



# City Research Online

## City St George's, University of London

**Citation:** Noble, R., Christodoulou, Z., Kyes, S., Pinches, R., Newbold, C. & Recker, M. (2013). The antigenic switching network of Plasmodium falciparum and its implications for the immuno-epidemiology of malaria. *Elife*, 2(2), e01074. doi: 10.7554/elife.01074

This is the published version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

**Permanent repository link:** <https://openaccess.city.ac.uk/id/eprint/24713/>

**Link to published version:** <https://doi.org/10.7554/elife.01074>

**Copyright and Reuse:** Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

# The antigenic switching network of *Plasmodium falciparum* and its implications for the immuno-epidemiology of malaria

Robert Noble<sup>1†</sup>, Zóe Christodoulou<sup>2†</sup>, Sue Kyes<sup>2</sup>, Robert Pinches<sup>2</sup>, Chris I Newbold<sup>2\*</sup>, Mario Recker<sup>1\*</sup>

<sup>1</sup>Department of Zoology, University of Oxford, Oxford, United Kingdom; <sup>2</sup>Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, Oxford, United Kingdom

**Abstract** Antigenic variation in the human malaria parasite *Plasmodium falciparum* involves sequential and mutually exclusive expression of members of the *var* multi-gene family and appears to follow a non-random pattern. In this study, using a detailed in vitro gene transcription analysis of the culture-adapted HB3 strain of *P. falciparum*, we show that antigenic switching is governed by a global activation hierarchy favouring short and highly diverse genes in central chromosomal location. Longer and more conserved genes, which have previously been associated with severe infection in immunologically naive hosts, are rarely activated, however, implying an in vivo fitness advantage possibly through adhesion-dependent survival rates. We further show that a gene's activation rate is positively associated sequence diversity, which could offer important new insights into the evolution and maintenance of antigenic diversity in *P. falciparum* malaria.

DOI: [10.7554/eLife.01074.001](https://doi.org/10.7554/eLife.01074.001)

\*For correspondence: chris.newbold@imm.ox.ac.uk (CIN); mario.recker@zoo.ox.ac.uk (MR)

†These authors contributed equally to this work

**Competing interests:** The authors declare that no competing interests exist.

**Funding:** See page 16

**Received:** 11 June 2013

**Accepted:** 19 August 2013

**Published:** 17 September 2013

**Reviewing editor:** Mercedes Pascual, University of Michigan, United States

© Copyright Noble et al. This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

## Introduction

Acquired antibody-mediated immunity against *Plasmodium falciparum* is directed against the disease-causing, intra-erythrocytic life-stage of the parasite's life-cycle. Continual exposure to *P. falciparum* infection can lead to a form of semi-immunity, whereby protection against life-threatening disease is acquired after a few infections only (Gupta et al., 1999) whereas sterile and long-lasting immunity against infection is never achieved through natural exposure (Greenwood et al., 1987; Marsh, 1992; Tran et al., 2013). The process of natural acquired immunity therefore coincides with a transition from acute and severe infection in young children to asymptomatic carriage in older individuals (Marsh and Snow, 1997). At the core of this variable outcome of infection and poor development of immunity lies the major parasite virulence factor and variant antigen, *P. falciparum* erythrocyte membrane protein 1 (PfEMP1). Members of this family of proteins are expressed on the surface of infected red blood cells and are involved in the binding of parasitised cells to other host cells and tissues (for an overview see Kraemer and Smith, (2006)). The conferred binding phenotypes are believed to enhance parasite survival by avoiding splenic clearance (Boone and Watters, 1995; Buffet et al., 2009) and also contribute to malaria pathology through parasite sequestration in the deep vasculature (Ockenhouse et al., 1991; Pongponratn et al., 1991). Mutually exclusive transcriptional switching between PfEMP1 variants during the course of an infection thus affords the parasite an immune evasion strategy and a means to exploit different host tissues (Roberts et al., 1992; Smith et al., 1995).

Each haploid parasite genome contains approximately 60 members of the *var* gene family (Gardner et al., 2002) encoding distinct PfEMP1 variants, with a high degree of divergence between the antigenic repertoires of any two parasites. The high diversity of *var* genes and *var* gene repertoires is predominantly

**eLife digest** Our ability to acquire immunity to a disease depends on our immune system learning to recognise foreign molecules—called antigens—that are specific to the disease-causing virus, bacterium or parasite. However, some pathogens, such as the malaria-causing parasite *Plasmodium falciparum*, get around this defence through a process called antigenic variation. This involves the parasite switching between different antigens over the course of an infection, preventing the host immune system from learning to recognise them and leading to infections that last many weeks or even months.

The main antigen in *P. falciparum* is a protein called PfEMP1, which is encoded by a family of genes called *var* ('variable'). *Var* genes have evolved to be highly diverse, and different parasites have different repertoires of around 50–60 *var* genes. This ensures that there are a huge number of distinct variants of the PfEMP1 antigen available within the population, allowing the malaria parasite to maintain long-lasting infections and also to infect the same individuals again and again.

Previous work has shown that the expression of *var* genes is not random, but it is not clear what determines which genes are expressed at any given time. Now, Noble et al. have performed a detailed investigation of antigenic switching in *P. falciparum*. Using clonal parasites, they closely monitored the expression of the entire *var* gene repertoire during many generations of parasite culture. They observed that although different cultures initially expressed distinct *var* genes, most of them ended up expressing two particular genes—*var27* and *var29*—at high levels, indicating a hard-wired gene 'activation hierarchy'.

Noble et al. found that whenever the parasites switched antigens, *var* genes that were centrally located on chromosomes—such as *var27* and *var29*—were more likely to be activated than those at the ends of chromosomes. Moreover, *var* genes that were highly diverse were more likely to be activated than more conserved genes: this is the first evidence linking *var* gene evolution with gene activation probabilities. Together, these factors gave rise to the proposed activation hierarchy, which favours genes optimised for immune evasion and aids their continued evolution and diversification. Further work is now needed to identify the molecular mechanisms that control antigenic switching and to determine whether these could represent new therapeutic targets for malaria.

DOI: [10.7554/eLife.01074.002](https://doi.org/10.7554/eLife.01074.002)

generated and maintained by frequent recombination and gene conversion events (*Deitsch et al., 1999; Ward et al., 1999; Freitas-Junior et al., 2000; Taylor et al., 2000; Kraemer et al., 2007; Frank et al., 2008*), which has also resulted in a high degree of mosaicism within these genes (*Ward et al., 1999; Taylor et al., 2000; Bull et al., 2008*). The modular and highly polymorphic, extracellular portion of PfEMP1 is composed of a variable number of Duffy-binding-like domains (DBL; at least two and up to seven per gene) and cysteine-rich interdomain regions (CIDR; up to two per gene) (*Kraemer et al., 2007*). Despite the overall diversity of *var* genes at the sequence level, distinct associations with severe malaria have been identified for several such genes (*Salanti et al., 2003; Jensen et al., 2004*), with the involvement of *var2csa* in pregnancy-associated malaria being the clearest example to date (*Salanti et al., 2004*), as well as associations with particular DBL domains (*Warimwe et al., 2009*) and combinations of DBL and CIDR (*Avril et al., 2012; Claessens et al., 2012; Lavtsen et al., 2012*).

*Var* genes can also be classified according to a conserved upstream promoter sequence into four types termed UpsA, UpsB, UpsC, and UpsE (*Lavtsen et al., 2003*). Of particular interest are the UpsA-type genes, which appear to have diverged from the other groups (*Kraemer et al., 2007; Bull et al., 2008*) and have been shown to be upregulated during severe infections in young hosts (*Jensen et al., 2004; Bull et al., 2005; Kaestli et al., 2006; Kyriacou et al., 2006; Rottmann et al., 2006; Warimwe et al., 2009*). As a possible consequence, acquisition of anti-PfEMP1 immunity appears to follow a particular order, with UpsA variants generally being the first to be broadly recognised in older children (*Warimwe et al., 2009; Cham et al., 2010*). Nevertheless, clear associations between the expression of particular genes or Ups-groups, host age and severity of disease are still missing and could in fact be strain- or location-specific. Understanding the underlying pattern of antigenic switching between *var* genes is therefore important to explain not only the mechanisms and dynamics of persistent

infections but also the relationship between immune-mediated expression of particular PfEMP1 sub-types and infection outcomes.

Various studies have investigated *var* gene expression during natural or experimental infections (Peters et al., 2002; Lavstsen et al., 2005; Wunderlich et al., 2005; Blomqvist et al., 2010). However, as the diversity and order of PfEMP1 variants in the human host is influenced by immune responses and other host factors (Kyriacou et al., 2006; Warimwe et al., 2009), attempts to elucidate the underlying patterns of antigenic change and estimate related switching parameters are commonly based on in vitro cultured parasites in the absence of selection pressure. In this setting, Horrocks et al. (Horrocks et al., 2004) found that the rates at which *var* genes activate or deactivate are non-random, gene-specific and highly dissimilar. It has furthermore been suggested that *var* genes occupying subtelomeric loci tend to switch off faster than those positioned in central chromosomal regions (Frank et al., 2007), offering a possible explanation for the rapid decline in the transcription of subtelomeric genes in parasites during culture adaptation (Peters et al., 2007; Zhang et al., 2011). More recently, Recker et al. (2011) proposed that *var* gene transcriptional change involves a highly structured pathway that has evolved in response to a trade-off between the within-host and the between-host level fitness. However, most of these previous studies were limited in analytical depth due to restrictions in the number of variants considered and/or time points at which gene transcription levels were measured, and could therefore describe only small fragments of the whole switching network.

In this study, we used a novel approach to provide the first detailed characterisation of antigenic switching in the HB3 isolate of *P. falciparum*. We analysed pooled *var* gene transcription data from verified quantitative real time PCR measurements of a diverse set of clonal parasite cultures using a statistically rigorous method of parameter estimation. This revealed a global hierarchy in *var* gene activation favouring a highly diverse set of genes in central chromosomal locations. Our results further suggest a role of active gene transcription in the generation of antigenic diversity and will have important implications for understanding the age- and exposure-dependent pathology of *P. falciparum* malaria.

## Results

11 subclones of the lab-adapted HB3 isolate of *P. falciparum* were obtained by serial dilution of the parent culture. The clones were cultivated and transcript levels of 38 *var* genes measured by quantitative PCR, first at 29–33 parasite generations (8–9 weeks) after cloning and then at several time points over the following 60–79 generations in each culture (Table 1); a detailed description can be found in the 'Materials and methods' section. These 38 genes comprised almost all known functional HB3 *var* genes (Kraemer et al., 2007; Rowe et al., 2009) and one partial gene. The HB3 parent culture was also observed at seven time points. At the first observation, each of the 11 clones predominantly transcribed one of the nine different *var* genes, which were most likely the genes transcribed at the time of cloning. Importantly, these nine genes—hereafter referred to as 'starter genes'—were representative of the *var* repertoire diversity in terms of chromosomal location and Ups type.

### Var gene transcription dynamics

The proportional transcript levels of the different starter genes declined at notably different rates over the course of the experiment (Figure 1, Figure 1—figure supplements 1 and 2). For example, the centrally located, UpsC starter gene *var29* in culture #8 accounted for more than 90% of the total transcript level at the first measurement and remained the dominant variant over the following 60 generations. In contrast, the transcript level of the *var35* starter gene of culture #11 (subtelomeric, UpsA promoter type) was already below 50% of the total transcript level at the point of the first observation (approximately 30 generations post cloning) and continued to decline during the time course of culture, to be eventually replaced by other genes as dominant transcripts. Similar patterns were seen in cultures #3 and #7, which both had centrally located, UpsC starter genes, whereas the remaining five starter genes showed intermediate rates of decline.

Of particular interest was the apparent convergence of most cultures towards high transcription of *var27* and/or *var29* (Figure 1, Figure 1—figure supplements 1 and 2), which are both situated in central chromosomal location but have different promoter types (UpsB and UpsC, respectively). In light of the very different initial conditions, in terms of *var* gene transcription levels at which these cultures were established, this indicates that antigenic switching in *P. falciparum* might be governed by an activation hierarchy.

**Table 1.** Parasite culture and RNA sampling information

Culture	Starter gene	Observed generations
1	<i>var11</i>	29, 48, 59, 69, 89
2	<i>var13</i>	29, 48, 69, 89, 99, 108
3	<i>var13</i>	30, 49, 70, 90, 100, 109
4	<i>var5</i>	30, 39, 49, 60, 70, 90
5	<i>var27</i>	30, 49, 70, 90
6	<i>var27</i>	29, 38, 48, 69, 89
7	<i>var28</i>	33, 52, 63, 73, 93
8	<i>var29</i>	30, 49, 70, 80
9	<i>var30</i>	29, 48, 59, 69, 89
10	<i>var31</i>	29, 48, 59, 69, 89, 99, 108
11	<i>var35</i>	32, 51, 65, 72, 92
Parent	–	0, 9, 19, 40, 60, 70, 79

DOI: [10.7554/eLife.01074.003](https://doi.org/10.7554/eLife.01074.003)

To fully understand the patterns of antigenic switching underlying the observed transcriptional changes in our parasite cultures, we estimated relevant switch parameters using a previously described Markov chain Monte Carlo (MCMC) method (**Noble and Recker, 2012**) ('Materials and methods'). We have previously proposed that *var* gene switching can be fully described dynamically by two sets of parameters ('Materials and methods' and [**Recker et al., 2011**]). First, each gene *i* has an off-rate,  $\omega_i$ , which is the per generation probability of switching from active to silenced. Second, each gene has a switch bias,  $\beta_{ij}$ , which is the probability that when gene *i* switches off, gene *j* becomes activated. We applied our algorithm to the full 38-gene data set and separately to a reduced set of the 16 most transcribed genes, and both models generally gave good and similar fits between the data and predicted transcript levels (**Figure 1—figure supplements 1 and 2**), with deviations resembling stochastic measurement errors.

Our method provided estimated switch parameters for the entire *var* repertoire, which are illustrated in **Figure 2** for the 16 most highly transcribed genes. Due to higher signal to noise ratios, parameter estimates of the starter genes (indicated in red) were more accurate and resulted in much narrower credible intervals, as indicated by less fuzzy rings that represent respective switch probabilities. As clearly demonstrated in **Figure 2**, there was wide distribution in switch biases and off-rates, and we next analysed for specific associations between these parameters and other gene characteristics.

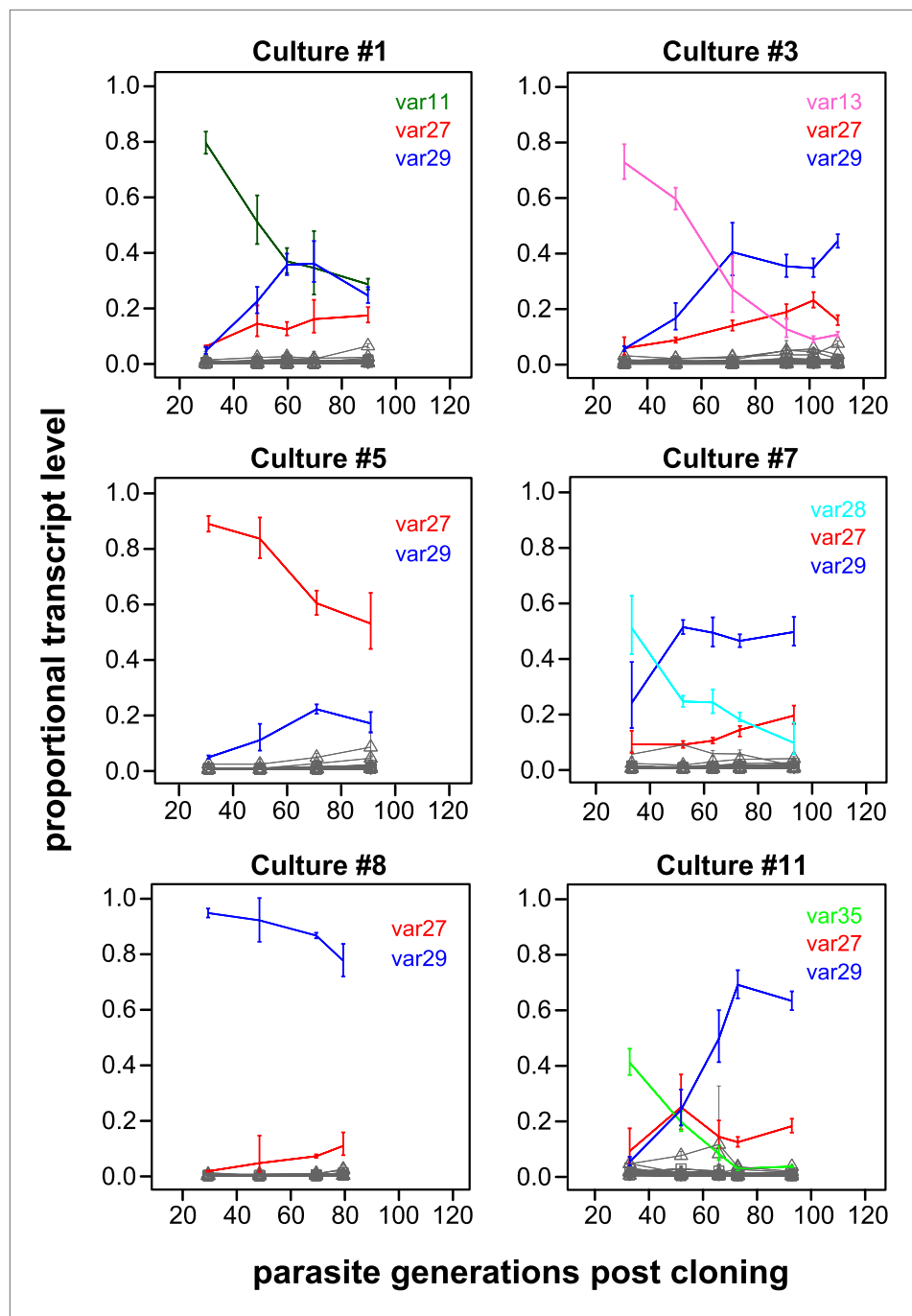
### Off-rate estimates suggest a role of chromosomal location, not promoter type

Estimated off-rates ranged from 0.3% to more than 5% per generation, with a mean of approximately 2.9% (**Table 2**). Notably, the two most transcribed genes, *var29* and *var27* (both centrally located on chromosome 4), had exceptionally low off-rates of approximately 0.3% and 0.8%, respectively, whereas the other starter genes switched off at rates between 1.4% and 3.5% per generation. In general we found off-rates of centrally located genes to be significantly lower than of genes in subtelomeric regions (Welch's  $t = 3.3$ ,  $p = 0.002$ ), as shown in **Figure 3**. This result remained significant ( $p = 0.035$ ) when we accounted for uncertainty in our estimates ('Materials and methods').

Previous studies reported that UpsA *var* genes from patient isolates were down-regulated faster than other *var* genes during adaptation to culture (**Peters et al., 2007; Zhang et al., 2011**). The HB3 repertoire contains seven UpsA genes, all but one of which are subtelomeric. We compared the off-rates of the subtelomeric UpsA genes with those of other subtelomeric genes and found no significant difference (Welch's  $t = 0.65$ ,  $p = 0.54$ ; accounting for uncertainty  $p = 0.55$ ), as shown in **Figure 3**. Among centrally located genes we compared UpsB with UpsC and again found no significant difference (**Figure 3**; Welch's  $t = 0.44$ ,  $p = 0.44$ ; accounting for uncertainty  $p = 0.57$ ). This lends support to a previous study suggesting that the rate at which transcriptionally active *var* genes become silenced is less dependent on promoter type than on chromosomal location (**Frank et al., 2007**).

### Var gene switching is governed by an activation hierarchy

In each of our cultures, gene transcription levels spanned several orders of magnitude, and only a minority of genes were ever highly transcribed during the experiment. As mentioned earlier, despite the diversity of the starter genes, the relative transcript levels of the activated genes appeared to converge towards similar levels across all cultures (**Figure 1, Figure 1—figure supplements 1 and 2**). To test the degree of similarity in gene activation, we excluded the nine starter genes and ranked the rest of the repertoire according to average transcript levels at later time points (more than 65 generations) for each culture. Pair-wise comparison showed that the ranking was highly correlated between



**Figure 1.** Proportional var transcript levels for six in vitro cultures. The parasite cultures initially expressed a variety of dominant 'starter genes', which switched off at notably different rates. Nevertheless, most cultures converged towards high level transcription of two centrally located genes *var27* and *var29* (red and blue lines, respectively), whereas most other gene transcripts (grey lines) remained at relatively low levels.

DOI: [10.7554/eLife.01074.004](https://doi.org/10.7554/eLife.01074.004)

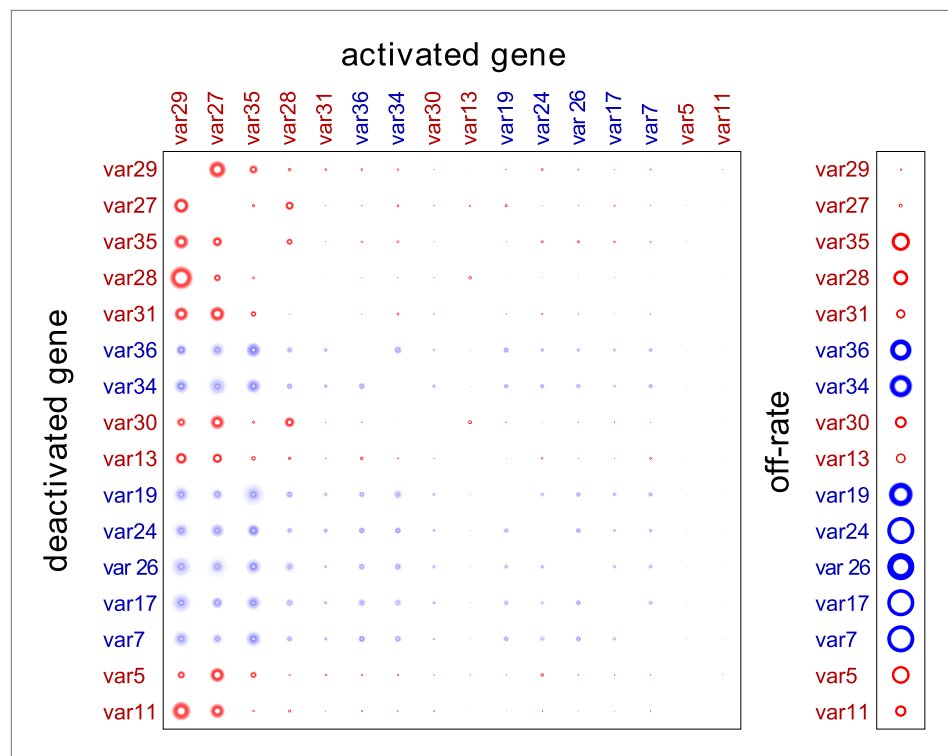
The following figure supplements are available for figure 1:

**Figure supplement 1.** Proportional var transcript levels and model output for cultures 1–6.

DOI: [10.7554/eLife.01074.005](https://doi.org/10.7554/eLife.01074.005)

**Figure supplement 2.** Proportional var transcript levels and model output for cultures 7–11.

DOI: [10.7554/eLife.01074.006](https://doi.org/10.7554/eLife.01074.006)



**Figure 2.** Estimation of switch parameters of highly transcribed genes. Parameter estimations for the 16 most transcribed var genes are represented as a switch matrix and an off-rate vector. The diameter of a circle in the  $i$ th row and  $j$ th column of the matrix is proportional to the switch bias  $\beta_{ij}$  from gene  $i$  to gene  $j$ , and the diameter of a circle in the off-rate vector is proportional to the rate  $\omega_i$  at which gene  $i$  becomes silenced. The fuzziness indicates uncertainty in the estimate, such that the darkness of each concentric ring is proportional to the likelihood that the parameter is within the corresponding range (Noble and Recker, 2012). Starter gene parameters, which are more precisely estimated, are in red, and genes are arranged from left to right, and top to bottom, in order of average transcript level.

DOI: [10.7554/eLife.01074.007](https://doi.org/10.7554/eLife.01074.007)

The following source data and figure supplements are available for figure 2:

**Source data 1.** Estimated switch biases and 95% credible intervals (CI).

DOI: [10.7554/eLife.01074.008](https://doi.org/10.7554/eLife.01074.008)

**Figure supplement 1.** Validation of estimated switching parameters.

DOI: [10.7554/eLife.01074.009](https://doi.org/10.7554/eLife.01074.009)

the cultures (Spearman's  $\rho > 0.8$ ,  $p=10^{-5}$  for all pairs), suggesting that the outcome of antigenic switching in *P. falciparum*, or at least within the HB3 isolate, is largely governed by an activation hierarchy that is independent of the currently activated gene.

Consistent with this hypothesis, the estimated switch biases between genes,  $\beta_{ij}$ , were remarkably similar, so that  $\beta_{ij} \approx \beta_{kj}$  for each  $i, j$  and  $k$ . (Figure 2, Figure 2—source data 1, Figure 2—figure supplement 1). All genes were found to have high switch biases towards the same few genes, which explain the observed convergence in var transcription in our cultures. As expected, the highest biases across the repertoire were found towards var29 and var27, with some genes also frequently switching towards var35 and/or var28 (both UpsC-types in central chromosomal location).

To better describe the observed activation hierarchy, we used our MCMC method to estimate the average switch bias towards each gene, which we hereafter refer to as the gene's activation bias. That is, the activation bias of gene  $j$  is defined as the average of the switch biases  $\beta_{ij}$  for all deactivating genes  $i$ . It can be understood as a gene specific, per switch activation probability and is therefore distinct from the previously defined on-rate (Horrocks et al., 2004), in that it is invariant to transcript levels. As shown in Table 2, there was a clear hierarchy in gene activation biases spanning over three orders of magnitude. The ranking also appeared to be non-random so that, for example, most genes within central chromosomal location were situated in the top half of the table. Analysing for possible associations with genetic attributes

**Table 2.** Parameter estimates for the HB3 var repertoire with 95% credible intervals

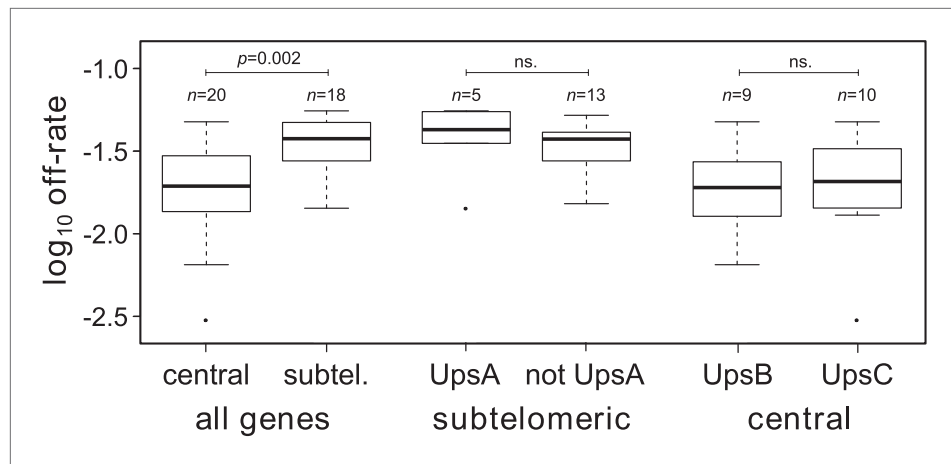
Gene	Location	Domains*	Ups	Off-rate (95% CI)	Activation bias (95% CI)
<i>var29</i>	Central	4	C	0.3% (0.1, 0.5)	27% (22, 32)
<i>var27</i>	Central	4	B	0.8% (0.5, 1.2)	23% (19, 28)
<i>var35</i>	Central	4	C	3.3% (2.2, 4.7)	8.0% (5.8, 11)
<i>var28</i>	Central	4	C	2.2% (1.4, 3.1)	6.9% (5.2, 9.1)
<i>var32</i>	Central	4	C	4.7% (2.5, 5.9)	2.5% (1.5, 3.3)
<i>var36</i>	Central	4	C	1.8% (0.2, 4.2)	2.5% (1.5, 4.1)
<i>var31</i>	Central	4	C	1.4% (0.9, 2.0)	2.2% (1.7, 2.9)
<i>var34</i>	Central	6	C	1.3% (0.1, 3.4)	2.1% (1.4, 3.6)
<i>var13</i>	Subtelomeric	4	B	1.9% (1.5, 2.4)	2.1% (1.6, 2.8)
<i>var30</i>	Central	4	C	1.9% (1.3, 2.7)	1.9% (1.4, 2.5)
<i>var10</i>	Subtelomeric	5	B	3.4% (1.1, 5.7)	1.7% (0.94, 2.7)
<i>var17</i>	Central	6	B	3.0% (0.8, 5.6)	1.7% (0.90, 2.8)
<i>var24</i>	Central	6	B	1.9% (0.1, 5.2)	1.6% (0.85, 3.1)
<i>var7</i>	Subtelomeric	7	B	2.2% (0.3, 5.1)	1.5% (0.83, 2.7)
<i>var19</i>	Central	4	B	1.3% (0.1, 3.2)	1.4% (0.93, 2.5)
<i>var26</i>	Central	4	B	1.5% (0.2, 3.4)	1.4% (0.90, 2.3)
<i>var25</i>	Central	6	B	2.7% (0.4, 5.6)	1.2% (0.60, 2.1)
<i>var1csa</i>	Subtelomeric	8	A	5.5% (4.4, 6.0)	1.2% (0.89, 1.5)
<i>var33</i>	Central	4	C	2.7% (0.5, 5.8)	1.1% (0.57, 1.9)
<i>var16</i>	Subtelomeric	4	B	2.8% (1.0, 5.3)	0.91% (0.52, 1.5)
<i>var4</i>	Subtelomeric	8	A	4.3% (2.1, 5.9)	0.86% (0.52, 1.3)
<i>var14</i>	Subtelomeric	4	B	4.0% (2.2, 5.9)	0.73% (0.45, 1.1)
<i>var22</i>	Central	7	B	1.9% (0.2, 5.1)	0.72% (0.42, 1.3)
<i>var21</i>	Central	7	B	4.7% (2.7, 5.9)	0.70% (0.46, 0.94)
<i>var8</i>	Subtelomeric	7	B	3.7% (1.8, 5.8)	0.66% (0.39, 1.0)
<i>var11</i>	Subtelomeric	5	B	1.5% (0.9, 2.2)	0.60% (0.45, 0.80)
<i>var18</i>	Subtelomeric	4	B	5.2% (3.8, 6.0)	0.60% (0.42, 0.79)
<i>var5</i>	Subtelomeric	6	A	3.5% (2.7, 4.3)	0.53% (0.39, 0.69)
<i>var39p†</i>	Subtelomeric	2	B	4.7% (2.7, 6.0)	0.48% (0.30, 0.66)
<i>var50Ψ‡</i>	Central	6	C	4.4% (2.0, 5.9)	0.46% (0.26, 0.67)
<i>var23</i>	Central	6	B	0.6% (0.0, 2.0)	0.38% (0.26, 0.55)
<i>var9</i>	Subtelomeric	6	B	2.8% (0.7, 5.3)	0.36% (0.20, 0.58)
<i>var2csa</i>	Subtelomeric	6	E	4.7% (2.3, 5.9)	0.28% (0.16, 0.40)
<i>var2</i>	Subtelomeric	6	A	1.4% (0.2, 4.2)	0.27% (0.17, 0.50)
<i>var12</i>	Subtelomeric	4	B	3.8% (1.5, 5.8)	0.26% (0.15, 0.40)
<i>var20</i>	Subtelomeric	4	B	4.1% (1.8, 5.9)	0.25% (0.14, 0.37)
<i>var1</i>	Subtelomeric	7	A	5.5% (4.2, 6.0)	0.024% (0.018, 0.031)
<i>var6</i>	Central	8	A	2.9% (0.4, 5.9)	0.018% (0.0086, 0.032)

\*Number of encoded DBL (Duffy binding-like) and CIDR (Cys rich inter-domain region).

†Partial gene.

‡Psuedogene.

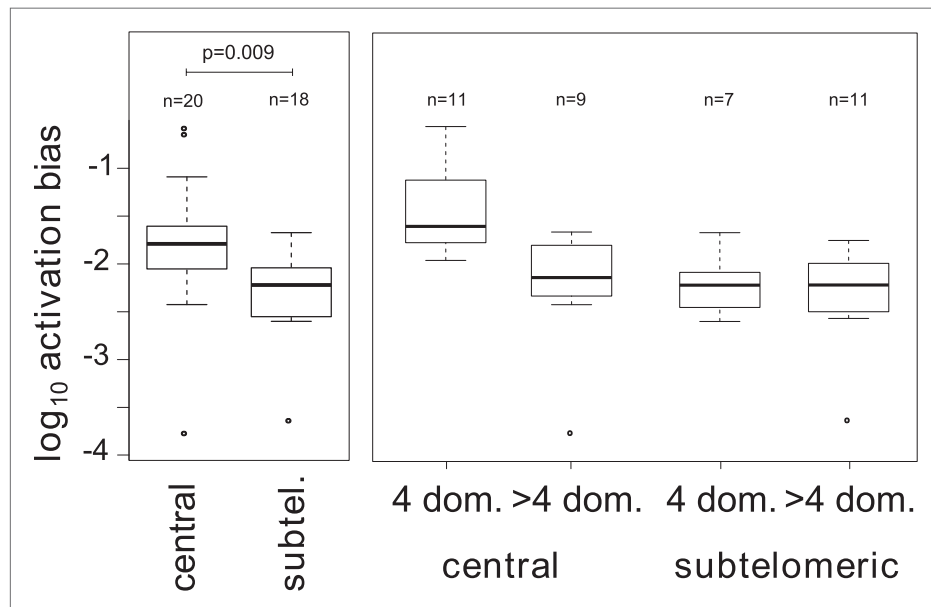
DOI: [10.7554/eLife.01074.010](https://doi.org/10.7554/eLife.01074.010)



**Figure 3.** Off-rate estimates show strong dependency on chromosomal location. The estimated rates at which active var genes become silenced are significantly lower for centrally located var genes than for those in subtelomeric locations. There was no significant effect of upstream promoter type when differentially testing off-rates of genes in subtelomeric location (UpsA vs non-UpsA) or in central location (UpsB vs UpsC).  
DOI: 10.7554/eLife.01074.011

confirmed that activation biases of centrally located var genes are significantly higher than those of subtelomeric genes (Figure 4, Welch’s  $t = 2.8$ ,  $p=0.009$ ; accounting for uncertainty  $p=0.011$ ). In fact, of the highest four activation biases, three belonged to neighbouring genes on chromosome 4: var29, var27 and var28. Importantly, no significant and independent effect of promoter type was found, which again points towards a strong influence of chromosomal location on var gene activation.

Additionally, we considered whether gene architecture had an effect on the observed activation biases, as recent research suggests that gene length, in terms of the number of binding domains, is another important var gene characteristic besides chromosomal location and promoter type (Buckee and Recker, 2012). Grouping genes depending on whether they encoded only four domains (short



**Figure 4.** Associations between activation biases and genetic attributes. The mean activation bias is higher for centrally located var genes. Within the set of centrally located genes, those encoding only four binding domains have higher activation biases than longer genes. Together, short, central genes have activation biases  $\approx 10$  times higher than the rest of the repertoire.  
DOI: 10.7554/eLife.01074.012

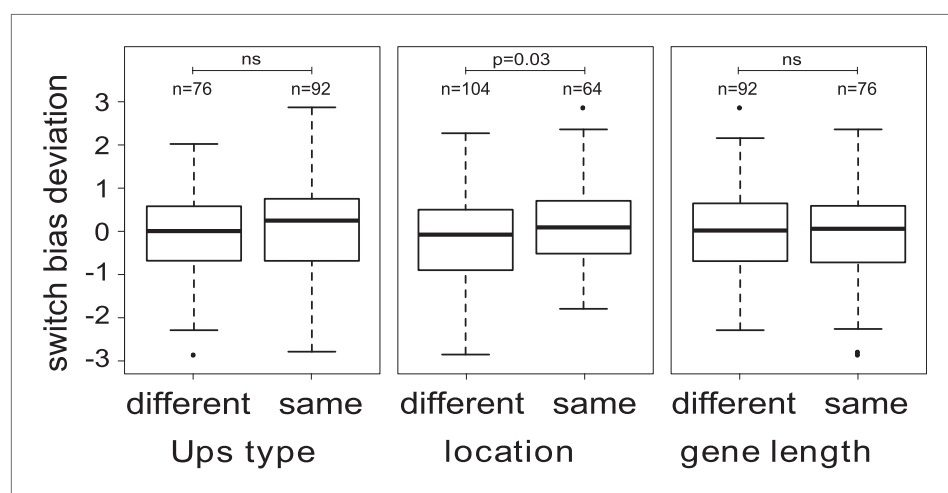
genes) or more than four domains (long genes) showed that activation biases differed significantly depending on gene length (Figure 4,  $F = 9.7$ ,  $p=0.004$ ) as well as chromosomal location ( $F = 8.1$ ,  $p=0.007$ ) and the interaction of these two factors ( $F = 5.0$ ,  $p=0.032$ ). As such, central *var* genes encoding only four binding domains (also known as Type 1 *var* genes [Gardner et al., 2002]) had on average a 10 times higher activation bias than the rest of the repertoire. Results were similar when we accounted for uncertainty in our estimates ( $p=0.005$ ,  $p=0.0096$ ,  $p=0.037$ , respectively).

### Var gene switching is influenced by the deactivating gene

We next tested explicitly whether the currently active gene has an influence on the direction of switching, that is if in general  $\beta_i \neq \beta_j$ . We used an optimisation algorithm ('Materials and methods') to fit two alternative models to the data for the 16 most highly transcribed *var* genes: (i) a general model including off-rates and individual switch biases (240 independent parameters), and (ii) a simpler model, in which all sets of switch biases were required to be identical, that is where  $\beta_{ij} = \beta_j$  for all  $i$  and  $j$  (31 independent parameters). A likelihood ratio test indicated that the model with individual switch biases was significantly more likely than the second model ( $F = 1.84$ ,  $p<0.0001$ ), confirming that antigenic switching in *P. falciparum* is not only controlled by an activation hierarchy but also influenced by the currently active gene. We again tested for associations with chromosomal location and promoter type ('Materials and methods'), which indicated that switch biases were higher between *var* genes with matching chromosomal locations than between differently located genes (Figure 5; Welch's  $t = 2.2$ ,  $p=0.029$ ), whereas no significant association was found between switch biases and matching Ups types or gene lengths. This result needs to be confirmed with larger data sets, however, as statistical significance was lost when we accounted for uncertainty in our parameter estimates.

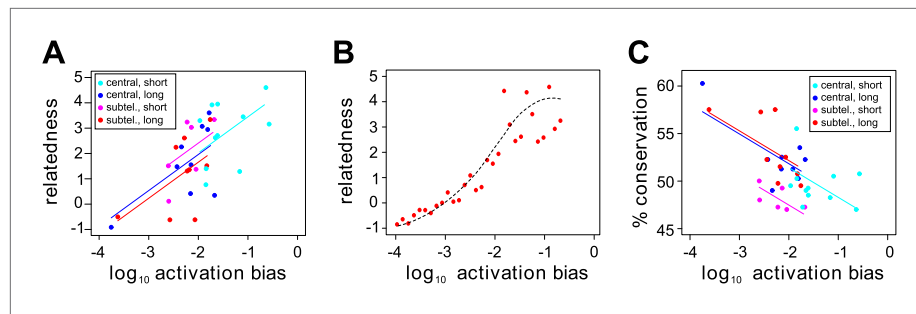
### Var gene sequences suggest a role of activation bias in antigenic evolution

*Var* gene sequences can be described using a set of 628 conserved homology blocks, with an average length of 19 amino acids (Rask et al., 2010). Using homology blocks found only within the DBL1 and CIDR1 domains, which are present in all HB3 *var* genes except *var1csa*, *var2csaA*, *var2csaB* and the partial gene *var39*, we constructed a *var* gene homology network, in which edges between genes are defined and weighted by the number of shared homology blocks (Figure 6—figure supplement 1). We then calculated the centrality of each gene within this network. This is a distance measure between all pairs of nodes (i.e., genes) and can therefore be used as a proxy for the relatedness of a gene to the rest of the repertoire. The inclusion of network centrality as an explanatory variable in the previous ANOVA model confirmed that the degree by which *var* genes share sequence blocks is positively correlated with activation bias, independently of chromosomal location and gene length (Figure 6A;  $F = 8.8$ ,  $p=0.006$  for the model comparison; accounting for uncertainty  $p=0.006$ ).



**Figure 5.** Variation in switch biases (deviation from the mean). Antigenic switching between *var* genes is more frequent between genes with matching chromosomal locations, for example from central to central or from subtelomeric to subtelomeric. No significant associations between switch bias and matching Ups type or gene length are found.

DOI: [10.7554/eLife.01074.013](https://doi.org/10.7554/eLife.01074.013)



**Figure 6.** Effect of *var* gene activation on sequence evolution. **(A)** *Var* gene activation biases are significantly correlated with each gene's relatedness to the rest of the HB3 repertoire, as measured by the gene's centrality within a shared homology block network, independently of chromosomal location or gene length. Trend lines show the fit of a linear regression model. **(B)** Simulation of *var* gene evolution by gene conversion, whereby homology blocks are swapped among pairs of genes chosen at random according to their activation biases, shows a similar correlation between the genes' activation biases and their relatedness to the rest of the repertoire. The red points show the outcome of one simulation, and the dashed line is the smoothed average of 50 runs. **(C)** Activation biases are negatively correlated with the gene's average domain sequence conservation, as a measure of population-level diversity, independently of chromosomal location and gene length. Trend lines show the fit of a linear regression model.

DOI: [10.7554/eLife.01074.014](https://doi.org/10.7554/eLife.01074.014)

The following figure supplements are available for figure 6:

**Figure supplement 1.** Relatedness network of the HB3 repertoire.

DOI: [10.7554/eLife.01074.015](https://doi.org/10.7554/eLife.01074.015)

We hypothesised that this could indicate a possible role of activation bias, or gene transcription, in *var* gene sequence evolution. Under the assumption that gene transcription facilitates recombination we would expect more frequently activated genes to share more sequence blocks with other genes, simply because they more often act as recombination sites. To test this we explicitly simulated gene conversion, in which pairs of genes were chosen at random, weighted by their activation bias and homology blocks copied between them ('Materials and methods'). As expected, we found a strong positive correlation between activation bias and our measure of relatedness in terms of network centrality (**Figure 6B**).

Under the same assumption we would also expect genes with higher activation biases to be overall more diverse. Indeed, using the average sequence conservation of all domains as a measure of population level diversity (**Buckee and Recker, 2012**), we again found this to be significantly and positively correlated, independently of chromosomal location and gene length (**Figure 6C**;  $F = 13.6$ ,  $p = 0.0009$  for the model comparison; accounting for uncertainty  $p = 0.0006$ ), suggesting a role of active gene transcription in generating antigenic diversity in *P. falciparum* malaria.

## Discussion

Several studies have noted that *var* genes located near the centres of chromosomes tend to be more highly transcribed in vitro than those in subtelomeric location (**Frank et al., 2007; Peters et al., 2007; Enderes et al., 2011; Zhang et al., 2011; Fastman et al., 2012**), and it has been suggested that this might be due to low deactivation rates (**Frank et al., 2007**). In this study, we found that high transcription levels are confined to a subset of central genes that encode only four PfEMP1 binding domains (termed Type 1 *var* genes [**Gardner et al., 2002**]) and, importantly, are the result of high activation biases rather than low off-rates. That is, even though some of the most highly transcribed genes had very low off-rates, the total sum of other genes switching off with high bias towards these genes significantly outweighs the effect of off-rates. This is consistent with our previous study using artificial transcription data, which also showed that off-rates, at least within a biologically plausible range, can have only a modest effect on the total transcript levels within parasite populations (**Noble and Recker, 2012**).

Reanalysis of data from a recent study on the NF54 parasite strain by **Fastman et al. (2012)** supports these findings. Consistent with our results for HB3, centrally located *var* genes encoding only four binding domains were more highly transcribed than the rest of the repertoire (Welch's  $t = 2.3$ ,  $p = 0.02$ ), with PF3D7\_0809100, PF3D7\_0421300 and PF3D7\_0800100 being the most highly transcribed genes. According to our model, these are predicted to have high activation bias, and indeed, PF3D7\_0809100 has previously been reported as the most transcribed *var* gene in NF54 bulk culture, whereas PF3D7\_0421300

and PF3D7\_0800100 were found to be the two genes most commonly activated by *var* gene switching in previous studies ([Dzikowski et al., 2006](#); [Frank et al., 2007](#); [Enderes et al., 2011](#)). Also in agreement with our results, NF54 gene transcription levels were positively correlated with relatedness to the rest of the repertoire in terms of shared DBL1 and CIDR1 homology blocks (Spearman's  $\rho = 0.3$ ,  $p=0.044$ ).

Our results further indicate that the direction of transcriptional switching between *var* genes depends not only on intrinsic activation biases but also on the particular preferences of each deactivating gene. Even though these variations were found to have only a modest effect on in vitro switching, they are expected to play a much greater role during infection, where factors such as immune responses and binding affinities can significantly alter the *var* gene expression profiles. Furthermore, the combination of an activation hierarchy and these specific variations could explain our previously proposed source-sink model underlying antigenic switching in *P. falciparum* ([Recker et al., 2011](#)), in which genes with high in-degrees, i.e., those most commonly switched to, correspond to the genes here identified as having high activation biases. Intriguingly, also, our results suggest that switch biases may be stronger between genes in similar chromosomal locations: central genes prefer to switch to other central genes, and subtelomeric genes prefer to switch to other subtelomeric genes. More data are needed to fully establish the factors influencing variation in switch biases, however, especially for less activated genes, where parameter estimations are more prone to uncertainties.

Despite recent advances in elucidating the genetics and epigenetics of *var* gene activation and silencing, not much is known about how mutually exclusive switching between the 60 *var* genes is controlled at the molecular level. The discovery of intrinsic *var* gene activation biases and their associations with chromosomal locations and other genetic features could therefore have important implications for future studies. These factors are clearly correlated, however, and it is uncertain how they are linked by cause and effect. For example, in a recent study it was found that gene length is positively correlated with sequence conservation ([Buckee and Recker, 2012](#)), and long *var* genes are more often positioned in subtelomeric regions. Nevertheless, the correlation between activation bias and relatedness, which was found to be independent of chromosomal location, gene length and upstream promoter, suggests a role of gene transcription in facilitating recombination and therefore in generating and maintaining antigenic diversity in *P. falciparum*. Transcription-associated recombination has been observed in other eukaryotes and may result from increased accessibility to recombination proteins or from stalled replication forks ([Prado and Aguilera, 2005](#); [Gottipati and Helleday, 2009](#)). Importantly, for *P. falciparum* this phenomenon could help to explain why different groups of *var* genes within the repertoire display very different levels of diversity within the population ([Buckee et al., 2009](#); [Buckee and Recker, 2012](#)).

It is further possible that some *var* genes are under balancing selection, whereby the relative sequence conservation of longer genes could be due to optimised functionality, for example in terms of binding avidity and/or affinity. This assumption would also help to reconcile apparent contradictions between our observations and *var* gene transcription in vivo. That is, various in vivo studies have reported UpsA *var* genes to be over-expressed during infections in individuals with little previous exposure ([Jensen et al., 2004](#); [Bull et al., 2005](#); [Lavstsen et al., 2005](#); [Kyriacou et al., 2006](#); [Warimwe et al., 2009](#)), despite their low activation rates in vitro. Each *var* repertoire consists of a set of relatively long and conserved genes, such as those predominantly found with in the UpsA group, and another, short and diverse set of genes, such as those predominantly found within the UpsB and UpsC groups ([Buckee and Recker, 2012](#)). Under the assumption that more conserved genes have an in vivo survival advantage, possibly due to higher binding affinities ([Avril et al., 2012](#); [Claessens et al., 2012](#); [Lavstsen et al., 2012](#)), these variants will dominate early infections in naive hosts, regardless of activation rates. Their lower sequence diversity, however, will then lead to a more rapid acquisition of protective immunity, causing a transition towards the expression of more diverse variants, facilitated by their inherent activation biases. In light of this we would also speculate that the proposed order in which individuals acquire immunity (anti-UpsA followed by anti-UpsB/C), as well as the proposed associations between the expression of UpsA *var* genes and severe disease outcome, especially in young children, is less determined by the Ups group itself and is more a consequence of within-host selection of long, conserved genes, which are often but not exclusively found within the UpsA group.

Together, these results demonstrate how selection pressures operating on multiple ecological scales have shaped the phenotypic plasticity embodied within the antigenic repertoire of *P. falciparum*. Within this setup, *var* genes and the molecular mechanisms that underlie their sequential activation and silencing have co-evolved to allow the parasite to express the most advantageous phenotype in response to its current environment.

## Materials and methods

### Parasite culture and cDNA preparation

Parasites were cultured and sorbitol-synchronised using standard techniques (Trager and Jensen, 1976; Lambros and Vanderberg, 1979), and RNA was extracted from saponin lysed, mid-to-late ring stage parasites with Trizol as previously described (Kyes et al., 2000). For cDNA preparation, 5 µg RNA was treated in 20 µl reactions with TurboFree DNase (Ambion®; following manufacturer's recommendations except incubation was at 16°C for 40 min). All DNase treated RNA samples were tested by PCR with primers to fructose-bisphosphate aldolase (PF14\_0425) (Salanti et al., 2003) to ensure complete removal of gDNA. Duplicate 20 µl reactions consisting of 6 µl RNA were reversed transcribed with 100 ng random hexamers (Invitrogen) and Bioscript reverse transcriptase (Bioline; following manufacturer's instructions) at 40°C for 40 min. The resulting cDNA was stored at -20°C in single-use aliquots and diluted 1:6 for real time PCR analysis. cDNA was synthesised in triplicate (84% of all data points) or duplicate, and real time PCR was done in duplicate for each primer pair on each cDNA.

### Primer design and initial testing

Primers specific for the HB3 var exon 1 repertoire were optimised to meet several basic criteria: that they distinguished between the known var gene sequences within the HB3 genome; that they were specific for HB3 var genes with no cross-reactivity to 3D7 var genes (for future application to studies on 3D7 × HB3 cross progeny); and that they reflected measurement of full-length var transcript. Primer pairs were designed using Primer3plus (Kraemer et al., 2007; Untergasser et al., 2007), with parameters manually set for product size range (100–150 nt), primer length (20 nt), T<sub>m</sub> (60°C), GC clamp (1 nt), salt correction formula and thermodynamic parameters (SantaLucia 1998); all other parameters were default settings (see Table 3 for primer sequences). To avoid general cross-detection of all var types, the first kilobase of exon 1 containing the conserved DBLα domain was avoided. The final kilobase was also excluded to avoid sterile transcripts (Su et al., 1995; Kyes et al., 2007). Each primer pair was specific for a single gene except, unavoidably, the var3 primer pair and the two pairs designed to the highly similar copies of var2csa. Each HB3 primer pair was tested five times on gDNA; pairs that varied more than ±1.5 threshold cycles (C<sub>t</sub>) values from the median or had efficiency value (Amplification) less than 1.85 were redesigned. Primer pairs that amplified 3D7 genomic DNA were discarded and redesigned so that they were specific for HB3 var genes only.

### qRT-PCR

Quantitative real time PCR (qRT-PCR) reactions were prepared with a Corbett CAS1200 liquid handling system and run on a Corbett Rotor-Gene 6000. Each 10 µl reaction contained 1 µl template, 5 µl SensiMix SYBR (Bioline), 3 µl water and 1 µl primer mix (1.6 µM of each primer). The cycling conditions were as follows: hold, 95°C/10 min; 45 cycles (95°C/25 s, 58°C/25 s, 68°C/30 s); melt, 55°C–99°C. Amplicons were checked for a single product of the expected size by agarose gel electrophoresis. Melt analysis confirmed single products.

### Data collection, processing and verification

Transcription levels for each replicate were calculated by the comparative quantification method of Pfaffl (2001). Comparative quantification results are given as expression levels relative to a control (house-keeping) gene, seryl-tRNA synthetase (Seryl-tRNAF agctacctcagaacaaccattatgtgc, Seryl-tRNAR atcctttc-catgtgccctgc), which is arbitrarily set at 1.0. Minor differences in Take Off values between primer pairs were accounted for using a correction factor derived from five replicates on gDNA. Expression was determined using  $A^{T.O.-CF}$ , where A is the average amplification, T.O. is the average C<sub>t</sub> and CF is the correction factor. To express results relative to the control gene, the following formula was used:  $A_{ref}^{T.O.-CF}/A_x^{T.O.-CF}$ . Expression levels were converted to proportions before calculating geometric means of the replicates. Replicates were excluded as outliers (1.1%) if they deviated from the mean by a factor of more than 10<sup>1.5</sup> or had an amplification efficiency below 1.5 (most excluded measurements met both criteria).

Dominant var gene expression detected by qRT-PCR was verified as both full-length and dominant by northern blot analysis, comparing specific probes for each dominant gene vs a generic exon 2 probe (data not shown). The primer pairs for var genes 5, 11, 13, 27, 28, 29, 31 and 35 were thus verified as specific for detecting full-length transcripts. The var3 primer pair, which was known to also detect var5 and pseudogene var46, indicated dominant var3 alongside var5 in culture #4. However, we were unable to

**Table 3.** HB3 qPCR primers and cross-referenced var gene identifiers

Gene name	Contig name	Broad locus name	F oligonucleotide	R oligonucleotide
<i>HB3 var16</i>	HB3-1000-1	PFHG_03232.1	ccctgtccacaaccatcagc	cgctcgtcatcagtgctcc
<i>HB3 var1</i>	HB3-1000-2	PFHG_03234.1	caaaggagaaggcaccacc	acctatggcaccctctcac
<i>HB3 var12</i>	HB3-1040	PFHG_03416.1	gatgctacaaccacccacc	gtgttaccactcgccactc
<i>HB3 var27</i>	HB3-1704_1	PFHG_03476.1	gctccaaccaccagcttcc	gcttctgctggtggctgc
<i>HB3 var28</i>	HB3-1074-2	PFHG_03478.1	gatggcacaagaagtggcgg	tggtctgggtcgacctctc
<i>HB3 var29</i>	HB3-1074-3	PFHG_03480.1	aagaagatggcgcgaagggc	tccggtgatcccttcttg
<i>HB3 var13</i>	HB3-1107	PFHG_03516.1	tggtaaatgcaagggtgatacagg	tgcatcgttatcactaccagc
<i>HB3 var1csa</i>	HB3-1108	PFHG_03521.1	cgcaatatgcaactaatgac	acttgccaattctgaacg
<i>HB3 var2</i>	HB3-1210	PFHG_03840.1	cgaggacaccacggaggagg	ttggtgctgctggttggc
<i>HB3 var5</i>	HB3-1235	PFHG_03671.1	aggctgctccttcagatgctg	tgtttccctaccatgacaaggatgcc
<i>HB3 var36</i>	HB3-1296	PFHG_04012.1	atggacaatgatggttaagg	taggagtaggtgtgctgctc
<i>HB3 var17</i>	HB3-1296-2	PFHG_04014.1	agatggcgcacaaggccaag	ttgggtttggcaccactagc
<i>HB3 var35</i>	HB3-1296-3	PFHG_04015.1	aaacggaaaacctggcctcc	tcgtcttgcccttggcttc
<i>HB3 var14</i>	HB3-1308	PFHG_04035.1	gggtggtggtgccgatcccgcc	tgtgacgcctccgtttagtgccc
<i>HB3 var8</i>	HB3-1334	PFHG_04081.1	ggcgggtgtctgtattccacc	gctgcctcaccacctgtag
<i>HB3 var9</i>	HB3-1408	PFHG_04057.1	tgctatgacgtgtaatgcccc	acttacatgagctccactctggtg
<i>HB3 var32</i>	HB3-1459	Fbadf	ggaaaccggtggactcac	acttgggtgcttgggggc
<i>HB3 var11</i>	HB3-1499	PFHG_04491.1	attgatgatgctgtcgcc	ggcacctggtttagtggtgg
<i>HB3 var19</i>	HB3-1514	PFHG_04620.1	aaactgacaatggccccgac	gttgttgaggggtcttcgg
<i>HB3 var18</i>	HB3-1523	PFHG_04593.1	gcggctcaccgacatcttc	gccgcctcgtcttctctgc
<i>HB3 var10</i>	HB3-1587	PFHG_04749.1	accactcgtgccaccacctc	gagttgtacctggcaccccc
<i>HB3 var7</i>	HB3-1604-1	PFHG_04769.1	agcagagtgttactcaggagg	gatggaccacgagatgtgcc
<i>HB3 var20</i>	HB3-1604-2	PFHG_04770.1	acgaagaagacgatgccacc	gaagtcttcggagcgaccac
<i>HB3 var4</i>	HB3-1703	PFHG_04861.1	tggtgccaaagaccctccc	ggccactcgtgtgtctgtg
<i>HB3 var2csaA</i>	HB3-1727	PFHG_05046.1	gggggaaatgtgggtgccg	gggggatacccccactcattaccag
<i>HB3 var3</i>	HB3-1737	PFHG_05052.1	aaagtgcgaagcacctcccc	cgccactgcagggattagctg
<i>HB3 var2csaB</i>	HB3-1817	PFHG_05155.1	tggtacagctgatggtgtactccg	tgtgcccgctttacggttctg
<i>HB3 var39p*</i>	HB3-2007	efbe	agccattactcgtcgaagctgg	agcggcacatcggcattttg
<i>HB3 var34</i>	HB3-209	PFHG_00592.1	agtgggtgctgtagagccaaagac	cctgcggcgggtgctgtaagg
<i>HB3 var23</i>	HB3-699-1	PFHG_02272.1	gaacccttgacgacgacac	ctcaacacagtcaaaggcg
<i>HB3 var30</i>	HB3-699-2	PFHG_02273.1	aagacgacaaacctggcacc	gtcgttgcctttggctcgg
<i>HB3 var6</i>	HB3-699-3	PFHG_02274.1	attcacagcactgaaagtcc	tcacaatcattaaagcatcc
<i>HB3 var26</i>	HB3-699-4	PFHG_02276.1	aagcagctgatggaacggac	tggttgtgtgggtcttggc
<i>HB3 var31</i>	HB3-699-5	PFHG_02277.1	cgcgaagacgaaaacgtcac	gtttcatccggaccgtctc
<i>HB3 var50Ψ†</i>	HB3-752-1	PFHG_02419.1	tggtaatgatgaagatgacg	gaattgcttactttgttc
<i>HB3 var24</i>	HB3-752-2	PFHG_02421.1	gctcgtctttaccaccgc	ttccgtctcctctcggcg
<i>HB3 var25</i>	HB3-752-3	PFHG_02423.1	agtgggtccaaaactgtcgg	accacaaaagtgcctcccc
<i>HB3 var22</i>	HB3-752-4	PFHG_02425.1	ccaccacaaaaccctccag	tccgcttgggtctgcttc
<i>HB3 var33</i>	HB3-752-5	PFHG_02429.1	acagaaagttggacaggatg	atggttgtttgagaattgc

\*Partial gene.

†Pseudogene.

DOI: [10.7554/eLife.01074.016](https://doi.org/10.7554/eLife.01074.016)

detect *var3* expression by northern blot. As this sample was positive by northern blot probed with *var5*, and it was the only one with 'dominant' *var3* expression by qRT-PCR, we discarded all *var3* qRT-PCR data as ambiguous. We also discarded all the data from one of the two *var2csaA/var2csaB* primer pairs.

## Parameter estimation

Following *Recker et al. (2011)*, we assumed that *var* transcription profiles can be described by a time-discrete model:

$$v_{i,c,t+1} = (1 - \omega_i)v_{i,c,t} + \sum_{j \neq i} \omega_j \beta_{ji} v_{j,c,t}$$

where  $v_{i,c,t}$  is the relative transcript level of variant  $i$  in culture  $c$  at time  $t$ ,  $\omega_i$  is the variant specific off-rate and  $\beta_{ji}$  is the switch bias (probability of a switch) from variant  $j$  to variant  $i$ . We assumed a maximum off-rate of 6% per generation, which is consistent with previous studies and theory (*Roberts et al., 1992; Horrocks et al., 2004; Frank et al., 2007; Recker et al., 2011*).

We used a Markov chain Monte Carlo (MCMC) method to obtain posterior distributions of likely values for the switch biases and off-rates, using the pooled data from all observed cultures as input. For model comparisons we used a simulated annealing algorithm to find maximum likelihood parameter sets. These methods have previously been described in detail and tested using diverse artificial data sets (*Noble and Recker, 2012*). For some of our analyses it was desirable to use a model of reduced dimension. Previously we have shown that reliable parameter estimates can be obtained from data for the 16 most highly transcribed *var* genes (*Noble and Recker, 2012*), which in this case comprised the 9 starter genes and the 7 others with highest average transcript levels. These 16 genes together accounted, on average, for more than 92% of the sum of transcript levels in each culture (excluding the culture's starter gene), whereas none of the excluded genes accounted for more than 1%. To estimate activation biases, we modified the model so that the direction of switching was independent of the previously active gene (i.e.,  $\beta_{ji} = \beta_{ki}$  for all  $i, j$  and  $k$ ).

Before conducting statistical tests, we transformed switch biases using the logit function  $\text{logit}(x) = \log(x/(1-x))$  to map from the range (0,1) to the range  $(-\infty, \infty)$ . Additionally, to account for uncertainty in our parameter estimates, we performed tests on the collected output of our MCMC method. This method samples from the posterior distribution, which means it selects numerous parameter sets such that the probability of each being chosen depends on how likely it is to explain the data. We performed our statistical tests on all the selected parameter sets and calculated mean p-values. We used the same parameter sets to obtain credible intervals for our estimates, as shown in *Table 2*.

## Variation in switch biases

We found that all *var* genes had similar sets of switch biases, such that  $\beta_{ij} \approx \beta_{kj}$  for each  $i, j$  and  $k$ . A model comparison indicated that these sets were not identical but varied depending on the deactivating gene. To estimate these variations we took a subset of the MCMC switch bias estimates and standardised each set  $\{\beta_{1j}, \beta_{2j}, \dots, \beta_{nj}\}$  by subtracting the mean and dividing by the standard deviation (after rescaling each set to compensate for  $\beta_{jj} = 0$ ). Negative values then corresponded to biases that were smaller than average and positive values to those that were larger. We used the subset of switch biases from the 9 starter genes to the 25 most transcribed genes, omitting the diagonal terms.

## Relatedness network

We constructed a network in which each node represented a gene and each edge was weighted by the number of shared variable DBL1 and CIDR1 homology blocks. Each gene's centrality within the network was calculated using a weighted closeness measure (*Newman, 2001*) implemented by the *tnet* package for R (*Opsahl, 2009*). We applied variance stabilising transformations to the data prior to analysis. Specifically, the centrality measure was rescaled and squared, and to each switch bias we applied the logit transformation, added five and then square-rooted the magnitude. These transformations made the data conform better to the assumptions of our statistical model; test results would have been more significant had we instead used the untransformed data.

## Gene conversion model

We simulated gene conversion in a repertoire of 30 genes, each comprising 40 homology blocks, with 30 possible variants for each homology block. Genes were initially constructed of randomly chosen homology blocks and were assigned activation biases according to a geometric sequence. For 30,000

iterations, pairs of genes were selected at random, with the probability of selection being proportional to activation bias. A randomly chosen homology block was copied from the first of the chosen genes to replace the corresponding block in the second gene. Homology blocks in all genes also underwent random mutation at a rate 10 times lower than the gene conversion rate. We constructed a relatedness network and calculated centrality in the same way as for the HB3 data.

### Initial conditions of the model

Our switching model required a set of initial conditions. Assuming that each culture, excluding the parent, began with a single parasite and that populations expanded approximately fivefold per generation, we deduced that the first transcriptional switch most likely occurred, on average, two or three generations after cloning. Therefore we assumed that each of these cultures remained clonal for the first three generations. For the parent culture, we used the first observation as the initial condition. If a switch had occurred in one or more of the initially clonal cultures within the first few generations, when the population was very small, then our assumed initial conditions would be inaccurate. Therefore we also ran our algorithms using the first observations as the initial conditions for all cultures. The parameter estimates were highly consistent with those obtained using clonal initial conditions.

### Cross-validation of parameter estimates

The simplest way to estimate switch biases and off-rates for a single *var* gene is to obtain a clone transcribing the gene, grow it in culture and analyse changes in transcript levels over time. A particular strength of our method was that it also estimated parameters for genes we did not clone. To assess the accuracy of results obtained in this way, we performed cross-validation by sequentially excluding data sets. For example, to derive alternative estimates for *var29* we ran the MCMC algorithm without the data for culture #8 (which initially expressed *var29*), and to derive alternative estimates for *var27* we excluded data for cultures #5 and #6. The results, shown in **Figure 2—figure supplement 1**, confirmed our method's ability to estimate switch biases of non-starter genes with sufficient accuracy to determine the general switching pattern. Off-rate estimates were however not always accurate, especially for genes with very low transcript levels.

We used a similar method to test the accuracy of our method for estimating variations in switch biases. Cultures #2 and #3 initially expressed the same starter gene *var13*, and both cultures #5 and #6 initially expressed *var27*. By excluding the data from culture #2 we derived estimates for the *var13* switch biases mostly dependent on the culture #3 data, and vice versa. Similarly we excluded data from each of cultures #5 and #6 to obtain two sets of estimates for the *var27* switch biases. For each of the two different data sets we calculated the deviations from the mean switch biases and found these to be positively correlated (Pearson's  $\rho = 0.43$ ,  $p=0.015$ ), confirming the validity of our method.

### Validation of model assumptions

Our likelihood function, used for parameter estimation and model comparisons, assumed that input data were subject to measurement errors following a log-normal distribution. We have previously shown that our MCMC algorithm reliably determines network structures when the measurement error distribution is log-normal (base 2) with standard deviation  $\sigma \leq 1$ , and that  $\sigma$  can be reliably estimated using our simulated annealing algorithm (**Noble and Recker, 2012**). In the 16-gene model,  $\sigma \approx 0.73$ , which was well within the required range.

To test whether the measurement errors obeyed a log-normal distribution, we analysed the distribution of the deviations between the data and the simulated annealing output, after applying log transformations. The distribution for the 16-gene model was unimodal and only slightly asymmetrical (skewness 0.4) but was more peaked than a normal distribution (excess kurtosis 2.4), so that it somewhat resembled a Laplace distribution. We tried running the simulated annealing algorithm with an error function that assumed log-Laplace-distributed noise, and this made very little difference to the results. The deviations for the constrained 38-gene model were similarly distributed, excluding a very small minority of outliers.

Unusually large deviations occurred at 13 of the 64 observed time points, 8 of which were between generations 48 and 52 and therefore belonged to the same round of measurements. Transcript levels deviated more for longer genes than for shorter genes, and northern blotting of the samples taken at the affected time points showed unusually low levels of long, for example *var* length, RNA, confirming that the cause was not PCR error but most likely an as yet unidentified artefact of RNA preparation. Excluding the 13 data points affected by this phenomenon made very little difference to our parameter estimates and statistical test results.

Analysis of the qRT-PCR replicates revealed that measurements of lower transcript levels had higher standard errors, but the average standard error increased only threefold for each 1000-fold difference in transcript level.

## Growth rates

It has been suggested that *var* gene transcript levels in culture may be affected by small differences in parasite growth rates (Horrocks *et al.*, 2004), although no evidence has been found to support this hypothesis. To test whether such differences might affect our estimates we added to our model an excess growth rate parameter  $\phi$  for each gene  $i$ :

$$v_{i,c,t+1} = (1 - \omega_i)(1 + \phi_i)v_{i,c,t} + \sum_{j \neq i} \omega_j \beta_{ji} (1 + \phi_j)v_{j,c,t}.$$

When we allowed growth rates to vary by up to 0.5% per generation (equivalent to approximately 50% over the duration of the experiment), the extended model was not significantly more likely than the model with only switch biases and off-rates. When we increased the upper bound on the excess growth rates to 1% per generation, the extended model became significantly more likely. However, such large differences in growth rates in vitro are biologically implausible and have never been observed. Importantly, the inclusion of growth rate differences had very little effect on our switch bias and off-rate estimates.

## Acknowledgements

We would like to thank Caroline O Buckee for providing estimates of *var* gene domain sequence conservation. The work was funded by the Wellcome Trust (Grant No. 082130/Z/07/Z to CN), the Biotechnology and Biological Sciences Research Council (studentship to RN) and the Royal Society (University Research Fellowship to MR).

## Additional information

### Funding

Funder	Grant reference number	Author
Wellcome Trust	082130/Z/07/Z	Chris I Newbold
BBSRC		Robert Noble
Royal Society		Mario Recker

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

### Author contributions

RN, Analysis and interpretation of data, Drafting or revising the article; ZC, Acquisition of data, Analysis and interpretation of data; SK, Acquisition of data, Analysis and interpretation of data, Drafting or revising the article; RP, Acquisition of data, Analysis and interpretation of data; CIN, MR, Conception and design, Analysis and interpretation of data, Drafting or revising the article

## References

- Avril M, Tripathi AK, Brazier AJ, Andisi C, Janes JH, Soma VL, *et al.* 2012. A restricted subset of *var* genes mediates adherence of *Plasmodium falciparum*-infected erythrocytes to brain endothelial cells. *Proc Natl Acad Sci USA* **109**:E1782–90. doi: [10.1073/pnas.1120534109](https://doi.org/10.1073/pnas.1120534109).
- Blomqvist K, Normark J, Nilsson D, Ribacke U, Orikiriza J, Trillkott P, *et al.* 2010. *var* gene transcription dynamics in *Plasmodium falciparum* patient isolates. *Mol Biochem Parasitol* **170**:74–83. doi: [10.1016/j.molbiopara.2009.12.002](https://doi.org/10.1016/j.molbiopara.2009.12.002).
- Boone KE, Watters DA. 1995. The incidence of malaria after splenectomy in Papua New Guinea. *BMJ* **311**:1273. doi: [10.1136/bmj.311.7015.1273](https://doi.org/10.1136/bmj.311.7015.1273).
- Buckee CO, Recker M. 2012. Evolution of the multi-domain structures of virulence genes in the human malaria parasite, *Plasmodium falciparum*. *PLOS Comput Biol* **8**:e1002451. doi: [10.1371/journal.pcbi.1002451](https://doi.org/10.1371/journal.pcbi.1002451).
- Buckee CO, Bull PC, Gupta S. 2009. Inferring malaria parasite population structure from serological networks. *Proc Biol Sci* **276**:477–85. doi: [10.1098/rspb.2008.1122](https://doi.org/10.1098/rspb.2008.1122).

- Buffet PA**, Safeukui I, Milon G, Mercereau-Puijalon O, David PH. 2009. Retention of erythrocytes in the spleen: a double-edged process in human malaria. *Curr Opin Hematol* **16**:157–64. doi: [10.1097/MOH.0b013e32832a1d4b](https://doi.org/10.1097/MOH.0b013e32832a1d4b).
- Bull PC**, Berriman M, Kyes S, Quail MA, Hall N, Kortok MM, et al. 2005. *Plasmodium falciparum* variant surface antigen expression patterns during malaria. *PLoS Pathog* **1**:202–13. doi: [10.1371/journal.ppat.0010026](https://doi.org/10.1371/journal.ppat.0010026).
- Bull PC**, Buckee CO, Kyes S, Kortok MM, Thathy V, Guyah B, et al. 2008. *Plasmodium falciparum* antigenic variation. Mapping mosaic var gene sequences onto a network of shared, highly polymorphic sequence blocks. *Mol Microbiol* **68**:1519–34. doi: [10.1111/j.1365-2958.2008.06248.x](https://doi.org/10.1111/j.1365-2958.2008.06248.x).
- Cham GK**, Turner L, Kurtis JD, Mutabingwa T, Fried M, et al. 2010. Hierarchical, domain type-specific acquisition of antibodies to *Plasmodium falciparum* erythrocyte membrane protein 1 in Tanzanian children. *Infect Immun* **78**:4653–9. doi: [10.1128/IAI.00593-10](https://doi.org/10.1128/IAI.00593-10).
- Claessens A**, Adams Y, Ghumra A, Lindergard G, Buchan CC, Andisi C, et al. 2012. 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**:E1772–81. doi: [10.1073/pnas.1120461109](https://doi.org/10.1073/pnas.1120461109).
- Deitsch KW**, del Pinal A, Wellems TE. 1999. Intra-cluster recombination and var transcription switches in the antigenic variation of *Plasmodium falciparum*. *Mol Biochem Parasitol* **101**:107–16. doi: [10.1016/S0166-6851\(99\)00062-6](https://doi.org/10.1016/S0166-6851(99)00062-6).
- 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. doi: [10.1371/journal.ppat.0020022](https://doi.org/10.1371/journal.ppat.0020022).
- Enderes C**, Kombila D, Dal-Bianco M, Dzikowski R, Kremsner P, Frank M. 2011. Var gene promoter activation in clonal *Plasmodium falciparum* isolates follows a hierarchy and suggests a conserved switching program that is independent of genetic background. *J Infect Dis* **204**:1620–31. doi: [10.1093/infdis/jir594](https://doi.org/10.1093/infdis/jir594).
- Fastman Y**, Noble R, Recker M, Dzikowski R. 2012. Erasing the epigenetic memory and beginning to switch—the onset of antigenic switching of var genes in *Plasmodium falciparum*. *PLOS ONE* **7**:e34168. doi: [10.1371/journal.pone.0034168](https://doi.org/10.1371/journal.pone.0034168).
- Frank M**, Dzikowski R, Amulic B, Deitsch K. 2007. Variable switching rates of malaria virulence genes are associated with chromosomal position. *Mol Microbiol* **64**:1486–98. doi: [10.1111/j.1365-2958.2007.05736.x](https://doi.org/10.1111/j.1365-2958.2007.05736.x).
- Frank M**, Kirkman L, Costantini D, Sanyal S, Lavazec C, Templeton TJ, et al. 2008. Frequent recombination events generate diversity within the multi-copy variant antigen gene families of *Plasmodium falciparum*. *Int J Parasitol* **38**:1099–109. doi: [10.1016/j.ijpara.2008.01.010](https://doi.org/10.1016/j.ijpara.2008.01.010).
- Freitas-Junior LH**, Bottius E, Pirrit LA, Deitsch KW, Scheidig C, Guinet F, et al. 2000. Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*. *Nature* **407**:1018–22. doi: [10.1038/35039531](https://doi.org/10.1038/35039531).
- Gardner MJ**, Hall N, Fung E, White O, Berriman M, Hyman RW, et al. 2002. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419**:498–511. doi: [10.1038/nature01097](https://doi.org/10.1038/nature01097).
- Gottipati P**, Helleday T. 2009. Transcription-associated recombination in eukaryotes: link between transcription, replication and recombination. *Mutagenesis* **24**:203–10. doi: [10.1093/mutage/gen072](https://doi.org/10.1093/mutage/gen072).
- Greenwood BM**, Bradley AK, Greenwood AM, Byass P, Jammeh K, Marsh K, et al. 1987. Mortality and morbidity from malaria among children in a rural area of the Gambia, West Africa. *Trans R Soc Trop Med Hyg* **81**:478–86. doi: [10.1016/0035-9203\(87\)90170-2](https://doi.org/10.1016/0035-9203(87)90170-2).
- Gupta S**, Snow RW, Donnelly CA, Marsh K, Newbold C. 1999. Immunity to non-cerebral severe malaria is acquired after one or two infections. *Nat Med* **5**:340–3. doi: [10.1038/6560](https://doi.org/10.1038/6560).
- Horrocks P**, Pinches R, Christodoulou Z, Kyes SA, Newbold CI. 2004. Variable var transition rates underlie antigenic variation in malaria. *Proc Natl Acad Sci USA* **101**:11129–34. doi: [10.1073/pnas.0402347101](https://doi.org/10.1073/pnas.0402347101).
- Jensen AT**, Magistrado P, Sharp S, Joergensen L, Lavstsen T, et al. 2004. *Plasmodium falciparum* associated with severe childhood malaria preferentially expresses PfEMP1 encoded by group a var genes. *J Exp Med* **199**:1179–90. doi: [10.1084/jem.20040274](https://doi.org/10.1084/jem.20040274).
- Kaestli M**, Cockburn IA, Cortes A, Baea K, Rowe JA, Beck HP. 2006. Virulence of malaria is associated with differential expression of *Plasmodium falciparum* var gene subgroups in a case-control study. *J Infect Dis* **193**:1567–74. doi: [10.1086/503776](https://doi.org/10.1086/503776).
- Kraemer SM**, Kyes SA, Aggarwal G, Springer AL, Nelson SO, Christodoulou Z, et al. 2007. Patterns of gene recombination shape var gene repertoires in *Plasmodium falciparum*: comparisons of geographically diverse isolates. *BMC Genomics* **8**:45. doi: [10.1186/1471-2164-8-45](https://doi.org/10.1186/1471-2164-8-45).
- Kraemer SM**, Smith JD. 2006. A family affair: var genes, PfEMP1 binding, and malaria disease. *Curr Opin Microbiol* **9**:374–80. doi: [10.1016/j.mib.2006.06.006](https://doi.org/10.1016/j.mib.2006.06.006).
- Kyes S**, Christodoulou Z, Pinches R, Kriek N, Horrocks P, Newbold C. 2007. *Plasmodium falciparum* var gene expression is developmentally controlled at the level of RNA polymerase II-mediated transcription initiation. *Mol Microbiol* **63**:1237–47. doi: [10.1111/j.1365-2958.2007.05587.x](https://doi.org/10.1111/j.1365-2958.2007.05587.x).
- Kyes S**, Pinches R, Newbold C. 2000. A simple RNA analysis method shows var and rif multigene family expression patterns in *Plasmodium falciparum*. *Mol Biochem parasitol* **105**:311–5. doi: [10.1016/S0166-6851\(99\)00193-0](https://doi.org/10.1016/S0166-6851(99)00193-0).
- Kyriacou HM**, Stone GN, Challis RJ, Raza A, Lyke KE, Thera MA, et al. 2006. Differential var gene transcription in *Plasmodium falciparum* isolates from patients with cerebral malaria compared to hyperparasitaemia. *Mol Biochem Parasitol* **150**:211–8. doi: [10.1016/j.molbiopara.2006.08.005](https://doi.org/10.1016/j.molbiopara.2006.08.005).
- Lambros C**, Vanderberg JP. 1979. Synchronization of *Plasmodium falciparum* erythrocytic stages in culture. *J Parasitol* **65**:418–20. doi: [10.2307/3280287](https://doi.org/10.2307/3280287).
- Lavstsen T**, Magistrado P, Hermsen CC, Salanti A, Jensen AT, Sauerwein R, et al. 2005. Expression of *Plasmodium falciparum* erythrocyte membrane protein 1 in experimentally infected humans. *Malar J* **4**:21. doi: [10.1186/1475-2875-4-21](https://doi.org/10.1186/1475-2875-4-21).

- Lavstsen T, Salanti A, Jensen ATR, Arnot DE, Theander TG. 2003. Sub-grouping of *Plasmodium falciparum* 3D7 var genes based on sequence analysis of coding and non-coding regions. *Malar J* 2:27. doi: 10.1186/1475-2875-2-27.
- Lavstsen T, Turner L, Saguti F, Magistrado P, Rask TS, Jespersen JS, et al. 2012. *Plasmodium falciparum* erythrocyte membrane protein 1 domain cassettes 8 and 13 are associated with severe malaria in children. *Proc Natl Acad Sci USA* 109:E1791–800. doi: 10.1073/pnas.1120455109.
- Marsh K. 1992. Malaria—a neglected disease? *Parasitology* 104(Suppl):S53–69. doi: 10.1017/S0031182000075247.
- Marsh K, Snow RW. 1997. Host-parasite interaction and morbidity in malaria endemic areas. *Philos Trans R Soc Lond B Biol Sci* 352:1385–94. doi: 10.1098/rstb.1997.0124.
- Newman MEJ. 2001. Scientific collaboration networks. II. shortest paths, weighted networks, and centrality. *Phys Rev E Stat Nonlin Soft Matter Phys* 64:016132. doi: 10.1103/PhysRevE.64.016132.
- Noble R, Recker M. 2012. A statistically rigorous method for determining antigenic switching networks. *PLOS ONE* 7:e39335. doi: 10.1371/journal.pone.0039335.
- Ockenhouse CF, Ho M, Tandon NN, Vansenter GA, Shaw S, White NJ, et al. 1991. Molecular-basis of sequestration in severe and uncomplicated *Plasmodium falciparum* malaria - differential adhesion of infected erythrocytes to CD36 and ICAM-1. *J Infect Dis* 164:163–169. doi: 10.1093/infdis/164.1.163.
- Opsahl T. 2009. *Structure and evolution of weighted networks*. p. 104–122 [Online]. Available: <http://toreopsahl.com/publications/thesis/>.
- 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 USA* 99:10 689–94. doi: 10.1073/pnas.162349899.
- Peters JM, Fowler EV, Krause DR, Cheng Q, Gatton ML. 2007. Differential changes in *Plasmodium falciparum* var transcription during adaptation to culture. *J Infect Dis* 195:748–55. doi: 10.1086/511436.
- Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* 29:e45. doi: 10.1093/nar/29.9.e45.
- Pongponratn E, Riganti M, Punpoowong B, Aikawa M. 1991. Microvascular sequestration of parasitized erythrocytes in human falciparum-malaria - a pathological study. *Am J Trop Med Hyg* 44:168–75.
- Prado F, Aguilera A. 2005. Impairment of replication fork progression mediates RNA polII transcription-associated recombination. *EMBO J* 24:1267–76. doi: 10.1038/sj.emboj.7600602.
- Rask TS, Hansen DA, Theander TG, Pedersen AG, Lavstsen T. 2010. *Plasmodium falciparum* erythrocyte membrane protein 1 diversity in seven genomes - divide and conquer. *Plos Comput Biology* 6:e1000933. doi: 10.1371/journal.pcbi.1000933.
- Recker M, Buckee CO, Serazin A, Kyes S, Pinches R, Christodoulou Z, et al. 2011. Antigenic variation in *Plasmodium falciparum* malaria involves a highly structured switching pattern. *Plos Pathog* 7:e1001306. doi: 10.1371/journal.ppat.1001306.
- Roberts DJ, Craig AG, Berendt AR, Pinches R, Nash G, Marsh K, et al. 1992. Rapid switching to multiple antigenic and adhesive phenotypes in malaria. *Nature* 357:689–92. doi: 10.1038/357689a0.
- Rottmann M, Lavstsen T, Mugasa JP, Kaestli M, Jensen AT, Muller D, et al. 2006. Differential expression of var gene groups is associated with morbidity caused by *Plasmodium falciparum* infection in Tanzanian children. *Infect Immun* 74:3904–11. doi: 10.1128/IAI.02073-05.
- Rowe JA, Claessens A, Corrigan RA, Arman M. 2009. Adhesion of *Plasmodium falciparum*-infected erythrocytes to human cells: molecular mechanisms and therapeutic implications. *Expert Rev Mol Med* 11:e16. doi: 10.1017/S1462399409001082.
- Salanti A, Dahlback M, Turner L, Nielsen MA, Barfod L, Magistrado P, et al. 2004. Evidence for the involvement of VAR2CSA in pregnancy-associated malaria. *J Exp Med* 200:1197–203. doi: 10.1084/jem.20041579.
- Salanti A, Staalsoe T, Lavstsen T, Jensen ATR, Sowa MPK, Arnot DE, et al. 2003. Selective upregulation of a single distinctly structured var gene in chondroitin sulphate A-adhering *Plasmodium falciparum* involved in pregnancy-associated malaria. *Mol Microbiol* 49:179–191. doi: 10.1046/j.1365-2958.2003.03570.x.
- Smith JD, Chitnis CE, Craig AG, Roberts DJ, Hudson-Taylor DE, Peterson DS, et al. 1995. Switches in expression of *Plasmodium falciparum* var genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. *Cell* 82:101–10. doi: 10.1016/0092-8674(95)90056-X.
- Su XZ, Heatwole VM, Wertheimer SP, Guinet F, Herrfeldt JA, Peterson DS, et al. 1995. The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of *Plasmodium falciparum*-infected erythrocytes. *Cell* 82:89–100. doi: 10.1016/0092-8674(95)90055-1.
- Taylor HM, Kyes SA, Newbold CI. 2000. Var gene diversity in *Plasmodium falciparum* is generated by frequent recombination events. *Mol Biochem Parasitol* 110:391–7. doi: 10.1016/S0166-6851(00)00286-3.
- Trager W, Jensen JB. 1976. Human malaria parasites in continuous culture. *Science* 193:673–675. doi: 10.1126/science.781840.
- Tran TM, Li S, Doumbo S, Doumtabe D, Huang CY, Dia S, et al. 2013. An intensive longitudinal cohort study of Malian children and adults reveals no evidence of acquired immunity to *Plasmodium falciparum* infection. *Clin Infect Dis* 57:40–7. doi: 10.1093/cid/cit174.
- Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JAM. 2007. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res* 35:W71–4. doi: 10.1093/nar/gkm306.
- Ward CP, Clotney GT, Dorris M, Ji DD, Arnot DE. 1999. Analysis of *Plasmodium falciparum* PfEMP-1/var genes suggests that recombination rearranges constrained sequences. *Mol Biochem Parasitol* 102:167–77. doi: 10.1016/S0166-6851(99)00106-1.

- Warimwe GM**, Keane TM, Fegan G, Musyoki JN, Newton CR, Pain A, et al. 2009. *Plasmodium falciparum* var gene expression is modified by host immunity. *Proc Natl Acad Sci USA* **106**:21 801–6. doi: [10.1073/pnas.0907590106](https://doi.org/10.1073/pnas.0907590106).
- Wunderlich G**, Alves FP, Golnitz U, Tada MS, Camargo EF, Pereira-da Silva LH. 2005. Rapid turnover of *Plasmodium falciparum* var gene transcripts and genotypes during natural non-symptomatic infections. *Rev Inst Med Trop Sao Paulo* **47**:195–201. doi: [10.1590/S0036-46652005000400004](https://doi.org/10.1590/S0036-46652005000400004).
- Zhang Q**, Zhang Y, Huang Y, Xue X, Yan H, Sun X, et al. 2011. 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. doi: [10.1371/journal.pone.0020591](https://doi.org/10.1371/journal.pone.0020591).