonally variant genes at the onset of a blood infection in malaria-naïve volunteers.

Pickford AK, Michel-Todó L, Dupuy F,Mayor A, Alonso PL, Lavazec C & Cortés A. Expression patterns of *Plasmodium falciparum* clonally variant genes at the onset of a blood infection in malaria-naive humans. *mBio* 12: e01636-21. (2021) https://doi.org/10.1128/mBio.01636-21.

AMERICAN SOCIETY FOR MICROBIOLOGY | mBio®

RESEARCH ARTICLE

Expression Patterns of *Plasmodium falciparum* Clonally Variant Genes at the Onset of a Blood Infection in Malaria-Naive Humans

Anastasia K. Pickford,[a] Lucas Michel-Todó,[a] Florian Dupuy,[b] Alfredo Mayor,[a,c,d] Pedro L. Alonso,[a,c*] Catherine Lavazec,[b] Alfred Cortés[a,e]

[a]ISGlobal, Hospital Clínic-Universitat de Barcelona, Barcelona, Catalonia, Spain
[b]INSERM U1016, Centre National de la Recherche Scientifique (CNRS) Unité Mixte de Recherche (UMR) 8104, Université de Paris, Institut Cochin, Paris, France
[c]Centro de Investigação em Saúde de Manhiça, Manhiça, Mozambique
[d]Consorcio de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain
[e]ICREA, Barcelona, Catalonia, Spain

ABSTRACT Clonally variant genes (CVGs) play fundamental roles in the adaptation of *Plasmodium falciparum* to fluctuating conditions of the human host. However, their expression patterns under the natural conditions of the blood circulation have been characterized in detail for only a few specific gene families. Here, we provide a detailed characterization of the complete *P. falciparum* transcriptome across the full intraerythrocytic development cycle (IDC) at the onset of a blood infection in malaria-naive human volunteers. We found that the vast majority of transcriptional differences between parasites obtained from the volunteers and the parental parasite line maintained in culture occurred in CVGs. In particular, we observed a major increase in the transcript levels of most genes of the *pfmc-2tm* and *gbp* families and of specific genes of other families, such as *phist*, *hyp10*, *rif*, or *stevor*, in addition to previously reported changes in *var* and *clag3* gene expression. Increased transcript levels of individual *pfmc-2tm*, *rif*, and *stevor* genes involved activation in small subsets of parasites. Large transcriptional differences correlated with changes in the distribution of heterochromatin, confirming their epigenetic nature. Furthermore, the similar expression of several CVGs between parasites collected at different time points along the blood infection suggests that the epigenetic memory for multiple CVG families is lost during transmission stages, resulting in a reset of their transcriptional state. Finally, the CVG expression patterns observed in a volunteer likely infected by a single sporozoite suggest that new epigenetic patterns are established during liver stages.

IMPORTANCE The ability of malaria parasites to adapt to changes in the human blood environment, where they produce long-term infection associated with clinical symptoms, is fundamental for their survival. CVGs, regulated at the epigenetic level, play a major role in this adaptive process, as changes in the expression of these genes result in alterations in the antigenic and functional properties of the parasites. However, how these genes are expressed under the natural conditions of the human circulation and how their expression is affected by passage through transmission stages are not well understood. Here, we provide a comprehensive characterization of the expression patterns of these genes at the onset of human blood infections, which reveals major differences with *in vitro*-cultured parasites. We also show that, during transmission stages, the previous expression patterns for many CVG families are lost, and new patterns are established.

KEYWORDS *Plasmodium falciparum*, chromatin, clonally variant genes, controlled human malaria infection, epigenetics, heterochromatin, malaria, transcription

Citation Pickford AK, Michel-Todó L, Dupuy F, Mayor A, Alonso PL, Lavazec C, Cortés A. 2021. Expression patterns of *Plasmodium falciparum* clonally variant genes at the onset of a blood infection in malaria-naive humans. mBio 12: e01636-21. https://doi.org/10.1128/mBio.01636-21.

Editor Thomas E. Wellems, National Institute of Allergy and Infectious Diseases

Address correspondence to Alfred Cortés, alfred.cortes@isglobal.org.

* Present address: Pedro L. Alonso, Global Malaria Program, World Health Organization, Geneva, Switzerland.

Received 8 June 2021
Accepted 2 July 2021
Published 3 August 2021

Plasmodium spp. are responsible for the globally important disease malaria, which causes over 200 million clinical cases and almost half a million deaths per year (1). The complex life cycle of malaria parasites is split between different hosts and cell types. The invertebrate mosquito host injects the infective parasite forms known as sporozoites into the vertebrate host during a blood meal. Sporozoites travel to the liver and multiply asexually within hepatocytes, generating merozoites that are then liberated into the bloodstream, where the intraerythrocytic development cycle (IDC) begins. This cycle consists of erythrocyte invasion followed by asexual multiplication, including the ring, trophozoite, and multinucleated schizont stages, and release of new merozoites. Repeated rounds of the IDC enable parasites to rapidly increase their biomass and establish an enduring infection in the vertebrate host. However, some parasites convert into sexual forms termed gametocytes and abandon the IDC. Mature male and female gametocytes are infective to mosquitoes. Mating occurs within the mosquito midgut, and after several complex development steps, new sporozoites are generated, closing the cycle (2).

Life cycle progression in *Plasmodium* spp. is controlled mainly at the transcriptional level such that each stage is characterized by a specific gene expression program (3–5). This enables parasites to thrive in the multiple different environments they are exposed to, which entail dramatic differences in conditions such as temperature, pH, and nutrient availability. However, in addition to environmental diversity associated with life cycle progression, malaria parasites also have to confront fluctuating conditions within the same niche. These fluctuations, which can occur between individual hosts of the same species or even within the course of a single blood infection, may derive, for instance, from changes in the host's physiological or immunological state or from the effects of antimalarial treatment (6, 7). Adaptation to such diverse conditions occurs through various genetic and nongenetic mechanisms. While in malaria parasites genetic changes play a major role in species evolution and long-term adaptation to new conditions, as in any other organism, rapid adaptation to conditions that fluctuate frequently requires reversible, dynamic mechanisms that provide phenotypic plasticity (i.e., alternative phenotypes from the same genome) (8). One such mechanism is clonally variant gene (CVG) expression, which refers to genes that can be found in a different state (active or silent) in different individual parasites with identical genomes and at the same stage of life cycle progression. While both states are heritable, CVGs undergo low-frequency stochastic switches between the active and silent states, which constantly generates transcriptional heterogeneity within parasite populations (9–11). Changes in the expression of these genes can result in phenotypic variation; therefore, when the conditions of the environment change, natural selection can operate upon this preexisting diversity and eliminate from the population parasites with CVG expression patterns that do not confer sufficient fitness under the new conditions. This is considered a bet-hedging adaptive strategy (12), a type of adaptive strategy commonly observed in many microbial species (13–15).

In *P. falciparum*, the most virulent human malaria parasite species, CVGs include gene families such as *var*, *rif*, *stevor*, *pfmc-2tm*, *hyp1* to *hyp17*, *phist*, and *surfin* linked to pathogenesis, antigenic variation, and host cell remodeling; *mspdbl2*, *eba140*, and *pfrh4* genes linked to erythrocyte invasion; *clag* genes involved in solute transport; *acs* and *acbp* families linked to acyl-CoA metabolism; and *pfap2-g*, the master regulator of sexual conversion (9–12), among others. Specific adaptive roles have been demonstrated for changes in the expression of *var* (16–18) and *clag3* (19–24) genes.

The active or silenced state of CVGs is regulated at the epigenetic level (25). These genes are mainly located in subtelomeric bistable chromatin domains, in which both the active (euchromatin) and the silenced (facultative heterochromatin) states can be stably transmitted for several generations of asexual growth. Spontaneous transitions between the two states underlie the transcriptional switches (8–11). For all CVGs analyzed so far, the silenced state is characterized by the posttranslational histone modification histone H3 lysine 9 trimethylation (H3K9me3) and heterochromatin protein 1 (HP1), whereas the

active state is associated with acetylation of H3K9 (H3K9ac). Transmission of these histone modifications through asexual replication constitutes the epigenetic memory for the transcriptional state of CVGs (8–11, 26–30). In *var* and *clag3* genes, the only *P. falciparum* genes that are known to show mutually exclusive expression (i.e., only one member of the family is active at a time in individual parasites) (31–35), this epigenetic memory is erased during transmission stages (here including gametocyte, mosquito, and liver stages) (23, 36–38).

The expression patterns of CVGs under culture conditions have been characterized for several specific gene families and also at a genome-wide scale (12, 32, 33, 39, 40). However, understanding how CVGs are expressed under the natural conditions of human infection is complicated by multiple factors, including common occurrence of polyclonal infections and genetic diversity among isolates, which mainly affects CVGs. The expression of only some specific CVG families such as *var*, *clag*, and genes involved in erythrocyte invasion has been characterized in some detail in natural infections (8, 16, 23, 41–43). This is an important gap of knowledge because the conditions of the environment influence the expression patterns of CVGs that prevail. Indeed, the limited data available suggest that the transcriptome of malaria parasites grown *in vitro* differs substantially from that observed *in vivo*, including differences in CVG expression (44–47).

Parasites obtained from controlled human malaria infection (CHMI) trials provide many of the advantages of both cultured parasites and parasites from natural human infections because they are exposed to "real" human host conditions, but they have a well-defined genetic background. Furthermore, the conditions of the host and the time of infection are well controlled, which reduces the number of variables and facilitates the interpretation of the results. Therefore, CHMI trials provide a valuable system to study malaria parasite biology, including *in vivo* expression of CVGs. So far, the analysis of gene expression in CHMI samples has focused mainly on the *var* and *clag* families (23, 36–38, 48–50). Two more recent studies analyzed the transcriptome of parasites obtained from CHMI volunteers at a genome-wide level. However, these studies only included ring-stage parasites, precluding the characterization of transcriptional patterns for genes expressed at other stages of the IDC (51, 52).

To provide a complete view of *P. falciparum* CVG expression patterns during the initial phase of a blood infection in malaria-naive humans, we performed a genome-wide transcriptomic comparison across the full IDC between parasites obtained from volunteers participating in a CHMI trial and the parental line maintained in culture. With this controlled approach, we identified transcriptional differences between parasites growing under *in vitro* culture conditions or growing in the human circulation after passage through transmission stages. To confirm the epigenetic nature of the differences observed, we mapped the genome-wide distribution of heterochromatin. We also tested the hypothesis that CVGs other than the mutually exclusively expressed *var* and *clag3* genes undergo an epigenetic reset during transmission stages.

RESULTS

Transcriptomic comparison between parasites obtained from CHMI volunteers and the parental line reveals changes in CVG expression. We performed a time course genome-wide transcriptomic analysis across the full IDC of *P. falciparum* parasites obtained from a CHMI trial in which cryopreserved Sanaria NF54 sporozoites were injected into naive human volunteers (53). Parasites were cryopreserved on day 9 after infection and upon microscopy diagnosis on days 11 to 14. Transcriptomic analysis was performed using parasites collected from four different volunteers (V18, V35, V48, and V63 lines, here collectively termed vNF54) on the day of diagnosis. Parasites were thawed and cultured for the minimum number of cycles needed to obtain sufficient material (4 replication cycles) and then tightly synchronized (involving an additional cycle of replication) before harvesting RNA at defined time points of the IDC (10 to 15, 20 to 25, 30 to 35, and 40 to 45 h postinvasion [hpi]). In parallel, we obtained RNA at the same time points from two independent biological replicates of tightly synchronized cultures of the parental (premosquito) NF54 line (pNF54) (Fig. 1A). Transcript levels were determined

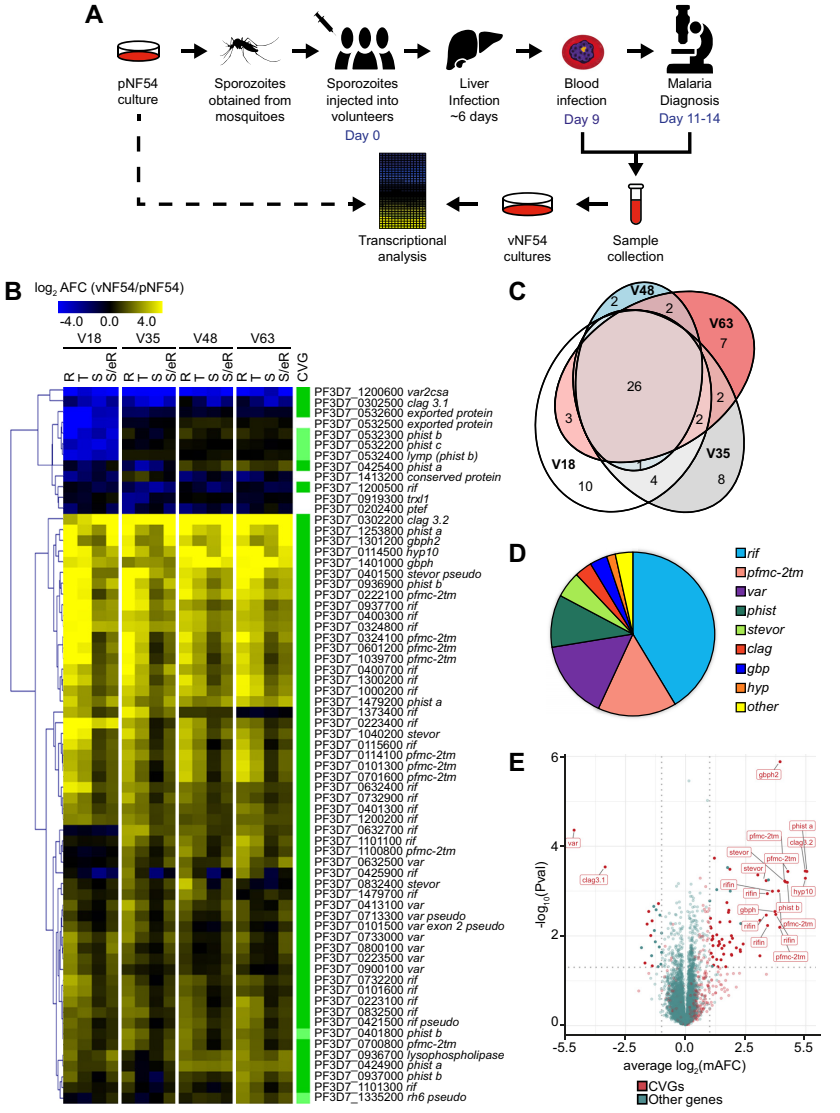


FIG 1 Genes differentially expressed between vNF54 and pNF54 lines. (A) Schematic of the study design comparing transcript levels between vNF54 lines obtained from volunteers participating in a controlled human malaria infection (CHMI) trial and the parental (premosquito) pNF54 line. (B) Heatmap of genes differentially expressed between vNF54 and pNF54 lines, ordered by hierarchical clustering. V18, V35, V48, and V63 are vNF54 lines obtained from different volunteers. Values are the $\log_2$ of the average fold change (AFC) of vNF54 versus pNF54 lines over four overlapping time intervals corresponding to the stages indicated (R, rings; T, trophozoites; S, schizonts; S/eR, schizonts and early rings). Only genes with an absolute value of the $\log_2$ of the maximum AFC (mAFC) of >2 (relative to pNF54) in at least one of the vNF54 lines are shown (see Materials and Methods for exclusion criteria). The

(Continued on next page)

using two-channel, long-oligonucleotide microarrays in which samples were hybridized against a common reference pool to obtain relative expression values (Cy5/Cy3). To quantify transcript-level differences between vNF54 and pNF54 lines, we calculated, for each gene, the maximum average fold change among overlapping time intervals of half the duration of the IDC (mAFC) (12) (see Materials and Methods).

There were 67 genes with a high-confidence mAFC of >4 (see Materials and Methods) in either direction [absolute value of the $\log_2$(mAFC) > 2] between at least 1 of the vNF54 lines and the pNF54 line, 21 of which had a mAFC of >16 (Fig. 1B and Data Set S1 in the supplemental material). The vast majority of differentially expressed genes had been previously identified as CVGs (Fig. 1B), either based on variant expression between isogenic lines (12) or because they carried the H3K9me3 or HP1 heterochromatin marks (54–57) (Data Set S2). Overall, the transcriptional changes (relative to pNF54) observed in the four vNF54 lines were highly similar (Fig. 1B), with the exception of a few genes showing a different pattern only in the V18 line. As a consequence, there was a large overlap in the genes differentially expressed (relative to pNF54) in each of the four vNF54 lines (Fig. 1C). Closer analysis revealed only a small number of genes with large transcript-level differences among the four vNF54 lines, and many of them showed a different pattern only in the V18 line, including a cluster of five downregulated neighbor genes in the distal subtelomeric region of chromosome 5 (Fig. S1).

The majority of genes differentially expressed in all vNF54 lines compared to pNF54 were expressed at higher levels in vNF54 lines, but a small number of genes were expressed at lower levels. The two most downregulated genes (mAFC > 10) were *var2csa* and *clag3.1*. The most upregulated genes in vNF54 lines (mAFC > 32) were a *hyp10* gene encoding an exported protein (PF3D7_0114500) (58), a *phist-a* gene (PF3D7_1253800), *clag3.2* (PF3D7_0302200), and a *pfmc-2tm* gene (PF3D7_0324100). Overall, the majority of genes that showed changes in expression between pNF54 and vNF54 lines belong to the large *pfmc-2tm*, *rif*, *var*, *stevor*, and *phist* CVG families encoding exported proteins (16–18, 58, 59) or to the smaller families *clag*, involved in solute transport (20), and glycophorin binding protein (*gbp*), encoding exported proteins of unknown function (58, 60, 61) (Fig. 1D). Given the overall similarity between the different vNF54 lines, we performed an additional analysis in which the different vNF54 lines were treated as replicates, which enables statistical analysis of the expression differences observed. The genes identified as most differentially expressed in this analysis were roughly the same as in the analysis of each vNF54 line separately, and *gpbh2* (PF3D7_1301200) was the most significantly upregulated gene in vNF54 lines (Fig. 1E and Data Set S1).

Transcriptional changes in specific CVG families. To analyze the expression patterns of CVGs within specific gene families, in addition to the transcript levels relative to a common reference pool (Cy5/Cy3 values), we also used the sample signal (Cy5 channel). The two-channel microarray approach used here is designed for hybridization against a common reference pool, which enables robust comparison of relative transcript levels between samples (3, 12, 62). However, the sample signal alone provides a semiquantitative estimate of the expression intensity of the genes, and it enables the identification of the predominantly expressed members of specific families (12), rather than informing only about the relative levels between samples.

Analysis of the transcriptional differences between pNF54 and vNF54 lines in specific CVG families revealed different scenarios for different families (Fig. 2A and B and

FIG 1 Legend (Continued)
column at the right indicates whether a gene was previously classified as a CVG (dark green; see Data Set S2) or belongs to a gene family in which other genes are CVGs (light green). Twelve genes had values out of the color scale range displayed. (C) Euler diagram showing the overlap between genes with an absolute value of the $\log_2$(mAFC) >2 (relative to pNF54) between the different volunteers. (D) Pie chart showing the distribution of genes with an absolute value of the $\log_2$(mAFC) of >2 between different gene families. (E) Volcano plot representing expression differences between vNF54 and pNF54 lines. Expression fold change values for each gene are the average of the $\log_2$(mAFC) in the four vNF54 lines. *P* values were calculated using an unpaired two-sided *t* test. CVGs are shown in red and other genes in green. Name labels are provided for genes with a mAFC of >10 and a *P* value of <0.01. Vertical dotted lines mark a $\log_2$(mAFC) of 1 or −1, whereas the horizontal dotted line marks a *P* value of 0.05.

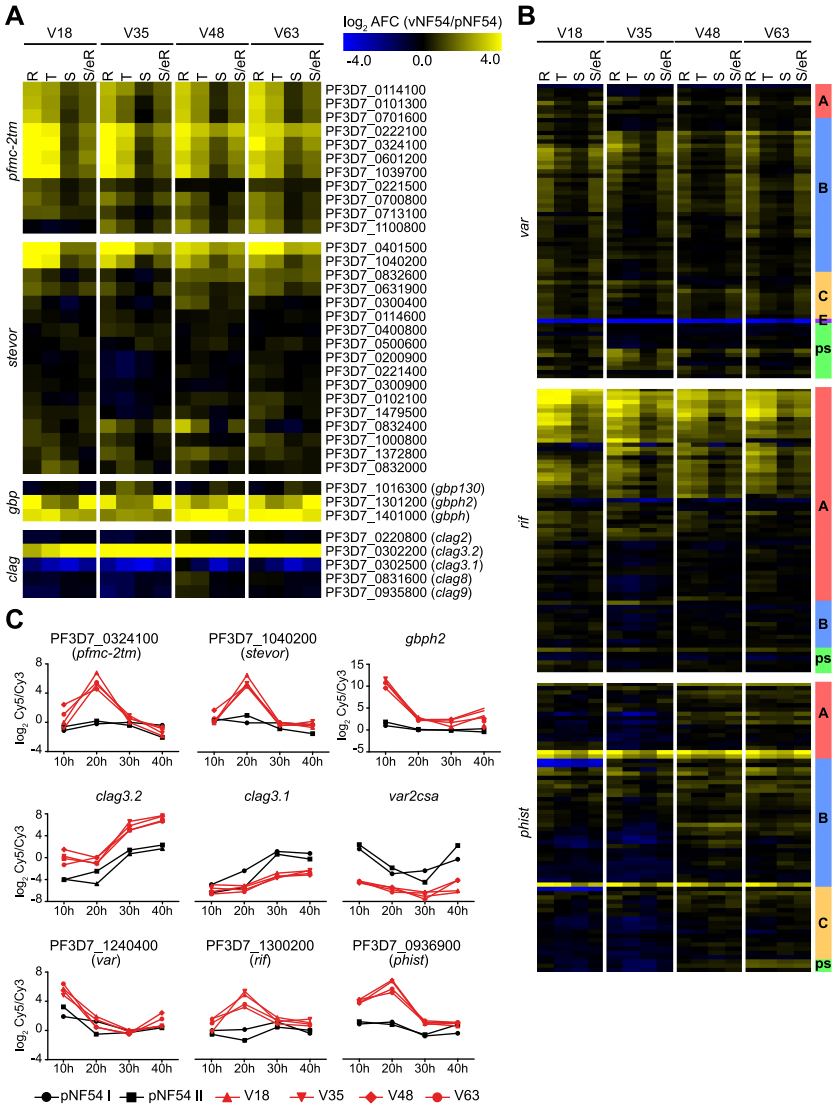


FIG 2 Transcriptional differences in specific CVG families. (A and B) Transcript-level changes in vNF54 lines relative to pNF54, as in Fig. 1B, for the CVG families indicated. See Materials and Methods for exclusion criteria. Eighteen genes (between the two panels) had values out of the color scale range displayed. In panel B, the subfamily of each gene (as annotated in PlasmoDB or in a previous study [12]) and pseudogenes (ps) are indicated in the column at the right. (C) Time course expression plots of selected genes of different CVG families. Values are the $\log_2$ of the sample versus reference pool ratio (Cy5/Cy3) in the vNF54 and pNF54 lines (*y* axis) and time after Percoll-sorbitol synchronization at which the sample was collected (*x* axis).

Fig. S2). Essentially all members of the *pfmc-2tm* family (39, 59, 63) were strongly upregulated in vNF54 lines (mAFC > 8 in all but three genes) (Fig. 2A and C), and this was confirmed by analyzing the Cy5 signal only, which revealed a large increase in the expression of the majority of *pfmc-2tm* genes (Fig. S2A). The analysis of a specific PfMC-2TM protein for which antibodies were available also revealed clearly increased abundance in a vNF54 line compared to pNF54 (Fig. S2B). In contrast, in the *var*, *rif*, *phist*, and *stevor* families, several specific genes were upregulated in vNF54 lines, but many other genes did not change, and, in some families, a few were downregulated. In the case of the mutually exclusively expressed *var* genes, vNF54 lines showed upregulation of many *var* genes, especially of type B, whereas the type E *var2csa* gene was strongly downregulated (Fig. 2B and C). Analysis of the Cy5 values revealed that *var2csa* was the predominantly expressed *var* gene in pNF54, whereas in vNF54 lines, multiple *var* genes were expressed at intermediate levels, similar to findings from previous CHMI studies (36, 38, 48, 52) (Fig. S2C). Considering that individual parasites express a single *var* gene, this result indicates that the parental population was relatively homogeneous, such that the vast majority of individual parasites expressed the same *var* gene (*var2csa*), whereas after passage through transmission stages the population became heterogeneous, with different individual parasites expressing different *var* genes. Some of the most highly expressed *var* genes in vNF54 lines coincided with the most highly expressed genes in a recent CHMI study using the 3D7 line to infect volunteers (52) (Fig. S3A), and preferential upregulation of type B *var* genes in parasites from the volunteers is also consistent with previous reports from other CHMI studies (36, 48, 52).

In the *rif* family (18, 64–67), a large subset of genes encoding type A RIFINS were upregulated in all vNF54 lines, whereas no general differences were observed in genes encoding type B RIFINS. In the *stevor* and *phist* families (18, 58, 68), only a small subset of specific genes showed strong upregulation in vNF54 lines, essentially the same genes in all four lines (Fig. 2A to C). In the *rif*, *stevor*, and *phist* families, which participate in antigenic variation and show expression switches but do not show mutually exclusive expression (39, 67, 69), the analysis of the Cy5 signal revealed that, in spite of changes in the expression of some specific genes, the predominantly expressed genes are similar between pNF54 and vNF54 lines (Fig. S2D to F). None of the genes that showed increased transcript levels in vNF54 lines became the dominantly expressed gene in any of these families. Of note, the most highly expressed *rif* gene in our data set was the same as in a previous CHMI study (52) (Fig. S3A).

Two small CVG families, *gbp* (60) and *clag* (20), showed major changes in the expression of a large proportion of their genes. Two of the three members of the *gbp* family, *gbph* and *gbph2*, were among the most highly upregulated genes (mAFC > 16) in vNF54 lines (Fig. 2A and C and Fig. S2G). In all vNF54 lines, there was increased expression of *clag3.2* and reduced expression of *clag3.1*, consistent with our previous reverse transcriptase quantitative PCR (RT-qPCR) results showing that, in the pNF54 population, essentially all parasites express *clag3.1*, whereas in vNF54 lines, there is a mixture of some parasites expressing *clag3.1* and a majority of parasites expressing *clag3.2* (23). There was no major change in the expression of the other CVG of the *clag* family, *clag2* (32) (Fig. 2A and C and Fig. S2H).

The expression of other CVG families such as acyl-CoA synthetase (*acs*), acyl-CoA binding protein (*acbp*), lysophospholipase, exported protein kinase (*fikk*), *hyp1* to *hyp 17*, *surfin*, and families linked to erythrocyte invasion (*eba*, *pfrh*) was almost identical between pNF54 and vNF54 lines, with the exception of a *hyp10* (PF3D7_0114500) and a lysophospholipase gene (PF3D7_0936700) (Fig. S4A).

Most changes in CVG expression between vNF54 and pNF54 lines are determined at the epigenetic level. The observation that the majority of genes differentially expressed between pNF54 and vNF54 lines are CVGs, which have been previously shown to be regulated by truly epigenetic mechanisms (25), strongly suggests that parasites recovered from the volunteers mainly differ from the parental line in their epigenetic makeup, rather than at the genetic level. To confirm this view, we sequenced the whole genome of the pNF54 and two vNF54 lines (V18 and V63) and

found no genetic differences in either coding or noncoding regions that are likely to explain any of the transcriptional changes common to all vNF54 lines (Data Set S3). However, we observed one duplicated region and several large subtelomeric deletions in V18 that explain the reduced expression of several genes specifically in this parasite line, including the cluster of neighbor downregulated genes in the distal subtelomeric region of chromosome 5 (Data Set S3). Unexpectedly, in the V18 line, there were essentially no reads mapping to the deleted regions, indicating that these large deletions occur with a 100% prevalence. Parasites in the parental pNF54 line (as in any other isolate maintained in culture for a long time) are expected to be a mixture of individual parasites that have accumulated different mutations (70). Therefore, the presence of deletions with a 100% prevalence in the V18 line suggests that this line is genetically homogeneous because it originated from a single sporozoite infecting the liver. Consistent with this view, the majority of differential single nucleotide polymorphisms (SNPs) and small indels identified among the pNF54, V63, and V18 lines were unique to V18 and had a 100% prevalence in this line (Data Set S3). Of note, the volunteer from which the V18 line originated was part of a group of the CHMI trial that was inoculated under a scheme that resulted in blood infection in only one out of six volunteers (group 2, 2,500 *P. falciparum* sporozites [PfSPZ] in 50 μl, intramuscular injection), indicating that very few sporozoites were viable. In contrast, the other parasite lines analyzed (V35, V48, and V63) are from volunteers that received direct venous inoculation resulting in infection of 100% of the volunteers (53). Altogether, these results indicate that the V18 line is genetically homogeneous for the mutations because it likely originated from infection by a single sporozoite.

Next, we performed comparative H3K9me3 chromatin immunoprecipitation followed by sequencing (ChIP-seq) analysis to determine the distribution of heterochromatin in one of the vNF54 lines (V63) and pNF54. In the genes showing the largest transcript-level differences between the two lines, higher expression was associated with reduced levels of heterochromatin in the upstream region and beginning of the coding sequence, whereas lower expression was associated with increased levels of heterochromatin (Fig. 3A and C and Data Set S4). This pattern was observed in 7 out of 8 genes with a mAFC of $>$16 between V63 and pNF54, and it occurred in genes from multiple families, including *stevor*, *gbp*, *clag*, *var*, *phist*, *hyp10*, and lysophospholipase. However, there were no apparent differences in heterochromatin prevalence between V63 and pNF54 at the putative regulatory regions of many differentially expressed genes from the large CVG families *pfmc-2tm*, *rif*, *var*, and *stevor* (Fig. 3B and C). The most plausible interpretation for this result is that individual genes of these families were activated in only small subsets of the parasites in V63, accounting for the increased transcript levels observed, but they remained silenced and in a heterochromatic state in the majority of parasites. Indeed, the analysis of a collection of subclones from the vNF54 line V33 (23), which also expressed *pfmc-2tm* genes at much higher levels than pNF54 (Fig. 4A), revealed that individual *pfmc-2tm* genes were silenced in the majority of subclones and expressed at very high levels in others (Fig. 4B). A similar scenario was observed for *rif* and *stevor* genes (Fig. S5). The expression patterns in recent subclones reflect the expression in individual parasites (12). Thus, these results indicate that individual *pfmc-2tm*, *rif*, and *stevor* genes were activated in only small subsets of parasites in V33 and also demonstrate population heterogeneity for the expression of these genes in a vNF54 line. For *var* genes, it has been previously established that in a population with a broad expression pattern, each individual gene is silenced in the majority of individual parasites (31, 36, 37). Together, these results indicate that, in the CVGs for which an association between increased transcript levels and lower levels of heterochromatin was not observed, this was explained by activation of the genes in only a small fraction of the parasites.

Changes in the expression of many CVGs are determined by a reset of epigenetic patterns during transmission stages. For *var* and *clag3* genes, an epigenetic reset during transmission stages results in transcriptional heterogeneity in the parasite population at the onset of a blood infection (23, 36–38, 48, 50). To determine if the epigenetic

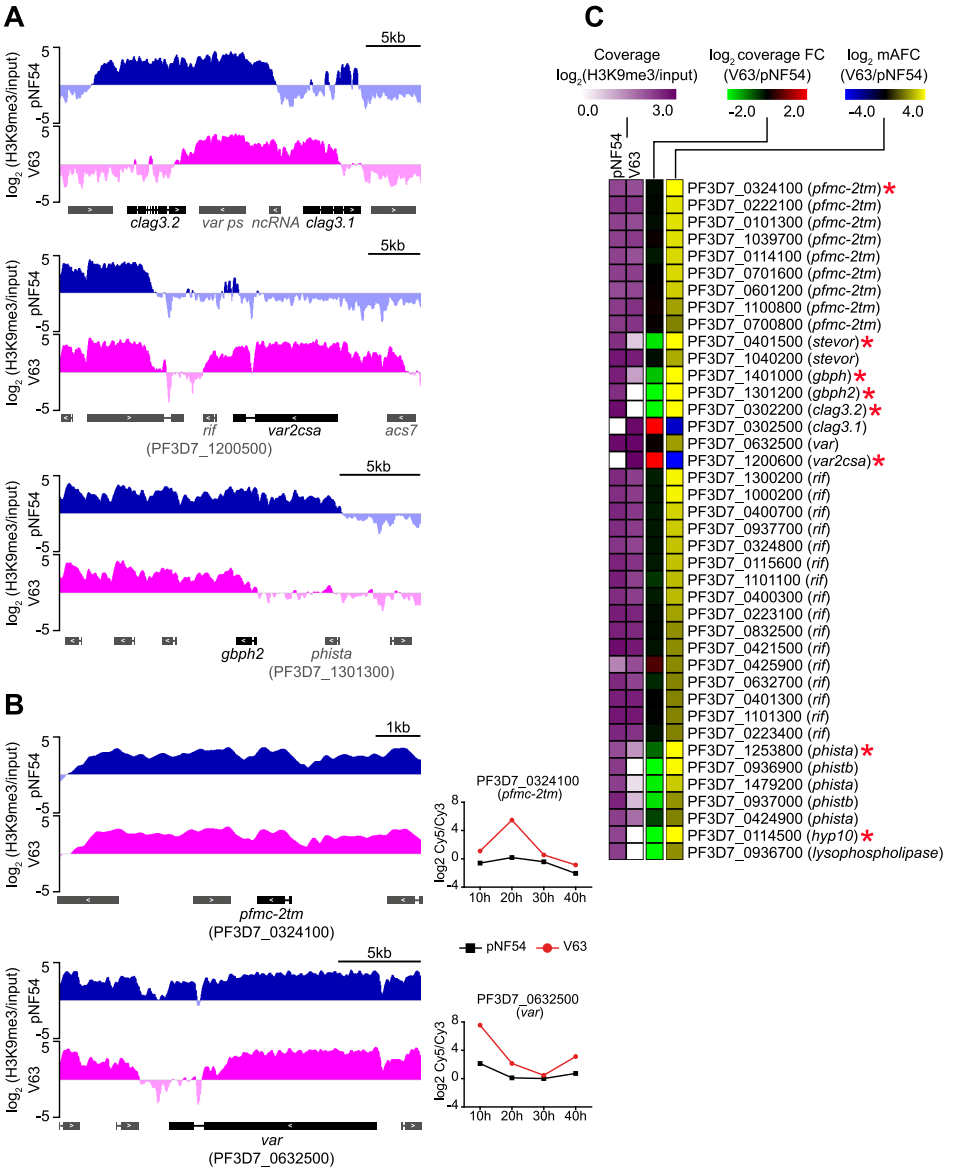


FIG 3 ChIP-seq analysis of vNF54 and pNF54 lines. (A and B) Distribution of normalized H3K9me3 signal relative to input in the pNF54 line and the vNF54 line V63. Representative genes are shown in which transcriptional changes between pNF54 and V63 are (*clag3.1*, *clag3.2*, *var2csa*, and *gbph2*) (A) or are not (*pfmc-2tm* and *var*) (B) accompanied by differences in heterochromatin distribution at their upstream regions. Genes are shown as rectangles (introns displayed as thin lines), with the direction of transcription indicated by arrowheads. The names of neighbor genes (in gray) with different levels

(Continued on next page)

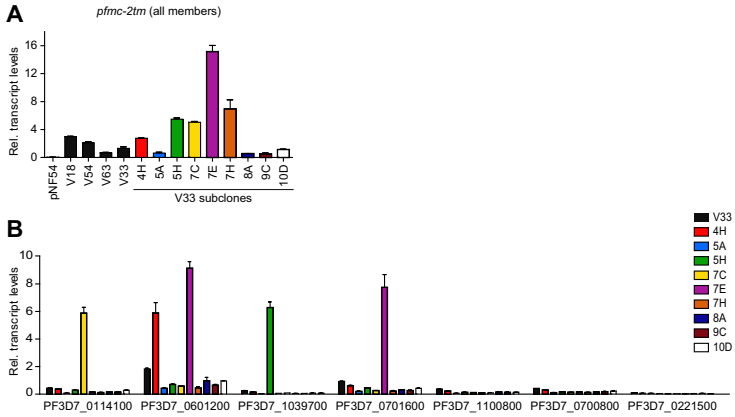


FIG 4 Expression of *pfmc-2tm* genes in subclones of a vNF54 line. (A) Transcript levels of *pfmc-2tm* genes (all members, analyzed with primers that amplify all the genes of the family) in pNF54, different vNF54 lines, and subclones of the V33 line at 20 to 25 h postinvasion (hpi). (B) Transcript levels of selected individual *pfmc-2tm* genes in V33 and V33 subclones. Transcript levels are normalized against serine-tRNA ligase (*serrs*). Data are presented as the average and SEM of two independent biological replicates.

memory for the transcriptional state of other CVGs is also lost during transmission stages, we compared the expression of *gbph2*, *pfmc-2tm*, and *stevor* genes between samples collected from the volunteers on day 9 or the day of microscopy diagnosis (for the samples used, day 11 or 14). Since liver stage development lasts ~6 days (36, 48, 49, 71), day 9 samples correspond to parasites that multiplied in the human circulation for only one round of the IDC after egress from the liver (second-generation blood stages), whereas day 11 or 14 samples are from parasites that replicated for 1 or 3 additional cycles, respectively (third- or fifth-generation blood stages), according to previous estimations (49). We reasoned that if only within-host selection during the IDC was responsible for the differences between pNF54 and vNF54 lines and there was no epigenetic reset during transmission stages, day 9 samples would show expression levels intermediate between pNF54 and day 11 or 14 samples, as parasites with unfavorable expression patterns would be eliminated progressively. In contrast, if an epigenetic reset is the main determinant of the expression patterns observed, similar transcript levels would be expected between day 9 and day 11 or 14 samples (23).

Because the parasitemia of day 9 samples was very low, obtaining sufficient material for RT-qPCR analysis required ~7 cycles of *in vitro* growth, whereas only ~4 cycles were needed for day 11 or 14 samples. Therefore, we first determined whether extended culture affected the expression levels of these genes. After 5 weeks in culture, transcript levels of *gbph2*, one *pfmc-2tm* gene, one *stevor* gene, and total *pfmc-2tm* family transcripts remained stable in both pNF54 and a vNF54 line (V63) (Fig. 5A). This result indicates that growth under *in vitro* conditions does not rapidly alter the expression of these genes and does not represent a confounding factor for the comparison of day 9 with day 11 or 14 samples. These experiments also confirmed the much higher transcript levels for these genes in vNF54 than in pNF54, and they revealed that they are virtually silenced in the latter.

FIG 3 Legend (Continued)
of H3K9me3 between pNF54 and V63 are shown. The time course expression plot for the genes in panel B, not included in Fig. 2C, is shown. These two genes had the largest expression fold increase in V63 relative to pNF54 in the entire *pfmc-2tm* and *var* families. (C) H3K9me3 ChIP-seq coverage (from −1,000 bp or closest upstream gene to +500 bp from the ATG), coverage fold change (FC) between V63 and pNF54, and mAFC between V63 and pNF54 for genes showing an absolute value of the $\log_2$(mAFC) of >2 between V63 and pNF54. Genes with an absolute value of the $\log_2$(mAFC) >4 (values out of the color scale range displayed) are marked with an asterisk.

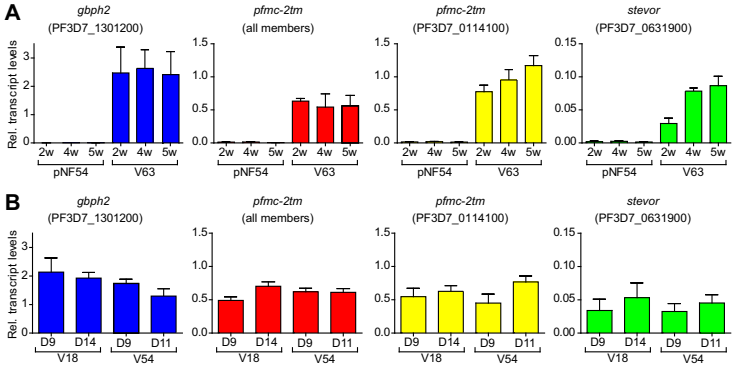


FIG 5 Changes in transcript levels associated with time in culture or time in the human circulation. (A) Effect of duration of growth under culture conditions on the expression of selected CVGs with different transcript levels between V63 (a vNF54 line) and pNF54. Relative transcript levels were determined by RT-qPCR after different times in culture, up to 5 weeks (w), in the pNF54 and V63 lines. RNA for transcriptional analysis was collected at the ring stage (10 to 15 h postinvasion [hpi]) for the *gbph2* gene and at the early trophozoite stage (20 to 25 hpi) for *pfmc-2tm* and *stevor* genes. For "*pfmc-2tm* (all members)", primers that amplify all genes of the family were used. (B) Comparison of relative transcript levels between parasites collected at day 9 postinfection or at days 14 (V18) or 11 (V54) (day of microscopy diagnosis), determined as in the previous panel. Transcript levels are normalized against serine-tRNA ligase (*serrs*). Data are presented as the average and SEM of three independent biological replicates.

Next, we compared the transcript levels of these genes between parasites collected on day 9 or the day of microscopy diagnosis in two vNF54 lines (V18 and V54, with microscopy diagnosis on days 14 and 11, respectively), which revealed no large differences (Fig. 5B). These results indicate that multiplication for a small number of rounds of the IDC in the blood of a naive human host does not have a major impact on the expression of *gbph2*, *pfmc-2tm*, and *stevor* genes. This is consistent with the idea that their increased expression in vNF54 lines compared to pNF54 was established before parasites reached the blood circulation, i.e., during transmission stages, as a consequence of an epigenetic reset. The results from a recent CHMI study in which blood-stage parasites from a volunteer infected via mosquito bite were blood passaged to new volunteers also support this idea (52, 72). Transcriptional comparison of parasites collected before and after blood growth in the new volunteers (three to five cycles), analogous to our day 9 versus day 11 or 14 comparison, revealed no difference in the expression of *var* or *rif* genes (52). We analyzed, in that data set, the expression of other CVGs and also observed no consistent changes (Fig. S3B). The majority of differences between samples collected before and after blood growth in the new volunteers occurred in genes that are not CVGs (Fig. S3C).

Phenotypic comparison of pNF54 and vNF54 lines. We performed exploratory experiments to determine if the transcriptional differences observed between pNF54 and vNF54 lines result in measurable functional differences. We focused on phenotypes that are mainly determined by CVGs and for which assays were available in our laboratories. Since many of the genes showing differential expression are exported to the erythrocyte cytoplasm or membrane (18, 58, 59) and some directly impact the mechanical properties of infected erythrocytes (i.e., *stevor*) (73), we compared the membrane deformability at the trophozoite stage of pNF54- and vNF54 (V63)-infected erythrocytes using a microsphiltration assay. While V63 showed a higher retention rate, indicative of lower deformability, the difference was not significant (Fig. 6A).

Since we observed changes in the expression of *clag3* genes, which determine solute uptake in infected erythrocytes (20), we compared sorbitol permeability between pNF54 and vNF54 lines. Sorbitol uptake resulting in red cell lysis was observed at

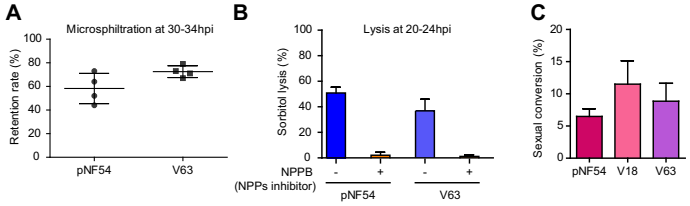


FIG 6 Phenotypic comparison of vNF54 and pNF54 lines. (A) Retention in microsphiltration assays of pNF54 and a vNF54 line (V63). Data are presented as the average, SEM, and individual data points of four independent biological replicates. (B) Sorbitol lysis assays for pNF54 and V63 in the absence or presence of NPPB (an inhibitor of new permeation pathways [NPPs]). Data are presented as the average and SEM of three independent biological replicates. (C) Sexual conversion rate (proportion of parasites that convert into sexual forms) of pNF54 and two vNF54 lines (V18 and V63). Data are presented as the average and SEM of three independent biological replicates. No significant difference ($P < 0.05$) was observed between pNF54 and vNF54 lines in any of the panels, using an unpaired two-sided *t* test.

similar levels in both lines (Fig. 6B) as expected, given that only the simultaneous silencing of the two *clag3* genes, which did not occur in either pNF54 or vNF54, is known to prevent sorbitol uptake (22). Lastly, we compared the sexual conversion rates of pNF54 and vNF54 lines (V18 and V63), which revealed no significant differences (Fig. 6C). This is consistent with the similar transcript levels of *pfap2-g*, the master regulator of sexual conversion (74), and of early gametocyte markers between pNF54 and vNF54 lines (Fig. S4B).

DISCUSSION

Here, we provide the first genome-wide transcriptional characterization across the full IDC of *P. falciparum* parasites obtained during the initial days of a blood infection in malaria-naive humans and compare it with the transcriptome of parasites with the same genome but maintained under *in vitro* culture conditions. We also compared the genome-wide distribution of heterochromatin between parasites obtained from the infected humans or in culture. We found that the largest expression differences occur in CVGs and are associated with epigenetic changes in the distribution of heterochromatin, and we provide an accurate view of how parasites use their CVGs when they establish a new blood infection after egress from the liver. We also show that passage through transmission stages results in a reset of the epigenetic memory involving multiple CVG families.

The majority of changes between premosquito parasites and parasites obtained from the infected volunteers occurred in gene families involved in processes such as antigenic variation, erythrocyte remodeling, or solute transport, whereas other CVG families showed few alterations. Even within the same family, subgroups of genes with different predicted functions showed different expression patterns: type A RIFINS, predicted to encode proteins localized at the infected erythrocyte surface and involved in rosetting (18, 67), were generally upregulated in parasites collected from the volunteers, but not type B RIFINS with different predicted function and localization. The most dramatic changes were observed in the *pfmc-2tm*, *gbp*, *var*, and *clag* families, whereas other families, such as *rif*, *stevor*, or *phist*, showed important transcript-level differences for several specific genes, but the predominantly expressed genes and global expression levels of the family were not altered. Together, these observations suggest that some CVG families have high plasticity in their expression patterns, whereas others have clearly "preferred" patterns.

While the epigenetic state of CVGs is transmitted from one generation of asexual parasites to the next (11, 17, 25), it has been proposed that an "epigenetic reset" may occur during transmission stages such that the epigenetic memory for the transcriptional state of all CVGs is erased and new patterns of expressed and silenced CVGs are established stochastically (75). In the murine malaria parasite *Plasmodium chabaudi*,

mosquito-transmitted parasites were less virulent than those transmitted by blood passage and expressed a broader repertoire of CVGs (75, 76). This led some authors to postulate that repeated cycles of the IDC progressively select for parasites with CVG expression patterns associated with increased fitness, and passage through transmission stages results in a reset of the epigenetic patterns and lower virulence (75), although evidence for this model in human malaria is lacking. In *P. falciparum*, an epigenetic reset during transmission stages was previously demonstrated for *var* and *clag3* genes (23, 36–38). Here, we observed clearly distinct expression patterns between premosquito parasites and parasites obtained from the volunteers for genes from multiple additional CVG families (*gbp*, *pfmc-2tm*, *stevor*, *rif*, *phist*, lysophospholipase, and *hyp10*) and similar expression patterns between day 9 and day 11 or 14 samples in all the CVG families analyzed (*gbp*, *pfmc-2tm*, and *stevor*). Similarly, previous analyses of parasites from CHMI trials showed no consistent CVG expression changes associated with an increasing number of multiplication cycles in the blood of naive humans (36, 52). These results are consistent with a general epigenetic reset of CVGs during transmission stages, although it is also possible that in CVGs for which no transcriptional changes were observed between pNF54 and vNF54 lines, the epigenetic memory is maintained throughout all transmission stages.

We postulate that during transmission stages, the epigenetic memory for multiple CVG families is erased, and later on, new epigenetic patterns are established stochastically such that they differ among individual parasites. However, while the state of a gene in an individual parasite is established stochastically, the probability of the gene acquiring an active or a silenced state may be hardwired and dictated by the underlying DNA sequence. This probabilistic scenario could explain the similar CVG expression patterns observed among the different volunteers and also the relatively constant patterns observed in some gene families. For some CVGs, the active or the silenced state may be strongly favored. Given that in the V18 line, which appears to have originated from a single sporozoite invading the liver, we observed CVG expression patterns similar to the other vNF54 lines, including a broad expression pattern of *var* genes indicative of transcriptional heterogeneity, we postulate that new epigenetic patterns are established during liver development. The stage at which the previous epigenetic patterns are erased is not known, but given the profound heterochromatin remodeling already observed in gametocytes and mosquito stages, it is likely to occur before parasites reach the liver (55, 77, 78).

Of note, in many genes from large multigene families, we observed large transcript-level fold differences that were not accompanied by measurable changes in the levels of H3K9me3 occupancy. We found that in the parasite populations obtained from the volunteers, individual genes of these families were active in only a small subset of the parasites, which still could result in a large fold increase in transcript levels if fewer parasites expressed it in the premosquito population (e.g., a 10-fold increase if a gene was active in 10% of the parasites collected from the volunteers but only in 1% of the premosquito population). This would result in a minor difference in the heterochromatin levels (e.g., 90% versus 99% of the parasites would have the gene from the example above in a heterochromatic state), which is not detectable in a comparative ChIP-seq analysis. Therefore, our results do not indicate that activation of many *var*, *rif*, or *pfmc-2tm* genes occurred independently of heterochromatin changes but, rather, reflect that activation occurred in only a small fraction of the parasites in the population, and reductions in heterochromatin levels were undetectable. Indeed, in the parental line, the *var2csa* gene was in an active state in essentially all the parasites, and this was clearly associated with the absence of heterochromatin at this locus, consistent with previous reports for the epigenetic regulation of *var* genes (28, 30).

Altogether, a model emerges in which a blood infection in a new host starts with a transcriptionally heterogeneous parasite population, consisting of a mixture of individual parasites with different combinations of active and silenced CVGs from multiple gene families. This constitutes the basis of a bet-hedging adaptive strategy because

transcriptional heterogeneity results in phenotypic and antigenic diversity that increases the chances of survival of the parasite population in the unpredictable conditions of the blood of a new human host. Our data provide an accurate picture of how parasites use their CVGs in this initial phase of a blood infection, revealing that a higher proportion of individual parasites (compared to culture conditions) have in an active state the majority of genes of the *pfmc-2tm* and *gbp* families and specific genes of the *rif* (type A), *phist*, *stevor*, and a few other CVG families. Additionally, we demonstrate transcriptional heterogeneity for the expression of *rif*, *stevor*, and *pfmc-2tm* genes in addition to *var* and *clag3* genes (23, 36, 38, 48, 52). While in our study we could not characterize the long-term expression patterns of CVGs because infections were terminated as soon as parasites were detected by light microscopy or symptoms appeared, current knowledge suggests that during chronic blood infection, the expression of CVGs is governed by low-frequency switches and dynamic natural selection of parasites with expression patterns that confer high fitness as the conditions of the host fluctuate (9–11, 50). Future research will be needed to fully characterize the precise phenotypic and antigenic differences that result from specific changes in CVG expression.

MATERIALS AND METHODS

Human samples and parasite culture. The *P. falciparum* NF54 line at Sanaria (pNF54) and lines derived from human volunteers participating in a CHMI study with Sanaria NF54 sporozoites (vNF54 lines) (53) were cultured in B+ erythrocytes at 3% hematocrit in RPMI 1640-based culture medium supplemented with 10% human serum under continuous shaking at 100 rpm and in a 5% CO_2, 3% O_2, and 92% N_2 atmosphere.

Transcriptomic analysis using microarrays. Time course transcriptomic analysis was performed using cultures tightly synchronized to a 5-h age window, which was achieved by Percoll purification of schizonts followed by sorbitol lysis 5 h later (79). Sixty-milliliter cultures at 3% hematocrit and ~10% parasitemia were split into four flasks that were cultured undisturbed for different periods of time (10, 20, 30, or 40 h) before harvesting in TRIzol and freezing at −80°C. Rather than splitting into identical cultures, larger volumes (3- to 4-fold) were used for the early time points because young (ring-stage) parasites contain a smaller amount of RNA per parasite. Cultures for RNA collection at late stages were diluted to ~5% parasitemia with fresh erythrocytes (just after splitting) to prevent culture collapse.

RNA was purified using the TRIzol method, and cDNA was synthesized by reverse transcription (starting with 5 to 10 μg of RNA), purified, and labeled as previously described (62). Samples were analyzed using two-color long oligonucleotide-based custom Agilent microarrays. The microarray design was based on Agilent design AMADID 037237 (62), modified by adding new probes for genes lacking unique probes and for some ncRNAs and reporter genes (new designs, AMADID 084561 and AMADID 085763) and reannotated according to a BLAST analysis (80). All samples (200 to 500 ng), labeled with Cy5, were hybridized against an equal amount of a common reference pool labeled with Cy3, consisting of a mixture of equal amounts of cDNA from rings, trophozoites, and schizonts of pNF54. Microarray hybridization was performed as previously described (62). Images were acquired using a microarray scanner (G2505C; Agilent Technologies) located in a low-ozone hood.

Microarray data analysis. Initial processing of microarray data, including linear and locally weighted scatterplot smoothing (LOWESS) normalization, was performed using the Feature Extraction software (Agilent) with default options. The next steps of the analysis were performed using Bioconductor in an R environment (R version 3.5.3). For each individual sample and channel (Cy3 and Cy5), background signal was calculated as the median of the 100 lowest signal probes. Probes with both Cy3 and Cy5 signals below three times the array background in all samples were excluded from further analysis. Gene-level $\log_2$(Cy5/Cy3) values and statistical estimation of parasite age (81) were computed as previously described (12). For each gene, the $\log_2$(Cy5/Cy3) values were plotted against the statistically estimated culture age (in hpi), and the plots were divided into four overlapping time intervals of identical length that roughly corresponded to the ring, trophozoite, schizont, and late-schizont/early-ring stages (12). For each gene and time interval, the average expression fold change (AFC) between each vNF54 and its control pNF54 line (the replicate of pNF54 analyzed in parallel) was calculated from the difference in the area under the curve in the $\log_2$(Cy5/Cy3) versus estimated age plots. The maximum AFC (mAFC) was the value of the AFC in the time interval at which it had the highest absolute value. All tRNAs were excluded from further analysis because our method is not suitable for the analysis of tRNA expression (it tends to show large technical variability). For the identification of genes differentially expressed among pNF54 and vNF54 lines (i.e., Fig. 1B to D, Fig. 2, Fig. S1 in the supplemental material, and Data Set S1), genes with expression intensity (Cy5 channel) values that were in all samples within the lowest 20th percentile were excluded because expression differences in genes expressed at near-background levels are of low confidence. Based on this criterion, we excluded 17 out of 103 genes with an absolute value of the $\log_2$ mAFC (vNF54/pNF54) of >2 (Fig. 1B to D), 131 out of 889 genes with a value of >1 (Data Set S1), and 4 out of 42 genes with differential expression among vNF54 lines (Fig. S1; see below). Additionally, we excluded transcripts that lack a PlasmoDB ID and genes with apparent artifacts according to visual inspection (i.e., genes with large expression differences observed only at time intervals that do not correspond to their peak expression) from the list of top differentially expressed genes presented

in Fig. 1B to D (6 and 13 genes, respectively) and Fig. S1 (7 and 4 genes, respectively). For the analysis presented in Fig. S1, we compared transcript levels directly between the different vNF54 lines. Since differences observed between lines that were not analyzed in parallel may potentially be attributable to a batch effect, we only included genes with an absolute value of the $\log_2$ mAFC of >2 (in any of the possible pairwise comparisons) among the vNF54 lines compared both directly and after normalizing against the pNF54 line analyzed in parallel. Heatmaps and hierarchical clustering based on Euclidean distance were generated using TMEV 4.9 (82). For the generation of volcano plots, samples from the different volunteers were treated as replicates, and an unpaired two-sided *t* test was performed between volunteer and parental NF54 replicates.

RT-qPCR transcriptional analysis. RNA was purified from parasite samples collected in TRIzol (Invitrogen) using the RNeasy minikit (Qiagen) as previously described (23, 79). Next, purified RNA was reverse transcribed using the reverse transcription system (Promega) alongside parallel reactions without reverse transcriptase to exclude gDNA contamination. Quantitative PCR to analyze cDNAs was performed in triplicate wells using the Power SYBR green master mix (Applied Biosystems) in a StepOnePlus real-time PCR system, essentially as previously described (26, 79). Relative transcript levels were calculated using the standard curve method and the normalizing gene serine-tRNA ligase (*serrs*), which shows stable transcript levels across the IDC. The primers used for qPCR are described in Data Set S5.

Whole-genome sequencing. PCR-free whole-genome Illumina sequencing was used to sequence the whole genome of the parental NF54 line alongside two vNF54 lines (V18 and V63). Genomic DNA was sheared to ~150 to 400 bp using a Covaris S220 ultrasonicator, and the NEBNext Ultra DNA library prep kit for Illumina was used for library preparation with specific paired-end TruSeq Illumina adaptors for each sample. Due to the high AT content of the *P. falciparum* genome, the end repair incubation step at 65°C was omitted. After quality control using the Bioanalyzer DNA high sensitivity kit (Agilent) and quantification using the Kapa library quantification kit (Roche), libraries were sequenced using the Illumina HiSeq2500 system. Over 6 million 125-bp paired-end reads were obtained for each sample.

For the data analysis, read quality was checked (FastQC program), and adaptors were trimmed (Cutadapt program) before mapping sequenced reads to the PlasmodDB *P. falciparum* 3D7 reference genome version 46 (https://plasmodb.org/plasmo/) using the BWA-MEM alignment algorithm. Next, GATK UnifiedGenotyper was used to perform variant calling based on GATK best practices to identify SNPs and small indels. GATK Variant Filtration was used to filter out variants with low calling quality (Phred QUAL < 20) or low read depth (DP < 20). Differences in SNP/indel allelic frequency between the different strains were calculated for each SNP/indel, and those showing a <50% difference were filtered out. GenomeBrowse (Golden Helix) was used to visualize alignments and variants, as well as for the detection of large subtelomeric deletions or duplications.

ChIP-seq experiments and analysis. ChIP-seq experiments were performed essentially as previously described (80). In brief, chromatin was extracted from saponin-lysed synchronous parasite cultures at the late-trophozoite/early-schizont stage using the MAGnify chromatin immunoprecipitation system (Life Technologies) (34). After cross-linking and washing, chromatin was sonicated with an M220 sonicator (Covaris) at 10% duty factor, 200 cycles per burst, and 140 W of peak incident power for 10 min. Next, 4 μg of chromatin were immunoprecipitated overnight at 4°C with 8 μg of antibody against H3K9me3 (Diagenode; catalog no. C15410193) previously coupled to protein A/G magnetic beads provided in the kit. Samples were then washed, de-cross-linked, and eluted following the MAGnify ChIP system recommendations but, at all times, avoiding high temperatures that could result in the denaturation of extremely AT-rich intergenic regions. De-cross-linking, proteinase K treatment, and elution were performed at 45°C (for 2 h, overnight, and 1.5 h, respectively).

Using a protocol adapted to a genome with an extremely high AT richness (83), libraries for Illumina sequencing were prepared from 5 ng of immunoprecipitated DNA. After end repair and addition of 3′ A overhangs, NEBNext multiplex oligos for Illumina (NEB; catalog nos. E7335 and E7500) were ligated. Agencourt AMPure XP beads (Beckman Coulter) were used for the purification steps, and the libraries were amplified (9 amplification cycles) with the Kapa HiFi PCR kit (Kapa Biosystems) in Kapa HiFi fidelity buffer (5×). Finally, 0.9× AMPure XP beads were used to purify the amplified libraries and remove adapter dimers. After library size analysis using a 4200 TapeStation system (Agilent Technologies) and quantification using the Kapa library quantification kit (Roche), sequencing was performed with a HiSeq2500 system (Illumina), obtaining 6 to 10 million 125-bp paired-end reads per sample.

Reads were mapped to the 3D7 reference genome using the Bowtie2 local alignment algorithm. Differential peak calling was performed using MACS2 (v2.2.7.1) following the author's recommendations. A first round of peak calling for every sample was performed using the MACS2 callpeak command with parameters as follows: -f BAMPE -g 2.41e7 –fe-cutoff 1.5 –nomodel –extsize 150, and then differential peaks were called using the bdgdiff command with parameters as follows: -g 250 -l 300 –cutoff 5. Results were annotated using custom scripts in Python against the *P. falciparum* 3D7 annotation in PlasmoDB (v46). The Integrative Genomics Viewer (IGV) was used to visualize the data.

H3K9me3 coverage across the genome for each sample was calculated using the DeepTools (v.3.5.0) BamCompare command. After normalizing to reads per kilobase per million (RPKMs), coverage was defined as the $\log_2$(IP/input) and computed for 100-bp intervals. For each gene, we calculated the average coverage for the 1,000 bp upstream plus the first 500 bp of the coding sequence because this is the region in which heterochromatin is typically associated with a silenced state.

Determination of sexual conversion rates. Sexual conversion rates were measured by treating sorbitol synchronized ring-stage cultures with 50 mM *N*-acetyl-D-glucosamine (GlcNAc) (Sigma-Aldrich; catalog no. A32869) for 5 days to eliminate asexual parasites. The sexual conversion rate was calculated as the gametocytemia at day 5 relative to the initial rings parasitemia at day 0 (the day of adding GlcNAc),

as previously described (84). Initial parasitemia was measured by flow cytometry (85) and gametocytemia by light microscopy quantification of Giemsa-stained smears.

Microsphiltration assay. In order to compare the deformability of erythrocytes infected with pNF54 and vNF54 (V63), we used a microsphiltration assay (86). Calibrated metal microspheres (96.50% tin, 3.00% silver, and 0.50% copper; Industrie des Poudres Sphériques) with two different size distributions (5 to 15 μm and 15 to 25 μm in diameter) were used to create a matrix within an inverted 1,000-μl antiaerosol pipette tip (Neptune). For this, 4 g of dry microspheres of each size range were resuspended in 12 ml of parasite culture medium (with 10% human serum), and 400 μl of this microsphere suspension was added into the tip and left to settle and form a 3- to 4-mm-thick layer above the tip filter. Next, 600 μl of tightly synchronized 30- to 34-hpi cultures (1 to 6% parasitemia) were placed on top of the microsphere layer and then perfused through the microsphere matrix at a flow rate of 60 ml/h using an electric pump (Syramed SP6000; Arcomed Ag), followed by a wash with 5 ml of culture medium.

Samples collected after perfusion through the matrix (in triplicate) and the original culture (not passed through the matrix) were analyzed by flow cytometry. For this, 2 μl of erythrocytes pellet were collected in 200 μl of RPMI and washed twice with 200 μl of phosphate-buffered saline (PBS), incubated for 25 min at room temperature with SYBR green (1:2,000 dilution of Sigma S9430 solution), washed twice again with 200 μl of PBS, and finally resuspended in 1.5 ml of PBS. The parasitemia of each sample was then measured with a BD Accuri C6 cytometer.

Sorbitol lysis assay. To test sorbitol sensitivity, tightly synchronized 20- to 24-hpi pNF54 and vNF54 (V63) cultures were treated with a sorbitol-containing isosmotic solution (300 mM sorbitol supplemented with 10 mM HEPES, 5 mM glucose, and adjusted to pH 7.4) or the same isosmotic solution with 100 μM of the general anion channel inhibitor 5-nitro-2-(3-phenylpropylamino) benzoic acid (NPPB, Sigma-Aldrich), which inhibits malarial new permeation pathways (87). Parasitemia was determined before (time zero) and after (time 60) incubation at 37°C for 1 h, using flow cytometry. The lysis percentage was calculated for each condition using the following formula: % lysis = [1 − (parasitemia time 60 min/parasitemia time zero)] × 100.

Western blotting. Synchronous cultures containing mainly ~40-hpi schizonts were purified by magnetic isolation, divided into pellets of approximately 2.5×10^6 schizonts, and stored frozen at −80°C. After thawing, proteins were denaturized in SDS-PAGE protein loading buffer for 5 min at 95°C and resolved by SDS-PAGE on 4 to 12% bis-Tris Criterion XT precast gels (Bio-Rad), transferred to a polyvinylidene difluoride (PVDF) membrane, and blocked for at least 1 h in 1% casein blocking buffer (Sigma). Membranes were incubated overnight at 4°C with the following primary antibodies: purified mouse antiserum against a PfMC-2TM protein (PF3D7_0114100) at 1:200 (kindly provided by Catherine Braun-Breton) (88) and mouse anti-HSP70 antibody (89) at 1:2,000. Membranes were then incubated for 1 h at room temperature with horseradish peroxidase-conjugated anti-mouse IgG secondary antibody (Promega) at 1:10,000, and peroxidase was detected using the Pierce chemiluminescence system (Pierce) following the manufacturer's instructions. To control for equal loading, parts of the membranes corresponding to different molecular weight ranges were separately hybridized with different antibodies. Signal quantification was performed using ImageJ software.

Data availability. The microarray and ChIP-seq data have been deposited in the GEO database with accession numbers GSE166258 and GSE166390, respectively. ChIP-seq data can be visualized at the UCSC genome browser using the following link: http://genome.ucsc.edu/s/apickford/apickford_lmichel. Whole-genome sequence data have been deposited at the Sequence Read Archive (SRA) database with accession number PRJNA699845. The scripts used for microarray data analysis are available in GitHub (https://github.com/CortesMalariaLab/A_Pickford_CHMItranscriptomic).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

DATA SET S1, XLS file, 0.2 MB.
DATA SET S2, XLS file, 0.7 MB.
DATA SET S3, XLS file, 1.6 MB.
DATA SET S4, XLS file, 1.2 MB.
DATA SET S5, XLS file, 0.03 MB.
FIG S1, PDF file, 0.2 MB.
FIG S2, PDF file, 0.3 MB.
FIG S3, PDF file, 0.3 MB.
FIG S4, PDF file, 0.2 MB.
FIG S5, PDF file, 0.1 MB.

ACKNOWLEDGMENTS

We thank Catherine Braun-Breton (Université de Montpellier) for providing antibodies against a PfMC-2TM protein; the Genomics Unit at the CRG for assistance with genome sequencing; Gloria P. Gómez-Pérez, José Muñoz (ISGlobal), the volunteers and clinical staff who participated in the CHMI in Barcelona, Sofía Mira-Martínez, and Ariel Magallón-Tejada (ISGlobal) for their contribution to the collection of samples for transcriptional

analysis in the CHMI study; Adam J. Reid (Wellcome Sanger Institute), J. Alexandra Rowe, and Philip J. Spence (University of Edinburgh) for providing processed data from the Milne et al. article; and Stephen L. Hoffman and Kim Lee Sim (Sanaria) for providing cryopreserved *P. falciparum* sporozoites.

This work was supported by grants from the Spanish Ministerio de Ciencia e Innovación (MCI)/Agencia Estatal de Investigación (AEI) (SAF2016-76190-R and PID2019-107232RB-I00 to A.C.), cofunded by the European Regional Development Fund (ERDF, European Union), and grants by the Instituto de Salud Carlos III and Fundación Ramón Areces to P.L.A. for the CHMI trial. F.D. and C.L. were supported by CNRS, INSERM, and the Fondation pour la Recherche Médicale under award number Equipe FRM EQ20170336722D. A.K.P. is supported by a fellowship from the Secretary for Universities and Research, Catalan Government (FI_B 00373), cofunded by the European Social Fund (ESF), European Commission. L.M.-T. is supported by a fellowship from the Spanish Ministry of Economy and Competitiveness (BES-2017-081079), cofunded by the European Social Fund (ESF). Our research is part of ISGlobal's Program on the Molecular Mechanisms of Malaria, which is partially supported by the Fundación Ramón Areces. We acknowledge support from the Spanish Ministry of Science and Innovation through the Centro de Excelencia Severo Ochoa 2019-2023 program (CEX2018-000806-S) and support from the Generalitat de Catalunya through the CERCA Program. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

REFERENCES

1. World Health Organization. 2020. World malaria report 2020. World Health Organization, Geneva, Switzerland.
2. Venugopal K, Hentzschel F, Valkiūnas G, Marti M. 2020. *Plasmodium* asexual growth and sexual development in the haematopoietic niche of the host. Nat Rev Microbiol 18:177–189. https://doi.org/10.1038/s41579-019-0306-2.
3. Bozdech Z, Llinas M, Pulliam BL, Wong ED, Zhu J, DeRisi JL. 2003. The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. PLoS Biol 1:E5. https://doi.org/10.1371/journal.pbio.0000005.
4. Le Roch KG, Zhou Y, Blair PL, Grainger M, Moch JK, Haynes JD, De La Vega P, Holder AA, Batalov S, Carucci DJ, Winzeler EA. 2003. Discovery of gene function by expression profiling of the malaria parasite life cycle. Science 301:1503–1508. https://doi.org/10.1126/science.1087025.
5. Howick VM, Russell AJC, Andrews T, Heaton H, Reid AJ, Natarajan K, Butungi H, Metcalf T, Verzier LH, Rayner JC, Berriman M, Herren JK, Billker O, Hemberg M, Talman AM, Lawniczak MKN. 2019. The Malaria Cell Atlas: single parasite transcriptomes across the complete *Plasmodium* life cycle. Science 365:eaaw2619. https://doi.org/10.1126/science.aaw2619.
6. Mackinnon MJ, Marsh K. 2010. The selection landscape of malaria parasites. Science 328:866–871. https://doi.org/10.1126/science.1185410.
7. Zuzarte-Luis V, Mota MM. 2018. Parasite sensing of host nutrients and environmental cues. Cell Host Microbe 23:749–758. https://doi.org/10.1016/j.chom.2018.05.018.
8. Llora-Batlle O, Tinto-Font E, Cortes A. 2019. Transcriptional variation in malaria parasites: why and how. Brief Funct Genomics 18:329–341. https://doi.org/10.1093/bfgp/elz009.
9. Cortés A, Deitsch KW. 2017. Malaria epigenetics. Cold Spring Harb Perspect Med 7:a025528. https://doi.org/10.1101/cshperspect.a025528.
10. Guizetti J, Scherf A. 2013. Silence, activate, poise and switch! Mechanisms of antigenic variation in *Plasmodium falciparum*. Cell Microbiol 15:718–726. https://doi.org/10.1111/cmi.12115.
11. Voss TS, Bozdech Z, Bartfai R. 2014. Epigenetic memory takes center stage in the survival strategy of malaria parasites. Curr Opin Microbiol 20:88–95. https://doi.org/10.1016/j.mib.2014.05.007.
12. Rovira-Graells N, Gupta AP, Planet E, Crowley VM, Mok S, Ribas de Pouplana L, Preiser PR, Bozdech Z, Cortés A. 2012. Transcriptional variation in the malaria parasite *Plasmodium falciparum*. Genome Res 22:925–938. https://doi.org/10.1101/gr.129692.111.
13. Levy SF, Ziv N, Siegal ML. 2012. Bet hedging in yeast by heterogeneous, age-correlated expression of a stress protectant. PLoS Biol 10:e1001325. https://doi.org/10.1371/journal.pbio.1001325.
14. Veening JW, Smits WK, Kuipers OP. 2008. Bistability, epigenetics, and bet-hedging in bacteria. Annu Rev Microbiol 62:193–210. https://doi.org/10.1146/annurev.micro.62.081307.163002.
15. Starrfelt J, Kokko H. 2012. Bet-hedging–a triple trade-off between means, variances and correlations. Biol Rev Camb Philos Soc 87:742–755. https://doi.org/10.1111/j.1469-185X.2012.00225.x.
16. Smith JD, Rowe JA, Higgins MK, Lavstsen T. 2013. Malaria's deadly grip: cytoadhesion of *Plasmodium falciparum*-infected erythrocytes. Cell Microbiol 15:1976–1983. https://doi.org/10.1111/cmi.12183.
17. Scherf A, Lopez-Rubio JJ, Riviere L. 2008. Antigenic variation in *Plasmodium falciparum*. Annu Rev Microbiol 62:445–470. https://doi.org/10.1146/annurev.micro.61.080706.093134.
18. Wahlgren M, Goel S, Akhouri RR. 2017. Variant surface antigens of *Plasmodium falciparum* and their roles in severe malaria. Nat Rev Microbiol 15:479–491. https://doi.org/10.1038/nrmicro.2017.47.
19. Mira-Martinez S, Pickford AK, Rovira-Graells N, Guetens P, Tinto-Font E, Cortes A, Rosanas-Urgell A. 2019. Identification of antimalarial compounds that require CLAG3 for their uptake by *Plasmodium falciparum*-infected erythrocytes. Antimicrob Agents Chemother 63:e00052-19. https://doi.org/10.1128/AAC.00052-19.
20. Nguitragool W, Bokhari AA, Pillai AD, Rayavara K, Sharma P, Turpin B, Aravind L, Desai SA. 2011. Malaria parasite *clag3* genes determine channel-mediated nutrient uptake by infected red blood cells. Cell 145:665–677. https://doi.org/10.1016/j.cell.2011.05.002.
21. Pillai AD, Nguitragool W, Lyko B, Dolinta K, Butler MM, Nguyen ST, Peet NP, Bowlin TL, Desai SA. 2012. Solute restriction reveals an essential role for *clag3*-associated channels in malaria parasite nutrient acquisition. Mol Pharmacol 82:1104–1114. https://doi.org/10.1124/mol.112.081224.
22. Mira-Martínez S, Rovira-Graells N, Crowley VM, Altenhofen LM, Llinás M, Cortés A. 2013. Epigenetic switches in *clag3* genes mediate blasticidin S resistance in malaria parasites. Cell Microbiol 15:1913–1923. https://doi.org/10.1111/cmi.12162.
23. Mira-Martínez S, van Schuppen E, Amambua-Ngwa A, Bottieau E, Affara M, Van Esbroeck M, Vlieghe E, Guetens P, Rovira-Graells N, Gómez-Perez GP, Alonso PL, D'Alessandro U, Rosanas-Urgell A, Cortés A. 2017. Expression of the *Plasmodium falciparum* clonally variant *clag3* genes in human infections. J Infect Dis 215:938–945. https://doi.org/10.1093/infdis/jix053.
24. Sharma P, Wollenberg K, Sellers M, Zainabadi K, Galinsky K, Moss E, Nguitragool W, Neafsey D, Desai SA. 2013. An epigenetic antimalarial resistance mechanism involving parasite genes linked to nutrient uptake. J Biol Chem 288:19429–19440. https://doi.org/10.1074/jbc.M113.468371.

25. Cortés A, Crowley VM, Vaquero A, Voss TS. 2012. A view on the role of epigenetics in the biology of malaria parasites. PLoS Pathog 8:e1002943. https://doi.org/10.1371/journal.ppat.1002943.
26. Crowley VM, Rovira-Graells N, de Pouplana LR, Cortés A. 2011. Heterochromatin formation in bistable chromatin domains controls the epigenetic repression of clonally variant *Plasmodium falciparum* genes linked to erythrocyte invasion. Mol Microbiol 80:391–406. https://doi.org/10.1111/j.1365-2958.2011.07574.x.
27. Jiang L, Lopez-Barragan MJ, Jiang H, Mu J, Gaur D, Zhao K, Felsenfeld G, Miller LH. 2010. Epigenetic control of the variable expression of a *Plasmodium falciparum* receptor protein for erythrocyte invasion. Proc Natl Acad Sci U S A 107:2224–2229. https://doi.org/10.1073/pnas.0913396107.
28. Lopez-Rubio JJ, Gontijo AM, Nunes MC, Issar N, Hernandez Rivas R, Scherf A. 2007. 5' flanking region of *var* genes nucleate histone modification patterns linked to phenotypic inheritance of virulence traits in malaria parasites. Mol Microbiol 66:1296–1305. https://doi.org/10.1111/j.1365-2958.2007.06009.x.
29. Brancucci NM, Bertschi NL, Zhu L, Niederwieser I, Chin WH, Wampfler R, Freymond C, Rottmann M, Felger I, Bozdech Z, Voss TS. 2014. Heterochromatin protein 1 secures survival and transmission of malaria parasites. Cell Host Microbe 16:165–176. https://doi.org/10.1016/j.chom.2014.07.004.
30. Chookajorn T, Dzikowski R, Frank M, Li F, Jiwani AZ, Hartl DL, Deitsch KW. 2007. Epigenetic memory at malaria virulence genes. Proc Natl Acad Sci U S A 104:899–902. https://doi.org/10.1073/pnas.0609084103.
31. Scherf A, Hernandez-Rivas R, Buffet P, Bottius E, Benatar C, Pouvelle B, Gysin J, Lanzer M. 1998. Antigenic variation in malaria: in situ switching, relaxed and mutually exclusive transcription of *var* genes during intra-erythrocytic development in *Plasmodium falciparum*. EMBO J 17:5418–5426. https://doi.org/10.1093/emboj/17.18.5418.
32. Cortés A, Carret C, Kaneko O, Yim Lim BY, Ivens A, Holder AA. 2007. Epigenetic silencing of *Plasmodium falciparum* genes linked to erythrocyte invasion. PLoS Pathog 3:e107. https://doi.org/10.1371/journal.ppat.0030107.
33. Dzikowski R, Frank M, Deitsch K. 2006. Mutually exclusive expression of virulence genes by malaria parasites is regulated independently of antigen production. PLoS Pathog 2:e22. https://doi.org/10.1371/journal.ppat.0020022.
34. Rovira-Graells N, Crowley VM, Bancells C, Mira-Martínez S, Ribas de Pouplana L, Cortés A. 2015. Deciphering the principles that govern mutually exclusive expression of *Plasmodium falciparum clag3* genes. Nucleic Acids Res 43:8243–8257. https://doi.org/10.1093/nar/gkv730.
35. Voss TS, Healer J, Marty AJ, Duffy MF, Thompson JK, Beeson JG, Reeder JC, Crabb BS, Cowman AF. 2006. A *var* gene promoter controls allelic exclusion of virulence genes in *Plasmodium falciparum* malaria. Nature 439:1004–1008. https://doi.org/10.1038/nature04407.
36. Bachmann A, Petter M, Krumkamp R, Esen M, Held J, Scholz JA, Li T, Sim BK, Hoffman SL, Kremsner PG, Mordmuller B, Duffy MF, Tannich E. 2016. Mosquito passage dramatically changes *var* gene expression in controlled human *Plasmodium falciparum* infections. PLoS Pathog 12:e1005538. https://doi.org/10.1371/journal.ppat.1005538.
37. Peters J, Fowler E, Gatton M, Chen N, Saul A, Cheng Q. 2002. High diversity and rapid changeover of expressed *var* genes during the acute phase of *Plasmodium falciparum* infections in human volunteers. Proc Natl Acad Sci U S A 99:10689–10694. https://doi.org/10.1073/pnas.162349899.
38. Dimonte S, Bruske EI, Hass J, Supan C, Salazar CL, Held J, Tschan S, Esen M, Flotenmeyer M, Koch I, Berger J, Bachmann A, Sim BK, Hoffman SL, Kremsner PG, Mordmuller B, Frank M. 2016. Sporozoite route of infection influences in vitro *var* gene transcription of *Plasmodium falciparum* parasites from controlled human infections. J Infect Dis 214:884–894. https://doi.org/10.1093/infdis/jiw225.
39. Lavazec C, Sanyal S, Templeton TJ. 2007. Expression switching in the *stevor* and *Pfmc-2TM* superfamilies in *Plasmodium falciparum*. Mol Microbiol 64:1621–1634. https://doi.org/10.1111/j.1365-2958.2007.05767.x.
40. Mok BW, Ribacke U, Winter G, Yip BH, Tan CS, Fernandez V, Chen Q, Nilsson P, Wahlgren M. 2007. Comparative transcriptomal analysis of isogenic *Plasmodium falciparum* clones of distinct antigenic and adhesive phenotypes. Mol Biochem Parasitol 151:184–192. https://doi.org/10.1016/j.molbiopara.2006.11.006.
41. Gomez-Escobar N, Amambua-Ngwa A, Walther M, Okebe J, Ebonyi A, Conway DJ. 2010. Erythrocyte invasion and merozoite ligand gene expression in severe and mild *Plasmodium falciparum* malaria. J Infect Dis 201:444–452. https://doi.org/10.1086/649902.
42. Tonkin-Hill GQ, Trianty L, Noviyanti R, Nguyen HHT, Sebayang BF, Lampah DA, Marfurt J, Cobbold SA, Rambhatla JS, McConville MJ, Rogerson SJ, Brown GV, Day KP, Price RN, Anstey NM, Papenfuss AT, Duffy MF. 2018. The *Plasmodium falciparum* transcriptome in severe malaria reveals altered expression of genes involved in important processes including surface antigen-encoding *var* genes. PLoS Biol 16:e2004328. https://doi.org/10.1371/journal.pbio.2004328.
43. Cortés A. 2008. Switching *Plasmodium falciparum* genes on and off for erythrocyte invasion. Trends Parasitol 24:517–524. https://doi.org/10.1016/j.pt.2008.08.005.
44. Brown AC, Guler JL. 2020. From circulation to cultivation: *Plasmodium* in vivo versus in vitro. Trends Parasitol 36:914–926. https://doi.org/10.1016/j.pt.2020.08.008.
45. Mackinnon MJ, Li J, Mok S, Kortok MM, Marsh K, Preiser PR, Bozdech Z. 2009. Comparative transcriptional and genomic analysis of *Plasmodium falciparum* field isolates. PLoS Pathog 5:e1000644. https://doi.org/10.1371/journal.ppat.1000644.
46. Daily JP, Le Roch KG, Sarr O, Ndiaye D, Lukens A, Zhou Y, Ndir O, Mboup S, Sultan A, Winzeler EA, Wirth DF. 2005. In vivo transcriptome of *Plasmodium falciparum* reveals overexpression of transcripts that encode surface proteins. J Infect Dis 191:1196–1203. https://doi.org/10.1086/428289.
47. Bachmann A, Petter M, Tilly AK, Biller L, Uliczka KA, Duffy MF, Tannich E, Bruchhaus I. 2012. Temporal expression and localization patterns of variant surface antigens in clinical *Plasmodium falciparum* isolates during erythrocyte schizogony. PLoS One 7:e49540. https://doi.org/10.1371/journal.pone.0049540.
48. Wang CW, Hermsen CC, Sauerwein RW, Arnot DE, Theander TG, Lavstsen T. 2009. The *Plasmodium falciparum var* gene transcription strategy at the onset of blood stage infection in a human volunteer. Parasitol Int 58:478–480. https://doi.org/10.1016/j.parint.2009.07.004.
49. Lavstsen T, Magistrado P, Hermsen CC, Salanti A, Jensen AT, Sauerwein R, Hviid L, Theander TG, Staalsoe T. 2005. Expression of *Plasmodium falciparum* erythrocyte membrane protein 1 in experimentally infected humans. Malar J 4:21. https://doi.org/10.1186/1475-2875-4-21.
50. Abdi AI, Hodgson SH, Muthui MK, Kivisi CA, Kamuyu G, Kimani D, Hoffman SL, Juma E, Ogutu B, Draper SJ, Osier F, Bejon P, Marsh K, Bull PC. 2017. *Plasmodium falciparum* malaria parasite *var* gene expression is modified by host antibodies: longitudinal evidence from controlled infections of Kenyan adults with varying natural exposure. BMC Infect Dis 17:585. https://doi.org/10.1186/s12879-017-2686-0.
51. Hoo R, Bruske E, Dimonte S, Zhu L, Mordmüller B, Sim BKL, Kremsner PG, Hoffman SL, Bozdech Z, Frank M, Preiser PR. 2019. Transcriptome profiling reveals functional variation in *Plasmodium falciparum* parasites from controlled human malaria infection studies. EBioMedicine 48:442–452. https://doi.org/10.1016/j.ebiom.2019.09.001.
52. Milne K, Ivens A, Reid AJ, Lotkowska ME, O'Toole A, Sankaranarayanan G, Munoz Sandoval D, Nahrendorf W, Regnault C, Edwards NJ, Silk SE, Payne RO, Minassian AM, Venkatraman N, Sanders MJ, Hill AV, Barrett M, Berriman M, Draper SJ, Rowe JA, Spence PJ. 2021. Mapping immune variation and *var* gene switching in naive hosts infected with *Plasmodium falciparum*. Elife 10:e62800. https://doi.org/10.7554/eLife.62800.
53. Gomez-Perez GP, Legarda A, Munoz J, Sim BK, Ballester MR, Dobano C, Moncunill G, Campo JJ, Cistero P, Jimenez A, Barrios D, Mordmuller B, Pardos J, Navarro M, Zita CJ, Nhamuave CA, Garcia-Basteiro AL, Sanz A, Aldea M, Manoj A, Gunasekera A, Billingsley PF, Aponte JJ, James ER, Guinovart C, Antonijoan RM, Kremsner PG, Hoffman SL, Alonso PL. 2015. 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 14:306. https://doi.org/10.1186/s12936-015-0817-x.
54. Flueck C, Bartfai R, Volz J, Niederwieser I, Salcedo-Amaya AM, Alako BT, Ehlgen F, Ralph SA, Cowman AF, Bozdech Z, Stunnenberg HG, Voss TS. 2009. *Plasmodium falciparum* heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors. PLoS Pathog 5:e1000569. https://doi.org/10.1371/journal.ppat.1000569.
55. Fraschka SA, Filarsky M, Hoo R, Niederwieser I, Yam XY, Brancucci NMB, Mohring F, Mushunje AT, Huang X, Christensen PR, Nosten F, Bozdech Z, Russell B, Moon RW, Marti M, Preiser PR, Bartfai R, Voss TS. 2018. Comparative heterochromatin profiling reveals conserved and unique epigenome signatures linked to adaptation and development of malaria parasites. Cell Host Microbe 23:407–420. https://doi.org/10.1016/j.chom.2018.01.008.
56. Lopez-Rubio JJ, Mancio-Silva L, Scherf A. 2009. Genome-wide analysis of heterochromatin associates clonally variant gene regulation with perinuclear repressive centers in malaria parasites. Cell Host Microbe 5:179–190. https://doi.org/10.1016/j.chom.2008.12.012.
57. Salcedo-Amaya AM, van Driel MA, Alako BT, Trelle MB, van den Elzen AM, Cohen AM, Janssen-Megens EM, van de Vegte-Bolmer M, Selzer RR,

Iniguez AL, Green RD, Sauerwein RW, Jensen ON, Stunnenberg HG. 2009. Dynamic histone H3 epigenome marking during the intraerythrocytic cycle of *Plasmodium falciparum*. Proc Natl Acad Sci U S A 106:9655–9660. https://doi.org/10.1073/pnas.0902515106.

58. Sargeant TJ, Marti M, Caler E, Carlton JM, Simpson K, Speed TP, Cowman AF. 2006. Lineage-specific expansion of proteins exported to erythrocytes in malaria parasites. Genome Biol 7:R12. https://doi.org/10.1186/gb-2006-7-2-r12.
59. Templeton TJ. 2009. The varieties of gene amplification, diversification and hypervariability in the human malaria parasite, *Plasmodium falciparum*. Mol Biochem Parasitol 166:109–116. https://doi.org/10.1016/j.molbiopara.2009.04.003.
60. Nolte D, Hundt E, Langsley G, Knapp B. 1991. A *Plasmodium falciparum* blood stage antigen highly homologous to the glycophorin binding protein GBP. Mol Biochem Parasitol 49:253–264. https://doi.org/10.1016/0166-6851(91)90069-i.
61. van Schravendijk MR, Wilson RJ, Newbold CI. 1987. Possible pitfalls in the identification of glycophorin-binding proteins of *Plasmodium falciparum*. J Exp Med 166:376–390. https://doi.org/10.1084/jem.166.2.376.
62. Painter HJ, Altenhofen LM, Kafsack BF, Llinas M. 2013. Whole-genome analysis of *Plasmodium* spp. utilizing a new Agilent Technologies DNA microarray platform. Methods Mol Biol 923:213–219. https://doi.org/10.1007/978-1-62703-026-7_14.
63. Sam-Yellowe TY, Florens L, Johnson JR, Wang T, Drazba JA, Le Roch KG, Zhou Y, Batalov S, Carucci DJ, Winzeler EA, Yates JR, III. 2004. A *Plasmodium* gene family encoding Maurer's cleft membrane proteins: structural properties and expression profiling. Genome Res 14:1052–1059. https://doi.org/10.1101/gr.2126104.
64. Weber JL. 1988. Interspersed repetitive DNA from *Plasmodium falciparum*. Mol Biochem Parasitol 29:117–124. https://doi.org/10.1016/0166-6851(88)90066-7.
65. 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. https://doi.org/10.1073/pnas.96.16.9333.
66. Fernandez V, Hommel M, Chen Q, Hagblom P, Wahlgren M. 1999. Small, clonally variant antigens expressed on the surface of the *Plasmodium falciparum*-infected erythrocyte are encoded by the *rif* gene family and are the target of human immune responses. J Exp Med 190:1393–1404. https://doi.org/10.1084/jem.190.10.1393.
67. Petter M, Haeggstrom M, Khattab A, Fernandez V, Klinkert MQ, Wahlgren M. 2007. Variant proteins of the *Plasmodium falciparum* RIFIN family show distinct subcellular localization and developmental expression patterns. Mol Biochem Parasitol 156:51–61. https://doi.org/10.1016/j.molbiopara.2007.07.011.
68. 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. https://doi.org/10.1016/s0166-6851(98)00144-3.
69. Witmer K, Schmid CD, Brancucci NM, Luah YH, Preiser PR, Bozdech Z, Voss TS. 2012. Analysis of subtelomeric virulence gene families in *Plasmodium falciparum* by comparative transcriptional profiling. Mol Microbiol 84:243–259. https://doi.org/10.1111/j.1365-2958.2012.08019.x.
70. Frank M, Kirkman L, Costantini D, Sanyal S, Lavazec C, Templeton TJ, Deitsch KW. 2008. Frequent recombination events generate diversity within the multi-copy variant antigen gene families of *Plasmodium falciparum*. Int J Parasitol 38:1099–1109. https://doi.org/10.1016/j.ijpara.2008.01.010.
71. Sauerwein RW, Roestenberg M, Moorthy VS. 2011. Experimental human challenge infections can accelerate clinical malaria vaccine development. Nat Rev Immunol 11:57–64. https://doi.org/10.1038/nri2902.
72. McCarthy JS, Sekuloski S, Griffin PM, Elliott S, Douglas N, Peatey C, Rockett R, O'Rourke P, Marquart L, Hermsen C, Duparc S, Möhrle J, Trenholme KR, Humberstone AJ. 2011. A pilot randomised trial of induced blood-stage *Plasmodium falciparum* infections in healthy volunteers for testing efficacy of new antimalarial drugs. PLoS One 6:e21914. https://doi.org/10.1371/journal.pone.0021914.
73. Sanyal S, Egee S, Bouyer G, Perrot S, Safeukui I, Bischoff E, Buffet P, Deitsch KW, Mercereau-Puijalon O, David PH, Templeton TJ, Lavazec C. 2012. *Plasmodium falciparum* STEVOR proteins impact erythrocyte mechanical properties. Blood 119:e1–8–e8. https://doi.org/10.1182/blood-2011-08-370734.
74. Kafsack BF, Rovira-Graells N, Clark TG, Bancells C, Crowley VM, Campino SG, Williams AE, Drought LG, Kwiatkowski DP, Baker DA, Cortes A, Llinas M. 2014. A transcriptional switch underlies commitment to sexual development in malaria parasites. Nature 507:248–252. https://doi.org/10.1038/nature12920.
75. Spence PJ, Brugat T, Langhorne J. 2015. Mosquitoes reset malaria parasites. PLoS Pathog 11:e1004987. https://doi.org/10.1371/journal.ppat.1004987.
76. Spence PJ, Jarra W, Levy P, Reid AJ, Chappell L, Brugat T, Sanders M, Berriman M, Langhorne J. 2013. Vector transmission regulates immune control of *Plasmodium* virulence. Nature 498:228–231. https://doi.org/10.1038/nature12231.
77. Zanghì G, Vembar SS, Baumgarten S, Ding S, Guizetti J, Bryant JM, Mattei D, Jensen ATR, Rénia L, Goh YS, Sauerwein R, Hermsen CC, Franetich JF, Bordessoulles M, Silvie O, Soulard V, Scatton O, Chen P, Mecheri S, Mazier D, Scherf A. 2018. A specific PfEMP1 is expressed in *P. falciparum* sporozoites and plays a role in hepatocyte infection. Cell Rep 22:2951–2963. https://doi.org/10.1016/j.celrep.2018.02.075.
78. Gomez-Diaz E, Yerbanga RS, Lefevre T, Cohuet A, Rowley MJ, Ouedraogo JB, Corces VG. 2017. Epigenetic regulation of *Plasmodium falciparum* clonally variant gene expression during development in *Anopheles gambiae*. Sci Rep 7:40655. https://doi.org/10.1038/srep40655.
79. Casas-Vila N, Pickford AK, Portugaliza HP, Tintó-Font E, Cortés A. Transcriptional analysis of tightly synchronized *Plasmodium falciparum* intraerythrocytic stages by RT-qPCR. Methods Mol Biol, in press. https://doi.org/10.1007/978-1-0716-1681-9_10.
80. Llorà-Batlle O, Michel-Todó L, Witmer K, Toda H, Fernández-Becerra C, Baum J, Cortés A. 2020. Conditional expression of PfAP2-G for controlled massive sexual conversion in *Plasmodium falciparum*. Sci Adv 6:eaaz5057. https://doi.org/10.1126/sciadv.aaz5057.
81. Lemieux JE, Gomez-Escobar N, Feller A, Carret C, Amambua-Ngwa A, Pinches R, Day F, Kyes SA, Conway DJ, Holmes CC, Newbold CI. 2009. Statistical estimation of cell-cycle progression and lineage commitment in *Plasmodium falciparum* reveals a homogeneous pattern of transcription in ex vivo culture. Proc Natl Acad Sci U S A 106:7559–7564. https://doi.org/10.1073/pnas.0811829106.
82. Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, Howe EA, Li J, Thiagarajan M, White JA, Quackenbush J. 2006. TM4 microarray software suite. Methods Enzymol 411:134–193. https://doi.org/10.1016/S0076-6879(06)11009-5.
83. Kensche PR, Hoeijmakers WA, Toenhake CG, Bras M, Chappell L, Berriman M, Bartfai R. 2016. The nucleosome landscape of *Plasmodium falciparum* reveals chromatin architecture and dynamics of regulatory sequences. Nucleic Acids Res 44:2110–2124. https://doi.org/10.1093/nar/gkv1214.
84. Bancells C, Llora-Batlle O, Poran A, Notzel C, Rovira-Graells N, Elemento O, Kafsack BFC, Cortes A. 2019. Revisiting the initial steps of sexual development in the malaria parasite *Plasmodium falciparum*. Nat Microbiol 4:144–154. https://doi.org/10.1038/s41564-018-0291-7.
85. Rovira-Graells N, Aguilera-Simon S, Tinto-Font E, Cortes A. 2016. New assays to characterise growth-related phenotypes of *Plasmodium falciparum* reveal variation in density-dependent growth inhibition between parasite lines. PLoS One 11:e0165358. https://doi.org/10.1371/journal.pone.0165358.
86. Lavazec C, Deplaine G, Safeukui I, Perrot S, Milon G, Mercereau-Puijalon O, David PH, Buffet P. 2013. Microsphiltration: a microsphere matrix to explore erythrocyte deformability. Methods Mol Biol 923:291–297. https://doi.org/10.1007/978-1-62703-026-7_20.
87. Kirk K, Horner HA, Elford BC, Ellory JC, Newbold CI. 1994. Transport of diverse substrates into malaria-infected erythrocytes via a pathway showing functional characteristics of a chloride channel. J Biol Chem 269:3339–3347. https://doi.org/10.1016/S0021-9258(17)41868-0.
88. Vincensini L, Richert S, Blisnick T, Van Dorsselaer A, Leize-Wagner E, Rabilloud T, Braun Breton C. 2005. Proteomic analysis identifies novel proteins of the Maurer's clefts, a secretory compartment delivering *Plasmodium falciparum* proteins to the surface of its host cell. Mol Cell Proteomics 4:582–593. https://doi.org/10.1074/mcp.M400176-MCP200.
89. Naissant B, Dupuy F, Duffier Y, Lorthiois A, Duez J, Scholz J, Buffet P, Merckx A, Bachmann A, Lavazec C. 2016. *Plasmodium falciparum* STEVOR phosphorylation regulates host erythrocyte deformability enabling malaria parasite transmission. Blood 127:e42–e53. https://doi.org/10.1182/blood-2016-01-690776.

SUPPLEMENTARY FIGURES

Expression patterns of *Plasmodium falciparum* clonally variant genes at the onset of a blood infection in malaria-naïve humans.

Anastasia K. Pickford, Lucas Michel-Todó, Florian Dupuy,

Alfredo Mayor, Pedro L. Alonso, Catherine Lavazec & Alfred Cortés

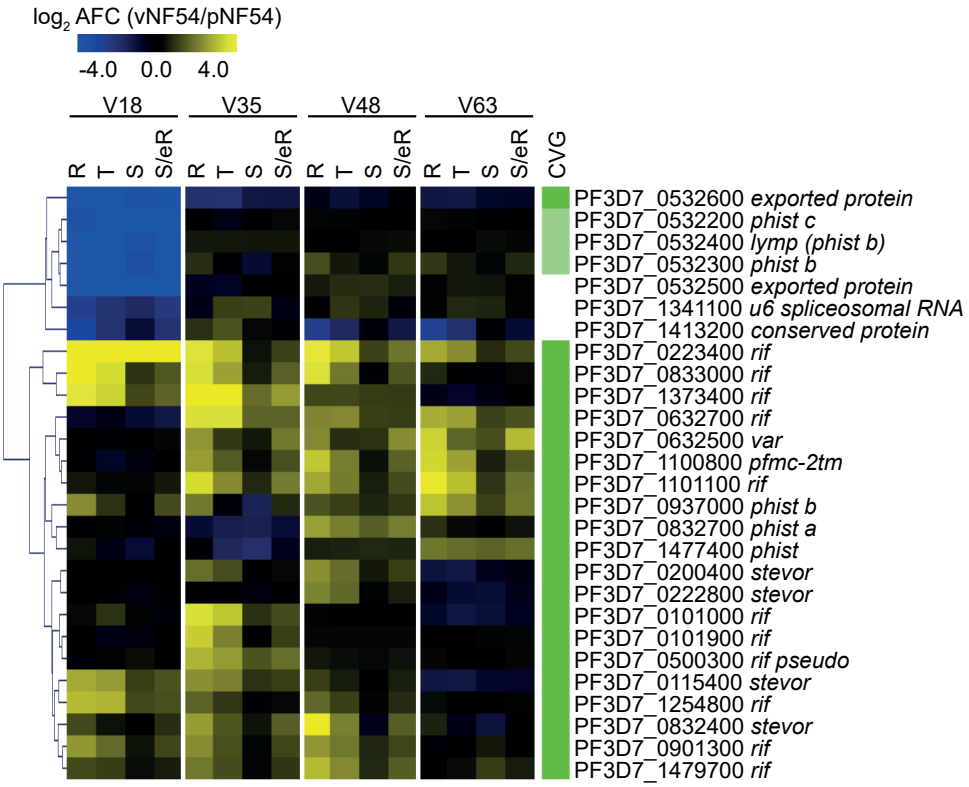
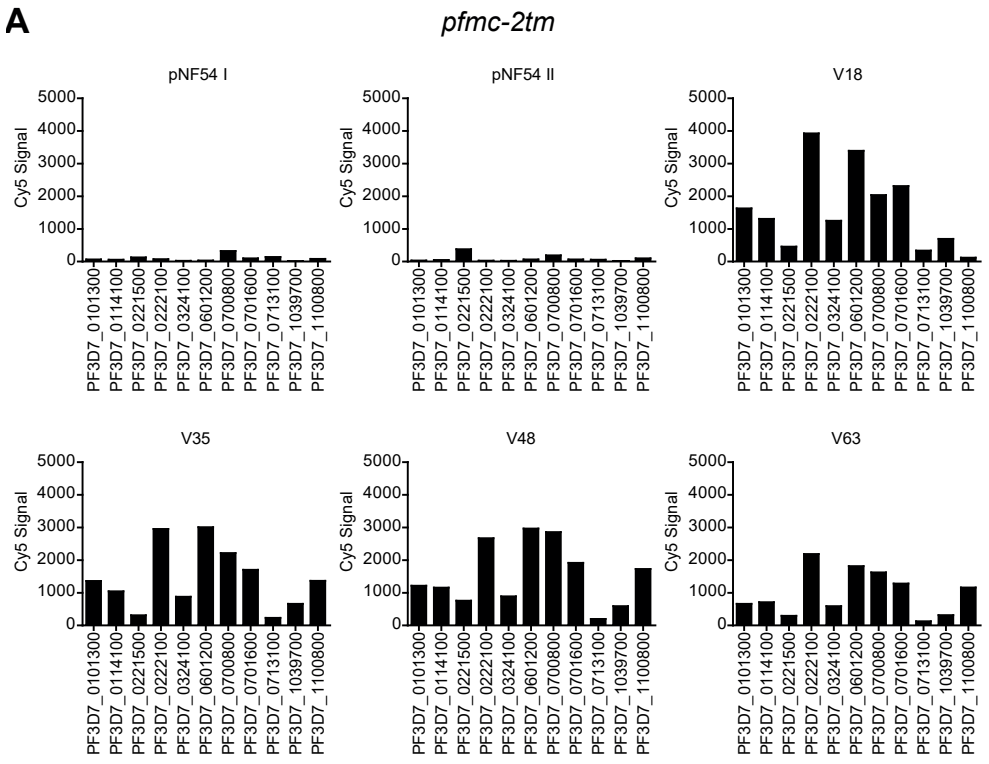
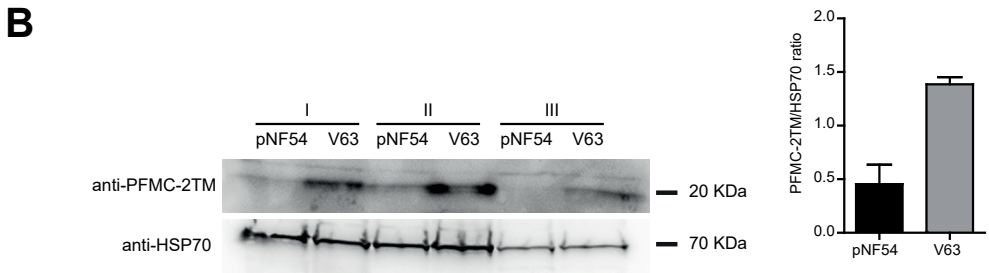
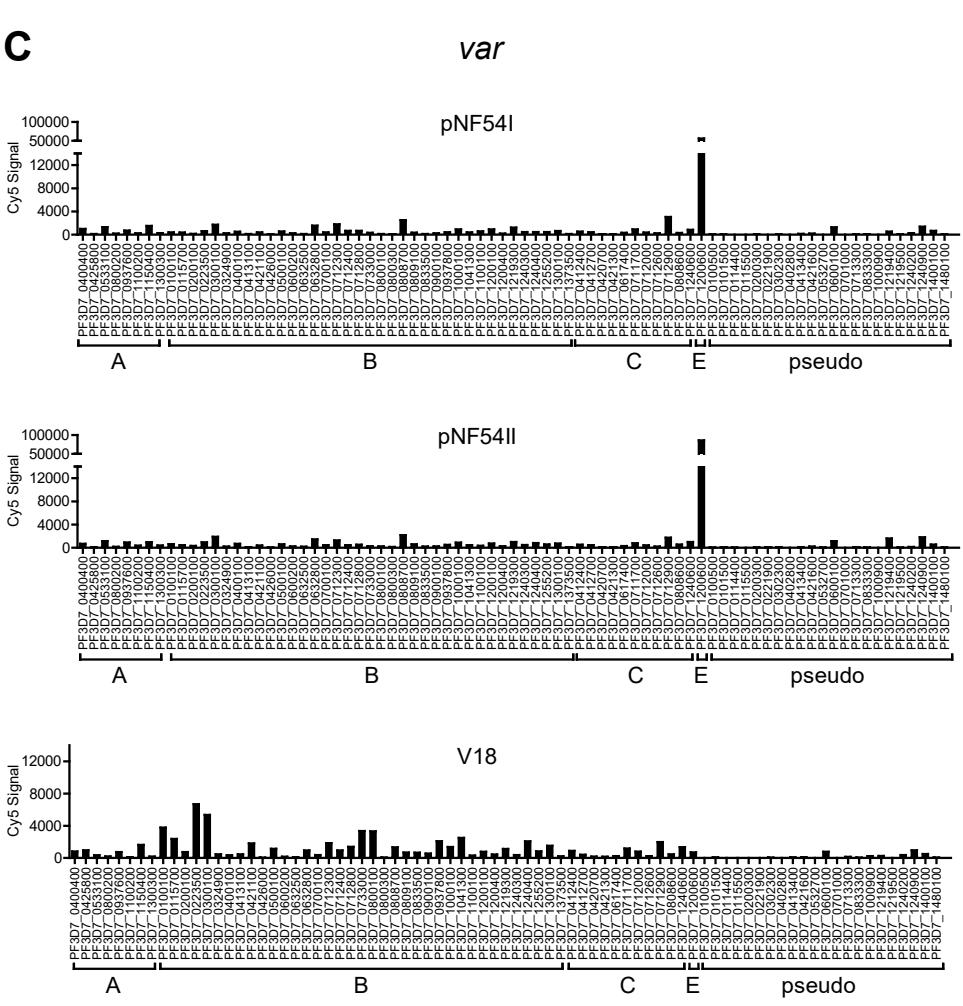
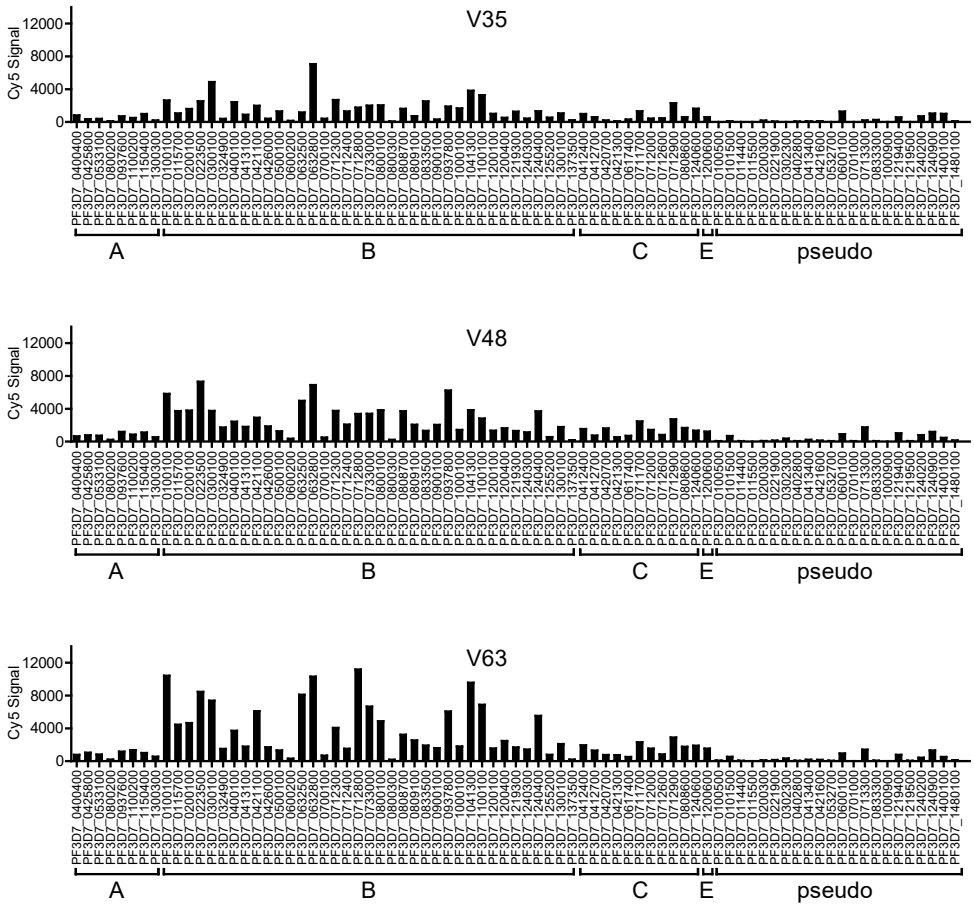


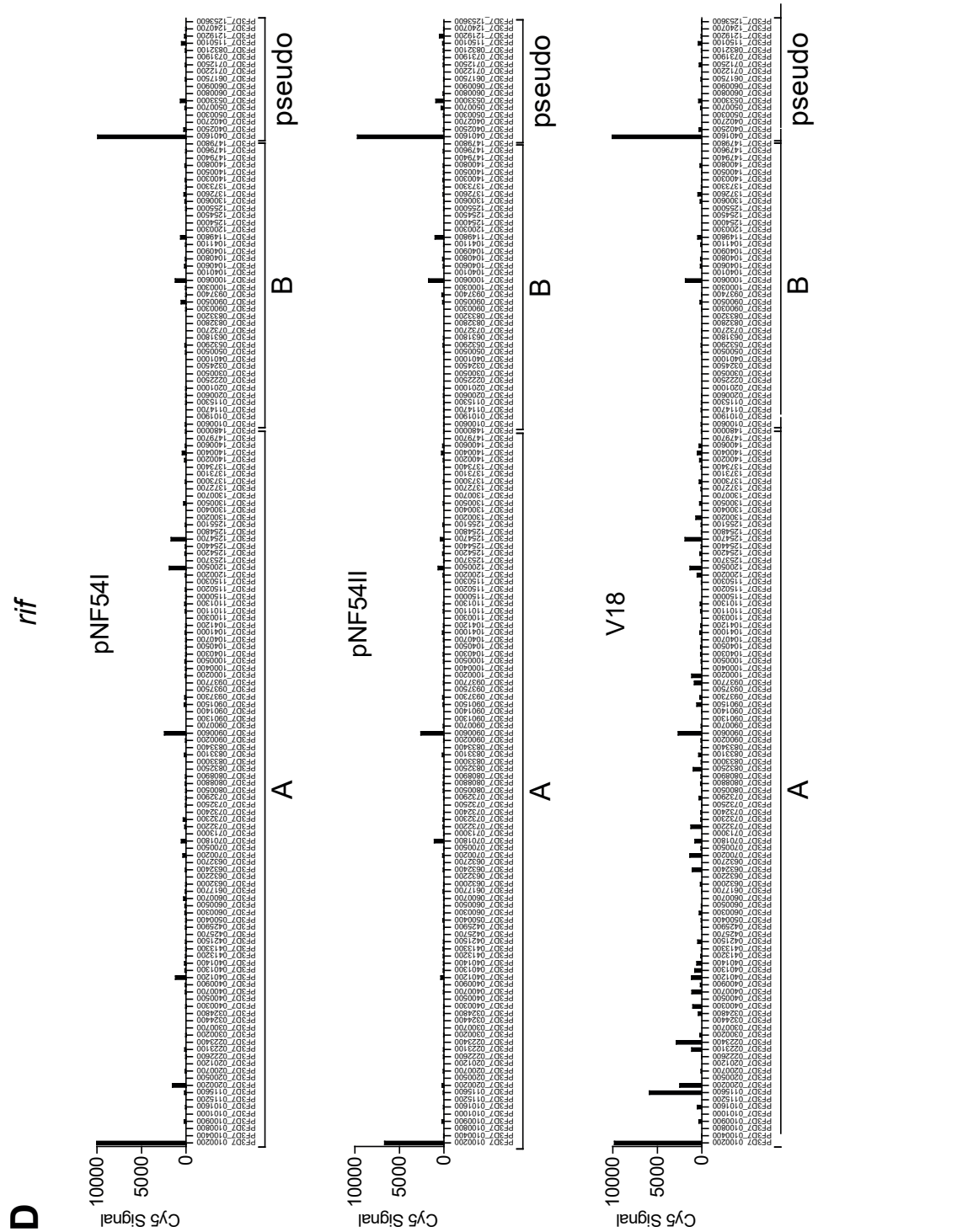
Fig. S1. Genes differentially expressed between parasites obtained from the different infected volunteers. Heatmap of the most differentially expressed genes between the four vNF54 lines (V18, V35, V48, and V63, obtained from four different volunteers), ordered by hierarchical clustering. Values are the $\log_2$ of the average fold-change (AFC) of vNF54 vs pNF54 lines over four overlapping time intervals corresponding to the stages indicated (R: rings, T: trophozoites, S: schizonts, S/eR: schizonts and early rings). Only genes with an absolute value of the $\log_2$(mAFC) >2 in at least one of the possible pairwise comparisons among the vNF54 lines (in both a direct comparison between vNF54 lines and using values normalized against the pNF54 line analyzed in parallel) are shown. See Methods for exclusion criteria. Four genes had values out of the range of the color scale displayed. The column at the right indicates CVGs according to the list in Data Set S2 (dark green) or genes that belong to a gene family in which other genes are CVGs (light green).

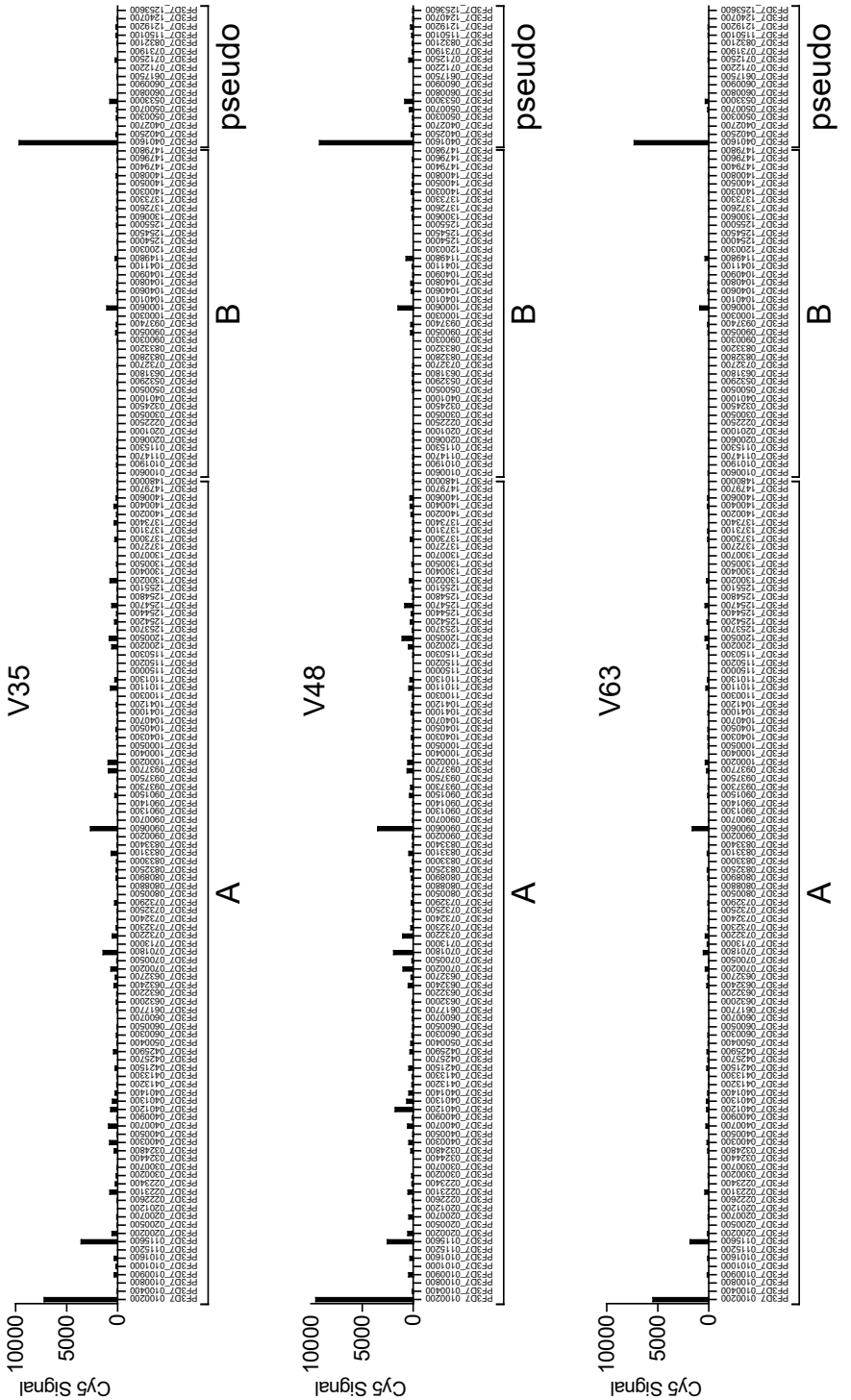

A
pfmc-2tm
pNF54 I
pNF54 II
V18
V35
V48
V63
Cy5 Signal
5000
4000
3000
2000
1000
0
PF3D7_0101300
PF3D7_0114100
PF3D7_0221500
PF3D7_0222100
PF3D7_0324100
PF3D7_0601200
PF3D7_0700800
PF3D7_0701600
PF3D7_0713100
PF3D7_1039700
PF3D7_1100800

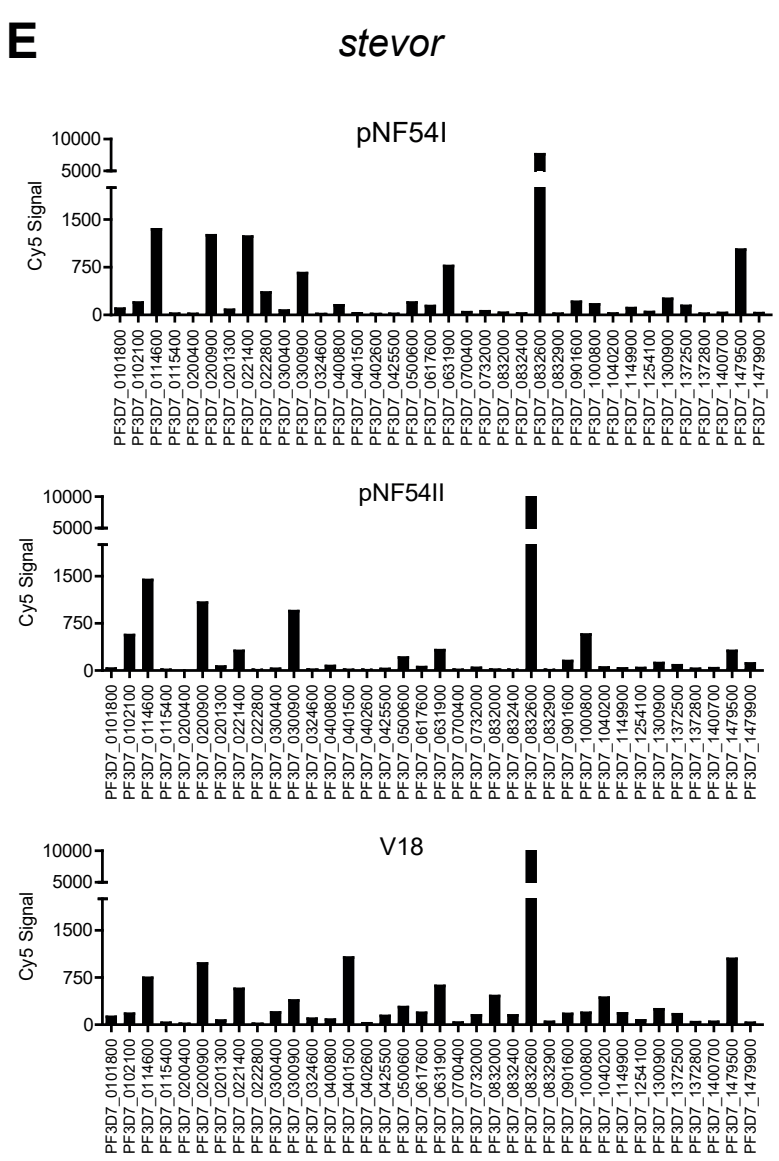

B
I
II
III
pNF54
V63
pNF54
V63
pNF54
V63
anti-PFMC-2TM
anti-HSP70
20 KDa
70 KDa
PFMC-2TM/HSP70 ratio
2.0
1.5
1.0
0.5
0.0
pNF54
V63

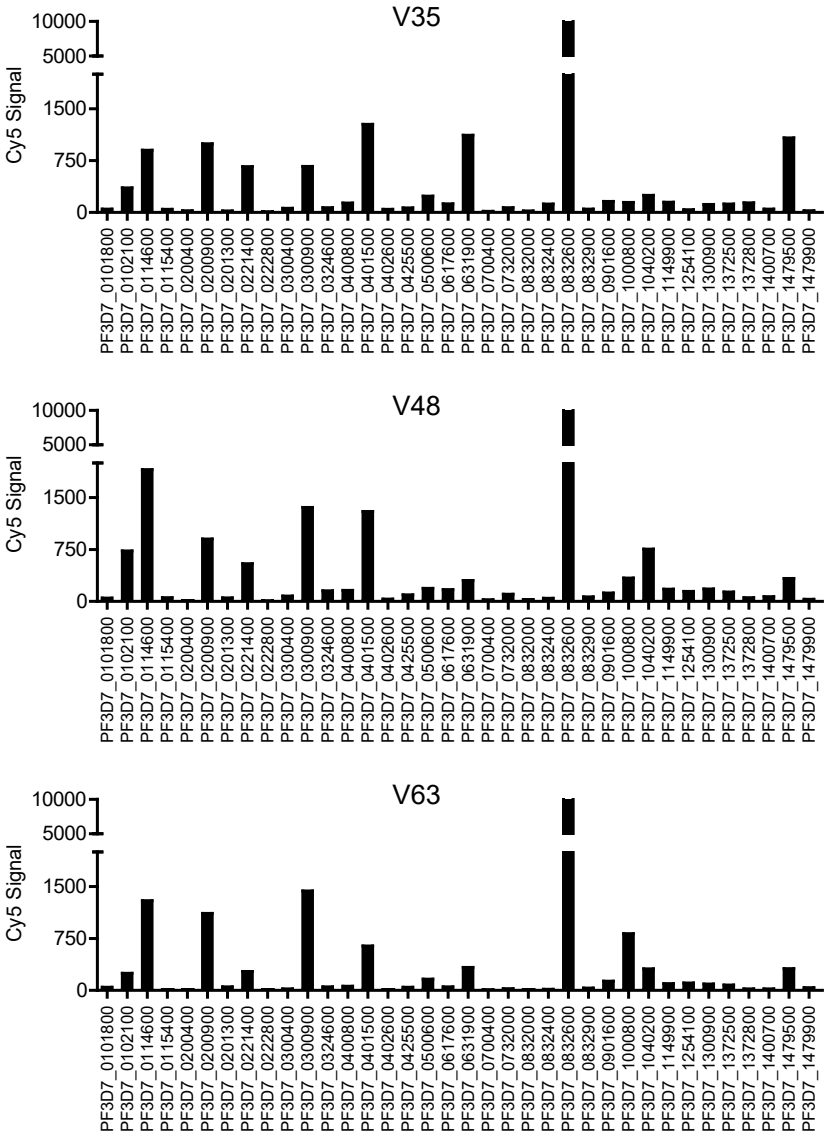

C
var
pNF54I
pNF54II
V18
Cy5 Signal
100000
50000
12000
8000
4000
0
A
B
C
E
pseudo


V35
V48
V63
Cy5 Signal
12000
8000
4000
0
A
B
C
E
pseudo

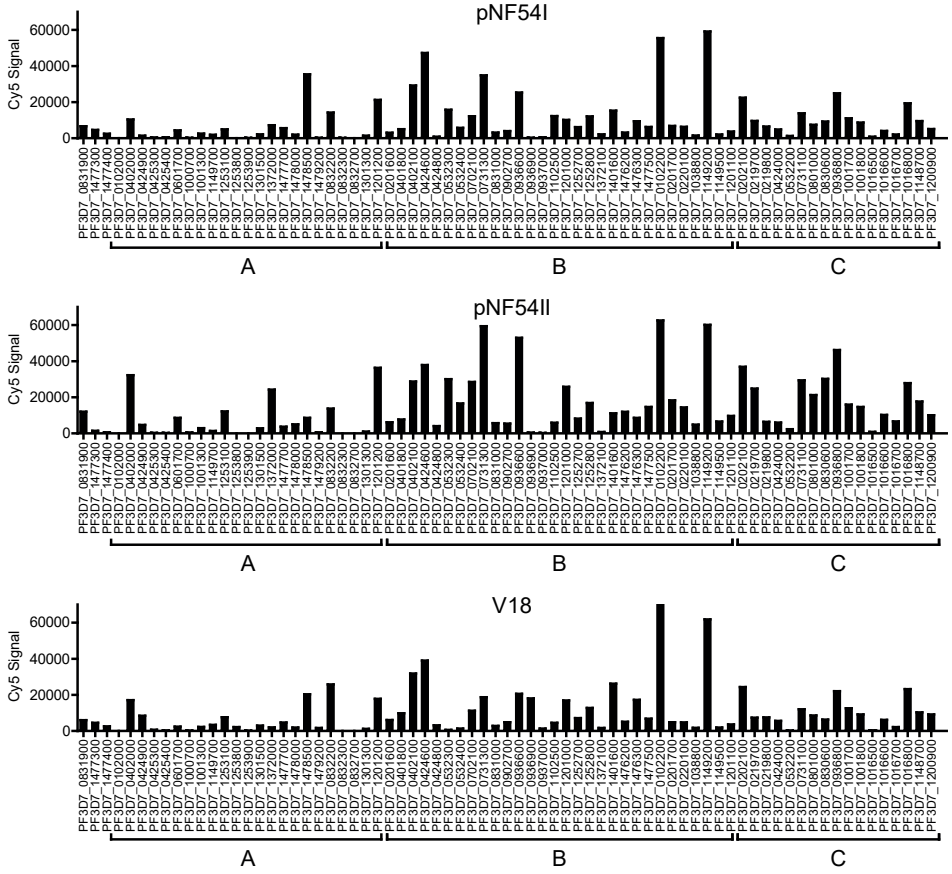

D
rif
pNF541
pNF541I
V18
Cy5 Signal
10000
5000
0
A
B
pseudo

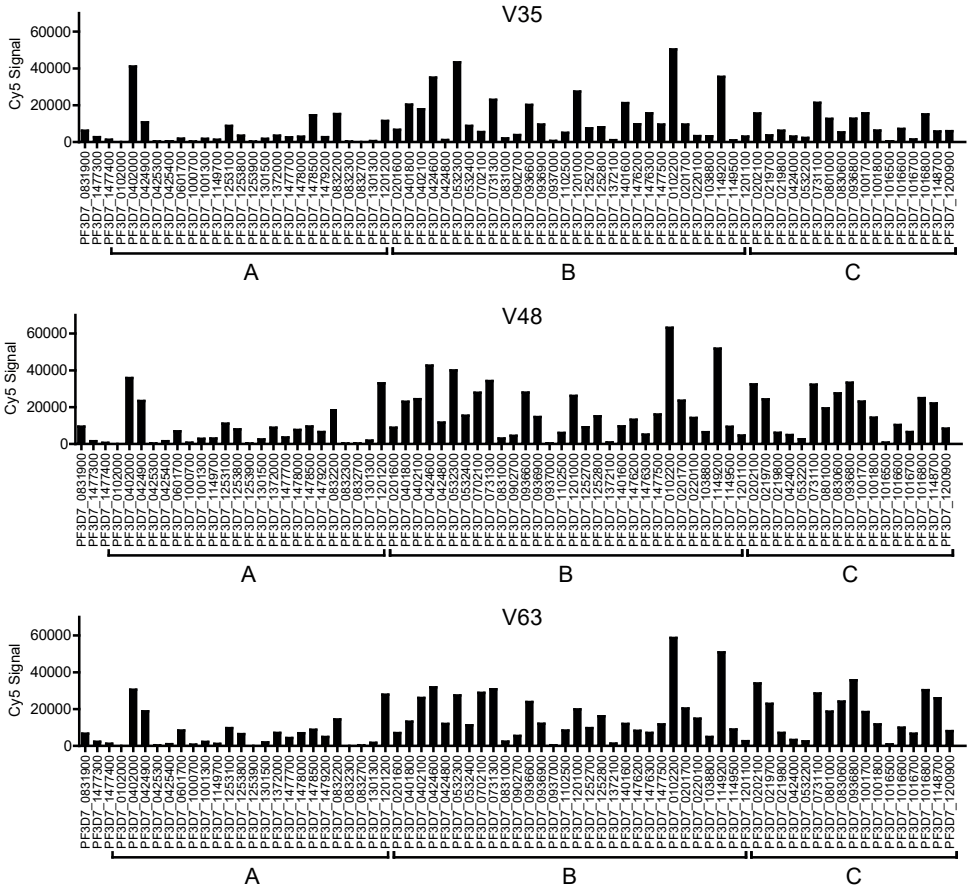

V35
V48
V63
Cy5 Signal
10000
5000
0
A
B
pseudo

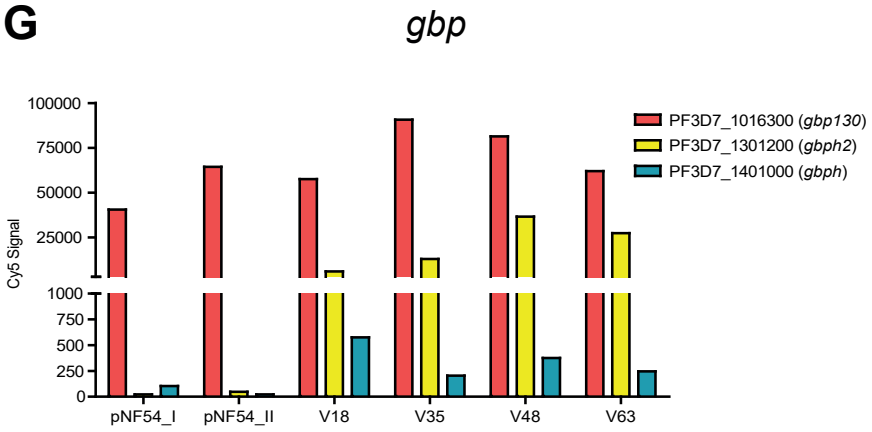

E
stevor
pNF54I
pNF54II
V18
Cy5 Signal
10000
5000
1500
750
0
PF3D7_0101800
PF3D7_0102100
PF3D7_0114600
PF3D7_0115400
PF3D7_0200400
PF3D7_0200900
PF3D7_0201300
PF3D7_0221400
PF3D7_0222800
PF3D7_0300400
PF3D7_0300900
PF3D7_0324600
PF3D7_0400800
PF3D7_0401500
PF3D7_0402600
PF3D7_0425500
PF3D7_0500600
PF3D7_0617600
PF3D7_0631900
PF3D7_0700400
PF3D7_0732000
PF3D7_0832000
PF3D7_0832400
PF3D7_0832600
PF3D7_0832900
PF3D7_0901600
PF3D7_1000800
PF3D7_1040200
PF3D7_1149900
PF3D7_1254100
PF3D7_1300900
PF3D7_1372500
PF3D7_1372800
PF3D7_1400700
PF3D7_1479500
PF3D7_1479900


V35
V48
V63
Cy5 Signal
10000
5000
1500
750
0
PF3D7_0101800
PF3D7_0102100
PF3D7_0114600
PF3D7_0115400
PF3D7_0200400
PF3D7_0200900
PF3D7_0201300
PF3D7_0221400
PF3D7_0222800
PF3D7_0300400
PF3D7_0300900
PF3D7_0324600
PF3D7_0400800
PF3D7_0401500
PF3D7_0402600
PF3D7_0425500
PF3D7_0500600
PF3D7_0617600
PF3D7_0631900
PF3D7_0700400
PF3D7_0732000
PF3D7_0832000
PF3D7_0832400
PF3D7_0832600
PF3D7_0832900
PF3D7_0901600
PF3D7_1000800
PF3D7_1040200
PF3D7_1149900
PF3D7_1254100
PF3D7_1300900
PF3D7_1372500
PF3D7_1372800
PF3D7_1400700
PF3D7_1479500
PF3D7_1479900

F *phist*




V35
V48
V63
Cy5 Signal
60000
40000
20000
0
A
B
C


G
gbp
100000
75000
50000
25000
1000
750
500
250
0
Cy5 Signal
PF3D7_1016300 (gbp130)
PF3D7_1301200 (gbph2)
PF3D7_1401000 (gbph)
pNF54_I
pNF54_II
V18
V35
V48
V63

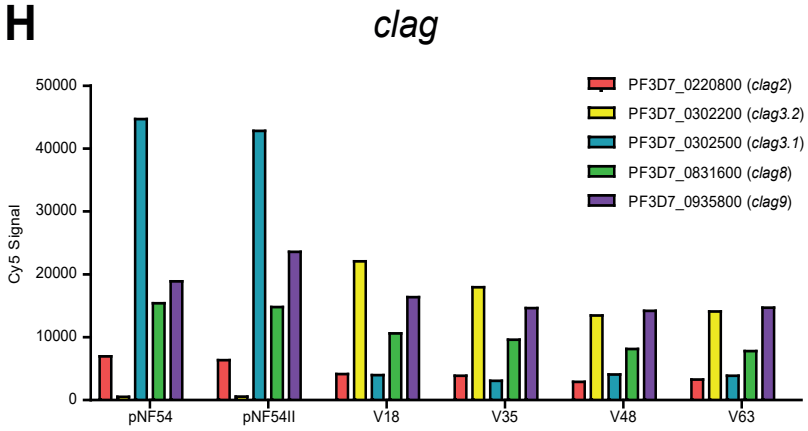

H
clag
50000
40000
30000
20000
10000
0
Cy5 Signal
PF3D7_0220800 (clag2)
PF3D7_0302200 (clag3.2)
PF3D7_0302500 (clag3.1)
PF3D7_0831600 (clag8)
PF3D7_0935800 (clag9)
pNF54
pNF54II
V18
V35
V48
V63

Fig. S2. Expression intensity in the CVG families with the largest transcriptional differences between vNF54 and pNF54 lines. (**A**) Normalized Cy5 signal for *pfmc-2tm* genes in the four vNF54 lines (V18, V35, V48, and V63) and two replicates of the pNF54 line. (**B**) Western blot analysis of pNF54 and V63 schizonts with antibodies against a specific member of the PFMC-2TM family (PF3D7_ 0114100) and the loading control heat shock protein 70 (HSP70), in three independent biological replicates (I, II and III). Bar chart shows the quantification of the relative PFMC-2TM band intensity (normalized by HSP70 band intensity). Data are presented as the average and SEM of three independent biological replicates. The difference between pNF54 and V63 was significant ($P<0.05$) using an unpaired two-sided t-test. (**C to H**) Same as panel A, but for genes of the *var* (**C**), *rif* (**D**), *stevor* (**E**), *phist* (**F**), *gbp* (**G**) and *clag* (**H**) families. Subfamilies are indicated for *var*, *rif* and *phist* genes. In panels A, C, D and H, expression values are at the time point of maximal expression of the gene family, i.e., 10 to 15 hpi for *var*, 20 to 25 hpi for *pfmc-2tm* and *rif*, and 30 to 35 hpi for *clag*. In panels E, F and G, values are at the time of maximal expression for each line and gene, because different genes of the *stevor*, *phist* and *gbp* families have a different time of maximal expression.

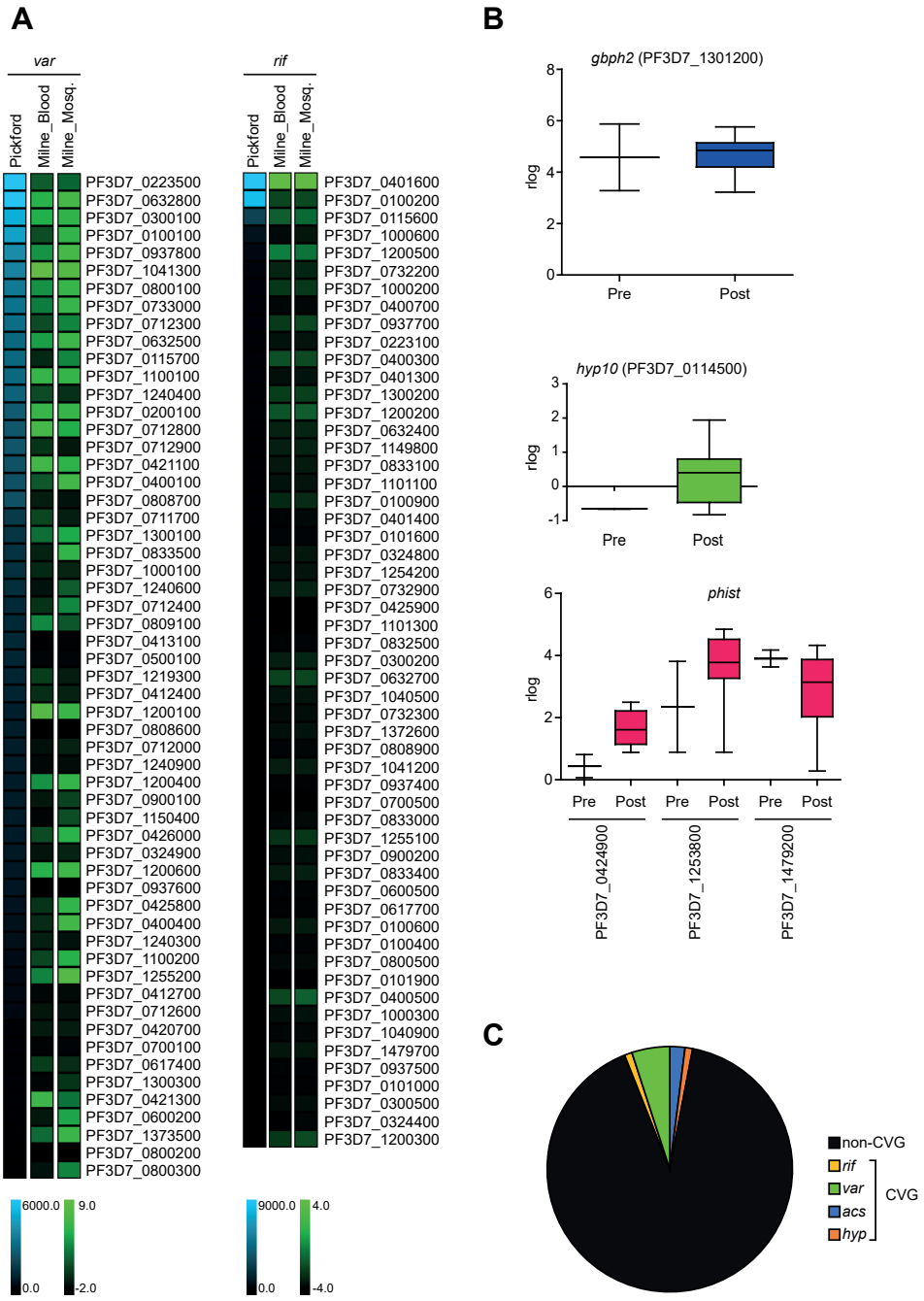

A
var
Pickford
Milne_Blood
Milne_Mosq.
PF3D7_0223500
PF3D7_0632800
PF3D7_0300100
PF3D7_0100100
PF3D7_0937800
PF3D7_1041300
PF3D7_0800100
PF3D7_0733000
PF3D7_0712300
PF3D7_0632500
PF3D7_0115700
PF3D7_1100100
PF3D7_1240400
PF3D7_0200100
PF3D7_0712800
PF3D7_0712900
PF3D7_0421100
PF3D7_0400100
PF3D7_0808700
PF3D7_0711700
PF3D7_1300100
PF3D7_0833500
PF3D7_1000100
PF3D7_1240600
PF3D7_0712400
PF3D7_0809100
PF3D7_0413100
PF3D7_0500100
PF3D7_1219300
PF3D7_0412400
PF3D7_1200100
PF3D7_0808600
PF3D7_0712000
PF3D7_1240900
PF3D7_1200400
PF3D7_0900100
PF3D7_1150400
PF3D7_0426000
PF3D7_0324900
PF3D7_1200600
PF3D7_0937600
PF3D7_0425800
PF3D7_0400400
PF3D7_1240300
PF3D7_1100200
PF3D7_1255200
PF3D7_0412700
PF3D7_0712600
PF3D7_0420700
PF3D7_0700100
PF3D7_0617400
PF3D7_1300300
PF3D7_0421300
PF3D7_0600200
PF3D7_1373500
PF3D7_0800200
PF3D7_0800300
6000.0
0.0
9.0
-2.0
rif
Pickford
Milne_Blood
Milne_Mosq.
PF3D7_0401600
PF3D7_0100200
PF3D7_0115600
PF3D7_1000600
PF3D7_1200500
PF3D7_0732200
PF3D7_1000200
PF3D7_0400700
PF3D7_0937700
PF3D7_0223100
PF3D7_0400300
PF3D7_0401300
PF3D7_1300200
PF3D7_1200200
PF3D7_0632400
PF3D7_1149800
PF3D7_0833100
PF3D7_1101100
PF3D7_0100900
PF3D7_0401400
PF3D7_0101600
PF3D7_0324800
PF3D7_1254200
PF3D7_0732900
PF3D7_0425900
PF3D7_1101300
PF3D7_0832500
PF3D7_0300200
PF3D7_0632700
PF3D7_1040500
PF3D7_0732300
PF3D7_1372600
PF3D7_0808900
PF3D7_1041200
PF3D7_0937400
PF3D7_0700500
PF3D7_0833000
PF3D7_1255100
PF3D7_0900200
PF3D7_0833400
PF3D7_0600500
PF3D7_0617700
PF3D7_0100600
PF3D7_0100400
PF3D7_0800500
PF3D7_0101900
PF3D7_0400500
PF3D7_1000300
PF3D7_1040900
PF3D7_1479700
PF3D7_0937500
PF3D7_0101000
PF3D7_0300500
PF3D7_0324400
PF3D7_1200300
9000.0
0.0
4.0
-4.0
B
gbph2 (PF3D7_1301200)
rlog
Pre
Post
hyp10 (PF3D7_0114500)
rlog
Pre
Post
phist
rlog
Pre
Post
Pre
Post
Pre
Post
PF3D7_0424900
PF3D7_1253800
PF3D7_1479200
C
non-CVG
rif
var
acs
hyp
CVG

Fig. S3. Analysis of CVG expression patterns in another CHMI study. (**A**) Expression levels of *var* and *rif* genes in our study (Pickford) and in the Milne *et al.* study (52), ordered from highest to lowest expression in our dataset. For our dataset, the median Cy5 signal of the vNF54 samples V18, V35, V48 and V63 at 10 to 15 hpi (*var*) or 20 to 25 hpi (*rif*) is shown. For the Milne study, the median of the Rlog expression value in volunteers (n = 12) infected with blood-stage parasites (Milne_Blood) and volunteers (n = 5) infected with sporozoites (Milne_Mosq.) is shown. Only genes for which data are available in the three datasets are shown. Eighteen genes had values out of the range of the color scale displayed. (**B**) CVG expression after different numbers of rounds of the IDC in the human circulation in the Milne et al. study. Box-and-whiskers plots showing the median Rlog values (boxes are quartiles, and whiskers are range) in the inoculum used to infect volunteers with blood-stage parasites ("Pre," corresponding to a blood sample collected at day 13 from a volunteer infected by mosquito bite, i.e., after liver development and three rounds of the IDC; n = 2 technical replicates) and in volunteers infected with the inoculum ("Post," corresponding to samples collected 7.5 to 10.5 days after infection by blood challenge with the inoculum; n=12 volunteers). Genes with large differences between pNF54 and vNF54 lines and peak transcript levels at the ring stage (the only stage available in the Milne dataset) were included. (**C**) Pie chart showing the distribution between clonally variant genes (CVGs; specific families indicated) and nonclonally variant genes among the top 100 genes with the largest changes in median Rlog values between "Pre" (inoculum) and "Post" samples in the Milne *et al.* study.

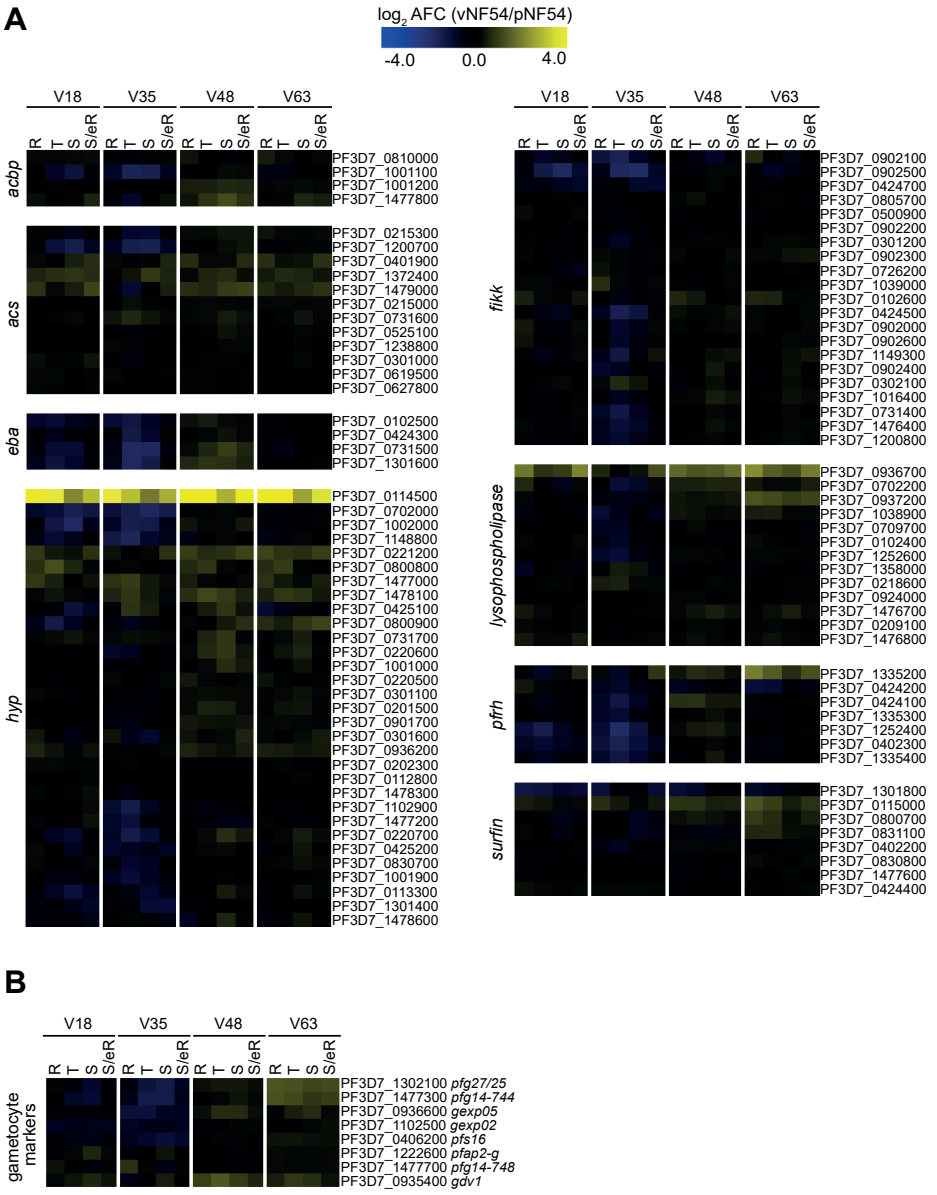


Fig. S4. Transcriptional patterns in clonally variant gene families that show few differences between vNF54 and pNF54 lines. (**A**) Transcript level changes in vNF54 relative to pNF54 lines, as in Fig. 2, for the clonally variant gene families indicated. Genes are ordered according to hierarchical clustering. One gene had values out of the color scale range displayed. (**B**) Same as in panel A, but for well-established markers of sexual commitment or early gametocytes.

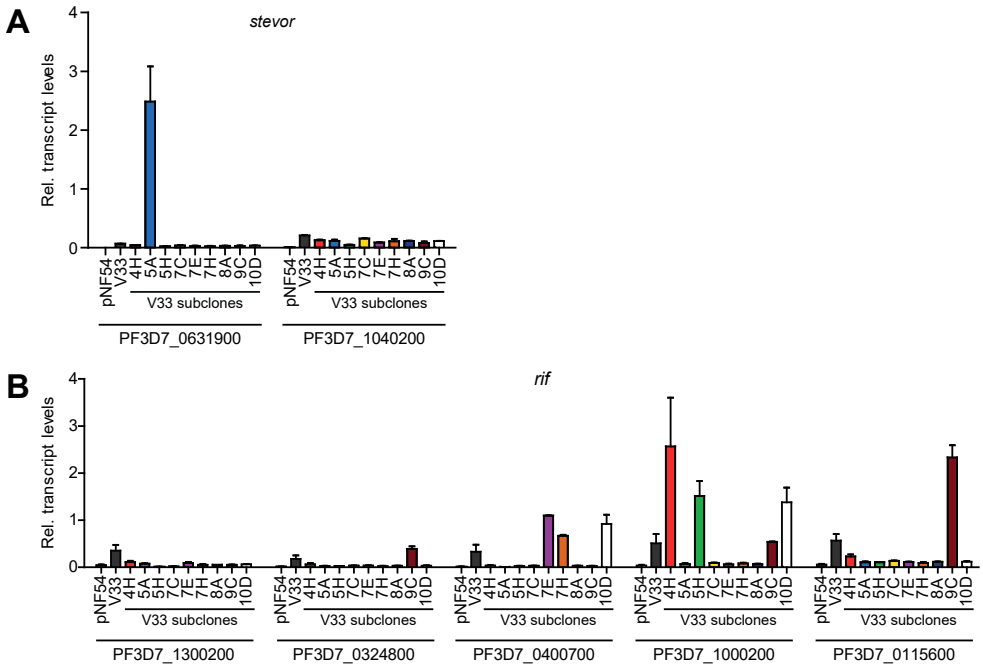


Fig. S5. Expression of stevor and rif genes in subclones of a vNF54 line. Transcript levels of selected *stevor* (**A**) and *rif* (**B**) genes in pNF54, the V33 vNF54 line and subclones of the V33 line at 20 to 25 h post-invasion (hpi). All *stevor* genes included in the analysis had maximal expression at the trophozoite stage. Transcript levels are normalized against *serine-tRNA ligase* (*serrs*). Data are presented as the average and SEM. of two independent biological replicates.

Article 3: Adaptation of *Plasmodium falciparum* to medium lacking a lipids supplement is associated with mutations in the *pfndh2* gene

Pickford AK, Michel-Todó L, & Cortés A. Adaptation of *Plasmodium falciparum* to medium lacking a lipids supplement is associated with mutations in the *pfndh2* gene. (**Under review**)

Adaptation of *Plasmodium falciparum* to medium lacking a lipids supplement is associated with mutations in the *pfndh2* gene.

Anastasia K. Pickford[1], Lucas Michel-Todó[1], Alfred Cortés[1,2*]

[1]ISGlobal, Hospital Clínic - Universitat de Barcelona, Barcelona, Catalonia, Spain
[2]ICREA, Barcelona, Catalonia, Spain
*For correspondence: alfred.cortes@isglobal.org

ABSTRACT

The complex life cycle of the malaria parasite *Plasmodium falciparum* entails many different environments. Additionally, the parasite encounters fluctuating conditions within the same niche. Clonally variant gene expression regulated at the epigenetic level provides an enormous potential for the adaptation of *P. falciparum* asexual blood stages to changes in their environment via bet-hedging strategies. However, currently there are very few examples of fluctuating conditions for which this type of adaptation has been demonstrated. Here we studied the adaptation of *P. falciparum* cultures to absence of an external source of lipids. In three independent selection experiments, cultures were able to adapt to lipid-free medium, but full transcriptome analysis did not reveal any changes in clonally variant gene expression consistently associated with adaptation. However, in the three selected lines the ability to grow in lipid-free medium was linked to acquisition of deletions or non-sense mutations in the *pfndh2* gene, which encodes a component of the mitochondrial electron transport chain. These results reveal an unexpected link between lipid metabolism and the electron transport chain, and suggest that adaptation to lipid-free medium involved metabolic rewiring.

INTRODUCTION

The protozoan parasite *Plasmodium falciparum* is responsible for the most severe form of malaria, a globally important disease that still causes over 200 million clinical cases and almost half a million deaths per year[1]. This pathogen has a complex life cycle that transitions between humans and anopheline mosquitoes, and also entails growth in various organs and tissues within each host. A malaria infection initiates when a female *Anopheles* mosquito injects parasites into a human host during a bloodmeal. After initial infection of the liver and exponential multiplication in this organ, parasites are released into the blood circulation, where they establish an intraerythrocytic development cycle (IDC) that is repeated periodically every 48 h. During this cycle, replication within red blood cells (RBCs) results in the generation of new parasites that are released back into the bloodstream, where they invade new RBCs and perpetuate the infection. At each cycle of replication, some parasites convert into sexual forms called gametocytes, which are the only form transmissible to mosquitoes. Within the mosquito, parasites undergo sexual reproduction and complex development and multiplication steps that eventually result in infection of the salivary glands, from where the parasites can be injected into another human host[2].

To deal with the environmental diversity their life cycle entails, malaria parasites use different transcriptional programmes at different developmental stages[3–5]. However, these pathogens are also confronted with fluctuating conditions within the same niche. The majority of fluctuating conditions occur during blood stages, which are responsible for long-term, chronic infection. Fluctuations occur both during the course of the same infection and between different individual hosts of the same species. For example, within the human blood circulation *P. falciparum* is exposed to changes in temperature during febrile episodes, in nutrient availability related to the host's physiological or nutritional state, in acquired immune responses and in presence of antimalarial drugs[6,7]. Consequently, similar to other organisms, malaria parasites have evolved mechanisms to adapt to the conditions of their environment[8]. Whilst genetic mutations (i.e., long-lasting changes in the primary DNA sequence) play a key role in adaptation to stable conditions and resistance to nearly all antimalarial drugs[9], recent research has shown that *P. falciparum* is also able to produce rapid directed transcriptional responses to some environmental conditions[10–12]. Additionally, *P. falciparum* uses an adaptive strategy based on epigenetic mechanisms that falls

between the other two with regards to speed and stability. This strategy involves clonally variant genes (CVGs), which in the absence of any selection can be found in either active or silent transcriptional states in parasites with the same genome and at the same stage of life cycle progression. Both states are transmitted from one generation of asexual parasites to the next by epigenetic mechanisms, with stochastic switches between the two states occurring with low-frequency[8,13-15]. In this manner, different individuals can present different CVG expression patterns, which constitutes a spontaneous source of transcriptional and phenotypical heterogeneity within parasite populations. Upon unexpected changes in the environment, natural selection can act on this pre-existing variability and select individual parasites with CVG expression patterns that confer high fitness under the new conditions. This is a bet-hedging adaptive strategy[16], a survival mechanism commonly observed in many microbial species[17-20]. Bet-hedging adaptive strategies, which dispense the need to invest in costly sensing mechanisms, are considered specially favourable for the adaptation of unicellular populations to environment conditions that change infrequently and in unpredictable ways[21].

In *P. falciparum*, the adaptive role of CVGs has been clearly demonstrated for two specific processes: antigenic variation to evade the host's immune response[22-24] and changes in infected RBC membrane permeability to acquire resistance to some antimalarial compounds[25-28]. Variation in the use of alternative RBC invasion pathways and in sexual conversion rates, which are processes linked to CVGs, is also likely to play important adaptive roles[13,15]. Nevertheless, given the high number of CVGs in the *P. falciparum* genome[13,16], it is likely that adaptation to several other fluctuating conditions also involves CVGs.

One fluctuating condition to which malaria parasites may adapt using bet-hedging strategies and CVGs is nutrient availability in the blood. In particular, *P. falciparum* may use this strategy to adapt to changes in the availability of lipids in the human host blood, because several CVGs are linked to lipid metabolism. Among them are a *lysophospholipase* (PF3D7_0936700)[16,29] and genes of the acyl-CoA synthetase (*acs*) and acyl-CoA binding protein (*acbp*) CVG families[16], comprising 13 and 4 genes, respectively. Genes of the *acs* family catalyse the activation of fatty acids to acyl-CoA and facilitate their transport across biological membranes[30,31]. *P. falciparum* has a large demand of lipids for membrane biogenesis, but in malaria endemic areas, where malaria and malnutrition often coexist[32], parasites often come across low concentrations of lipids

in the human blood. While some lipids are synthesised by the parasite, others are scavenged from the host. The latter can be obtained from the RBC membrane or from the extracellular environment[33,34]. Under culture conditions, the only essential lipids that must be externally supplied to *P. falciparum* are fatty acids, which are key building blocks for membrane lipid synthesis and for other processes. The minimal combinations of externally supplied fatty acids that can support normal parasite growth have been defined, and they must include at least oleic acid and a saturated fatty acid[35–38]. *Plasmodium spp.* genomes encode all the enzymes necessary to synthesise fatty acids, via a FASII pathway located in a non-photosynthetic plastid organelle called apicoplast, but during blood stages this pathway is inactive (at least under nutrient-rich conditions) and scavenging from the host is the primary source of fatty acids[31,33,39–41]. Indeed, FASII enzymes are dispensable and expressed at very low levels during asexual blood stages, and they are essential only for mosquito or liver development[42–44].

Here we studied how *P. falciparum* asexual blood stages adapt to low availability of lipids in the extracellular environment and tested the hypothesis that CVGs and bet-hedging strategies play a role in the adaptive process. For this end, we performed in vitro selection of *P. falciparum* cultures with lipid-free media and analysed the adapted populations by whole genome sequencing and transcriptomic analysis. While we did not observe consistent changes in the expression of CVGs in the adapted population, in three independent selections we unexpectedly found that adaptation to lipid starvation was associated with deletion or truncation of an enzyme of the mitochondrial electron transport chain, suggesting a link between lipid metabolism and mitochondrial function.

RESULTS

P. falciparum can adapt to grow in the absence of a lipids supplement in the culture medium

The standard *P. falciparum* culture medium consists of an RPMI-based solution supplemented with either 10% (v/v) human serum or 0.5% (w/v) AlbuMAX II, a lipid-rich bovine serum albumin (BSA), as the only source of albumin and lipids such as cholesterol and fatty acids[45–47]. Both supplements support long-term asexual parasite cultures with similar

multiplication rates. Whilst human serum resembles in vivo conditions more closely and provides some additional nutrients (e.g., AlbuMAX II medium requires addition of hypoxanthine[45,46,48]), AlbuMAX II has a lower cost and lower batch to batch variability, often making it the preferred option for *P. falciparum* cultures[45,46].

To determine if *P. falciparum* cultures can adapt to grow in the absence of exogenous lipids, we selected cultures of the transcriptionally-diverse 3D7-A line[16,49] with media containing no lipids. We first attempted to adapt cultures to grow without any supplement, but this was not possible, likely because albumin is essential for parasite growth[35]. In contrast, when cultures were selected with medium supplemented with 0.5% (w/v) lipid-free bovine serum albumin (low endotoxin, lyophilised, Cohn Fraction V, ≥98% purity) and no human serum or AlbuMAX II as a source of lipids, hereafter called lipid-free medium (LF), adapted cultures were readily achieved. In three independent selection experiments, cultures were selected in LF medium for 24-60 days in parallel with control cultures maintained with 0.5% AlbuMAX II (standard concentration), with dilution to a 1% parasitaemia at the beginning of each cycle (ring stage) and parasitaemia measured by light microscopy analysis of Giemsa-stained smears (selections 1 and 3) or flow cytometry (selection 2) every 48 h (before diluting the cultures) (**Fig. 1**, **Supplementary Fig. 1**). We observed that cultures in LF medium initially multiplied at relatively high rates for two cycles (~60% of controls), then growth was severely compromised for several cycles (multiplication rates ≤20% of controls) and, from cycle 8-10 onwards, relatively high growth rates were observed again (30-80% of control cultures). These results demonstrate that *P. falciparum* cultures can adapt to stable, long term-growth in the absence of a lipids supplement (human serum or Albumax II), albeit with multiplication rates always lower than in lipid-supplemented medium. While in selection 1 lipids derived from Albumax II were present in the 50% RBCs solution used to dilute cultures (final Albumax II concentration <0.015%, as the haematocrit was 3%), in selections 2 and 3 the RBCs solution was spun and resuspended in LF medium before use. Therefore, in selections 2 and 3 only traces of lipids derived from Albumax II were present in the culture medium.

Validation of the LF-selected lines

To validate that the selected lines were adapted to grow in LF medium, we used several different approaches. We tested the parasite line obtained after selection 2 (3D7-A-LF-2) and the line maintained in parallel under

control conditions (3D7-A-Ctrl-2), because this was the longest selection experiment and the selected line obtained had the highest multiplication rate in LF medium. First, we compared growth of 3D7-A-LF-2 and 3D7-A-Ctrl-2 over four replication cycles under either standard 0.5% AlbuMAX II or LF conditions (**Fig. 2A**). Cultures were diluted to 1% parasitaemia at the ring stage of each cycle and multiplication rates were measured by flow cytometry. Multiplication rates of both lines remained stable over the four cycles in cultures with 0.5% AlbuMAX II medium. In contrast, in cultures with LF medium, multiplication rates dropped significantly in the 3D7-A-Ctrl-2 line from cycle 3 onwards, but not in the 3D7-A-LF-2 line. In a second approach, cultures of the same two parasite lines were diluted to 0.1% or 0.01% parasitaemia and left undisturbed (no medium changes) in the two different media for two or three cycles, respectively, before measuring parasitaemia by flow cytometry (**Fig. 2B**). In both the 3D7-A-LF-2 and the 3D7-A-Ctrl-2 lines, growth was much lower in LF than in 0.5% AlbuMAX II medium, and the reduction of growth was similar between the two lines. Therefore, this result raised doubts about the level of adaptation of the 3D7-A-LF-2 line to grow in LF medium. In a third approach, rather than maintaining the 3D7-A-LF-2 and 3D7-A-Ctrl-2 lines continuously in the two different types of media, cultures were continuously maintained in LF medium and, at the ring stage of each replication cycle, cultures were diluted to 1% parasitaemia and split into dishes with 0.5% AlbuMAX II or LF medium to compare growth under the two conditions (**Fig. 2C**). After two cycles, the multiplication rate of the 3D7-A-Ctrl-2 line dropped severely even in the dishes with 0.5% AlbuMAX II, indicating that the absence of lipids had produced irreversible damage. In contrast, stable growth was observed throughout the experiment for the 3D7-A-LF-2 line.

Altogether, these results indicate that the selected lines adapted to growth in the absence of a lipids supplement and could grow under these conditions for long periods of time. While the selected lines could withstand LF conditions clearly better than control lines, adaptation was somehow unstable and the ability to grow in LF medium depended on factors such as the regularity of media changes and/or the initial parasitaemia of the cultures, and multiplication rates were always lower than in lipid-rich medium. We did not specifically validate that the lines obtained after selections 1 and 3 were adapted, but given that they also grew continuously in LF medium for multiple generations and they showed a similar growth dynamic to 3D7-A-LF-2 during selection (**Fig. 1**), we consider that they are indeed adapted. Furthermore, they show molecular defects in the same gene as 3D7-A-LF-2 (see below).

Transcriptomic characterisation of the LF-selected lines

To determine the molecular mechanism of adaptation to LF conditions, we performed a transcriptomic analysis across the full IDC of LF-selected and control lines from selections 2 and 3 (3D7-A-Ctrl-2, 3D7-A-LF-2, 3D7-A-Ctrl-3 and 3D7-A-LF-3) (**Fig. 3A**). For each selection experiment, three cultures were tested: control line (always maintained in 0.5% AlbuMAX II), LF0 (LF-selected cultures always maintained in LF medium), and LFCM (LF-selected cultures maintained in medium with 0.5% AlbuMAX II from the beginning of the cycle of RNA harvesting). The latter condition was included to distinguish potential directed transcriptional responses from stable transcriptome alterations independent from the external cue. We reasoned that if a transcriptional alteration (relative to control cultures) was the result of a directed response that is active only while the external cue (absence of lipids) is maintained, it would be observed only in LF0 cultures, whereas stable alterations would be observed in both LF0 and LFCM cultures. Cultures were tightly synchronised to a 5 h age window and RNA collected every eight hours (8-13, 16-21, 24-29, 32-37 and 40-45 h post-invasion [hpi]) for transcript levels analysis using custom two-channel long oligonucleotide microarrays[50]. All samples were hybridised against a common reference pool to obtain relative expression values. For each gene, the maximum average fold change among overlapping time intervals of half the duration of the IDC (mAFC) was calculated[16,29] (see Methods) to quantify the transcript level changes between the different conditions (Ctrl, LF0 and LFCM) (**Supplementary Table 1**).

When comparing the transcriptome of LF-selected vs control lines, there was a large overlap between LF0 and LFCM cultures in the genes that showed a mAFC>4 in either direction [absolute value of the $\log_2$(mAFC) >2] (**Fig. 3B-C**). As expected from this result, very few genes (mainly spliceosomal RNAs differentially expressed in selection 3) showed large differences between LF0 and LFCM conditions (**Fig. 3B** bottom). The significance of the differences in the expression of these few genes between LF0 and LFCM is unclear, because they occurred only in selection 3. These results indicate that the vast majority of transcriptional differences between control and selected lines reflect stable alterations rather than directed responses that depend on presence or absence of the stress condition (lipids deprivation).

Unexpectedly, there was very little overlap between the genes differentially expressed in selections 2 and 3 (in any of the possible pairwise comparisons between Ctrl, LF0 and LFCM) (**Fig. 3B,D**). The most differentially expressed genes in LF-selected vs control lines from selection 2 did not show large transcript levels changes in selection 3. Similarly, the genes with the most pronounced transcriptional changes in selection 3, many of which were non-coding RNAs and genes encoded in the mitochondrial genome, did not show differential expression in selection 2 (**Fig. 3B**). As expected from these results, analysis of the differentially expressed genes between the different conditions treating the two selection experiments as replicates, which enables statistical analysis, revealed only one significantly differentially expressed gene (mAFC > 4, Pval < 0.05) in the Ctrl vs LF0 and Ctrl vs LFCM comparisons (*etramp12*, PF3D7_1240100), and none in the LF0 vs LFCM comparisons (**Supplementary Fig. 2**).

In both selection experiments, some of the differentially expressed genes between the different conditions were CVGs of the *var*, *rif*, *stevor* and *pfmc-2tm* families linked to antigenic variation (**Fig. 3B**). This observation most likely reflects the population bottleneck experienced by LF-selected lines during adaptation, as these genes are unlikely to be directly linked to survival in lipid-free medium. None of the genes with large transcriptional differences in either selection experiment were predicted to be related with lipid metabolism, with the exception of a CVG of the *acbp* family (PF3D7_1477800) that was upregulated in LF-selected cultures only in selection 3. Furthermore, gene set enrichment analysis (GSEA) did not reveal any metabolic pathway related with lipid metabolism that was enriched in genes differentially expressed in both selection experiments (**Supplementary Table 2**). The majority of enriched pathways were related with CVGs. Closer inspection of genes linked to the apicoplast-based FASII pathway for fatty acids synthesis revealed no obvious changes common to both replicates (**Supplementary Fig. 3**).

Together, these results show that different LF-selected cultures had different transcriptional alterations, and none of the alterations is clearly linked to lipid metabolism. Therefore, a causal relationship cannot be established between the transcriptional changes observed and adaptation to LF conditions.

Genetic changes in cultures adapted to LF conditions

Since the transcriptomic analysis did not reveal any changes consistently associated with adaptation to LF conditions, we sequenced the whole genome of control and LF-selected lines to explore the possibility that adaptation was linked to selection of specific genetic variants. In the 3D7-A-LF-2 line, there was a large deletion including the PF3D7_0914900 (BSD-domain protein) and PF3D7_0915000 (type II NADH:ubiquinone oxidoreductase [*pfndh2*]) genes that did not occur in any of the other selected or control parasite lines (**Fig. 4A**). This deletion explains the severely reduced transcript levels of these two genes in the 3D7-A-LF-2 line (**Fig. 3B**). Analysis of SNPs and small indels (**Supplementary Table 3**) revealed no mutations common to all three LF-selected lines, but both 3D7-A-LF-1 and 3D7-A-LF-3 carried an insertion of an adenine at position 898 (codon 300) of the coding region of *pfndh2* (deleted in 3D7-A-LF-2) just after a track of 11 consecutive adenines, which resulted in a frameshift leading to a premature stop codon and a truncated protein of approximately half its original size (**Fig. 4B**). This mutation, which was validated by Sanger sequencing (Supplementary Fig. 4), was present in approximately 80-90% of the reads from LF-selected parasites and almost completely absent from the control lines (3D7-A-Ctrl-2 and 3D7-A-Ctrl-3) (**Fig. 4C**). However, close inspection of the reads at this position revealed the presence of rare reads carrying the insertion in both control lines (<1% frequency), indicating that this mutation was possibly present in the parental population, albeit at very low frequency, and this enabled rapid selection of mutant parasites during the adaptation process. Together, our full genome sequencing analysis revealed that, in the three independent LF-selection experiments, adaptation to lipids deprivation was linked to loss-of-function of PfNDH2, via either a deletion or a frameshift mutation. PfNDH2 catalyses the transfer of electrons from NADH to ubiquinone as part of the mitochondrial electron transport chain (**Fig. 4D**)[51,52].

Analysis of mutations from publicly available genome sequences from field isolates (http://www.malariagen.net)[53,54] showed that the very same mutation (insertion at position 898) and a small number of other non-sense mutations in *pfndh2* occur in some isolates, although with very low allelic frequency in the population and always with an incomplete prevalence in the isolates where they occurred (i.e., mutant alleles always coexisted with wild type alleles) (**Supplementary Table 4**). Polymorphism in the length of the adenines track where we observed the insertion mutation in the LF-selected lines was the most common *pfndh2* mutation in field isolates.

DISCUSSION

Here we demonstrate that *P. falciparum* can adapt to grow in the absence of any external source of lipids. Unexpectedly, we found that adaptation was associated with selection of parasites carrying mutations in the *pfndh2* gene. The observation that in three independent selection experiments mutants for this gene were selected, including two independent genetic defects (a large deletion and a single nucleotide insertion), strongly suggests a causal relationship between loss of function of this gene and the ability of the parasites to thrive in medium without lipids.

While adaptation to several fluctuating conditions in the human blood likely depends on bet-hedging strategies[8,13,15,16], we did not observe transcriptional changes in the expression of CVGs consistently associated with adaptation to LF medium. The transcriptional changes observed were different between replicate selection experiments and mainly affected genes unrelated to lipids metabolism. Therefore, we did not obtain any evidence supporting our initial hypothesis, which was that adaptation to absence of an external source of lipids may occur by changes in the expression of CVGs and bet-hedging strategies. Currently, we cannot completely rule out the possibility that adaptation to this condition may occur by changes in the expression of CVGs but in our experiments this was masked by more efficient adaptation by selection of the *pfndh2* mutants. However, we consider this an unlikely possibility; if this was the case, we would expect that specific changes in the expression of CVGs would provide an advantage additional to the mutations and they would be consistently observed in our long-term selected lines.

One study reported that *P. falciparum* cultures multiplied normally with a concentration of AlbuMAX II as low as 0.01% (50-fold lower than the standard concentration) over short periods of time (one week)[48], but the majority of previous studies found that supplementation of the culture medium with at least some specific fatty acids, in addition to albumin[35-37], was essential for *P. falciparum* cultures. Therefore, it is surprising that our adapted lines were capable of growing in the absence of any external source of lipids for up to 60 days, and in the line obtained in selection 2 multiplication rates were ~80% as high as in regular medium. A fundamental difference with previous studies is that we selected cultures with LF medium for a long period of time, whereas the majority of previous studies assessed the requirement for specific lipids for a small number of

cycles that do not provide the opportunity for parasite adaptation. However, in one study cultures were maintained with combinations of fatty acids that do not support normal growth for up to 28 days and no indication of adaptation to these conditions was observed[36]. We hypothesise that the discrepancy may be due to the fact that our parental population contained rare parasites carrying mutations in the *pfndh2* gene, which facilitated rapid adaptation, whereas such parasites may be absent from the starting population used in the previous studies.

The genome of *P. falciparum* encodes all genes needed to synthesise fatty acids via a FASII pathway located in the apicoplast. However, this pathway is essential only for sporozoite and liver stages, whereas in blood stage parasites it is dispensable and possibly inactive, as fatty acids are scavenged from the host [31,33,39,41-44]. One study found that under conditions of limited availability of lipids in the media, the activity of the FASII pathway is rapidly increased, possibly by posttranscriptional mechanisms[40]. Activation of the FASII pathway under lipids deprivation conditions may contribute to parasite survival in LF medium, and if it depends on posttranscriptional modifications, as proposed, it would have escaped detection by our transcriptomic analysis. However, rapid posttranscriptional activation of a metabolic pathway alone is not consistent with the adaptation pattern that we observed, which required >8 growth cycles. Therefore, the adaptation of our parasite lines to survive in LF medium involved other changes, as revealed by full genome sequencing, notwithstanding the possibility that posttranslational activation of the FASII pathway may have occurred.

Unexpectedly, adaptation to absence of external lipids was linked to selection of parasites with non-functional PfNDH2. This enzyme, anchored in the mitochondrial cristae membrane, participates in the mitochondrial electron transport chain. Its main function is to transfer an electron from NADH to ubiquinone, and it has no proton-pumping activity. PfNDH2 was initially considered a promising antimalarial drug target, but later research showed that it is essential only in mosquito stages and non-essential in intraerythrocytic blood stages, and therefore not a good target for drug development[51,52]. Furthermore, the mitochondrial electron transport chain is intact in parasites in which *pfndh2* was disrupted[51,52], likely because four other dehydrogenases can transfer electrons from different substrates to ubiquinone (**Fig. 4D**). The only essential role of the malarial electron transport chain in blood stage parasites is to maintain the continuous reoxidation of ubiquinol to ubiquinone to sustain the activity of dihydroorotate dehydrogenase (DHOD), which is essential for de novo

pyrimidine synthesis[55]. In *Plasmodium* blood stages, energy metabolism depends mainly on glucose degradation by glycolysis in the cytoplasm, and the mitochondrial tricarboxylic acids cycle (TCA), which requires an active electron transport chain, is dispensable[56,57]. In fact, *P. falciparum* mitochondria appear to lack cristae during asexual blood stages[58].

Because PfNDH2 is essential during transmission stages, mutations in this gene are in principle unlikely to mediate adaptation to lipid restriction in natural malaria infections. However, the occurrence of the insertion mutation in an homopolymeric track of adenines and the observation that this mutation appears to occur in many field isolates in a low proportion of the parasites raises the intriguing possibility that this region may display hypermutability, such that new alleles with different number of adenines are constantly generated. In this scenario, frameshift alleles resulting in truncated PfNDH2 would be transiently selected under severe lipid restriction, but new variants expressing the complete protein would constantly arise, securing transmission.

While PfNDH2 is not necessary for the activity of the mitochondrial electron transport chain because other enzymes can provide electrons to ubiquinone, PfNDH2 is the only known enzyme that uses NADH as an electron donor. Therefore, disruption of the activity of PfNDH2 may alter the balance between NADH and its oxidised form, NAD+. We speculate that this may in turn alter metabolic fluxes in the mitochondria and in the full cell. Since fatty acids synthesis via the FASII pathway requires reduced NADH (or NADPH)[31,39], absence of an enzyme that recycles NADH to NAD+ may promote this pathway and enable parasite survival when extracellular fatty acids are not available. Malaria parasites have a remarkable metabolic plasticity[56], which makes it plausible that changes in a mitochondrial enzyme affect the whole metabolic network of the parasite. Future studies using metabolomics and other approaches will be needed to confirm this hypothesis and to precisely define the mechanistic link between mutations in PfNDH2 and adaptation to lipids deprivation.

MATERIALS AND METHODS

Parasite cultures

The *P. falciparum* line 3D7-A[59] was cultured in B+ RBCs [obtained from the Catalan official blood bank ('Banc de Sang i Teixits') after ethical approval by the Hospital Clínic (Barcelona) ethics committee] at 3% haematocrit in RPMI-1640-based culture medium containing standard concentrations of $NaHCO_3$, HEPES, hypoxanthine and gentamycin, and supplemented either with 0.5% AlbuMAX II (control conditions) or with 0.5% lipid-free BSA (Sigma # A1470) (LF conditions). Cultures were maintained in static conditions in a 5% CO_2, 3% O_2 and 92% N_2 atmosphere.

Selection and validation of adaptation experiments

For the selection experiments parasites were maintained either in control (0.5% AlbuMAX II supplemented media) or LF medium for the duration of the selection experiment, which varied between 30 and 60 days. At the beginning of every cycle, parasitaemia was adjusted to 1% by dilution with uninfected RBCs, and 48 h later (i.e., before the next dilution) parasitaemia was determined again. To determine parasitaemia, in the first and third selection experiments we used light microscopy examination of Giemsa-stained smears, whereas in the second selection we used flow cytometry. For flow cytometry measurements, we used SYTO 11 to stain nucleic acids and a FACScalibur flow cytometer (Becton Dickinson), as previously described[60].

To validate the LF-adapted lines, cultures were maintained in either 0.5% AlbuMAX II or LF medium, with dilution to 1% parasitaemia at the beginning of each cycle and determination of parasitaemia 48 h later using flow cytometry (validation experiment 1) or light microscopy (validation experiment 3), or dilution to 0.1 or 0.01% parasitaemia at the beginning of the assay and flow cytometry determination of the final parasitaemia two or three cycles later, respectively (validation experiment 2).

Transcriptomic analysis using microarrays

Time-course transcriptomic analysis was performed using cultures tightly synchronised to a 5 h age window by Percoll purification followed

by sorbitol lysis 5 h later[61]. After Percoll purification, cultures of LF-selected lines were split into two flasks with different conditions: LF0 that was always maintained in LF media and LFCM that was maintained in AlbuMAX II-supplemented media for the rest of the experiment. Control cultures were always kept in AlbuMAX II supplemented media. After sorbitol lysis, 0-5 hpi cultures from each condition (Ctrl, LF0 and LFCM) were divided into five separate dishes that were left undisturbed for different times (8, 16, 24, 32 and 40 h) until RNA was collected in TRIzol and frozen at -80ºC. RNA purification and cDNA synthesis and labelling by reverse transcription was performed as previously described[50]. For microarray hybridization, we used two-colour, long oligonucleotide-based, custom Agilent microarrays with a design based on Agilent AMADID #037237[50] modified by the addition of new probes for genes without unique probes and some reporter genes, and reannotated (new designs: AMADID-084561 and AMADID-085763)[62]. The different samples were labelled with Cy5 and hybridised against a common reference pool labelled with Cy3, as previously described[50]. The reference pool consisted of a mixture of equal amounts of cDNA from rings, trophozoites and schizonts of the 3D7-A line. Finally, images were acquired with a Microarray Scanner (no. G2505C, Agilent Technologies) placed in a low ozone hood.

Microarray data analysis

The microarray data was initially processed using the default options of the Feature Extraction software (Agilent) that included linear and LOWESS normalisation. The next steps of the analysis were carried out using Bioconductor in an R environment (R version 3.5.3). The background signal was calculated for each individual sample and channel (Cy3 and Cy5) as the median of the 100 lowest signal probes, and probes with both Cy3 and Cy5 signals below three times the array background were eliminated from further analysis. Both statistical estimation of parasite age (in hpi)[63] and gene level $\log_2$(Cy5/Cy3) were computed as previously described[16,29]. Next, estimated age vs $\log_2$(Cy5/Cy3) plots were generated for each gene and divided into four overlapping time intervals of the same-length that roughly corresponded to the ring, trophozoite, schizont and late schizont/early ring stages[16,29]. Using the difference in the area under the curve in the $\log_2$(Cy5/Cy3) vs estimated age plots, the average expression fold-change (AFC) was calculated between the three different conditions (Control, LF0 and LFCM) for each gene and time interval. The time interval at which the AFC had the highest absolute value was used as the maximum AFC (mAFC). All

tRNAs were excluded from further analysis, and genes whose expression intensity (Cy5 channel) values fell within the lowest 20th percentile in all samples were excluded from the lists of differentially expressed genes because differences in genes expressed at near background levels are of low confidence. Expression differences based on potentially cross-reactive probes that recognise groups of related genes (e.g., probes that recognise several rRNA genes) and expression differences in ncRNAs not annotated in PlasmoDBv53 are listed in Supplementary Table 1 but were excluded from Fig. 3. TMEV 4.9 was used to generate heatmaps with hierarchical clustering based on Euclidean distance[64] and gene set enrichment analysis (GSEA) was performed using GSEA v3.0 Preranked[65].

Whole genome sequencing and data analysis

The whole genome of the LF-selected lines 3D7-A-LF-1, 3D7-A-LF-2 and 3D7-A-LF-3 and the control lines 3D7-A-Ctrl-2 and 3D7-A-Ctrl-3 were sequenced via PCR-free Illumina sequencing approximately as described[29]. In brief, after shearing DNA to ~150-400 bp with a Covaris S220 ultrasonicator, library preparation was performed with the NEBNext® Ultra™ DNA Library Prep Kit for Illumina using specific paired-end TruSeq Illumina adaptors for each sample. The End Repair incubation step at 65°C was omitted due to the high AT content of the *P. falciparum* genome. Libraries underwent quality control using the Bioanalyzer DNA High Sensitivity kit (Agilent), followed by quantification using the KAPA Library Quantification Kit (Roche). The Illumina HiSeq2500 System was used to sequence the libraries and over 6 million 125 bp paired-end reads were obtained for each sample.

Raw data was processed by checking the read quality (FastQC program) and trimming the adaptors (Cutadapt program), before the reads were mapped to the PlasmodDB *P. falciparum* 3D7 reference genome version 46 (https://plasmodb.org/plasmo/) using the Bowtie2 local alignment algorithm. Next, to identify SNPs and small indels based on GATK best practices, we performed variant calling using the GATK-UnifiedGenotyper. Variants with low calling quality (Phred QUAL<20) or low read depth (DP<20) were filtered out with the GATKVariant Filtration tool. In addition, for each SNP/indel the differences in frequency between control and LF-selected lines were calculated and only those showing >50% difference were retained. Genome Browse (Golden Helix) was used to visualise alignments and variants as well as to detect large deletions.

Preparation of samples for Sanger Sequencing

A fragment of the PF3D7_0915000 gene including the mutation detected by Illumina sequencing was amplified by PCR using primers *pfndh2*_Fw (CAGGTGTCGAAGTTACCGC) and *pfndh2*_Rv (AAGTATATCCAAATGGGATTATG), and sequenced with primer *pfndh2*_Fw.

Data availability

The microarray and whole genome sequencing data have been deposited in the GEO and SRA databases with accession numbers GSE183526 (token for reviewer access: irubumimlrghpof) and PRJNA759734 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA759734), respectively.

REFERENCES

1. World_Health_Organization. World malaria report 2020. ISBN 978-992-974-001579-001571 (2020).
2. Phillips, M. A. *et al.* Malaria. *Nat Rev Dis Primers* **3**, 17050 (2017).
3. Bozdech, Z. *et al.* The Transcriptome of the Intraerythrocytic Developmental Cycle of *Plasmodium* falciparum. *PLoS Biol* **1**, E5 (2003).
4. Howick, V. M. *et al.* The Malaria Cell Atlas: Single parasite transcriptomes across the complete *Plasmodium* life cycle. *Science* **365** (2019).
5. Le Roch, K. G. *et al.* Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science* **301**, 1503-1508 (2003).
6. Mackinnon, M. J. & Marsh, K. The selection landscape of malaria parasites. *Science* **328**, 866-871 (2010).
7. Zuzarte-Luis, V. & Mota, M. M. Parasite Sensing of Host Nutrients and Environmental Cues. *Cell Host Microbe* **23**, 749-758 (2018).
8. Llora-Batlle, O., Tinto-Font, E. & Cortes, A. Transcriptional variation in malaria parasites: why and how. *Brief Funct Genomics* **18**, 329-341 (2019).
9. Haldar, K., Bhattacharjee, S. & Safeukui, I. Drug resistance in *Plasmodium*. *Nat Rev Microbiol* **16**, 156-170 (2018).
10. Brancucci, N. M. B. *et al.* Lysophosphatidylcholine Regulates Sexual Stage Differentiation in the Human Malaria Parasite *Plasmodium falciparum*. *Cell* **171**, 1532-1544 (2017).
11. Mancio-Silva, L. *et al.* Nutrient sensing modulates malaria parasite virulence. *Nature* **547**, 213-216 (2017).

12. Tintó-Font, E. *et al.* A heat-shock response regulated by the PfAP2-HS transcription factor protects human malaria parasites from febrile temperatures. *Nat Microbiol* **6**, 1163-1174 (2021).
13. Cortés, A. & Deitsch, K. W. Malaria Epigenetics. *Cold Spring Harb Perspect Med* **7**, a025528 (2017).
14. Guizetti, J. & Scherf, A. Silence, activate, poise and switch! Mechanisms of antigenic variation in *Plasmodium falciparum*. *Cell Microbiol* **15**, 718-726 (2013).
15. Voss, T. S., Bozdech, Z. & Bartfai, R. Epigenetic memory takes center stage in the survival strategy of malaria parasites. *Curr Opin Microbiol* **20**, 88-95 (2014).
16. Rovira-Graells, N. *et al.* Transcriptional variation in the malaria parasite *Plasmodium falciparum*. *Genome Res* **22**, 925-938 (2012).
17. Levy, S. F., Ziv, N. & Siegal, M. L. Bet hedging in yeast by heterogeneous, age-correlated expression of a stress protectant. PLoS *Biol* **10**, e1001325 (2012).
18. Starrfelt, J. & Kokko, H. Bet-hedging, a triple trade-off between means, variances and correlations. *Biol Rev Camb Philos Soc* **87**, 742-755 (2012).
19. Veening, J. W., Smits, W. K. & Kuipers, O. P. Bistability, epigenetics, and bet-hedging in bacteria. *Annu Rev Microbiol* **62**, 193-210 (2008).
20. Satory, D., Gordon, A. J., Halliday, J. A. & Herman, C. Epigenetic switches: can infidelity govern fate in microbes? *Curr Opin Microbiol* **14**, 212-217 (2011).
21. Kussell, E. & Leibler, S. Phenotypic diversity, population growth, and information in fluctuating environments. *Science* **309**, 2075-2078 (2005).
22. Scherf, A., Lopez-Rubio, J. J. & Riviere, L. Antigenic variation in *Plasmodium falciparum*. *Annu Rev Microbiol* **62**, 445-470 (2008).
23. Smith, J. D., Rowe, J. A., Higgins, M. K. & Lavstsen, T. Malaria's deadly grip: cytoadhesion of *Plasmodium falciparum*-infected erythrocytes. *Cell Microbiol* **15**, 1976-1983 (2013).
24. Wahlgren, M., Goel, S. & Akhouri, R. R. Variant surface antigens of *Plasmodium falciparum* and their roles in severe malaria. *Nat Rev Microbiol* **15**, 479-491 (2017).
25. Mira-Martinez, S. *et al.* Identification of Antimalarial Compounds That Require CLAG3 for Their Uptake by *Plasmodium falciparum*-Infected Erythrocytes. *Antimicrob Agents Chemother* **63**, e00052-00019 (2019).
26. Mira-Martínez, S. *et al.* Epigenetic switches in *clag3* genes mediate blasticidin S resistance in malaria parasites. *Cell Microbiol* **15**, 1913-1923 (2013).
27. Mira-Martínez, S. *et al.* Expression of the *Plasmodium falciparum* Clonally Variant *clag3* Genes in Human Infections. *J Infect Dis* **215**, 938-945 (2017).
28. Sharma, P. *et al.* An epigenetic antimalarial resistance mechanism involving parasite genes linked to nutrient uptake. *J Biol Chem* **288**, 19429-19440 (2013).
29. Pickford, A. K. *et al.* Expression Patterns of *Plasmodium falciparum* Clonally Variant Genes at the Onset of a Blood Infection in Malaria-Naive Humans. *mBio*, e0163621 (2021).
30. Bethke, L. L. *et al.* Duplication, gene conversion, and genetic diversity in the species-specific acyl-CoA synthetase gene family of *Plasmodium falciparum*. *Mol Biochem Parasitol* **150**, 10-24 (2006).
31. Shears, M. J., Botté, C. Y. & McFadden, G. I. Fatty acid metabolism in the *Plasmodium*

apicoplast: Drugs, doubts and knockouts. *Mol Biochem Parasitol* **199**, 34-50 (2015).
32. Das, D. *et al.* Complex interactions between malaria and malnutrition: a systematic literature review. *BMC Med* **16**, 186 (2018).
33. Ramakrishnan, S., Serricchio, M., Striepen, B. & Bütikofer, P. Lipid synthesis in protozoan parasites: a comparison between kinetoplastids and apicomplexans. *Prog Lipid Res* **52**, 488-512 (2013).
34. Gulati, S. *et al.* Profiling the Essential Nature of Lipid Metabolism in Asexual Blood and Gametocyte Stages of *Plasmodium falciparum*. *Cell Host Microbe* **18**, 371-381 (2015).
35. Mitamura, T., Hanada, K., Ko-Mitamura, E. P., Nishijima, M. & Horii, T. Serum factors governing intraerythrocytic development and cell cycle progression of *Plasmodium falciparum*. *Parasitol Int* **49**, 219-229 (2000).
36. Mi-Ichi, F., Kita, K. & Mitamura, T. Intraerythrocytic *Plasmodium falciparum* utilize a broad range of serum-derived fatty acids with limited modification for their growth. *Parasitology* **133**, 399-410 (2006).
37. Mi-Ichi, F., Kano, S. & Mitamura, T. Oleic acid is indispensable for intraerythrocytic proliferation of *Plasmodium falciparum*. *Parasitology* **134**, 1671-1677 (2007).
38. Asahi, H., Izumiyama, S., Tolba, M. E. & Kwansa-Bentum, B. *Plasmodium falciparum*: differing effects of non-esterified fatty acids and phospholipids on intraerythrocytic growth in serum-free medium. *Exp Parasitol* **127**, 708-713 (2011).
39. Ben Mamoun, C., Prigge, S. T. & Vial, H. Targeting the Lipid Metabolic Pathways for the Treatment of Malaria. *Drug Dev Res* **71**, 44-55 (2010).
40. Botté, C. Y. *et al.* Atypical lipid composition in the purified relict plastid (apicoplast) of malaria parasites. *Proc Natl Acad Sci USA* **110**, 7506-7511 (2013).
41. Tokumasu, F., Hayakawa, E. H., Fukumoto, J., Tokuoka, S. M. & Miyazaki, S. Creative interior design by *Plasmodium falciparum*: Lipid metabolism and the parasite's secret chamber. *Parasitol Int* **83**, 102369 (2021).
42. Yu, M. *et al.* The fatty acid biosynthesis enzyme FabI plays a key role in the development of liver-stage malarial parasites. *Cell Host Microbe* **4**, 567-578 (2008).
43. Vaughan, A. M. *et al.* Type II fatty acid synthesis is essential only for malaria parasite late liver stage development. *Cell Microbiol* **11**, 506-520 (2009).
44. van Schaijk, B. C. *et al.* Type II fatty acid biosynthesis is essential for *Plasmodium falciparum* sporozoite development in the midgut of *Anopheles* mosquitoes. *Eukaryot Cell* **13**, 550-559 (2014).
45. Cranmer, S. L., Magowan, C., Liang, J., Coppel, R. L. & Cooke, B. M. An alternative to serum for cultivation of *Plasmodium falciparum* in vitro. *Trans R Soc Trop Med Hyg* **91**, 363-365. (1997).
46. Brown, A. C. & Guler, J. L. From Circulation to Cultivation: *Plasmodium* In Vivo versus In Vitro. *Trends Parasitol* **36**, 914-926 (2020).
47. Garcia-Gonzalo, F. R. & Izpisúa Belmonte, J. C. Albumin-associated lipids regulate human embryonic stem cell self-renewal. *PLoS One* **3**, e1384 (2008).
48. Srivastava, K., Singh, S., Singh, P. & Puri, S. K. In vitro cultivation of *Plasmodium falciparum*: studies with modified medium supplemented with ALBUMAX II and various animal sera. *Exp Parasitol* **116**, 171-174 (2007).

49. Cortés, A. *et al.* Epigenetic silencing of *Plasmodium falciparum* genes linked to erythrocyte invasion. *PLoS Pathog* **3**, e107 (2007).
50. Painter, H. J., Altenhofen, L. M., Kafsack, B. F. & Llinas, M. Whole-genome analysis of *Plasmodium spp.* utilizing a new agilent technologies DNA microarray platform. *Methods Mol Biol* **923**, 213-219 (2013).
51. Boysen, K. E. & Matuschewski, K. Arrested oocyst maturation in *Plasmodium* parasites lacking type II NADH:ubiquinone dehydrogenase. *J Biol Chem* **286**, 32661-32671 (2011).
52. Ke, H. *et al.* Mitochondrial type II NADH dehydrogenase of *Plasmodium falciparum* (PfNDH2) is dispensable in the asexual blood stages. *PLoS One* **14**, e0214023 (2019).
53. Ahouidi, A. *et al.* An open dataset of *Plasmodium falciparum* genome variation in 7,000 worldwide samples. *Wellcome Open Res* **6**, 42 (2021).
54. Manske, M. *et al.* Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. *Nature* **487**, 375-379 (2012).
55. Painter, H. J., Morrisey, J. M., Mather, M. W. & Vaidya, A. B. Specific role of mitochondrial electron transport in blood-stage *Plasmodium falciparum*. *Nature* **446**, 88-91 (2007).
56. Ke, H. *et al.* Genetic investigation of tricarboxylic acid metabolism during the *Plasmodium falciparum* life cycle. *Cell Rep* **11**, 164-174 (2015).
57. MacRae, J. I. *et al.* Mitochondrial metabolism of sexual and asexual blood stages of the malaria parasite *Plasmodium falciparum*. *BMC Biol* **11**, 67 (2013).
58. Evers, F. *et al.* Composition and stage dynamics of mitochondrial complexes in *Plasmodium falciparum*. *Nat Commun* **12**, 3820 (2021).
59. Cortés, A., Benet, A., Cooke, B. M., Barnwell, J. W. & Reeder, J. C. Ability of *Plasmodium falciparum* to invade Southeast Asian ovalocytes varies between parasite lines. *Blood* **104**, 2961-2966 (2004).
60. Rovira-Graells, N., Aguilera-Simon, S., Tinto-Font, E. & Cortes, A. New Assays to Characterise Growth-Related Phenotypes of *Plasmodium falciparum* Reveal Variation in Density-Dependent Growth Inhibition between Parasite Lines. *PLoS ONE* **11**, e0165358 (2016).
61. Casas-Vila, N., Pickford, A. K., Portugaliza, H. P., Tintó-Font, E. & Cortés, A. Transcriptional Analysis of Tightly Synchronized *P. falciparum* Intraerythrocytic Stages by RT-qPCR. *Methods Mol Biol* **2369**, 165-185 (2021).
62. Llorà-Batlle, O. *et al.* Conditional expression of PfAP2-G for controlled massive sexual conversion in *Plasmodium falciparum*. *Sci Adv* **6**, eaaz5057 (2020).
63. Lemieux, J. E. *et al.* Statistical estimation of cell-cycle progression and lineage commitment in *Plasmodium falciparum* reveals a homogeneous pattern of transcription in ex vivo culture. *Proc Natl Acad Sci USA* **106**, 7559-7564 (2009).
64. Saeed, A. I. *et al.* TM4 microarray software suite. *Methods Enzymol* **411**, 134-193 (2006).
65. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* **102**, 15545-15550 (2005).
66. Xie, T. *et al.* The global motion affecting electron transfer in *Plasmodium falciparum* type II NADH dehydrogenases: a novel non-competitive mechanism for quinoline ketone derivative inhibitors. *Phys Chem Chem Phys* **21**, 18105-18118 (2019).

ACKNOWLEDGMENTS

This publication uses data from the MalariaGEN *Plasmodium falciparum* Community Project as described in "An open dataset of *Plasmodium falciparum* genome variation in 7,000 worldwide samples". Wellcome Open Research 2021642 DOI: 10.12688/wellcomeopenres.16168.1. We thank the Genomics Unit at the CRG for assistance with genome sequencing. This work was supported by grants from the Spanish Ministerio de Ciencia e Innovación (MCI)/ Agencia Estatal de Investigación (AEI) [SAF2016-76190-R and PID2019-107232RBI00 to A.C.], co-funded by the European Regional Development Fund (ERDF, European Union). A.K.P. is supported by a fellowship from the Secretary for Universities and Research, Catalan Government (FI_B 00373), co-funded by the European Social Fund (ESF), European Commission. L.M.-T. is supported by a fellowship from the Spanish Ministry of Economy and Competitiveness (BES-2017-081079), co-funded by the European Social Fund (ESF). Our research is part of ISGlobal's Program on the Molecular Mechanisms of Malaria, which is partially supported by the Fundación Ramón Areces. We acknowledge support from the Spanish Ministry of Science and Innovation through the "Centro de Excelencia Severo Ochoa 2019-2023" Program (CEX2018-000806-S), and support from the Generalitat de Catalunya through the CERCA Program. Funding for open access charge: Agencia Estatal de Investigación.

AUTHOR CONTRIBUTIONS

AKP and AC conceived the project, designed the experiments and interpreted the results. AKP performed the experiments. LM-T performed bioinformatic analyses. AKP and AC wrote the manuscript, with input from LM-T.

COMPETING INTERESTS STATEMENT

The authors declare no competing interests.

FIGURES

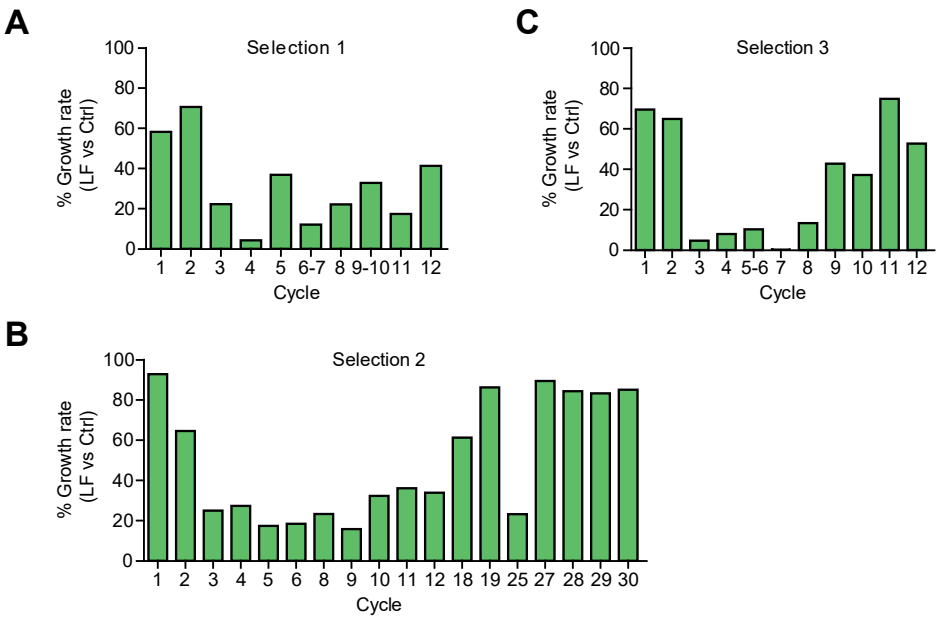


Fig. 1. Selection of cultures in lipid-free (LF) medium. (**A-C**) Growth rates of 3D7-A cultures maintained in LF medium relative to control cultures (maintained in parallel in standard medium) during three independent selection experiments (panels A to C). Cultures were diluted to a 1% parasitaemia at the ring stage of each cycle, and growth rates calculated as the fold-increase in parasitaemia at the ring stage relative to the previous cycle. In selection experiments 1 and 3, parasitaemias were measured by light-microscopy analysis of Giemsa-stained smears, whereas in selection experiment 2 they were measured by flow cytometry.

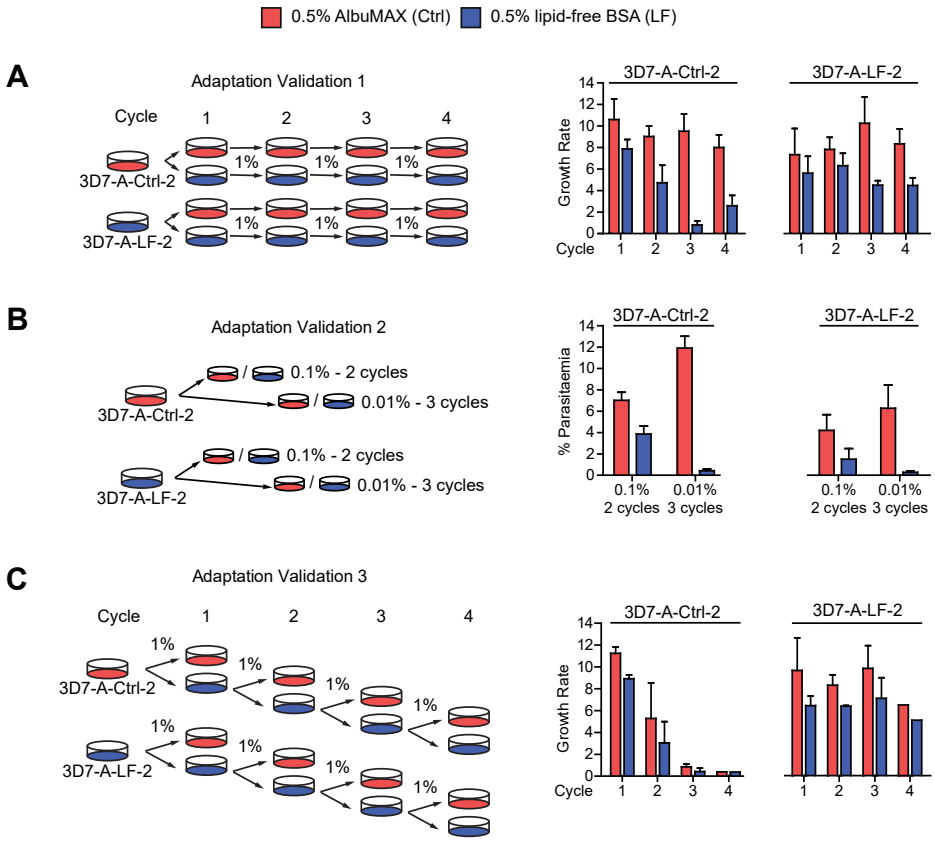


Fig. 2. Adaptation validation experiments. (**A**) Cultures of the 3D7-A-Ctrl-2 (control) and 3D7-A-LF-2 (LF-selected) lines were split and continuously maintained in media supplemented with 0.5% AlbuMAX II (red, control condition) or 0.5% lipid-free BSA (blue, LF condition). Cultures were diluted to a 1% parasitaemia at the ring stage of each cycle, and growth rates were calculated as the increase in parasitaemia (measured by flow cytometry at the ring stage of each cycle, before dilution) relative to the previous cycle. (**B**) Same as in panel A, but initial parasitaemias were adjusted to 0.1% or 0.01% and cultures were left undisturbed for 2 or 3 cycles, respectively, until the final parasitaemia was measured by flow cytometry. (**C**) Same as in panel A, but after each cycle the culture maintained in LF medium was used to continue the experiment and compare growth between control and LF conditions (by preparing cultures at 1% parasitaemia and splitting into dishes with one or the other medium). In this experiment, parasitaemias were measured by light-microscopy analysis of Giemsa-stained smears. Data are presented as the average and s.e.m of two (C) or three (A, B) independent biological replicates.

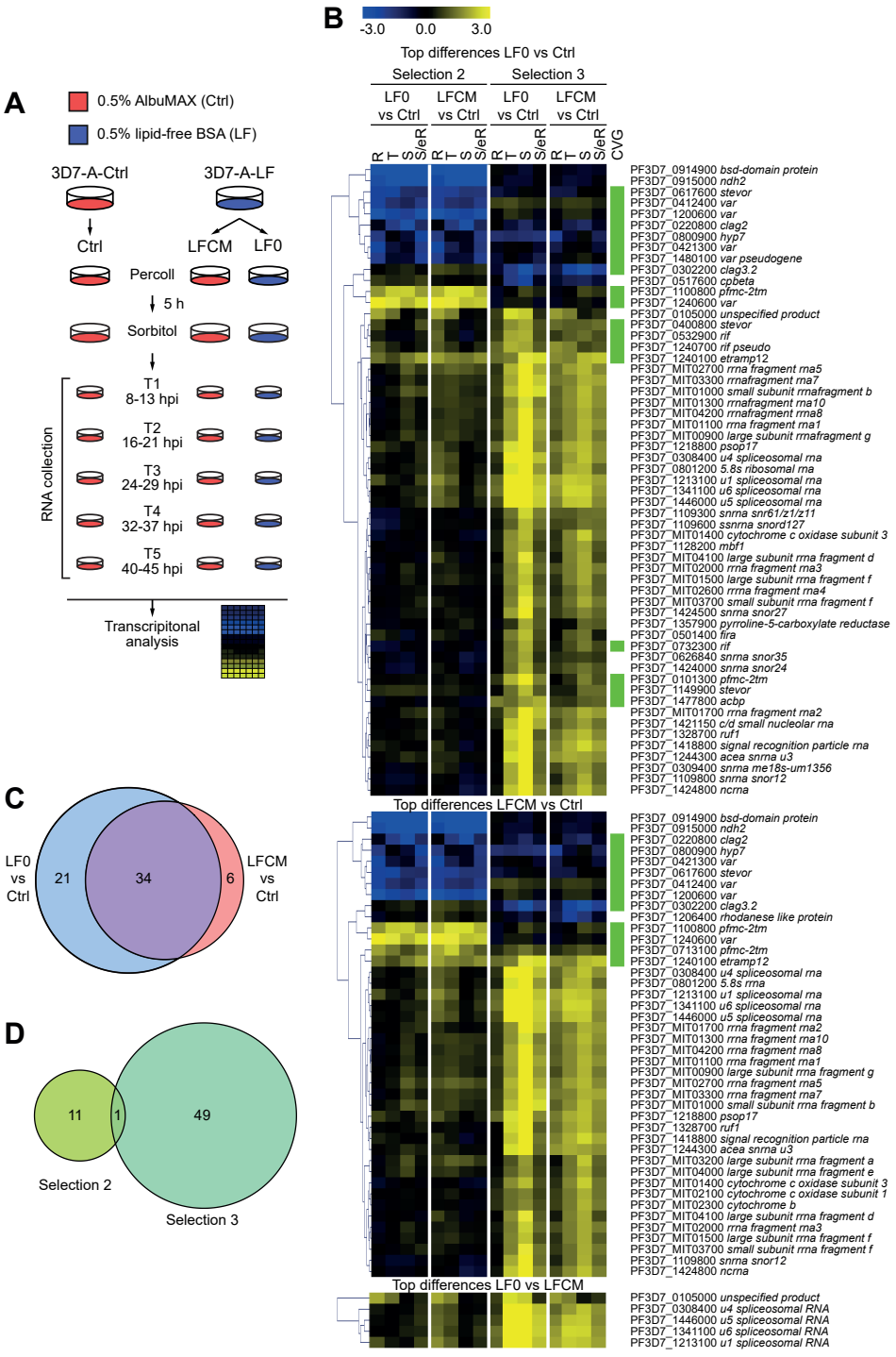

A
0.5% AlbuMAX (Ctrl)
0.5% lipid-free BSA (LF)
3D7-A-Ctrl
3D7-A-LF
Ctrl
LFCM
LF0
Percoll
5 h
Sorbitol
RNA collection
T1
8-13 hpi
T2
16-21 hpi
T3
24-29 hpi
T4
32-37 hpi
T5
40-45 hpi
Transcripitonal analysis
B
-3.0
0.0
3.0
Top differences LF0 vs Ctrl
Selection 2
Selection 3
LF0 vs Ctrl
LFCM vs Ctrl
LF0 vs Ctrl
LFCM vs Ctrl
R
T
S
S/eR
CVG
PF3D7_0914900 bsd-domain protein
PF3D7_0915000 ndh2
PF3D7_0617600 stevor
PF3D7_0412400 var
PF3D7_1200600 var
PF3D7_0220800 clag2
PF3D7_0800900 hyp7
PF3D7_0421300 var
PF3D7_1480100 var pseudogene
PF3D7_0302200 clag3.2
PF3D7_0517600 cpbeta
PF3D7_1100800 pfmc-2tm
PF3D7_1240600 var
PF3D7_0105000 unspecified product
PF3D7_0400800 stevor
PF3D7_0532900 rif
PF3D7_1240700 rif pseudo
PF3D7_1240100 etramp12
PF3D7_MIT02700 rrna fragment rna5
PF3D7_MIT03300 rrnafragment rna7
PF3D7_MIT01000 small subunit rrnafragment b
PF3D7_MIT01300 rrnafragment rna10
PF3D7_MIT04200 rrnafragment rna8
PF3D7_MIT01100 rrna fragment rna1
PF3D7_MIT00900 large subunit rrnafragment g
PF3D7_1218800 psop17
PF3D7_0308400 u4 spliceosomal rna
PF3D7_0801200 5.8s ribosomal rna
PF3D7_1213100 u1 spliceosomal rna
PF3D7_1341100 u6 spliceosomal rna
PF3D7_1446000 u5 spliceosomal rna
PF3D7_1109300 snrna snr61/z1/z11
PF3D7_1109600 ssnrna snord127
PF3D7_MIT01400 cytochrome c oxidase subunit 3
PF3D7_1128200 mbf1
PF3D7_MIT04100 large subunit rrna fragment d
PF3D7_MIT02000 rrna fragment rna3
PF3D7_MIT01500 large subunit rrna fragment f
PF3D7_MIT02600 rrna fragment rna4
PF3D7_MIT03700 small subunit rrna fragment f
PF3D7_1424500 snrna snor27
PF3D7_1357900 pyrroline-5-carboxylate reductase
PF3D7_0501400 fira
PF3D7_0732300 rif
PF3D7_0626840 snrna snor35
PF3D7_1424000 snrna snor24
PF3D7_0101300 pfmc-2tm
PF3D7_1149900 stevor
PF3D7_1477800 acbp
PF3D7_MIT01700 rrna fragment rna2
PF3D7_1421150 c/d small nucleolar rna
PF3D7_1328700 ruf1
PF3D7_1418800 signal recognition particle rna
PF3D7_1244300 acea snrna u3
PF3D7_0309400 snrna me18s-um1356
PF3D7_1109800 snrna snor12
PF3D7_1424800 ncrna
Top differences LFCM vs Ctrl
PF3D7_0914900 bsd-domain protein
PF3D7_0915000 ndh2
PF3D7_0220800 clag2
PF3D7_0800900 hyp7
PF3D7_0421300 var
PF3D7_0617600 stevor
PF3D7_0412400 var
PF3D7_1200600 var
PF3D7_0302200 clag3.2
PF3D7_1206400 rhodanese like protein
PF3D7_1100800 pfmc-2tm
PF3D7_1240600 var
PF3D7_0713100 pfmc-2tm
PF3D7_1240100 etramp12
PF3D7_0308400 u4 spliceosomal rna
PF3D7_0801200 5.8s rrna
PF3D7_1213100 u1 spliceosomal rna
PF3D7_1341100 u6 spliceosomal rna
PF3D7_1446000 u5 spliceosomal rna
PF3D7_MIT01700 rrna fragment rna2
PF3D7_MIT01300 rrna fragment rna10
PF3D7_MIT04200 rrna fragment rna8
PF3D7_MIT01100 rrna fragment rna1
PF3D7_MIT00900 large subunit rrna fragment g
PF3D7_MIT02700 rrna fragment rna5
PF3D7_MIT03300 rrna fragment rna7
PF3D7_MIT01000 small subunit rrna fragment b
PF3D7_1218800 psop17
PF3D7_1328700 ruf1
PF3D7_1418800 signal recognition particle rna
PF3D7_1244300 acea snrna u3
PF3D7_MIT03200 large subunit rrna fragment a
PF3D7_MIT04000 large subunit rrna fragment e
PF3D7_MIT01400 cytochrome c oxidase subunit 3
PF3D7_MIT02100 cytochrome c oxidase subunit 1
PF3D7_MIT02300 cytochrome b
PF3D7_MIT04100 large subunit rrna fragment d
PF3D7_MIT02000 rrna fragment rna3
PF3D7_MIT01500 large subunit rrna fragment f
PF3D7_MIT03700 small subunit rrna fragment f
PF3D7_1109800 snrna snor12
PF3D7_1424800 ncrna
Top differences LF0 vs LFCM
PF3D7_0105000 unspecified product
PF3D7_0308400 u4 spliceosomal RNA
PF3D7_1446000 u5 spliceosomal RNA
PF3D7_1341100 u6 spliceosomal RNA
PF3D7_1213100 u1 spliceosomal RNA
C
LF0 vs Ctrl
21
34
6
LFCM vs Ctrl
D
11
1
49
Selection 2
Selection 3

Fig. 3. Transcriptomic analysis of control and lipid-free (LF)-selected lines. (**A**) Schematic of the experimental set up. Control (3D7-A-Ctrl-2 and 3D7-A-Ctrl-3) and LF-selected (3D7-A-LF-2 and 3D7-A-LF-3) lines were tightly synchronised to a 0-5 h age window by Percoll purification followed by sorbitol lysis 5 h later. After Percoll purification, 3D7-A-Ctrl cultures were maintained in 0.5% AlbuMAX II (red, Ctrl medium), whereas 3D7-A-LF cultures were split into two separate conditions: LF0 always maintained in 0.5% lipid-free BSA (blue, LF medium) and LFCM switched to Ctrl medium. RNA samples were collected every 8 h at the indicated time points (in h post invasion [hpi]) and used for transcriptomic analysis using microarrays. (**B**) Heatmap of genes differentially expressed between LF0 and Ctrl (top), LFCM and Ctrl (middle) or LF0 and LFCM (bottom) cultures, ordered by hierarchical clustering. Values are the $\log_2$ of the average fold-change (AFC) of LF0 or LFCM vs the Ctrl cultures at four overlapping time intervals corresponding to the stages indicated (R: rings, T: trophozoites, S: schizonts, S/eR: schizonts and early rings). Only genes with an absolute value of the $\log_2$ of the maximum average fold-change (mAFC) >2 (LF0 vs Ctrl and LFCM vs Ctrl comparisons) or a difference in mAFC >2 (LF0 vs LFCM comparison) in at least one of the two selection experiments (selections 2 and 3) are shown. The column at the right indicates (in green) whether a gene was previously classified as a CVG. (**C**) Euler diagram showing the overlap of genes with an absolute value of the $\log_2$(mAFC) >2 (relative to Ctrl cultures) between the LF0 and LFCM conditions. (**D**) Euler diagram showing the overlap of genes with an absolute value of the $\log_2$(mAFC) >2 (in either of the possible comparisons between Ctrl, LF0 or LFCM conditions) between selections 2 and 3.

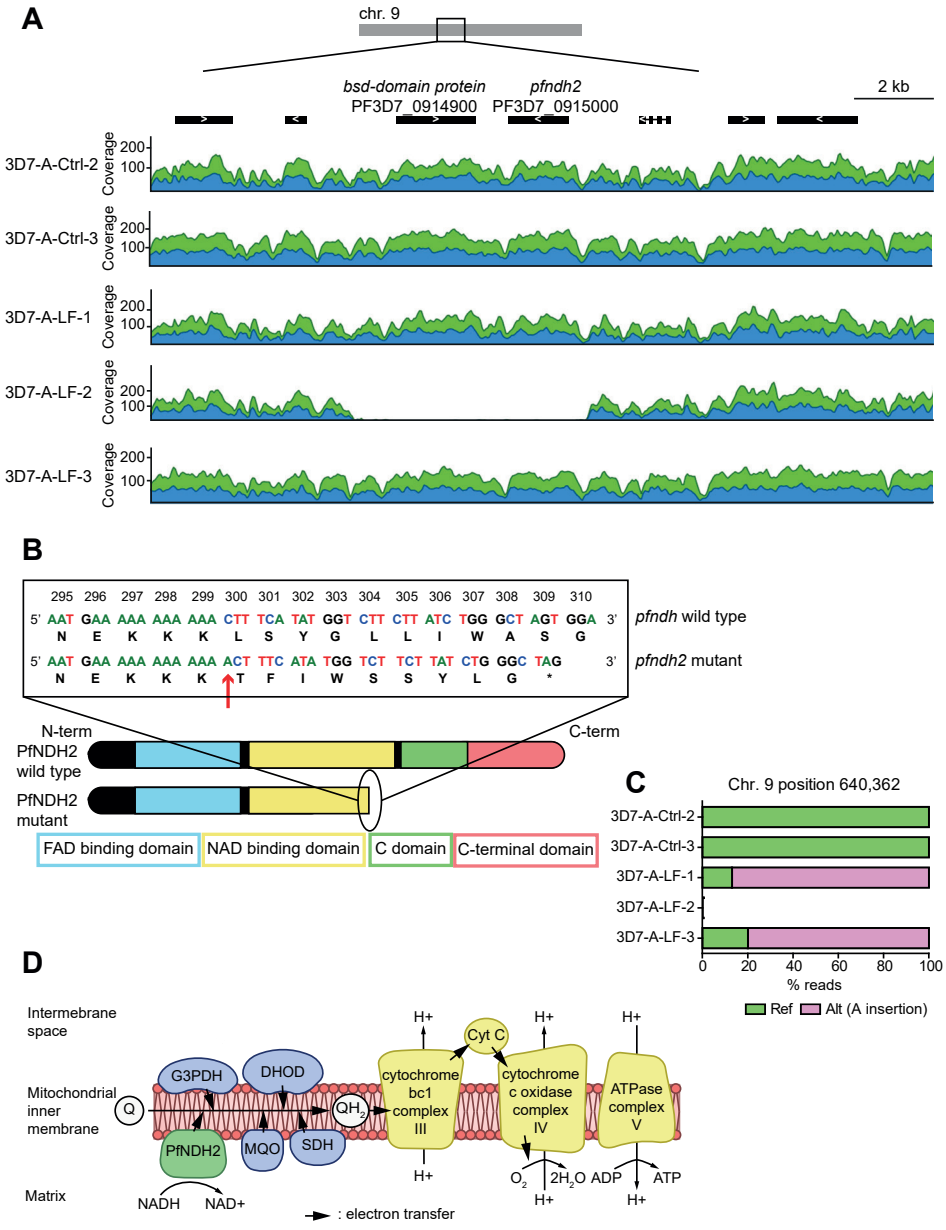

A
chr. 9
bsd-domain protein
PF3D7_0914900
pfndh2
PF3D7_0915000
2 kb
3D7-A-Ctrl-2
3D7-A-Ctrl-3
3D7-A-LF-1
3D7-A-LF-2
3D7-A-LF-3
Coverage
200
100
B
295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310
5' AAT GAA AAA AAA AAA CTT TCA TAT GGT CTT CTT ATC TGG GCT AGT GGA 3'
N E K K K L S Y G L L I W A S G
pfndh wild type
5' AAT GAA AAA AAA AAA ACT TTC ATA TGG TCT TCT TAT CTG GGC TAG 3'
N E K K K T F I W S S Y L G *
pfndh2 mutant
N-term
C-term
PfNDH2 wild type
PfNDH2 mutant
FAD binding domain
NAD binding domain
C domain
C-terminal domain
C
Chr. 9 position 640,362
3D7-A-Ctrl-2
3D7-A-Ctrl-3
3D7-A-LF-1
3D7-A-LF-2
3D7-A-LF-3
0 20 40 60 80 100
% reads
Ref
Alt (A insertion)
D
Intermembrane space
Mitochondrial inner membrane
Matrix
G3PDH
DHOD
PfNDH2
MQO
SDH
Q
QH2
cytochrome bc1 complex III
Cyt C
cytochrome c oxidase complex IV
ATPase complex V
H+
O2
2H2O
ADP
ATP
NADH
NAD+
: electron transfer

Fig. 4. Whole genome sequencing of lipid-free (LF)-selected and control lines. (**A**) Deletion of a region of chromosome 9 including the genes PF3D7_0914900 and PF3D7_0915000 (*pfndh2*) in the 3D7-A-LF-2 line. The coverage of aligned whole genome sequencing reads at this region is shown for control (3D7-A-Ctrl-2 and 3D7-A-Ctrl-3) and LF-selected (3D7-A-LF-1, 3D7-A-LF-2 and 3D7-A-LF-3) lines. Blue and green represent the coverage of aligned forward and reverse reads, respectively. (**B**) A single nucleotide insertion in the 3D7-A-LF-1 and 3D7-A-LF-3 lines results in a frameshift mutation that leads to a premature STOP codon (*) in the PfNDH2 protein. The position of the known functional domains, according to Xie *et al.*[66], is indicated. (**C**) Proportion of Illumina reads with (Alt) or without (Ref) the single base insertion at position 640,362 of chromosome 9 (position 898 of the *pfndh2* gene) in the control and LF-selected lines. This gene is deleted in the 3D7-A-LF-2 line. (**D**) Schematic of the mitochondrial electron transport chain. PfNDH2 transfers electrons from NADH to ubiquinone (Q) to form ubiquinol (QH_2).

SUPPLEMENTARY FIGURES

Adaptation of *Plasmodium falciparum* to medium lacking a lipids supplement is associated with mutations in the *pfndh2* gene

Anastasia K. Pickford, Lucas Michel-Todó & Alfred Cortés

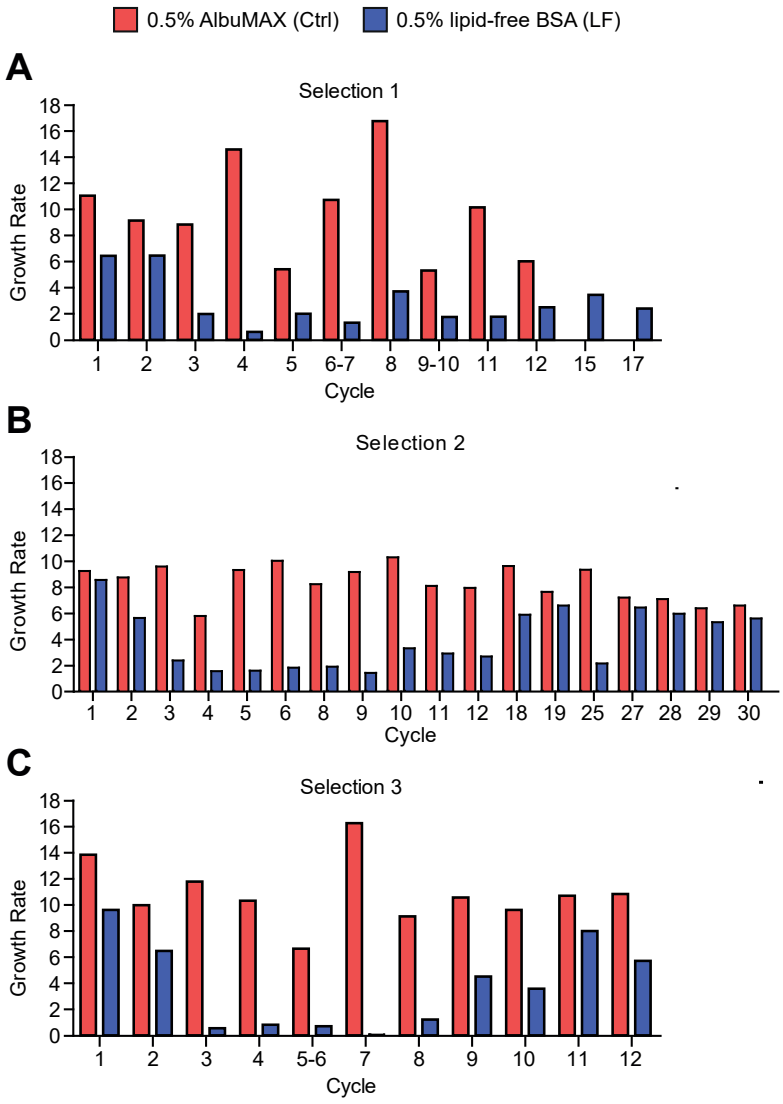


Fig. S1. Selection of cultures in lipid-free (LF) medium. (**A-C**) Growth rates of 3D7-A control (red) and LF-selected (blue) cultures during three independent selection experiments (panels A to C). Cultures were diluted to a 1% parasitaemia at the ring stage of each cycle, and growth rates calculated as the fold-increase in parasitaemia at the ring stage relative to the previous cycle. In selection experiments 1 and 3, parasitaemias were measured by light-microscopy analysis of Giemsa-stained smears, whereas in selection experiment 2 they were measured by flow cytometry. In selection 1, the control culture was maintained for only 12 cycles.

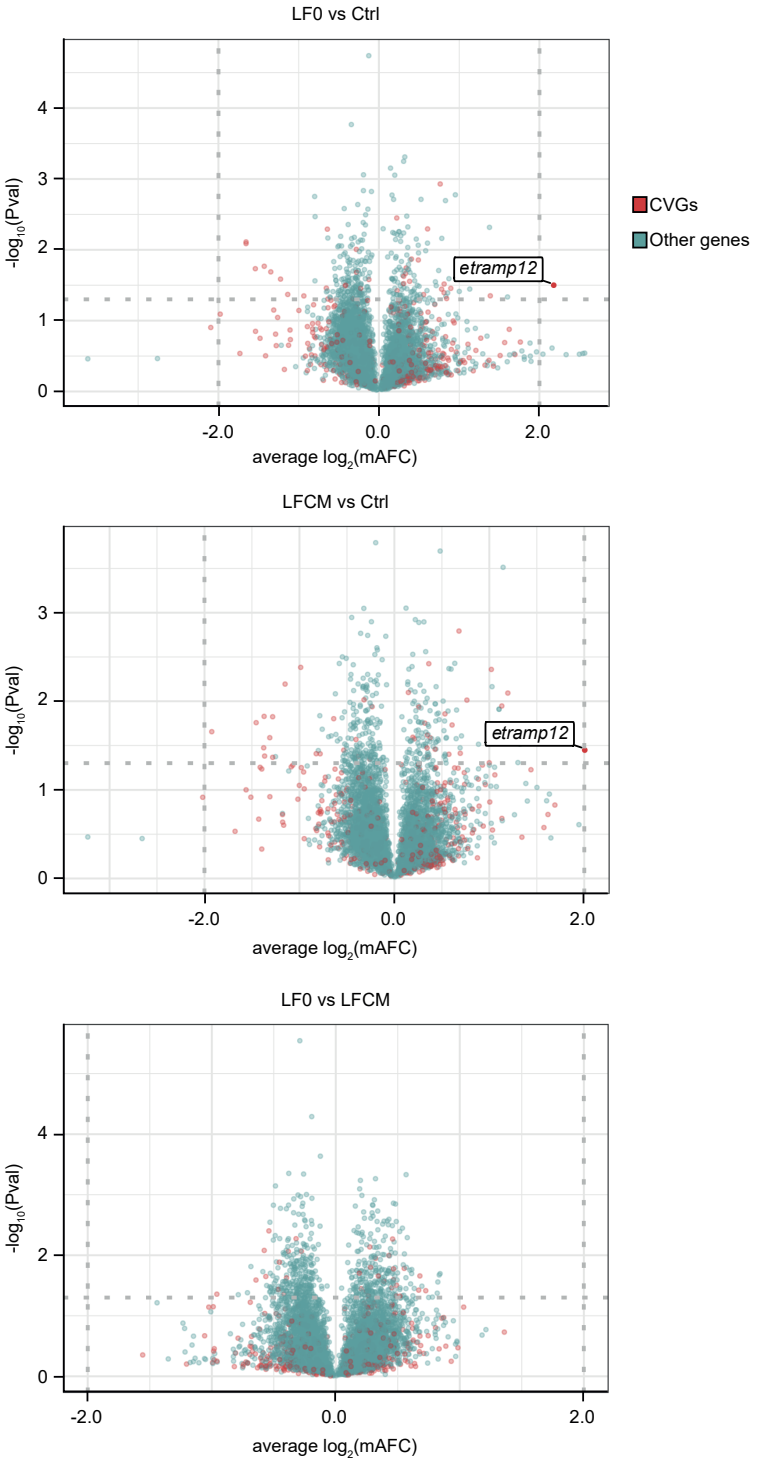

LF0 vs Ctrl
-log10(Pval)
4
3
2
1
0
etramp12
-2.0
0.0
2.0
average log2(mAFC)
CVGs
Other genes
LFCM vs Ctrl
-log10(Pval)
3
2
1
0
etramp12
-2.0
0.0
2.0
average log2(mAFC)
LF0 vs LFCM
-log10(Pval)
4
2
0
-2.0
0.0
2.0
average log2(mAFC)

Fig. S2. Volcano plots of the transcriptomic analysis of lipid-free (LF)-selected and control cultures. Expression fold-change values for the LF0 vs Ctrl (top), LFCM vs Ctrl (middle) and LF0 vs LFCM (bottom) comparisons, expressed as the $\log_2$(mAFC), are the average of the results for selections 2 and 3. P-values were calculated using an unpaired two-sided t-test. Clonally variant genes (CVGs) are shown in red and others in blue. The position of genes with an absolute value of the $\log_2$(mAFC) >2 and a P-value <0.05 is shown. Vertical dotted lines delimitate regions of the plot with a log2(mAFC)<2 or >2 (sides), whereas the horizontal dotted line delimitates the region with a P value <0.05 [$-\log_{10}$(Pval) > 1.30, upper part].

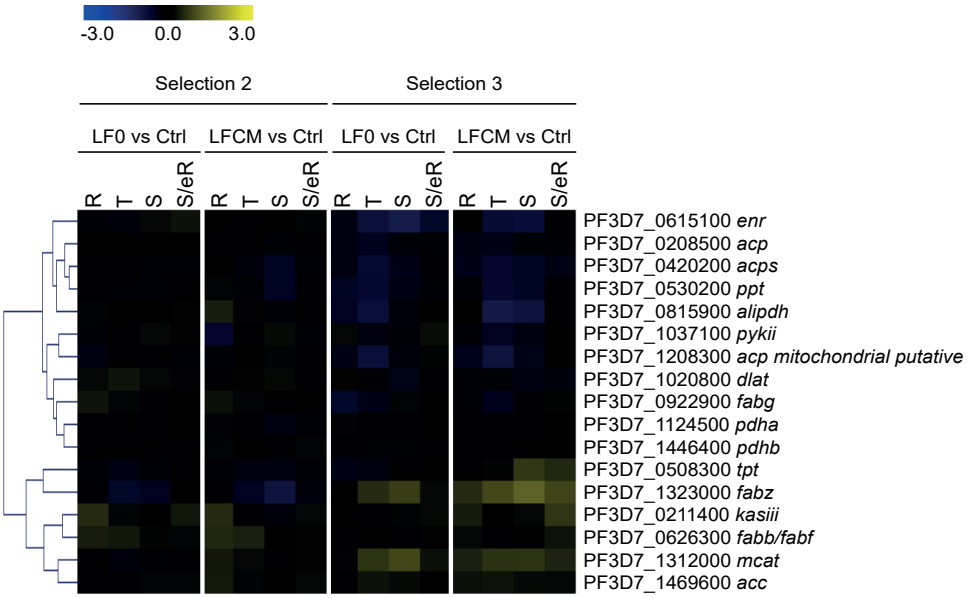


Fig. S3. Expression of genes linked to the FASII pathway for fatty acids synthesis in control and lipid-free (LF)-selected cultures. Values are the expression fold-change in LF0 or LFCM vs Ctrl cultures in selection experiments 2 or 3, as in main Fig. 3, for genes linked to the FASII pathway, according to previous reviews (Shears MJ *et al*, 2015, *Mol Biochem Parasitol* 199:34-50; Mazumdar, J & Striepen, B, 2007, *Eukaryot Cell* 6:1727-1735). Gene names were obtained from PlasmoDB. Genes are ordered by hierarchical clustering.

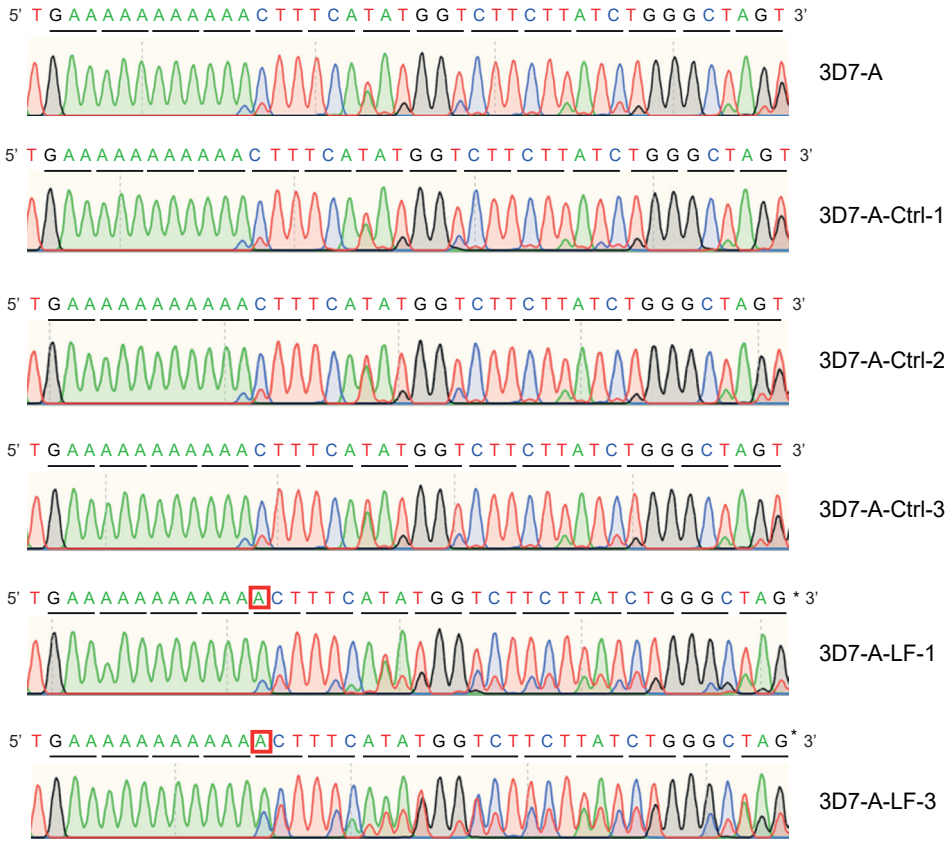


Fig. S4. Sanger sequencing of the mutation in *pfndh2*. Sanger sequencing confirmation of the mutation in the *pfndh2* gene identified in some lipid-free (LF)-selected lines by Illumina sequencing. Results for the parental 3D7-A line, lipid-free (LF)-selected lines 3D7-A-LF-1 and 3D7-A-LF-3, and control lines 3D7-A-Ctrl-1, 3D7-A-Ctrl-2 and 3D7-A-Ctrl-3 are shown. The 3D7-A-LF-2 line was not included because the *pfndh2* gene is deleted in this line. Underlines indicate codons, showing how the single base insertion in 3D7-A-LF-1 and 3D7-A-LF-3 (red boxes) results in a frameshift mutation and a premature STOP codon (*). Double peaks after the poly(A) track correspond to molecules losing one nucleotide likely during PCR amplification (both in *pfndh2* wild type and mutant lines), which commonly occurs at homopolymeric sequences.

Other results 1: Selection of *Plasmodium falciparum* with sub-optimal glucose concentrations

Pickford AK & Cortés A. Selection of *Plasmodium falciparum* with sub-optimal glucose concentrations. (**Unpublished**).

1. Introduction

One of the main objectives of this thesis was to explore which fluctuating conditions of the human host circulation *P. falciparum* can adapt to by the means of CVG-mediated bet-hedging adaptive strategies. In this manner, we selected nutrient availability as one of the conditions to explore as in the field malaria and malnutrition often coexist[192] and it is probable that *Plasmodium* spp. encounter limiting concentrations of different types of nutrients within their host's serum. In Article 3 of this thesis, we studied how *P. falciparum* adapted to external lipid deprivation and found that a genetic mutation was involved rather than epigenetic changes in CVG expression. However, following this line of research, we set out to establish if CVGs play a role in the adaptation of *P. falciparum* to glucose starvation.

Glucose is an essential nutrient for malaria parasites, especially for asexual blood stages[422]. Generally, glucose levels in the blood are tightly regulated except in the case of diseases such as diabetes[423] (which has a high prevalence in some malaria endemic countries[191]) and the average glucose concentration in human sera ranges from 0.7 to 1.3 mg/mL[424]. Nonetheless, in cases of hypoglycaemia associated to severe malaria this concentration has been reported to be as low as 0.12 mg/mL[193]. Moreover, it has been described that when parasites sequester, microcapillary concentrations of glucose drop dramatically[425]. As a result, *Plasmodium* spp. can potentially be exposed to variations in the availability of this nutrient and, therefore, it is reasonable to hypothesise that they may possess a mechanism with which to adapt to these fluctuations.

Our objectives, therefore, were first to establish whether *P. falciparum* parasites could adapt to low glucose concentrations and second, to determine the possible mechanisms behind this potential adaptation.

2. Results

2.1. Selection I

We performed an initial selection experiment to determine whether the transcriptionally-diverse parasite line 3D7-A[322] could grow in media with limited glucose concentration. We established four conditions: 2 g/L of glucose (control with standard in vitro glucose concentration), 0.5 g/L of glucose (just below the normoglycaemic level), 0.1 g/L of glucose (similar to the hypoglycaemic levels in severe malaria) and 0 g/L of glucose. 3D7-A parasites were cultured in each of these conditions for a total of almost two weeks (9 cycles). Cultures were diluted to 1% parasitaemia at the ring stage of each cycle, and growth rates calculated as the fold-increase in parasitaemia at the ring stage relative to the previous cycle as measured by smears and flow cytometry (see methods). Almost immediately, we observed that parasites cultured in media with 0.1 and 0 g/L of glucose did not survive past the first cycle. However, in the 0.5 g/L condition, growth rates declined progressively until the fourth cycle and then slowly started to recover over the next cycles, though in an unstable manner (**Fig. 26**). Throughout the whole selection the control parasites grew steadily.

From this first selection, we concluded that parasites were able to grow in 0.5 g/L of glucose though stable growth rates were not achieved after 9 cycles of selection. Consequently, a longer selection period was needed to confirm if parasites could truly adapt to this low concentration of glucose.

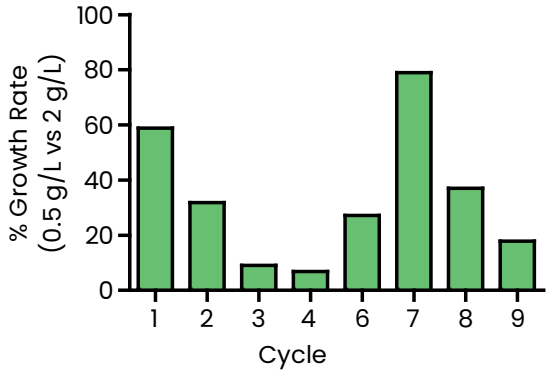


Fig. 26. Growth rate of 3D7-A cultures maintained in 0.5 g/L of glucose relative to control parasites kept in 2 g/L over a selection period of 9 cycles.

2.2. Selection II

In order to address the questions raised with the first selection, we performed a second selection experiment with a few differences (**Fig. 27**). Firstly, we decided to try out a new condition of 1 g/L of glucose just in case in the first selection we had gone too low with the range of glucose concentrations to which *Plasmodium* parasites can potentially adapt to. Secondly, in addition to a replicate of the 0.5 g/L condition, we also thawed the parasite culture maintained at this concentration of glucose in the first selection in order to see if longer exposure to these conditions resulted in a more stable adaptation. Finally, this selection lasted a total of three weeks (11 cycles) and parasite growth rates were monitored in the same manner as in the first (see methods).

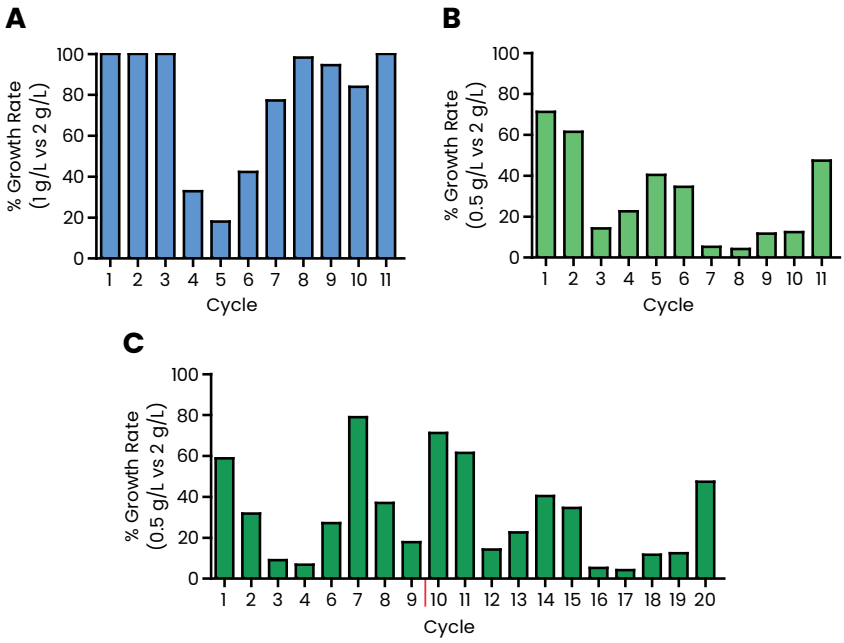


Fig. 27. Growth rate of 3D7-A cultures maintained in 1 g/L (A) or 0.5 g/L (B) of glucose relative to control parasites kept in 2 g/L over a selection period of 11 cycles. (C) Total growth rates of 3D7-A cultures maintained in 0.5 g/L of glucose relative to control parasites from the first selection (cycles 1-9) and then continued in the second selection (cycles 10-20). The red line separates the two selection periods.

In this second selection, we observed in the 1 g/L concentration that parasite growth rates dropped to 20-40% of those of the control between the fourth and sixth cycle but afterwards completely recovered their initial growth rate (similar to that of the control culture) (**Fig. 27A**). In the replicate of the 0.5 g/l condition we saw once again that growth rates fluctuated during the whole selection (**Fig. 27B**). Finally, in the continuation of the initial 0.5 g/L culture from the first selection we also observed that growth rates continued to oscillate between 0% and 50% of those of the control despite being exposed to this concentration of glucose for approximately a month and a half (20 cycles) (**Fig. 27C**). Once more, control parasites showed stable growth throughout the selection.

All in all, from this second selection experiment we were able to conclude that the 3D7-A parasite line was capable of growing in media with limited glucose concentrations but clear evidence of adaptation was still needed as the 1 g/L condition did not appear to exert a strong enough selective pressure and in the 0.5 g/L condition growth rates did not stabilise after more than a month in culture.

2.3. Adaptation validation experiments

In order to establish whether the selected parasite lines were truly adapted to grow in sub-optimal glucose concentrations we performed a final set of adaptation validation experiments. The parasite lines grown in 2 g/L (control conditions), 1 g/L and 0.5 g/L from the second selection (the 0.5 g/L parasites were the ones kept for a total of 20 cycles between the two selections) were each divided into three cultures supplemented with either 2 g/L, 0.75 g/L or 0.5 g/L of glucose. We then measured their growth at every cycle for a total of four cycles in two independent biological replicates (**Fig. 28A**). Growth rates were calculated in the same manner as in the selection assays (see methods). After four cycles, we did not observe any significant differences between the three parasite lines (control, 1 g/L selected and 0.5 g/L selected) in any of the three glucose concentrations (2 g/L, 0.75 g/L and 0.5 g/L) (**Fig. 28B**). Therefore, we concluded that the parasite lines selected at lower glucose concentrations were not stably adapted to these conditions and that it was not worth continuing this line of research.

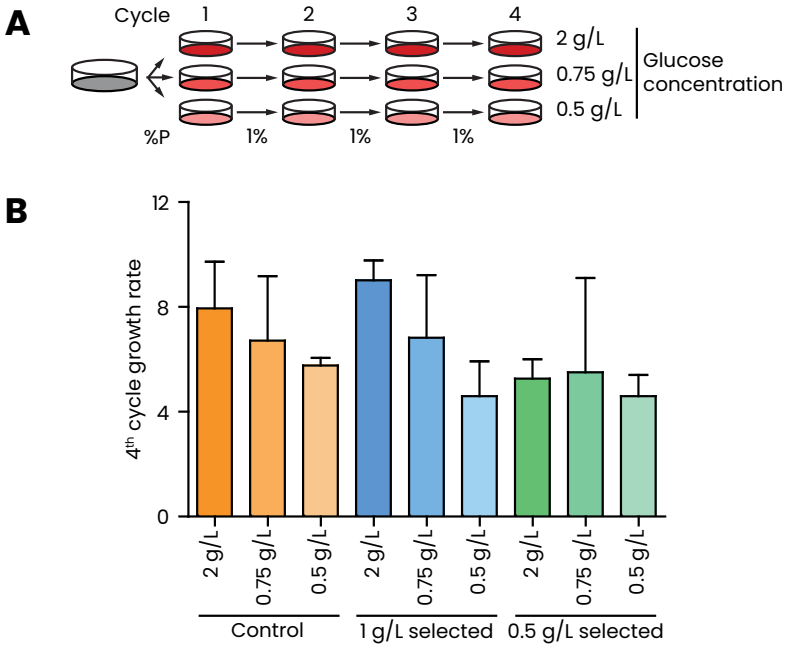


Fig. 28. (A) Experimental design of adaptation validation experiments. The grey culture represents either the control, 1 g/L or 0.5 g/L selected parasites. B) Growth rates of the different parasite lines (control, 1 g/L selected and 0.5 g/L selected) after four cycles in different glucose concentrations (2 g/L, 0.75 g/L and 0.5 g/L). Data are presented as the average and s.e.m of two independent biological replicates.

3. Discussion

We selected 3D7-A parasites with different concentrations of glucose and studied their growth rates over a varying number of cycles. We found that parasites were able to grow in two sub-optimal concentrations of glucose (1 g/L and 0.5 g/L) but when exposed to concentrations below 0.1 g/L, parasites died after the first cycle. Parasites selected with 1 g/L glucose media suffered only a short-lasting growth alteration from which they quickly and completely recovered. On the other hand, parasites selected with 0.5 g/L of glucose over a period of almost 1.5 months did not show clear signs of stable adaptation as they survived albeit with fluctuating growth rates that were also always lower than those of the control cultures. In the additional experiments performed to prove adaptation, no differences were observed between control and selected parasite lines, indicating that there had not been a stable adaptation.

To our knowledge, no other studies have been performed in which *P. falciparum* parasites have been selected with low concentrations of glucose over long periods of time. However, various lines of research have looked into the effects of short-term glucose deprivation on malaria parasites[230,424,426,427]. In these studies, parasite cultures were maintained in glucose-free media or in media with substandard glucose concentrations for periods of time ranging between 2 to 6 h and transcriptional alterations were measured for specific genes or at the full-genome level. Consequently, these studies did not focus on the adaptation of malaria parasites to low glucose concentrations but rather assessed direct transcriptional effects of reducing the available glucose in the media (i.e., protective transcriptional responses). Some studies tried to "recover" some of these parasite lines by supplementing the medium with normal glucose levels after the deprivation period[230,427]. However, it appears that once a certain level of glucose starvation is surpassed, recovery is not possible. Another study established that parasite growth was impaired after 24 to 48 h exposure to glucose concentrations below 0.1 g/L[428], corresponding with our observations of parasites subjected to the same concentration. The parasite's apparent inability to grow with less than 0.1 g/L of glucose in the culture medium does not fit with the low blood sugar levels found in patients with severe malaria. A possible explanation for how malaria parasites are able to survive these hypoglycaemic conditions is that in these patients an important compensatory increase in glucose production also occurs[429]. Therefore, instead of parasites adapting to a lower availability of glucose, they may simply benefit from their host's physiological reaction of increasing glucose production, with this not translating into a rise in blood sugar levels due to the rapid glucose consumption by both host and parasite.

In conclusion, we were not able to obtain parasites that showed clear signs of adaptation to low glucose concentrations and consequently it is unlikely that a bet-hedging adaptive strategy mediated by CVGs is involved in parasite survival to prolonged glucose deprivation. It is possible that a directed transcriptional response or adjustments at the metabolic level may enable parasites to grow under these conditions but this does not fall within the field of interest of this thesis, which is why this line of research was not continued.

4. Methods

4.1. Parasite culture

The *P. falciparum* line 3D7-A was cultured in B+ erythrocytes at 3% haematocrit in glucose-free RPMI-1640-based culture medium (Biowest P0883-N1L) supplemented with 0.5% AlbuMAX and 0.23% $NaHCO_3$. Depending on the experimental condition, this culture medium was then supplemented with different amounts of glucose (Sigma G8270-100G) resulting in a final concentration that ranged between 0 g/L and 2 g/L. Cultures were maintained in static conditions in a 5% CO_2, 3% O_2 and 92% N_2 atmosphere.

4.2. Selection and adaptation comprobation assays

For both these assays, parasites were maintained in media with different concentrations of glucose for varying periods of time. At the beginning of each cycle, cultures were adjusted to 1% parasitaemia by dilution with uninfected RBCs, and 48 h (i.e., before diluting again) parasitaemia was determined once more by flow cytometry and contrasted with Giemsa smears. For flow cytometry measurements a FACScalibur flow cytometer (Becton Dickinson) was used with SYTO 11 as a marker to stain nucleic acids as previously described[430]. Throughout the selection period, parasites were synchronized on a weekly basis using sorbitol lysis.

Other results 2: Development of a triple transgenic parasite line to study the expression dynamics of *clag3* genes.

Pickford AK & Cortés A. Development of a triple transgenic parasite line to study the expression dynamics of *clag3* genes.(**Unpublished**).

1. Background and objectives

Given the focus of this thesis on the role of CVGs in the adaptation of *P. falciparum* to fluctuating conditions within the human host[291], *clag3* genes are of particular interest because they are one of the few CVGs for which there is actual evidence of an adaptive role, besides *var* genes. While switches in *var* gene expression mediate adaptation to the host's immune response[183,311,323,324], different *clag3* expression patterns have been associated with changes in the permeability of iRBC with regards to nutrients[389,401] and various antimalarial compounds[325–327]. As a result, selection of *clag3* gene expression patterns is easier to study in vitro, making these genes good candidates with which to analyse epigenetic switching dynamics and further characterise bet-hedging adaptive strategies in *P. falciparum*.

To date, the principal tools for analysing *clag3* expression patterns are RT-qPCR, microarrays, RNA-seq and western blot. These techniques measure transcript or protein levels as an average of the parasite population and can therefore only provide information on the tendency of *clag3* expression patterns (i.e., the predominantly expressed *clag3* gene under a certain set of conditions). Subsequently, they cannot inform on what happens at the single-cell level and neither are they sensitive enough to analyse subtle switching events or to accurately follow the temporal transcriptional dynamics of these genes. Single-cell RNA-seq could be a solution to this limitation but it is not yet a widespread technique in *Plasmodium* spp. research[431]. Moreover, due to its high cost it is mainly used for full-genome analysis rather than to look at changes in specific genes. In Article 1 of this thesis, we developed an immunofluorescence assay (IFA) that uses an anti-CLAG3 antibody to detect CLAG3 expression in individual parasites. Nevertheless, this assay does not distinguish between CLAG3.1 and CLAG3.2 and it only allows for the analysis of a low number of cells as it requires manual counting at a fluorescence microscope, making it time-consuming and not a high-throughput technique. Consequently, there is a need for a tool that enables the analysis of CLAG3 expression levels at the single-cell level, that can distinguish between CLAG3.1 and CLAG3.2 and that also permits the analysis of a large number of parasites in a relatively short period of time in order to study switching dynamics at the populational level.

In this project we have designed a tool that addresses all the aforementioned requirements and consists in a triple transgenic parasite line. In this line, *clag3.1*, *clag3.2* and a control gene are all individually

tagged with a different fluorescent protein using CRISPR-Cas9 technology. The selected tags each have distinct fluorescence spectra which can be detected by flow cytometry. In this manner, a high number of parasites can be analysed in a short period of time, distinguishing at the single-cell level if they express *clag3.1*, *clag3.2*, neither or both (an expression pattern yet to be confirmed to exist) *clag3* genes (**Fig. 29**). The control gene is expressed at the same time or slightly later than *clag3* genes and enables the distinction between schizonts that do not express *clag3* genes because they have these genes epigenetically silenced (control gene is expressed) or because they are too immature to do so (control gene is not expressed).

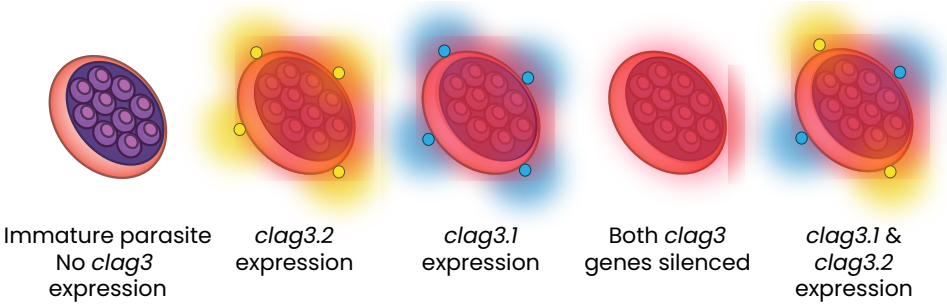


Fig. 29. *clag3* expression patterns detectable with our triple transgenic parasite line. Red fluorescence indicates control gene expression, yellow *clag3.2* expression and blue *clag3.1* expression. Furthermore, DNA staining can be used to identify iRBCs and to distinguish late stages from rings with the intensity of the DNA staining signal.

In the literature, there are previous taggings of *clag3* genes with fluorescent proteins but only of a *clag3* hybrid[432] and of a chimeric *clag3.1* protein[433]. Furthermore, the objective of these taggings was to study the trafficking of the RhopH complex and not *clag3* expression dynamics. Therefore, to our knowledge this is the first design in which both *clag3* genes are separately tagged within the same parasite line with the objective of studying their expression patterns.

With this triple transgenic parasite line, we plan to monitor and directly demonstrate the dynamic selection of *clag3* epigenetic variants with single-cell resolution. In order to do so, this project has been divided into three stages: 1) design and generation of the *clag3* triple transgenic line using CRISPR-Cas9 technology; 2) use this tool to monitor *clag3* expression patterns across time within a parasite population maintained under

different conditions by the means of flow cytometry; and 3) employ the gathered information to generate a predictive mathematical model of *clag3* expression dynamics, switching rates and relative fitness advantage of different expression patterns under different conditions.

At present, we have successfully designed and tagged each gene of this project in individual parasite lines and we have managed to produce a triple transgenic line with all three editions simultaneously. However, we have not been able to continue with the second stage of this project because of two main problems that have led to our triple transgenic line presenting a series of defects that need to be addressed. The first issue revolves around the difficulties we have had to perform sequential gene editions in the same parasite line given the limited number of selectable markers in *P. falciaprum*[434]. The second involves the integration of multiple copies of one of the CRISPR-Cas9 plasmids into the *clag3.2* locus resulting in the disruption of the natural expression dynamics of this gene. While we have found ways to overcome the first problem, we are still working on how to deal with the second.

All in all, in this results summary we will: 1) fully describe the CRISPR-Cas9 designs for our *clag3* triple transgenic parasite line; 2) detail the main transgenic lines we have created; 3) explain the two main issues we have encountered alongside the different solutions we have used to try and resolve them; 4) present two exploratory experiments we performed to find new conditions with which to select different *clag3* expression patterns; and 5) define the next steps we believe to be necessary in order to obtain this tool as well as the future experiments we would like to perform once it has been generated. Finally, we will also portray all the methods used to carry out the experiments involved in these different sections.

2. CRISPR-Cas9 strategies

Briefly, in *Plasmodium* spp., genomic edition using the CRISPR-Cas9 system involves providing parasites with the following three elements: the Cas9 nuclease that can produce a double stranded break (DSB) in the parasite genome, a single guide RNA (sgRNA) that leads the Cas9 to a specific target in the genome where the DSB is to be made and a donor template for the cellular repair machinery to utilise in order to close this DSB and avoid parasite death[434,435]. There are many different strategies that have been used to deliver these elements to parasites that involve the use

of different plasmids with different components. For our triple transgenic parasite line, we have used a CRISPR-Cas9 strategy based on the one designed by Ghorbal *et al.*[435] that was later modified by Lim *et al.*[436] and then by Knuepfer *et al.*[437]. This strategy entails the use of two plasmids, referred to hereafter as the donor plasmid and the Cas9 plasmid (**Fig. 30**). Although the specific details of each tagging design will be detailed in this section, the basic components of each of these two plasmids were more or less the same in all cases.

In our designs, the donor plasmid includes the fluorescent protein tag or other elements that are to be incorporated into the parasite genome and it is flanked by two homology regions (HR). These HRs consist of sequences of approximately 300 to 500 bp that can be found in the parasite genome, one upstream of the Cas9 cleavage site and the other downstream, and which act as the repair template in order for the DSB to be fixed. On the other hand, the Cas9 plasmid used in our experiments derives from the pDC2-Cas9-hDHFRyFCU plasmid from Knuepfer *et al.*[437] and carries three different expression cassettes, each with their own promoter and terminator, that control the expression of the following elements: the Cas9 nuclease driven by a CAM promoter, the sgRNA driven by a short U6 promoter and the dual selectable marker hDHFR-yFCU driven by the PbDT promoter.

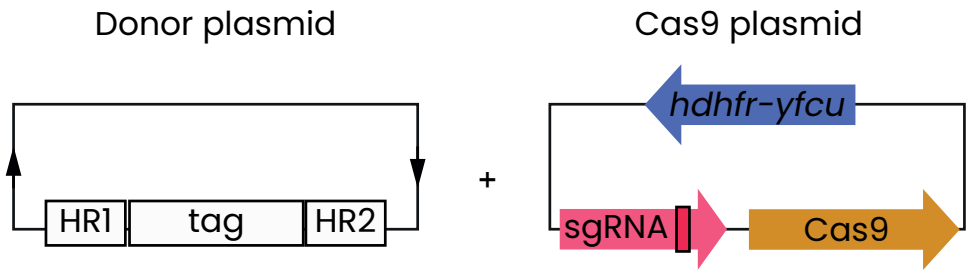


Fig.30. Plasmids used in our CRISPR-CAS9 strategy. On the left is the donor plasmid with the tag to be incorporated into the parasite genome flanked by the two homology regions. On the right is the Cas9 plasmid with its three expression cassettes including the Cas9 nuclease, the sgRNA (whose sequence conferring specificity for its target site is represented by the red box) and the fusion gene hdhfr-yfcu dual selectable marker. These plasmids derive from those designed by Ghorbal *et al.*[435], which Lim *et al.*[436] later modified by adding the sgRNA expression cassette to the Cas9 plasmid and that were then edited by Knuepfer *et al.*[437] who added the hDHFR-yFCU dual selectable marker also to the Cas9 plasmid.

The dual selectable marker results from the fusion of the positive selectable marker hDHFR (human dihydrofolate reductase) and the negative selectable marker yFCU (yeast cytosine deaminase/uridyl phosphoribosyl transferase) and allows for sequential editions of the same parasite line[435]. First, the positive selectable marker hDHFR confers resistance to the antifolate drug WR99210 and enables the selection of those parasites that have successfully uptaken the Cas9 plasmid (and most of them also the donor plasmid). Second, the negative selectable marker yFCU makes parasites susceptible to the drug 5-fluorocytosine (clinical grade Ancotil, Mylan N.V.) and therefore, serves to eliminate parasites that despite being correctly edited maintain the Cas9 plasmid in an episomal form and express the yFCU marker (negative selection). More details of this dual selection will be given in section 4.1 as its failure to work efficiently has been one of the two main problems of this project. This is because the presence of residual episomal Cas9 plasmid due to a not 100% efficient negative selection with 5-fluorocytosine renders parasites resistant to WR99210 so that this dual selection system cannot be reutilised in the same parasite line.

2.1. C-terminal tagging of *clag3.2* with eYFP

For this design we used a sgRNA that recognises a sequence located at -30 to -11 bp upstream from the *clag3.2* stop codon. On either side of the Cas9 cleavage site we defined two homology regions: HR1 (-827 to -34 bp upstream from the *clag3.2* stop codon) and HR2 (+1 to +485 bp downstream from the *clag3.2* stop codon). Together, HR1 and HR2 flank an in-frame recodonized version of the sequence from the end of HR1 to the stop codon (excluded), followed by the eYFP coding sequence (**Fig. 31**).

The donor plasmid for this tagging (pHR1clag3.2-eYFP-HR2clag3.2) was derived from the pHRap2g-eYFP plasmid[403] in two steps. Firstly, the *pfap2-g* HR1 was removed using the restriction sites Spe I and Bgl II and replaced with the *clag3.2* HR1 that was amplified with primers P1 and P2 (which included the recodonized fragment after HR1). Second, the *pfap2-g* HR2 was eliminated with restriction sites Aat II and Xho I and replaced by the *clag3.2* HR2, that was amplified with primers P3 and P4. The Cas9 plasmid for this design (pDC2-cam-Cas9-U6-hDHFRyFCU_sgRNA_clag3.2) derived from the pDC2-Cas9-hDHFRyFCU plasmid[437], in which we cloned our *clag3.2* sgRNA target sequence (included in annealed oligonucleotides P5 and P6) at the BtgZ I restriction site.

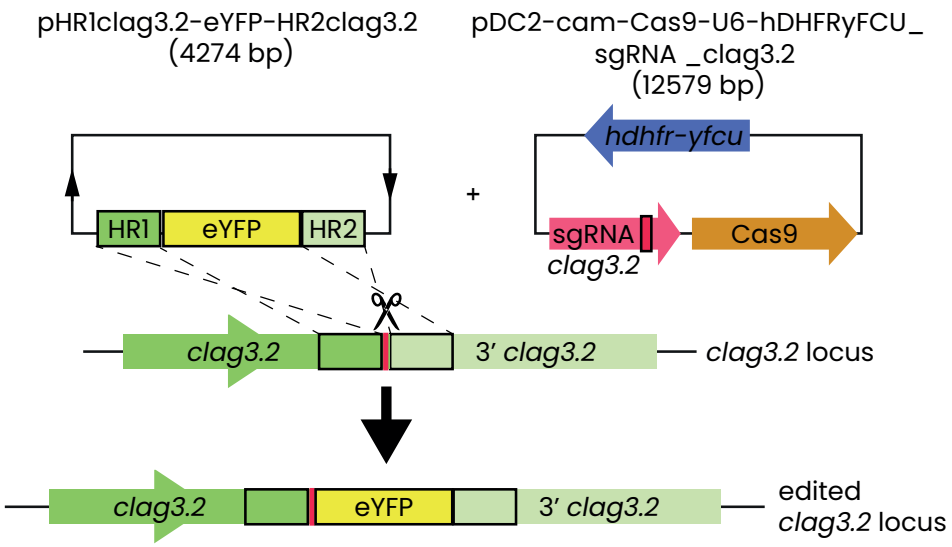


Fig.31. CRISPR-Cas9 design for *clag3.2* C-terminal tagging with eYFP. The donor plasmid on the left and the Cas9 plasmid on the right have the same structure as those in Fig.30. The red box in the *clag3.2* gene and the scissors mark the sgRNA sequence where Cas9-mediated cleavage is expected. The *clag3.2* downstream region (3' *clag3.2*) includes the endogenous terminator of the gene that remains intact after editing.

2.2. C-terminal tagging of *clag3.1* with mCerulean

For this tagging, we selected a sgRNA that recognises a sequence located at -30 to -11 bp upstream from the *clag3.1* stop codon. The sgRNA sequences of both *clag3* genes are highly similar and are only different in three nucleotides, one of which is located -4 bp from the PAM site. This alongside the differences in the HRs (especially HR2) should provide sufficient specificity to our design. Once again, on either side of the Cas9 cleavage site, HR1 (-840 to -34 bp upstream from the *clag3.1* stop codon) and HR2 (+27 to +692 bp downstream from the *clag3.1* stop codon) flank an in frame recodonized version of the sequence from the end of HR1 to the stop codon (excluded), followed by the mCer coding sequence and, after this, the 26 bp from the stop codon to the beginning of HR2 (**Fig. 32**).

The donor plasmid for this tagging (pUC19_HR1_clag3.1_ mCer_HR2_ clag3.1) was derived from the pUC19 plasmid in two steps. First, mCer was

amplified from the mCeruleanC1 plasmid (kindly donated by A. Vaquero at the Josep Carreras Leukaemia Research Institute) using primers P7 (which included the recodonized fragment after HR1, the Bgl II restriction site and a linker formed of 6 alanines with which to join the tag to the CLAG3.1 protein) and P8 (which added a stop codon and the restriction site Xho I). This mCer fragment alongside the *clag3.1* HR1 (amplified with primers P9 and P10) was cloned into PUC19 at the BamH I restriction site, which was maintained. Secondly, the *clag3.1* HR2 that was amplified with primers P11 (that included the 26bp from the stop codon to the beginning of HR2) and P12, was introduced into this intermediate plasmid directly after the mCer fragment at the Xba I restriction site, which was also maintained. The Cas9 plasmid (pDC2-cam-Cas9-U6-hDHFRyFCU_sgRNA_clag3.1) was derived from the pDC2-Cas9-hDHFRyFCU plasmid[437] in which we cloned our *clag3.1* sgRNA target sequence (included in annealed oligonucleotides P13 and P14) at the BtgZ I restriction site.

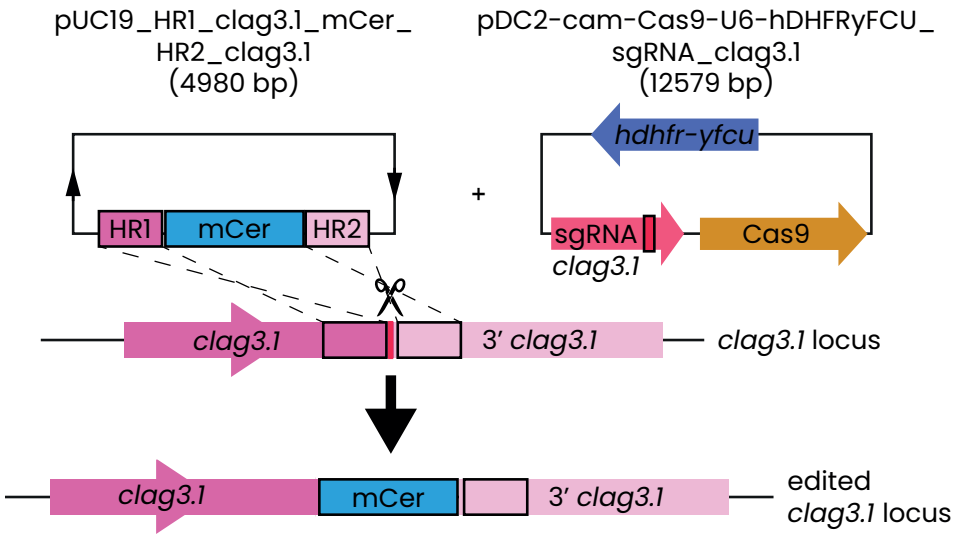


Fig.32. CRISPR-Cas9 design for *clag3.1* C-terminal tagging with mCer. The donor plasmid on the left and the Cas9 plasmid on the right have the same structure as those in Fig.30. The red box in the *clag3.1* gene and the scissors mark the sgRNA sequence where Cas9-mediated cleavage is expected. The *clag3.1* downstream region (3′ *clag3.1*) includes the endogenous terminator of the gene that remains intact after editing

2.3. Control gene strategies

For our control gene we had to select a gene with a similar time of expression to that of *clag3* genes, preferably slightly later, as this would enable us to differentiate between parasites mature enough to express *clag3* genes and parasites too immature to do so. We selected various control gene candidates and we attempted two different tagging approaches.

2.3.1. Strategy I: Control gene C-terminal tagging

This first approach consisted in a C-terminal tagging of the control gene with the fluorescent tag tdTomato (tdTom), similar to that described for the two *clag3* genes. The potential control gene candidates for this strategy were *rhoph2*, *rhoph3* and *ama1*. The reason behind the selection of the first two was that both RhopH2 and RhopH3 proteins form part of the RhopH complex alongside CLAG3[377,378] and therefore they all have a very similar time of expression and trafficking within the parasite. With regards to *ama1*, this gene is a schizont marker that is expressed later than *clag3* genes during the IDC and is frequently used as a maturity control for *clag3* expression in our lab, both for RT-qPCR analysis and for the IFA described in Article 1 of this thesis. In the end we selected *rhoph3* as our control gene for this strategy because the distribution of introns in *rhoph2* made the CRISPR-Cas9 design too complicated and with regards to *ama1*, prior episomal C-terminal tags have been found to interfere with the function of this protein[438,439].

For the *rhoph3* tagging, we chose a sgRNA that recognises a sequence located immediately upstream from the *rhoph3* stop codon (at -1 to -20bp). On either side of the Cas9 cleavage site, HR1 (-507 to -24 bp upstream from the *rhoph3* stop codon) and HR2 (+1 to +465 bp downstream from the *rhoph3* stop codon) flank an in frame recodonized version of the sequence from the end of HR1 to the stop codon (excluded) followed by the tdTom coding sequence (**Fig. 33**).

The donor plasmid for this tagging (pHR1rhoph3-TdTomato-HR2rhoph3) was derived from the pHRap2g-tdTom plasmid[403] in two steps. Firstly, the *pfap2-g* HR1 was removed using the restriction sites Spe I and Bgl II and replaced with the *rhopH3* HR1 that was amplified with primers P15 and P16.

Second, the *pfap2-g* HR2 was eliminated with restriction sites Aat II and Xho I and replaced by the *rhoph3* HR2, that was amplified with primers P17 and P18. The Cas9 plasmid for this design (pDC2-cam-Cas9-U6-hDHFRyFCU_sgRNA_rhoph3) derived from the pDC2-Cas9-hDHFRyFCU plasmid[437] in which we cloned our *rhoph3* sgRNA target sequence (included in the annealed oligonucleotides P19 and P20) at the BtgZ I restriction site.

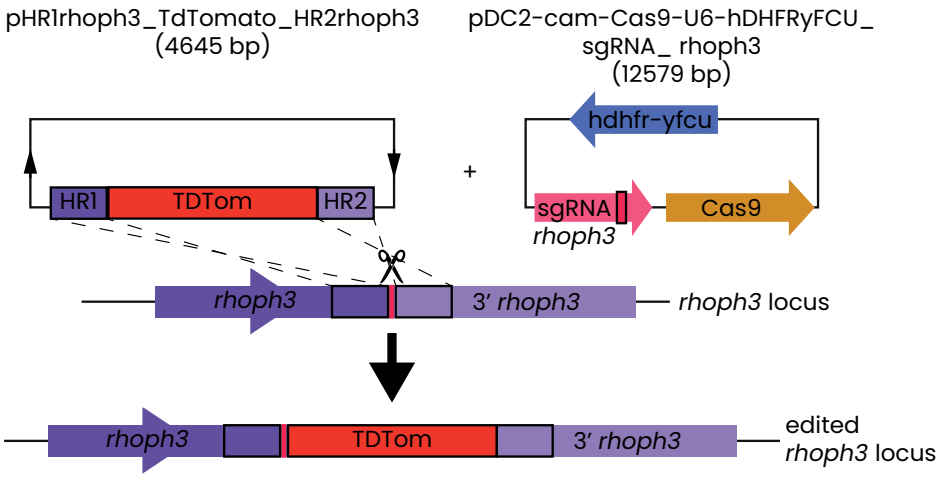


Fig.33. CRISPR-Cas9 design for *rhoph3* C-terminal tagging with tdTomato. The donor plasmid on the left and the Cas9 plasmid on the right have the same structure as those in Fig.30. The red box in the *rhopH3* gene and the scissors mark the sgRNA sequence where Cas9-mediated cleavage is expected. The *rhopH3* downstream region (3' *rhopH3*) includes the endogenous terminator of the gene that remains intact after editing

2.3.2. Strategy II: Control gene promoter controlling tdTom expression integrated in the *lisp1* locus

This second strategy is based on a previous design from our lab[440] that consists in the integration of the promoter of the control gene regulating the expression of tdTom into the *liver specific protein 1* (*lisp1* - PF3D7_1418100) locus that is dispensable in asexual blood stages[441]. In this manner, when the control gene's endogenous promoter is activated and the endogenous protein is expressed, the integrated protomer in *lisp1* is also activated and tdTom is expressed (**Fig. 34**).

To select the control gene for this strategy, we used the Bozdech & Llinas

et al. microarray dataset[217] and the Bártfai *et al*. RNA seq dataset[442] (both available on PlasmoDB) in order to find genes with peak expression just after *clag3.1*, that had similar transcript levels to those of *clag3.1* at the time of peak expression, that did not present clonally variant expression and that had a known function so as to not select a non-essential gene. Taking all of this into account, we chose the following three genes: *rap2* (PF3D7_0501600), *gap45* (PF3D7_1222700) and *ama1* (PF3D7_1133400).

The next step was to define the promoter regions of these genes and to design primers to amplify them. In order to do so, we took into account the available transcription start sites (TSS) on PlasmoDB from Adjalley *et al*.[443]. For *rap2* the TSS is around -1000 bp from the ATG, therefore we used primers (P21 and P22) that amplified from -1671 to -1 bp. For *gap45* there appear to be two TSS clusters, one at approximately -800 bp from the ATG and the other at around -2000 bp from the ATG. As a result, we did two approaches: "GAP45-short" amplifying from -1760 to -1 bp (P23 and P25) and "GAP45-long" amplifying from -3010 to -1 bp (P24 and P25). Finally, for *ama1* there are predicted TSSs that reach as far a -600 bp from this gene's ATG, so we designed primers (P26 and P27) that amplified from -1570 to -1 bp.

The plasmid construction for this strategy was relatively simple as it only required a single modification of the pfs16-tdTomato-lisp1 donor plasmid[440]. This modification consisted in the removal of the *pfs16* promoter using restriction sites Sal I and Not I and substituting it for the promoter of the different control gene candidates. Finally, the *lisp1* sgRNA and the Cas9 plasmid containing it (pDC2-Cas9-hDHFRyFCU-lisp1) were the same as previously described[440]. However, we also used two other versions of this plasmid generated by fellow members of our lab, in which the hDHFR-yFCU cassette was substituted for a blasticidin deaminase or a PAC cassette enabling us to use a different selection drug. The resulting Cas9 plasmids were pDC2-Cas9-BSD_lisp1 and pDC2-Cas9-PACyFCU_lisp1.

2.3.3. Strategy II: Unexpected *rif* sequence

During the generation of the plasmids for this second strategy, we found that before the tdTom sequence there was an unidentified sequence that corresponded to the initial fragment of the *rif* gene PF3D7_0115300. This fragment was traced back to the original plasmid from which the pfs16-tdTomato-lisp1 plasmid was derived[444], and appears to have been accidentally maintained in the other intermediate versions of this

plasmid[445] which were then kindly given to our lab. Despite the fact that this fragment was integrated in all the donor plasmids initially generated for this control strategy, fortunately one of the *E. coli* colonies produced for the *ama1* promoter cloning spontaneously lost this fragment. Therefore, we were able to repeat the production of the donor plasmids in exactly the same manner as previously described but using our *ama1* maxiprep 3 as the initial plasmid and swapping the *ama1* promoter for that of *gap45* (both long and short versions). To save time, we abandoned the *rap2* promoter strategy. The final plasmids designed for strategy II were: ama1_TdTomato_LISP1_HR1_HR2, gap45_short_TdTomato_LISP1_HR1_HR2 and gap45_long_TdTomato_LISP1_HR1_HR2.

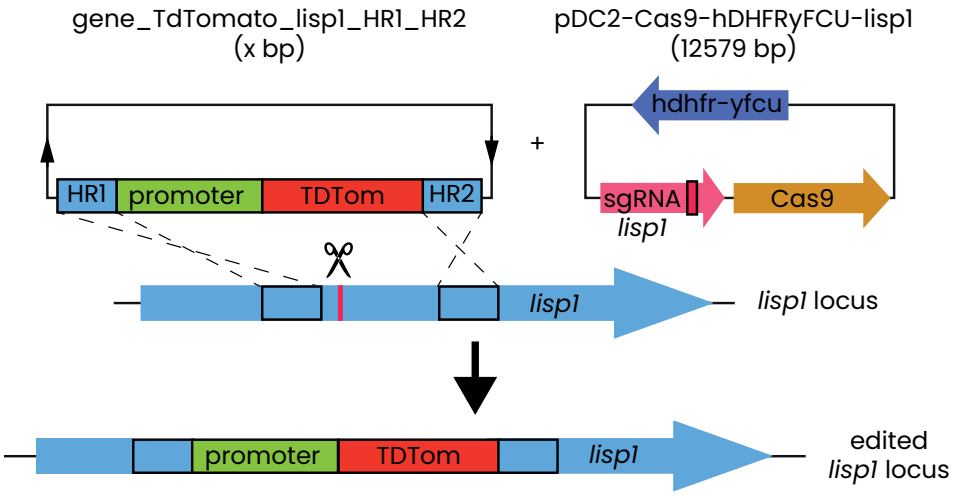


Fig.34. CRISPR-Cas9 design for the integration of a promoter controlling tdTomato expression in the *lisp1* locus. The donor plasmid on the left and the Cas9 plasmid on the right have the same structure as those in Fig.30. The size (bp) of the donor plasmid varies depending on the length of the selected gene promotor. The red box in the *lisp1* gene and the scissors mark the sgRNA sequence where Cas9-mediated cleavage is expected.

2.4. CRISPR-Cas9 strategies for eliminating parasites with multiple copies of donor plasmid in the *clag3.2* locus

Finally, in addition to the designs for each of the three taggings of our triple transgenic parasite line, we also created two additional CRISPR-Cas9 strategies in an attempt to solve our issue with multiple integrations.

Although this problem will be described in further detail in section 4.2, in brief it consists in multiple copies of the pHR1clag3.2-eYFP-HR2clag3.2 donor plasmid being integrated for unknown reasons into the *clag3.2* locus and hindering the normal expression of this gene.

2.4.1. Plasmids with sgRNAs targeting the backbone of the donor plasmid

This strategy consists in using a sgRNA to target a sequence in the backbone of the pHR1clag3.2-eYFP-HR2clag3.2 donor plasmid that is not normally integrated into the parasite genome when the eYFP tag is incorporated as expected (a single copy). However, if multiple copies of this plasmid are integrated into the *clag3.2* locus, this backbone-targeting sgRNA can lead the Cas9 to produce a DSB that would kill those parasites that present this defect due to the absence of HRs to act as a template to repair it. For this strategy we selected two sgRNA targets. The first called sgRNA-episome3 targeted a sequence already present in the donor plasmid backbone, located at the very end of HR2 (with the PAM cleavage site located outside of this region) and it was generated using the annealed oligonucleotides P28 and P29. As we were concerned that this proximity between the sgRNA target and HR2 could lead to cleavage of correctly edited parasites, we decided to design a new target for a sgRNA-episome4 that we had to clone into the donor plasmid. In order to clone the "episome 4" target into the pHR1clag3.2-eYFP-HR2clag3.2 donor plasmid, we started by linearising this plasmid at the Pvu II restriction site. Next, the sgRNA-episome4 target was originated using annealed oligonucleotides P32 and P33 that included both the target as well as maintaining the Pvu II and adding the Bgl II restriction sites. The annealed oligonucleotides were then introduced into the linearised donor plasmid, generating the final donor plasmid pHR1clag3.2-eYFP-HR2clag3.2_PvuII_sgRNA4. Finally, the targets of both sgRNA-episome3 and sgRNA_episome4 were each incorporated into a different pDC2-Cas9-hDHFRyFCU plasmid[437], as described with the other Cas9 plasmids of this project, generating plasmids pDC2-cam-Cas9-U6-hDHFRyFCU_sgRNA_episome3 and pDC2-cam-Cas9-U6-hDHFRyFCU_sgRNA_episome4.

2.4.2. Incorporation of the negative selectable marker yFCU into the pHR1clag3.2-eYFP-HR2clag3.2 donor plasmid

The other strategy to eliminate parasites with multiple integrations

consists in introducing the yFCU negative selectable marker cassette into the pHR1clag3.2-eYFP-HR2clag3.2 donor plasmid. In this manner, if multiple copies of this plasmid are incorporated into the *clag3.2* locus, the yFCU cassette will also be incorporated and treatment with 5-fluorocytosine will kill these parasites. To incorporate the yFCU negative selectable marker cassette into the pHR1clag3.2-eYFP-HR2clag3.2 donor plasmid, first the yFCU cassette had to be isolated from the plasmid pL-6_eGFP[435] using restriction sites Not I and Sac II. Next, these restriction sites had to be added to the pHR1clag3.2-eYFP-HR2clag3.2 donor plasmid so that the yFCU cassette could be incorporated into it. Therefore, the pHR1clag3.2-eYFP-HR2clag3.2 donor plasmid was linearised at the restriction site Pvu II and an oligonucleotide containing the restriction sites Not I, Sac II and maintaining the original Pvu II (originated by annealing P34 and P35), were introduced into said linearised plasmid. This new intermediate plasmid was then sequentially cut at the restriction sites Not I and Sac II to clone the yFCU selectable marker cassette, resulting in the final plasmid: pHR1clag3.2-eYFP-HR2clag3.2_PvuII _yFCU.

3. Transgenic parasite lines

So far during this project, a total of 44 transfections (summarised in **Supplementary Table 1**) have been performed in the attempt to obtain the desired triple transgenic line with all three genes correctly edited. In this section, however, we will only describe the most important lines that have been generated so far: the single transgenic parasite lines that each have one of the taggings described in this project and that serve as gating controls for flow cytometry, two double transgenic parasite lines with two simultaneous editions and the only triple transgenic line (with three simultaneous editions) that we have generated to date. A visual summary indicating the origin, characteristics and problems of each of these lines can be found at the end of this chapter.

Unless stated otherwise, the routine procedure for the generation and characterisation of these lines was as follows. First, all transfections were performed by standard electroporation at the ring stage followed by a 4-day treatment with WR99210 (see Methods). Next, if a few weeks later parasites were detectable by light microscopy, genomic DNA (gDNA) was routinely extracted and analysed by PCR to determine if parasites had been correctly edited. If this was the case, the transgenic lines were then characterised by RT-qPCR analysis to determine *clag3* expression

patterns and by IFA and / or flow cytometry to detect the fluorescence of the different tags present in each line. Also, additional experiments and treatments were performed in order to address the two main problems of this project: residual episomal Cas9 plasmid and multiple integrations of donor plasmid in the *clag3.2* locus. These experiments will be further detailed in sections 4.1 and 4.2 respectively, but we will briefly describe them here in order to facilitate the understanding of this section.

In order to eliminate episomal Cas9 plasmid, parasites were submitted to a 10-day treatment with 5-fluorocytosine. Presence of this episome was then measured by RT-qPCR and then actions such as big dilutions were performed in order to obtain a subpopulation of episome-free parasites. With regards to multiple integrations in the *clag3.2* locus, these were detected by performing two types of PCRs: one with external primers that fall outside of the edited area of the locus (whose resulting amplicon informs on the length of the sequence integrated into the parasite genome), and the other with an external primer and a primer targeting the plasmid backbone (that only results in amplification if multiple integrations are present). Depending on the results of these PCRs, actions such as sorting and subcloning were performed in order to obtain subpopulations of multiple integration-free parasites.

Finally, it is important to mention that all the experiments performed to characterise these transgenic parasite lines generally consisted of only one biological replicate and, therefore, could not be used to draw firm biological conclusions. Subsequently, we have only used the obtained results as a guide to help direct the development of this tool and not as a means to make definitive conclusions on the characteristics of each parasite line. When we finally obtain our definitive triple transgenic line, all characterisation experiments will be performed in triplicate so that the obtained results can be considered conclusive.

3.1. Single transgenic I: *clag3.2* + eYFP (T2)

Parasite line T2 was generated when the 3D7-A parasite line (a transcriptionally diverse stock of the 3D7 clonal line)[322] was transfected with 12 µg of donor plasmid pHR1clag3.2_eYFP_HR2clag3.2 linearised at restriction site Pvu I in the plasmid backbone and 60 µg of the pDC2-cam-Cas9-U6-hDHFRyFCU_sgRNA_clag3.2 plasmid. Parasites were detected by light microscopy approximately 3 weeks after transfection

and PCR analysis of extracted gDNA showed that our construct had been correctly integrated in all parasites as no wild type band was detected (**Fig. 35**).

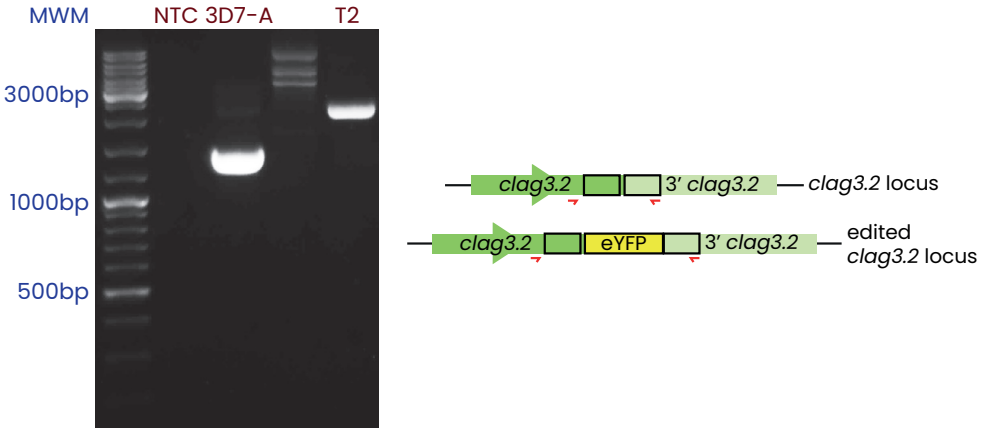


Fig. 35. Diagnostic PCR analysis to validate the integration of eYFP in the *clag3.2* locus in the T2 transgenic line. On the left, agarose gel with the expected amplicons of 1446 bp and 2190 bp in 3D7-A (parental line) and T2, respectively. Amplified with primers P66 and P57. NTC = negative control, MWM = molecular weight marker. On the right, schematic indicating the position of the primers (red) within the wild-type and edited *clag3.2* locus. Boxes indicate position of HR1 (left) and HR2 (right).

Next, we attempted to visualise eYFP fluorescence directly by live-cell fluorescence analysis but were unsuccessful. Therefore, we performed an IFA in which we used α-CLAG3 and α-GFP antibodies. Due to the sequence similarity between GFP and eYFP[446], it is possible to use the same antibody to detect both fluorescent proteins. To our surprise, only 6% of the CLAG3-expressing parasites presented α-GFP signal (**Fig. 36**). In the original 3D7-A parasite line, there is approximately a 50-50 distribution of parasites that express *clag3.1* and *clag3.2*. Therefore, we expected the T2 line to have around 50% of *clag3.2*+eYFP expressing parasites as post-transfection there were no wild-type parasites with an untagged *clag3.2*. The fact that this value was much lower could either be because the majority of the T2 parasite population had switched to expressing *clag3.1* (possibly an indirect consequence of the transfection bottleneck) or that there was an issue with the expression of the *clag3.2*+eYFP construct.

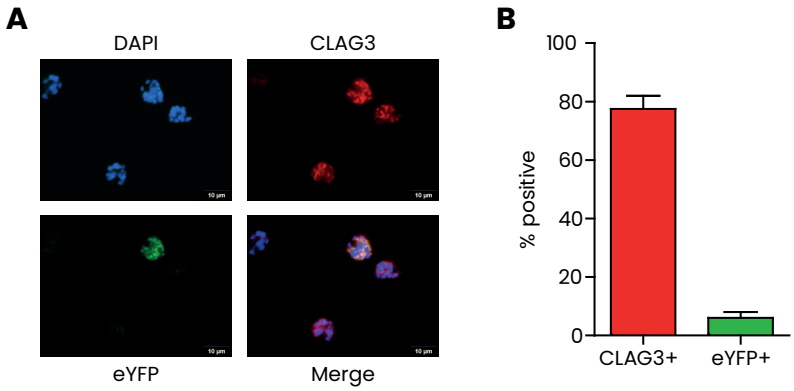


Fig. 36. CLAG3 and eYFP expression in the T2 transgenic line. (A) IFA analysis of mature schizont-infected erythrocytes with α-CLAG3 and α-GFP antibodies. DAPI was used to mark the parasite nuclei. The merge is of the three channels. (B) Proportion of parasites that were positive for CLAG3 expression (red) or positive for eYFP fluorescence (green). All eYFP positive parasites were also CLAG3 positive. Shown are the average and s.e.m of two technical replicates.

To determine whether there had been a significant change in *clag3* expression patterns, we performed a RT-qPCR analysis of *clag3.2* and *clag3.1* expression levels in the T2 and the parental 3D7-A parasite lines (**Fig. 37**). We found that there was very little difference between the two parasite lines and consequently a switch from *clag3.2* to *clag3.1* was not the reason behind the low percentage of *clag3.2*+eYFP-expressing parasites in our IFA experiment.

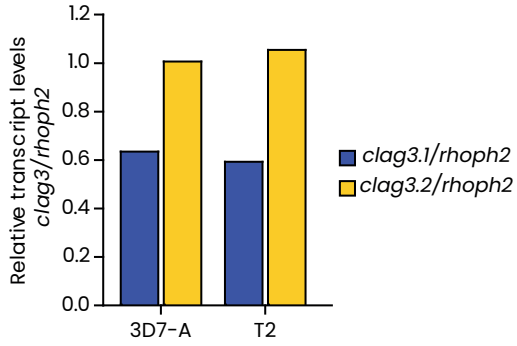


Fig. 37. RT-qPCR analysis of *clag3* expression in the T2 transgenic line. Shown are the relative transcript levels of *clag3.1* (blue) and *clag3.2* (yellow) normalized against *rhoph2* in both the parental 3D7-A line and the T2 line. (Primers: P66, P50, P65, P86 and P87).

In further PCR analysis we found that in the T2 line there were multiple integrations of the donor plasmid in the *clag3.2* locus and therefore it is possible that said integrations could be interfering with the correct expression of our tagged CLAG3.2. Moreover, the T2 line also had residual episomal Cas9 plasmid which we were unable to eliminate by our standard procedure of 5-fluorocytosine treatment followed by a big dilution (see section 4.1). Therefore, in an attempt to solve all these problems in one go, we performed a sorting experiment with the objective of obtaining a T2 parasite line with eYFP fluorescence in all *clag3.2* expressing parasites, without multiple integrations and without residual episomal Cas9 plasmid. Using strict gates to sort for parasites expressing eYFP, we started four separate cultures with either 1000, 100, 10 or 1 sorted parasite(s) (**Fig. 38**).

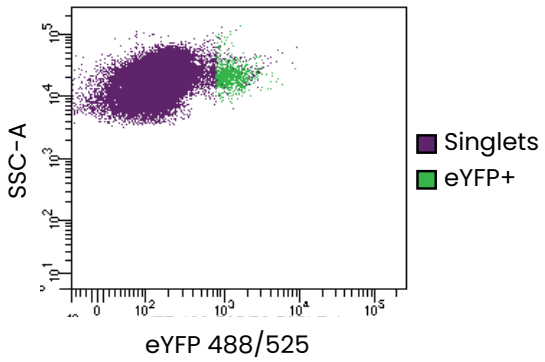


Fig. 38. Sorting of the T2 transgenic line for eYFP expressing parasites. On the X axis eYFP fluorescence values and on the Y axis side scatter values. Parasites that fall within the strict eYFP expression gate are highlighted in green.

Once the sorted parasites had grown sufficiently to be characterised, we confirmed by PCR that they had the *clag3.2* locus edited. Next, we tested for the presence of multiple integrations and residual episomal Cas9 plasmid and found that only T2-1 (originated from one sorted parasite) was free of both. Subsequently, we continued with the determination of eYFP expression in this line. First, we performed the same IFA as before but only using anti-GFP antibodies (not anti-CLAG3) and compared the eYFP expression in the T2-1 line with that of the other sorted

lines (T2-1000, T2-100, T2-10, T2-1). While the proportion of GFP-positive parasites was higher than in the original T2 line, for reasons that we do not understand T2-1 had the lowest percentage of eYFP expressing schizonts and also the signal intensity of these schizonts was lower than that of the other parasite lines at the same exposure level (**Fig. 39**).

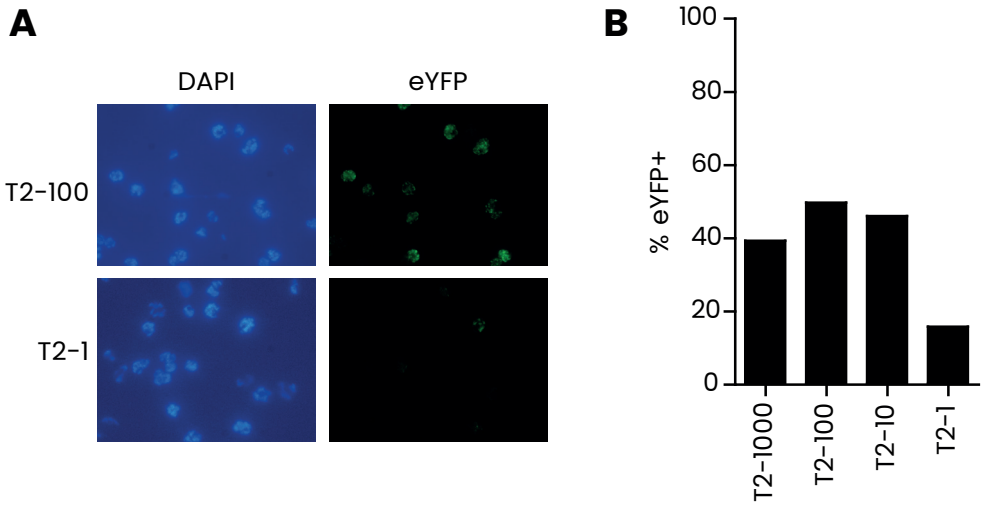


Fig. 39. eYFP expression in sorted T2 transgenic lines. (A) IFA analysis of mature schizont-infected erythrocytes with anti-GFP antibodies. DAPI was used to mark the parasite nuclei. (B) Proportion of parasites that were positive for eYFP expression

Second, to confirm these results, we analysed eYFP fluorescence of the T2-1 line by flow cytometry. We found that approximately only 10% of the analysed schizonts showed eYFP fluorescence (**Fig. 40**). We cannot discard in either experiment that this low percentage of eYFP-expressing parasites could be due to parasites not being mature enough to express *clag3.2* in the T2-1 line as we did not have a stage control. However, given that parasite cultures were tightly synchronised to a 5 h age window we took this as an indication that something in this line was not working correctly rather than it being due to parasite immaturity. Taking all these results into account, we decided that neither the initial T2 line nor the sorted T2-1 line were adequate for continuing developing our triple transgenic line with. That said, the initial T2 line can be used as a positive eYFP flow cytometry gating control in future experiments.

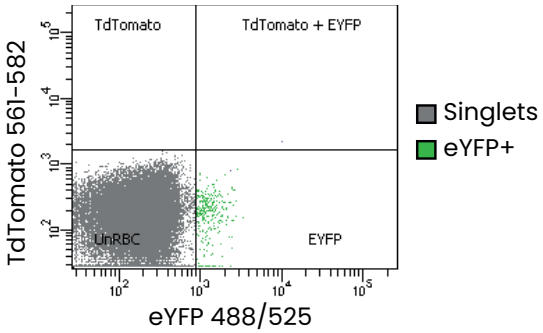


Fig. 40. Flow cytometry analysis of T2-1. On the X axis eYFP fluorescence values and on the Y axis tdTomato fluorescence values. eYFP positive parasites are highlighted in green. Parasites identified by MitoTracker staining.

3.2. Single transgenic II: *clag3.1* + mCer (T13)

Although it took us four attempts, the next single transgenic line T13 was achieved by transfecting 3D7-A parasites with 15 µg of donor plasmid pHR1clag3.1_mCer_HR2clag3.1 linearised at the Xmn I restriction site in the plasmid backbone and 60 µg of the pDC2-cam-Cas9-U6-hDHFRyFCU_sgRNA_clag3.1 plasmid. Parasites were detected by light microscopy 2 weeks after transfection and PCR analysis of extracted gDNA showed that our construct appeared to have been correctly integrated in all parasites (**Fig. 41**).

Next, as with T2, we set out to detect mCer fluorescence by IFA and flow cytometry (**Fig. 42**). In the IFA (identical to that performed for the T2 line), around 50% of the CLAG3-expressing parasites also showed mCer fluorescence (once again, due to their high sequence similarity, α-GFP antibodies can be used against the mCer tag[446]). Similarly, in the flow cytometry analysis, we found that there was mCer fluorescence in approximately 40% of schizonts.

To check that the observed mCer fluorescence corresponded with *clag3* expression patterns we performed a RT-qPCR analysis (**Fig. 43**). In both T13 and the parental 3D7-A line, there were similar transcript levels of both *clag3* genes, corresponding nicely with the IFA and flow cytometry results in which approximately 50% of parasites expressed CLAG3.1-mCer.

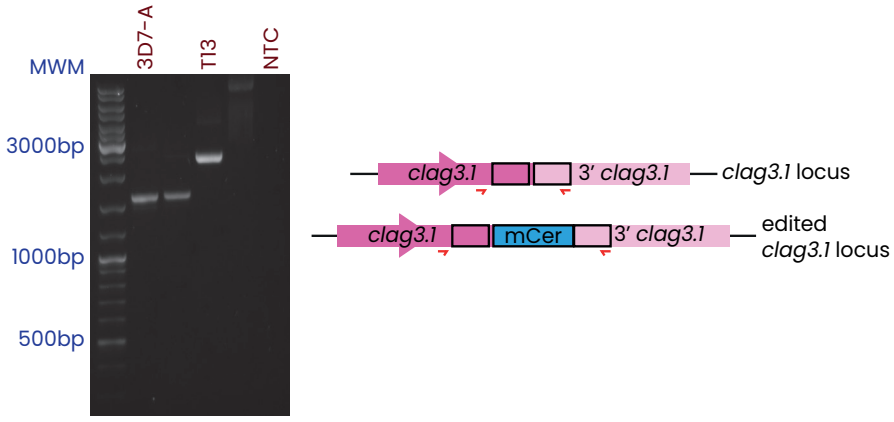


Fig. 41. Diagnostic PCR analysis to validate the integration of mCer in the *clag3.1* locus in the T13 transgenic line. On the left, agarose gel with the expected amplicons of 1662 bp and 2421 bp in 3D7-A (parental line) and T13, respectively. Amplified with primers P66 and P48. NTC = negative control, MWM = molecular weight marker. On the right, schematic indicating the position of the primers (red) within the wild-type and the edited *clag3.1* locus. Boxes indicate position of HR1 (left) and HR2 (right).

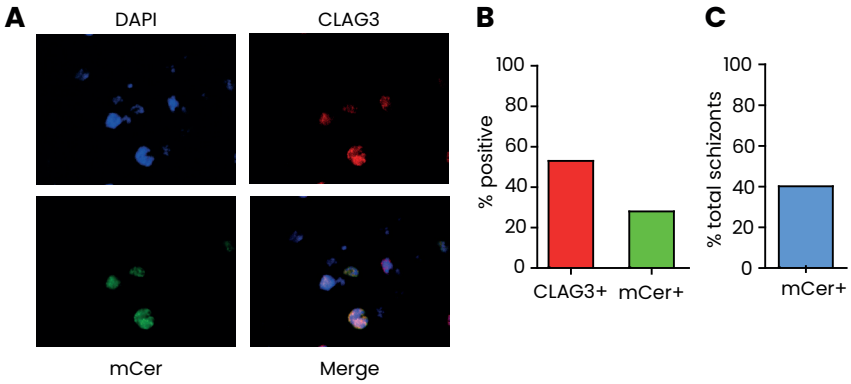


Fig. 42. CLAG3 and mCer expression in the T13 transgenic line. (A) IFA analysis of mature schizont-infected erythrocytes with α-CLAG3 and α-GFP antibodies. DAPI was used to mark the parasite nuclei. The merge is of the three channels. (B) Proportion of parasites that were positive for CLAG3 expression (red) or positive for mCer fluorescence (green). All mCer positive parasites were also CLAG3 positive. (C) Flow cytometry analysis showing the proportion of mCer+ schizonts relative to total schizonts identified by MitoTracker staining.

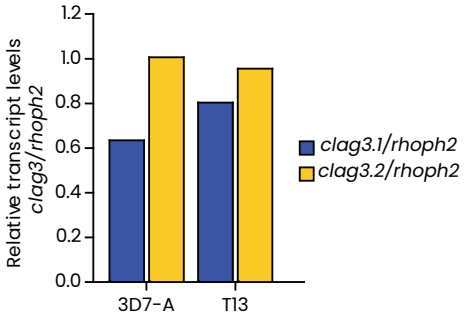


Fig. 43. RT-qPCR analysis of *clag3* expression in the T13 transgenic line. Shown are the relative transcript levels of *clag3.1* (blue) and *clag3.2* (yellow) normalized against *rhoph2* in both the parental 3D7-A line and the T13 line. (Primers: P66, P50, P65, P86 and P87).

Finally, PCR analysis determined that there were not multiple integrations of the pUC19_HR1_clag3.1_mCer_HR2_clag3.1 donor plasmid in the *clag3.1* locus but that the T13 line did have unedited parasites (we ignore why this did not show up in the first diagnostic PCR). Furthermore, RT-qPCR analysis (not shown) revealed that there was residual episomal Cas9 plasmid. Therefore, we treated the T13 parasite line for 10 days with 5-fluorocytosine and then performed big dilutions to 100 and 10 parasites (see section 4.1). This served to eliminate residual episomal Cas9 plasmid from this line and, in the case of the dilution to 10 parasites, to eliminate the rare unedited parasites from this line (**Fig. 44**). The resulting T13-10 line was free of residual episomal Cas9 plasmid, multiple integrations in the *clag3.1* locus and unedited parasites. Furthermore, this line showed adequate *clag3* expression and fluorescence profiles. Subsequently, we considered that this line could be used to continue developing our triple transgenic line with as well as to serve as a flow cytometry gating control in future experiments.

3.3. Single transgenic III: control gene + tdTom

3.3.1. C-terminal tagging of *rhoph3* with tdTomato

For this first control gene strategy, we transfected parasites with the pHR1rhoph3-TdTomato-HR2rhoph3 and pDC2-cam-Cas9-U6-hDHFRyFCU_sgRNA_rhoph3 plasmids. In total we performed 10 different transfection attempts using different parasite lines (3D7-A or the CLAG3.2-

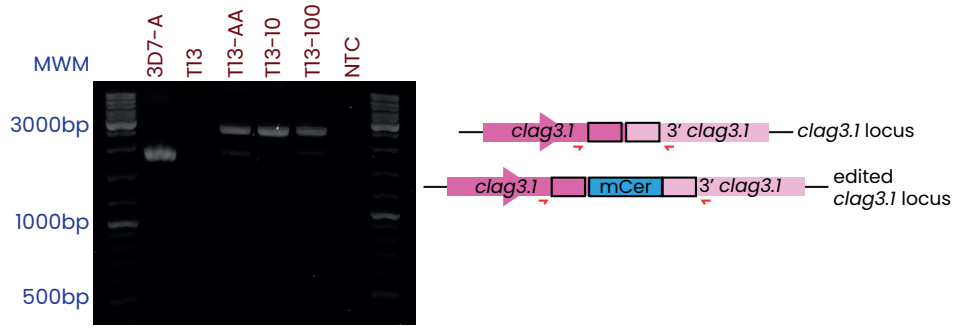


Fig. 44. Diagnostic PCR analysis to validate the integration of mCer in the *clag3.1* locus in the T13-derived transgenic lines. On the left, agarose gel with the expected amplicons of 1928 bp and 2688 bp in the 3D7-A (parental line) and T13-derived lines, respectively. Amplified with primers P66 and P49. T13-AA = T13 post 5-fluorcytosine treatment, T13-10/100 = T13 post 5-fluorocytosine and dilution to 10 or 100 parasites, NTC = negative control, MWM = molecular weight marker. On the right, schematic indicating the position of the primers (red) within the wild-type and edited *clag3.1* locus. Boxes indicate position of HR 1 (left) and HR2 (right).

eYFP transgenic line T2), with different amounts of donor plasmid (12, 15 or 30 µg) and performing both single or double transfections (i.e., simultaneously editing the *rhopH3* locus and one of the *clag3* loci). In 7 of the 10 transfections there was parasite growth of only unedited parasites and in the remaining 3 there was no parasite growth at all. We ignore why this strategy did not work but we hypothesise that the cause could be an inefficient sgRNA or that the tag itself interfered with the function of the RhopH3 protein which is essential in *P. falciparum*[394]. Taking all the above into account, this strategy was abandoned.

3.3.2. *ama1* promoter controlling tdTomato within the *lisp1* locus

With this strategy involving plasmids ama1_TdTomato_LISP1_HR1_HR2 and pDC2-Cas9-hDHFRyFCU-lisp1, we generated two transgenic parasite lines: T17 and T21. The first was a single transgenic line originated from the 3D7-A line and the second was a double transgenic line generated by transfecting the CLAG3.1-mCer transgenic line T13. In both cases, the construct was correctly integrated into the *lisp1* locus. However, we were unable to detect tdTom fluorescence in either of the two lines by IFA or flow cytometry and concluded that this construct was not functional.

3.3.3. *gap45* promoter controlling tdTomato within the *lisp1* locus

This final control gene strategy was the same as the previous one except that the *ama1* promoter was substituted with either the long or the short version of the *gap45* promoter. The transgenic lines T30 and T31 were generated with the long and short version of this promoter, respectively, and analysis were performed to determine which was the better control. In both cases, 15 µg of donor plasmid (gap45_short_TdTomato_LISP1_HR1_HR2 or gap45_long_TdTomato_LISP1_HR1_HR2) linearised at restriction site Sca I in the plasmid backbone and 60 µg of Cas9 plasmid (pDC2-Cas9-hdhfryfcu-lisp1) were used. Parasites were detected by light microscopy approximately 2 weeks post-transfection and diagnostic PCR showed that both transgenic lines had been edited correctly and there were not any unedited parasites (**Fig. 45**).

Next, we checked we checked tdTom fluorescence by flow cytometry and saw that both T30 and T31 had a similar signal intensity (**Fig. 46**).

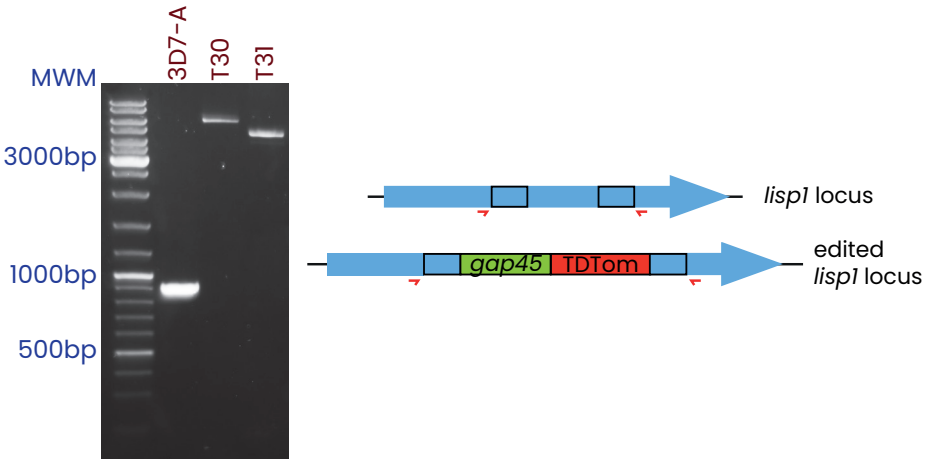


Fig. 45. Diagnostic PCR analysis to validate the integration of the short / long *gap45* promoter controlling tdTomato expression in the *lisp1* locus in the T30 and T31 transgenic lines. On the left, agarose gel with the expected amplicons of 918 bp, 5524 bp and 4274 bp in 3D7-A (parental line), T30 and T31, respectively. Amplified with primers P77 and P78. MWM = molecular weight marker. On the right, schematic indicating the position of the primers (red) within the wild-type and edited *lisp1* locus. Blue boxes indicate position of HRs and the green box with "*gap45*" indicates where the short or long version of the *gap45* promoter is located within this construct.

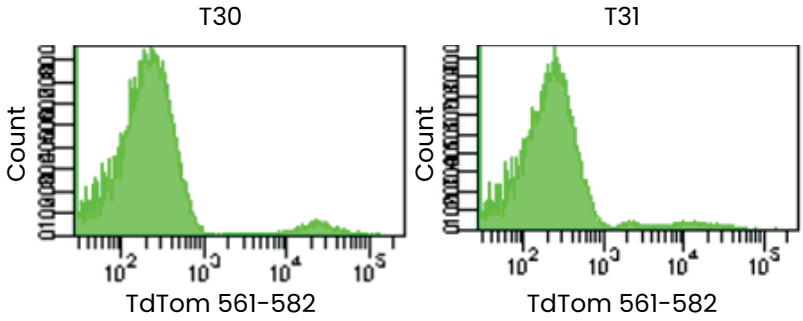


Fig. 46. T30 and T31 tdTom fluorescence histograms. On the X axis tdTom fluorescence values and on the Y axis cell count. Differences in the distribution of the tdTom fluorescence values between the two lines are probably due to stage differences.

Finally, we also analysed *clag3* and *tdTom* transcript levels by RT-qPCR (**Fig. 47**) and found that once again they were similar between the two lines. As a result, we concluded that both versions of the *gap45* promoter controlling tdTom expression were adequate to use as a stage control and that either of these lines could serve as tdTom expression gating controls for flow cytometry.

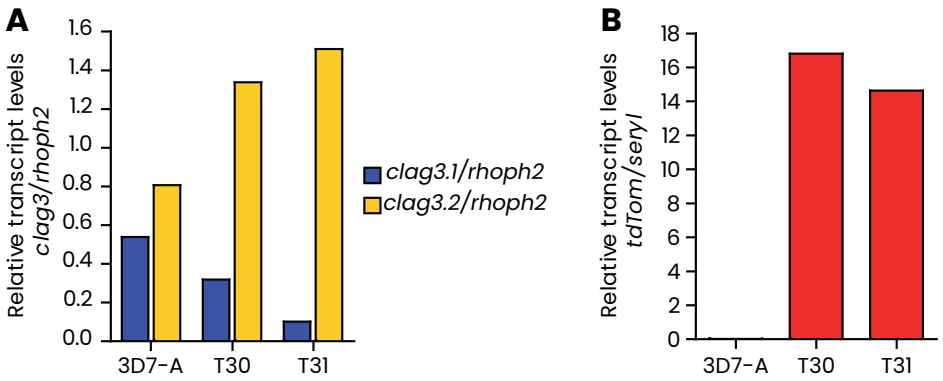


Fig. 47. RT-qPCR analysis of *clag3* and *tdTom* expression in the T30 and T31 transgenic lines. (A) Relative transcript levels of *clag3.1* (blue) and *clag3.2* (yellow) normalized against *rhoph2* in 3D7-A, T30 and T31. (B) Relative transcript levels of tdTom normalized against *serine-tRNA ligase* (*serrs*) in the same parasite lines. (Primers: P66, P50, P65, P86, P87, P88, P89, P91 and P92).

3.4. Double transgenic lines

We attempted to generate various different double transgenic lines throughout this project (all of which can be found in **Supplementary table 1**) but the majority either did not grow, were not correctly edited or did not show fluorescence. Although we tried to do two simultaneous editions to produce double transgenic lines, this never worked and, in the end, only sequential transfections were successful. Overall, the only functional double transgenic line we obtained was the line named T20 and it was generated after transfecting the CLAG3.1-mCer transgenic line T13-10 (submitted to 5-fluorocytosine treatment and big dilution to 10 parasites) with 15 µg of donor plasmid pHR1clag3.2_eYFP_HR2clag3.2 linearised at restriction site Pvu I in the plasmid backbone and 60 µg of the pDC2-cam-Cas9-U6-hDHFRyFCU_sgRNA_clag3.2 plasmid. Parasites were detected by light microscopy 2.5 weeks post-transfection and the first diagnostic PCR showed that all parasites had been correctly edited at the *clag3.2* locus. However, in this PCR we also observed an additional amplicon of a higher molecular weight indicating the possible presence of multiple integrations of donor plasmid in the *clag3.2* locus. We also checked if the *clag3.1* locus was still correctly edited with the mCer tag and this was the case (**Fig. 48**).

Next, we submitted the T20 parasite line to our routine 10-day treatment with 5-fluorocytosine to eliminate residual episomal Cas9 and generated the line T20-post. We then evaluated the fluorescence of the T20-post parasite line by both IFA and flow cytometry. In the IFA (identical to the ones performed for the T2 and T13 single transgenic lines), we found that in two separate replicates all CLAG3 positive schizonts using an α-CLAG3 antibody were also positive with an antibody against GFP (recognising both eYFP and mCer). This indicated that all parasites expressing CLAG3 were expressing a tagged version of these genes. However, the use of the α-GFP antibody meant that we could not distinguish between parasites expressing CLAG3.2-eYFP or CLAG3.1-mCer. In order to address this issue, we analysed our parasite line by flow cytometry and to our surprise we found that 70% of the analysed schizonts expressed CLAG3.1-mCer and only 3% CLAG3.2-eYFP (**Fig. 49**). Therefore, it appeared that parasites had switched to expressing predominantly CLAG3.1 or that there was a problem with the expression of CLAG3.2-eYFP, possibly due to the multiple integrations we had observed in our diagnostic PCR. The remaining 27% of schizonts that did not show *clag3* expression were probably not mature enough to express these genes.

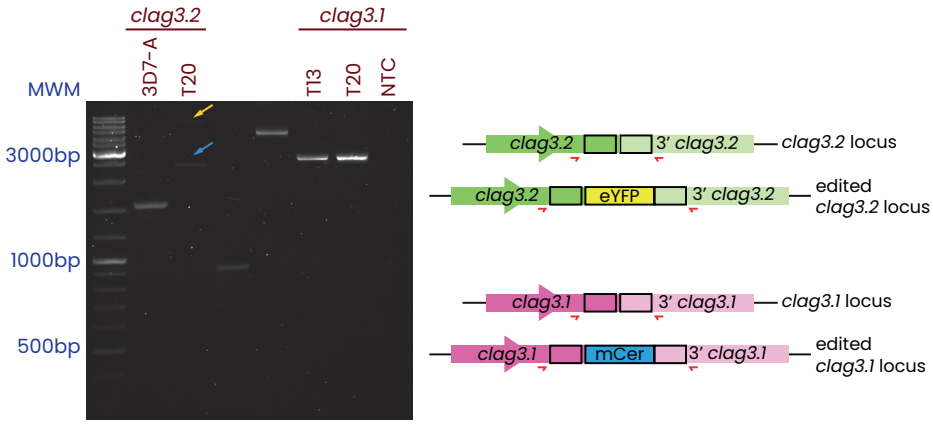


Fig. 48. Diagnostic PCR analysis to validate the integration of eYFP in the *clag3.2* locus and of mCer in the *clag3.1* locus in the T20 transgenic line. On the left, agarose gel with analysis of the *clag3.2* locus on the left and of the *clag3.1* locus on the right. In the *clag3.2* locus, expected amplicons of 1587 bp in 3D7-A (wild-type) and of 2331 bp in T20 (indicated with blue arrow) were observed. T20 presented an additional amplicon of 8000 bp (indicated with a yellow arrow) possibly due to the presence of multiple integrations of donor plasmid in this locus. (All amplified with primers P66 and P63). In the *clag3.1* locus, the expected amplicon of 2688 bp (corresponding to a correctly edited locus) was detected in both T13 (parental line) and T20 (amplified with primers P66 and P49). NTC = negative control, MWM = molecular weight marker. On the right, schematic indicating the position of the primers (red) within the wild-type and edited *clag3.2* and *clag3.1* loci. Boxes indicate position of HR 1 (left) and HR2 (right).

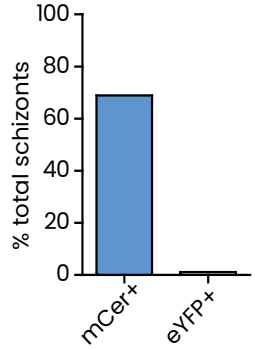


Fig. 49. Flow cytometry analysis of mCer and eYFP fluorescence in the T20-post transgenic line. Proportion of mCer+ and eYFP+ schizonts relative to total schizonts identified by MitoTracker staining.

In order to address why CLAG3.1-mCer expression was much higher than CLAG3.2-eYFP expression in the T20-post line, we performed a RT-qPCR analysis of these genes in this line and its parental lines (T20, T13 and 3D7-A) (**Fig. 50**). To our surprise, we found that post-transfection T20 parasites presented much lower transcript levels of *clag3.1* compared to the parental T13 line and that after the standard 5-fluorocytosine treatment to eliminate residual episomal Cas9 plasmid, this same T20 line had undergone a large switch from parasites expressing *clag3.2* to parasites expressing *clag3.1*. This explained the predominance of mCer fluorescence in the flow cytometry experiment. In order to determine whether the 5-fluorocytosine treatment was responsible for this switch we performed further experiments (see section 5.1) and found no relation between the two. Therefore, we ignore what produced this unexpected switch in *clag3* expression patterns between before and after 5-fluorocytosine treatment. Also, we observed that the overall *clag3* transcript levels in the T20 lines was slightly lower than in the parental lines, although, as we only performed one biological replicate, we cannot discard that this was due to a technical issue. However, if these lower *clag3* expression levels were real, we also ignore what could have caused them.

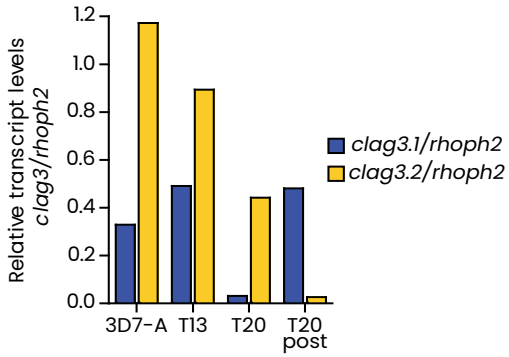


Fig. 50. RT-qPCR analysis of *clag3* expression in the T20 transgenic line. Shown are the relative transcript levels of *clag3.1* (blue) and *clag3.2* (yellow) normalized against *rhoph2* in the parental 3D7-A and T13 lines and the T20 transgenic line before (T20) and after (T20-post) 5-fluorocytosine treatment. (Primers: P66, P50, P65, P86 and P87).

With the overall general lower expression levels of both *clag3* genes and the presence of multiple integrations in the *clag3.2* locus in the T20 lines, we decided to perform a sorting experiment in order to obtain double-edited parasites that expressed CLAG3.2-eYFP, that did not have multiple

integrations nor residual episomal Cas9 plasmid and which could be used for the last sequential transfection to obtain our triple transgenic parasite line. As with the T2 line, we applied a strict gate for eYFP-expressing parasites and employed the original T20 line (before 5-fluorocytosine treatment) that predominantly expressed *clag3.2* as determined by RT-qPCR analysis, to sort 10 (x3 replicates) and 100 parasites (**Fig. 51**).

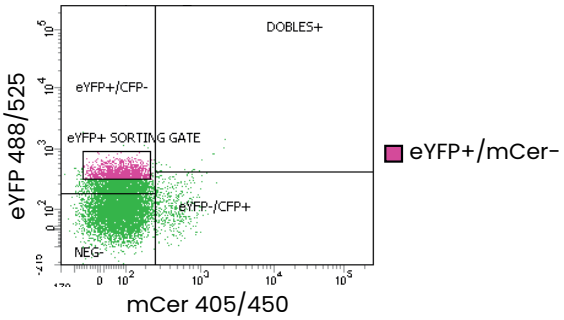


Fig. 51. Sorting of the T20 transgenic line for eYFP expressing parasites. On the X axis mCer fluorescence values and on the Y axis eYFP fluorescence values. Parasites that fall within the strict eYFP expression gate are highlighted in pink.

After that, we characterised the sorted T20 lines (10a, 10b, 10c and 100) by RT-qPCR and flow cytometry and found that they predominantly expressed CLAG3.2-eYFP and presented mostly eYFP fluorescence (**Fig. 52**).

Finally, we analysed the sorted T20 cultures for presence of residual episomal Cas9 plasmid by RT-qPCR and for presence of multiple integrations of donor plasmid in the *clag3.2* locus by PCR and unfortunately found both in all of these lines. Subsequently, we concluded that the T20-sorted lines were not adequate for continuing the development of our triple transgenic line. However, we did observe an inconsistency between the two RT-qPCR analyses performed for this double transgenic line, as in the second analysis (Fig. 52A) the original T20 line presented more similar transcript levels of both *clag3* genes than in the first analysis (Fig. 50). Furthermore, in the second flow cytometry experiment in which we analysed the fluorescence of this line (Fig. 52B) (in the first experiment – Fig. 49 - we only analysed the fluorescence of the T20-post line that had been treated with 5-fluorocytosine) we observed both mCer and eYFP signal possibly indicating that in the population there were parasites expressing *clag3.1* and parasites expressing *clag3.2*. Therefore, although all these results have to be considered with "caution" as they are

all based on single biological replicates, these last observations led us to believe that the original T20 line may present correct *clag3* expression profiles. As a result, we decided to use this line to continue constructing our triple transgenic line, with the idea of addressing the remaining issues of residual episomal Cas9 plasmid and multiple integrations of donor plasmid in the *clag3.2* locus in the final step of this process.

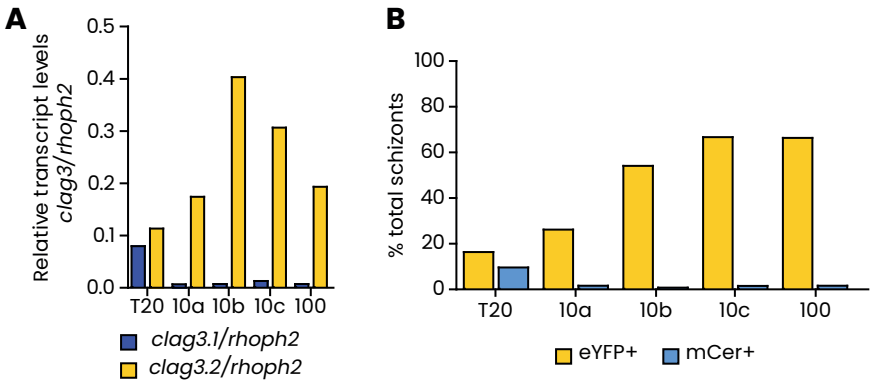


Fig. 52. Characterisation of sorted T20 parasites. (A) RT-qPCR analysis of *clag3* expression patterns of the T20-sorted parasites. Shown are the relative transcript levels of *clag3.1* (blue) and *clag3.2* (yellow) normalized against *rhoph2* in the initial T20 line and in the different sorted cultures (10a, 10b, 10c and 100). (Primers: P66, P50, P65, P86 and P87). (B) Flow cytometry analysis of sorted T20 parasites. Shown are the proportion of eYFP+ and mCer+ schizonts relative to total schizonts identified by MitoTracker staining.

3.5. Triple transgenic line

The final edition to be performed in order to obtain our triple transgenic parasite line was the introduction of the *gap45* promoter + tdTom construct into the *lisp1* locus. Using the original double transgenic T20 line (before 5-fluorocytosine treatment and sorting), we performed two transfections, one with each version of the *gap45* promoter (long and short) as we had not detected any significant differences between the two (detailed in section 3.3.3). For these transfections, we used 15 µg of donor plasmid (gap45_ short_TdTomato_LISP1_HR1_HR2 or gap45_long_TdTomato_ LISP1_HR1_HR2) linearised at the Sca I restriction site in the plasmid backbone and 60 µg of the Cas9 plasmid (pDC2-Cas9-BSD-lisp1). In the Cas9 plasmid we decided to switch from the selectable marker hDHFR to blasticidin deaminase as in

the original T20 line there was residual episomal Cas9 plasmid containing the hDHFR selectable marker (more details in section 4.1.2). The lines generated were T32 (*gap45*_long) and T33 (*gap45*_short) and parasites were detected by light microscopy 2 weeks after transfection.

First, we performed two diagnostic PCRs with which to check if T32 and T33 had been correctly edited (**Fig. 53**). The first PCR used external primers (flanking the edited region in the *lisp1* locus) and the second used an external primer and an internal primer (within the tdTom sequence). In the first PCR, we only obtained the amplicon corresponding to an unedited locus in the triple transgenic lines but in the second PCR we did get the amplicon corresponding to an edited locus. This indicated that our construct had been introduced but that there were most likely multiple integrations of the donor plasmid in the *lisp1* locus resulting in a sequence too long to be amplified in the first PCR. In this case, we did not consider multiple integrations of the donor plasmid a major issue because unlike with the *clag3.2* locus, integration of multiple copies of the plasmid would result in higher expression of the fluorescent stage reporter without altering the expression of the genes under study. Finally, we found that we did not have a pure parasite population in either of the two transgenic lines due to the presence of unedited parasites detected also in the first PCR.

Next, to further characterise these triple transgenic lines, we performed RT-qPCR analysis and flow cytometry experiments. In the RT-qPCR analysis, we found that only T32 showed tdTom expression and that its expression levels were lower than those of the single transgenic lines with the same control edition (**Fig. 54A**). The presence of unedited parasites in the triple transgenic lines could be responsible for these lower tdTom transcript levels. Also, we checked *clag3* expression patterns as they can be altered by exposure to the toxic compound BSD[326,327], as blasticidin deaminase was used as the selectable marker in these transfections (**Fig. 54B**). Both T32 and T33 presented similar *clag3* expression patterns to those of the parental 3D7-A line, so that it appeared that the post-transfection selection with BSD had not severely altered *clag3* expression patterns. Finally, in the flow cytometry analysis we found that T32 seemingly had higher tdTom fluorescence values than T33, although more biological replicates would be needed to confirm this (**Fig. 54C**).

Overall, we concluded that T32 was better than T33 as it showed higher tdTom fluorescence. However, this line still had the following issues: 1) presence of parasites without the *lisp1* edition and therefore lacking

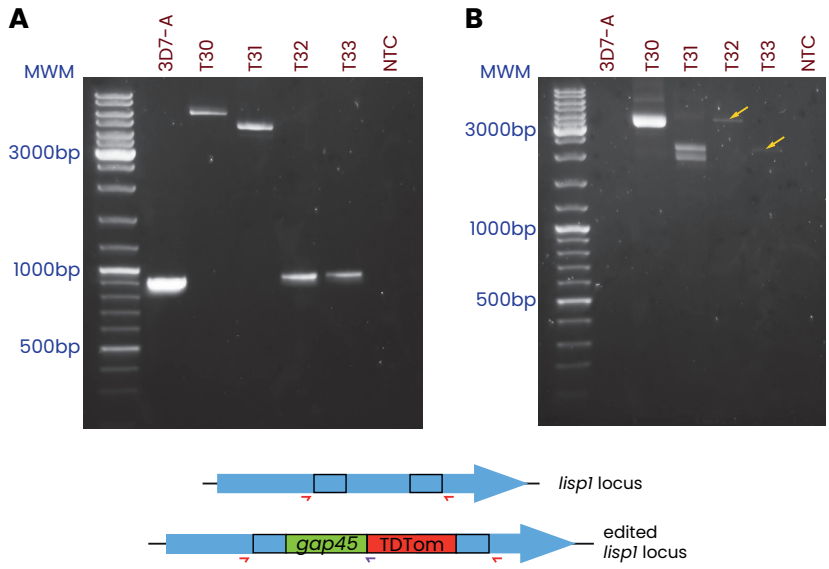


Fig. 53. Diagnostic PCR analysis to validate the integration of the *gap45* promoter controlling tdTomato expression in the *lisp1* locus in the T32 and T33 triple transgenic lines and the T30 and T31 single transgenic lines with the same respective editions in the *lisp1* locus. Above, agarose gels of: (A) External PCR in which 3D7-A, T32 and T33 all present a 918 bp amplicon corresponding to an unedited locus, while the single transgenic lines with the same respective editions as T32 and T33 present the expected amplicon of an edited locus (see Fig. 46). Amplified with primers P77 and P78. (B) External + internal PCR in which the expected amplicon is of 3484 bp for T30 and T32 or of 2234 bp for T31 and T33. Amplified with primers P77 and P90. No amplification is expected in 3D7-A (wild-type). Yellow arrows indicate the position of amplicons in T32 and T33 due to poor visibility. MWM = molecular weight marker. Below is a schematic indicating the position of the external (red) and internal (purple) primers within the wild-type and edited *lisp1* locus. Blue boxes indicate position of HR 1 (left) and HR2 (right).

the *gap45*-long promoter + tdTom construct; 2) presence of multiple integrations of the pHR1clag3.2-eYFP-HR2clag3.2 donor plasmid in the *clag3.2* locus originated in the previous T20 line (as we have mentioned, the multiple integrations of the gap45_longt_TdTomato_LISP1_HR1_HR2 donor plasmid in the *lisp1* locus were not of concern); and 3) possible presence of residual episomal Cas9 plasmid carrying the blasticidin deaminase selectable marker that could interfere with future experiments in which BSD could be used as a drug with which to select parasites with different *clag3* expression patterns. In order to tackle all these problems at once, we decided to first

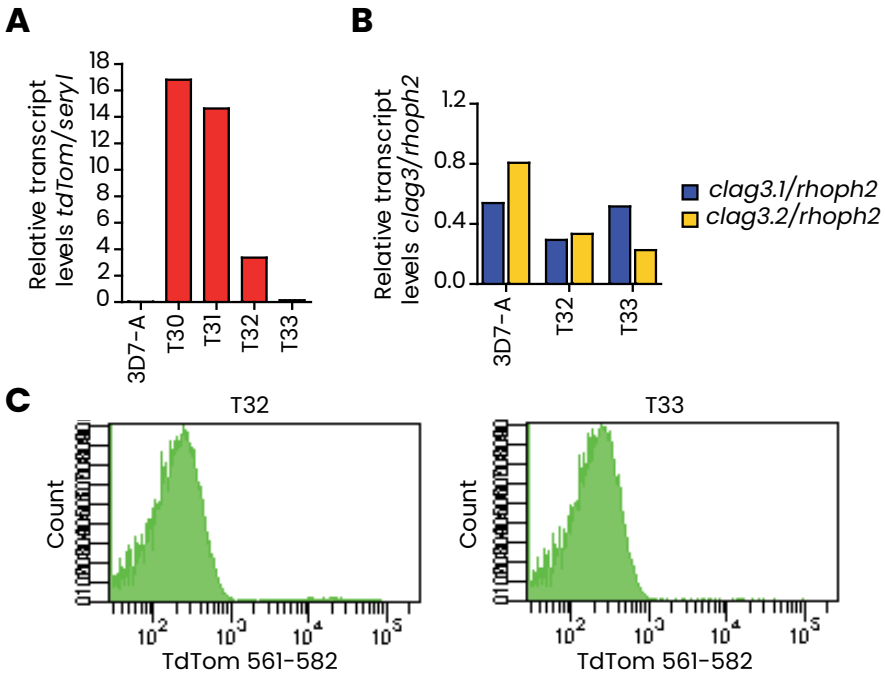


Fig. 54. RT-qPCR and flow cytometry analysis of the triple transgenic T32 and T33 lines. (A) RT-qPCR analysis of *tdtom* expression in the parental 3D7-A line, the single transgenic T30 and T31 lines and the triple transgenic T32 and T33 lines with the same respective editions. Shown are the relative transcript levels of *tdTom* normalized against *serine-tRNA ligase* (*serrs*). (Primers: P88, P89, P91 and P92). (B) RT-qPCR analysis of *clag3* expression in the parental 3D7-A line and the triple transgenic T32 and T33 lines. Shown are the relative transcript levels of *clag3.1* (blue) and *clag3.2* (yellow) normalized against *rhoph2* in the same parasite lines. (Primers: Primers: P66, P50, P65, P86 and P87). (C) TdTom fluorescence histograms of T32 and T33 with tdTom fluorescence values on the X axis and the cell count on the Y axis.

sort parasites by flow cytometry in order to obtain a pure triple transgenic parasite population that had all three editions and second, subclone this sorted population in order to obtain a parasite population originated from an individual parasite without multiple integrations of donor plasmid in the *clag3.2* locus, without residual episomal Cas9 plasmid with the blasticidin deaminase selectable marker and that predominantly expressed *clag3.2* (as currently there are no known compounds that select for parasites expressing this particular *clag3* gene, more details in section 5).

In a first attempt, we sorted 1000 T32 parasites with both tdTom and eYFP fluorescence (**Fig. 55A**) with the objective of eliminating parasites without the third edition in the *lisp1* locus as well as obtaining parasites predominantly expressing *clag3.2*. Next, we subcloned these sorted parasites and obtained 11 subclones which we analysed by PCR and qPCR. We found that all subclones had multiple integrations in the *clag3.2* locus and therefore could not be used as our final triple transgenic line. Given that, up until now, all attempts to obtain parasites without multiple integrations of donor plasmid in the *clag3.2* locus by sorting for eYFP-expressing parasites had been unsuccessful, we decided to change our strategy and we did a new sorting experiment of the T32 line in which we collected all parasites that showed tdTom fluorescence (**Fig. 55B**). In this manner, we hoped to eliminate parasites that did not have the third edition of the *lisp1* locus, without favouring the collection of parasites with multiple integrations in the *clag3.2* locus, even if this meant sacrificing obtaining a parasite population expressing predominantly *clag3.2*. In this second sorting, we collected 150,000 parasites with tdTom fluorescence but unfortunately diagnostic PCR revealed that in this population there were still parasites with multiple integrations in the *clag3.2* locus, residual episomal Cas9 plasmid with the blasticidin deaminase selectable marker and parasites without the third edition in the *lisp1* locus (**Fig. 56**).

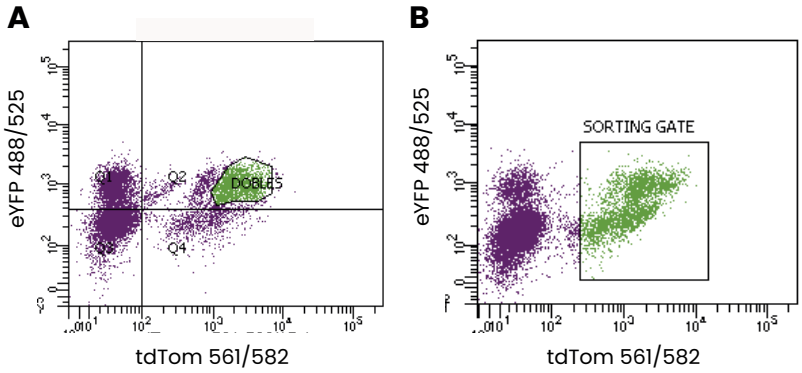


Fig. 55. Sorting of the T32 triple transgenic line for (A) eYFP and tdTom expressing parasites or (B) tdTom expressing parasites. Shown on the X axis are tdTom fluorescence values and on the Y-axis eYFP fluorescence values in both panels. Parasites that fall within the different sorting gates are highlighted in green.

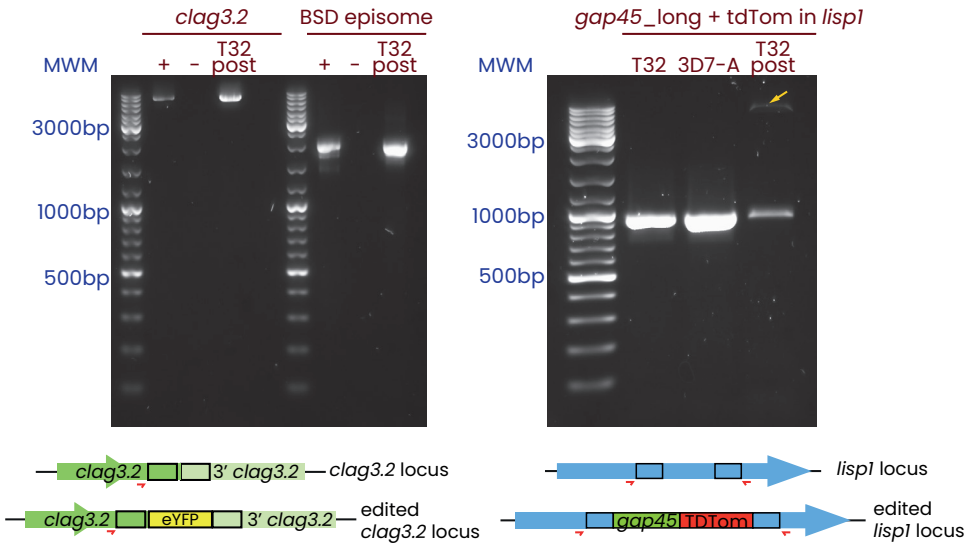


Fig. 56. Diagnostic PCR analysis of the T32 line after sorting (T32-post). Above, agarose gels showing: on the left, external + plasmid backbone PCR in which amplification is only expected if there are multiple integrations of donor plasmid in the *clag3.2* locus (amplified with primers P51 and P38); in the middle, residual episomal Cas9 plasmid (with blasticidin deaminase selectable marker) PCR in which amplification is only expected if this plasmid is present (amplified with primers P41 and P38); and on the right, third edition comprobation PCR in which an edited amplicon is of 5524 bp and an unedited amplicon of 918 bp. Amplified with primers P77 and P78). In this last PCR, the yellow arrow indicates the position of an amplicon possibly corresponding to multiple integrations of donor plasmid in the *lisp1* locus, however an amplicon is not always expected in this case. + = positive control (on the left transgenic line with multiple integrations of donor plasmid in *clag3.2* locus, in the middle maxiprep of Cas9 plasmid). - = negative control, MWM = molecular weight marker. Below is a schematic showing the position of primers (red) that are present in the wild-type and edited *clag3.2* and *lisp1* loci. Not shown are the primers that are located in unintegrated parts of the donor or Cas9 plasmid. Boxes indicate the position of HRs in both loci.

Next, we subcloned this sorted line with the hope of obtaining a parasite population without any of the aforementioned issues. In a first attempt we only obtained 7 subclones, which is why we performed a second attempt (this time with two 96-well plates) that produced 24 subclones. All these subclones were analysed by PCR and qPCR and once again we found that they all had multiple integrations of donor plasmid in the *clag3.2* locus with the exception of 3 subclones in which it appeared that both *clag3* genes had recombined generating a *clag3* hybrid that has been previously described[380].

On the other hand, the residual episomal Cas9 plasmid with the blasticidin deaminase selectable marker had been lost in 9 subclones. Therefore, it would appear that this experimental strategy combining sorting and subcloning was adequate for eliminating residual episomal Cas9 plasmid but not for obtaining parasites without multiple integrations of donor plasmid in the *clag3.2* locus. However, this could be because the initial T20 line either did not have or only had a very low proportion parasites with a correct edition of the *clag3.2* locus. Given these negative results, we did not check which subclones had the third edition in the *lisp1* locus. Finally, we analysed some of these subclones by flow cytometry to determine their fluorescence profiles and evaluate if they corresponded with plausible *clag3* expression patterns (**Fig. 57**). We selected three subclones that presented multiple integrations of donor plasmid in the *clag3.2* locus and one subclone that by PCR appeared to have a hybrid *clag3* gene. Unexpectedly, we found that the three subclones with multiple integrations did not show CLAG3.1+mCer fluorescence, suggesting that the multiple integrations of donor plasmid in the *clag3.2* locus affect the normal expression of both *clag3* genes. In contrast, the *clag3*-hybrid subclone did present both eYFP and mCer fluorescence but as we were unable to amplify either *clag3* gene individually by long PCR[380], it was unclear how the expression of these fluorescence tags was being controlled. The most plausible explanation is that complex recombination events occurred in this parasite line. Furthermore, in all the analysed subclones we detected a high percentage of parasites that only presented control tdTom fluorescence possibly indicating that both their *clag3* genes were silenced. This tied in with a prior observation of these subclones showing partial resistance to sorbitol lysis, an event previously linked to simultaneous silencing of the two *clag3* genes[326]. It is possible that the presence of multiple integrations in the *clag3.2* locus was also responsible for the silencing of both *clag3* genes in these parasites although further experimental evidence is needed to demonstrate this.

In conclusion, our T32 triple transgenic line and its respective subclones were not adequate to be used as our final tool with which to analyse *clag3* expression patterns, given that the regulating mechanisms controlling *clag3* expression appeared to have been compromised. However, by producing the T32 line we have demonstrated that it is possible to generate a triple transgenic line, although a revaluation of our CRISPR-Cas9 design and transfection strategy is needed.

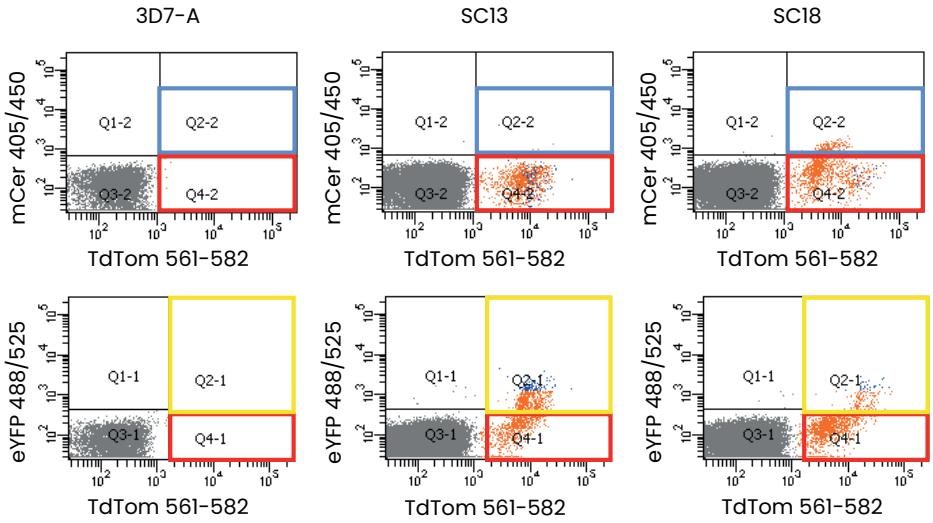


Fig. 57. Flow cytometry analysis of T32 subclones. From left to right: 3D7-A wild-type line, subclone 13 with multiple integrations in the *clag3.2* locus and subclone 18 with a *clag3* hybrid. On the X axis, in both rows, tdTom fluorescence values are shown. On the Y axis of the top row, mCer fluorescence values are shown, and on the bottom row eYFP fluorescence values are shown. Blue boxes highlight tdTom- and mCer-expressing parasites, yellow boxes tdTom- and eYFP-expressing parasites and red boxes only tdTom-expressing parasites.

As indicated at the start of this chapter, on the next page is a schematic of all the transgenic parasite lines described in this chapter (**Fig. 58**).

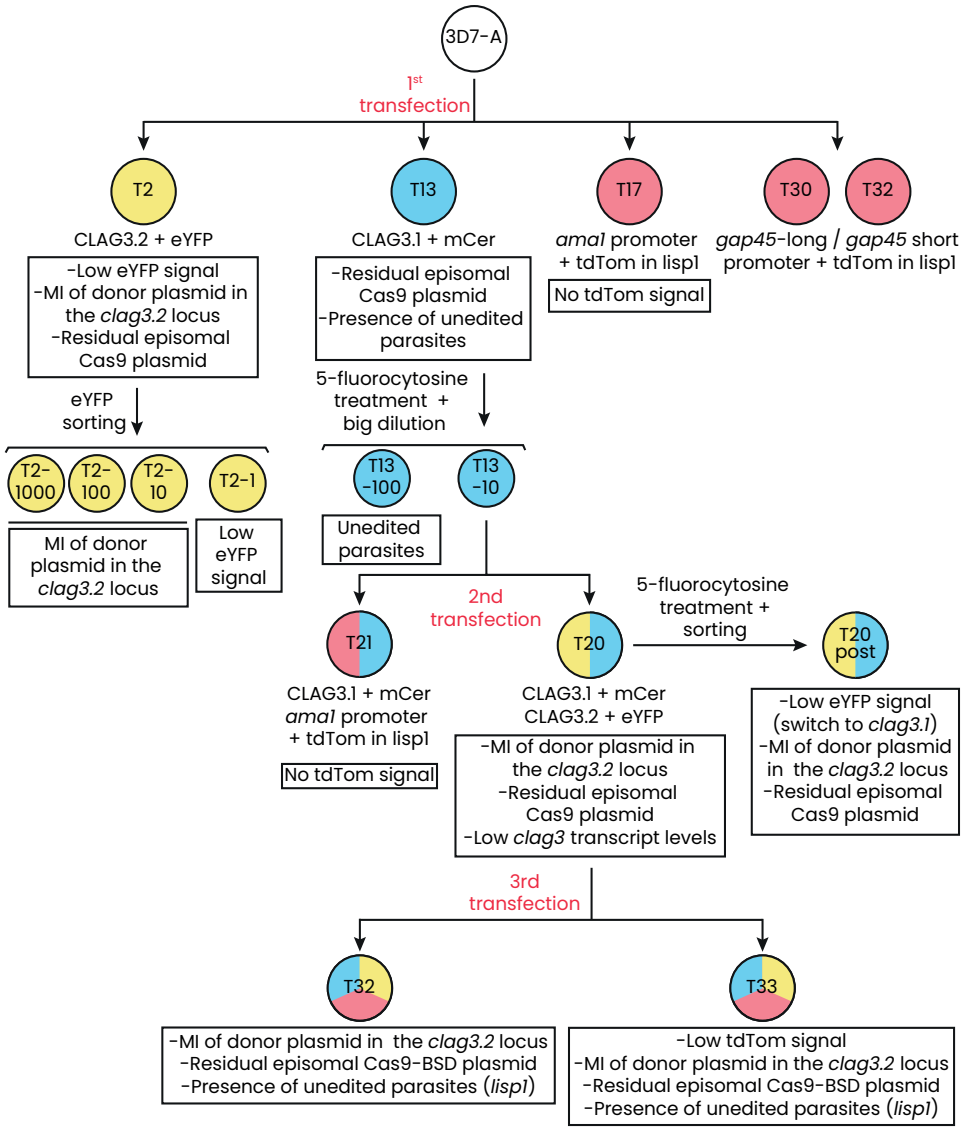


Fig. 58. Schematic of all described transgenic lines produced during the generation of our *clag3* triple transgenic parasite line. Below each transgenic parasite line (portrayed by colour-coded circles in which yellow represents eYFP, blue mCer and red tdTom) there is information on the editions each one presents and, in the boxes, the principal problems encountered in each of these lines. Alongside the arrows there is information on the different treatments or experimental procedures performed to generate each line. MI = multiple integrations

4. Challenges encountered during the generation of the triple transgenic line and strategies used to overcome them

4.1. Residual Cas9 episome

As previously described, the use of the hDHFR-yFCU dual positive-negative selectable marker enables marker-free integrations that allow for sequential editions of the same parasite line[435] (**Fig. 59**).

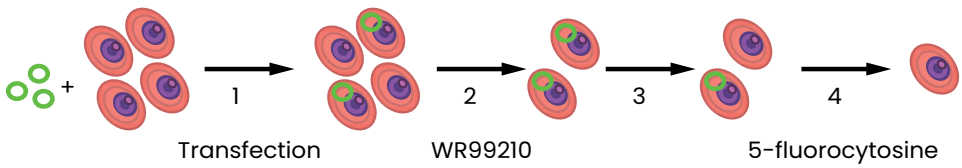


Fig. 59. Model of dual positive-negative selection of "marker-free" transgenic parasite lines. First, parasites are transfected with donor and Cas9 plasmids (green) (1). Positive selection commences 12-24 h later and consists in a 4-day treatment with the antifolate drug WR99210 (2). This treatment selects those parasites that have successfully uptaken the Cas9 plasmid with the hDHFR selectable marker (the majority of these parasites also typically uptake the donor plasmid). On average, a couple of weeks pass until parasites are detected by light microscopy and PCR analysis is performed to check if they have been correctly edited. During this time, some parasites spontaneously lose episomal (non-integrated) plasmids (3). Finally, if parasites have been correctly edited, negative selection entails and consists in a 10-day treatment with 5-fluorocytosine (4), selecting those parasites that have spontaneously lost the Cas9 plasmid and therefore are no longer susceptible to this treatment having lost the yFCU selectable marker.

In all the analysed transgenic lines of this project, we have found that post-transfection there is always residual episomal Cas9 plasmid, therefore, we routinely submit all parasite lines destined to be re-transfected to the 10-day 5-fluorocytosine treatment. In general, this dual positive-negative selection works fairly well in our lab and we are able to obtain transgenic parasite lines with minimal amounts of residual episomal Cas9 plasmid. Subsequently, this episomal plasmid is only a problem for sequential transfections when the number of parasites that retain this plasmid is superior to the number of parasites that uptake new Cas9 plasmids in the posterior transfections (typically 1 in 10^7 transfected

parasites), as this means that they can expand more rapidly during the selection with WR99210 and predominate in the selected population. We ignore why the 5-fluorocytosine treatment is not 100% efficient and suspect that the variability in its outcome may be due to events such as the Cas9 plasmid becoming spontaneously integrated in the genome of some parasites or that, after the positive selection with WR99210, the hDHFR-yFCU cassette may be silenced by unknown mechanisms and impede the negative selection with 5-fluorocytosine. Regardless of why this treatment does not always work, the result is that the parasites that retain this episomal Cas9 plasmid are resistant to WR99210 what makes it impossible to re-transfect them with the same selectable marker without the risk of selecting unedited parasites.

All in all, the variability in the efficacy of 5-fluorocytosine treatment in combination with the limited number of available selectable markers in *P. falciparum*[434], have constituted an important challenge for the generation of our triple transgenic line that entails three sequential transfections of the same parasite line. As a result, we have been forced to find alternative solutions besides this dual positive-negative selection in order to obtain re-transfectable parasite lines. Before detailing these alternative solutions, it is important to describe the method used to detect residual episomal Cas9 plasmid in our transgenic parasite lines. This method consists in a qPCR analysis performed on parasite gDNA with primers that amplify a region of the hDHFR-yFCU cassette (P75 and P76). Using the standard curve method[447], it is possible to quantify the average number of residual episome copies per parasite within the population using our lab-derived line W4-2 as a reference, given that it has one copy of *hdhfr* integrated in its genome as determined by southern blot[315]. Finally, to assess whether the quantification of residual episome copies, as determined by qPCR, corresponds with a resistance phenotype to WR99210, on occasions we subject these analysed lines to the standard 4-day WR99210 treatment. In general, when residual episome is not detected by qPCR, these lines do not survive the WR99210 treatment.

4.1.1. Obtaining subpopulations of episome-free parasites

Our main approach to obtain re-transfectable parasite lines consists in combining the use of the positive-negative selectable marker with big dilutions or sorting by flow cytometry in order to procure subpopulations of parasites that have spontaneously lost residual episomal Cas9 plasmid.

Although this objective can also be achieved by subcloning, we prefer these other methods as they are less time-consuming and do not require reducing parasite populations to one individual parasite which significantly compromises the transcriptional diversity of the population. Normally, we perform large dilutions or sort parasites to obtain subpopulations of 10, 100, 1,000 or 10,000 parasites. The rationale behind these numbers is that based on qPCR results, if we estimate, for instance, that 1 in 10^6 parasites carry the residual episomal Cas9 plasmid after 5-fluorocytosine treatment, then a subset of 10,000 will have a 1% chance of including a parasite with this episomal plasmid. Therefore, by reducing the population to 10,000 parasites there is a high probability of obtaining only episome-free parasites.

After 5-fluorocytosine treatment we performed big dilutions on the CLAG3.2-eYFP single transgenic line T2, the CLAG3.1-mCer single transgenic line T13, the CLAG3.2-eYFP + CLAG3.1-mCer the double transgenic line T20 and on the triple transgenic line T32. This was followed by qPCR analysis to determine the amount of residual episomal Cas9 post-big dilution (**Fig. 60A**). Firstly, neither 5-fluorocytosine treatment nor dilution to 10,000 parasites served to eliminate this episomal plasmid from the T2 line, which explains why all posterior transfections performed on this line were unsuccessful. Next, in the T13 line the 5-fluorocytosine treatment was successful in eliminating episomal Cas9 plasmid and the additional dilution to 10 parasites served to eliminate unedited parasites from this line. Subsequently, T13-10 is the only parasite line we have successfully re-transfected with the same selectable marker. In the case of T20, 5-fluorocytosine treatment was insufficient to completely eliminate episomal Cas9 plasmid but the posterior dilution to 100 parasites was and the resulting T20-100 line did not survive treatment with WR99210. However, other issues prevented it from being used for further re-transfections. Finally, as with T2, neither 5-fluorocytosine treatment nor dilution to 10, 100 or 1,000 parasites served to eliminate residual episomal Cas9 plasmid from the T32 line. Consequently, all re-transfection attempts performed on this line (related to the elimination of multiple integrations of donor plasmid in the *clag3.2* locus described in section 4.2.2) using the same selectable markers also failed. Similar results were observed with sorting because out of all the sorted T2 (to 1, 10, 100 and 1000 parasites) and T20 (to 3x10 and 100 parasites) lines, only the T2-1 line was found to be free of residual episomal Cas9 plasmid (**Fig. 60B**).

In summary, we found that the combination of 5-fluorocytosine with big dilutions or sorting can be sufficient to eliminate parasites with

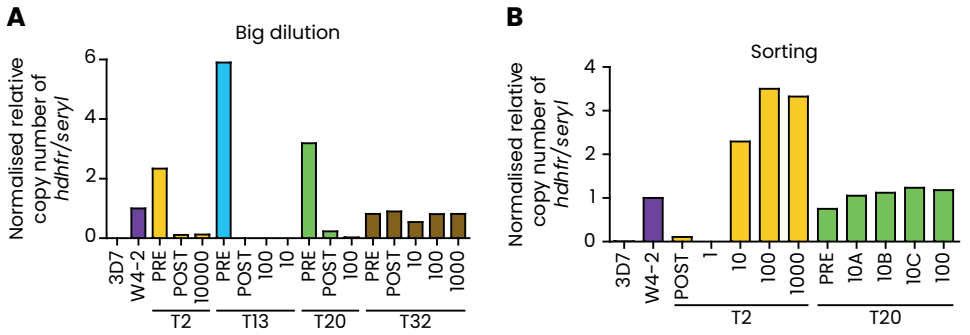


Fig. 60. qPCR analysis of the average number of residual episomal Cas9 plasmid copies in different parasite lines after treatment with 5-fluorocytosine combined with big dilutions (A) or sorting (B). Shown are relative copy numbers of *hdhfr* normalized against *serine-tRNA ligase* (*serrs*) and normalised with the value of the control W4-2 line. The analysed parasite lines include: the parental 3D7-A line, the control W4-2 line with one copy of *hdhfr* integrated in its genome and the different transgenic lines (T2, T13, T20, T32) before (PRE) or after (POST) 5-fluorocytosine treatment and after big dilution to 10, 100 or 10,000 parasites or sorting to 1, 10, 100 or 1,000 parasites. (Primers: P75, P76, P88 and P89).

residual episomal Cas9 plasmid. However, the variability in the outcome of this strategy and the time it entails (various weeks of parasite growth in addition to the time needed to prepare samples to perform the qPCR analysis) mean that it is not an ideal solution for this problem.

4.1.2. Switching selectable markers

For the final transfection to generate a triple transgenic line we replaced the positive selectable marker hDHFR for other markers with different modes of action. This was complicated by the fact that there are only a few selectable markers available for *Plasmodium* research[434] and that one of the other main markers that is used is the blasticidin deaminase gene that renders parasites resistant to blasticidin S (BSD), a compound known to affect *clag3* expression patterns[326,327]. In the end we tested both blasticidin deaminase and another marker, puromycin-N-acetyltransferase (PAC), which renders parasites resistant to puromycin. We and other members of our lab generated copies of the Cas9 plasmid, substituting the hDHFR-yFCU cassette with that of either of these two selectable markers. It is important to note that the blasticidin deaminase marker was not fused to yFCU and

therefore the possibility for a dual positive-negative selection to eliminate episomal Cas9 plasmid was lost. However, as all these different drugs have distinct modes of action, it would be possible to re-transfect and select parasites using a different drug each time without the risk of selecting resistant parasites from the previous transfection.

We performed 7 transfections with the PAC marker but none were successful as in 4 transfections we never observed parasite growth and in 3 only unedited parasites grew. The majority of these transfections were performed in parallel with others involving the same edition but with the hDHFR selectable marker. All of the simultaneous hDHFR transfections worked, therefore, we concluded that the problem was the puromycin selection and we stopped using the PAC selectable marker.

On the other hand, our 2 transfections using the blasticidin deaminase as a selectable marker were both successful and enabled us to obtain our triple transgenic line. However, as mentioned, exposure to BSD has been show to select for different *clag3* expression patterns[326] and the concentrations of this compound needed to do this are much lower than those used in the 6-day post-transfection treatment (0.2-0.6 µg/mL vs 2.5 µg/mL). Although the selectable resistance marker should render parasites resistant to this compound without it being necessary for different *clag3* expression patterns to be selected, in a previous study it was observed that this can happen[326]. Subsequently, not only were we concerned that the selection with BSD had altered the *clag3* expression patterns of our triple transgenic line, but also that the fact that this line had retained residual Cas9 episome with the blasticidin deaminase selectable marker meant that future selection experiments with BSD could be compromised (and removing parasites with the episome would then be required). Consequently, blasticidin deaminase is not an optimal selectable marker to be used in this project that revolves around studying the expression dynamics of *clag3* genes.

All in all, our results suggest that the best strategy is to try and obtain episome-free parasites after each transfection by combining 5-fluorocytosine treatment with big dilutions and therefore recycle the hDHFR selectable marker, as it is the most effective marker in our lab and WR99210 does not affect *clag3* expression patterns. Nevertheless, this method of obtaining episome-free parasites is time consuming and not always successful, therefore there is room for innovation and optimization in order to solve this problem.

4.2. Multiple integrations of donor plasmid in the *clag3.2* locus

The other major issue of this project consists in multiple copies of the pHR1clag3.2_eYFP_HR2clag3.2 donor plasmid becoming integrated into the *clag3.2* locus during the reparation of the DSB generated by the Cas9 nuclease. While we have not found this to occur when editing the *clag3.1* locus, it has happened in the *lisp1* locus. However, as we have described, we do not consider this to be a problem because *lisp1* is not expressed in asexual blood stages[441] and integrations of multiple copies of the plasmid containing the tdTom expression cassette may simply increase the expression levels of this marker. On the contrary, this is not the case with the *clag3.2* locus, where the presence of multiple integrations of donor plasmid could significantly alter the expression of *clag3* genes. In the *P. falciparum* genome, the *clag3.2* locus is located upstream to the *clag3.1* locus and between them lies a *var* pseudogene (PF3D7_0302300) and a ncRNA (PF3D7_0302400). Like other CVGs, the expression of *clag3* genes is regulated at the epigenetic level with the distribution of facultative heterochromatin determining their transcriptional state[297,300,315]. In this manner, the presence of large foreign elements within the *clag3* loci could affect the distribution of heterochromatin and alter the natural activation or silencing of these genes.

All of the 9 CLAG3.2 + eYFP taggings we have performed have all resulted in the integration of multiple copies of donor plasmid into the locus of this gene. Furthermore, in these transgenic lines, eYFP fluorescence did not always correlate with *clag3.2* transcription levels. For instance, in the CLAG3.2-eYFP single transgenic line T2 we observed that despite apparently normal *clag3.2* transcript levels, eYFP fluorescence could not be detected. However, on the occasions CLAG3.2-eYFP expression did correlate with *clag3.2* transcription levels, we have observed other strange events. For example, in the double transgenic line T20 there was a large unexplained switch of parasites expressing *clag3.2* to parasites expressing *clag3.1* after 5-fluorocytosine treatment. Also, some of the subclones of the triple transgenic line T32 appeared to have undergone strange recombination events leading to a hybrid *clag3* gene expressing both tags (eYFP and mCer). Furthermore, all of the analysed T32 subclones had a large proportion of parasites in which both *clag3* genes were apparently silenced as they only presented control gene expression indicated by tdTom fluorescence. Therefore, although we do not have any direct evidence linking these events to the presence of multiple copies of donor plasmid integrated in the *clag3.2* locus, this is the only common factor that joins

them. As a result, our current main priority is to obtain transgenic parasite lines without this alteration. Therefore, we will proceed to describe how we detect this problem and the strategies we have tried to solve it.

4.2.1. Detection of multiple copies of donor plasmid integrated in an edited locus

There are two types of analysis we perform to detect multiple integrations using extracted parasite gDNA: PCR and qPCR.

Generally, we perform two types of PCR analysis (**Fig. 61**). The first combines the use of external primers that flank the edited region of the locus with a PCR elongation time of 10 minutes. This enables the amplification of the longer amplicons that result from these multiple integrations (e.g., 3 copies of a donor plasmid of 4,500 bp produces an amplicon of 13,5000 bp), albeit sometimes they are too long to be amplified and no amplicon is obtained. The second PCR we perform utilises one of the aforementioned external primers in addition to a plasmid backbone primer that is not present in the tag nor HRs. This PCR only amplifies if plasmid backbone has been integrated into the locus. All in all, the results of these two PCRs give us an indication of whether there are multiple integrations present in the analysed locus. However, as in each case we ignore exactly how many copies of donor plasmid are integrated, in which orientation and what parts of the plasmid are involved, the results of these PCRs can sometimes be hard to interpret without information from other more complex and expensive techniques, such as whole genome sequencing or southern blot, which we do not routinely perform with our transgenic lines.

With regards to the qPCR analysis (**Fig. 62**), we use primers that amplify a fragment of HR2 or the tag and compare the number of copies we obtain with those of a line known to only have one copy of said fragment. The problem with this method is that it does not allow us to distinguish between multiple integrated copies and presence of residual episomal donor plasmid.

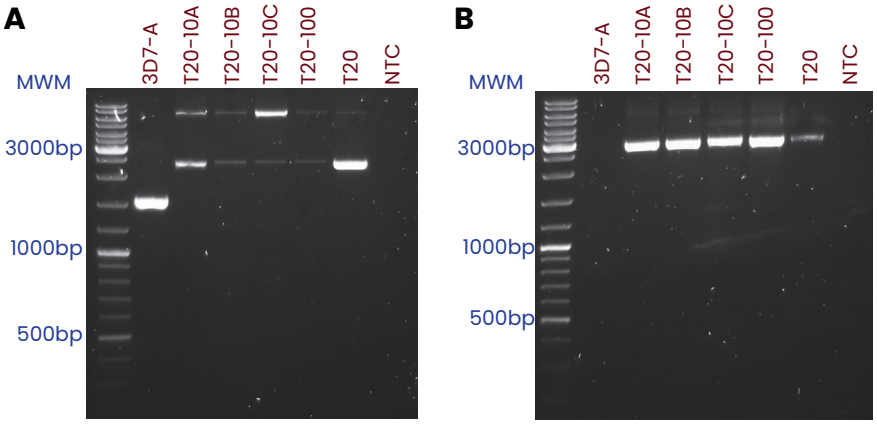


Fig. 61. Examples of a PCR analysis to detect the presence of multiple copies of donor plasmid in transgenic lines. (A) External PCR of wild-type 3D7-A, double transgenic T20 and its sorted lines (10a, 10b, 10c, 100). The wild type amplicon is of 1587 bp and that of correctly edited parasites is 2331 bp (amplified with primers P66 and P63). All larger amplicons correspond to multiple integrations. (B) External + plasmid backbone PCR of the same parasite lines. In this case, the presence of amplicon indicates that there are multiple integrations in the *clag3.2* locus (amplified with primers P66 and P38). MWM = molecular weight marker. NTC = negative control.

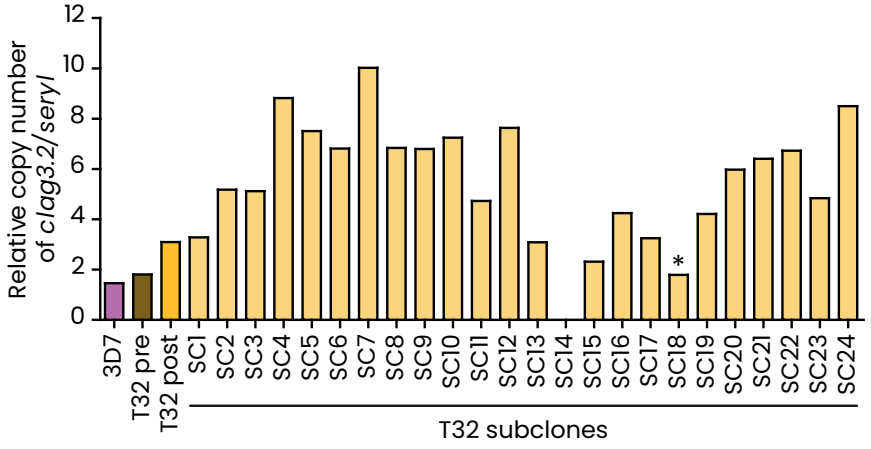


Fig. 62. Examples of a qPCR analysis to detect the presence of multiple copies of donor plasmid in transgenic lines. Shown are relative copy numbers of a fragment located in the HR2 of our *clag3.2* tagging design normalized against *serine-tRNA ligase* (*serrs*) in wild-type 3D7-A, the T32 triple transgenic line before (pre) and after (post) sorting and in all the subclones of the post-sorting culture. (Primers: P55, P56, P88 and 89). *Although SC18 presents similar relative copy numbers to the wild-type 3D7-A line, this subclone has a hybrid *clag3* (described in section 3.5) and therefore is not correctly edited.

4.2.2. Strategies for obtaining parasites without multiple integrations of donor plasmid in the *clag3.2* locus

As we have no control over how these copies of donor plasmid are integrated into the *clag3.2* locus, all the following strategies are based on isolating subpopulations of parasites that do not have this issue as we hypothesise that due to the stochasticity of this event, they may be present within the general parasite population. In this manner, our first approach for obtaining multiple integration-free subpopulations has been to perform sorting, subcloning or both. However, none of these strategies have proved to be successful. On the one hand, sorting for eYFP expression with the premise that parasites expressing this tag will have the *clag3.2* locus correctly edited, only served to select for parasites with this alteration (example in Fig. 61). This may be because these parasites express various copies of the eYFP tag and consequently give off a stronger fluorescence signal so that they are sorted more efficiently by flow cytometry. On the other hand, subcloning has not worked either and this may be because the proportion of parasites with multiple integrations is much higher than those without (if they do indeed exist) and therefore the probability of obtaining a multiple integration-free subclone is very low (example in Fig. 62).

Another strategy we have tried consists in using sgRNAs that target donor plasmid backbone with the objective of directing the Cas9 nuclease to produce a DSB in those parasites that have integrated plasmid backbone into their genome and cause their death due to the lack of a repair template. As described in section 2.4.1, we generated two new Cas9 plasmids that included either the sgRNA-episome3 or sgRNA-episome4 cassette. While the target of sgRNA-episome3 was located at the very end of HR2 of the CLAG3.2-eYFP donor plasmid, the target of the sgRNA-episome4 was cloned into this donor plasmid generating the pHR1clag3.2-eYFP-HR2clag3.2_PvuII_ sgRNA4 plasmid. First, we tried transfecting the triple transgenic T32 line with the plasmids of the sgRNA-episome3 strategy (the original donor and Cas9 plasmids of the CLAG3.2+eYFP design as well as the pDC2-cam-Cas9-U6-hDHFRyFCU_sgRNA_episome3 Cas9 plasmid) in order to eliminate all the parasites from this line with multiple integrations of the donor plasmid in the *clag3.2* locus. However, due to the presence of residual episomal Cas9 plasmid in the T32 line that we were unable to eliminate via 5-fluorocytosine treatment and big dilutions, this transfection was unsuccessful. Next, we tried transfecting the CLAG3.1-mCer single transgenic line T13 with the sgRNA-episome4 strategy plasmids (the original CLAG3.2+eYFP Cas9 plasmid in addition to the newly generated pHR1clag3.2-eYFP-HR2clag3.2_PvuII_

sgRNA4 and pDC2-cam-Cas9-U6-hDHFRyFCU_sgRNA_episome4 plasmids). Unfortunately, this approach did not work and although we obtained edited parasites, they still had multiple integrations in the *clag3.2* locus.

The final approach we have taken to obtain parasites without multiple integrations in the *clag3.2* locus consists in incorporating the yFCU selectable marker cassette into the *clag3.2* donor plasmid (pHR1clag3.2-eYFP-HR2clag3.2_PvuII_yFCU) (see section 2.4.2). In this manner, 5-fluorocytosine treatment could serve not only to select parasites without residual episomal Cas9 plasmid but also to select those without multiple integrations in the *clag3.2* locus. Subsequently, we transfected the CLAG3.1-mCer T13 line with this new CLAG3.2-eYFP-yFCU donor plasmid and with the original CLAG3.2-eYFP Cas9 plasmid, generating two replicate lines named T42 and T43. At the same time, we tried combining this strategy with the previous one and generated the T44 line by transfecting T13 parasites with the aforementioned plasmids in addition to the sgRNA-episome3 Cas9 plasmid. Once these lines were detectable by light microscopy, we subjected them to a 10-day 5-fluorocytosine treatment and gDNA was extracted before and after this treatment for PCR analysis. To check for the presence of multiple integrations in the *clag3.2* locus and determine whether the yFCU negative selectable marker had worked, we performed the routine external and external + plasmid backbone PCRs with the gDNA from these three lines (**Fig. 63**). With the external PCR we found that none of these lines presented parasites with the correctly edited locus and that this did not change after the 5-fluorocytosine treatment (**Fig. 63a**). Therefore, this hinted at multiple integrations being present in these lines. However, in the external + plasmid backbone PCR, there was only amplification in the T44 line and it seemed that the treatment with 5-fluorocytosine did have a partial effect in reducing the total number of parasites with multiple donor plasmid integrations (**Fig. 63b**). Therefore, although we did not obtain parasites without multiple copies of donor plasmid in the *clag3.2* locus with any of these strategies, the effect of the 5-fluorocytosine treatment on the T44 line suggests that this strategy could work in the future. As a result, we wish to see that if by initiating the 5-fluorocytosine treatment directly after the post-transfection positive selection with WR99210, we can increase the chances of obtaining multiple integration-free parasites. However, given that all the *clag3.2*+eYFP transfections we have performed to date have resulted in multiple integrations of donor plasmid in the *clag3.2* locus, it would appear that our current design favours this problem. Consequently, perhaps the most effective solution would be to change the design of this construct and use a different sgRNA.

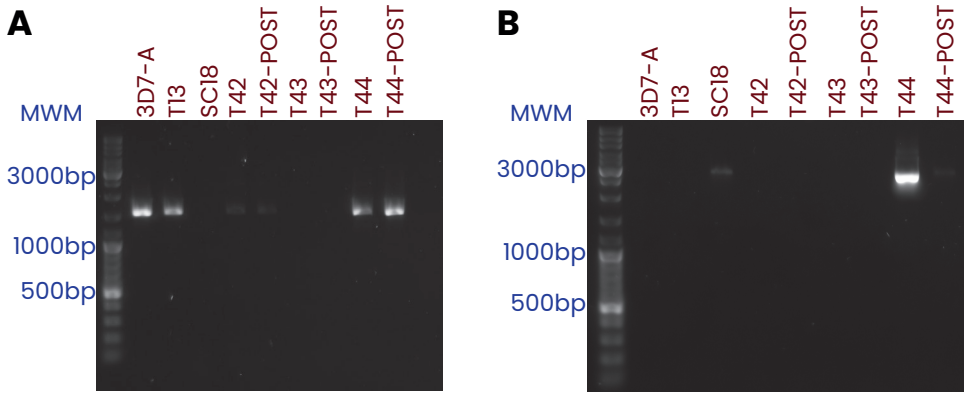


Fig. 63. Diagnostic PCR analyses of the presence of multiple integrations at the *clag3.2* locus before and after 5-fluorocytosine treatment in the T42, T43 and T44 transgenic lines. (A). External PCR in which the control unedited parasite lines (3D7-A and T13) show the expected amplicon of 1587 bp and none of the transgenic lines present the expected amplicon of 2331 bp (amplified with primers: P66 and P63). (B) External + plasmid backbone PCR in which there is only amplification if there are multiple integrations (amplified with primers P66 and P38). The lines analysed in both PCRs include the negative unedited controls 3D7-A and T13, the positive control with multiple integrations SC18 and the T42, T43 and T44 transgenic lines before and after (POST) 5-fluorocytosine treatment. MWM = molecular weight marker.

5. Other experiments

In addition to all the experiments performed to obtain a correctly edited triple transgenic line, we performed two additional exploratory assays that involved studying the effect of different compounds on *clag3* expression patterns. In the lab we can select parasites that express *clag3.1* by treating them with BSD[326] or we can select parasites that have both *clag3* genes silenced by treating them with T3, T16 or higher concentrations of BSD[325,326]. However, we have yet to identify compounds that select for parasites expressing *clag3.2*. As later on in this project we wish to use our triple transgenic line to study *clag3* expression dynamics in parasites grown under normal conditions and under different selective pressures, it would be useful to have compounds with which to select parasites that express each *clag3* gene. As a result, we performed two selection experiments with different compounds to see if they had an effect on *clag3* expression patterns.

5.1. Selection with 5-fluorocytosine

Based on the observation that post-5-fluorocytosine treatment and big dilution to 100 parasites our T20 double transgenic parasite line switched from predominantly expressing *clag3.1* to expressing mainly *clag3.2* (Fig. 50), we decided to determine if 5-fluorocytosine treatment had an effect on *clag3* expression patterns. In order to do so, we took T20 and wild-type 3D7-A parasites and selected them over a period of 2 weeks with 2 different concentrations of 5-fluorocytosine: 1 µM and 10 µM. RNA samples were collected on days 0, 7 and 14 and RT-qPCR analysis was performed to study *clag3* expression levels (**Fig. 64**). In a single biological replicate, we found that in general, 5-fluorocytosine treatment does not appear to affect *clag3* expression patterns regardless of the concentration of drug used and the duration of the treatment. However, additional biological replicates would be needed to confirm these results.

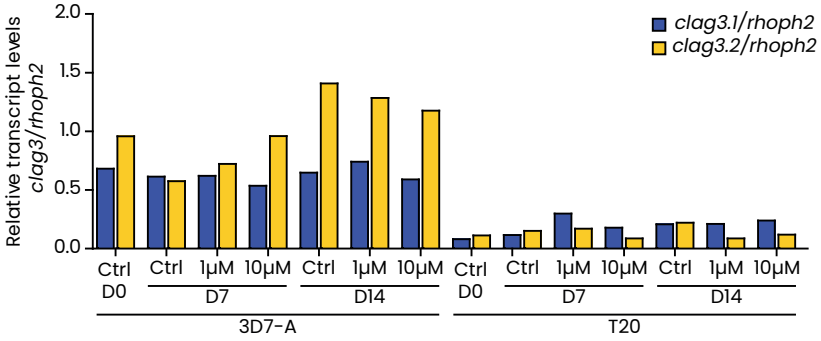


Fig. 64. RT-qPCR analysis of *clag3* expression in the 3D7-A and double transgenic T20 parasite lines after selection with 1 µM or 10 µM of 5-fluorocytosine during 7 or 14 days. Shown are the relative transcript levels of *clag3.1* (blue) and *clag3.2* (yellow) normalized against *rhopH2*. (Primers: P66, P50, P65, P86 and P87)

5.2. Selection with human serum

It has been found that in vivo *P. falciparum* parasites express predominantly *clag3.2*[335]. Therefore, we decided to investigate if growth in culture medium supplemented with human serum instead of AlbuMAX II as the external source of lipids had an effect on *clag3* expression patterns. In order to do so, we split a 3D7-A culture into two dishes and grew them in parallel in culture medium supplemented with either 0.5% AlbuMAX II or 10%

human serum for a total of 3 to 4 weeks.

Two biological replicates were performed and RNA samples taken at the beginning and end of each selection, followed by RT-qPCR analysis (**Fig. 65**). We found that the parasites grown in human serum presented overall lower levels of *clag3* expression with a similar balance between *clag3.1* and *clag3.2* expressing parasites compared to the AlbuMAX II culture that continued to express predominantly *clag3.2*. Consequently, we were able to conclude that human serum supplemented medium would not be an adequate condition with which to select *clag3.2* expressing parasites in vitro.

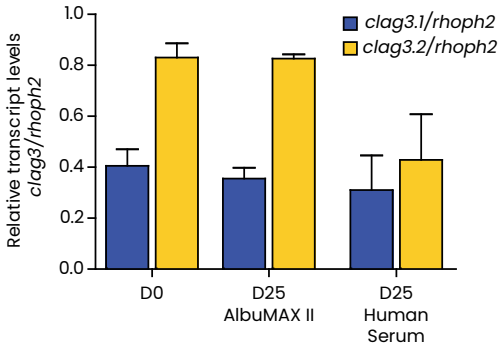


Fig. 65. RT-qPCR analysis of *clag3* expression in the 3D7-A and double transgenic T20 parasite lines after culture in AlbuMAX II or human serum supplemented medium for 25 days. Shown are the relative transcript levels of *clag3.1* (blue) and *clag3.2* (yellow) normalized against *rhoph2*. Data are presented as the average and s.e.m of two independent biological replicates. (Primers: P66, P50, P65, P86 and P87)

6. Future steps

In order to complete this project, the first step will be to generate a correctly edited triple transgenic line. We have found that our *clag3.1*+mCer and *gap45*_promoter (long) + tdTom in *lisp1* constructs work and we have obtained parasites with both of these editions correctly incorporated. Furthermore, we have found different approaches that enable us to perform sequential transfections and overcome the problem of residual Cas9 episome. However, the main challenge that remains to be addressed is that of incorporating the eYFP tag into the *clag3.2* locus without the integration of multiple copies of the donor plasmid. As mentioned above, at present

one of the two clearest ways of achieving this is repeating the transfection in which the 5-fluorocytosine selectable marker cassette is incorporated into the *clag3.2* donor plasmid and to treat parasites with 5-fluorocytosine directly after the WR99210 treatment. This has proved to be successful for eliminating multiple integrations in other transgenic lines of our lab and therefore it is worth trying it again for this project. The other option for solving the multiple integration problem is to change the *clag3.2*+eYFP design with the hope that the new design with a different sgRNA does not favour the incorporation of multiple integrations like our current design appears to do. We are aware that there are many different CRISPR-Cas9 strategies optimised for *P. falciparum* research[434], but given that our current strategy has worked perfectly well on many occasions, we do not believe that adapting a new strategy for this project is the most straightforward solution. Nevertheless, it is always a possibility if the other aforementioned options do not work.

Once we have found a solution for the multiple integration problem, the most direct line of action would be to first transfect our CLAG3-mCer single transgenic T13 line with the *clag3.2* construct using the hDHFR selectable marker (as T13 is free of residual episomal Cas9 plasmid). If this second edition is successful, the next step would be to try and eliminate residual Cas9 episome from this new line using preferably the combination of 5-fluorocytosine treatment with a big dilution. In this manner, a third sequential edition could be performed reusing the hDHFR selectable marker. If this was not possible, then the blasticidin deaminase selectable marker could be used for the third sequential transfection with the control construct. Once obtained, this triple transgenic line would have to be validated by PCR to confirm the correct integration of all tags, by qPCR to assess *clag3* expression levels and by flow cytometry to determine the fluorescence profiles of this line. All these screenings would include at least three independent biological replicates in order to conclusively characterise this line. If needed, FACS-based sorting, large dilutions or subcloning could be used to eliminate remaining subsets of parasites with an unedited locus.

With a characterised and functional triple transgenic line, we would first monitor by flow cytometry (and also validating by RT-qPCR) *clag3* expression patterns under normal culture conditions over a period of approximately 5 weeks and starting with a FACS-sorted pure subpopulation of parasites expressing *clag3.1*, *clag3.2* or both genes. The objective would be to determine the dynamic stochastic creation of expression variability

within the parasite population. By detecting the stochastic switches between active and silent transcriptional states of both *clag3* genes, it would be possible to determine the switching rates of these genes and the fidelity of transmission of these transcriptional states under standard culture conditions. Next, the objective would be to see how all the above varies when parasites are cultured in conditions that alter *clag3* expression patterns. By selecting parasites with different compounds such as BSD or T3 over a similar period of 5 weeks and analysing samples every 2 to 4 days, it would be possible to monitor how *clag3* expression patterns vary in response to these selective pressures and how they recover once said pressures have been removed. In this manner, we could determine the relative fitness of each of the different *clag3* expression patterns under each of the different conditions as well as study the kinetics of each selection. This would inform on the fitness advantage that each *clag3* expression pattern confers under each selective condition and help further understand the role of these genes in malaria parasite adaptation. Furthermore, this analysis would enable the identification of the spontaneous frequency of parasites that present both *clag3* genes silenced under basal conditions and determine if parasites with the two genes simultaneously active exist.

The final objective would be to generate a predictive mathematical model of *clag3* expression dynamics using all the information obtained. This model would help estimate the fundamental parameters behind *clag3* adaptation mechanisms, such as epigenetic switching rates or relative fitness rates associated to different expression patterns under different conditions; and once more, help characterise the adaptive potential of *clag3* genes.

7. Methods

7.1. Cloning

Guides were designed using the EuPaDGT Web-based tool and the sgRNA database generated by Ribeiro *et al.*[448]. The LA Taq DNA Polymerase (Takara) was used for PCR amplifications from 3D7-A *P. falciparum* genomic DNA. All cloning steps were performed using the In-Fusion HD Cloning Kit (Clontech). To clone the plasmids, we used *Escherichia coli* DH5α or Stellar Competent cells (Takara) for difficult cloning. All primers

and oligonucleotides were from Integrated DNA Technologies and are described in **Supplementary Table 2**. All the plasmids generated and used in the different cloning strategies are detailed in **Supplementary Table 3**.

7.2. Transfection

Ring-stage parasite cultures were transfected by standard electroporation using a BioRad apparatus, typically with 15 µg of donor plasmid linearised at a restriction site located in the plasmid backbone and 60 µg of Cas9 plasmid. Edited parasites were selected with 10 nM of WR99210 typically for 4 days, as previously described[437], followed by treatment with 1 µM of 5-fluorocytosine (clinical grade Ancotil, Mylan N.V.) for 10 days, unless stated otherwise. Alternative selections were performed with puromycin at 100 ng/ml for a variable number of days or with blasticidin S at 2.5 µg/mL for 6 days.

7. 3. PCR

To assess the correct edition of the transgenic lines, diagnostic PCR was performed using the LA Taq DNA Polymerase (Takara) on parasite gDNA. The primers used for diagnostic PCR are detailed in **Supplementary Table 4**.

7.4. RT-qPCR and qPCR

Parasite samples were collected in TRIzol (Invitrogen) and RNA was purified using the RNeasy Mini Kit (Qiagen) as previously described[335,447]. Next, the Reverse Transcription System (Promega) was used to reverse transcribe the purified RNA alongside parallel reactions without reverse transcriptase to exclude gDNA contamination. Quantitative PCR to analyse cDNAs or gDNAs was performed in triplicate wells using the Power SYBR Green Master Mix (Applied Biosystems) in a StepOnePlus Real Time PCR System, essentially as previously described[300,447]. The standard curve method was used to calculate relative transcript levels with the normalizing gene *serine-tRNA ligase* (*serrs*), that possess stable transcript levels across the intraerythrocytic development cycle. For *clag3* genes, *rhoph2* was used for normalization because it has similar temporal expression dynamics. The primers used for qPCR are detailed in **Supplementary Table 4**.

7.5. IFA

Air-dried smears were made using Percoll-purified schizont-stage parasites. Next, smears were fixed with 1% paraformaldehyde for 10 min and then permeabilized with 0.1% Triton X-100 in phosphate-buffered saline (PBS) for a further 10 min. Smears were incubated with any of the following primary polyclonal antibodies: mouse anti-CLAG3 (1:2,000; from mouse 167#2, kindly provided by Sanjay A. Desai, NIAID-NIH)[390] and rabbit anti-GFP (1:1,000; Invitrogen #A11122, lot #1828014; this antibody also reacts with eYFP and mCer given their sequence similarity[446]). This was followed by incubation with the following secondary antibodies: anti-mouse conjugated with Alexa Fluor 546 (Life Technologies, A-10036) and anti-rabbit conjugated with Alexa Fluor 488 (Life Technologies, A-11034). DAPI (4=,6=-diamidino-2-phenylindole) was used to stain the nuclei and preparations were mounted with Vectashield medium (Palex Medical) and visualized under an Olympus IX51 epifluorescence microscope. Images were acquired with an Olympus DP72 camera using CellSens Standard 1.11 software, and processed using ImageJ.

7.6. Flow cytometry and sorting

For flow cytometry, parasites were tightly synchronized to a 5 h age window the previous cycle(s) and then analysed approximately at 40 to 45 hpi, 1 or 2 cycles later. Samples (1 µL of parasite culture in 1 mL of sterile 1% PBS) were analysed using a BD LSR Fortessa™ apparatus which examined typically 100,000 cells per sample detecting eYFP (laser: 488 nM; Filter: 525/50-505LP; power: 50 mW), mCer (laser: 405 nM; Filter: 405-450/50-A; power: 50 mW) and / or tdTom (laser: 561 nm; filter: 582/15; power: 50 mW). Initial gating was performed using the side scatter and forward scatter areas (SSC-A versus FSC-A plot) to define the RBC population. Next, singlets were gated using the forward scatter height (FSC-H) versus FSC-A plot. Finally, the fluorescence profiles of the different parasite lines were analysed using the different lasers. Flow cytometry data was collected using BD FASCDiVa v.6.1.3 software and downstream analysis was performed using Flowing Software version 2.5.1 (Perttu Terho). As a stage control to differentiate between infected and non-infected RBCs and ring and late stage parasites, we used MitoTracker as a marker of active mitochondria as previously described[243].

For sorting, parasites were tightly synchronized to a 5 h age window the previous cycle(s) and then before the sorting experiment, schizonts were Percoll-purified at approximately 38 to 43 hpi, 1 or 2 cycles later. A FACSAria SORP (BD Biosciences; 5 lasers, 18 parameters) cell sorter was used to sort eYFP+ or tdTom+ parasites, using the 3D7-A wild-type line as a parallel negative control. eYFP measurements were made with the 488 nm-laser at 100 MW and tdTom measurements with the 561 nm-laser at 100 MW. The cell sorting conditions were as follows: flow chip diameter: 100 µm, sheath solution: ISOTON (Beckman Coulter), sheath pressure: 20 PSI and sort mode: purity. The erythrocytes population was identified and gated on SSC-A vs FSC-A plots. To avoid the sorting of cell doublets or cell aggregates, single cells were selected on FSC-H vs FSC-A plots. eYFP-positive parasites were selected from plots combining the fluorescence channels 488-525/50-505LP-A and 488-670/14-635LP-A and tdTom-positive parasites were selected from plots combining the fluorescence channels 561-582/15-A and 488-525/50-505LP-A. Sorted parasites were collected into a tube with culture media and 2 to 5 µL of packed erythrocytes at 37ºC and then placed back in culture.

7.7. Subcloning by limiting dilution

Before subcloning, late-stage cultures were kept under shaking conditions (100 rpm) for 24 h in order to reduce the number of double-infected erythrocytes. Next, ring-stage parasites were synchronized by sorbitol lysis and then parasitaemia was measured using flow cytometry as previously described[430]. After this, cultures were sequentially diluted to a final concentration of 0.3 parasites/100 µL and the diluted culture was divided into 40 wells (each containing 100 µL) in a 96-well plate that also included controls of 1000 and 100 parasites/100 µL. Media was changed once a week and parasitaemia was checked by Giemsa-stained thin smears from day 10 onwards, with growth expected in less than 50% of the wells.

8. Supplementary tables

Supplementary table 1. List of all the transfections performed to date during the development of the *clag3* triple transgenic line. For each transfection the following information is shown: ID, original parasite line used for transfection, the CRISPR-Cas9 construct used, name of the donor plasmid(s) used, amount of donor plasmid used, the state the donor plasmid was transfected in (L = linearised, Lx2 = linearised in two places and C = circular or non-linearised), name of the Cas9 plasmid(s) used, the amount of Cas9 plasmid used, the selectable marker used for the selection of edited parasites (H = hDHFR, P= puromycin-N-acetyl transferase and B = blasticidin deaminase), date of the transfection, date that parasites were detected by light microscopy post-transfection, presence of edited parasites (Y = yes, N = no), presence of unedited parasites (Y/N), presence of residual episomal Cas9 plasmid (Y/N), presence of multiple integrations of donor plasmid in the edited loci (Y/N) and other comments.

ID	Original Parasite line	Construct	Donor Plasmid	µg Donor plasmid	Donor plasmid state	Cas9 plasmid	µg Cas9 plasmid	Selectable marker	Date of transfection	Date of visible parasites	Edited	Unedited	Episomal Cas9 plasmid	Multiple integrations	Comments
T1	3D7-A	*rhoph3* + tdTom	pHR1rhoph3-TdTomato-HR2rhoph3	12	L	pDC2-cam-Cas9-U6-hDHFRyFCU_sgRNA_rhoph3	60	H	03/10/17	23/10/17	N	Y			
T2	3D7-A	*clag3.2* + eYFP	pHR1clag3.2-eYFP-HR2clag3.2	12	L	pDC2-cam-Cas9-U6-hDHFRyFCU_sgRNA_clag3.2	60	H	03/10/17	23/10/17	Y	N	Y	Y	Single transgenic line
T3	3D7-A	*rhoph3* + tdTom & *clag3.2* + eYFP	pHR1rhoph3-TdTomato-HR2 rhoph3 + pHR1clag3.2-eYFP-HR2clag3.2	12	L	pDC2-cam-Cas9-U6-hDHFRyFCU_sgRNA_rhoph3 + pDC2-cam-Cas9-U6-hDHFRyFCU_sgRNA_clag3.2	60	H	03/10/17	23/10/17	N	Y			
T4	T2	*clag3.1* + mCer	pUC19_HR1_clag3.1_mCer_HR2_clag3.1	30	L	pDC2-cam-Cas9-U6-hDHFRyFCU_sgRNA_clag3.1	60	H	02/05/18	15/05/18	N	Y			
T5	T2	*rhoph3* + tdTom	pHR1rhoph3-TdTomato-HR2rhoph3	30	L	pDC2-cam-Cas9-U6-hDHFRyFCU_sgRNA_rhoph3	60	H	02/05/18	15/05/18	N	Y			
T6	T2	*rhoph3* + tdTom & *clag3.2* + eYFP	pHR1rhoph3-tdTomato-HR2 rhoph3 + pHR1clag3.2-eYFP-HR2clag3.2	30	L	pDC2-cam-Cas9-U6-hDHFRyFCU_sgRNA_rhoph3 + pDC2-(...)-sgRNA_clag3.2	60	H	02/05/18	15/05/18	N	Y			

T7	3D7-A	*clag3.1* + mCer	pUC19_HR1_ clag3.1_mCer_ HR2_clag3.1	30	L	pDC2-cam-Cas9-U6-hDHFRyFCU_ sgRNA_clag3.1	60	H	25/06/18	x	N				
T8	3D7-A	*rhoph3* + tdTom	pHR1rhoph3-TdTomato-HR2rhoph3	30	L	pDC2-cam-Cas9-U6-hDHFRyFCU_ sgRNA_rhoph3	60	H	25/06/18	x	N				
T9	T2	*rhoph3* + tdTom	pHR1rhoph3-TdTomato-HR2rhoph3	30	L	pDC2-cam-Cas9-U6-hDHFRyFCU_ sgRNA_rhoph3	60	H	25/06/18	x	N				
T10	T2	*rhoph3* + tdTom	pHR1rhoph3-TdTomato-HR2rhoph3	30	L	pDC2-cam-Cas9-U6-hDHFRyFCU_ sgRNA_rhoph3	60	H	25/06/18	x	N				
T11	T2	*clag3.1* + mCer	pUC19_HR1_ clag3.1_mCer_HR2_clag3.1	15	L	pDC2-cam-Cas9-U6-hDHFRyFCU_ sgRNA_clag3.1	60	H	17/09/18	02/10/18	N	Y			
T12	T2	*rhoph3* + tdTom	pHR1rhoph3-TdTomato-HR2rhoph3	15	L	pDC2-cam-Cas9-U6-hDHFRyFCU_ sgRNA_rhoph3	60	H	17/09/18	02/10/18	N	Y			
T13	3D7-A	*clag3.1* + mCer	pUC19_HR1_ clag3.1_mCer_ HR2_clag3.1	15	L	pDC2-cam-Cas9-U6-hDHFRyFCU_ sgRNA_clag3.1	60	H	17/09/18	05/10/18	Y	Y	Y	N	Single transgenic line
T14	3D7-A	*rhoph3* + tdTom	pHR1rhoph3-TdTomato-HR2rhoph3	15	L	pDC2-cam-Cas9-U6-hDHFRyFCU_ sgRNA_rhoph3	60	H	17/09/18	12/10/18	N	Y			
T15	3D7-A	*rhoph3* + tdTom & *clag3.1* + mCer	pHR1rhoph3-TdTomato-HR2 rhoph3 + pUC19_HR1_ clag3.1_mCer_ HR2_clag3.1	15	L	pDC2-cam-Cas9-U6-hDHFRyFCU_ sgRNA_rhoph3 + pDC2-cam-Cas9-U6-hDHFRyFCU_ sgRNA_clag3.1	60	H	17/09/18	26/10/18	N	Y			
T16	3D7-A	*clag3.2* + eYFP	pHR1clag3.2-eYFP-HR2clag3.2	15	L	pDC2-Cas9-PACyFCU_clag3.2	60	P	04/03/19	15/03/19	N	Y			

T17	3D7-A	promoter + tdTom in *lisp1*	ama1_ TdTomato_ LISP_HR1_HR2	15	Lx2	pDC2-Cas9-hDHFRyFCU_ sgRNA_LISP1	60	H	04/03/19	22/03/19	Y	N			No tdTom signal
T18	3D7-A	promoter + tdTom in *lisp1*	ama1_ TdTomato_ LISP_HR1_HR2	15	Lx2	pDC2-Cas9-PACyFCU_lisp1	60	P	04/03/19	15/03/19	N	Y			
T19	T13	*clag3.2* + eYFP	pHR1clag3.2-eYFP-HR2clag3.2	15	L	pDC2-Cas9-PACyFCU_clag3.2	60	P	04/03/19	x	N				
T20	T13	*clag3.2* + eYFP	pHR1clag3.2-eYFP-HR2clag3.2	15	L	pDC2-cam-Cas9-U6-hDHFRyFCU_ sgRNA_clag3.2	60	H	04/03/19	19/03/19	Y	N	Y	Y	Double transgenic line
T21	T13	promoter + tdTom in *lisp1*	ama1_ TdTomato_ LISP_HR1_HR2	15	Lx2	pDC2-Cas9-hDHFRyFCU_ sgRNA_LISP1	60	H	04/03/19	22/03/19	Y	N	Y	Y	No tdTom signal
T22	T13	promoter + tdTom in *lisp1*	ama1_ TdTomato_ LISP_HR1_HR2	15	Lx2	pDC2-Cas9-PACyFCU_lisp1	60	P	04/03/19	x	N				
T23	3D7-A	promoter + tdTom in *lisp1*	ama1_ TdTomato_ LISP_HR1_HR2	15	Lx2	pDC2-Cas9-PACyFCU_lisp1	60	P	28/03/19	x	N				
T24	3D7-A	*clag3.2* + eYFP	pHR1clag3.2-eYFP-HR2clag3.2	15	L	pDC2-Cas9-PACyFCU_clag3.2	60	P	28/03/19	x	N				
T25	3D7-A	*clag3.2* + eYFP	pHR1clag3.2-eYFP-HR2clag3.2	15	L	pDC2-Cas9-PACyFCU_clag3.2	60	P	28/03/19	19/04/19	N	Y			
T26	3D7-A	promoter + tdTom in *lisp1*	ama1_ TdTomato_ LISP_HR1_HR2	15	L	pDC2-Cas9-hDHFRyFCU_ sgRNA_LISP1	60	H	14/08/19	04/09/19	Y	N			No tdTom signal
T27	3D7-A	promoter + tdTom in *lisp1*	rap2_TdTomato _LISP1_HR1_HR2 (RIF)	15	L	pDC2-Cas9-hDHFRyFCU_ sgRNA_LISP1	60	H	14/08/19	02/09/19	Y	Y			Wrong design (*rif* fragment)

T28	3D7-A	promoter + tdTom in *lisp1*	gap45_ short_ TdTomato_LISP _HR1_HR2 (RIF)	15	L	pDC2-Cas9-hDHFRyFCU_ sgRNA_LISP1	60	H	14/08/19	29/08/19	Y	N			Wrong design (*rif* fragment)
T29	3D7-A	promoter + tdTom in *lisp1*	gap45_long_ TdTomato_LISP1 _HR1_HR2 (RIF)	15	L	pDC2-Cas9-hDHFRyFCU_ sgRNA_LISP1	60	H	14/08/19	29/08/19	Y	N			Wrong design (*rif* fragment)
T30	3D7-A	promoter + tdTom in *lisp1*	gap45_long_ TdTomato_LISP1 _HR1_HR2	15	L	pDC2-Cas9-hDHFRyFCU_ sgRNA_LISP1	60	H	13/01/20	29/01/20	Y	N	Y	Y	Single transgenic line
T31	3D7-A	promoter + tdTom in *lisp1*	gap45_short_T dTomato_LISP1_ HR1_HR2	15	L	pDC2-Cas9-hDHFRyFCU_sgRNA _LISP1	60	H	13/01/20	27/01/20	Y	N	Y	Y	Single transgenic line
T32	T20	promoter + tdTom in *lisp1*	gap45_long_ TdTomato_LISP1 _HR1_HR2	15	L	pDC2-Cas9-bsd_lisp1	60	B	13/01/20	27/01/20	Y	Y	Y	Y	Triple transgenic line
T33	T20	promoter + tdTom in *lisp1*	gap45_short_ TdTomato_LISP1 _HR1_HR2	15	L	pDC2-Cas9-bsd_lisp1	60	B	13/01/20	27/01/20	Y	Y	Y	Y	Triple transgenic line
T34	T13	*clag3.2* + eYFP	pHR1clag3.2-eYFP-HR2clag3.2	15	C	pDC2-cam-Cas9-U6-hDHFRyFCU_ sgRNA_clag3.2	60	H	14/02/21	04/03/21	N	Y			
T35	T13	*clag3.2* + eYFP	pHR1clag3.2-eYFP-HR2clag3.2	15	L	pDC2-cam-Cas9-U6-hDHFRyFCU_ sgRNA_clag3.2	60	H	14/02/21	04/03/21	Y	Y		Y	
T36	T32		-	-	-	pDC2-cam-Cas9-U6-hDHFRyFCU_ sgRNA_episome3	60	H	18/03/21	31/03/21	N	Y			
T37	T32		-	-	-	pDC2-cam-Cas9-U6-hDHFRyFCU_ sgRNA_episome3	60	H	18/03/21	31/03/21	N	Y			

T38	T32		-	-	-	pDC2-cam-Cas9-U6-hDHFRyFCU_sgRNA_episome3	60	H	18/03/21	31/03/21	N	Y				
T39	T13	*clag3.2* + eYFP	pHR1clag3.2-eYFP-HR2clag3.2_PvuII_sgRNA4	15	L	pDC2-cam-Cas9-U6-hDHFRyFCU_sgRNA_ clag3.2 + pDC2-cam-Cas9-U6-hDHFRyFCU_sgRNA_episome4	60	H	30/03/21	17/04/21	Y	Y			Y	
T40	T13	*clag3.2* + eYFP	pHR1clag3.2-eYFP-HR2clag3.2_PvuII_sgRNA4	15	L	pDC2-cam-Cas9-U6-hDHFRyFCU_sgRNA_clag3.2 + pDC2-(...)-sgRNA_episome4	60	H	30/03/21	14/04/21	Y	Y			Y	
T41	T13	*clag3.2* + eYFP	pHR1clag3.2-eYFP-HR2clag3.2_PvuII_sgRNA4	15	C	pDC2-cam-Cas9-U6-hDHFRyFCU_sgRNA_clag3.2 + pDC2-(...)-sgRNA_episome4	60	H	30/03/21	17/04/21	Y	Y			Y	
T42	T13	*clag3.2* + eYFP	pHR1clag3.2-eYFP-HR2clag3.2_PvuII_yFCU	15	L	pDC2-cam-Cas9-U6-hDHFRyFCU_sgRNA_clag3.2	60	H	15/04/21	05/05/21	Y	Y			Y	
T43	T13	*clag3.2* + eYFP	pHR1clag3.2-eYFP-HR2clag3.2_PvuII_yFCU	15	L	pDC2-cam-Cas9-U6-hDHFRyFCU_sgRNA_clag3.2	60	H	15/04/21	05/05/21	Y	N			Y	
T44	T13	*clag3.2* + eYFP	pHR1clag3.2-eYFP-HR2clag3.2_PvuII_yFCU	15	L	pDC2-cam-Cas9-U6-hDHFRyFCU_sgRNA_clag3.2 + pDC2-(...)-sgRNA_episome3	60	H	15/04/21	03/05/21	Y	Y			Y	

Supplementary table 2: List of all the primers and oligonucleotides used for the generation of the CRISPR-CAS9 plasmids used in this project. For each primer the following information is shown: ID (with which it is referred to in the text), name, sequence in 5′ to 3′ orientation and a brief description of what it amplifies.

ID	Name	Sequence (5' → 3')	Details
P1	clag3.2_HR1_FW	GCGGGGAGGACTAGTGTTGGCCTC CTAGTTCAGA	Forward primer for *clag3.2* homology region 1
P2	clag3.2_HR1_RV	TGCTGCTGCAGATCTGTGGAACCGC GCAATCATCAACTCGTTATCGGGAGT ATCAAAATCCTTTTCATCG	Reverse primer for *clag3.2* homology region 1
P3	clag3.2_HR2_FW	TACAAGTAACTCGAGTATATATACATA CATACCTATGTA	Forward primer for *clag3.2* homology region 2
P4	clag3.2_HR2_RV	GTGCCACCTGACGTCGTTCATCTGA ATTTTCAATAATAT	Reverse primer for *clag3.2* homology region 2
P5	clag3.2_sgRNA	TAAGTATATAATATTGCGATCATTAATT CATTGTCGTTTTAGAGCTAGAA	Oligonucleotide with *clag3.2* sgRNA
P6	clag3.2_sgRNA (RV)	TTCTAGCTCTAAAACGACAATGAATT AATGATCGCAATATTATATACTTA	Reverse complement of *clag3.2* sgRNA oligonucleotide
P7	mCer-FW	CCAGATGACGAGTTGATGATTTCTCG TTTCCATAGATCTGCAGCAGCAGCA GCAGCA ATGGTGAGCAAGGGCGAGG	Forward primer for mCerulean
P8	mCer-RV	CGACTCTAGAGGATCTCTCGAGTTA CTTGTACAGCTCGTCCATGC	Reverse primer for mCerulean
P9	clag3.1_HR1_FW	CGGTACCCGGGGATCCTTCTTCACT CACGGACTTGC	Forward primer for *clag3.1* homology region 1
P10	clag3.1_HR1_RV	CAACTCGTCATCTGGAGTATCAGAAT CCTTTTCATCG	Reverse primer for *clag3.1* homology region 1
P11	clag3.1_HR2_FW	CGAGAGATCCTCTAGATATAAAAATA TAAAAAATATATATATATAATTTTGTAC ATAATAGCAATTC	Forward primer for *clag3.1* homology region 2
P12	clag3.1_HR2_RV	GCAGGTCGACTCTAGAAAGTATGGC TTAAATATTTTAATCT	Reverse primer for *clag3.1* homology region 2
P13	clag3.1_sgRNA	TAAGTATATAATATTGATATCATTAATT CATCGTCGTTTTAGAGCTAGAA	Oligonucleotide with *clag3.1* sgRNA
P14	clag3.1_sgRNA (RV)	TTCTAGCTCTAAAACGACGATGAATT AATGATATCAATATTATATACTTA	Reverse complement of *clag3.1* sgRNA oligonucleotide
P15	RhopH3_HR1_FW	GCGGGGAGGACTAGTCCACAAACTA TGGATGAAGC	Forward primer for *rhoph3* homology region 1
P16	RhopH3_HR1_RV	TGCTGCTGCAGATCTTAACTCGTTCT CCGAGGTGAACGTTTTAGATTCTTCT AATAATATTCT	Reverse primer for *rhoph3* homology region 1
P17	RhopH3_HR2_FW	TACAAGTAACTCGAGATTAAAAATTTA ATCCTACATGTAGA	Forward primer for *rhoph3* homology region 2
P18	RhopH3_HR2_RV	GTGCCACCTGACGTCCACTCTTTTAC ATATTACGGGG	Reverse primer for *rhoph3* homology region 2

P19	RhopH3_sgRNA	TAAGTATATAATATTCAATTCATTTTCAGAAGTAAGTTTTAGAGCTAGAA	Oligonucleotide with *rhoph3* sgRNA
P20	rhopH3_sgRNA (RV)	TTCTAGCTCTAAAACTTACTTCTGAAAATGAATTGAATATTATATACTTA	Reverse complement of *rhoph3* sgRNA oligonucleotide
P21	rap2-prom-tag Fw	CGTCCCAGAAGTCGACGTTGTAATTCTAAAACACTCCCA	Forward primer for *rap2* promoter
P22	rap2_prom_internal_RV	GCTCACCATGCGGCCGCTTTGTTGCTCTTTTTTTTTTTTTTT	Reverse primer for *rap2* promoter
P23	gap45_short-prom-tag Fw	CGTCCCAGAAGTCGACACACATTAACACATTATTCAGCT	Forward primer for *gap45* promoter (short)
P24	gap45_long-prom-tag Fw	CGTCCCAGAAGTCGACTCAAAGGATTATTTAAGCAAAGC	Forward primer for *gap45* promoter (long)
P25	gap45_prom_internal_RV	GCTCACCATGCGGCCGCTATTATATAATATTGAATACAGTAAAA	Reverse primer for *gap45* promoter
P26	ama1-prom-tag Fw	CGTCCCAGAAGTCGACCCACGATTTTTTAATCAGATCCC	Forward primer for *ama1* promoter
P27	ama1-prom-tag Rv	GCTCACCATGCGGCCGCTTTTGTACAATTTATAACAAGTACAT	Reverse primer for *ama1* promoter
P28	sgRNA3_Fw	TAAGTATATAATATTAAATTCAGATGAACGACGTCGTTTTAGAGCTAGAA	Oligonucleotide with episome 3 sgRNA
P29	sgRNA3_Rv	TTCTAGCTCTAAAACGACGTCGTTCATCTGAATTTAATATTATATACTTA	Reverse complement of episome 3 sgRNA oligonucleotide
P30	sgRNA4_Fw	TAAGTATATAATATTCTGATTGCAGATCTACTCACGTTTTAGAGCTAGAA	Oligonucleotide with episome 4 sgRNA
P31	sgRNA4_Rv	TTCTAGCTCTAAAACGTGAGTAGATCTGCAATCAGAATATTATATACTTA	Reverse complement of episome 4 sgRNA oligonucleotide
P32	Oligo_sgRNA4_PvuII_Fw	GATTCATTAATGCAGCTGCTGATTGCAGATCTACTCACGGGCTGGCACGACAGGTT	Oligonucleotide with episome 4 sgRNA target
P33	Oligo_sgRNA4_PvuII_Rv	AACCTGTCGTGCCAGCCCGTGAGTAGATCTGCAATCAGCAGCTGCATTAATGAATC	Reverse complement of episome 4 sgRNA target oligonucleotide
P34	pL1144400_NotI/SacII_frag.F	GATTCATTAATGCAGCTGGCGGCCGCACTTACCGCGGCTGGCACGACAGGTT	Oligonucleotide with Not I and Sac II restriction sites
P35	pL1144400_NotI/SacII_frag.R	AACCTGTCGTGCCAGCCGCGGTAAGTGCGGCCGCCAGCTGCATTAATGAATC	Reverse complement of oligonucleotide with Not I and Sac II restriction sites

Supplementary table 3: List of all the CRISPR-CAS9 plasmids used in this project. For each plasmid the following information is shown: name (with which it is referenced in the text), plasmid type (whether it is a donor or a Cas9 plasmid) and the CRISPR-CAS9 construct for which it is used.

	Plasmid name	Plasmid type	Construct
1	pHR1clag3.2-eYFP-HR2ap2G	Donor - intermediate version	*clag3.2* + eYFP
2	pHR1clag3.2-eYFP-HR2clag3.2	Donor - final version	
3	pDC2-cam-Cas9-U6-hDHFRyFCU_sgRNA_ clag3.2	Cas9	
4	pUC19_HR1_clag3.1_mCer	Donor - intermediate version	*clag3.1* + mCer
5	pUC19_HR1_clag3.1_mCer_HR2_clag3.1	Donor - final version	
6	pDC2-cam-Cas9-U6-hDHFRyFCU_sgRNA_clag3.1	Cas9	
7	pHR1rhoph3-TdTomato-HR2ap2G	Donor - intermediate version	*rhoph3* + tdTom
8	pHR1rhoph3-TdTomato-HR2rhoph3	Donor - final version	
9	pDC2-cam-Cas9-U6-hDHFRyFCU_sgRNA_rhoph3	Cas9	
10	ama1_TdTomato_LISP1_HR1_HR2	Donor - final version	*ama1* promoter + tdTom in lisp1
11	gap45_short_TdTomato_LISP1_HR1_HR2	Donor - final version	*gap45* promoter (short) + tdTom in lisp1
12	gap45_long_TdTomato_LISP1_HR1_HR2	Donor - final version	*gap45* promoter (long) + tdTom in lisp1
13	pDC2-Cas9-hDHFRyFCU1_sgRNA_LISP1	Cas9	All lisp1 constructs
14	pDC2-cam-Cas9-U6-hDHFRyFCU_sgRNA_episome3	Cas9	To cut plasmid backbone
15	pDC2-cam-Cas9-U6-hDHFRyFCU_sgRNA_episome4	Cas9	To cut plasmid backbone
16	pHR1clag3.2-eYFP-HR2clag3.2_PvuII_sgRNA4	Donor - final version	*clag3.2* + eYFP
17	pHR1clag3.2-eYFP-HR2clag3.2_PvuII_yFCU	Donor - final version	*clag3.2* + eYFP
18	pDC2-Cas9-PACyFCU_clag3.2	Cas9 with PAC marker	*clag3.2* + eYFP
19	pDC2-Cas9-bsd_lisp1	Cas9 with blasticidin deaminase marker	All *lisp1* constructs
20	pDC2-Cas9-PACyFCU_lisp1	Cas9 with PAC marker	All *lisp1* constructs

Supplementary table 4: List of all the primers used for the diagnostic PCR and qPCR analyses of this project. For each primer the following information is shown: ID (with which it is referred to in the text), name, sequence in 5' to 3' orientation, a brief description of what it targets (if it falls within a homology region of one the CRISPR-CAS9 designs it is indicated) and reference.

ID	Name	Sequence (5' - 3')	Target	Reference
P36	Promoter_AMA1_-460	TTGAAACCTTTACACAAA CGTTA	*ama1* promoter	-
P37	Promoter_AMA1_-943	GTTTTCATAAAAGAAGTTG CCTT	*ama1* promoter	-
P38	AmpR_+234_R	CTTCGGTCCTCCGATCGT T	Ampicillin	Designed by Oriol Llorà-Batlle
P39	AmpR_+600_R	TGCCTGACTCCCCGTCG T	Ampicillin	Designed by Oriol Llorà-Batlle
P40	AmpR_+62_F	GCAACTCGGTCGCCGCA T	Ampicillin	Designed by Oriol Llorà-Batlle
P41	bsdF	TCTCTGAAGACTACAGCG TC	Blasticidin deaminase	Rovira-Graells *et al.*, 2015
P42	bsdR	ACGATACAAGTCAGGTTG CC	Blasticidin deaminase	Rovira-Graells *et al.*, 2015
P43	3.1 Fw	TGTGCAATATATCAAAGT GTACATGC	*clag3.1* promoter	Iriko *et al.*, 2008
P44	3.1 RV	TAGAAAATATTAGAATTGC TATTATGTAC	*clag3.1* terminator (Inside HR2)	Iriko *et al.*, 2008
P45	clag3.1_6F_+4335	GGTTCCCCAAGTTCTACT GC	*clag3.1* CDS	Designed by Núria Rovira-Graells
P46	clag3.1_HR2_qPCR_Fw	TTGTAATTCTTTAATGTGTA TAAGA	*clag3.1* terminator (Inside HR2)	-
P47	clag3.1_HR2_qPCR_RV	AAGTATGGCTTAAATATTT TAATCT	*clag3.1* terminator (Inside HR2)	-
P48	clag3.1_HR2_RV	GCAGGTCGACTCTAGAA AGTATGGCTTAAATATTTT AATCT	*clag3.1* terminator (Inside HR2))	-
P49	clag3.1_terminator_Rv	AACACAATATATATAAAC GATTAAG	*clag3.1* terminator (Inside HR2)	-
P50	clag3.1-6R	ATAAATATTTGGATGCTTC AGCA	*clag3.1* CDS (Inside HR1)	Crowley *et al.*, 2011
P51	3.2 Fw	AATAGTTGAGTACGCACT AATATGTC	*clag3.2* promoter	Iriko *et al.*, 2008

P52	3.2 RV	ACACAAATTCTTAATAATTATATAAAACC	*clag3.2* terminator (Inside HR2)	Iriko *et al.*, 2008
P53	clag3.2_+4430_Fw	AAAAATAGTACTAGTTATAAATCTC	*clag3.2* CDS	-
P54	Clag3.2_HR2_mid_Rv	CAATATTCTAAAAATTAATATTACGA	*clag3.2* terminator (Inside HR2)	-
P55	clag3.2_HR2_qPCR_Fw	TAAAATTAAAAGAGTCACATTCATT	*clag3.2* terminator (Inside HR2)	-
P56	clag3.2_HR2_qPCR_RV	GTTCATCTGAATTTTCAATAATAT	*clag3.2* terminator (Inside HR2))	-
P57	clag3.2_HR2_RV	GTGCCACCTGACGTCGTTCATCTGAATTTTCAATAATAT	*clag3.2* terminator (Inside HR2)	-
P58	Clag3.2_terminator_+0_Rv	ATACATAGGTATGTATGTATATATA	*clag3.2* terminator (Inside HR2)	-
P59	Clag3.2_terminator_+1254_Rv	CGATTAGTACTAATCGTGATATG	*clag3.2* terminator (Inside HR2)	-
P60	Clag3.2_terminator_+461_Rv	ATATTATTGAAAATTCAGATGAAC	*clag3.2* terminator (Inside HR2)	-
P61	Clag3.2_terminator_+601_Rv	CATTTAATAATGTTTACAAAACACA	*clag3.2* terminator	-
P62	Clag3.2_terminator_+941_Rv	CGTAAGTTATACACACATGTCA	*clag3.2* terminator	-
P63	clag3.2_terminator_Rv	TGTGTTTTGTAAACATTATTAAATG	*clag3.2* terminator	-
P64	clag3.2-6F_+4380	ATCCCAAAAGATGTACTCCTG	*clag3.2* CDS	Designed by Núria Rovira-Graells
P65	clag3.2-6R	ACAAATATGTTTCTGAACTAGGA	*clag3.2* CDS (Inside HR1)	Crowley *et al.*, 2011
P66	clag3.1 clag3.2-6F	TAGTAATGAGAATTAGTTGGACA	*clag3.2* & *clag3.1* (CDS of both)	Crowley *et al.*, 2011
P67	eYFP_+180_Fw	CGTGACCACCTTCGGCTAC	eYFP	-
P68	eYFP_+417_Rv	CAAGCTGGAGTACAACTACAA	eYFP	-

P69	eYFP_+685_Rv	TCACTCTCGGCATGGACGA	eYFP	Designed by Núria Rovira-Graells
P70	eYFP_NR_Fw	AGCGCACCATCTTCTTCAAGG	eYFP	Designed by Núria Rovira-Graells
P71	GAP45_PROM_internal_RV	GCTCACCATGCGGCCGCTATTATATAATATTGAATACAGTAAAA	*gap45* promoter	-
P72	GAP45_PROMOTER-1165_FW	TGCAATAAAAGTGCATAGAAAAC	*gap45* promoter	-
P73	GAP45_PROMOTER-2170_FW	ATCAAATGAAATGATATAACGGG	*gap45* promoter	-
P74	GAP45_PROMOTER-426_FW	TCATGCCTTTTAAACAATAATATAT	*gap45* promoter	-
P75	hdhfrF	AGTAGAAGGTAAACAGAATCTG	*hdhfr*	Rovira-Graells *et al.*, 2015
P76	hdhfrR	GGCATCATCTAGACTTCTGG	*hdhfr*	Rovira-Graells *et al.*, 2015
P77	LISP1_+5088_F	TATGAAGAATATATTGAACGAATC	*lisp1* CDS	Llorà-Batlle *et al.*, 2020
P78	LISP1_+6006_R	AGTATACCCAGGAGTGGATAA	*lisp1* CDS	Llorà-Batlle *et al.*, 2020
P79	Dir/M13_Fw (UP)	GTTTTCCCAGTCACGAC	Plasmid Backbone	-
P80	Episome_3'_qPCR_F	ATGTATCCGCTCATGAGACAA	Plasmid Backbone	Designed by Oriol Llorà-Batlle
P81	Episome_3'_qPCR_R	TGAGCAAAAACAGGAAGGCAA	Plasmid Backbone	Designed by Oriol Llorà-Batlle
P82	Episome_5'_qPCR_F	TTACACTTTATGCTTCCGGCT	Plasmid Backbone	Designed by Oriol Llorà-Batlle
P83	Episome_5'_qPCR_R	GAGTATTCTATAGTGTCACCTA	Plasmid Backbone	Designed by Oriol Llorà-Batlle
P84	Rer/M13_Rv (UP)	CAGGAAACAGCTATGAC	Plasmid Backbone	-
P85	SP6	GCCAAGCTATTTAGGTGACAC	Plasmid Backbone	-
P86	rhopH2_qRT_P2F	TGTTGCTGTCCATATTTAGTTTT	*rhopH2* CDS	Crowley *et al.*, 2011

P87	rhopH2_qRT_P2R	AATATATCGCTACATAACTTCGT	*rhopH2* CDS	Crowley *et al.*, 2011
P88	PF07_0073_Fw	AAGTAGCAGGTCATCGTGGTT	*seryl tRNA synthetase* CDS	Salanti *et al.*,2003
P89	PF07_0073_Rv	TTCGGCACATTCTTCCATAA	*seryl tRNA synthetase* CDS	Salanti *et al.*,2003
P90	TDT_+0_Rv	TGACCTCCTCGCCCTTGC	tdTomato	-
P91	TdTomato_+879_R	CTTGGTCACCTTCAGCTTGG	tdTomato	Designed by Elisabet Tintó Font
P92	TdTomato_+726_F	TCCGAGGACAACAACATGGC	tdTomato	Designed by Elisabet Tintó Font
P93	yfCU_+1069_Fw	CAAGACTTCTGTAGCCATGAT	yFCU	-

Global results summary

Supervisor report 2.4 (PhD student's contribution to each of the articles included in this thesis.

Supervisor report 2.5 (Certification of the admission and publication of the articles included in this thesis)

Global results summary

Supervisor report 2.4: PhD student's contribution to each of the articles included in this thesis.

Alfred Cortés
ICREA Research Professor

ISGlobal
CEK Building
C./ Rosselló 153, 1st floor
08036 BARCELONA
Catalonia, Spain
Tel.: +34 93 2275400 ext. 4276
alfred.cortes@isglobal.org

21 September 2021

Report of the Thesis supervisor: contribution of the PhD candidate to the scientific articles included in this Thesis

Article 1. "Identification of Antimalarial Compounds That Require CLAG3 for Their Uptake by *Plasmodium falciparum*-Infected Erythrocytes"

Published in *Antimicrob. Agents Chemother.* (2019)

Impact Factor on the year of publication: 4.904 (first quartile, Q1)

Anastasia K. Pickford is co-first author in this article. Her contribution was to perform the IFA experiments presented in Fig. 1, the selection experiments with compounds T3 and T16 and subsequent transcriptional characterization of the selected lines presented in Fig. 3A, Fig. S1, Fig. S2, Fig. S3 and Fig. S4, and the characterization of the parasite lines selected with T3 using permeability assays and IFA presented in Fig. 3B-C and Fig. 4.

A preliminary version of this article was included in the PhD thesis of Sofía Mira-Martínez. However, the final version of the article includes a lot of new results not included in the preliminary version. The majority of the new experiments were performed by Anastasia K. Pickford. These new results are additional replicates for several experiments, the detailed characterization of the T3-selected parasite lines at multiple levels, selection experiments with different concentrations of the compound, and experiments to determine if the changes occurring in the parasites after selection with T3 are reversible when the pressure is removed. Therefore, since the final version of the article included in this thesis is substantially different from the preliminary version included in the Thesis of Sofía Mira-Martínez, and after consultation with the Doctorate School academic commission (Biomedicine program), I consider that this article can be included in the PhD Thesis of Anastasia K. Pickford.

Article 2. "Expression Patterns of *Plasmodium falciparum* Clonally Variant Genes at the Onset of a Blood Infection in Malaria-Naive Humans"

Published in *mBio* (2021).

Last Impact Factor available (2020): 7.867 (first quartile, Q1)

Anastasia K. Pickford is first author in this article. She did essentially all the experimental work, and together with a bioinformatician she analysed the data. She also prepared the figures, and wrote the manuscript together with the senior author.

Article 3. "Adaptation of *Plasmodium falciparum* to medium lacking a lipids supplement is associated with mutations in the *pfndh2* gene"

Article under review (submitted on 8th September 2021).

Anastasia K. Pickford is first author in this manuscript. She did all the experimental work, and together with a bioinformatician she analysed the data. She also prepared the figures, and wrote the manuscript together with the senior author.

In addition, Anastasia K. Pickford is co-first author in a Methods article not included in this PhD thesis: "**Transcriptional Analysis of Tightly Synchronized *Plasmodium falciparum* Intraerythrocytic Stages by RT-qPCR**". Casas-Vila N*, Pickford AK*, Portugaliza HP, Tintó-Font E, Cortés A. *Methods Mol. Biol.* 2021;2369:165-185. doi: 10.1007/978-1-0716-1681-9_10.

* Equal contribution as first authors

Firmado digitalmente
por CORTES CLOSAS
ALFRED - 46549133P
Fecha: 2021.09.21
23:53:01 +02'00'

Alfred Cortés Closas
ICREA Research Professor at ISGlobal

Supervisor report 2.5: Certification of the admission and publication of the articles included in this thesis

Alfred Cortés
ICREA Research Professor

ISGlobal
CEK Building
C./ Rosselló 153, 1st floor
08036 BARCELONA
Catalonia, Spain
Tel.: +34 93 2275400 ext. 4276
alfred.cortes@isglobal.org

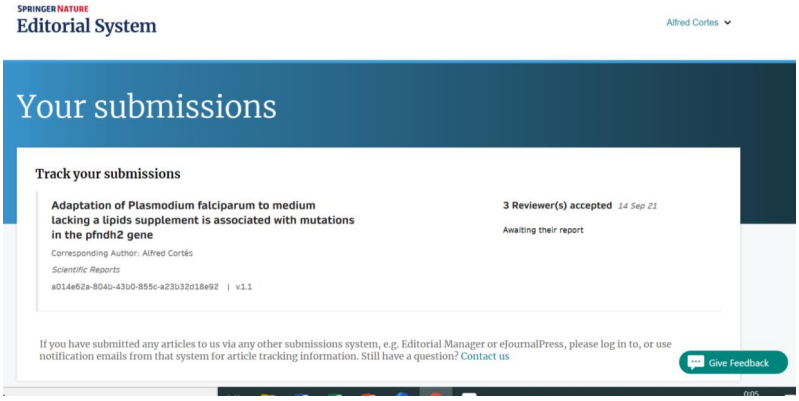
21 September 2021

Report of the Thesis supervisor: admission of the papers included in Anastasia K. Pickford's PhD thesis in a journal

I certify that the articles that are included in the PhD thesis of Anastasia K. Pickford have either been published (Articles 1 and 2) or are currently under review (Article 3).

As a proof of publication of Articles 1 and 2, the website of the journals can be used (https://journals.asm.org/doi/10.1128/AAC.00052-19?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%20%200pubmed and https://journals.asm.org/doi/10.1128/mBio.01636-21 for Articles 1 and 2, respectively).

Article 3 (Anastasia K. Pickford is first author) has been submitted to *Scientific Reports*. As a proof that this article is under review (Submission ID a014e62a-804b-43b0-855c-a23b32d18e92), please see the snapshot below:



Sincerely,

Firmado digitalmente por CORTES CLOSAS ALFRED - 46549133P
Fecha: 2021.09.22 00:09:29 +02'00'

Alfred Cortés Closas

ICREA Research Professor at ISGlobal

Global summary

The results of this thesis include two published articles, one article under review and two collections of unpublished results. The latter have not been submitted for publication because they are either incomplete or inconclusive but they have been included in this thesis as they form part of the body of work destined to address the main objectives of this dissertation and contribute to the overall discussion and conclusions. In the following chapter, the main findings of each article / result section will be summarised.

Article 1

The first article of this thesis is entitled "Identification of antimalarial compounds that require CLAG3 for their uptake by *Plasmodium falciparum*-infected erythrocytes" and it was published in *Antimicrobial Agents and Chemotherapy* in April 2019. In this article we set out to further characterise the role of *clag3* genes in the formation of a functional *Plasmodium* surface anion channel (PSAC) by determining if other antimalarial compounds, in addition to the previously described blasticidin S (BSD) and leupeptin, require *clag3* expression for their uptake and, therefore, if they are susceptible to the drug resistance mechanism regulated at the epigenetic level involving these genes.

Firstly, we validated an immunofluorescence assay (IFA) with which to detect the presence or absence of CLAG3 proteins using an anti-CLAG3 antibody that recognised both CLAG3.1 and CLAG3.2[390]. We used this assay to visualise how selection of parasites with increasing concentrations of BSD results in the disappearance of CLAG3-expressing parasites. In this manner, we were able to demonstrate that the 10G-0.6-2 parasite line, that was later used in other experiments, lacked CLAG3 proteins.

Secondly, we selected a series of drugs to test whether they needed CLAG3 proteins for their transport into the iRBC. These drugs were either predicted to require a channel for their uptake into the iRBC[449] or there was previous evidence for their transport via new permeation pathways (NPPs) that include the PSAC[326,327,450–453]. Next, we performed a growth assay in which we compared the growth of the 10G-0.6-2 parasite line with that of the control 10G line in the presence of the different selected drugs. The main difference between these two parasite lines is the presence or absence of CLAG3 proteins due to their epigenetic silencing in the 10G-0.6-2 line. Therefore,

those drugs that only affected the growth of the control 10G line required CLAG3 genes for their uptake. Out of all the tested drugs, we only found significant growth differences between the two parasite lines with the bis-thiazolium salts T3 and T16, in addition to the controls leupeptin and BSD. These results demonstrate that there are other compounds that require a CLAG3-dependent PSAC for their uptake.

Thirdly, to determine the effect of these drugs on *clag3* expression patterns, we selected 10G parasites with sub-curative concentrations of T3 and T16, as well as with a few of the other drugs previously demonstrated to require NPPs for their transport. After selecting parasites for approximately three weeks and analysing *clag3* expression patterns via RT-qPCR, we found that only T3 and T16-selected parasites presented a significant decrease in overall *clag3* expression levels. Therefore, parasites that expressed neither *clag3.1* nor *clag3.2* had been selected. None of the analysed drugs, except the control BSD, selected for parasites expressing a particular *clag3* gene. Using the aforementioned IFA and a sorbitol resistance assay, we demonstrated that the low *clag3* expression levels in the T3-selected parasites corresponded with an absence of CLAG3 protein. Furthermore, when we removed the drug, we found that *clag3*-expressing parasites were progressively selected and the initial *clag3* expression patterns were recovered. All in all, this represents another case of *clag3*-mediated drug resistance regulated at the epigenetic level and further demonstrates the adaptive role of these CVGs.

Finally, as there were drugs that were not affected by the absence of CLAG3 proteins in our growth assay despite presenting evidence of requiring NPPs for their uptake, we performed experiments to determine if PSAC-mediated transport could occur in the absence of CLAG3 expression. For this, we used the reporter compound 5-aminolevulinic acid (5-ALA) that requires a functional PSAC for its uptake[394]. Inside the iRBC, 5-ALA is processed to the fluorescent compound protoporphyrin IX that is detectable via fluorescent microscopy[454]. We studied the uptake of 5-ALA in the 10G-0.6-2 and T3-selected parasite lines and found that despite the absence of CLAG3, this compound was still able to enter the iRBCs. Consequently, with these experiments we show that there is both CLAG3-dependent and CLAG3-independent PSAC activity which mediates the permeability of the iRBC to different compounds.

Article 2

The second article of this thesis is entitled "Expression patterns of *Plasmodium falciparum* clonally variant genes at the onset of a blood infection in malaria-naïve humans" and it was published in *mBio* in August 2021. The main objectives of this paper were to determine how malaria parasites use their CVGs at the onset of a blood infection and further explore the hypothesis that these genes undergo an epigenetic reset during passage through transmission stages. Furthermore, we also identified the transcriptional differences between same-genome parasites maintained exclusively in vitro or exposed to the conditions of the human circulation.

Firstly, we began by performing the first detailed genome-wide transcriptomic characterisation of the complete intraerythrocytic development cycle (IDC) of *P. falciparum* parasites at the onset of a blood infection in malaria-naïve human volunteers. We used parasites obtained from four different volunteers who participated in a controlled human malaria infection (CHMI) trial and compared them with the parental NF54 line maintained exclusively under culture conditions. We found that the main transcriptional differences between the parasites obtained from the infected volunteers (vNF54) and the parental line (pNF54) were in CVGs, which were predominantly expressed at higher levels in the vNF54 lines and belonged to the following gene families: *pfmc-2tm*, *rif*, *var*, *stevor*, *phist*, *clag*, *gbp*, *hyp10* and *lysophospholipase*. In contrast, *clag3.1* and *var2csa* were strongly downregulated in the vNF54 lines. In general, the transcriptional profile of the four different vNF54 lines was highly similar.

Secondly, we analysed the transcriptional changes in each CVG family in further detail as they presented different scenarios. For the *pfmc-2tm* family, essentially all its members were expressed at much higher levels in the vNF54 parasite lines than in the pNF54 line. In the case of mutually-exclusively expressed *var* genes, the parental parasite line homogenously expressed *var2csa* while the vNF54 lines expressed multiple different *var* genes at intermediate levels, corresponding with different individuals of these parasite populations expressing different *var* genes of mainly type B. In the *rif* family, both vNF54 and pNF54 parasites expressed the same predominant genes but in the vNF54 lines, additional type A RIFINs were expressed and no changes were observed in type B RIFINs. In both *stevor* and *phist* families the same predominantly expressed genes were also shared by both vNF54 and pNF54 parasites but a few additional specific genes were upregulated in vNF54 lines. In the three-member *gbp* family, both *gbph* and *gbph2* were highly

upregulated in vNF54 parasites, with *gbph2* constituting the most differentially expressed gene in our dataset. In the *clag* family, parasites went from expressing predominantly *clag3.1* in the pNF54 line to expressing both *clag3.1* and *clag3.2* in the vNF54 lines. No changes were observed in *clag2*. Finally, no other CVG families presented any common major transcriptional changes with only a few individual genes like one *hyp10* and one *lysophospholipase* showing distinct upregulation in vNF54 lines.

Thirdly, we discarded that the observed transcriptional patterns were due to genetic changes, as they could not be explained by any of the mutations detected in the whole genome sequencing of the pNF54 parasites and two of the vNF54 lines. However, we did find that one of the vNF54 lines had various large subtelomeric deletions that were present in the entire parasite population and that the majority of its differential SNPs also showed a 100% prevalence in this line. Of note, the experimental conditions with which this particular volunteer was infected make it plausible that only one sporozoite reached the liver, which would explain the homogenous mutations in the blood-stage parasite population.

Fourthly, we performed H3K9me3 ChIP-seq in order to study the distribution of heterochromatin throughout the pNF54 and vNF54 genomes, as CVGs are regulated at the epigenetic level and these were the most differentially expressed genes in our dataset. Indeed, the genes with the largest transcriptional differences from various different CVG families showed a corresponding change in heterochromatin distribution (higher expression entailed absence of heterochromatin and lower expression presence of heterochromatin). In the case of the larger CVG families, many transcriptional changes were not accompanied by the expected changes in heterochromatin. In order to address this disparity, we analysed subclones of one vNF54 line as an approximation to single-cell analysis. We found that in the case of *pfmc-2tm*, *stevor* and *rif* gene families, each analysed gene was highly expressed in either one or a few of the subclones and silenced in the remaining majority. This goes to show that different individual parasites express different members of these families resulting in detectable transcriptional changes that are not in detectable variations in heterochromatin levels at the populational level. All in all, the changes in CVG expression correlated with transitions between euchromatic and heterochromatic states.

Fifthly, we aimed to establish whether the observed changes in CVG expression patterns were due to an epigenetic reset or whether they were

a consequence of within-host selection. In order to do so, we compared by RT-qPCR the expression of various different genes between volunteer samples collected after an estimated single cycle of intraerythrocytic asexual replication post-liver egress and samples collected after an additional one to three cycles. We observed that the transcriptional levels of these genes showed no significant differences between the two sample collection times. We also re-analysed another CHMI dataset from Milne *et al.*, in which *var* and *rif* gene expression was compared between samples collected before and after 3 to 5 cycles of development in the blood and no differences were found. Consequently, we were able to conclude that at the onset of a human-blood infection, an epigenetic reset and not selection by the human host is the main driver of the changes in CVG expression and that not only the previously described mutually-exclusively expressed *var* and *clag3* genes undergo this reset.

Lastly, we compared between the pNF54 and a vNF54 lines, various phenotypic traits associated with CVG expression such as membrane deformability, solute permeability and sexual conversion. We did not find any significant phenotypic differences between the two lines.

Article 3

The third article of this thesis is entitled "Adaptation of *Plasmodium falciparum* to growth in medium lacking a lipids supplement is associated with mutations in the *pfndh2* gene" and it is currently under revision at the journal *Scientific reports* (September 2021). In this article our main objective was to establish whether *P. falciparum* parasites were able to adapt to fluctuations in the availability of lipids by the means of a CVG-mediated bet-hedging adaptive strategy.

In order to achieve this objective, we first performed a total of three selection experiments in which 3D7-A *P. falciparum* cultures were grown in a culture medium lacking an external source of lipids for a period of time that varied between 24 and 60 days. This lack of lipids was achieved by substituting the standard culture lipid source AlbuMAX II with lipid-free bovine serum albumin (BSA). In all three selections, we found that parasite growth was severely compromised from the 3rd cycle onwards until the 8th-10th cycle when growth rates recovered up to 80% of the growth rates of the control culture maintained in parallel with AlbuMAX II-supplemented media. We then performed a series of experiments to validate the adaptation of the

parasites selected without an external source of lipids. These experiments consisted in comparing the growth of control (3D7-A-Ctrl) and selected (3D7-A-LF) parasites between standard or lipid-free medium. Overall, the 3D7-A-LF parasites showed better growth in the lipid-free medium than the 3D7-A-Ctrl parasites and we took this as validation that the 3D7-A-LF parasites were adapted to grow without an external source of lipids.

Next, in order to understand the mechanisms behind the aforementioned adaptation, we performed a transcriptome-wide microarray analysis on the 3D7-A-Ctrl and 3D7-A-LF parasite lines from two different selections. At the beginning of this experiment, the 3D7-A-LF parasites were divided into two conditions: LF0 (always maintained in lipid-free medium) and LFCM (switched from lipid-free to standard medium at the beginning of this experiment) in order to distinguish between selected transcriptional patterns and directed transcriptional responses derived from the exposure to the absence of an external lipid source. We found there was a large overlap of the genes differentially expressed in the LF0 and LFCM conditions with respect to the control. This indicated that the observed transcriptional differences were principally due to stable changes after the different selection periods. On the contrary, we saw very little overlap in the transcriptional differences of each selection. The main differentially expressed genes were CVGs, though the exact CVGs in question differed between the two selections, indicating that this was an indirect effect of the population passing through a bottleneck rather than a transcriptional change mediating adaptation. All in all, the results of the microarray analysis showed that adaptation to grow without an external source of lipids involved different transcriptional alterations in independent selection experiments, none of which could be clearly linked to lipid metabolism.

Finally, we performed whole genome sequencing on 3D7-A-Ctrl and 3D7-A-LF parasites from all three selections to see if any genomic mutations could explain the adaptation observed in the 3D7-A-LF parasites. This analysis revealed that the 3D7-A-LF parasites from the second selection presented a large deletion including the PF3D7_0914900 (BSD-domain protein) and PF3D7_0915000 (type II NADH:ubiquinone oxidoreductase [*pfndh2*]) genes. Furthermore, the 3D7-A-LF parasites from the first and third selections carried an insertion of an adenine at position 898 of the coding region of *pfndh2* that resulted in a frameshift leading to a premature stop codon and a truncated protein of approximately half its original size. All in all, our full genome sequencing analysis revealed that in the three independent selection experiments, adaptation to lipid deprivation was associated with the loss of a

functional PfNDH2, via either a deletion or frameshift mutation. This reveals a link between lipid metabolism and mitochondrial electron transport in asexual blood stages that will require further research in order to understand the exact mechanism behind this association.

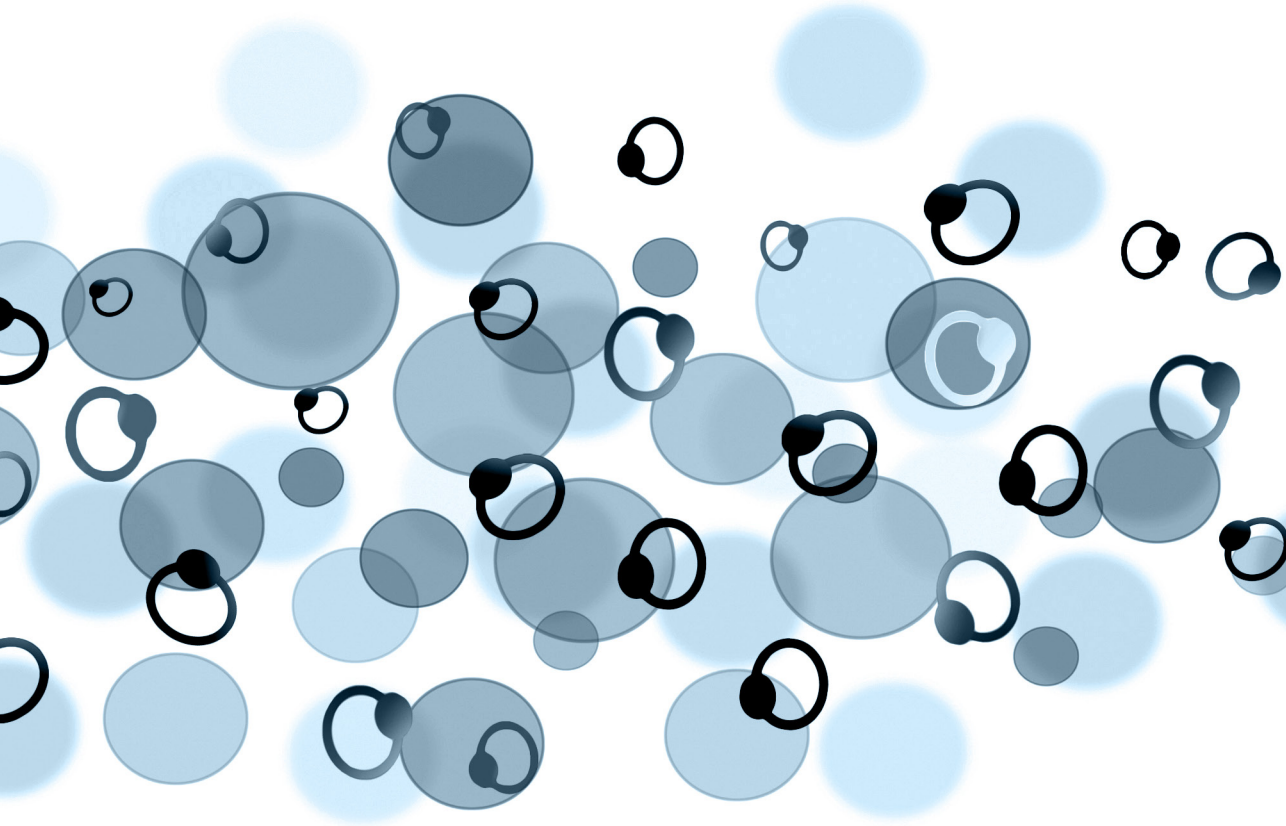
Other results 1

We performed a series of experiments with the objective of determining if *P. falciparum* parasites utilise CVG-mediated bet-hedging adaptive strategies to adapt to fluctuations in the availability of glucose. In a similar manner to that of Article 3, we attempted to adapt parasites to grow in media with sub-optimal glucose concentrations. However, we found that parasites were unable to stably adapt to this condition and therefore this line of research was abandoned.

Other results 2

These results form part of our ongoing project involving the generation of a triple transgenic parasite line using CRISPR-Cas9 technology. The objectives of this project are to study *clag3* expression dynamics at the single-cell level within a parasite population and to measure parameters such as the relative fitness of each *clag3* expression pattern under different conditions, *clag3* switching rates and the spontaneous presence and frequency of different *clag3* expression patterns in the parasite population, among others. In order to do so, in this transgenic line, *clag3.1*, *clag3.2* and a control gene are each tagged with a different fluorescent protein and flow cytometry will be used to detect their fluorescence. We have successfully designed and performed the tagging of each gene separately, however we have yet to generate a fully-functional triple-transgenic line with all three simultaneous editions. One of our two main problems is associated with residual episomal plasmid containing the Cas9 gene and the selectable marker. Presence of this episome in our transgenic lines leads to resistance to the drugs we use as selective markers and therefore complicates the performance of sequential transfections. The other main problem is the integration of multiple copies of the donor plasmid into the *clag3.2* locus that hinders the natural expression of these genes. We are currently working to overcome these challenges and hopefully we will reach our objective of creating this triple transgenic tool and use it to study *clag3* expression dynamics under standard and different environmental conditions. Nonetheless, this will not be achieved before the termination of this thesis.

DISCUSSION



In this thesis, we set out to shed further light on how epigenetic variation mediated by CVG expression enables *P. falciparum* to adapt to the fluctuating conditions of the human host circulation. In order to do so, we searched for new compounds that were susceptible to the drug resistance mechanism mediated by *clag3* genes, we designed a tool with which to study the expression dynamics of *clag3* genes under standard and selective conditions, we characterised the expression patterns of CVGs at the onset of a blood infection using samples from a CHMI trial and we searched for further fluctuating conditions to which malaria parasites can adapt to by the means of CVG-mediated bet-hedging adaptive strategies. Taking into account all the results that we have generated in this thesis, in the following discussion we will analyse what we now know about the role of CVGs in the adaptation of malaria parasites to the fluctuating conditions of the human host circulation as well as the epigenetic mechanisms that regulate their expression patterns throughout these adaptive processes. Also, we will discuss the other adaptive mechanism we have uncovered which enables malaria parasites to adapt to fluctuating conditions of the human host circulation but that is not mediated by changes in CVG expression. Next, we will examine how our results contribute to what is known about how the expression patterns of *clag3* genes affect the permeability of the iRBC membrane given their role in the formation of the *plasmodium* surface anion channel (PSAC). Finally, we will weigh up the differences between in vivo and in vitro studies in addition to analysing the translatability between the two settings with regards to the study of CVG expression.

1. Role of CVGs in the adaptation of Plasmodium falciparum to fluctuating conditions in the human host circulation

As described in the introduction of this thesis, malaria parasites find themselves in many different environments throughout their complex life cycle and have to adapt not only to changes between these different surroundings, but also to the variations within each one. This dissertation is focused on the environment constituted by the human host circulation which malaria parasites encounter briefly during the sporozoite stage, throughout the repeated cycles of asexual replication within human RBCs, during sexual conversion and as mature gametocytes. As we have seen, the fluctuating conditions of the human bloodstream can exert important selective pressures on the parasite population[186,455]. In this first discussion chapter, we will analyse in the light of our results, how CVGs can mediate bet-hedging adaptive strategies that enable parasite survival within this environment. We will distinguish between the adaptive role of the epigenetic reset that occurs during transmission stages and affects CVG expression patterns at the onset of a blood infection, as well as the role of stochastic switching in CVG expression during the course of a blood infection. Also, we will specifically analyse the adaptive role of *clag3* genes.

1.1. Adaptive role of the epigenetic reset of CVG expression patterns at the onset of a blood infection in a new human host

In Article 2 of this thesis, we found that numerous CVGs show evidence of undergoing an epigenetic reset during transmission stages (i.e., mosquito and liver stages). This reset consists of the loss of the CVG expression patterns selected during the infection in the previous human host and the establishment of "fresh expression patterns" before the onset of a new blood infection in a new human host. These "fresh expression patterns" can be understood as increased transcript levels of various CVGs, as well as a higher diversity among individual parasites in the expression profile of various CVG families due to the activation in some individual parasites of genes that were previously silenced. This epigenetic reset of CVG expression patterns can be considered a bet-hedging adaptive strategy for three main reasons.

Firstly, one of the key characteristics of a bet-hedging strategy is for the population involved to present phenotypic diversity that precedes changes

in the environment[456]. We observed that at the onset of a blood infection there is an increase in the transcriptional diversity of CVGs, with different individual parasites presenting different CVG expression patterns. Many CVGs are predicted to have adaptive roles[291], therefore this transcriptional diversity could likely translate into different parasites presenting different phenotypes with different adaptive potential. However, we did not find evidence of this phenotypic heterogeneity with the exploratory assays we performed on parasites collected at the onset of a blood infection, although this can probably be explained by two factors. On the one hand, we did not perform an exhaustive phenotypic characterisation that included well-established phenotypes associated with changes in CVG expression (e.g., cytoadherence and *var* genes[183,323,324], drug resistance and *clag3* genes[325–327] or invasion pathways and *eba* and *pfrh* genes[315]). On the other, we only performed bulk population assays that may have not been sensitive enough to detect alternative phenotypes of small subsets of parasites. Therefore, by analysing the aforementioned phenotypes in parasite populations obtained at the onset of a blood infection or in transcriptionally-homogenous subclones of these populations, it may be possible to confirm this phenotypic diversity corresponding to the transcriptional heterogeneity of CVGs after they have been epigenetically reset.

Secondly, another important attribute of bet hedging is for the population to be exposed to an unpredictable environment that albeit fluctuates infrequently[137,144,146–149]. After their egress from the liver, malaria parasites are suddenly exposed to the bloodstream environment of a new human host that presents fluctuating conditions such as the immune response[171], the host's nutritional[186] or physiological state (i.e., pregnancy, comorbidities…), variations in red blood cell receptors[199] or other characteristics[202] and the presence of antimalarial drugs due to preventive or direct treatment of the disease[455]. These conditions can vary over time within the same human host as well as be different from those of the previous human hosts the parasite population has infected. However, it is important to note that some of these conditions only fluctuate infrequently as otherwise this could represent a severe challenge for host homeostasis.

Finally, it is important to note that bet-hedging strategies lend themselves particularly well to large populations with exponential replication rates, in which the focus is not on the survival of the individuals but rather on the long-term survival of the population over numerous generations[151]. This is the case of malaria parasites as thousands of

merozoites per invading sporozoite are liberated from the liver into the bloodstream[457], where they face many rounds of asexual replication within human RBCs in order to establish an infection, increment their biomass and ensure the transmission of sexual forms to the invertebrate mosquito host.

In summary, at the onset of a blood infection in a new human host, *Plasmodium* spp. have to be prepared to face the conditions that can fluctuate unpredictably in this new environment. By diversifying their CVG expression patterns beforehand, parasites exit the liver as a more phenotypically heterogeneous population. This increases their chances of surviving the different selective pressures active in the human bloodstream as those individual parasites that stochastically have the most beneficial phenotype are selected, enabling the survival of the population (**Fig. 66**).

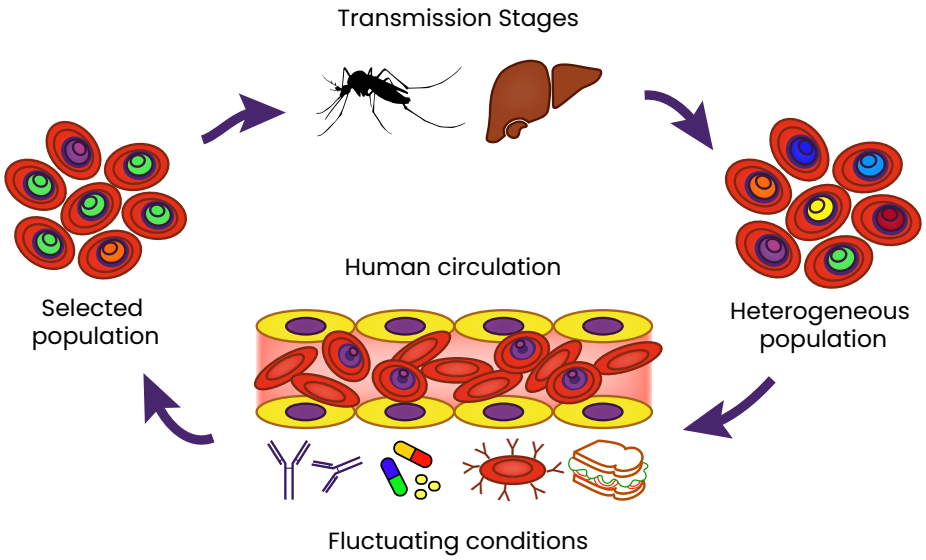


Fig. 66: Epigenetic reset of CVGs as a bet-hedging adaptive strategy. The expression patterns of CVGs are epigenetically reset during transmission stages. Subsequently, at the onset of a blood infection, the malaria parasite population is more transcriptionally and phenotypically diverse (represented by different colour parasites). The fluctuating selective pressures of the human circulation can act upon this pre-existing variability and select those parasites with the fittest phenotypes.

1.2. Adaptive role of CVG expression during the course of a blood infection

After analysing how CVG expression patterns vary at the onset of a blood infection, the next step is to evaluate how they change over the course of a blood infection and find further examples of how they mediate bet-hedging adaptive strategies in this environment. Although in this thesis we have not obtained any results that provide further evidence on this subject, given that our main objective has been to further characterise the adaptive role of CVGs within the fluctuating environment of the human host circulation, we believe it is important to discuss what is known and what further experiments could help shed more light on this topic.

At present, there is very little information available on how CVGs mediate adaptation of malaria parasites to the different selective pressures they can encounter over the course of a natural blood infection. On the one hand, it is well-established that stochastic switches in *var* expression mediate immune evasion[183] and, on the other, it is known that field isolates express predominantly *clag3.2*[335], although it has yet to be demonstrated if this is related to solute uptake in vivo. Nonetheless, there is currently only one study that we are aware of which has attempted to analyse CVG expression patterns during a long-term infection[458]. In this longitudinal study that monitored chronically-infected, asymptomatic children over a period of 4 months, *var* gene expression was found to change rapidly or persist for up to 10 weeks depending on the gene. This revealed a complex scenario in which the expression patterns of these genes are influenced by their differences in switching rates, the selective pressures of the immune response and splenic clearance, the presence of mono- or polyclonal infections and variations in the prior exposure to malaria parasites. Therefore, even though these various confounding factors inherent in natural infections make it difficult to establish exactly how CVG expression patterns are shaped by the exposure to the selective pressures of the human host, it does appear that dynamic selection of parasites with different CVG expression patterns actually occurs in vivo.

Next, there are a few CHMI trials which have also provided some insight into this topic even though they only show what happens during the initial phases of the disease given that it is unethical to maintain human volunteers with untreated infections for long periods of time[459,460]. One of these studies is that of Milne *et al*.[172], which set out to determine if group A and DC8 *var* variants, that mediate more efficient cytoadherence and

are associated with severe disease[461], expand more rapidly in non-immune hosts than other *var* variants, given the fitness advantage associated with their resistance to splenic clearance. However, the analysis of the samples from this CHMI revealed that there was no indication of any *var* or *rif* genes being selected in parasites exposed to host conditions during approximately the first 2 to 3 weeks of a blood infection. This could be because these *var* variants did not present a sufficiently high fitness advantage for them to be selected or because the selective pressure of the host's immune response or splenic clearance may take longer than expected to shape the parasite population in malaria-naïve volunteers. On the contrary, in another CHMI trial that analysed immune volunteers with prior exposure to malaria parasites, it was found that the antibodies naturally acquired before the study appeared to preferentially protect against group A and DC8 *var* genes[462]. Consequently, this prior exposure resulted in other *var* genes becoming the predominantly expressed members in these volunteers, demonstrating how, in this case, the immune response can shape CVG expression patterns over the course of an infection[462].

Finally, experiments with the rodent malaria species *P. chabaudi* have also demonstrated how the selective pressures of the host circulation can influence CVG expression patterns. This is based on the observation that serial blood-passaged parasites that express a subset of CVGs produce a more severe infection than parasites transmitted via mosquito, which express a broader repertoire of CVGs as they have been epigenetically reset during transmission stages[328,330]. The difference in disease severity between these two transmission routes can be explained by the fact that serially passaged parasites are constantly exposed to host circulation conditions and the CVG expression patterns associated with a higher fitness are under constant selection. Subsequently, when these parasites are directly inoculated into a new host, they already present "fit" CVG expression patterns and are able to expand more rapidly and produce more severe disease than if these patterns had been epigenetically reset and required de novo selection (**Fig. 67**). With regards to human malaria species, the CHMI trails that have compared these two transmission routes did not observe these differences in disease severity[463,464]. Nevertheless, this can be explained by the fact that the blood-stage parasites used for these studies were not repeatedly passaged between volunteers and therefore had only been exposed to human host circulation conditions for no more than a few cycles. A more accurate comparison with the rodent malaria observations could be made with data obtained from the malariotherapy treatments performed in the first half of the 20th century[465]. In these

therapies, neurosyphilis patients were inoculated with malaria parasites with the objective of inducing periodic fevers to eliminate the bacteria responsible for this disease. To this end, malaria parasites were both serially passaged between patients as well as transmitted via mosquito. However, although much data was recorded from these therapies which has contributed greatly to malariology, there are no formal analyses on whether the serially passaged parasites produced more severe disease as observed in the aforementioned *P. chabaudi* studies.

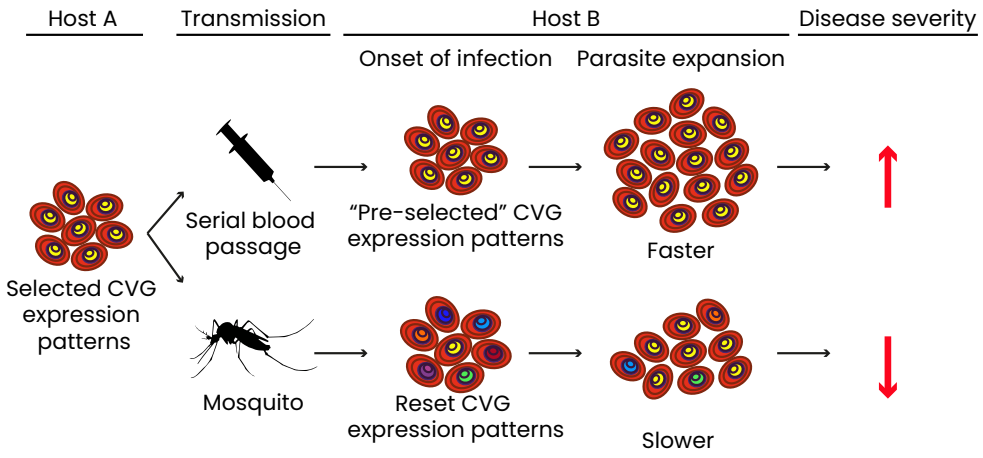


Fig. 67: Association between the route of transmission and the severity of a malaria infection. More details in the text. Each colour represents a different phenotype associated with a specific CVG expression pattern. Yellow parasites are those whose CVG expression patterns are associated with the fittest phenotype.

All in all, these studies provide different pieces of evidence of how the fluctuating conditions of the host circulation can shape CVG expression patterns and how these genes can mediate bet-hedging adaptive strategies. Nevertheless, clearer in vivo research is required to determine exactly how CVG expression patterns are selected by the pressures of the human host circulation. An ideal experimental set up would be to perform a long-term CHMI trial in malaria-naïve volunteers and periodically take samples in order to analyse how parasite CVG expression varies over time. Furthermore, by simultaneously monitoring volunteer circulation parameters such as antimalarial antibodies or standard haematological parameters, it may be possible to determine the selective pressures shaping these expression patterns. However, as mentioned, this type of

study is unethical[459,460]. Another alternative could be to perform a more controlled study of natural infections using asymptomatic, untreated individuals with monoclonal infections. Albeit, this type of cohort is not easy to gather. Furthermore, both of these experimental approaches present a series of technical challenges that derive from the use of in vivo samples that will be described later on in this discussion and which further complicate the characterisation of CVG expression patterns over the course of a blood infection. In summary, in vitro experiments such as the ones performed during this thesis are currently the most feasible and informative approach with which to identify conditions that can affect CVG expression patterns and then hypothesise how these conditions can be translated to the in vivo setting.

1.3. Adaptive role of *clag3* genes in relation to the fluctuating environment of the human host circulation

The proteins encoded by *clag3* genes are involved in the formation of the PSAC channel[390], a broad selectivity channel that enables the uptake of various types of solutes across the iRBC membrane[396,466]. Genetic mutations and changes at the epigenetic level that affect the expression patterns of *clag3* genes have been found to modify the solute selectivity of the PSAC towards antimalarial compounds[325–327] and nutrients[389,401]. As a result, *clag3* genes have the potential to mediate key adaptation processes in vivo with regards to fluctuations in antimalarial drug pressures and nutrient availability as we will proceed to describe.

1.3.1. Role of *clag3* genes in the adaptation of *P. falciparum* to fluctuating concentrations of antimalarial compounds

One of the key examples of a bet-hedging adaptive strategy mediated by CVGs is that of *clag3* genes and their involvement in a drug resistance mechanism regulated at the epigenetic level[291]. This mechanism consists of the selection of parasites with *clag3* expression patterns that render parasites resistant to different antimalarial compounds. Prior to our results, only BSD had been identified as being susceptible to this resistance mechanism, as low concentrations of this drug select *clag3.1*-expressing parasites[326], while high concentrations select parasites with both *clag3* genes silenced[326,327]. However, in Article 1 of this thesis, we have identified two additional compounds – T3 and T16 – that are also susceptible to this drug

resistance mechanism and that select parasites with both *clag3* genes silenced. So far, all of these results have served to further consolidate the adaptive role of *clag3* genes in vitro but currently there is not any evidence of this drug resistance mechanism occurring in the in vivo setting. That said, the characteristics of this resistance mechanism suggest that it could be clinically relevant and facilitate the adaptation of *P. falciparum* parasites to fluctuations in antimalarial drug concentrations in the human host circulation, as we will proceed to discuss.

To begin, unlike BSD that is not a candidate for the clinical treatment of malaria infections, T16 and especially T3 have been considered potential antimalarial compounds, with the latter, otherwise known as albitiazolium, having reached phase II clinical trials[467]. Consequently, the fact that potential clinically-relevant drugs are susceptible to this resistance mechanism make it noteworthy of being considered in the in vivo setting. That said, it is also important to mention that none of the antimalarial drugs routinely used to treat malaria are susceptible to this mechanism.

Next, *clag3*-mediated resistance arises when parasites are exposed to subcurative concentrations of the aforementioned susceptible compounds[325,326]. This is a situation that can easily occur throughout the course of an antimalarial treatment, as antimalarial drug concentrations can fluctuate in the host's plasma due to reasons such as the natural decline associated to degradation and excretion of the drug from the body, incomplete dosing, incorrect patient compliance or even substandard medicine quality in developing countries and tropical environments[5,73,208–211].

Moreover, in vitro we observe that when continuously exposed to subcurative concentrations of BSD, T3 or T16, parasite populations take 2 to 3 weeks to show signs of stable adaptation (almost complete recovery of pre-drug exposure growth rates). This reflects the time it takes for the selected parasites with both *clag3* genes silenced to expand within the population and therefore, depends on the initial proportion of parasites with this expression pattern within said population. It is unknown if parasites with both *clag3* genes silenced exist or are indeed viable under in vivo conditions. However if they were to be present in similar proportions as in vitro, it is feasible for the aforementioned timeframe to occur as malaria infections have in some cases been reported to last over a year[468] and relapse of malaria infections due to factors such as drug resistance or infection with multiple *Plasmodium* spp. can entail treatment durations of various weeks[73].

Last but not least, given that this resistance mechanism is regulated at the epigenetic level, one of its principal characteristics is that it is reversible. Since CLAG3 proteins are predicted to be essential when parasites are grown in media resembling human plasma[401], it is highly probable that the double silencing of both *clag3* genes comes with an associated fitness cost such as arrested growth or lower multiplication rates which have been observed in vitro[326]. Therefore, as this mechanism is regulated at the epigenetic level, once the drug pressure disappears this fitness cost is rapidly lost and *clag3*-expressing parasites can expand once more and lead to a relapse of the infection. As a result, under fluctuating drug concentrations, this reversible epigenetic mechanism is more advantageous than a more irreversible genetic one (**Fig. 68**). Also, given the epigenetic nature of this mechanism, the diagnosis of disease relapse and consequent treatment failure would be much harder in the clinical setting. This is because while genetic mutations can be relatively easy to detect by sequencing parasite genomic DNA[469,470], the accurate analysis of transcriptional changes requires RNA samples of synchronised parasite populations which are a lot more complicated to obtain in vivo[471]. Consequently, the detection and monitoring of resistant parasites by epigenetic mechanisms would be much harder in the field.

All in all, given the complexity and cost of developing new antimalarial drugs[472] and taking into account the arguments discussed above, the *clag3*-mediated resistance mechanism could be considered worthy of being monitored during in vitro preclinical trials and/or throughout in vivo clinical trials. In this manner, even if a drug is described to be effective at high concentrations, if it is found to be susceptible to this resistance mechanism the necessary measures can be taken in order to avoid future treatment failure (e.g., prescription with another non-susceptible drug or adjustment of the formulation or dosage to avoid long exposure of parasites to subcurative drug concentrations). Consequently, our growth assay with the 10G-0.6-2 parasite line could be used as a screening tool with which to detect drugs susceptible to this resistance mechanism.

1.3.2. Role of *clag3* genes in the adaptation of *Plasmodium falciparum* to fluctuations in the availability of nutrients

Uptake of nutrients has also been linked to *clag3* expression because when parasites are grown in media with physiological concentrations of certain nutrients (isoleucine, glutamine and hypoxanthine)[389], it has been observed, on the one hand, that PSAC inhibitor efficacy increases[389] and,

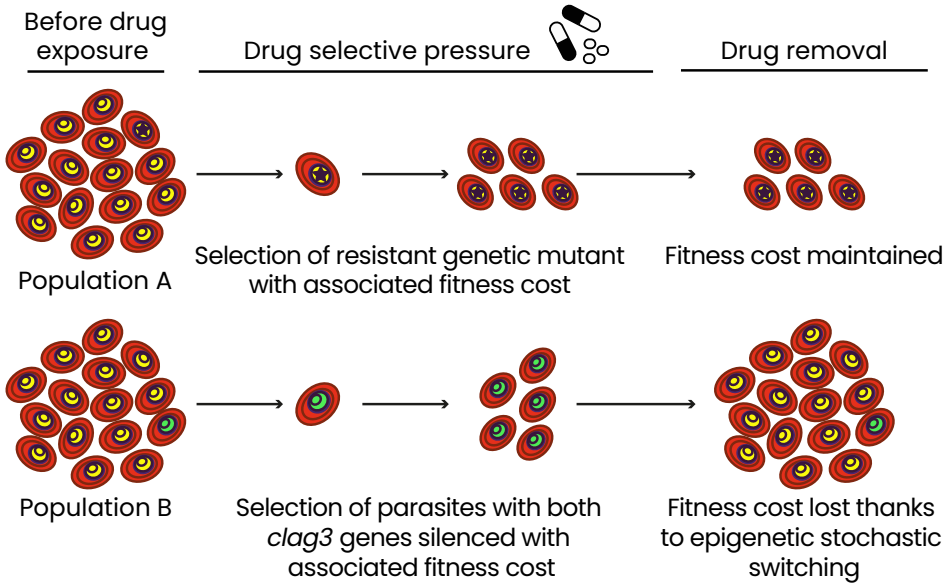


Fig. 68: Advantage of a reversible drug resistance mechanism mediated at the epigenetic level vs. a non-reversible mechanism mediated at the genetic level under fluctuating antimalarial drug concentrations. When population A is exposed to an antimalarial drug, those parasites that present a genetic mutation that renders them resistant are selected. When the drug pressure disappears, given the irreversibility of genetic mutations, the fitness cost associated to the selected mutant persists. On the other hand, in population B, resistance to the antimalarial drug is achieved through the selection of parasites with both *clag3* genes silenced. As this is regulated at the epigenetic level, upon drug removal, the fitness cost associated with this expression pattern is lost as stochastic switching enables parasites that express *clag3* to expand once more in the population.

on the other, that parasites lacking CLAG3 proteins are unable to grow[401]. Subsequently, it has been suggested that *clag3* genes play an important role in the acquisition of nutrients in vivo where nutrient availability is much lower than in the standard RPMI-based culture medium used in vitro[258,473]. Nevertheless, there is currently no evidence of *clag3* expression being associated with changes in iRBC permeability to a specific type of nutrient.

In this thesis, we performed two types of experiments in which *clag3* expression patterns were analysed in relation to the type and availability of lipids in the culture media. The first consisted in comparing the *clag3* transcripts of parasites grown in media with either AlbuMAX II or human

serum as the main source of lipids in the culture medium (see other results 2, chapter 5.2). The rationale behind this was that as field parasite isolates have been observed to express predominantly *clag3.2*[335], we wanted to determine whether growth in human serum could select for parasites expressing this paralog. The second experiment consisted in analysing the full transcriptome of parasites adapted to grow without an external source of lipids (Article 3 of this thesis). As *clag3* genes were included in this genome-wide analysis we were able to determine whether their expression patterns were affected by these culture conditions. However, in neither of these experiments were there consistent or significant changes detected in *clag3* expression. As a result, it appears that there is no direct relationship between lipid uptake and *clag3*-mediated membrane permeability. Although further experiments would be needed to fully discard this, it is an expected result because fatty acids are not included among the type of solutes predicted to be transported via the PSAC such as amino acids, sugars, purines, some vitamins, organic cations and halide anions[390,466].

All things considered, it is possible that *clag3* genes or even the other clonally variant member of the clag family, *clag2*[322], may mediate the uptake of other specific nutrients or compounds that fluctuate in human blood and that their bet-hedging adaptive role may go beyond drug resistance. Our yet-to-be-obtained triple transgenic *clag3* parasite line could prove to be a useful tool with which to study changes in *clag3* expression patterns in response to different environmental pressures and hopefully will be able to shed further light on this subject. Finally, the adaptive potential of *clag3* genes and the other members of the *clag* family with regards to determining the solute selectivity of the PSAC will be further analysed in chapter 4 of this discussion.

2. Epigenetic mechanisms behind CVG-mediated adaptation

The bet-hedging adaptive strategies mediated by CVGs that have been mentioned in this thesis involve one or both of the following epigenetic mechanisms: stochastic switching between different transcriptional states and the epigenetic reset of CVG expression patterns during transmission stages. In this next discussion chapter, we will analyse what each of these mechanisms consist of, their relevance in CVG-mediated adaptation, the novelties our results provide and what fundamental research questions remain to be addressed.

2.1. Stochastic switching

The basis of any bet-hedging adaptive strategy involving CVGs are the low-frequency stochastic switches that occur between the active and silent transcriptional states of these genes. These switches result in the generation of transcriptional heterogeneity within the parasite population that translates into phenotypic diversity[293]. Therefore, they serve to increase the phenotypic diversity in the parasite population both before the action of different selective pressures, as well as to reintroduce it after a specific phenotype has been selected.

The switch in the transcriptional state of a CVG consists basically of the transition of bistable chromatin from active H3K9ac-marked euchromatin to silent H3K9me3- and HP1-marked heterochromatin or vice versa[268,291,294–302]. The essential nature of these switches is their stochasticity, derived from the observation that individual parasites kept under identical conditions, such as controlled in vitro culture, can randomly present these events in different CVGs or not[293]. The exact mechanisms behind CVG stochastic switching remain unknown and their low frequency hinders their experimental analysis[293]. In other organisms such as plants, stochastic epimutations regarding DNA methylation have been highly studied[474,475] and, therefore, can offer inspiration as to how CVG stochastic switching is regulated in malaria parasites. In plants, these epimutations involve mechanisms such as imperfect replication, mistakes in maintenance, localised off-targeting of DNA methyltransferases or transcriptional interference, among others. As changes in the transcriptional state of CVGs imply histone modifications, this has led Llorà-Batlle *et al*. to propose two potential mechanisms that could regulate this stochastic switching. On

the one hand, changes in the expression levels of epigenetic regulators in certain individual cells could randomly induce switches between different transcriptional states. On the other, errors in the transmission of epigenetic memory during asexual replication in the IDC could be responsible for switches in CVG expression patterns[293]. Nevertheless, in both cases experimental evidence is lacking and more research into the mechanisms behind the stochastic switches in CVG transcriptional states is required. With our triple transgenic *clag3* parasite line we wish to characterise stochastic switching in *clag3* genes and hopefully we will be able to shed more light on this process.

2.2. Epigenetic reset of CVGs

The other main epigenetic mechanism that we have found to play a fundamental role in the adaptation of *P. falciparum* to the fluctuating conditions of the human host circulation and that is mediated by CVGs, is the epigenetic reset of CVG expression patterns during transmission stages. This reset is comprised of two phases: the erasure of the epigenetic markers determining the CVG expression patterns that were selected during the previous human infection, followed by the de novo establishment of new CVG expression patterns in order to prepare parasites for the infection of the next human host. In the following sections we will continue to analyse what this epigenetic reset entails by examining how different CVGs are affected by this reset, which possible mechanisms are involved in each of its two phases and at what point of the parasite's life cycle these take place.

Before we begin, it is important to mention that the ideal experimental set-up with which to address these questions would be to follow variations in the transcriptome and epigenome of genetically and transcriptionally homogenous parasite lines (e.g., recently obtained subclones) over the course of their entire life cycle. However, this has yet to be performed in *Plasmodium* research. Subsequently, in this discussion we will take into account the transcriptomic and epigenomic datasets of those malaria stages that have been analysed[266,476–478], even though they cannot be directly compared as they look at different chromatin markers in different parasites strains and species. Furthermore, they also present the limitation of having been performed at the populational level on transcriptionally diverse parasite lines which can conceal subtle changes that occur in subsets of parasites. Nevertheless, these data alongside information on

similar events in other organisms can help to understand more about the epigenetic reset of CVGs.

2.2.1. CVGs involved in the epigenetic reset

In our analysis of CVG expression at the onset of a blood infection, one of our main novel findings is that other CVGs, in addition to the previously described *var*[332–334] and *clag3*[335] genes, undergo an epigenetic reset during transmission stages. These other CVGs include the *pfmc-2tm*, type A *rif*, *stevor*, *phist* and *gbp* families, as well as single members from the *hyp10* and *lysophospholipase* family. A summary of the transcriptional differences that we found each family to present at the onset of a blood infection in a new human host are summarised in **Table 5**.

With regards to the other CVGs in which we did not find evidence of an epigenetic reset, there are two possibilities that could explain their expression patterns at the onset of a blood infection: 1) these particular CVGs did not undergo an epigenetic reset or 2) they did undergo an epigenetic reset but their expression patterns after being lost and established de novo were very similar to those prior parasite passage through transmission stages. We believe this second option to be more plausible because the expression patterns of all CVGs depend on the presence or absence of facultative heterochromatin, as determined by specific epigenetic markers[268,291,294–302]. The only main difference between CVGs with regards to the regulation of their expression, is that *var* and *clag3* genes present mutually-exclusive expression[306,314,315,383]. However, as we found that other non-mutually exclusive CVGs were also reset, this property cannot explain why some CVGs present evidence of being epigenetically reset while others do not. Thus, it would be mechanistically simpler if all CVGs were to undergo an epigenetic reset rather than for there to be distinctions between different CVGs that are silenced by the same molecular mechanism.

To further support the argument that all CVGs are epigenetically reset during transmission stages, there were no clear patterns with regards to function and localization with which to distinguish CVGS that do or do not show evidence of being reset. All known CVGs functions are related to host-parasite interactions[291]. Therefore, this was the case of genes for which we did find evidence of an epigenetic reset (*var* genes and cytoadherence and antigenic variation[50,183,314,324], type A *rif* genes and rosetting[324,349], *clag3* genes

Gene family or gene	Expression patterns post-reset	
	Populational level	Individual parasites
pfmc-2tm	↑ **expression levels + ↑ diversity:** Activation of almost the entire family. Each gene is expressed in a higher number of parasites than in the pre-reset population, although it remains silent in the majority of the post-reset population.	Each parasite expresses 1/+ *pfmc-2tm* genes.
var	≈ **expression levels + ↑ diversity:** Post-reset population is composed of different individual parasites expressing different *var* genes that were not expressed in the pre-reset population.	Each parasite expresses a different *var* gene due to mutually-exclusive expression[306,314,479].
rif	≈ **expression levels + ↑ diversity:** Post-reset population expresses more type A *rif* genes but overall presents the same predominantly expressed *rif* genes as the pre-reset population. Each gene is silenced in the majority of the post-reset population.	Each parasite expresses 1/+ *rif* genes, with a preference for certain *rif* genes.
stevor	**≈ expression levels + ≈ diversity:** Post-reset population presents the same predominantly expressed *stevor* genes + the activation of some genes that were silenced in the pre-reset population. Each gene is silenced in the majority of the post-reset population.	Each parasite expresses 1/+ *stevor* genes, with a preference for certain *stevor* genes.
phist	**≈ expression levels + ≈ diversity:** Post-reset population presents the same predominantly expressed *phist* genes + the activation of some genes that were silenced in the pre-reset population. Each gene is silenced in the majority of the post-reset population.	Each parasite expresses 1/+ *phist* genes, with a preference for certain *phist* genes.
gbp	**Activation of *gbph2* and *gbph*** in the majority of the post-reset population. *gbp130* expression is the same as in the pre-reset population.	Majority of parasites express *gbph2* and *gbph*.
clag3	**Activation of *clag3.2*** in a large proportion of the post-reset population with consequently **less parasites expressing *clag3.1***.	Each parasite expresses one *clag3* due to mutually-exclusive expression[315,383].
hyp10 (PF3D7_0114500)	Activation of PF3D7_0114500 in some parasites of post-reset population.	Some parasites express PF3D7_0114500.
lysophospholipase (PF3D7_0936700)	Activation of PF3D7_0936700 in some parasites of post-reset population.	Some parasites express PF3D7_0936700.

Table 5: Populational and individual parasite CVG expression patterns at the onset of a human blood infection after undergoing an epigenetic reset during transmission stages.

and solute uptake[389,390] or *stevor* genes and regulation of iRBC membrane deformability[37,180,358,480]) as well as those for which we did not (*eba* and *pfrh* genes and erythrocyte invasion[404] or *6-cys* genes and immune evasion[420,421]). With regards to location, the proteins encoded by the epigenetically reset *var*[50,183,314,324], type A *rif*[324,349] and some members of the *stevor* family[356,357,481] are considered surface antigens and are situated on the exterior of the iRBC, the same as *surf* [482,483] and *6-cys* genes[420] which did not show signs of being epigenetically reset. Similarly, the epigenetically reset *pfmc-2tm* proteins[376] and members of the *phist* family[484–486] have more intracellular location, as do members of the *acs* and *acbp*[416] and *fikk* genes[411,413] families which did not show signs of being reset.

If all CVGs are epigenetically reset, the next question that arises is why do some CVGs show the same expression patterns as they did before the reset while others do not? This may be due to the fact that the underlying DNA sequence of some genes can influence their transcriptional state, in what is known as "facilitated epigenetic variation" and in which "the genotype directly potentiates the epigenotype in a probabilistic but not strictly deterministic manner"[487]. For example, it has been described that certain DNA sequences can impact nucleosome positioning[488] and hence potentially affect the interaction with transcription factors or histone modifying enzymes. Hence, something similar may occur in some CVGs, with their sequence facilitating or not the formation of facultative heterochromatin and, therefore, altering the probability of these genes being active or silent when their expression patterns are established de novo during the epigenetic reset. This probabilistic model could explain both why some CVG families did not show signs of being epigenetically reset after passage through transmission stages and why in some families such as *rif*, *phist* and *stevor*, the predominantly expressed members in the parasite population were the same as before the reset (**Fig. 69**).

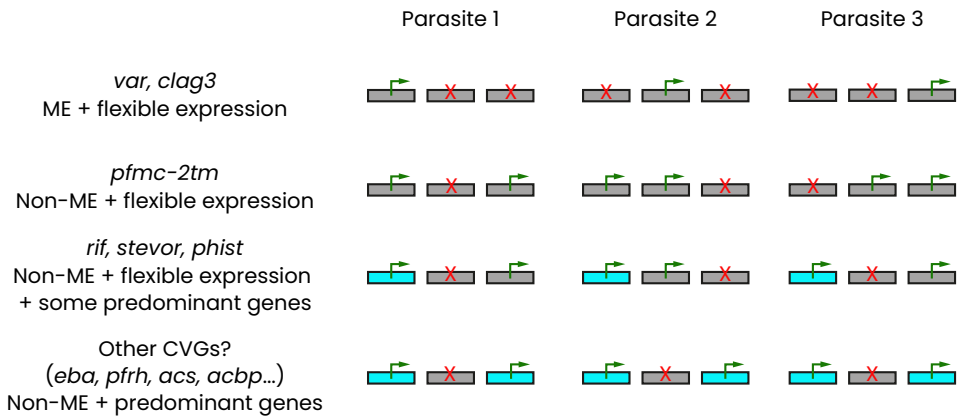


Fig. 69: Model to explain the differences observed in the de novo establishment of CVG expression patterns during transmission stages. Classification of CVGs based on whether they present mutually-exclusive (ME) expression and on the flexibility of their expression patterns depending on whether the same predominant genes are expressed pre and post transmission. Three genes (rectangles) are represented for each type of CVG in three different individual parasites. Green arrows indicate a transcriptionally active state and red crosses a silent state. By comparing the three parasites, it is possible to distinguish between those CVGs with more flexible expression patterns and those that may be more highly influenced by the underlying DNA sequence and consequently show preference for the active state in predominantly expressed genes (highlighted in blue).

2.2.2. How CVG expression patterns are reset

To understand how CVG expression patterns are reset in malaria parasites, it is possible to look at similar epigenetic reprogramming events in other organisms. During early embryonic development in mammals, epigenetic reprogramming occurs once before fertilization entailing profound changes in germline cells; and again directly after fertilization during early zygotic development ensuring the transition from a highly specific cell-type to a state of totipotency[489,490]. Broadly speaking, these reprogramming events are achieved through changes in nucleosome structure, deposition of histone variants, variations in histone marks and DNA demethylation throughout the entire genome[490–492]. Similarly, in dedifferentiation studies in plants, involving germline reprogramming and protoplast formation, extensive chromatin remodelling with changes in H3K9me and release of HP1 have been described[493,494].

Given that the expression patterns of CVGs are mediated by the distribution of facultative heterochromatin, which in itself is determined by specific histone modifications and stabilizing protein HP1[267,268,291,295–302], it is possible to hypothesise that an epigenetic reset of CVGs may consist of similar mechanisms to those mentioned above. In this manner, the initial erasure of previous expression patterns could be obtained through events such as a general retraction or expansion of heterochromatin (or both) in different chromosomal locations, followed by the establishment of new expression patterns through the recovery or elimination of heterochromatin in a more or less probabilistic manner depending on the underlying DNA sequence. Various ChIP-seq studies have analysed heterochromatin distribution in different parasite stages and found that its distribution, maintenance and inheritance is relatively stable throughout the life cycle except for a few subtle changes[266,476–478]. The comparison of asexual blood stages with either gametocytes[266] or sporozoites[476,478], has revealed that between these stages heterochromatin principally expands from subtelomeric to central chromosomal regions covering genes which encode exported and erythrocyte remodelling proteins. However, it also retracts in some regions which include genes that are expressed in these stages (e.g., *pfap2-g* in gametocytes)[266]. Whether this expansion is maintained from gametocytes to sporozoites or established de novo in between these stages is unknown. The only study that looks at the changes in CVG expression patterns between asexual blood stages, gametocytes and ookinetes was performed in *P. berghei* and did not find that there was a heterochromatin expansion in any of the analysed stages[477]. However, the authors of this research justify that this may be because unlike *P. falciparum*, *P. berghei* malaria parasites lack central chromosome islands of CVGs[266] and that the genes present in these islands in *P. falciparum* do not have known orthologues in this rodent species.

Overall, even though changes in heterochromatin distribution have been described between different stages of the malaria parasite life cycle, none appear to be major enough to justify a general reset of CVG expression patterns. Nevertheless, this may be because all of these datasets are from experiments carried out at the populational level in transcriptionally diverse parasite lines. Subsequently, changes of heterochromatin in genes that are activated in small subsets of parasites are hidden by the general heterochromatin pattern of the majority of the parasite population in which these genes are silenced (**Fig. 70**). This is what we found in our dataset where, for example, transcriptional changes in *pfmc-2tm*, *rif* and *stevor* genes were not reflected in changes in heterochromatin at the populational

level. However, transcriptional analysis of subclones obtained from the same parasite line revealed that the parasite population was composed of individual parasites expressing different members of these gene families and that each gene remained silent in the majority of the population. As a result, the changes in heterochromatin in the subset of the population in which each individual gene was activated were covered up by the general silenced pattern of the population. For future studies, we propose to directly study heterochromatin distribution in recently subcloned, transcriptionally-homogenous parasite lines across asexual and transmission stages in order to assess epigenomic changes without the confounding factor of populational diversity.

Finally, there are still no ChIP-seq data for other key stages such as early mosquito stages (especially gametes and zygotes) or liver stages in which major changes in heterochromatin might be observed. Moreover, it is also possible that other histone marks and epigenetic modifications, different from those analysed to date, may also play an important role in CVG resetting. Therefore, these are all key issues that need to be addressed in order to further investigate how this epigenetic mechanism occurs.

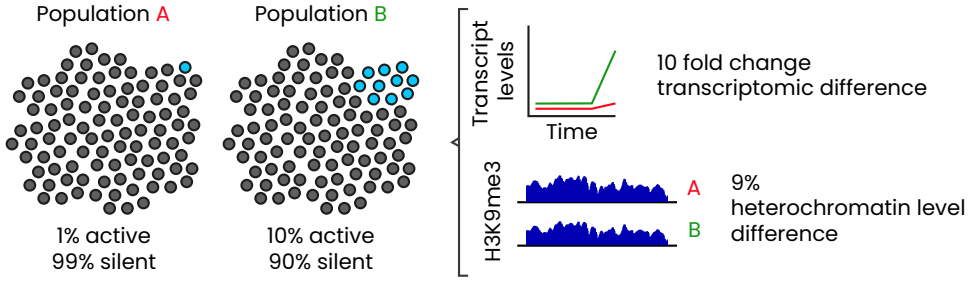


Fig. 70: Example of the confounding effect of analysing data at the populational level. If a gene passes from being active in 1% of population A to being active in 10% of population B, there is a transcriptional fold change of 10 which is easily detectable in transcriptomic analysis. However, when looking at ChIP-seq data, the 9% reduction in overall heterochromatin levels from 99% in population A to 90% in population B is a lot harder to detect. In this manner, the distribution of heterochromatin in CVGs in small subsets of parasites can easily be masked by that of the general population.

2.2.3. When CVG expression patterns are reset

The three main candidates to host the epigenetic reset, in part or in its entirety, are gametocytes, mosquito stages and liver stages. In support of gametocytes is the aforementioned finding that there are significant changes in the distribution of heterochromatin in these stages which are very similar to those later observed in sporozoite stages[266,476–478]. Thus, the mechanisms involved in the epigenetic reset could possibly commence in these sexual forms.

Next, the reason why mosquito stages are also likely candidates to host the epigenetic reset of CVGs because in other organisms epigenetic reprogramming events occur during fertilization[490] and the most similar event in the malaria parasite's life cycle is gamete fusion within the mosquito midgut[495]. However, there are significant differences between these two processes as in mammals fertilization is preceded by meiosis and the objective of epigenetic reprogramming during embryonic development is to establish cellular totipotency so that an initial cell produces multiple cell types to generate a diverse and multicellular organism[490]. On the contrary, in *Plasmodium* spp., meiosis occurs after fertilization[496] and although sexual recombination offers the opportunity to introduce diversity into the parasite's genome[497], this unicellular organism has sequential life stage progression that does not require the generation of multiple cell types from a single cell. Nevertheless, gamete fusion could be a key moment for epigenetic remodelling to occur in malaria parasites.

Finally, part of the epigenetic reset could also take place in liver stages given that in our results we observed that a single sporozoite was able to generate as much transcriptional variation in CVGs after passing through the liver as multiple sporozoites. This indicates that the establishment of new CVG expression patterns does not occur before this point in the parasite's life cycle. Very little is known about *P. falciparum* liver stages due to the technical difficulties associated with their study, although new technologies are helping to overcome this limitation[498,499]. What is certain is that upon infecting hepatocytes, sporozoites undergo severe morphological changes that enable them to transform from motile, invasive forms into replication-competent schizonts capable of utilising host resources in order to generate thousands of merozoites[500]. This metamorphosis may include other significant changes that could affect the heterochromatin distribution regulating CVG expression.

Notably, there is evidence of CVG expression in all of the aforementioned stages[37,349,354,476,478,501–503], although whether the proteins encoded by these CVGs play important functional roles in all of these stages is unclear. Furthermore, it is important to remember that just because a CVG is not expressed at a specific stage, it does not mean that it is epigenetically silenced. In *Plasmodium* spp., gene expression is tightly regulated predominantly at the transcriptional level by a transcription factor cascade. Subsequently, each parasite stage is characterized by a specific gene expression programme[216–218]. Therefore, until it is "time" for a particular CVG to be expressed it "does not matter" if it is epigenetically poised to do so or not[292]. As a result, general changes in the heterochromatin distribution of CVGs could occur at early mosquito stages and not become apparent until the onset of a blood infection when asexual intraerythrocytic parasites start to express the genes affected by these changes. Meanwhile, the CVGs expressed during transmission stages could still undergo the necessary epigenetic changes accompanied by changes in gene expression necessary for their expression at that specific point in the life cycle.

In conclusion, from our results and taking into account the information discussed above, we propose a model (**Fig. 71**) in which the first phase of the epigenetic reset of CVGs commences in sexual or mosquito stages based on the fact that significant changes in heterochromatin distribution have been observed in both gametocytes and sporozoites compared to asexual stages[266,476–478]. We hypothesise that this phase would entail the erasure of the epigenetic marks that determine the distribution of heterochromatin in these genes, resulting in a complete loss of the expression patterns selected in the previous human host. Next, when parasites reach the liver of a new human host, the second phase of this reset would commence and these epigenetic marks would be established de novo resulting in the heterochromatin distribution and corresponding expression patterns observed at the onset of a blood infection. In order to confirm this model, ideally it would be necessary to perform RNA-seq and ChIP-seq analysis of samples of asexual intraerythrocytic, gametocyte, mosquito and liver stages of recent subclones that were genetically and transcriptionally homogenous. As previously mentioned, in this way it would be possible to follow the expression patterns and the heterochromatin patterns of CVGs throughout the entire life cycle and avoid the confounding effect of analysing transcriptionally diverse parasite populations.

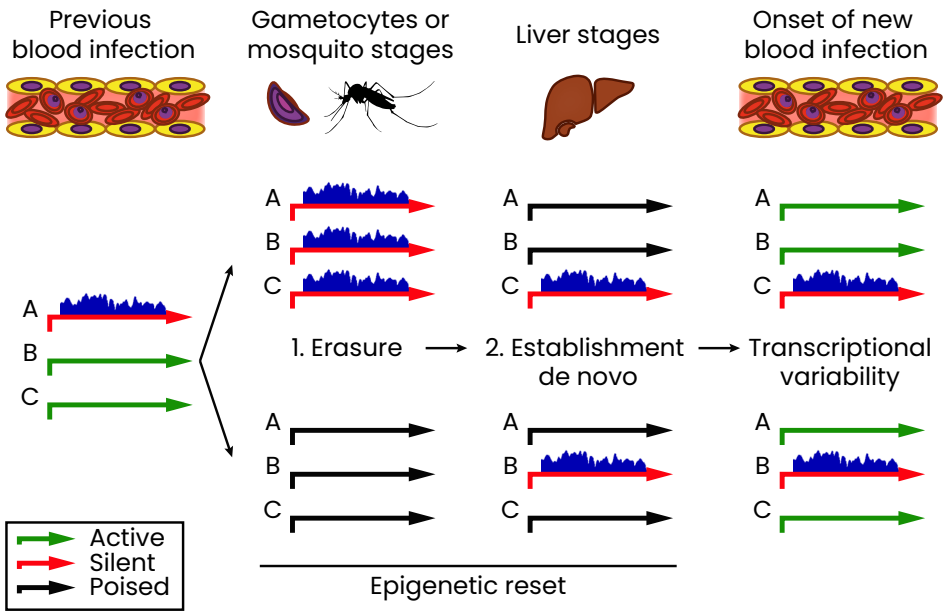


Fig. 71: Model for the epigenetic reset of CVGs during transmission stages. Shown is a hypothetical CVG family conformed of genes A, B and C, which are either active or silent depending on the absence or presence of heterochromatin. During gametocyte or mosquito stages, the expression patterns from the previous blood infection are erased by means of either a general expansion (top) or loss (bottom) of heterochromatin markers. Next, during liver stages, expression patterns are epigenetically established de novo by the loss (top) or gain (bottom) of these heterochromatin markers. In this manner, before being released into the bloodstream, genes A, B and C are either silent or "poised" at the epigenetic level and this expression pattern can vary between individual parasites. After egress from the liver, the epigenetically poised genes will become active, resulting in a transcriptionally heterogenous parasite population at the onset of a blood infection.

3. Other mechanisms that mediate adaptation of *Plasmodium falciparum* to fluctuating human host conditions

One of the objectives of this thesis was to identify further examples of CVG-mediated bet-hedging adaptive processes. While we did manage to do this with our experiments involving the selection of *clag3* expression patterns with the compounds T3 and T16 (described in Article 1 of this thesis), this was not the case with the two other adaptation experiments we performed involving fluctuations in nutrient availability. On the one hand, we found that parasites were did not show signs of stable adaptation when grown in media with sub-optimal concentrations of glucose due to the essentiality of this nutrient[422]. However, on the other, parasites did show signs of stable adaptation when cultured in media lacking an external source of lipids although this did not entail a bet-hedging strategy mediated by CVGs. In this section, we will discuss our findings on this new example of malaria parasite adaptation and reflect on how these pathogens possess an important arsenal of adaptive mechanisms with which to confront the fluctuating environment of the human bloodstream.

3.1. Adaptation to fluctuations in lipid availability

To begin, the rationale behind the hypothesis that CVGs could be involved in the adaptation of malaria parasites to fluctuations in lipid availability was based on the fact that there are CVGs involved in lipid metabolism such as members of the *acyl-CoA synthetase* (*acs*), *acyl-CoA binding protein* (*acbp*) and *lysophospholipase* families[322,414–416]. Furthermore, despite possessing the metabolic pathways necessary for lipid synthesis such as the FASII pathway in the apicoplast, malaria parasites mainly scavenge the lipids they need for membrane biogenesis directly from the host and therefore are susceptible to variations in the availability of this family of nutrients[415,504–507]. In this manner, as malaria and malnutrition often coexist in the field, parasites are most likely exposed to this selective pressure in their natural environment[192].

We put this hypothesis to the test in Article 3 of this thesis and found that the 3D7-A malaria parasite line was able to adapt to grow without an external source of lipids in the culture medium. By substituting AlbuMAX II with lipid-free bovine serum albumin (BSA), the only source of fatty acids these parasites had access to were those present in the cytoplasmic

membrane of the erythrocytes they infected[504,508]. This is surprising as there is no other in vitro evidence of malaria parasites being able to grow without supplementation with at least some specific fatty acids[509–511]. However, this has only predominantly been analysed over short periods of time (very few cycles) and the only study that did analyse the effect of growth in a medium with a combination of fatty acids that do not support parasite development for 28 days, did not detect any signs of adaptation[509]. We propose that this discrepancy with our results could derive from the fact that these studies used other parasite lines that did not present the mutation we later found to be associated to adaptation to lipid deprivation.

After performing full transcriptomic and genomic analysis, the only common factor between the adapted parasites from three independent selections were mutations (either a large deletion or a single nucleotide deletion) in the type II NADH:ubiquinone oxidoreductase or *pfndh2* (PF3D7_0915000) that resulted in an absent or truncated version of this component of the mitochondrial electron transport chain (METC)[512,513]. We can only speculate as to how a non-functional PfNDH2 enables adaptation to lipid starvation, as in asexual intraerythrocytic stages there is no obvious relation between lipid metabolism and the METC. This is because during these stages, the METC does not sustain an active mitochondrial tricarboxylic acids cycle (TCA) (unlike in other organisms or in malaria mosquito stages), as glucose degradation by glycolysis in the cytoplasm is the main source of energy[197,514]. Therefore, during the parasite's development in the blood, the sole function of the METC is to maintain the continuous reoxidation of ubiquinol to ubiquinone in order to sustain dihydroorotate dehydrogenase (DHOD) activity that is essential for de novo pyrimidine synthesis[515]. The other components of the METC can transfer electrons from different substrates to ubiquinone, therefore this role of PfNDH2 is redundant and explains why it is non-essential in asexual blood stages[512,513] and how our selected parasites can survive without it. The only distinguishing characteristic of PfNDH2 is that the substrate it reduces is NADH resulting in generation of NAD+[512,513]. As a result, it is possible that the loss of a functional PfNDH2 may alter the balance between NADH and NAD+ and indirectly affect the metabolic fluxes in the mitochondria and other cellular compartments, possibly resulting in a beneficial effect for those parasites that lack an external source of lipids.

Various PfNDH2 mutations (including our single nucleotide deletion) have been detected in field isolates[248,516], the majority in the same large homopolymeric track of adenines, albeit at low frequency, low confidence

and with all isolates presenting both mutant and wild-type alleles. However, if this poly-A track presents hypermutability it is possible for the adaptive mechanism mediated by PfNDH2 to occur in vivo despite this protein being essential for mosquito stages[512,513]. In this manner unpredictable fluctuations in lipid availability between or within different hosts could select parasites with mutations that result in a non-functional PfNDH2 and enable parasite survival under these conditions. However, in the same parasite population, the constant generation of new variants expressing the complete protein would secure transmission. That said, it is possible that this mechanism is more efficient in vitro and may mask others that only operate when the *pfndh2* mutation is not viable.

In addition to the mutations in PfNDH2, it is possible that other elements which we did not detect in our analyses also played a role in this adaptation mechanism, such as subtle transcriptional changes that were beyond the sensitivity of our microarray assay or changes at the metabolic or posttranscriptional level which we did not have time to analyse. One study found that when grown in a medium containing only two essential fatty acids, malaria parasites activated the FASII pathway in the apicoplast increasing the de novo synthesis of lipids[507]. In this case, the rapid activation of this pathway suggested that it was mediated by changes at the posttranscriptional level[507]. In our results, the inheritable adaptation pattern we observe takes approximately 8 cycles and is not consistent with a rapid posttranscriptional activation of a metabolic pathway alone. That said, it is possible that this pathway, which is normally only functional in sporozoites and liver stages[517–519], is activated in our adapted parasites and in combination with a non-functional PfNDH2 helps compensate lipid deprivation by increasing the de novo synthesis of fatty acids. Furthermore, as the FASII pathway requires reduced NADH (or NADPH)[415,505], we speculate that the absence of PfNDH2 that recycles NADH to NAD+, may promote this pathway and enable parasite survival under these conditions (**Fig. 72**). However, it is unclear if and how changes produced by a mitochondrial enzyme can translate into changes in the cytoplasm or the apicoplast, although given the metabolic plasticity of *Plasmodium* spp. this is entirely plausible[197].

In summary, this adaptive mechanism is yet another example of how malaria parasites can adapt to the fluctuating environment of the human host. In this case, adaptation occurred independently of changes in CVG expression. Although we ignore whether in absence of the *pfndh2* mutations changes in CVG expression could have had a more significant role, it seems

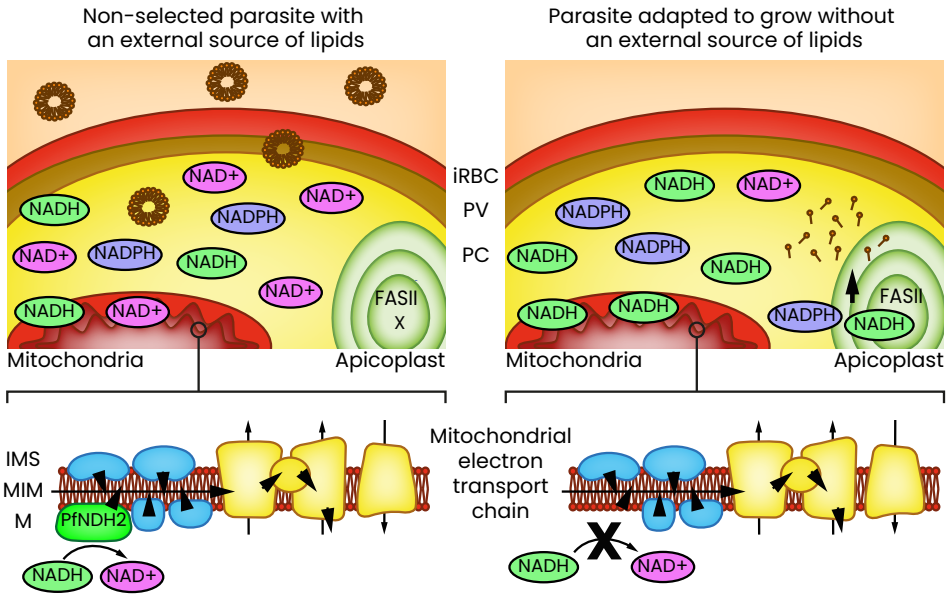


Fig. 72: Hypothetical model for the adaptation of malaria parasites to grow without an external source of lipids in the culture media. On the left, a parasite under normal culture conditions that presents a complete METC, an unaltered NADH/ NAD+ balance and an inactive FASII pathway in the apicoplast. On the right, a parasite adapted to grow without an external source of lipids that presents a METC lacking a functional PfNDH2 resulting in the disruption of the NADH/NAD+ balance in favour of NADH. Consequently, there is an increment of NADH/ NADPH in the cytoplasm/apicoplast which contributes to the de novo synthesis of fatty acids in the apicoplast where the FASII pathway has been stimulated by the deprivation of lipids. iRBC = infected red blood cell membrane, PV = parasitophorous vacuole membrane, PC = parasite cytoplasm, IMS = intermembrane space, MIM = mitochondrial internal membrane, M = mitochondrial matrix.

unlikely as the only changes we detected were inconsistent between the different adapted parasite lines and were most likely the by-product of the selection bottleneck. Nevertheless, this new mechanism could still be considered a bet-hedging adaptive strategy, albeit mediated at the genetic level, as the *pfndh2* poly-A track hypermutability could result in phenotypic diversity that precedes environmental fluctuations. Although it can be argued that most genetic mutations come before an environmental stress, in this case this hypermutability may facilitate the adaptation to a condition that fluctuates unpredictably and the selected phenotypes could be lost once the conditions change. On the contrary, most genetic

mutations are largely irreversible and mediate a longer lasting phenotypic change that plays a role in species evolution rather than in adaptive bet-hedging.

Lastly, it would be interesting to further explore the adaptive plasticity of *P. falciparum* by subjecting these parasites to other in vitro "simulations" of host circulation fluctuating conditions such as the ones we performed in this thesis with regards to glucose and lipid availability. In this manner, perhaps it would be possible to identify further examples of bet-hedging adaptive strategies mediated by CVG expression. Examples of conditions that could be tested include availability of other nutrients, different antimalarial drugs, oxidative stress or the use of red blood cells with genetic defects that have previously been demonstrated to be protective against malaria such as thalassemia and sickle cell disease[205,206].

4. Role of *clag3* genes in the formation and function of the PSAC

It is well-established that malaria parasites increase the permeability of the erythrocytes they infect by means of new permeation pathways (NPPs)[397] and that the PSAC (*Plasmodium* surface anion channel) is the main channel responsible for this increased iRBC permeability[396]. However, there is still much to be learnt about this channel with regards to its structure, functionality and its relation with *clag3* genes. In this chapter we will analyse how the results described in Article 1 of this thesis have contributed to provide further evidence on the involvement of CLAG3 genes in the formation of the PSAC and the adaptive potential that this offers to malaria parasites.

4.1. Characteristics of the compounds that require CLAG3-dependent PSAC activity for their uptake

The PSAC is a unique broad-selectivity channel that enables the uptake of a wide variety of solutes[390,520]. As it is currently the only identified parasite-induced channel, it constitutes an important alternative for those compounds that cannot cross the iRBC membrane via lipid-based diffusion[449]. A study that used a computational method that takes into account drug polar surface area, molecular weight and lipophilicity[521], predicted that only a low number of compounds require uptake via the PSAC, as the majority of antimalarials can easily diffuse through lipid membranes as they are small and hydrophobic[449]. However, although in silico predictions are useful, only experimental work can conclusively determine which compounds require CLAG3s for their uptake. For this reason, in Article 1 of this thesis we set out to identify further antimalarial compounds whose transport via the PSAC could be affected by changes in *clag3* expression and this also offered us the opportunity to analyse if these compounds had any common properties. A summary of the characteristics of the drugs we tested and the reason behind their inclusion in our study can be found in Table 6. However, in the end, in addition to the previously described BSD[326,327] and leupeptin[451,522], only T3 and T16 were found to require CLAG3 proteins for their uptake. These four compounds all have a similar molecular weight and are hydrophilic. However, neither of these properties are exclusive to these compounds. Subsequently, further research is necessary to define the type of compound that requires CLAG3 proteins for their uptake. In this manner, the growth assay we used to identify these compounds could be used to identify others, facilitating the identification of common physicochemical properties.

Drug	Properties	Criteria for inclusion in our study
Azithromycin (Macrolide antibiotic)	Weak base Hydrophobic MW: 749	Predicted to require channel-mediated transport in malaria parasites[449] and other microorganisms[523].
Blasticidin S (Nucleoside antibiotic)	Hydrophilic MW: 422,4	Requires PSAC for its uptake: exposure to BSD can alter PSAC permeability[524] and select parasites with different *clag3* expression patterns[326,327].
Doxycycline (Tetracycline antibiotic)	Acid Hydrophobic MW: 444,4	Predicted to require channel-mediated transport in malaria parasites[449] and other microorganisms[525].
Fosmidomycin (Antibiotic)	Acid Hydrophilic MW: 183	Evidence for NPP transport as a closely related compound was only uptaken in iRBCs with increased permeability via NPPs[450].
Leupeptin (Cysteine and serine protease inhibitor)	Base Hydrophilic MW: 426,6	Requires PSAC for its uptake: selection of parasites with leupeptin reduces PSAC permeability[451]. In one case, resistance found to be associated to a mutation in *clag3.2*[526].
Lumefantrine (Fluorene derivative)	Base Hydrophobic MW: 528,9	Antimalarial predicted to require channel mediated transport[449].
Pentamidine (Synthetic derivative of amidine)	Weak base Hydrophilic MW: 340,4	Evidence of NPP transport: not uptaken by uninfected RBCs and its transport is blocked by the NPP inhibitor furosemide[452].
T3 Albitiazolium (Choline analogue)	Cation Hydrophilic MW: 614	Evidence of NPP transport: very low uptake by uninfected RBCs and transport is PSAC-inhibitor sensitive. Also presents a residual non-saturable uptake mechanism that remains to be clarified[453].
T16 (Bis-quaternary ammonium compound)	Cation Hydrophilic MW: 676	Evidence of NPP transport: very low uptake by uninfected RBCs and transport is furosemide sensitive[453].

Table 6: Drugs analysed to determine whether they require CLAG3-dependent PSAC activity for their uptake. Properties as described in PubChem (https://pubchem.ncbi.nlm.nih.gov). The criteria for the inclusion in this study was prior evidence of NPP transport (compounds highlighted in blue) or to be clinically relevant compounds predicted to require a channel for their uptake given their physicochemical properties[449]. MW = molecular weight (g/mol).

4.2. Model that explains CLAG3-independent PSAC activity

Many studies have implicated CLAG3 proteins in the formation of the PSAC[390,391,527]. However, despite reaffirming the importance of a CLAG3-dependent PSAC for the uptake of certain compounds, in Article 1 of this thesis we also found evidence of CLAG3-independent PSAC activity based primarily on two observations. Firstly, the transport of drugs for which there was prior evidence of NPPs uptake (fosmidomycin[450] and pentamidine[452]) was not affected by the absence of CLAG3 proteins in our growth assay. Second, our T3- and BSD-selected parasite lines that consequently presented dramatically reduced expression of both *clag3* genes, were still able to uptake the reporter compound 5-ALA, known to be transported into iRBC via the PSAC[394,454]. Furthermore, a later study performed by Gupta *et al*. found that under the supraphysiological nutrient concentrations commonly found in standard RPMI-based culture medium[258,473], parasite growth was not affected by the genetic knockout of CLAG3[401]. As a result, all these findings support that CLAG3-independet PSAC activity exists and has led to new hypothesis being made on the structure of the PSAC and the role of CLAG3 proteins in its formation, as we will proceed to discuss.

To begin, it is important to note that CLAG3s are a key component of the PSAC even though they may not be essential for this channel to function. As previously described, CLAG3s conform the RhopH complex alongside RhopH2 and RhopH3[378], two proteins that have been found to be essential for both parasite development and PSAC activity[392–394]. Nevertheless, out of the RhopH complex, only CLAG3s possess a transmembrane domain[526,528] capable of participating in the formation of a pore that can cross the iRBC membrane[395]. Consequently, it would appear that without CLAG3s, RhopH2 and RhopH3 alone would not be able to form this channel. It has been suggested that the RhopH complex may interact with other host- or parasite-derived proteins that could form the channel pore. However, to date no other channel-like parasite proteins have been found to immunoprecipitate with RhopH2[392] and although there are host proteins that have been purified alongside the RhopH complex, there are many doubts regarding their specificity and actual involvement in NPPs[529,530]. As a result, in the absence of CLAG3 proteins, it would appear that another component would be necessary to participate in PSAC formation in order for its activity to be maintained. One possibility that we and others defend is that given their similarity, the other members of the *clag* family could also form part of this channel and, when necessary, compensate the

absence of CLAG3 proteins under certain conditions[325,377,379,391,401]. However, more experimental evidence is needed to support this claim as it has yet to be directly demonstrated if these other paralogues really do play a role in solute uptake in *P. falciparum*. So far, only *clag2* has been linked to changes in iRBC permeability, as BSD has been found to select parasites that have this gene silenced[327,526] regardless of *clag3* switching[315]. That said, not much more is known about *clag2* and even less has been found out about *clag8*, as no knockouts or knockdowns have been performed to assess the essentiality of either of these two genes. In contrast, *clag9*, the most distant member of this family[391], has been found to be non-essential in vitro and been described to play roles in merozoite binding to erythrocytes[531] and cytoadherence[532], although the latter role has been questioned[391]. Nonetheless, no studies have linked this paralog to solute uptake or participation in NPPs activity.

All in all, despite the need for more experimental evidence, the observations of CLGA3-independent PSAC activity in combination with recent cryoelectron microscopy studies[395,533] have led to the following proposed model of the structure of the PSAC (**Fig. 73**). In this model, the PSAC has a higher-order stoichiometry and is conformed of various RhopH complexes (CLAG3, RhopH2 and RhopH3) in conjunction with other CLAG paralogs. While the CLAG proteins would form the channel pore thanks to their transmembrane domain[395,526,528], the two RhopH proteins would constitute intracellular components of the channel with a yet to be described function[393,528] as that is their predicted location[393,528]. The rationale behind the higher order stoichiometry of the PSAC structure, is that a single RhopH complex does not possess sufficient transmembrane domains to form a pore[534] and, even if they were sufficient, it would be necessary for the complex to undergo significant conformational changes during its insertion in the iRBC membrane in order for this pore to be formed[395]. Furthermore, the fact that CLAG3 proteins have been found to oligomerize at the iRBC surface[527] also suggest that various CLAG proteins form this channel. It is unclear how the other CLAG paralogs may integrate the PSAC, as only CLAG3 proteins have been found to be continuously associated to RhopH2 and RhopH3 throughout their trafficking from the rhoptries to the iRBC membrane[395,432]. However, as previously discussed, the other CLAGs may be able to interact with the RhopH proteins and maintain minimal PSAC activity in the case of the absence of CLAG3 genes[325,401]. All things considered, future experiments are necessary to continue deciphering the structure and activity of the PSAC, especially with regards to the characterisation of the other CLAG paralogs.

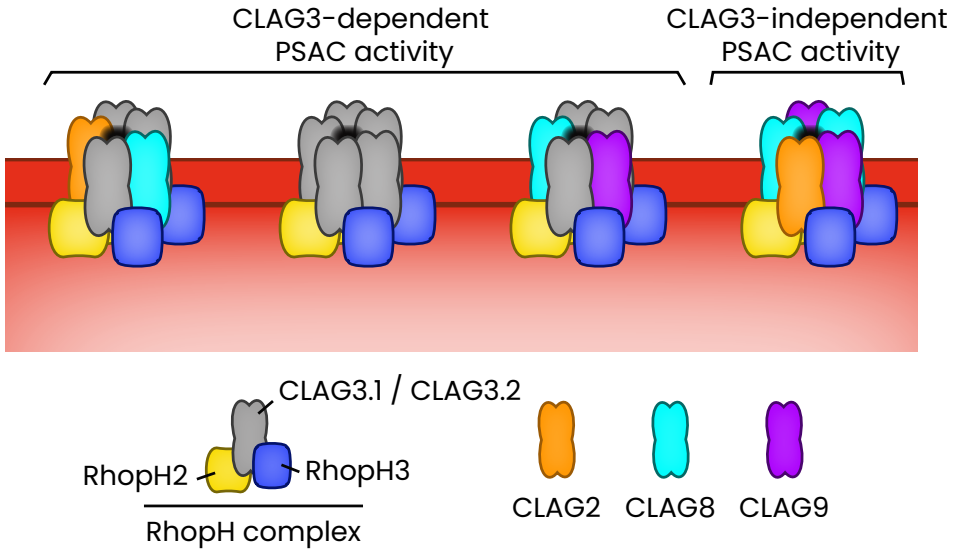


Fig. 73: Hypothetical model of the PSAC structure including other CLAG paralogs. In this model this parasite-induced channel possesses a higher stoichiometry as it is formed by various RhopH complexes (conformed of CLAG3, RhopH2 and RhopH3) possibly in combination with other CLAG paralogs. The CLAG proteins form the pore that crosses the iRBC membrane while the RhopH proteins are intracellular. When CLAG3s are not expressed or absent, the other CLAG paralogs can interact with RhopH2 and RhopH3 to maintain PSAC activity.

4.3. Adaptive potential of CLAG-mediated PSAC solute selectivity

As described in the section above, recent evidence from different laboratories, including ours, suggests that the other members of the *clag* multigene family in addition to CLAG3s may be involved in the formation of the PSAC and that this channel may consist of various CLAG proteins. Subsequently, different combinations of the CLAG paralogs could form PSACs with different solute selectivity and possibly explain how this channel has been reported to transport such a wide-range of solutes[390,520]. Furthermore, the clonally variant members of this multigene family, *clag3* and *clag2*[315,322], may show different expression patterns in different individual parasites resulting in phenotypic diversity with regards to iRBC membrane permeability. In this manner, the permeability phenotypes of parasite populations could be fine-tuned to the fluctuating environment of the human circulation, as conditions such as nutrient levels or presence

of antimalarial drugs could select for parasites showing different iRBC permeability profiles (**Fig. 74**). In the following section we will discuss what existing evidence supports the adaptive potential of a multi-CLAG PSAC and how parasites may overcome the fitness cost of having both CLAG3 proteins absent in the in vitro and possibly the in vivo setting.

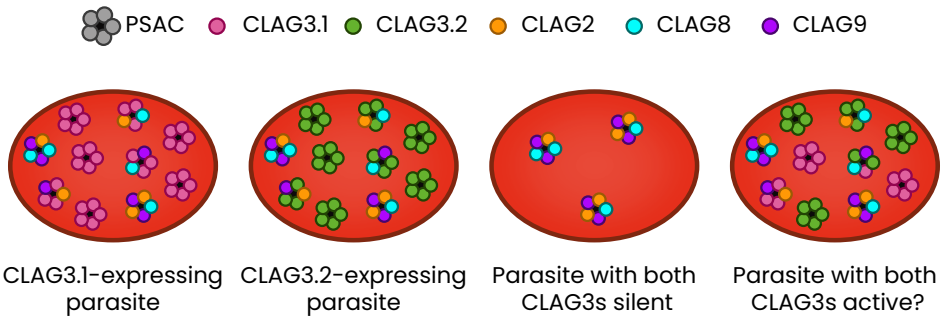


Fig. 74: Adaptive potential of a multi-CLAG PSAC combined with clonally variant expression of *clag3* genes. More details in the text. Shown are four *P. falciparum*-infected RBCs with different PSAC channels on their surface. These channels consist of different CLAG paralogs with the differences in CLAG3 determined by the expression patterns of these genes in each individual parasite (it has yet to be determined if parasites can express both *clag3* genes simultaneously).

Most current evidence of CLAG-mediated PSAC solute selectivity is associated with CLAG3 proteins. Both genetic mutations in the sequences of these proteins[451,522,524] as well as epigenetic changes affecting the expression patterns of the genes that encode them[325–327] have been found to be responsible for alterations in PSAC solute selectivity. On the one hand, absence of CLAG3 proteins has been linked to a loss of iRBC membrane permeability towards the antimalarial compounds BSD[326,327], leupeptin[325], T3[325] and T16[325]; in addition to other various solutes such as sorbitol or L-Alanine[326]. On the other hand, differences in PSAC solute selectivity have also been observed between parasites expressing *clag3.1* and *clag3.2*. Currently, the only example of this is that parasites from the 3D7 background that express *clag3.1* are more resistant to low concentrations of BSD than those expressing *clag3.2*[326]. However, it has also been observed that field isolates predominantly express *clag3.2* in vivo and then switch to express *clag3.1* when grown in vitro[335]. Furthermore, when these isolates were exposed to BSD, in some cases *clag3.1*-expressing parasites were

selected and in others *clag3.2*-expressing ones were, suggesting that this solute selectivity is strain specific[335]. How the two almost identical CLAG3 paralogs[380] can confer the PSAC with different solute selectivity is surprising and the sequence differences that confer each phenotype are not known.

With regards to the other CLAG paralogs, although there is currently no evidence linking them directly with NPPs activity and solute uptake, the enticing possibility that they may come together in different combinations to form PSACs with different solute selectivities could offer malaria parasites a huge adaptive potential as mentioned above. This hypothesis could be supported by the fact that the multigene *clag* family is conserved across most of the *Plasmodium* genus[391]. Furthermore, it shows clade-specific expansion in the *Laverania* subgenus, with some ape *Plasmodium* spp. possessing up to seven times the number of *clag* genes found in the *P. falciparum* genome[535]. However, if the different *clag* paralogs were to provide different solute selectivity, why *P. falciparum* is one of the species with one of the lowest numbers of paralogs[535] when it is more exposed to antimalarial drug pressure than its ape counterparts is a question that would need to be addressed. Nevertheless, we and others hypothesise that the other CLAGs help maintain PSAC activity when CLAG3s are absent[325,377,379,391,401] and when this absence is due to antimalarial compounds selecting for parasites with both *clag3* genes silenced, it would appear that these other paralogs can mediate the transport of essential nutrients while simultaneously preventing the uptake these toxic compounds into the iRBC[325]. While this difference in nutrient and drug transport could be due to the solute selectivity of CLAG3-free PSACs, another possibility is that this type of PSAC is generally less abundant on the iRBC surface and that the absence of CLAG3 results in an overall decrease in solute transport. This second scenario could explain the lower growth rates sometimes associated with CLAG3 absence in vitro[326,401] as the general reduction in solute and consequently nutrient transport would affect parasite development but the supraphysiological nutrient concentrations in standard RPMI-based culture media[258,473] would ensure parasite survival.

It has yet to be determined whether CLAG3-independent PSAC activity is enough to sustain parasites in vivo, although it was recently found that a CLAG3 knockout was unable to grow in media that more closely resembled the human plasma[401]. Nonetheless, even though PSAC activity mediated solely by other CLAGs may not be sufficient to overcome the deficit in nutrient uptake derived from the combination of CLAG3 absence and lower nutrient availability in vivo, parasites could be able to survive

by the means of a different mechanism. Recently, it has been found that in Kelch13-mediated artemisinin resistance ring stage parasites can survive drug pressure by entering into a state of arrested development[254]. This can be explained by the fact that mutations in the Kelch13 protein compromise an endocytosis pathway responsible for haemoglobin uptake and, therefore, result in a decrease in amino acid availability that delays parasite growth at the ring stage[254]. If absence of CLAG3 proteins could result in a similar arrest in development as that observed in Kelch13-mediated artemisinin resistance, albeit through a different mechanism, it is possible that parasites without these proteins could be viable in vivo. Furthermore, the drug resistance mechanism mediated by the selection of parasites with both *clag3* genes silenced could also take place in this setting, as parasites would remain in a state of arrested development until drug elimination, when parasite growth could resume thanks to epigenetic stochastic switching resulting in *clag3*-expressing parasites reappearing in the population (**Fig. 75**).

In summary, there is a growing body of evidence that supports the hypothesis that different *clag* genes may mediate different PSAC solute selectivity and that this could help the malaria parasite to survive within the fluctuating environment of the human host circulation. However, as we have mentioned throughout this chapter, further experimental evidence is required to demonstrate the involvement of other CLAG paralogs in PSAC formation and solute selectivity and to translate all the described in vitro findings to the in vivo setting. When our *clag3* triple transgenic line is ready, we hope to be able to help address this gap of knowledge by identifying further compounds that select for different *clag3* expression patterns and by determining how these genes can further contribute to altering the solute selectivity of the PSAC.

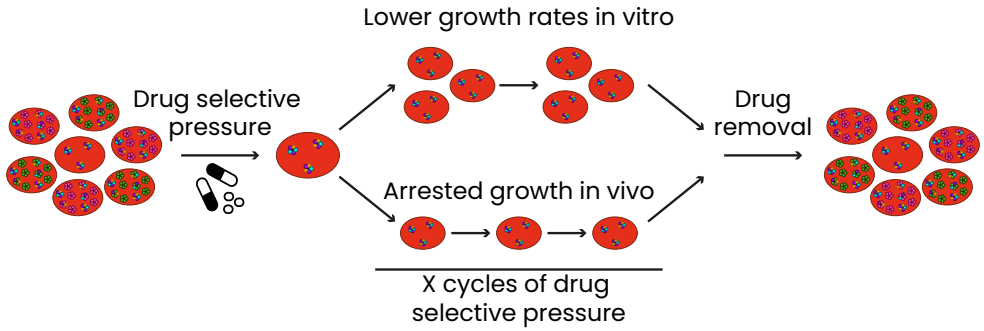


Fig. 75: Model of how malaria parasites overcome the fitness cost of having both *clag3* genes silenced in vivo and in vitro due to the selection with antimalarial drugs. The exposure to certain toxic compounds selects for parasites that have both *clag3* genes silenced enabling the parasite population to survive the drug pressure. In the absence of CLAG3s, PSAC channels composed of other CLAG paralogs that are less abundant on the iRBC surface maintain minimal solute uptake and under standard in vitro conditions are sufficient to maintain the transport of essential nutrients as these are in abundance in the culture medium. In this manner parasites continue to develop but at lower growth rates. On the other hand, in vivo CLAG3-independent PSAC activity is not enough to maintain the transport of essential nutrients as these are less abundant in the human plasma. Therefore, parasites enter into a state of arrested growth in order to survive and overcome the fitness cost of having both *clag3* genes silenced. Upon drug removal, stochastic switching reintroduces *clag3*-expressing parasites into the population and normal nutrient uptake and parasite development are resumed.

5. Differences and translatability of CVG expression between in vitro and in vivo settings

The conditions *Plasmodium* spp. are exposed to in vitro differ significantly from those they encounter in vivo with regards to nutrient availability, haematocrit, temperature, gas levels, rheology and exposure to host factors such as the immune response and hormones, among others[258,536]. This has been found to substantially affect parasite biology with changes at the metabolic, transcriptional and genetic level having been detected between parasites from these two growth environments[258,536]. Here we will discuss how CVG expression varies between the in vitro and in vivo conditions, in addition to the benefits and limitations that each setting offers with regards to the study of these genes in malaria parasites.

5.1. Differences in CVG expression between in vitro and in vivo conditions

Various studies have directly compared the in vivo and in vitro transcriptome of malaria parasites and found that they differ significantly. One of the major differences that consistently appears between the two settings is that of CVGs showing higher and more varied expression in vivo[258,264,537,538]. We observed something similar in Article 2 of this thesis, when comparing same-genome parasites kept either entirely in culture conditions or exposed to various cycles of asexual replication in the bloodstream of naïve human hosts after passage through transmission stages. We found that parasites exposed to in vivo conditions presented overall higher and more heterogenous CVG expression. That said, our results led us to conclude that these CVG expression patterns observed at the onset of a blood infection were primarily the result of an epigenetic reset during transmission stages rather than a consequence of exposure to the selective forces of the human host circulation.

Another type of study that has compared the in vivo and in vitro transcriptome of malaria parasites, is that which involve the adaptation of *P. falciparum* field isolates to culture conditions. Once again, in these studies, in addition to metabolic and genetic differences[255,539,540], CVGs were among the genes to show the most notable transcriptional changes during this adaptive process. More specifically, clonally variant surface antigens (especially *var* genes[541–543]) as well as the clonally variant *pfap2-g* (in addition to other gametocyte markers[264,542,544,545]) were found to be

significantly downregulated after adaptation to culture conditions.

All of these different studies have made it increasingly apparent that CVGs are among the most environmentally-sensitive genes of malaria parasites and suggest that they play a more fundamental role in vivo than in vitro. Not only is this because the role of some CVGs is more relevant under the in vivo setting (e.g., cytoadherence mediated by *var* encoded PfEMP1 proteins[183,323,324]) but also because compared to tightly-controlled culture conditions[536], the in vivo environment is much more variable. For instance, in vivo factors such as the immune response and heterogeneity of host receptors (e.g., on the surface of RBC or endothelium, among others) have been suggested to select for the diverse phenotypes encoded by different clonally variant surface antigens[546]. Moreover, less obvious factors such as changes in fluid dynamics have also been found to affect the expression of exported antigens when comparing static and shaking conditions in vitro and this could also translate to the in vivo setting[544]. However, more experimental evidence is needed to understand the role of higher CVG expression in vivo, especially with regards to the less-studied families.

An ideal experimental set-up with which to compare the effects of in vitro and in vivo conditions on CVG expression could be to take a transcriptionally-diverse, in vitro culture of asexual intraerythrocytic parasites and use it to directly inject malaria-naïve human volunteers (direct blood challenge) as well as to maintain a parallel in vitro culture. Next, by regularly taking blood samples from the volunteers and comparing them with the in vitro culture, the differences in CVG expression between the in vivo and in vitro conditions could be analysed at both the transcriptomic and epigenomic levels (**Fig. 76**). With this approach, it would be possible to avoid the confounding effect of an epigenetic reset of CVG expression in the in vivo setting, as parasites would not be passaged through mosquitoes or liver stages. However, this experimental set up would still suffer from a series of limitations that derive from the study of in vivo infections, as will be analysed in the following section.

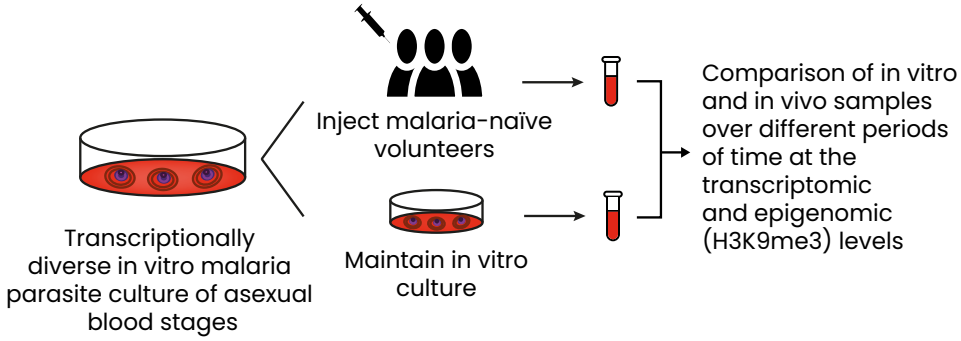


Fig. 76: Experimental set-up with which to compare CVG expression between in vitro and in vivo settings. More details in the text.

5.2. Challenges of studying CVG expression in vivo

Despite the fact that studying *P. falciparum* in its natural environment can often provide more physiologically relevant information on its biology, the in vivo study of these pathogens can be extremely complex. One of the principal reasons for this complexity is the variability that arises from factors such as the immune response, the physiological state or presence of antimalarial drugs with regards to the human host, as well as from other factors like the multiplicity of parasite strains involved, the possibility of mixed infections with other *Plasmodium* spp. or even the amount of sporozoites received in the moment of infection[547–550]. This is further complicated in the case of CVGs by the abundance of polyclonal infections and genetic diversity among field isolates. CHMI trials constitute an interesting compromise between the two settings as they combine the physiological conditions of an in vivo infection with the more controlled environment of an in vitro experiment by using malaria-naïve volunteers of defined characteristics that receive a controlled dose of the same parasite strain. In this manner, CHMI trials can be considered a valuable tool with which to study malaria parasite biology and have been used on various occasions to study CVG expression[172,331–335,462,542,551]. For example, in Article 2 of this thesis, the use of CHMI samples provided us with fundamental insight into CVG expression at the onset of a blood infection under natural yet "controlled" conditions that we would not have been able to obtain solely in vitro.

In addition to the variability that derives from in vivo studies which can partially be solved by CHMI, there are two other main limitations that complicate the design and interpretation of these studies. The first

limitation revolves around the ethical considerations of injecting volunteers with malaria parasites[459,460] and the maintenance of induced (or natural) infections for sufficient time to study the biological aspects of interest. As described in the first chapter of this discussion, CHMI trials only allow for the study of the initial stages of a malaria infection as participants are promptly treated with antimalarials as soon as they present symptoms. Likewise, naturally-infected individuals are typically treated if they are positive for *P. falciparum*, although in some studies with careful clinical monitoring asymptomatic individuals can be left without treatment for some days. Therefore, it is extremely difficult to obtain experimental evidence on how CVG expression varies over the course of a long-term infection as voluntary or natural infections cannot be left untreated for long periods of time.

The other main limitation of in vivo studies is that the samples obtained from these infections often have low parasitaemias and they only contain ring-stage parasites as late stages are sequestered in the microvasculature[49]. Limited parasitaemia and subsequent low amounts of parasite material (e.g., DNA and RNA, among others) can be addressed by using ultra-sensitive methods with extra amplification cycles. However, the only way to analyse the full IDC cycle of in vivo parasites is to submit them to ex vivo culture for at least one cycle (this can also be used to increase parasitaemia), which could have an effect on CVG expression patterns. While we demonstrated that 5 weeks of in vitro development did not affect the expression patterns of *gbph2*, *pfmc-2tm* or various *stevor* genes, it has been shown that *clag3* gene expression can rapidly change between in vivo and in vitro conditions[335]. In some cases, it is also possible to mature parasites ex vivo until the stage of interest without reinvasion, as has often been done to study the expression of invasion genes[552]. Nevertheless, although ex vivo culture is the only current option to analyse the full IDC at the transcriptomic level, another alternative could be to analyse CVG expression at the epigenetic level. As heterochromatin distribution patterns have been found to be maintained throughout the different stages of the IDC[266], it would be possible to determine which CVGs show active or silent expression patterns using only ring stages. Although the limiting factor for this would be obtaining sufficient parasite material for ChIP-seq analysis, this can be more easily addressed by technological advances rather than solving the problem of obtaining in vivo late-stage parasite samples. In this manner, the differences in CVG expression between in vivo and in vitro conditions could be analysed by monitoring changes in the epigenome, constituting a form of "epigenetic epidemiology"[553].

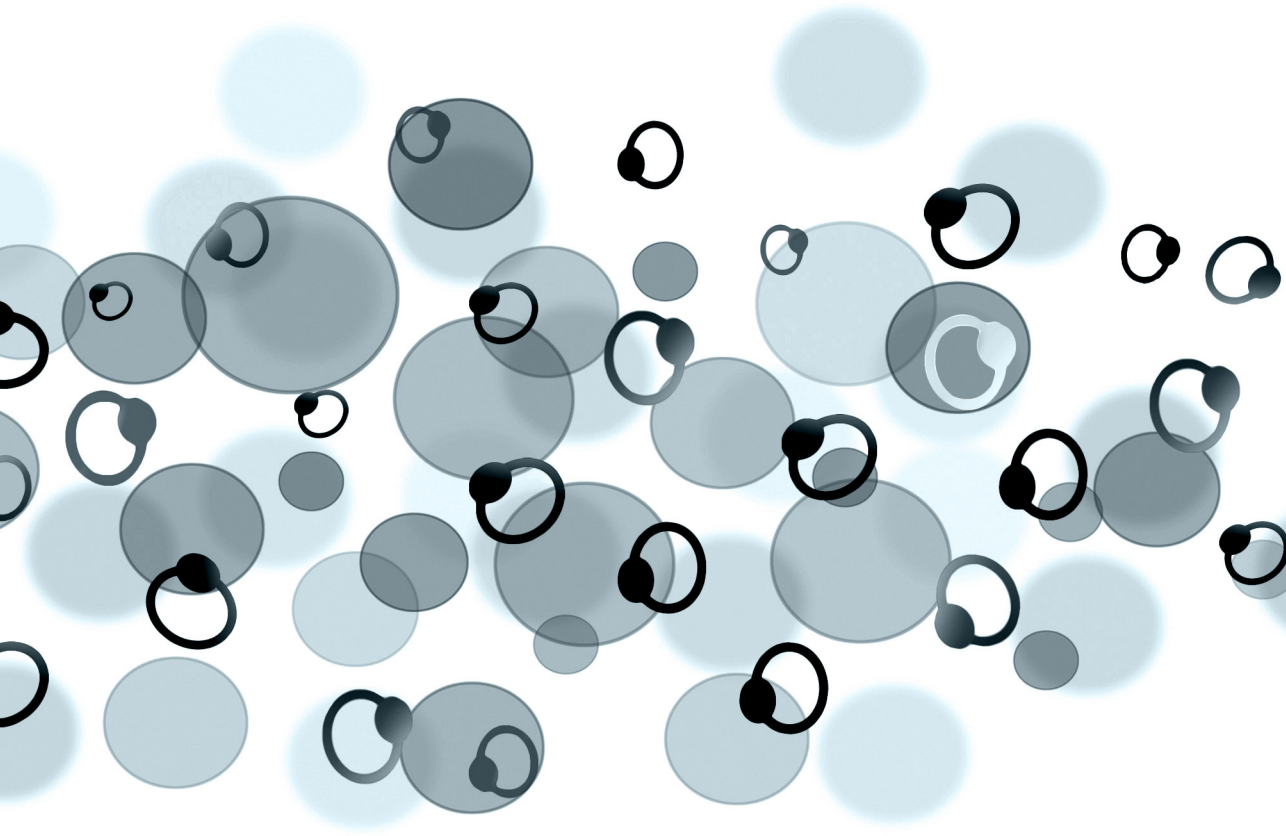
5.3. Importance of in vitro studies on CVG expression

While the study of natural infections can be extremely complex, in vitro approaches represent a much more controlled environment where it is possible to manipulate one variable at a time and reach more direct conclusions. Consequently, this makes the in vitro setting extremely useful for exploring questions on basic malaria parasite biology that are too difficult to address in vivo. For instance, this was the case with our experimental approach in Article 3 of this thesis, in which we set out to determine whether *P. falciparum* utilises CVG-mediated bet-hedging adaptive strategies to adapt to fluctuations in lipid availability in the human host's serum. Whilst in vitro we were able to control the exact amount of the external lipid source our parasite lines were exposed to in the culture media, this would have been virtually impossible in an in vivo setting. Not only would there be ethical considerations of maintaining volunteers on a completely lipid-free diet but also, given the complexity of lipid metabolism in the human host, it would be extremely difficult to establish what sources of lipids parasites would actually be utilising.

On the other hand, the major limitation of in vitro experiments is their in vivo translatability. This, for instance, has been found to be an important issue in drug screening assays as in vitro and in vivo parasites have been found to show different sensitivity to the same antimalarial drugs[554,555]. These variations in sensitivities have been linked to differences in genetic backgrounds, fitness costs associated to drug resistance that are more apparent in a particular growth environment, changes in their transport properties or the influence of the availability of nutrients on drug uptake and activation, among others[258]. This could be applicable to the *clag3* drug resistance mechanism[325–327] for which there is currently only in vitro evidence. That said, as previously discussed, this resistance mechanism has the potential to be at least considered relevant in vivo and to be further characterised in this setting although it is possible that there may be differences in what is observed in vitro.

In conclusion, even though in vitro studies do not always provide an exact picture of what occurs in vivo, they are extremely useful for understanding different aspects of parasite biology, including CVG expression, which can then be developed to be studied or applied in the field. Moreover, the cost in time, money or even human or experimental animal suffering is not always justified by the desire to obtain an in vivo result and therefore the value of in vitro studies should not be forgotten.

CONCLUSIONS



1. Exposure of *Plasmodium falciparum* to subcurative concentrations of the antimalarial compounds T3 and T16 selects for parasites with both *clag3* genes silenced and constitutes a new example of the *clag3* drug resistance mechanism that is regulated at the epigenetic level and is a CVG-mediated bet-hedging adaptive strategy.

2. T3 and T16, in addition to the previously described blasticidin and leupeptin, require a CLAG3-dependent PSAC for their uptake.

3. PSAC activity can be maintained in the absence of *clag3* expression.

4. After passage through transmission stages and at the onset of a blood infection in a new human host, *P. falciparum* presents higher and more diverse CVG expression patterns.

5. CVGs present the most significant transcriptional differences between same-genome *P. falciparum* parasites maintained in vitro and parasites passaged through transmission stages and exposed to the conditions of the human host circulation, supporting the importance of these genes in vivo.

6. The CVGs that show the biggest transcriptional differences after passage through transmission stages are: the previously described *var* and *clag3* genes, practically the entire *pfmc-2tm* and *gbp* families, type A *rif* genes, various members of the *stevor* and *phist* families, and single members from the *hyp10* and *lysophospholipase* families.

7. Epigenetic changes in the distribution of H3K9me3-marked heterochromatin are responsible for the transcriptional patterns of CVGs after passage through transmission stages.

8. The transcriptional patterns of CVGs at the onset of a blood infection are mainly determined by an epigenetic reset during transmission stages rather than by within-host selection.

9. New epigenetic patterns of CVG expression are not established before liver stages, as a single sporozoite can generate the same transcriptional diversity in CVG expression as multiple sporozoites, after passage through the liver.

10. CVG transcriptional diversity at the onset of a blood infection constitutes the basis of a bet-hedging adaptive strategy.

11. *P. falciparum* parasites can adapt to grow without an external source of lipids and after a selection period of 24-60 days present up to 80% growth relative to control parasites without lipid deprivation.

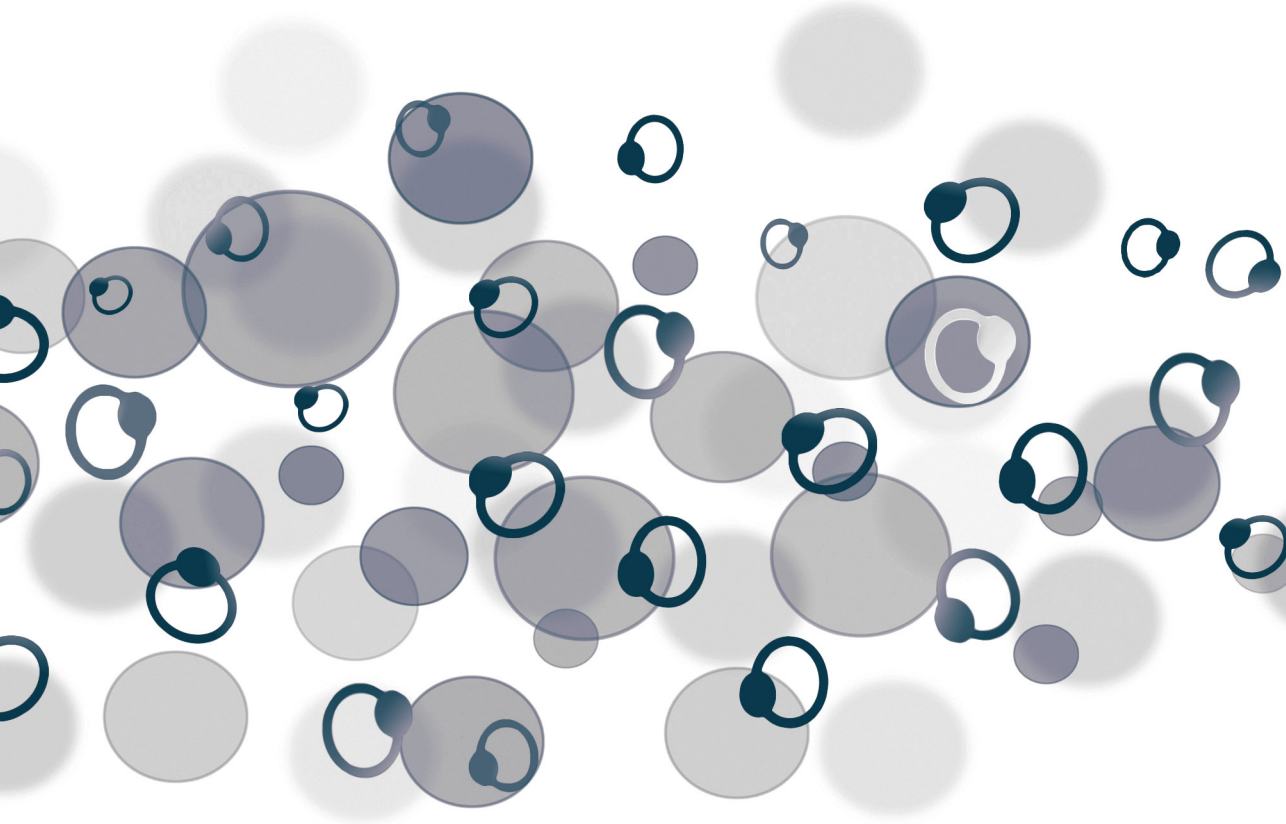
12. Transcriptional changes do not appear to be responsible for *P. falciparum* adaptation to grow without an external source of lipids and consequently this adaptation does not involve CVG-mediated bet-hedging adaptive strategies.

13. Mutations that result in the loss of a functional type II NADH:ubiquinone oxidoreductase (*pfndh2*) are associated with the adaptation of *P. falciparum* to lipid starvation indicating that there is a link between lipid metabolism and the mitochondrial electron transport chain under these conditions.

14. *P. falciparum* parasites do not show signs of stable adaptation when selected with sub-optimal concentrations of glucose.

15. We have designed a tool with which to characterise *clag3* expression dynamics in vitro consisting in a triple transgenic parasite line generated using CRISPR-Cas9 technology.

BIBLIOGRAPHY



1. Loy, D. E. *et al*. Out of Africa: origins and evolution of the human malaria parasites *Plasmodium falciparum* and *Plasmodium vivax*. *Int. J. Parasitol*. **47**, 87–97 (2017).
2. Cox, F. E. History of the discovery of the malaria parasites and their vectors. *Parasit. Vectors* **3**, 5 (2010).
3. Salvi, C. History of malaria elimination in the European Region. *WHO*. (2016).
4. WHO. Elimination of malaria. *World Malaria Report*. (2009).
5. Phillips, M. A. *et al*. Malaria. *Nat. Rev. Dis. Prim*. **3**, 17050 (2017).
6. Adl, S. M. *et al*. Revisions to the Classification, nomenclature and diversity of eukaryotes. *J. Eukaryot. Microbiol*. **66**, 4–119 (2019).
7. Morrison, D. A. Evolution of the Apicomplexa: where are we now? *Trends Parasitol.* **25**, 375–382 (2009).
8. White, M. W. & Suvorova, E. S. Apicomplexa cell cycles: something old, borrowed, lost, and new. *Trends Parasitol*. **34**, 759–771 (2018).
9. Seeber, F. & Steinfelder, S. Recent advances in understanding apicomplexan parasites. *F1000Research* **5**, 1369 (2016).
10. Gibson, A. R. & Striepen, B. *Cryptosporidium*. *Curr. Biol*. **28**, 193–194 (2018).
11. Halonen, S. K. & Weiss, L. M. Toxoplasmosis. *Handb clin Neurol.* **114**, 125–145 (2013).
12. O'Donoghue, P. Haemoprotozoa: Making biological sense of molecular phylogenies. *Int. J. Parasitol. Parasites Wildl*. **6**, 241–256 (2017).
13. Florin-Christensen, M. Piroplasmids and ticks: a long-lasting intimate relationship. *Front. Biosci*. **1**, 3064–3073 (2009).
14. Galen, S. C. *et al*. The polyphyly of *Plasmodium*: Comprehensive phylogenetic analyses of the malaria parasites (Order *Haemosporida*) reveal widespread taxonomic conflict. *R. Soc. Open Sci*. **5**, 171780 (2018).
15. Wykes, M. N. & Good, M. F. What have we learnt from mouse models for the study of malaria? *Eur. J. Immunol*. **39**, 2004–2007 (2009).
16. Faust, C. & Dobson, A. P. Primate malarias: Diversity, distribution and insights for zoonotic *Plasmodium*. *One Heal*. **1**, 66–75 (2015).
17. Mayxay, M. *et al*. Mixed-species malaria infections in humans. *Trends Parasitol*. **20**, 233–240 (2004).
18. Paul, R. E. L. *et al*. The evolutionary ecology of *Plasmodium*. *Ecol. Lett*. **6**, 866–880 (2003).
19. WHO. World Malaria Report 2020. *WHO Global report* (2020).
20. Milner, D. A. Malaria pathogenesis. *Cold Spring Harb. Perspect. Med*. **8**, a025569 (2018).
21. Dayananda, K. K. *et al*. Epidemiology, drug resistance, and pathophysiology of *Plasmodium vivax* malaria. *J. Vector Borne Dis*. **55**, 1-8 (2018).
22. Howes, R. E. *et al*. Global epidemiology of *Plasmodium vivax*. *Am. J. Trop. Med. Hyg*. **95**, 15–34 (2016).
23. Kawamoto, F. *et al*. How prevalant are *Plasmodium ovale* and *P. malariae* in East Asia? *Parasitol. Today* **15**, 422–426 (1999).
24. Collins, W. E. & Jeffery, G. M. *Plasmodium malariae*: parasite and disease. *Clin. Microbiol. Rev*. **20**, 579–592 (2007).

25. Sutherland, C. J. *et al*. Two nonrecombining sympatric forms of the human malaria parasite *Plasmodium ovale* occur globally. *J. Infect. Dis*. **201**, 1544–1550 (2010).
26. Tomar, L. R. *et al*. Complicated malaria: a rare presentation of *Plasmodium ovale. Trop. Doct.* **45**, 140–142 (2015).
27. Ahmed, M. A. & Cox-Singh, J. *Plasmodium knowlesi* - an emerging pathogen . *ISBT Sci. Ser*. **10**, 134–140 (2015).
28. Lee, K. S. *et al*. *Plasmodium knowlesi*: Reservoir hosts and tracking the emergence in humans and macaques. *PLoS Pathog*. **7**, e1002015 (2011).
29. Singh, B. *et al*. A large focus of naturally acquired *Plasmodium knowlesi* infections in human beings. *Lancet* **363**, 1017–1024 (2004).
30. Tavares, J. *et al*. Role of host cell traversal by the malaria sporozoite during liver infection. *J. Exp. Med*. **210**, 905–915 (2013).
31. Amino, R. *et al*. Quantitative imaging of *Plasmodium* transmission from mosquito to mammal. *Nat. Med*. **12**, 220–224 (2006).
32. Sturm, A. Manipulation of host hepatocytes by the malaria parasite for delivery into liver sinusoids. *Science* **313**, 1287–1290 (2006).
33. Garcia, L. S. Malaria. *Clin. Lab. Med*. **30**, 93–129 (2010).
34. Venugopal, K. *et al*. *Plasmodium* asexual growth and sexual development in the haematopoietic niche of the host. *Nat. Rev. Microbiol*. **18**, 177–189 (2020).
35. Kafsack, B. F. C. *et al*. A transcriptional switch underlies commitment to sexual development in malaria parasites. *Nature* **507**, 248–252 (2014).
36. Hawking, F. *et al*. Evidence for cyclic development and short-lived maturity in the gametocytes of *Plasmodium falciparum. Trans. R. Soc. Trop. Med. Hyg*. **65**, 549–559 (1971).
37. Tibúrcio, M. *et al*. A switch in infected erythrocyte deformability at the maturation and blood circulation of *Plasmodium falciparum* transmission stages. *Blood* **119**, 172–180 (2012).
38. Dearnley, M. K. *et al*. Origin, composition, organization and function of the inner membrane complex of *Plasmodium falciparum* gametocytes. *J. Cell Sci*. **125**, 2053–2063 (2012).
39. Aingaran, M. *et al*. Host cell deformability is linked to transmission in the human malaria parasite *Plasmodium falciparum. Cell. Microbiol*. **14**, 983–993 (2012).
40. Sologub, L. *et al*. Malaria proteases mediate inside-out egress of gametocytes from red blood cells following parasite transmission to the mosquito. *Cell. Microbiol*. **13**, 897–912 (2011).
41. Aly, A. S. I. *et al*. Malaria parasite development in the mosquito and infection of the mammalian host. *Annu. Rev. Microbiol*. **63**, 195–221 (2009).
42. Nilsson, S. K. *et al*. Targeting human transmission biology for malaria elimination. *PLoS Pathog*. **11**, e1004871 (2015).
43. Bhatt, S. *et al*. The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. *Nature* **526**, 207–211 (2015).

44. Sinka, M. E. *et al*. The dominant *Anopheles* vectors of human malaria in Africa, Europe and the Middle East: Occurrence data, distribution maps and bionomic précis. *Parasites and Vectors* **3**, 117 (2010).
45. Ashley, E. A. *et al*. Malaria. *Lancet* **391**, 1608–1621 (2018).
46. Cowman, A. F. *et al*. Malaria: Biology and disease. *Cell* **167**, 610–624 (2016).
47. Miller, L. H. *et al*. The pathogenic basis of malaria. *Nature* **415**, 673–679 (2002).
48. English, M. C. *et al*. Hyponatraemia and dehydration in severe malaria. *Arch. Dis. Child*. **74**, 201–205 (1996).
49. Coban, C. *et al*. Tissue-specific immunopathology during malaria infection. *Nat. Rev. Immunol.* **18**, 266–278 (2018).
50. Su, X. *et al*. The large diverse gene family *var* encodes proteins involved in cytoadherence and antigenic variation of *Plasmodium falciparum*-infected erythrocytes. *Cell* **82**, 89–100 (1995).
51. Leech, J. H. *et al*. Identification of a strain-specific malarial antigen exposed on the surface of *Plasmodium falciparum*-infected erythrocytes. *J. Exp. Med*. **159**, 1567–1575 (1984).
52. Snow, B. *et al*. Receptor-specific adhesion and clinical disease in *Plasmodium falciparum*. *Am. J. Trop. Med. Hyg*. **57**, 389–398 (1997).
53. Pain, A. *et al*. Platelet-mediated clumping of *Plasmodium falciparum*-infected erythrocytes is a common adhesive phenotype and is associated with severe malaria. *Proc. Natl. Acad. Sci. USA*. **98**, 1805–1810 (2001).
54. McQuaid, F. & Rowe, J. A. Rosetting revisited: A critical look at the evidence for host erythrocyte receptors in *Plasmodium falciparum* rosetting. *Parasitology* **147**, 1–11 (2020).
55. Salanti, A. *et al*. Evidence for the involvement of VAR2CSA in pregnancy-associated malaria. *J. Exp. Med*. **200**, 1197–1203 (2004).
56. Smith, J. D. *et al*. Identification of a *Plasmodium falciparum* intercellular adhesion molecule-1 binding domain: A parasite adhesion trait implicated in cerebral malaria. *Proc. Natl. Acad. Sci. USA*. **97**, 1766–1771 (2000).
57. Turner, L. *et al*. Severe malaria is associated with parasite binding to endothelial protein C receptor. *Nature* **498**, 502–505 (2013).
58. Aird, W. C. *et al*. Blood Spotlight *Plasmodium falciparum* picks (on) EPCR. *Blood* **123**, 163–168 (2018).
59. Barros Pinto, M. P. & Marques, G. Severe malaria. *Infection* **48**, 143–146 (2020).
60. Bernabeu, M. & Smith, J. D. EPCR and malaria severity: The center of a perfect storm. *Trends Parasitol*. **33**, 295–308 (2017).
61. Wassmer, S. C. & Grau, G. E. R. Severe malaria: what's new on the pathogenesis front? *Int. J. Parasitol*. **47**, 145–152 (2017).
62. Millar, S. B. & Cox-Singh, J. Human infections with *Plasmodium knowlesi*—zoonotic malaria. *Clin. Microbiol. Infect*. **21**, 640–648 (2015).
63. Layland, L. E. & Specht, S. Helpful or a Hindrance: Co-infections with Helminths During Malaria. *Adv. Exp. Med. Biol.* **828**, 99–129 (2014).

64. González, R. *et al*. HIV and malaria interactions: where do we stand? *Expert Rev. Anti. Infect. Ther*. **10**, 153–165 (2012).
65. Ezechi, O. *et al*. HIV/AIDS, tuberculosis, and malaria in pregnancy. *J. Pregnancy* **2012**, 140826 (2012).
66. Langhorne, J. *et al*. Immunity to malaria: More questions than answers. *Nat. Immunol.* **9**, 725–732 (2008).
67. Mathison, B. A. & Pritt, B. S. Update on malaria diagnostics and test utilization. *J. Clin. Microbiol.* **55**, 2009–2017 (2017).
68. Bousema, T. *et al*. Asymptomatic malaria infections: detectability, transmissibility and public health relevance. *Nat. Rev. Microbiol*. **12**, 833–840 (2014).
69. Ochola, L. *et al*. The reliability of diagnostic techniques in the diagnosis and management of malaria in the absence of a gold standard. *Lancet Infect. Dis*. **6**, 582–588 (2006).
70. Olliaro, P. Mode of action and mechanisms of resistance for antimalarial drugs. *Pharmacol. Ther.* **89**, 207–219 (2001).
71. Plewes, K. *et al*. Malaria: What's new in the management of malaria? *Infect. Dis. Clin. North Am*. **33**, 39–60 (2019).
72. Ashley, E. A. *et al*. Spread of artemisinin resistance in *Plasmodium falciparum* malaria. *N. Engl. J. Med*. **371**, 411–423 (2014).
73. WHO. Guidelines for the treatment of malaria. (2015).
74. White, N. J. Review series Antimalarial drug resistance. *J. Clin. Invest.* **113**, 1084-1092 (2004).
75. Ter Kuile, F. O. *et al*. Predictors of mefloquine treatment failure: A prospectivestudy of 1590 patients with uncomplicated *falciparum* malaria. *Trans. R. Soc. Trop. Med. Hyg*. **89**, 660–664 (1995).
76. Wang, J. *et al*. Artemisinin, the magic drug discovered from traditional Chinese medicine. *Engineering* **5**, 32–39 (2019).
77. Hiebsch, R. R., Raub, T. J. & Wattenberg, B. W. Primaquine blocks transport by inhibiting the formation of functional transport vesicles: Studies in a cell-free assay of protein transport through the Golgi apparatus. *J. Biol. Chem*. **266**, 20323–20328 (1991).
78. Baggish, A. L. & Hill, D. R. Antiparasitic agent atovaquone. *Antimicrob. Agents Chemother*. **46**, 1163–1173 (2002).
79. Sullivan, D. J. *et al*. On the molecular mechanism of chloroquine's antimalarial action. *Proc. Natl. Acad. Sci. USA*. **93**, 11865–11870 (1996).
80. Hoppe, H. C. *et al*. Antimalarial quinolines and artemisinin inhibit endocytosis in *Plasmodium falciparum*. *Antimicrob. Agents Chemother*. **48**, 2370–2378 (2004).
81. Khan, J. *et al*. Molecular mechanisms of action and resistance of antimalarial drugs. *Bacterial Adaptation to Co-resistance* (eds. Mandal, S. M. & Paul, D.) 267–296 (Springer Nature Singapore Pte Ltd., 2019).
82. Price, R. N. *et al*. Global extent of chloroquine-resistant *Plasmodium vivax*: a systematic review and meta-analysis. *Lancet Infect. Dis*. **14**, 982–991 (2014).

83. Chu, C. S. & White, N. J. Management of relapsing *Plasmodium vivax* malaria. *Expert Rev. Anti. Infect. Ther*. **14**, 885–900 (2016).
84. Ashley, E. A. & Phyo, A. P. Drugs in development for malaria. *Drugs* **78**, 861–879 (2018).
85. malERA: An updated research agenda for diagnostics, drugs, vaccines, and vector control in malaria elimination and eradication. *PLoS Med*. **14**, e1002455 (2017).
86. Burt, A. Heritable strategies for controlling insect vectors of disease. *Phil. Trans. R. Soc. B*. **369**, 20130432 (2014).
87. Esvelt, K. M. *et al*. Concerning RNA-guided gene drives for the alteration of wild populations. *eLife* **3**, e03401 (2014).
88. Desai, M. *et al*. Prevention of malaria in pregnancy. *Lancet Infect. Dis*. **18**, e119–e132 (2018).
89. Ashley, E. A. & Yeka, A. Seasonal malaria chemoprevention: closing the know–do gap. *Lancet* **396**, 1778–1779 (2020).
90. Laurens, M. B. The promise of a malaria vaccine—are we closer? *Annu. Rev. Microbiol.* **72**, 273–292 (2018).
91. Cockburn, I. A. & Seder, R. A. Malaria prevention: from immunological concepts to effective vaccines and protective antibodies. *Nat. Immunol*. **19**, 1199–1211 (2018).
92. Laurens, M. B. RTS,S/AS01 vaccine (MosquirixTM): an overview. *Hum. Vaccin. Immunother*. **16**, 480–489 (2020).
93. Schuerman, L. RTS,S malaria vaccine could provide major public health benefits. *Lancet* **394**, 735–736 (2019).
94. Chandramohan, D. et al. Seasonal malaria vaccination with or without seasonal malaria chemoprevention. *N. Engl. J. Med*. **385**, 1005–1017 (2021).
95. Barry, A. E. & Arnott, A. Strategies for designing and monitoring malaria vaccines targeting diverse antigens. *Front. Immunol*. **5**, 359 (2014).
96. WHO. Global technical strategy for malaria 2016-2030. (2015).
97. Alonso, P. L. Malaria: a problem to be solved and a time to be bold. *Nat. Med*. **27**, 1506–1509 (2021).
98. Brooks, A. N. *et al*. Adaptation of cells to new environments. *Wiley Interdiscip. Rev. Syst. Biol. Med*. **3**, 544–561 (2011).
99. Agozzino, L. *et al*. How Do Cells Adapt? Stories Told in Landscapes. *Annu. Rev. Chem. Biomol. Eng*. **11**, 155–182 (2020).
100. Yona, A. H. *et al*. A Relay race on the evolutionary adaptation spectrum. *Cell* **163**, 549–559 (2015).
101. Gasch, A. P. *et al*. Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell* **11**, 4241–4257 (2000).
102. Causton, H. C. *et al*. Remodeling of yeast genome expression in response to environmental changes. *Mol. Biol. Cell* **12**, 323–337 (2001).
103. Goldberg, A. D. *et al*. Epigenetics: A landscape takes shape. *Cell* **128**, 635–638 (2007).
104. Cavalli, G. & Heard, E. Advances in epigenetics link genetics to the environment and disease. *Nature* **571**, 489–499 (2019).

105. Devaskar, S. U. & Raychaudhuri, S. Epigenetics - A science of heritable biological adaptation. *Pediatr. Res*. **61**, 2–5 (2007).
106. Weaver, I. C. G. *et al*. Stress and the emerging roles of chromatin remodeling in signal integration and stable transmission of reversible phenotypes. *Front. Behav. Neurosci.* **11**, 41 (2017).
107. Li, B. *et al*. The role of chromatin during transcription. *Cell* **128**, 707–719 (2007).
108. Cedar, H. & Bergman, Y. Linking DNA methylation and histone modification: Patterns and paradigms. *Nat. Rev. Genet.* **10**, 295–304 (2009).
109. Klose, R. J. & Bird, A. P. Genomic DNA methylation: The mark and its mediators. *Trends Biochem. Sci.* **31**, 89–97 (2006).
110. Mukherjee, K. *et al*. Insects as models to study the epigenetic basis of disease. *Prog. Biophys. Mol. Biol*. **118**, 69–78 (2015).
111. Breiling, A. & Lyko, F. Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. *Epigenetics Chromatin* **8**, 24 (2015).
112. Smith, Z. D. *et al*. DNA methylation dynamics of the human preimplantation embryo. *Nature* **511**, 611–615 (2014).
113. Takayama, S. *et al*. Genome methylation in *D. melanogaster* is found at specific short motifs and is independent of DNMT2 activity. *Genome Res.* **24**, 821–830 (2014).
114. Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693–705 (2007).
115. Lawrence, M., Daujat, S. & Schneider, R. Lateral thinking: How histone modifications regulate gene expression. *Trends Genet*. **32**, 42–56 (2016).
116. Ragunathan, K. *et al*. Epigenetic inheritance uncoupled from sequence-specific recruitment. *Science* **348**, 132-135 (2015).
117. Audergon, P. N. C. B. *et al*. Restricted epigenetic inheritance of H3K9 methylation. *Science* **348**, 132–135 (2015).
118. Holoch, D. & Moazed, D. RNA-mediated epigenetic regulation of gene expression. *Nat. Rev. Genet*. **16**, 71–84 (2015).
119. Zhou, L. *et al*. Histone acetylation promotes long-lasting defense responses and longevity following early life heat stress. *PLoS Genet*. **15**, e1008122 (2019).
120. Volpe, T. A. *et al*. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**, 1833–1837 (2002).
121. Rinn, J. L. & Chang, H. Y. Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem*. **81**, 145–166 (2012).
122. Yan, W. Potential roles of noncoding RNAs in environmental epigenetic transgenerational inheritance. *Mol. Cell. Endocrinol*. **398**, 24–30 (2014).
123. Rakyan, V. K. *et al*. Metastable epialleles in mammals. *Trends Genet*. **18**, 348–351 (2002).
124. Frías-Lasserre, D. & Villagra, C. A. The importance of ncRNAs as epigenetic mechanisms in phenotypic variation and organic evolution. *Front. Microbiol*. **8**, 2483 (2017).
125. Boyko, A. & Kovalchuk, I. Genome instability and epigenetic modification-heritable responses to environmental stress? *Curr. Opin. Plant Biol*. **14**, 260–266 (2011).

126. Rechavi, O. *et al*. Transgenerational inheritance of an acquired small RNA-based antiviral response in *C. elegans. Cell* **147**, 1248–1256 (2011).
127. Rosenberg, S. M. Evolving responsively: adaptive mutation. *Nat. Rev. Genet*. **2**, 504–515 (2001).
128. Galhardo, R. S. *et al*. Mutation as a stress response and the regulation of evolvability. *Crit. Rev. Biochem. Mol. Biol*. **42**, 399–435 (2007).
129. Leigh, E. G. Natural selection and mutability. *Am. Nat*. **104**, 301–305 (1970).
130. Perry, G. H. *et al*. Diet and the evolution of human amylase gene copy number variation. *Nat. Genet*. **39**, 1256–1260 (2007).
131. Tang, Y. C. & Amon, A. Gene copy-number alterations: A cost-benefit analysis. *Cell* **152**, 394–405 (2013).
132. Kaya, A. *et al*. Molecular signatures of aneuploidy-driven adaptive evolution. *Nat. Commun*. **11**, 588 (2020).
133. Selmecki, A. Aneuploidy and isochromosome formation in drug-Resistant *Candida albicans. Science* **313**, 367–370 (2006).
134. Loewe, L. & Hill, W. G. The population genetics of mutations: good, bad and indifferent. *Philos. Trans. R. Soc. Lond. B. Biol. Sci*. **365**, 1153–1167 (2010).
135. Henikoff, S. & Shilatifard, A. Histone modification: Cause or cog? *Trends Genet.* **27**, 389–396 (2011).
136. Sheltzer, J. M. & Amon, A. The aneuploidy paradox: Costs and benefits of an incorrect karyotype. *Trends Genet.* **27**, 446–453 (2011).
137. Seco-Hidalgo, V. *et al*. To bet or not to bet: Deciphering cell to cell variation in protozoan infections. T*rends Parasitol*. **31**, 350–356 (2015).
138. Forsman, A. Rethinking phenotypic plasticity and its consequences for individuals, populations and species. *Heredity*. **115**, 276–284 (2015).
139. Peltier, E. *et al*. Dissection of the molecular bases of genotype x environment interactions: A study of phenotypic plasticity of *Saccharomyces cerevisiae* in grape juices. *BMC Genomics* **19**, 772 (2018).
140. Denlinger, D. & Armbruster, P. Mosquito diapause. *Annu. Rev. Entomol*. **59**, 73–93 (2014).
141. Whitman DW, A. A. What is phenotypic plasticity and why is it important. in Phenotypic plasticity of insects: mechanisms and consequences 1–63 (Enfield, NH: CRC Press., 2009).
142. Auld, J. R. *et al*. Re-evaluating the costs and limits of adaptive phenotypic plasticity. *Proc. Biol. Sci*. **277**, 503–511 (2010).
143. DeWitt, T. J. *et al*. Costs and limits of phenotypic plasticity. *Trends Ecol. Evol*. **13**, 77–81 (1998).
144. Simons, A. M. Playing smart vs. playing safe: The joint expression of phenotypic plasticity and potential bet hedging across and within thermal environments. *J. Evol. Biol.* **27**, 1047–1056 (2014).
145. Philippi, T. & Seger, J. Hedging one's evolutionary bets, revisited. *Trends Ecol. Evol.* **4**, 41–44 (1989).

146. Herman, J. J. *et al*. How stable 'should' epigenetic modifications be? Insights from adaptive plasticity and bet hedging. *Evolution*. **68**, 632–643 (2014).
147. Levy, S. F. *et al*. Bet hedging in yeast by heterogeneous, age-correlated expression of a stress protectant. *PLoS Biol*. **10**, e1001325 (2012).
148. Starrfelt, J. & Kokko, H. Bet-hedging-a triple trade-off between means, variances and correlations. *Biol. Rev.* **87**, 742–755 (2012).
149. Veening, J.-W. *et al*. Bistability, epigenetics, and bet-hedging in bacteria. *Annu. Rev. Microbiol*. **62**, 193–210 (2008).
150. Viney, M. & Reece, S. E. Adaptive noise. *Proc. Biol. Sci*. **280**, 20131104 (2013).
151. Satory, D. *et al*. Epigenetic switches: Can infidelity govern fate in microbes? *Curr. Opin. Microbiol*. **14**, 212–217 (2011).
152. Balázsi, G. *et al*. Cellular decision making and biological noise: From microbes to mammals. *Cell* **144**, 910–925 (2011).
153. Veening, J. W. *et al*. Bet-hedging and epigenetic inheritance in bacterial cell development. *Proc. Natl. Acad. Sci. USA* **105**, 4393–4398 (2008).
154. Beaumont, H. J. E. *et al*. Experimental evolution of bet hedging. *Nature* **462**, 90–93 (2009).
155. Burd, M. *et al*. Ovule number per flower in a world of unpredictable pollination. *Am. J. Bot.* **96**, 1159–1167 (2009).
156. VanWallendael, A. *et al*. A Molecular view of plant local adaptation: Incorporating stress-response networks. *Annu. Rev. Plant Biol*. **70**, 559–583 (2019).
157. Guo, T. *et al*. Dynamic effects of interacting genes underlying rice flowering-time phenotypic plasticity and global adaptation. *Genome Res*. **30**, 673–683 (2020).
158. Childs, D. Z. *et al*. Evolutionary bet-hedging in the real world: empirical evidence and challenges revealed by plants. *Proc. Biol. Sci*. **277**, 3055–3064 (2010).
159. Franks, S. J. Plasticity and evolution in drought avoidance and escape in the annual plant *Brassica rapa. New Phytol*. **190**, 249–257 (2011).
160. Caljon, G. *et al*. Alice in microbes' land: Adaptations and counter-adaptations of vector-borne parasitic protozoa and their hosts. *FEMS Microbiol. Rev*. **40**, 664–685 (2016).
161. Capewell, P. *et al*. A co-evolutionary arms race: Trypanosomes shaping the human genome, humans shaping the trypanosome genome. *Parasitology* **142**, 108–119 (2015).
162. Vonlaufen, N. *et al*. Stress response pathways in protozoan parasites. *Cell. Microbiol*. **10**, 2387–2399 (2008).
163. Hasday, J. D. *et al*. The role of fever in the infected host. *Microbes Infect*. **2**, 1891–1904 (2000).
164. Oakley, M. S. *et al*. Clinical and molecular aspects of malaria fever. *Trends Parasitol*. **27**, 442–449 (2011).
165. Tintó-Font, E. *et al*. A heat-shock response regulated by the PfAP2-HS transcription factor protects human malaria parasites from febrile temperatures. *Nat. Microbiol.* **6**, 1163-1174 (2021).

166. Nepveu, F. & Turrini, F. Targeting the redox metabolism of *Plasmodium falciparum*. *Future Med. Chem*. **5**, 1993–2006 (2013).
167. Becker, K. *et al*. Oxidative stress in malaria parasite-infected erythrocytes: host–parasite interactions. *Int. J. Parasitol*. **34**, 163–189 (2004).
168. Ginsburg, H. & Atamina, H. The redox status of malaria-infected erythrocytes: an overview with an emphasis on unresolved problems. *Parasite* **1**, 5–13 (1994).
169. Oliveira, P. L. & Oliveira, M. F. Vampires, Pasteur and reactive oxygen species. *FEBS Lett*. **525**, 3–6 (2002).
170. Zambrano-Villa, S. *et al*. How protozoan parasites evade the immune response. *Trends Parasitol*. **18**, 272–278 (2002).
171. Brodin, P. & Davis, M. M. Human immune system variation. *Nat. Rev. Immunol*. **17**, 21–29 (2017).
172. Milne, K. *et al*. Mapping immune variation and *var* gene switching in naive hosts infected with *Plasmodium falciparum*. *eLife* **10**, e62800 (2021).
173. Gomes, P. S. *et al*. Immune Escape Strategies of Malaria Parasites. *Front. Microbiol*. **7**, 1617 (2016).
174. Anderson, H. L. *et al*. The evolving erythrocyte: red blood cells as modulators of innate immunity. *J. Immunol.* **201**, 1343–1351 (2018).
175. Mercier, F. E. *et al*. The bone marrow at the crossroads of blood and immunity. *Nat. Rev. Immunol*. **12**, 49–60 (2011).
176. Kanellopoulos-Langevin, C. *et al*. Tolerance of the fetus by the maternal immune system: role of inflammatory mediators at the feto-maternal interface. *Reprod. Biol. Endocrinol*. **1**, 121 (2003).
177. Hong, S. & Van Kaer, L. Immune privilege: Keeping an eye on natural killer T cells. *J. Exp. Med.* **190**, 1197–1200 (1999).
178. Russell, B. *et al*. Human ex vivo studies on asexual *Plasmodium vivax*: The best way forward. *Int. J. Parasitol*. **42**, 1063–1070 (2012).
179. Suwanarusk, R. *et al*. The deformability of red blood cells parasitized by *Plasmodium falciparum* and *P. vivax. J. Infect. Dis*. **189**, 190–194 (2004).
180. Sanyal, S. *et al. Plasmodium falciparum* STEVOR proteins impact erythrocyte mechanical properties. *Blood* **119**, 1–8 (2012).
181. Yam, X. Y. & Preiser, P. R. Host immune evasion strategies of malaria blood stage parasite. *Mol. Biosyst*. **13**, 2498–2508 (2017).
182. Recker, M. *et al*. Antigenic Variation in *Plasmodium falciparum* malaria involves a highly structured switching pattern. *PLoS Pathog*. **7**, e1001306 (2011).
183. Scherf, A. *et al*. Antigenic variation in *Plasmodium falciparum. Annu. Rev. Microbiol.* **62**, 445–470 (2008).
184. Tougan, T. *et al*. Molecular camouflage of *Plasmodium falciparum* merozoites by binding of host vitronectin to P47 fragment of SERA5. *Sci. Rep*. **8**, 5052 (2018).
185. Saito, F. *et al.* Immune evasion of *Plasmodium falciparum* by RIFIN via inhibitory receptors. *Nature* **552**, 101–105 (2017).

186. Zuzarte-Luís, V. & Mota, M. M. Parasite sensing of host nutrients and environmental cues. *Cell Host Microbe* **23**, 749–758 (2018).
187. Glew, R. H. *et al.* Lipid profiles and trans fatty acids in serum phospholipids of semi-nomadic Fulani in northern Nigeria. *J. Heal. Popul. Nutr*. **28**, 159–166 (2010).
188. Arama, C. *et al*. Ethnic differences in susceptibility to malaria: What have we learned from immuno-epidemiological studies in West Africa? *Acta Trop*. **146**, 152–156 (2015).
189. Zuzarte-Luís, V. *et al*. Dietary alterations modulate susceptibility to *Plasmodium* infection. *Nat. Microbiol*. **2**, 1600–1607 (2017).
190. Lain, K. Y. & Catalano, P. M. Metabolic changes in pregnancy. *Clin. Obstet. Gynecol.* **50**, 938–948 (2007).
191. Kalra, S. *et al*. Malaria and diabetes. *J. Pak. Med. Assoc.* **67**, 810–813 (2017).
192. Das, D. *et al*. Complex interactions between malaria and malnutrition: a systematic literature review. *BMC Med*. **16**, 186 (2018).
193. Das, B. S. *et al.* Hypoglycaemia in severe *falciparum* malaria. *Trans. R. Soc. Trop. Med. Hyg*. **82**, 197–201 (1988).
194. Planche, T. & Krishna, S. Severe malaria: metabolic complications. *Curr. Mol. Med.* **6**, 141-53 (2006).
195. Olszewski, K. L. *et al*. Host-parasite interactions revealed by *Plasmodium falciparum* metabolomics. Cell Host Microbe **5**, 191–199 (2009).
196. Polonais, V. & Soldati-Favre, D. Versatility in the acquisition of energy and carbon sources by the Apicomplexa. *Biol. Cell* **102**, 435–445 (2010).
197. Ke, H. *et al*. Genetic investigation of tricarboxylic acid metabolism during the *Plasmodium falciparum* life cycle. *Cell Rep*. **11**, 164–174 (2015).
198. Mancio-Silva, L. *et al*. Nutrient sensing modulates malaria parasite virulence. *Nature* **547**, 213–216 (2017).
199. Ararat-Sarria, M. *et al*. Parasite-related genetic and epigenetic aspects and host factors influencing *Plasmodium falciparum* invasion of erythrocytes. *Front. Cell. Infect. Microbiol.* **8**, 454 (2019)
200. Thiam, L. G. et al. Blood donor variability is a modulatory factor for *P. falciparum* invasion phenotyping assays. *Sci. Rep*. **11**, 7129 (2021).
201. Reid, M. E., & Lomas-Francis, C. *The blood group antigen facts book*. (Elsevier, 2004).
202. Lelliott, P. M. *et al*. The influence of host genetics on erythrocytes and malaria infection: Is there therapeutic potential? *Malar. J*. **14**, 289 (2015).
203. Leffler, E. M. *et al*. Resistance to malaria through structural variation of red blood cell invasion receptors. *Science* **356**, 6393 (2017).
204. Kepple, D. *et al*. Alternative invasion mechanisms and host immune response to *Plasmodium vivax* malaria: trends and future directions. *Microorganisms* **9**, 15 (2020).
205. Kwiatkowski, D. P. How malaria has affected the human genome and what human genetics can teach us about malaria. *Am. J. Hum. Genet*. **77**, 171–192 (2005).

206. Piel, F. B. The present and future global burden of the inherited disorders of hemoglobin. *Hematol. Oncol. Clin. North Am.* **30**, 327–341 (2016).
207. Goheen, M. M. *et al.*. The role of the red blood cell in host defence against *falciparum* malaria: an expanding repertoire of evolutionary alterations. *Br. J. Haematol*. **179**, 543–556 (2017).
208. Romphosri, S. *et al.* Role of a concentration gradient in malaria drug resistance evolution: A combined within- and between-hosts modelling approach. *Sci. Rep.* **10**, 6219 (2020).
209. Newton, P. N. *et al.* A collaborative epidemiological investigation into the criminal fake artesunate trade in South East Asia. *PLoS Med.* **5**, e32 (2008).
210. Dondorp, A. M. *et al.* Fake antimalarials in Southeast Asia are a major impediment to malaria control: multinational cross-sectional survey on the prevalence of fake antimalarials. *Trop. Med. Int. Heal.* **9**, 1241–1246 (2004).
211. Cui, L. *et al.* Malaria in the Greater Mekong Subregion: Heterogeneity and complexity. *Acta Trop.* **121**, 227–239 (2012).
212. Haldar, K. *et al.* Drug resistance in *Plasmodium*. *Nat. Rev. Microbiol.* **16**, 156–170 (2018).
213. Goldberg, D. E. *et al.* Outwitting evolution: Fighting drug-resistant TB, malaria, and HIV. *Cell* **148**, 1271–1283 (2012).
214. Rosenthal, P. J. The interplay between drug resistance and fitness in malaria parasites. *Mol. Microbiol.* **89**, 1025–1038 (2013).
215. Deitsch, K. *et al.* Mechanisms of gene regulation in *Plasmodium*. *Am. J. Trop. Med. Hyg.* **77**, 201–208 (2007).
216. Le Roch, K. G. *et al.* Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science* **301**, 1503–1508 (2003).
217. Bozdech, Z. *et al.* The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biol.* **1**, E5 (2003).
218. Howick, V. M. *et al.* The malaria cell atlas: Single parasite transcriptomes across the complete *Plasmodium* life cycle. *Science* **365**, eaaw2619 (2019).
219. Painter, H. J. *et al.* The Apicomplexan AP2 family: Integral factors regulating *Plasmodium* development. *Mol. Biochem. Parasitol.* **176**, 1–7 (2011).
220. Campbell, T. L. *et al.* Identification and genome-wide prediction of DNA binding specificities for the ApiAP2 family of regulators from the malaria parasite. *PLoS Pathog.* **6**, e1001165 (2010).
221. Santos, J. M. *et al.* Red blood cell invasion by the malaria parasite Is coordinated by the PfAP2-I transcription factor. *Cell Host Microbe* **21**, 731-741. (2017).
222. Ganesan, K. *et al.* A genetically hard-wired metabolic transcriptome in *Plasmodium falciparum* fails to mount protective responses to lethal antifolates. *PLoS Pathog.* **4**, e1000214 (2008).
223. Gunasekera, A. M. *et al. Plasmodium falciparum*: Genome wide perturbations in transcript profiles among mixed stage cultures after chloroquine treatment. *Exp. Parasitol.* **117**, 87–92 (2007).

224. Kato, N. *et al*. Gene expression signatures and small-molecule compounds link a protein kinase to *Plasmodium falciparum* motility. *Nat. Chem. Biol*. **4**, 347–356 (2008).
225. Kritsiriwuthinan, K. *et al*. Global gene expression profiling of *Plasmodium falciparum* in response to the anti-malarial drug pyronaridine. *Malar. J*. **10**, 242 (2011).
226. Le Roch, K. G. *et al*. A systematic approach to understand the mechanism of action of the bisthiazolium compound T4 on the human malaria parasite, *Plasmodium falciparum. BMC Genomics* **9**, 513 (2008).
227. Hu, G. *et al*. Transcriptional profiling of growth perturbations of the human malaria parasite *Plasmodium falciparum. Nat. Biotechnol.* **28**, 91–98 (2010).
228. Tamez, P. A. *et al*. An erythrocyte vesicle protein exported by the malaria parasite promotes tubovesicular lipid import from the host cell surface. *PLoS Pathog*. **4**, e1000118 (2008).
229. Dahl, E. L. *et al.* Tetracyclines specifically target the apicoplast of the malaria parasite *Plasmodium falciparum*. *Antimicrob. Agents Chemother*. **50**, 3124–3131 (2006).
230. Babbitt, S. E. *et al. Plasmodium falciparum* responds to amino acid starvation by entering into a hibernatory state. *Proc. Natl. Acad. Sci. USA.* **109**, E3278–E3287 (2012).
231. Torrentino-Madamet, M. *et al.* Global response of *Plasmodium falciparum* to hyperoxia: A combined transcriptomic and proteomic approach. *Malar. J*. **10**, 4 (2011).
232. Chaubey, S. *et al.* Endoplasmic reticulum stress triggers gametocytogenesis in the malaria parasite. *J. Biol. Chem*. **289**, 16662–16674 (2014).
233. Natalang, O. et al. Dynamic RNA profiling in *Plasmodium falciparum* synchronized blood stages exposed to lethal doses of artesunate. *BMC Genomics* **9**, 388 (2008).
234. Chou, E. S. *et al*. A high parasite density environment induces transcriptional changes and cell death in *Plasmodium falciparum* blood stages. *FEBS J*. **285**, 848–870 (2018).
235. Oakley, M. S. M. *et al*. Molecular factors and biochemical pathways induced by febrile temperature in intraerythrocytic *Plasmodium falciparum* parasites. *Infect. Immun.* **75**, 2012–2025 (2007).
236. Brancucci, N. M. B. *et al.* Lysophosphatidylcholine regulates sexual stage differentiation in the human malaria parasite *Plasmodium falciparum*. *Cell* **171**, 1532–1544 (2017).
237. Filarsky, M. *et al*. GDV1 induces sexual commitment of malaria parasites by antagonizing HP1-dependent gene silencing. *Science* **359**, 1259–1263 (2018).
238. Reece, S. E. *et al*. Sex ratio adjustment and kin discrimination in malaria parasites. *Nature* **453**, 609–614 (2008).
239. Trager, W. & Gill, G. S. Enhanced gametocyte formation in young erythrocytes by *Plasmodium falciparum* in vitro. *J. Protozool*. **39**, 429–432 (1992).
240. Bruce, M. C. *et al*. Commitment of the malaria parasite *Plasmodium falciparum* to sexual and asexual development. *Parasitology.* **100**, 191–200 (1990).
241. Schneider, P. *et al*. Virulence, drug sensitivity and transmission success in the rodent malaria, *Plasmodium chabaudi*. *Proc. Biol. Sci*. **279**, 4677–4685 (2012).

242. Bousema, T. *et al*. Human immune responses that reduce the transmission of *Plasmodium falciparum* in African populations. *Int. J. Parasitol.* **41**, 293–300 (2011).
243. Portugaliza, H. P. *et al*. Artemisinin exposure at the ring or trophozoite stage impacts *Plasmodium falciparum* sexual conversion differently. *eLife* **9**, e60058 (2020).
244. Gardner, M. J. et al. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419**, 498–511 (2002).
245. Lovett, S. T. Encoded errors: Mutations and rearrangements mediated by misalignment at repetitive DNA sequences. *Mol. Microbiol*. **52**, 1243–1253 (2004).
246. Miles, A. *et al.* Indels, structural variation, and recombination drive genomic diversity in *Plasmodium falciparum*. *Genome Res*. **26**, 1288–1299 (2016).
247. Hamilton, W. L. *et al*. Extreme mutation bias and high AT content in *Plasmodium falciparum. Nucleic Acids Res.* **45**, 1889–1901 (2017).
248. Manske, M. *et al*. Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. *Nature* **487**, 375–379 (2012).
249. Cheeseman, I. H. *et al*. Population structure shapes copy number variation in malaria parasites. *Mol. Biol. Evol.* **33**, 603–620 (2016).
250. Heinberg, A. *et al*. Direct evidence for the adaptive role of copy number variation on antifolate susceptibility in *Plasmodium falciparum. Mol. Microbiol*. **88**, 702–712 (2013).
251. Guler, J. L. *et al*. Asexual populations of the human malaria parasite, *Plasmodium falciparum*, use a two-step genomic strategy to acquire accurate, beneficial DNA amplifications. *PLoS Pathog*. **9**, e1003375 (2013).
252. Cowell, A. N. & Winzeler, E. A. The genomic architecture of antimalarial drug resistance. *Brief. Funct. Genomics* **18**, 314–328 (2019).
253. Siegel, S. *et al*. Mitochondrial heteroplasmy is responsible for Atovaquone drug resistance in *Plasmodium falciparum*. *bioRxiv* 232033 (2017).
254. Birnbaum, J. *et al.* A Kelch13-defined endocytosis pathway mediates artemisinin resistance in malaria parasites. *Science* **367**, 51–59 (2020).
255. Claessens, A. *et al.* Culture adaptation of malaria parasites selects for convergent loss-of-function mutants. *Sci. Rep*. **7**, 41303 (2017).
256. Day, K. P. *et al*. Genes necessary for expression of a virulence determinant and for transmission of *Plasmodium falciparum* are located on a 0.3-megabase region of chromosome 9. *Proc. Natl. Acad. Sci. USA.* **90**, 8292–8296 (1993).
257. Anderson, T. J. C. *et al*. Gene copy number and malaria biology. *Trends Parasitol.* **25**, 336–343 (2009).
258. Brown, A. C. & Guler, J. L. From circulation to cultivation: *Plasmodium* in vivo versus in vitro. *Trends Parasitol.* **36**, 914–926 (2020).
259. Rénia, L. & Goh, Y. S. Malaria parasites: The great escape. *Front. Immunol*. **7**, 463 (2016).
260. Anders, R. F. *et al*. Molecular variation in *Plasmodium falciparum*: Polymorphic antigens of asexual erythrocytic stages. *Acta Trop*. **53**, 239–253 (1993).

261. Drew, D. R. *et al*. Defining the antigenic diversity of *Plasmodium falciparum* apical membrane antigen 1 and the requirements for a multi-allele vaccine against malaria. *PLoS One* **7**, e51023 (2012).
262. Samarakoon, U. *et al.* The landscape of inherited and de novo copy number variants in a *Plasmodium falciparum* genetic cross. *BMC Genomics* **12**, 457 (2011).
263. Gonzales, J. M. *et al*. Regulatory hotspots in the malaria parasite genome dictate transcriptional variation. *PLoS Biol.* **6**, e238 (2008).
264. Mackinnon, M. J. *et al*. Comparative transcriptional and genomic analysis of P*lasmodium falciparum* field isolates. *PLoS Pathog.* **5**, e1000644 (2009).
265. Mok, S. *et al*. Structural polymorphism in the promoter of *pfmrp2* confers *Plasmodium falciparum* tolerance to quinoline drugs. *Mol. Microbiol*. **91**, 918–934 (2014).
266. Fraschka, S. A. *et al*. Comparative heterochromatin profiling reveals conserved and unique epigenome signatures linked to adaptation and development of malaria parasites. *Cell Host Microbe* **23**, 407-420.e8 (2018).
267. Flueck, C. *et al. Plasmodium falciparum* heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors. *PLoS Pathog*. **5**, e1000569 (2009).
268. Lopez-Rubio, J. J. *et al*. Genome-wide analysis of heterochromatin associates clonally variant gene regulation with perinuclear repressive centers in malaria parasites. *Cell Host Microbe* **5**, 179–190 (2009).
269. Miao, J. *et al*. The malaria parasite *Plasmodium falciparum* histones: organization, expression, and acetylation. *Gene* **369**, 53–65 (2006).
270. Biterge, B. & Schneider, R. Histone variants: key players of chromatin. *Cell Tissue Res*. **356**, 457–466 (2014).
271. Gupta, A. P. & Bozdech, Z. Epigenetic landscapes underlining global patterns of gene expression in the human malaria parasite *Plasmodium falciparum*. *Int. J. Parasitol*. **47**, 399–407 (2017).
272. Trelle, M. B. *et al*. Global histone analysis by mass spectrometry reveals a high content of acetylated lysine residues in the malaria parasite *Plasmodium falciparum. J. Proteome Res*. **8**, 3439–3450 (2009).
273. Sullivan, W. J. *et al*. Histones and histone modifications in protozoan parasites. *Cell. Microbiol*. **8**, 1850–1861 (2006).
274. Salcedo-Amaya, A. M. *et al*. Dynamic histone H3 epigenome marking during the intraerythrocytic cycle of *Plasmodium falciparum. Proc. Natl. Acad. Sci. USA*. **106**, 9655–9660 (2009).
275. Silberhorn, E. *et al. Plasmodium falciparum* Nucleosomes exhibit reduced stability and lost sequence dependent nucleosome positioning. *PLoS Pathog*. **12**, e1006080 (2016).
276. Petter, M. *et al*. H2A.Z and H2B.Z Double-variant nucleosomes define intergenic regions and dynamically occupy *var* gene promoters in the malaria parasite *Plasmodium falciparum. Mol. Microbiol*. **87**, 1167–1182 (2013).

277. Hoeijmakers, W. A. M. et al. H2A.Z/H2B.Z Double-variant nucleosomes inhabit the AT-rich promoter regions of the *P. falciparum* genome. *Mol. Microbiol.* **87**, 1061–1073 (2013).
278. Ay, F. *et al*. Three-dimensional modeling of the *P. falciparum* genome during the erythrocytic cycle reveals a strong connection between genome architecture and gene expression. *Genome Res*. **24**, 974–988 (2014).
279. Batugedara, G. *et al*. The role of chromatin structure in gene regulation of the human malaria parasite. *Trends Parasitol*. **33**, 364–377 (2017).
280. Ralph, S. A. *et al*. Antigenic variation in *Plasmodium falciparum* is associated with movement of *var* loci between subnuclear locations. *Proc. Natl. Acad. Sci. USA.* **102**, 5414–5419 (2005).
281. Weiner, A. *et al*. 3D Nuclear architecture reveals coupled cell cycle dynamics of chromatin and nuclear pores in the malaria parasite *Plasmodium falciparum. Cell. Microbiol.* **13**, 967–977 (2011).
282. Li, F. *et al*. Nuclear non-coding RNAs are transcribed from the centromeres of *Plasmodium falciparum* and are associated with centromeric chromatin. *J. Biol. Chem*. **283**, 5692–5698 (2008).
283. Raabe, C. A. *et al*. A global view of the nonprotein-coding transcriptome in *Plasmodium falciparum. Nucleic Acids Res*. **38**, 608–617 (2009).
284. Li, Y. *et al*. Noncoding RNAs in apicomplexan parasites: An update. *Trends Parasitol*. **36**, 835–849 (2020).
285. Mourier, T. *et al*. Genome-wide discovery and verification of novel structured RNAs in *Plasmodium falciparum. Genome Res*. **18**, 281–292 (2008).
286. Broadbent, K. M. *et al.* Strand-specific RNA sequencing in *Plasmodium falciparum* malaria identifies developmentally regulated long non-coding RNA and circular RNA. *BMC Genomics* **16**, 454 (2015).
287. Ponts, N. *et al.* Genome-wide mapping of DNA methylation in the human malaria parasite *Plasmodium falciparum. Cell Host Microbe* **14**, 696–706 (2013).
288. McInroy, G. R. *et al*. Enhanced methylation analysis by recovery of unsequenceable fragments. *PLoS One* **11**, e0152322 (2016).
289. Hammam, E. *et al*. Discovery of a new predominant cytosine DNA modification that is linked to gene expression in malaria parasites. *Nucleic Acids Res.* **48**, 184–199 (2020).
290. Jabeena, C. A. & Rajavelu, A. Epigenetic players of chromatin structure regulation in *Plasmodium falciparum. ChemBioChem* **20**, 1225–1230 (2019).
291. Cortés, A. & Deitsch, K. W. Malaria epigenetics. *Cold Spring Harb. Perspect. Med*. **7**, 1–23 (2017).
292. Guizetti, J. & Scherf, A. Silence, activate, poise and switch! Mechanisms of antigenic variation in *Plasmodium falciparum. Cell. Microbiol.* **15**, 718–726 (2013).
293. Llorà-Batlle, O. *et al*. Transcriptional variation in malaria parasites: why and how. *Brief. Funct. Genomics* **18**, 329–341 (2019).
294. Cortés, A. *et al*. A view on the role of epigenetics in the biology of malaria parasites. *PLoS Pathog*. **8**, e1002943 (2012).

295. Brancucci, N. M. B. *et al*. Heterochromatin protein 1 secures survival and transmission of malaria parasites. *Cell Host Microbe* **16**, 165–176 (2014).
296. Chookajorn, T. *et al*. Epigenetic memory at malaria virulence genes. *Proc. Natl. Acad. Sci. USA*. **104**, 899–902 (2007).
297. Comeaux, C. A. *et al*. Functional analysis of epigenetic regulation of tandem *RhopH1/clag* genes reveals a role in *Plasmodium falciparum* growth. *Mol. Microbiol.* **80**, 378–390 (2011).
298. Coleman, B. I. *et al.* A *Plasmodium falciparum* histone deacetylase regulates antigenic variation and gametocyte conversion. *Cell Host Microbe* **16**, 177–186 (2014).
299. Duraisingh, M. T. & Skillman, K. M. Epigenetic variation and regulation in malaria parasites. *Annu. Rev. Microbiol*. **72**, 355–375 (2018).
300. Crowley, V. M. *et al*. Heterochromatin formation in bistable chromatin domains controls the epigenetic repression of clonally variant *Plasmodium falciparum* genes linked to erythrocyte invasion. *Mol. Microbiol.* **80**, 391–406 (2011).
301. Jiang, L. *et al.* Epigenetic control of the variable expression of a *P. falciparum* receptor protein for erythrocyte invasion. *Proc. Natl. Acad. Sci. USA*. **107**, 2224–2229 (2010).
302. Lopez-Rubio, J. J. *et al.* 5′ Flanking region of *var* genes nucleate histone modification patterns linked to phenotypic inheritance of virulence traits in malaria parasites. *Mol. Microbiol*. **66**, 1296–1305 (2007).
303. Voss, T. S. *et al.* Epigenetic memory takes center stage in the survival strategy of malaria parasites. *Curr. Opin. Microbiol*. **20**, 88–95 (2014).
304. Coleman, B. I. *et al*. Nuclear repositioning precedes promoter accessibility and is linked to the switching frequency of a *Plasmodium falciparum* invasion gene. *Cell Host Microbe* **12**, 739–750 (2012).
305. Duraisingh, M. T. *et al*. Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in *Plasmodium falciparum*. *Cell* **121**, 13–24 (2005).
306. Dzikowski, R. *et al.* Mechanisms underlying mutually exclusive expression of virulence genes by malaria parasites. *EMBO Rep*. **8**, 959–965 (2007).
307. Zhang, Q. *et al*. Exonuclease-mediated degradation of nascent RNA silences genes linked to severe malaria. *Nature* **513**, 431–435 (2014).
308. Volz, J. C. *et al.* PfSET10, a *P. falciparum* methyltransferase, maintains the active *var* gene in a poised state during parasite division. *Cell Host Microbe* **11**, 7–18 (2012).
309. Ukaegbu, U. E. *et al.* Recruitment of PfSET2 by RNA polymerase II to variant antigen encoding loci contributes to antigenic variation in *P. falciparum. PLoS Pathog.* **10**, e1003854 (2014).
310. Jiang, L. *et al.* PfSETvs methylation of histone H3K36 represses virulence genes in *Plasmodium falciparum. Nature* **499**, 223–227 (2013).
311. Deitsch, K. W. & Dzikowski, R. Variant gene expression and antigenic variation by malaria parasites. *Annu. Rev. Microbiol.* **71**, 625–641 (2017).
312. Avraham, I. *et al*. Insulator-like pairing elements regulate silencing and mutually exclusive expression in the malaria parasite *P. falciparum. Proc. Natl. Acad. Sci. USA* . **109**, E3678-86 (2012).

313. Amit-Avraham, I. *et al.* Antisense long noncoding RNAs regulate *var* gene activation in the malaria parasite *Plasmodium falciparum. Proc. Natl. Acad. Sci. USA*. **112**, E982–E991 (2015).
314. Scherf, A. *et al*. Antigenic variation in malaria: In situ switching, relaxed and mutually exclusive transcription of *var* genes during intra-erythrocytic development in *Plasmodium falciparum. EMBO J.* **17**, 5418–5426 (1998).
315. Cortés, A. *et al*. Epigenetic silencing of *Plasmodium falciparum* genes linked to erythrocyte invasion. *PLoS Pathog.* **3**, 1023–1035 (2007).
316. Taylor, H. M. *et al*. *Var* gene diversity in *Plasmodium falciparum* is generated by frequent recombination events. *Mol. Biochem. Parasitol*. **110**, 391–397 (2000).
317. Freitas-Junior, L. H. *et al*. Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum. Nature* **407**, 1018–1022 (2000).
318. Ribacke, U. *et al*. Genome wide gene amplifications and deletions in *Plasmodium falciparum. Mol. Biochem. Parasitol.* **155**, 33–44 (2007).
319. Mok, B. W. *et al*. A highly conserved segmental duplication in the subtelomeres of *P. falciparum* chromosomes varies in copy number. *Malar. J.* **7**, 46(2008).
320. Claessens, A. *et al.* Generation of antigenic diversity in *P. falciparum* by structured rearrangement of *var* genes during mitosis. *PLoS Genet.* **10**, e1004812 (2014).
321. Waters, A. P. Epigenetic roulette in blood stream *Plasmodium*: gambling on sex. *PLoS Pathog*. **12**, e1005353 (2016).
322. Rovira-Graells, N. *et al*. Transcriptional variation in the malaria parasite *Plasmodium falciparum. Genome Res.* **22**, 925–938 (2012).
323. Smith, J. D. *et al*. Malaria's deadly grip: cytoadhesion of *Plasmodium falciparum* -infected erythrocytes. *Cell. Microbiol.* **15**, 1976–1983 (2013).
324. Wahlgren, M. *et al*. Variant surface antigens of *Plasmodium falciparum* and their roles in severe malaria. *Nat. Rev. Microbiol.* **15**, 479–491 (2017).
325. Mira-Martínez, S. *et al.* Identification of antimalarial compounds That require CLAG3 for their uptake by *Plasmodium falciparum*-infected erythrocytes. *Antimicrob. Agents Chemother.* **63**, e00052-19 (2019).
326. Mira-Martínez, S. *et al.* Epigenetic switches in *clag3* genes mediate blasticidin S resistance in malaria parasites. *Cell. Microbiol.* **15**, 1913-1923 (2013).
327. Sharma, P. *et al*. An epigenetic antimalarial resistance mechanism involving parasite genes linked to nutrient uptake. *J. Biol. Chem*. **288**, 19429–19440 (2013).
328. Spence, P. J. *et al*. Mosquitoes reset malaria parasites. *PLoS Pathog.* **11**, e1004987 (2015).
329. Brugat, T. *et al.* Antibody-independent mechanisms regulate the establishment of chronic *Plasmodium* infection. *Nat. Microbiol*. **2**, 16276 (2017).
330. Spence, P. J. *et al.* Vector transmission regulates immune control of *Plasmodium* virulence. *Nature* **498**, 228–231 (2013).
331. Wang, C. W. *et al.* The *P. falciparum var* gene transcription strategy at the onset of blood stage infection in a human volunteer. *Parasitol. Int*. **58**, 478–480 (2009).

332. Dimonte, S. *et al*. Sporozoite route of infection influences in vitro *var* gene transcription of *Plasmodium falciparum* parasites from controlled human infections. *J. Infect. Dis*. **214**, 884–894 (2016).
333. Bachmann, A. *et al*. Mosquito passage dramatically changes *var* gene expression in controlled human *P. falciparum* infections. *PLoS Pathog*. **12**, e1005538 (2016).
334. Peters, J. *et al*. 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*. **99**, 10689–10694 (2002).
335. Mira-Martínez, S. *et al.* Expression of the *Plasmodium falciparum* clonally variant *clag3* genes in human infections. *J. Infect. Dis.* **215**, 938–945 (2017).
336. Hviid, L. & Jensen, A. T. R. PfEMP1 – A parasite protein family of key importance in *P. falciparum* malaria immunity and pathogenesis. *Adv. Parasitol.* **88** 51–84 (2015).
337. Chen, Q. *et al*. Developmental selection of *var* gene expression in *Plasmodium falciparum. Nature* **394**, 392–395 (1998).
338. Perez-Toledo, K. *et al. Plasmodium falciparum* heterochromatin protein 1 binds to tri-methylated histone 3 lysine 9 and is linked to mutually exclusive expression of *var* genes. *Nucleic Acids Res.* **37**, 2596–2606 (2009).
339. Freitas-Junior, L. H. *et al.* Telomeric heterochromatin propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites. *Cell* **121**, 25–36 (2005).
340. Epp, C. *et al*. Chromatin associated sense and antisense noncoding RNAs are transcribed from the *var* gene family of virulence genes of the malaria parasite *Plasmodium falciparum. RNA* **15**, 116–127 (2008).
341. Bryant, J. M. *et al.* CRISPR/Cas9 genome editing reveals that the intron Is not essential for *var2csa* gene activation or silencing in *Plasmodium falciparum. mBio* **8**, e00729-17 (2017).
342. Guizetti, J. *et al*. Trans-acting GC-rich non-coding RNA at *var* expression site modulates gene counting in malaria parasite. *Nucleic Acids Res.* **44**, 9710-9718 (2016).
343. Barcons-Simon, A. *et al*. CRISPR interference of a clonally variant GC-rich noncoding RNA family leads to general repression of *var* genes in *Plasmodium falciparum. mBio* **11**, e03054-19 (2020).
344. Broadbent, K. M. *et al*. A global transcriptional analysis of *P. falciparum* malaria reveals a novel family of telomere-associated lncRNAs. *Genome Biol*. **12**, R56 (2011).
345. Al-Khedery, B. *et al*. Antigenic variation in malaria: A 3′ genomic alteration associated with the expression of a *P. knowlesi* variant antigen. *Mol. Cell* **3**, 131–141 (1999).
346. Barnwell, J. W. *et al*. Splenic requirement for antigenic variation and expression of the variant antigen on the erythrocyte membrane in cloned *Plasmodium knowlesi* malaria. *Infect. Immun*. **40**, 985–994 (1983).
347. Pain, A. *et al.* The genome of the simian and human malaria parasite *Plasmodium knowlesi*. *Nature* **455**, 799–803 (2008).

348. Cheng, Q. *et al*. *stevor* and *rif* are *Plasmodium falciparum* multicopy gene families which potentially encode variant antigens. *Mol. Biochem. Parasitol.* **97**, 161–176 (1998).
349. Petter, M. *et al*. Diverse expression patterns of subgroups of the *rif* multigene family during *Plasmodium falciparum* gametocytogenesis. *PLoS One* **3**, e3779 (2008).
350. Fernandez, V. *et al.* Small, Clonally variant antigens expressed on the surface of the *Plasmodium falciparum*–infected erythrocyte are encoded by the *rif* gene family and are the target of human immune responses. *J. Exp. Med.* **190**, 1393–1404 (1999).
351. Goel, S. *et al*. RIFINs are adhesins implicated in severe *Plasmodium falciparum* malaria. *Nat. Med.* **21**, 314–321 (2015).
352. Petter, M. *et al.* Variant proteins of the *Plasmodium falciparum* RIFIN family show distinct subcellular localization and developmental expression patterns. *Mol. Biochem. Parasitol.* **156**, 51–61 (2007).
353. Khattab, A. *et al*. *Plasmodium falciparum* variant STEVOR antigens are expressed in merozoites and possibly associated with erythrocyte invasion. *Malar. J*. **7**, 137 (2008).
354. McRobert, L. *et al.* Distinct trafficking and localization of STEVOR proteins in three stages of the *Plasmodium falciparum* life cycle. *Infect. Immun.* **72**, 6597–6602 (2004).
355. Kaviratne, M. *et al.* Small variant STEVOR antigen is uniquely located within Maurer's Clefts in *Plasmodium falciparum*-infected red blood cells. *Eukaryot. Cell* **1**, 926–935 (2002).
356. Niang, M. *et al*. The *Plasmodium falciparum* STEVOR multigene family mediates antigenic variation of the infected erythrocyte. *PLoS Pathog*. **5**, e1000307 (2009).
357. Niang, M. *et al*. STEVOR is a *Plasmodium falciparum* erythrocyte binding protein that mediates merozoite invasion and rosetting. *Cell Host Microbe* **16**, 81–93 (2014).
358. Naissant, B. *et al. P. falciparum* STEVOR phosphorylation regulates host erythrocyte deformability enabling malaria parasite transmission. *Blood* **127**, e42–e53 (2016).
359. Frech, C. & Chen, N. Variant surface antigens of malaria parasites: Functional and evolutionary insights from comparative gene family classification and analysis. *BMC Genomics* **14**, 427 (2013).
360. Sargeant, T. J. *et al.* Lineage-specific expansion of proteins exported to erythrocytes in malaria parasites. *Genome Biol.* **7**, R12 (2006).
361. Llinás, M. *et al*. Comparative whole genome transcriptome analysis of three *Plasmodium falciparum* strains. *Nucleic Acids Res*. **34**, 1166–1173 (2006).
362. Foth, B. J. *et al*. Quantitative time-course profiling of parasite and host cell proteins in the human malaria parasite *P. falciparum. Mol. Cell. Proteomics* **10**, 1–17 (2011).
363. Kumar, V. *et al*. *Plasmodium* helical interspersed subtelomeric family—an enigmatic piece of the *Plasmodium* biology puzzle. *Parasitol. Res.* **118**, 2753–2766 (2019).
364. Eksi, S. *et al*. Identification of a subtelomeric gene family expressed during the asexual-sexual stage transition in *Plasmodium falciparum. Mol. Biochem. Parasitol.* **143**, 90–99 (2005).
365. Silvestrini, F. *et al.* Protein export marks the early phase of gametocytogenesis of the human malaria parasite *Plasmodium falciparum. Mol. Cell. Proteomics* **9**, 1437–1448 (2010).

366. Joice, R. *et al. Plasmodium falciparum* transmission stages accumulate in the human bone marrow. *Sci. Transl. Med*. **6**, 244re5 (2014).
367. Warncke, J. D. *et al. Plasmodium* helical interspersed subtelomeric (PHIST) proteins, at the center of host cell remodeling. *Microbiol. Mol. Biol. Rev*. **80**, 905–927 (2016).
368. Mok, B. W. *et al.* Comparative transcriptomal analysis of isogenic *Plasmodium falciparum* clones of distinct antigenic and adhesive phenotypes. *Mol. Biochem. Parasitol.* **151**, 184–192 (2007).
369. Bertin, G. I. *et al.* Proteomic analysis of *Plasmodium falciparum* parasites from patients with cerebral and uncomplicated malaria. *Sci. Rep*. **6**, 26773 (2016).
370. Ndam, N. T. *et al. Plasmodium falciparum* transcriptome analysis reveals pregnancy malaria associated gene expression. *PLoS One* **3**, e1855 (2008).
371. Claessens, A. *et al*. A subset of group A-like *var* genes encodes the malaria parasite ligands for binding to human brain endothelial cells. *Proc. Natl. Acad. Sci. USA*. **109**, E1772–E1781 (2012).
372. Regev-Rudzki, N. *et al*. Cell-cell communication between malaria-infected red blood cells via exosome-like vesicles. *Cell* **153**, 1120–1133 (2013).
373. Lavazec, C. *et a*l. Expression switching in the *stevor* and *pfmc-2TM* superfamilies in *Plasmodium falciparum. Mol. Microbiol.* **64**, 1621–1634 (2007).
374. Kaur, J. & Hora, R. '2TM proteins': An antigenically diverse superfamily with variable functions and export pathways. *PeerJ* **6** e4757 (2018).
375. Wickert, H. *et a*l. Maurer's cleft organization in the cytoplasm of *Plasmodium falciparum*-infected erythrocytes: new insights from three-dimensional reconstruction of serial ultrathin sections. *Eur. J. Cell Biol*. **83**, 567–582 (2004).
376. Yadavalli, R. *et a*l. Trafficking and association of *Plasmodium falciparum* MC-2TM with the Maurer's Clefts. *Pathogens* **10**, 431 (2021).
377. Kaneko, O. *et a*l. The high molecular mass rhoptry protein, RhopH1, is encoded by members of the *clag* multigene family in *Plasmodium falciparum* and *Plasmodium yoelii. Mol. Biochem. Parasitol*. **118**, 223–231 (2001).
378. Holder, A. A. *et al*. Isolation of a *Plasmodium falciparum* rhoptry protein. *Mol. Biochem. Parasitol.* **14**, 293–303 (1985).
379. Kaneko, O. *et al.* Apical expression of three RhopH1/Clag proteins as components of the *P. falciparum* RhopH complex. *Mol. Biochem. Parasitol*. **143**, 20–28 (2005).
380. Iriko, H. *et al.* Diversity and evolution of the *rhoph1/clag* multigene family of *Plasmodium falciparum. Mol. Biochem. Parasitol.* **158**, 11–21 (2008).
381. Otto, T. D. *et al*. New insights into the blood-stage transcriptome of *Plasmodium falciparum* using RNA-Seq. *Mol. Microbiol*. **76**, 12–24 (2010).
382. Bowman, S. *et al*. The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum. Nature* **400**, 532–538 (1999).
383. Rovira-Graells, N. *et al.* Deciphering the principles that govern mutually exclusive expression of *Plasmodium falciparum clag3* genes. *Nucleic Acids Res*. **43**, 8243–8257 (2015).

384.Barnes, D. A. *et al*. Mapping the genetic locus implicated in cytoadherence of *Plasmodium falciparum* to melanoma cells. *Mol. Biochem. Parasitol.* **66**, 21–29 (1994).
385.Etzion, Z. *et al.* Isolation and characterization of rhoptries of *Plasmodium falciparum. Mol. Biochem. Parasitol.* **47**, 51–61 (1991).
386.Coppel, R. L. *et al.* A cDNA clone expressing a rhoptry protein of *Plasmodium falciparum. Mol. Biochem. Parasitol.* **25**, 73–81 (1987).
387.Cooper, J. A. *et al*. The 140/130/105 kilodalton protein complex in the rhoptries of *Plasmodium falciparum* consists of discrete polypeptides. *Mol. Biochem. Parasitol.* **29**, 251–260 (1988).
388.Gardiner, D. L. *et al.* CLAG 9 is located in the rhoptries of *Plasmodium falciparum. Parasitol. Res*. **93**, 64–67 (2004).
389.Pillai, A. D. *et al.* Solute restriction reveals an essential role for *clag3*-associated channels in malaria parasite nutrient acquisition. *Mol. Pharmacol*. **82**, 1104–1114 (2012).
390.Nguitragool, W. *et al.* Malaria parasite *clag3* genes determine channel-mediated nutrient uptake by infected red blood cells. *Cell* **145**, 665–677 (2011).
391. Gupta, A. *et al*. The conserved *clag* multigene family of malaria parasites: Essential roles in host–pathogen interaction. *Drug Resist. Updat*. **18**, 47–54 (2015).
392.Counihan, N. A. et al. *Plasmodium falciparum* parasites deploy RhopH2 into the host erythrocyte to obtain nutrients, grow and replicate. *eLife* **6**, e23217 (2017).
393.Ito, D. *et al* An essential dual-function complex mediates erythrocyte invasion and channel-mediated nutrient uptake in malaria parasites. *eLife* **6**, e23485 (2017).
394.Sherling, E. S. *et al*. The *Plasmodium falciparum* rhoptry protein RhopH3 plays essential roles in host cell invasion and nutrient uptake. *eLife* **6**, e23239 (2017).
395.Schureck, M. A. *et al*. Malaria parasites use a soluble RhopH complex for erythrocyte invasion and an integral form for nutrient uptake. *eLife* **10**, e65282 (2021).
396.Desai, S. A. Why do malaria parasites increase host erythrocyte permeability? *Trends Parasitol*. **30**, 151–159 (2014).
397.Counihan, N. A. *et al.* How malaria parasites acquire nutrients from their host. *Front. Cell Dev. Biol*. **9**, 649184 (2021).
398.Winterberg, M. *et al*. Chemical activation of a high-affinity glutamate transporter in human erythrocytes and its implications for malaria-parasite-induced glutamate uptake. *Blood* **119**, 3604–3612 (2012).
399.Staines, H. M. *et al*. Electrophysiological studies of malaria parasite-infected erythrocytes: Current status. *Int. J. Parasitol*. **37**, 475–482 (2007).
400.Bouyer, G. *et al*. Erythrocyte peripheral type benzodiazepine receptor/voltage-dependent anion channels are upregulated by *Plasmodium falciparum*. *Blood* **118**, 2305–2312 (2011).
401. Gupta, A. *et al*. Complex nutrient channel phenotypes despite Mendelian inheritance in a *Plasmodium falciparum* genetic cross. *PLOS Pathog*. **16**(2), e1008363 (2020).
402.Meibalan, E. & Marti, M. Biology of malaria transmission. *Cold Spring Harb. Perspect. Med*. **7**, a025452 (2017).

403. Bancells, C. *et al*. Revisiting the initial steps of sexual development in the malaria parasite *Plasmodium falciparum*. *Nat. Microbiol*. **4**, 144–154 (2019).
404. Cowman, A. F. *et al.* The molecular basis of erythrocyte invasion by malaria parasites. *Cell Host Microbe* **22**, 232–245 (2017).
405. Reed, M. B. *et al*. Targeted disruption of an erythrocyte binding antigen in *Plasmodium falciparum* is associated with a switch toward a sialic acid-independent pathway of invasion. *Proc. Natl. Acad. Sci. USA*. **97**, 7509–7514 (2000).
406. Wright, G. J. & Rayner, J. C. *Plasmodium falciparum* erythrocyte invasion: Combining function with immune evasion. *PLoS Pathog*. **10**(3), e1003943 (2014).
407. Stubbs, J. *et al*. Microbiology: Molecular mechanism for switching of *P. falciparum* invasion pathways into human erythrocytes. *Science* **309**, 1384–1387 (2005).
408. Duraisingh, M. T. *et al.* Phenotypic variation of *P. falciparum* merozoite proteins directs receptor targeting for invasion of human erythrocytes. *EMBO J*. **22**, 1047–1057 (2003).
409. Cortés, A. Switching *Plasmodium falciparum* genes on and off for erythrocyte invasion. *Trends Parasitol.* **24**, 517–524 (2008).
410. Ward, P. *et al.* Protein kinases of the human malaria parasite *Plasmodium falciparum*: The kinome of a divergent eukaryote. *BMC Genomics* **5**, 79 (2004).
411. Davies, H. *et al.* An exported kinase family mediates species-specific erythrocyte remodelling and virulence in human malaria. *Nat. Microbiol.* **5**, 848-863 (2020)
412. Siddiqui, G. *et al.* Identification of essential exported *P. falciparum* protein kinases in malaria-infected red blood cells. *Br. J. Haematol*. **188**, 774–783 (2020).
413. Nunes, M. C. *et al. Plasmodium falciparum fikk* kinase members target distinct components of the erythrocyte membrane. *PLoS One* **5**, e11747 (2010).
414. Sheokand, P. K. *et al.* Glm S mediated knock-down of a phospholipase expedite alternate pathway to generate phosphocholine required for phosphatidylcholine synthesis in *Plasmodium falciparum*. *Biochem. J*. **478**(18), 3429-3444 (2021)
415. Shears, M. J. *et al.* Fatty acid metabolism in the *Plasmodium* apicoplast: Drugs, doubts and knockouts. *Mol. Biochem. Parasitol.* **199**, 34–50 (2015).
416. Bethke, L. L. *et al.* Duplication, gene conversion, and genetic diversity in the species-specific acyl-CoA synthetase gene family of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **150**, 10–24 (2006).
417. Winter, G. *et a*l. SURFIN is a polymorphic antigen expressed on *Plasmodium falciparum* merozoites and infected erythrocytes. *J. Exp. Med*. **201**, 1853–1863 (2005).
418. Mphande, F. A. et al. SURFIN4.1, a schizont-merozoite associated protein in the SURFIN family of *Plasmodium falciparum. Malar. J*. **7**, 116 (2008).
419. Chan, J. A. *et al.* Surface antigens of *P. falciparum*-infected erythrocytes as immune targets and malaria vaccine candidates. *Cell. Mol. Life Sci*. **71**, 3633–3657 (2014).
420. Arredondo, S. A. & Kappe, S. H. I. The s48/45 six-cysteine proteins: mediators of interaction throughout the *Plasmodium* life cycle. *Int. J. Parasitol*. **47**, 409–423 (2017).
421. Molina-Cruz, A. *et al*. The human malaria parasite *Pfs47* gene mediates evasion of the mosquito immune system. *Science* **340**, 984–987 (2013).

422. Kirk, K. Glucose uptake in *Plasmodium falciparum*-infected erythrocytes is an equilibrative not an active process. *Mol. Biochem. Parasitol*. **82**, 195–205 (1996).
423. Ceriello, A. *et al*. Glycaemic variability in diabetes: clinical and therapeutic implications. *Lancet Diabetes Endocrinol*. **7**, 221–230 (2019).
424. Fang, J. *et al*. Ambient glucose concentration and gene expression in *Plasmodium falciparum*. *Mol. Biochem. Parasitol*. **133**, 125–129 (2004).
425. White, N. J. & Ho, M. The pathophysiology of malaria. *Advances in Parasitology* **31**, 83–173 (1992).
426. Fang, J. *et al.* The effects of glucose concentration on the reciprocal regulation of rRNA promoters in *Plasmodium falciparum*. *J. Biol. Chem.* **279**, 720–725 (2004).
427. Kumar, N. *et al*. Induction and localization of *P. falciparum* stress proteins related to the heat shock protein 70 family. *Mol. Biochem. Parasitol*. **48**, 47–58 (1991).
428. Humeida, H. *et al*. The effect of glucose and insulin on in vitro proliferation of *Plasmodium falciparum*. *J. Diabetol*. **2**, 6 (2011).
429. Thien, H. V. *et al.* Hypoglycemia in *falciparum* malaria: is fasting an unrecognized and insufficiently emphasized risk factor? *Trends Parasitol.* **22**, 410–415 (2006).
430. Rovira-Graells, N. *et al*. New assays to characterise growth-related phenotypes of *Plasmodium falciparum* reveal variation in density-dependent growth inhibition between parasite lines. *PLoS One* **11**, e0165358 (2016).
431. Nötzel, C. & Kafsack, B. F. C. There and back again: malaria parasite single-cell transcriptomics comes full circle. *Trends Parasitol*. **37**, 850-852 (2021)
432. Ahmad, M. *et al*. Live-cell FRET reveals that malaria nutrient channel proteins CLAG3 and RhopH2 remain associated throughout their tortuous trafficking. *mBio* **11**, e01354-20 (2020).
433. Ghoneim, A. M. Recombinant expression of the rhoptry protein CLAG3.1 as a green fluorescent chimeric protein in *Plasmodium falciparum*. *Delta J. Sci*. **36**, 201–211 (2013).
434. Lee, M. C. S. *et al.* Cutting back malaria: CRISPR/Cas9 genome editing of *Plasmodium*. *Brief. Funct. Genomics* **18**, 281–289 (2019).
435. Ghorbal, M. *et al*. Genome editing in the human malaria parasite *Plasmodium falciparum* using the CRISPR-Cas9 system. *Nat. Biotechnol.* **32**, 819–821 (2014).
436. Lim, M. Y. X. *et al*. UDP-galactose and acetyl-CoA transporters as *Plasmodium* multidrug resistance genes. *Nat. Microbiol*. **1**, 16166 (2016).
437. Knuepfer, E. *et al*. Generating conditional gene knockouts in *Plasmodium* – a toolkit to produce stable DiCre recombinase-expressing parasite lines using CRISPR/Cas9. *Sci. Rep*. **7**, 3881 (2017).
438. Tyler, J. S. *et al.* Focus on the ringleader: the role of AMA1 in apicomplexan invasion and replication. T*rends Parasitol*. **27**, 410–420 (2011).
439. Treeck, M. *et al.* Functional analysis of the leading malaria vaccine candidate AMA-1 reveals an essential role for the cytoplasmic domain in the invasion process. *PLoS Pathog.* **5**, e1000322 (2009).

440.Portugaliza, H. P. *et al.* Reporter lines based on the *gexp02* promoter enable early quantification of sexual conversion rates in the malaria parasite *Plasmodium falciparum*. *Sci. Rep.* **9**, 14595 (2019).
441. Ishino, T. *et al*. LISP1 is important for the egress of *Plasmodium berghei* parasites from liver cells. *Cell. Microbiol*. **11**, 1329–1339 (2009).
442.Bártfai, R. *et al*. H2A.Z demarcates intergenic regions of the *Plasmodium falciparum* epigenome that are dynamically marked by H3K9ac and H3K4me3. *PLoS Pathog.* **6**, e1001223 (2010).
443.Adjalley, S. H. *et al.* Landscape and dynamics of transcription initiation in the malaria parasite *Plasmodium falciparum*. *Cell Rep*. **14**, 2463–2475 (2016).
444.Marti, M. *et al*. Targeting malaria virulence and remodeling proteins to the host erythrocyte. *Science* **306**, 1930–3 (2004).
445.Buchholz, K. *et al.* A high-throughput screen targeting malaria transmission stages opens new avenues for drug development. *J. Infect. Dis*. **203**, 1445–1453 (2011).
446.Day, R. N. & Davidson, M. W. The fluorescent protein palette: tools for cellular imaging. *Chem. Soc. Rev*. **38**, 2887 (2009).
447.Casas-Vila, N. *et al*. Transcriptional analysis of tightly synchronized *Plasmodium falciparum* intraerythrocytic stages by RT-qPCR. *Methods Mol. Biol*. **2369**, 165–185 (2021).
448.Ribeiro, J. M. *et al.* Guide RNA selection for CRISPR-Cas9 transfections in *Plasmodium falciparum*. *Int. J. Parasitol*. **48**, 825–832 (2018).
449.Basore, K. *et al.* How do antimalarial drugs reach their intracellular targets? *Front. Pharmacol.* **6**, 91 (2015).
450.Baumeister, S. *et al*. Fosmidomycin uptake into *Plasmodium* and *Babesia*-infected erythrocytes is facilitated by parasite-induced new permeability pathways. *PLoS One* **6**, e19334 (2011).
451. Lisk, G. *et al*. Changes in the Plasmodial Surface Anion Channel reduce leupeptin uptake and can confer drug resistance in *Plasmodium falciparum*-infected erythrocytes. *Antimicrob. Agents Chemother*. **52**, 2346–2354 (2008).
452.Stead, A. M. W. *et al.* Diamidine compounds: Selective uptake and targeting in *Plasmodium falciparum*. *Mol. Pharmacol.* **59**, 1298–1306 (2001).
453.Wein, S. *et al*. Transport and pharmacodynamics of albitiazolium, an antimalarial drug candidate. *Br. J. Pharmacol.* **166**, 2263–2276 (2012).
454.Sigala, P. A. *et al.* Deconvoluting heme biosynthesis to target blood-stage malaria parasites. *eLife* **4**, e09143 (2015).
455.Mackinnon, M. J. & Marsh, K. The selection landscape of malaria parasites. *Science* **328**, 866–871 (2010).
456.Kussell, E. Phenotypic diversity, population growth, and information in fluctuating environments. *Science* **309**, 2075–2078 (2005).
457.Prudêncio, M., Rodriguez, A. & Mota, M. M. The silent path to thousands of merozoites: the *Plasmodium* liver stage. *Nat. Rev. Microbiol*. **4**, 849–856 (2006).

458.Kaestli, M. *et al.* Longitudinal assessment of *Plasmodium falciparum var* gene transcription in naturally infected asymptomatic children in Papua New Guinea. *J. Infect. Dis*. **189**, 1942–1951 (2004).

459.Jamrozik, E. & Selgelid, M. J. Ethical issues surrounding controlled human infection challenge studies in endemic low-and middle-income countries. *Bioethics* **34**, 797–808 (2020).

460.Njue, M. *et al*. Ethical considerations in Controlled Human Malaria Infection studies in low resource settings: Experiences and perceptions of study participants in a malaria Challenge study in Kenya. *Wellcome Open Res*. **3**, 39 (2018).

461. Jensen, A. R. *et al.* Cerebral *Plasmodium falciparum* malaria: The role of PfEMP1 in its pathogenesis and immunity, and PfEMP1-based vaccines to prevent it. *Immunol. Rev*. **293**, 230–252 (2020).

462.Abdi, A. I. *et al. Plasmodium falciparum* malaria parasite *var* gene expression is modified by host antibodies: longitudinal evidence from controlled infections of Kenyan adults with varying natural exposure. *BMC Infect. Dis*. **17**, 585 (2017).

463.Duncan, C. J. A. & Draper, S. J. Controlled human blood stage malaria infection: current status and potential applications. *Am. J. Trop. Med. Hyg*. **86**, 561–565 (2012).

464.Alkema, M. *et al.* A randomized clinical trial to compare *Plasmodium falciparum* gametocytemia and infectivity after blood-stage or mosquito bite–induced controlled malaria infection. *J. Infect. Dis*. **224**, 1257-1265 (2020)

465.Snounou, G. & Pérignon, J.-L. Malariotherapy – insanity at the service of malariology. *Adv. Paras.* **81** 223–255 (2013).

466.Desai, S. A. Ion and nutrient uptake by malaria parasite-infected erythrocytes. *Cell. Microbiol*. **14**, 1003–1009 (2012).

467.Wein, S. *et al*. High accumulation and in vivo recycling of the new antimalarial Albitiazolium lead torapid parasite death. *Antimicrob. Agents Chemother*. **61**, e00352-17 (2017).

468.Ashley, E. A. & White, N. J. The duration of *P. falciparum* infections. *Malar. J*. **13**, 500 (2014).

469.Koepfli, C. & Mueller, I. Malaria epidemiology at the clone level. *Trends Parasitol*. **33**, 974–985 (2017).

470.Auburn, S. & Barry, A. E. Dissecting malaria biology and epidemiology using population genetics and genomics. *Int. J. Parasitol.* **47**, 77–85 (2017).

471. Greischar, M. A. *et al*. The challenge of quantifying synchrony in malaria parasites. *Trends Parasitol*. **35**, 341-355 (2019).

472.Gelb, M. H. Drug discovery for malaria: a very challenging and timely endeavor. *Curr. Opin. Chem. Biol.* **11**, 440–445 (2007).

473.Desai, S. A. Insights gained from *P. falciparum* cultivation in modified media. *Sci. World J*. **2013**, 363505 (2013).

474.Johannes, F. & Schmitz, R. J. Spontaneous epimutations in plants. *New Phytol*. **221**, 1253–1259 (2019).

475. Biwer, C. *et al*. The role of stochasticity in the origin of epigenetic variation in animal populations. *Integr. Comp. Biol*. **60**, 1544–1557 (2020).
476. Gómez-Díaz, E. *et al*. Epigenetic regulation of *P. falciparum* clonally variant gene expression during development in *Anopheles gambiae*. *Sci. Rep*. **7**, 40655 (2017).
477. Witmer, K. *et al*. An epigenetic map of malaria parasite development from host to vector. *Sci. Rep*. **10**, 6354 (2020).
478. Zanghì, G. *et al*. A specific PfEMP1 is expressed in *P. falciparum* sporozoites and plays a role in hepatocyte infection. *Cell Rep*. **22**, 2951–2963 (2018).
479. Voss, T. S. *et al.* A *var* gene promoter controls allelic exclusion of virulence genes in *Plasmodium falciparum* malaria. *Nature* **439**, 1004–1008 (2006).
480. Lavazec, C. Molecular mechanisms of deformability of *Plasmodium*-infected erythrocytes. *Curr. Opin. Microbiol.* **40**, 138–144 (2017).
481. Bachmann, A. *et al*. A comparative study of the localization and membrane topology of members of the RIFIN, STEVOR and PfMC-2TM protein families in *Plasmodium falciparum*-infected erythrocytes. *Malar. J*. **14**, 274 (2015).
482. Quintana, M. D. P. *et al.* Antibodies in children with malaria to PfEMP1, RIFIN and SURFIN expressed at the *Plasmodium falciparum* parasitized red blood cell surface. *Sci. Rep*. **8**, 3262 (2018).
483. Kanoi, B. N. *et al*. Global repertoire of human antibodies against *Plasmodium falciparum* RIFINs, SURFINs, and STEVORs in a malaria exposed population. *Front. Immunol*. **11**, 893 (2020).
484. Kumar, V. *et al.* PHISTc protein family members localize to different subcellular organelles and bind *Plasmodium falciparum* major virulence factor PfEMP1. *FEBS J.* **285**, 294–312 (2018).
485. Oberli, A. *et al.* A *Plasmodium falciparum* PHIST protein binds the virulence factor PfEMP1 and comigrates to knobs on the host cell surface. *FASEB J*. **28**, 4420–4433 (2014).
486. Tarr, S. J. *et al*. A conserved domain targets exported PHISTb family proteins to the periphery of *Plasmodium*-infected erythrocytes. *Mol. Biochem. Parasitol*. **196**, 29–40 (2014).
487. Richards, E. J. Inherited epigenetic variation — revisiting soft inheritance. *Nat. Rev. Genet.* **7**, 395–401 (2006).
488. Struhl, K. & Segal, E. Determinants of nucleosome positioning. *Nat. Struct. Mol. Biol*. **20**, 267–273 (2013).
489. Hackett, J. A. *et al*. Parallel mechanisms of epigenetic reprogramming in the germline. *Trends Genet*. **28**, 164–174 (2012).
490. Zhou, L. & Dean, J. Reprogramming the genome to totipotency in mouse embryos. *Trends Cell Biol*. **25**, 82–91 (2015).
491. Ross, P. J. & Canovas, S. Mechanisms of epigenetic remodelling during preimplantation development. *Reprod. Fertil. Dev*. **28**, 25 (2016).
492. Hajkova, P. *et al.* Chromatin dynamics during epigenetic reprogramming in the mouse germ line. *Nature* **452**, 877–881 (2008).

493.Grafi, G. How cells dedifferentiate: a lesson from plants. *Dev. Biol.* **268**, 1–6 (2004).
494.Williams, L. *et al*. Chromatin reorganization accompanying cellular dedifferentiation is associated with modifications of histone H3, redistribution of HP1, and activation of E2F-target genes. *Dev. Dyn*. **228**, 113–120 (2003).
495.Baton, L. A. & Ranford-Cartwright, L. C. Spreading the seeds of million-murdering death: Metamorphoses of malaria in the mosquito. *Trends Parasitol*. **21**, 573–580 (2005).
496.Sinden, R. E. & Hartley, R. H. Identification of the meiotic division of malarial parasites. *J. Protozool*. **32**, 742–744 (1985).
497.Walliker, D. Implications of genetic exchange in the study of protozoan infections. *Parasitology* **99**, S49–S58 (1989).
498.Voorberg-van der Wel, A. *et al*. Modeling relapsing malaria: emerging technologies to study parasite-host interactions in the liver. *Front. Cell. Infect. Microbiol.* **10**, 606033 (2021).
499.Prudêncio, M. *et al*. A toolbox to study liver stage malaria. *Trends Parasitol.* **27**, 565–574 (2011).
500.Frevert, U. Sneaking in through the back entrance: the biology of malaria liver stages. *Trends Parasitol*. **20**, 417–424 (2004).
501. Llorà-Batlle, O. *et al.* Conditional expression of PfAP2-G for controlled massive sexual conversion in *Plasmodium falciparum*. *Sci. Adv*. **6**, eaaz5057 (2020).
502.Fougère, A. *et al*. Variant exported blood-stage proteins encoded by *Plasmodium* multigene families are expressed in liver stages where they are exported into the parasitophorous vacuole. *PLOS Pathog*. **12**, e1005917 (2016).
503.Caldelari, R. *et al*. Transcriptome analysis of *Plasmodium berghei* during exo-erythrocytic development. *Malar. J.* **18**, 330 (2019).
504.Ramakrishnan, S. *et al*. Lipid synthesis in protozoan parasites: A comparison between kinetoplastids and apicomplexans. *Prog. Lipid Res.* **52**, 488–512 (2013).
505.Ben Mamoun, C. *et al*. Targeting the lipid metabolic pathways for the treatment of malaria. *Drug Dev. Res.* 71, 44-55 (2009).
506.Tokumasu, F. *et al*. Creative interior design by *Plasmodium falciparum*: Lipid metabolism and the parasite's secret chamber. *Parasitol. Int.* **83**, 102369 (2021).
507.Botte, C. Y. *et al.* Atypical lipid composition in the purified relict plastid (apicoplast) of malaria parasites. *Proc. Natl. Acad. Sci. USA*. **110**, 7506–7511 (2013).
508.Gulati, S. *et al.* Profiling the essential nature of lipid metabolism in asexual blood and gametocyte stages of *Plasmodium falciparum*. *Cell Host Microbe* **18**, 371–381 (2015).
509.Mi-Ichi, F. *et al.* Intraerythrocytic *P. falciparum* utilize a broad range of serum-derived fatty acids with limited modification for their growth. *Parasitology* **133**, 399–410 (2006).
510. Mi-Ichi, F. *et al.* Oleic acid is indispensable for intraerythrocytic proliferation of *Plasmodium falciparum*. *Parasitology* **134**, 1671–1677 (2007).
511. Mitamura, T. & Palacpac, N. M. Q. Lipid metabolism in *Plasmodium falciparum*-infected erythrocytes: possible new targets for malaria chemotherapy. *Microbes Infect.* **5**, 545–552 (2003).

512. Ke, H. *et al.* Mitochondrial type II NADH dehydrogenase of *Plasmodium falciparum* (PfNDH2) is dispensable in the asexual blood stages. *PLoS One* **14**, e0214023 (2019).
513. Boysen, K. E. & Matuschewski, K. Arrested oocyst maturation in *Plasmodium* parasites lacking type II NADH:Ubiquinone Dehydrogenase. *J. Biol. Chem.* **286**, 32661–32671 (2011).
514. MacRae, J. I. *et al*. Mitochondrial metabolism of sexual and asexual blood stages of the malaria parasite *Plasmodium falciparum*. *BMC Biol*. **11**, 67 (2013).
515. Painter, H. J. *et al.* Specific role of mitochondrial electron transport in blood-stage *Plasmodium falciparum*. *Nature* **446**, 88–91 (2007).
516. Ahouidi, A. *et al.* An open dataset of *Plasmodium falciparum* genome variation in 7,000 worldwide samples. *Wellcome Open Res*. **6**, 42 (2021).
517. van Schaijk, B. C. L. *et al.* Type II fatty acid biosynthesis is essential for *Plasmodium falciparum* sporozoite development in the midgut of *Anopheles* mosquitoes. *Eukaryot. Cell* **13**, 550–559 (2014).
518. Vaughan, A. M. *et al.* Type II fatty acid synthesis is essential only for malaria parasite late liver stage development. *Cell. Microbiol*. **11**, 506–520 (2009).
519. Yu, M. *et al.* The fatty acid biosynthesis enzyme FabI plays a key role in the development of liver-stage malarial parasites. *Cell Host Microbe* **4**, 567–578 (2008).
520. Desai, S. A. *et al.* A voltage-dependent channel involved in nutrient uptake by red blood cells infected with the malaria parasite. *Nature* **406**, 1001–1005 (2000).
521. Egan, W. J. *et al*. Prediction of drug absorption using multivariate statistics. *J. Med. Chem*. **43**, 3867–3877 (2000).
522. Lisk, G. *et al.* Altered plasmodial surface anion channel activity and in vitro resistance to permeating antimalarial compounds. *Biochim. Biophys. Acta*. **1798**, 1679–1688 (2010).
523. Blais, J. *et al.* Azithromycin uptake and intracellular accumulation by *Toxoplasma gondii*-infected macrophages. *J. Antimicrob. Chemother*. **34**, 371–382 (1994).
524. Hill, D. A. *et al.* A blasticidin S-resistant *Plasmodium falciparum* mutant with a defective plasmodial surface anion channel. *Proc. Natl. Acad. Sci. USA*. **104**, 1063–1068 (2007).
525. Chopra, I. Molecular mechanisms involved in the transport of antibiotics into bacteria. *Parasitology* **96**, S25–S44 (1988).
526. Sharma, P. *et al.* A CLAG3 mutation in an amphipathic transmembrane domain alters malaria parasite nutrient channels and confers leupeptin resistance. *Infect. Immun*. **83**, 2566–2574 (2015).
527. Gupta, A. *et al*. CLAG3 self-associates in malaria parasites and quantitatively determines nutrient uptake channels at the host membrane. *mBio* **9**, e02293-17 1–18 (2018).
528. Nguitragool, W. *et al*. Proteolysis at a specific extracellular residue implicates integral membrane CLAG3 in malaria parasite nutrient channels. *PLoS One* **9**, e93759 (2014).

529.Ferru, E. *et al.* Regulation of membrane-cytoskeletal interactions by tyrosine phosphorylation of erythrocyte band 3. *Blood* **117**, 5998–6006 (2011).
530.Pantaleo, A. *et al*. Current knowledge about the functional roles of phosphorylative changes of membrane proteins in normal and diseased red cells. *J. Proteomics* **73**, 445–455 (2010).
531. Chourasia, B. K. *et al. P. falciparum* Clag9-associated PfRhopH complex is involved in merozoite binding to human erythrocytes. *Infect. Immun.* **88**, e00504-19 (2019).
532.Trenholme, K. R. *et al. clag9*: A cytoadherence gene in *P. falciparum* essential for binding of parasitized erythrocytes to CD36. *Proc. Natl. Acad. Sci*. **97**, 4029–4033 (2000).
533.Ho, C.-M. *et al.* Native structure of the RhopH complex, a key determinant of malaria parasite nutrient acquisition. *Proc. Natl. Acad. Sci*. **118**, e2100514118 (2021).
534.Unwin, N. The structure of ion channels in membranes of excitable cells. *Neuron* **3**, 665–676 (1989).
535.Otto, T. D. *et al*. Genomes of all known members of a *Plasmodium* subgenus reveal paths to virulent human malaria. *Nat. Microbiol*. **3**, 687–697 (2018).
536.LeRoux, M. *et al. Plasmodium falciparum* biology: analysis of in vitro versus in vivo growth conditions. *Trends Parasitol.* **25**, 474–481 (2009).
537.Bachmann, A. *et al.* Temporal expression and localization patterns of variant surface antigens in clinical *Plasmodium falciparum* isolates during erythrocyte schizogony. *PLoS One* **7**, e49540 (2012).
538.Daily, J. P. *et al.* In vivo Transcriptome of *P. falciparum* reveals overexpression of transcripts that encode surface proteins. *J. Infect. Dis*. **191**, 1196–1203 (2005).
539.Yamaga, K. M. *et al. Plasmodium falciparum*: Comparative analysis of antigens from continuous in vitro cultured and in vivo derived malarial parasites. *Exp. Parasitol.* **58**, 138–146 (1984).
540.Chen, K. *et al.* Competition between *Plasmodium falciparum* strains in clinical infections during in vitro culture adaptation. *Infect. Genet. Evol.* **24**, 105–110 (2014).
541. Peters, J. M. *et al.* Differential changes in *Plasmodium falciparum var* transcription during adaptation to culture. *J. Infect. Dis*. **195**, 748–755 (2007).
542.Hoo, R. *et al*. Transcriptome profiling reveals functional variation in *Plasmodium falciparum* parasites from controlled human malaria infection studies. *EBioMedicine* **48**, 442–452 (2019).
543.Zhang, Q. *et al.* From in vivo to in vitro: Dynamic analysis of *Plasmodium falciparum var* gene expression patterns of patient isolates during adaptation to culture. *PLoS One* **6**, e20591 (2011).
544.Nyarko, P. B. *et al.* Investigating a *Plasmodium falciparum* erythrocyte invasion phenotype switch at the whole transcriptome level. *Sci. Rep*. **10**, 245 (2020).
545.Lapp, S. A. *et al*. *Plasmodium knowlesi* gene expression differs in ex vivo compared to in vitro blood-stage cultures. *Malar. J.* **14**, 110 (2015).
546.Lee, H. J. *et al.* Transcriptomic studies of malaria: a paradigm for investigation of systemic host-pathogeninteractions. *Microbiol. Mol. Biol. Rev*. **82**, e00071-17 (2018).

547.Druilhe, P. *et al*. A primary malarial infection is composed of a very wide range of genetically diverse but related parasites. *J. Clin. Invest*. **101**, 2008–2016 (1998).
548.Yuan, L. *et al*. *P. falciparum* populations from northeastern Myanmar display high levels of genetic diversity at multiple antigenic loci. *Acta Trop*. **125**, 53–59 (2013).
549.Murray, L. *et al*. Microsatellite genotyping and genome-wide single nucleotide polymorphism-based indices of *Plasmodium falciparum* diversity within clinical infections. *Malar. J*. **15**, 275 (2016).
550.Nkhoma, S. C. *et al.* Close kinship within multiple-genotype malaria parasite infections. *Proc. Biol. Sci*. **279**, 2589–2598 (2012).
551. Lavstsen, T. *et al*. Expression of *Plasmodium falciparum* erythrocyte membrane protein 1 in experimentally infected humans. *Malar. J*. **4**, 21 (2005).
552.Nery, S. *et al*. Expression of *Plasmodium falciparum* genes involved in erythrocyte invasion varies among isolates cultured directly from patients. *Mol. Biochem. Parasitol.* **149**, 208–215 (2006).
553.Foley, D. L. *et al*. Prospects for epigenetic epidemiology. *Am. J. Epidemiol.* **169**, 389–400 (2008).
554.Ogutu, B. R. *et al*. Antimalarial drug sensitivity profile of Western Kenya *Plasmodium falciparum* field isolates determined by a SYBR Green I in vitro assay and molecular analysis. *Am. J. Trop. Med. Hyg*. **85**, 34–41 (2011).
555.Chaorattanakawee, S. *et al*. Attenuation of *Plasmodium falciparum* in vitro drug resistance phenotype following culture adaptation compared to fresh clinical isolates in Cambodia. *Malar. J*. **14**, 486 (2015).

ACKNOWLEDGMENTS

First and foremost, I would like to express my most profound gratitude to Dr Alfred Cortés for supervising my thesis and giving me his total support throughout my entire PhD. I would not be here today without his guidance, invaluable insight and unending patience. He has shown me how to maintain an optimistic outlook when facing all of the different challenges that come with scientific research. He has my utmost respect as a scientist and mentor.

I would like to extend my heartfelt appreciation to ISGlobal for providing such a favourable environment in which to carry out a PhD and for providing me with funding support including my mobility grant, for all of which I am extremely grateful. I would especially like to thank the ISGlobal personnel in the human resources, purchasing, legal, accounting and training and education departments for all of their help with the "non-scientific" aspects of my thesis. They have all made many day-to-day issues a lot easier to deal with.

With specific reference to the University of Barcelona, I would primarily like to express my sincerest gratitude to Dr Gemma Marfany Nadal for being my tutor and for her positive feedback and constructive criticism of my work. Similarly, I would also like to thank the other members of my follow-up committee Dr. Quique Bassat and Dr. Montse Corominas, as well as Àngels Serra at the Biomedicine doctorate programme for kindly helping me through the final difficult stages of depositing my thesis with all of the paperwork that entails.

I would like to thank the AGAUR and the European Social Fund for providing me with my FI grant that has funded most of my PhD.

I am also extremely grateful to Dr. Catherine Lavazec and the other members of the Lavazec team– Florian, Marie Esther, Cyrielle and Daniela, for making me feel so welcome during my research stay in Paris and for helping to make it such an incredible experience both professionally and personally.

A huge thank you to my teammates in the Malaria Epigenetics group: Eli, Oriol, Lucas, Núria R., Harvie, Cristina, Sofia, Núria C., Alba, Carla, Rafa and Rosa. Not only have you all helped me on countless occasions but you have also contributed to making the whole experience both fun and unforgettable. I will always fondly remember the good times we have had together both in and out of the lab.

Many thanks as well to all my other friends and workmates at the CEK building. Nanos, glycos, falci, micro, immuno and Ger, you too have contributed to creating a friendly and enjoyable working environment which has been a pleasure to be a part of. Thank you also to Cristina Canaleta and the other IDIBAPS workers who make working at the CEK that much easier. Also, a special mention to Isabel and Sara at the flow cytometry service for their help and kindness.

Finally, I would like to thank my friends and family for all their love and support with a special mention to my sister Cordelia Pickford who helped me with the graphic design of my thesis.

ABBREVIATIONS & ANNEX

Abbreviations

5-ALA	5-aminolevulinic acid
6-CYS	Six-cysteine
ACBP	Acyl-CoA binding protein
ACS	Acyl-CoA synthase
ACTs	Artemisinin-based combination therapies
AMA1	Apical membrane antigen 1
ApiAP2	Apicomplexan AP2 family
BSA	Bovine serum albumin
BSD	Blasticidin S
CHMI	Controlled human malaria infection
CLAG	Cytoadherence link asexual gene
CNV	Copy number variation
CVG	Clonally variant gene
DSB	Double-stranded break
DTR	Directed transcriptional response
EBA	Erythrocyte-binding antigen
eYFP	Yellow fluorescent protein
GBP	Glycophorin binding protein
gDNA	Genomic DNA
GFP	Green fluorescent protein
H3K9ac	Histone 3, lysine 9 acetylation
H3K9me3	Histone 3, lysine 9 trimethylation
hDHFR	Human dihydrofolate reductase
HP1	Heterochromatin protein 1
HR1	Homology region 1
HR2	Homology region 2
IDC	Intraerythrocytic development cycle
IFA	Immunofluorescence assay
Indel	Insertions and deletions
iRBC	Infected red blood cell
LISP1	Liver specific protein 1
lncRNA	Long non-coding RNA
LysoPC	Lysophosphatidylcholine
mCer	mCerulean
METC	Mitochondrial electron transport chain
ncRNA	Non-coding RNA
NPPs	New permeation pathways
PAC	Puromycin-N-acetyltransferase
PCR	Polymerase chain reaction

PfEMP1	*P. falciparum* erythrocyte membrane protein 1
PfNDH2	*P. falciparum* type II NADH:ubiquinone oxidoreductase
PfRH	*P. falciparum* reticulocyte-binding homologue
PHIST	*Plasmodium* helical interspersed subtelomeric
PSAC	*Plasmodium* surface anion channel
qPCR	Quantitative polymerase chain reaction
RBC	Red blood cell
RDTs	Rapid diagnostic tests
RIF	Rapid interspersed family
RNS	Reactive nitrogen species
ROS	Reactive oxidative species
RT-qPCR	Quantitative reverse transcription polymerase chain reaction
sgRNA	Single guide RNA
SNP	Single-nucleotide polymorphism
STEVOR	Subtelomeric variable open reading frame
SURFIN	Surface-associated interspersed
tdTom	tdTomato
TSS	Transcription start sight
yFCU	Yeast cytosine deaminase/uridyl phosphoribosyl transferase

Annex: other work

During my thesis I also participated in the writing of a methods article on RT-qPCR transcriptional analysis in Plasmodium falciparum, which has been published in *Methods in Molecular Biology*.

Transcriptional Analysis of Tightly Synchronized Plasmodium falciparum Intraerythrocytic Stages by RT-qPCR. Casas-Vila N*, Pickford AK*, Portugaliza HP, Tintó-Font E, Cortés A. *Methods Mol. Biol.* 2369:165-185. (2021) doi: 10.1007/978-1-0716-1681-9_10.
(*Equal contribution as first authors)