Accepted Manuscript

Global analysis of Plasmodium falciparum histidine-rich protein-2 (pfhrp2) and pfhrp3 gene deletions using whole-genome sequencing data and meta-analysis

Nuno Sepúlveda, Jody Phelan, Ernest Diez-Benavente, Susana Campino, Taane G. Clark, Heidi Hopkins, Colin Sutherland, Chris J. Drakeley, Khalid B. Beshir

PII: S1567-1348(18)30226-0
DOI: doi:10.1016/j.meegid.2018.04.039
Reference: MEEGID 3505

To appear in: *Infection, Genetics and Evolution*

Received date: 8 March 2018
Revised date: 24 April 2018
Accepted date: 29 April 2018

Please cite this article as: Nuno Sepúlveda, Jody Phelan, Ernest Diez-Benavente, Susana Campino, Taane G. Clark, Heidi Hopkins, Colin Sutherland, Chris J. Drakeley, Khalid B. Beshir , Global analysis of Plasmodium falciparum histidine-rich protein-2 (pfhrp2) and pfhrp3 gene deletions using whole-genome sequencing data and meta-analysis. The address for the corresponding author was captured as affiliation for all authors. Please check if appropriate. Meegid(2018), doi:10.1016/j.meegid.2018.04.039

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Global analysis of *Plasmodium falciparum* histidine-rich protein-2 (*pfhrp2*) and *pfhrp3* gene deletions using whole-genome sequencing data and meta-analysis

Nuno Sepúlveda[1,2,*], Jody Phelan[1], Ernest Diez-Benavente[1], Susana Campino[1], Taane G. Clark[1], Heidi Hopkins[1], Colin Sutherland[1], Chris J. Drakeley[1], Khalid B. Beshir[1]

[1] London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, United Kingdom

[2] Centre of Statistics and its Applications of University of Lisbon, Faculdade de Ciências da Universidade de Lisboa, Bloco C6 - Piso 4, Campo Grande, 1749-016 Lisboa, Portugal

* Corresponding author
London School of Hygiene and Tropical Medicine
Keppel Street
WC1E 7HT London
United Kingdom

E-mail: nuno.sepulveda@lshtm.ac.uk

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Abstract

Many rapid diagnostic tests (RDT) used on suspected malaria cases are based on the detection of the protein encoded by the *Plasmodium falciparum* histidine-rich protein-2 (*pfhrp2)* gene, which shares a high sequence homology with *pfhrp3* in the 3D7 reference genome. Parasite isolates showing *pfhrp2* and *pfhrp3* gene deletions have been emerging over the years, but a comprehensive genetic analysis of these variants is still lacking. With this purpose, genomic data from experimental *P. falciparum* genetic crosses between different laboratory lines (3D7, HB3, DD2, 7G8 and GB4) were first analysed (n=98). The frequency of *pfhrp2* deletions was consistent with a Mendelian prediction in HB3 x DD2 (56.7%; 95%CI=(39.5%-72.9%)). Moreover, the *pfhrp2* and *pfhrp3* deletions segregated independently of each other in the same genetic cross. Analysis of 3D7 x HB3 and 7G8 x GB4 estimated the probability of spontaneously generating a *pfhrp2* deletion during sexual recombination to be up to 6.2%. Next, whole genome sequence data from 1,970 *P. falciparum* isolates collected globally were analysed. Nine samples displayed depth of coverage consistent with *pfhrp2* deletions (0.5%), but the corresponding split-read analysis could not confirm deletions in seven of these samples. Twenty-eight isolates had evidence of *pfhrp3* deletions (1.4%), which are widespread in Southeast Asia. Finally, a meta-analysis of published data revealed a positive mean association between the frequencies of *pfhrp2* and *pfhrp3* deletions in Africa and South America. This result suggested a shared selective pressure acting on these genetic variants. In conclusion, evidence of genetic selection on both *pfhrp2* and *pfhrp3* deletions was presented, but experimental crosses do not provide evidence of a fitness cost of these variants. Further work is urgently needed to accurately determine the prevalence and the degree of association between these genetic variants, and the respective impact on diagnostic accuracy of many in-use RDT.

Keywords: genetic selection, malaria, rapid diagnostic test, gene deletion, genomic data.

ACCEPTED MANUSCRIPT

1. Introduction

The past decade has witnessed a significant decline in the number of malaria cases and related deaths across the globe (*WHO | World Malaria Report 2015*, 2016). The decline has been attributed to many factors including an increased frequency of diagnostic tests performed on suspected cases, leading to improvements in access to appropriate treatment. Among the diagnostic tools available, rapid diagnostic tests (RDT), which comprise a small chromatography cassette that detects parasite-specific antigens in patient blood, have the advantage of being easily applicable and reliable for use in routine clinical and community-based care, including in remote and underserved endemic settings. For *Plasmodium falciparum* infections, the most popular RDT are based on the detection of histidine-rich protein-2 (*pfhrp2*) antigen. However, this approach may be compromised by the emergence of *P. falciparum* parasites with deleted *pfhrp2*. South America was the first region to report the occurrence of false negative RDT results due to this gene deletion (Gamboa et al., 2010; Akinyi et al., 2013; Abdallah et al., 2015; Akinyi Okoth et al., 2015; Murillo Solano et al., 2015; Rachid Viana et al., 2017). Similar evidence was followed by more recent studies from Africa (Dolo et al., 2012; Wurtz et al., 2013; Parr et al., 2016; Beshir et al., 2017) and Asia (Kumar et al., 2013; Li et al., 2015; Bharti et al., 2016). Driven by the high impact of *pfhrp2*-detecting RDT use on malaria control, mathematical modelling exercises have identified low transmission regions of sub-Saharan Africa as the priority for *pfhrp2*-deletion investigations (Watson et al., 2017).

In the *P. falciparum* genome, *pfhrp2* is located in a sub-telomeric region of chromosome 8. Genetic variation is a hallmark of this gene due to diferent histidine-alanine-rich motifs repeated a different number of times in parasite samples collected across the globe (Baker et al., 2010). However, deletions of this gene appear to be mostly at low frequency in Africa, Asia and Pacific (Baker et al., 2010). Interestingly, in the 3D7 reference genome, this gene shares 85-90% homology in nucleotide sequence with *pfhrp3*, located in a sub-telomeric region of chromosome 13 (Wellems and Howard, 1986). Such sequence similarity suggests that they encode related protein isoforms that have a degree of functional redundancy. Thus, in absence of *pfhrp2*, RDT positivity could be retained due to *pfhrp3* expression (Baker et al., 2010; Beshir et al., 2017). If this hypothesis is true, both genes, although genetically unlinked, would be under a shared selective pressure in *P. falciparum* endemic countries.

ACCEPTED MANUSCRIPT

The location of the two loci in sub-telomeric regions places them among antigenic diversity gene families, such as those encoding STEVOR or rifin, which are prone to recombination and genomic instability (Bopp et al., 2013). In theory, deletions of *pfhrp2* and *pfhrp3* are then expected to occur more frequently than in other loci in the *P. falciparum* genome.

To better understand the segregation of these important deletions and their potential fitness cost, the present study first analysed whole-genome sequencing data from three experimental crosses: 3D7 x HB3, HB3 x DD2, and 7G8 x GB4 (Walliker et al., 1987; Wellems et al., 1990; Hayton et al., 2008). The analysis of this data was driven by the fact that the DD2 and HB3 lines show *pfhrp2* and *pfhrp3* deletions in their genome, respectively (Kemp et al., 1987; Wellems et al., 1987; Baker et al., 2010). The cross between 7G8 and GB4 lines, both with intact *pfhrp2* and *pfhrp3* loci was included to calculate the frequency of spontaneously generating deletions of these genes during sexual recombination. A global analysis of *pfhrp2* and *pfhrp3* gene deletions was then performed using a large collection of whole-genome sequencing data of field isolates from 20 countries. Finally, a meta-analysis of existing data was performed to estimate the linkage disequilibrium between *pfhrp2* and *pfhrp3* gene deletions.

2. Methods

2.1 Genomic data from lab strains and experimental crosses

A total of 98 publicly available samples was downloaded from the MalariaGEN *Pf* genetic crosses project website (Miles et al., 2016). These samples refer to the parental lines 3D7 (n=1), HB3 (n=2), DD2 (n=1), 7G8 (n=1) and GB4 (n=1) and the respective progeny from the following experimental crosses: 3D7 x HB3 (n=19), HB3 x DD2 (n=35), and 7G8 x GB4 (n=38). These lab strains are originally from West Africa (3D7), Honduras (HB3), Indochina (DD2), Brazil (7G8) and Ghana (GB4).

The MalariaGEN Pf community project provides an invaluable resource of whole-genome sequencing data of field samples (Manske et al., 2012; Miotto et al., 2015). In this study, a subset of 1,970 high-quality sequencing data from this project (Gomes et al., 2017; Ravenhall et al., 2016) was analysed. The samples represented a total of 20 countries (South America 2, Asia/Oceania 8, and Africa 10) with sample sizes ranging from 2 (India) to 399 (Cambodia). The raw data and alignment "bam" files were processed as described in detail elsewhere (Campino et al., 2016). In

ACCEPTED MANUSCRIPT

brief, each of these whole-genome samples was mapped onto the 3D7 reference genome (v3) using bwa-mem. The genome-wide and *pfhrp2/3* gene coverage was determined using samtools (http://samtools.sourceforge.net/).

All samples from the genetic crosses and the parental lines had an overall mean coverage of at least 20 reads per position (Table 1). Similar overall mean coverage was obtained for the field samples (Supplementary Data 1). Additionally, in both lab and field samples, there was an average coverage of 25 reads per position in the highly conserved Aldolase and L-lactate dehydrogenase LDH genes (*Pf3D7_1444800* and *Pf3D7_1324900*, respectively) which also produce antigens targeted by some malaria RDT.

2.2 Detection of *pfhrp2/pfhrp3* deletions

The genomic analysis was focused on 100-kb extended loci of the *pfhrp2* and *pfhrp3* genes (from 1,355,000 to 1,455,000 at chromosome 8 and 2,800,000 to 2,900,000 at chromosome 13, respectively). The number of reads mapped onto each position (e.g., coverage) of these extended loci were determined from the respective bam files using samtools (www.htslib.org). As mentioned in the Introduction, *pfhrp2* and *pfhrp3* share high sequence homology in the 3D7 reference genome and, therefore, coverage of each sample might be affected by reads that could be mapped onto either gene. To understand the extent of this potential problem, the 3D7 nucleotide sequences of the two genes were initially aligned using clustal multiple sequence alignment by muscle (www.ebi.ac.uk/Tools/msa/muscle/). The respective alignment showed differences scattered around these aligned sequences (Supplementary Figure 1A). In this scenario, the mapping quality scores were expected to correctly assign reads to the corresponding gene. This expectation was confirmed by a kmer analysis where almost every position of the 3D7 *pfhrp2* and *pfhrp3* sequences appears to be unique (Supplementary Figure 1B). Hence, the effect of multi-mappers on coverage appears to be negligible.

Since current methods to detect gene deletions using whole-genome sequencing data require the analysis of the coverage profile for the whole genome (Boeva et al., 2012; Sepúlveda et al., 2013), a simple algorithm was implemented in order to perform deletion calling of the *pfhrp2* and *pfhrp3* genes. For each sample, the percentage of positions with coverage $\leq$ two reads was calculated for the coding

ACCEPTED MANUSCRIPT

region of each gene; the two-read cut-off allowed for some coverage at putative deletions due to mapping or sequencing errors. Gene deletions were then called out in samples where that percentage was at least 75%. A second cut-off of 10 reads was used to identify putative polyclonal samples with deleted and non-deleted *pfhrp2* and *pfhrp3* genes (i.e., mixed genotype call). To strengthen the call of these mixed infections, this analysis was extended to the 1-kb downstream and upstream flanking regions of each gene under analysis. A mixed genotype of the *pfhrp2* and *pfhrp3* genes was only considered for samples where the same genotype call was obtained for the coding regions of these genes and at least one of the respective flanking regions. Genotype calls based on larger flanking regions (2.5kb, 5kb and 10kb) were also performed, but they only generated consistent results for samples with large deletions. To increase the evidence for mixed genotype calls, the alignment files of the detected samples were explored with the aim of identifying split reads mapped onto the flanking positions defining the putative deleted regions. In this analysis, a mix of split reads (e.g., evidence for a deletion) and non-split reads (e.g., evidence for a non-deleted region) was expected to map onto these genomic positions.

Since the above algorithm could generate highly stringent cutoff values to detect mixed infections in samples with high depth of coverage, two additional methods were used in the analysis. The first one determined a coverage cutoff for each sample based on 10% of the median coverage across the genome. This method did not perform well, because the cutoff values for calling a deletion or a mixed genotype were very close to each other for samples with moderate depth of coverage. As expected, a higher number of mixed infections was indeed identified in the field samples with a high depth of coverage. However, when the respective coverage profiles were visualised, they showed erratic random fluctuations that seemed inconsistent with a stable coverage depth expected to be observed in samples with mixed infections. The second method was based on a similar idea, but with the cutoff values being calculated in the 100-kb extended loci of the *pfhrp2* and *pfhrp3* genes. Again, this method did not show a good performance in the field samples, because the cutoff values generated for mixed infections tended to be very too low (median of 7 reads).

2.3 Meta-analysis of *pfhrp2/pfhrp3* deletion status

ACCEPTED MANUSCRIPT

A meta-analysis was performed on published data where the combined status of *pfhrp2* and *pfhrp3* deletions was reported for each sample. These studies were sought in PubMed using the following search terms: ('pfhrp2', 'pfhrp3' and 'deletion') or ('hrp2', 'hrp3' and 'deletion'). This search retrieved a total of 22 studies at the time of writing (April 13th, 2018). Of these, six studies were excluded from the analysis, because they did not evaluate *pfhrp2* and *pfhrp3* deletions in the field (Ramutton et al., 2012; Mendoza et al., 2013; Das et al., 2018; Parr et al., 2018) or referred to case studies (Houzé et al., 2011; Johora et al., 2017). Other six studies determined the combined genotypes of *pfhrp2* and *pfhrp3* status only on a subset of samples with negative RDT result (Kumar et al., 2013; Wurtz et al., 2013; Li et al., 2015; Bharti et al., 2016; Parr et al., 2016; Gupta et al., 2017). These studies were discarded from the analysis, because the respective data were intrinsically incomplete and biased by study design. Meta-analysis was then performed on the remaining ten studies. These studies consisted of data from Bolivia (Rachid Viana et al., 2017), Brazil (Rachid Viana et al., 2017), Colombia (Murillo Solano et al., 2015; Dorado et al., 2016), French Guiana (Trouvay et al., 2013), Guyana (Akinyi Okoth et al., 2015), Honduras (Abdallah et al., 2015), Peru (Gamboa et al., 2010), and Suriname (Akinyi Okoth et al., 2015) in South America, and from Eritrea (Menegon et al., 2017; Berhane et al., 2018), Ghana (Amoah et al., 2016), and Kenya (Beshir et al., 2017) in Africa .

All these studies provided evidence of *P. falciparum* infections using microscopy and PCR amplification. Initial evidence of deletion was confirmed by PCR amplification of *pfhrp2* and *pfhrp3* and their flanking regions to infer the extent of the respective deletions. To test the quality of DNA in the samples and confirm the gene deletions, further PCR amplification was carried out in other genes, such as the merozoite surface protein-1 or -2 genes (Gamboa et al., 2010; Murillo Solano et al., 2015; Dorado et al., 2016; Beshir et al., 2017; Rachid Viana et al., 2017). However, the recommended antigen analysis based on re-testing the samples with a different *pfhrp2*-based RDT or using an enzyme-linked immunosorbent assay to quantify the levels of the *pfhrp2* protein (Cheng et al., 2014) in a sample was only performed in three studies (Gamboa et al., 2010; Dorado et al., 2016; Rachid Viana et al., 2017).

The meta-analysis was conducted using Bayesian methods with the objective of estimating the mean proportion of *pfhrp2* and *pfhrp3* gene deletions and the respective log odds ratio. In this analysis, a Multinomial distribution was assumed for

ACCEPTED MANUSCRIPT

the 2 x 2 frequency table of each study. To describe study heterogeneity, a Dirichlet distribution was additionally assumed for the joint probabilities of the combined *pfhrp2* and *pfhrp3* genotypes from each study. The resulting statistical model was the Multinomial-Dirichlet distribution (MDD). The analysis was performed assuming a single MDD for all studies (global analysis), and a different MDD for studies from Africa and South America.

In terms of Bayesian specification, a Gamma distribution with both shape and scale hyperparameters equal to 0.001 was assumed for each parameter of the MDD in order to obtain non-informative and independent prior distributions. Markov Chain Monte Carlo methods via the software JAGS (http://mcmc-jags.sourceforge.net/) were used to perform posterior inference. A single chain of 110,000 iterations were generated for each analysis, where the first 10,000 iterations were considered the respective burn-in period. A lag of 10 iterations assured a total of 10,000 independent values of the posterior distributions. The respective script and output files are available as Supplementary Data 2 and 3.

2.4 Statistical inference

To analyse whole-genome sequencing data from experimental crosses, the Clopper-Pearson confidence interval at 95% was used to estimate the prevalence of *pfhrp2* and *pfhrp3* gene deletions. The Fisher's exact test was performed to compare the prevalence of each gene deletion across different experimental populations under a 5% significance level. Since whole-genome sequencing samples from the field were based on different sampling strategies (not entirely random), only the observed prevalence of *pfhrp2* and *pfhrp3* deletions was reported here.

With respect to meta-analysis, statistical inference was based on the mean prevalence of *pfhrp2* and *pfhrp3* deletions and the corresponding odds ratio in log scale (log Δ) using the posterior mean and the 95% credible interval of these parameters. In particular, the log Δ was calculated to measure the strength of association between *pfhrp2* and *pfhrp3* deletions. A positive value of log Δ provided evidence for a propensity of *pfhrp2* and *pfhrp3* deletions occurring together in the same infected individual, whereas a negative value suggested a propensity for a deleted *pfhrp2* gene occurring together with an intact *pfhrp3* gene (or the other way

ACCEPTED MANUSCRIPT

around). The whole statistical analysis was performed in the base R software version 3.3.2 (https://www.r-project.org/). No specific packages were used in this analysis.

3. Results

3.1 Deletions of *pfhrp2/pfhrp3* genes in lab strains

Genomic data from 3D7, HB3, DD2, 7G8, and GB4 lab lines were interrogated as proof-of-principle that the datasets captured the known *pfhrp2* and *pfhrp3* genotypes of each line. Read coverage of *pfhrp2* gene was effectively zero in DD2 with normal coverage for *pfhrp3* (Table 1 and Figure 1A). There was zero coverage of *pfhrp3* in both genome data sets from the HB3 line (Figure 1B). In the DD2 line, the deletion of *pfhrp2* was included in a regional loss spanning the positions of 1,368,994 and 1,383,070 (14kb) on chromosome 8. With respect to the HB3 line, there was evidence for a large deleted region of about 47kb spanning the positions 2,807,552 and 2,854,158 at chromosome 13, where *pfhrp3* is located. Although *pfhrp2* appears intact in the HB3 line, a possible deletion might also be present in an upstream region of this gene at chromosome 8 (1,392,075-1,398,230; 6kb). This region includes two genes encoding distinct PHISTa-like proteins with unknown function (PF_0832200 and PF_0832300). As positive controls for the presence of both *pfhrp2* and *pfhrp3* genes, the genomic data of the 7G8 and GB4 lines showed sufficient coverage for the extended *pfhrp2* and *pfhrp3* loci on their respective chromosomes (Supplementary Figure 2).

3.2 Deletions of *pfhrp2* and *pfhrp3* genes in the experimental crosses

Data from different experimental crosses were analysed in order to understand the inheritance of *pfhrp2* and *pfhrp3* deletions and to estimate the frequency of spontaneously occurring gene deletions during sexual recombination between parental strains with intact genes at both loci.

As expected, genomic data of HB3 x DD2 could identify single or double *pfhrp2* and *pfhrp3* deletions in the respective progeny (Table 2). In particular, the detection of clones with deletions of a single gene demonstrated the ability to uniquely map reads onto these genes sharing a high sequence homology. There were 21 out of 37 clones with *pfhrp2* deletion (95%CI=(39.5%;72.9%)). This result showed that this deletion could be maintained upon genetic crossing and its frequency did not depart from a Mendelian prediction. As expected from two genetically unlinked genes, the

ACCEPTED MANUSCRIPT

inheritance patterns of *pfhrp2* and *pfhrp3* deletions appears to be independent of each other (p=0.40; Fisher's exact test excluding the frequency of mixed calls). Note that there were two samples with possible mixed genotypes (i.e., both deleted and intact forms were present at one locus) for *pfhrp2* gene (Figure 2A; Table 2). Split-read analysis revealed the presence of split reads in the flanking position defining the beginning of the deletion (Supplementary Figure 3A). However, these split reads appeared to be orphaned of the respective second fragment, suggesting a possible insertion-deletion.

In 3D7 x HB3, there were no clones lacking *pfhrp3* (0 out of 19 clones; 95%CI=(0.0%;16.9%)). However, this result should be interpreted with caution due to the small number of clones analyzed (n=19). In the alternative HB3 x DD2, there were five out of 37 clones with evidence of *pfhrp3* deletions (95%CI=(4.3%;28.8%)). However, this frequency was not statistically different from the one observed in the 3D7 x HB3 cross (Fisher exact's test, p=0.15). Therefore, under these lab conditions, the segregation of the *pfhrp3* deletion derived from HB3 might not affected by the second parental strain used in the genetic crosses.

No evidence was found for any *pfhrp2* or *pfhrp3* deletion in 7G8 x GB4 (n=38; CI95%=(0.0%;9.2%)). Combining this result with the one from 3D7 x HB3 where no *pfhrp2* deletions were also found, the overall prevalence of spontaneously generating a *pfhrp2* deletion from parasites with intact genes was estimated to be less than 6% (n=58 clones; 95%CI=(0.0%;6.2%)). The spontaneous generation of *pfhrp2* and *pfhrp3* deletions in these experimental settings appears to be a relatively rare genomic event, but larger sample sizes are needed to investigate this further.

3.3 Global analysis of deletions in field Isolates

The analysis proceeded with the detection of *pfhrp2* and *pfhrp3* deletions in 1,940 field isolates (Table 3). There was a predominance of samples from Southeast Asia, namely, Cambodia (n=499), Thailand (n=219), and Vietnam (n=180), where the frequency of the deletions is currently unknown. There was also a total of 25 isolates from South America, namely, Colombia and Peru, two countries where the presence of parasites lacking *pfhrp2* and *pfhrp3* has been reported.

The stringent deletion calling algorithm detected a total of 34 isolates with evidence for deletion in either gene (Table 3; Supplementary Figure 4). There were nine isolates (0.5%) with evidence for *pfhrp2* deletions, where two Kenyan isolates were

ACCEPTED MANUSCRIPT

the only ones detected with almost zero coverage. The remaining seven isolates appear to be mixed genotype calls. Figure 2B illustrates two examples of coverage profiles from Cambodia and Vietnam with evidence for these mixed infections. However, when the read alignment files were mined, there was no clear evidence for the presence of split reads and, therefore, mixed calls could have just resulted from random fluctuations of coverage (Supplementary Figure 3B).

In contrast to *pfhrp2*, there was a higher frequency of samples with possible *pfhrp3* deletions (1.4%, 28 out of 1,940; Supplementary Figure 4). Interestingly, the geographical location of these samples was spread over Southeast Asia (India, n=1; Cambodia, n=14; Laos, n=1; Thailand, n=1; Vietnam, n=8). Outside Southeast Asia, only two isolates were found with this gene deletion, one from Peru and another from Kenya. Finally, five samples showed evidence for possible mixed genotype call for *pfhrp3* (Cambodia, n=3; Vietnam, n=2). Again, these mixed genotypes could not be confirmed by the split-read analysis.

3.4 Meta-analysis of *pfhrp2/pfhrp3* deletions

To further understand the level of association between *pfhrp2* and *pfhrp3* deletions, a meta-analysis was performed using data from ten studies of which six were from South America and four from Africa. These ten studies encompassed 15 different sites in total (Table 4). The sample sizes ranged from 25 (Bolivia) to 253 (Colombia). According to a separate MDD for data of each continent, the highest estimated probability of *pfhrp2* and *pfhrp3* deletions were observed in Ghindae/Massawa of Eritrea (π_{HRP2-}=0.58; 95%CI=(0.45;0.71) and Bolivia (π_{HRP3-}=0.64; 95%CI=(0.46;0.81)), respectively. The study from Guyana was the only one where no *pfhrp2* and *pfhrp3* deletions were observed. Interestingly, data from Puerto Lempira in Honduras showed evidence for the existence of a high percentage of *pfhrp3* deletions (π_{HRP3-}=0.44; 95%CI=(0.33;0.55)) in the presence of an intact *pfhrp2* gene (π_{HRP2-}=0.01; 95%CI=(0.00;0.03)). These data agreed with the Honduran origin of HB3 line (Bhasin and Trager, 1984) and a possible genetic selection of *pfhrp3* deletion not related to the use of HRP2-detecting RDT in this country.

At the continent level, the meta-analysis suggested a higher mean proportion of *pfhrp2* deletions in the African studies than in the ones from South America (posterior means of 0.28 and 0.14, respectively; Figure 3A). Interestingly, the mean

ACCEPTED MANUSCRIPT

proportion of *pfhrp3* deletions appeared not to substantially vary between studies from different continents (posterior means of 0.40 and 0.34, respectively; Figure 3A). Since *pfhrp2* and *pfhrp3* genes share high homology in nucleotide sequence (85-90%) in the 3D7 reference genome, a putative positive association was hypothesized between the frequencies of *pfhrp2* and *pfhrp3* deletions due to a shared selection pressure resulting from the use of HRP2-detecting RDT. The point estimates of the log odds ratio varied from -5.55 (Beni in Bolivia) to 10.86 (7 Departments of Colombia) (Table 4). No evidence for association between *pfhrp2* and *pfhrp3* deletions was found for Beni in Bolivia, Pará state in Brazil, French Guiana, Georgetown in Guyana, and Puerto Lempira in Honduras. This lack of association could be explained by the low frequency of *pfhrp2* deletions. Evidence for a positive association was in turn found for ten sites (Gash Barka/Debub and Ghindae/Massawa in Eritrea, Accra and Cape Coast in Ghana, Mbita in Kenya, Acre and Rondonia states from Brazil, both studies from Colombia, and Paramaribo in Suriname). The Peruvian study was the only one providing evidence for a negative association between the *pfhrp2* and *pfhrp3* deletions (n=148; log Δ=-1.05; 95%CI=(-1.74;-0.36)). When these results were combined using a separate MDD for the data of each continent, there was evidence for a positive mean association between *pfhrp2* and *pfhrp3* deletions in both Africa and South America (Figure 3A). Since the data from the studies analysed referred to the combined genotype of the *pfhrp2* and *pfhrp3* genes at the individual level, the positive association between the corresponding genotype frequencies was indicative of a propensity of the *pfhrp2* and *pfhrp3* deletions occurring in the same infected sample. A similar result was obtained when the whole data set was analysed using a single MDD. Finally, this positive association was negatively but weakly correlated with the mean log odds of *pfhrp2* deletions in the global analysis and the analysis of Africa and South America data (Pearson's correlation coefficient=-0.07, -0.11, and -0.09, respectively; Figure 3B).

When assessing the data to be included in the meta-analysis, the data of the Ghanaian study showed some inconsistencies with respect to the *pfhrp2* and *pfhrp3* gene deletions. In particular, there were some positive RDT samples with contradicting evidence for deletions of both *pfhrp2* and *pfhrp3* genes (see Figure 3 of Amoah et al., 2016). Meta-analysis was then repeated without the data from this study. There was evidence for a minor impact of this study on the posterior

ACCEPTED MANUSCRIPT

distributions of mean probability *pfhrp2* and *pfhrp3* deletions irrespective of the analysis (Figure 3A). The major impact was observed for the posterior distribution of average log odds ratio between probability of *pfhrp2* and *pfhrp3* deletions in studies from Africa. However, qualitatively, data still suggested a positive association between the probability of *pfhrp2* and *pfhrp3* deletions in those studies.

4. Discussion

This study demonstrates the potential use of whole-genome sequencing data to detect *pfhrp2* and *pfhrp3* deletions in laboratory and field isolates of *P. falciparum*. Genomic data from experimental crosses also showed the potential of identifying possible polyclonal samples with a mix of parasites exhibiting deleted and non-deleted genes. However, this potential could not be confirmed in the field samples where all samples with possible mixed genotypes failed to show evidence for the presence of split reads in the respective read alignments. Since sequence diversity is the hallmark of *pfhrp2* and *pfhrp3* (Baker et al., 2010), the low coverage of these samples could have resulted from monoclonal infections with parasites showing sequences of *pfhrp2* and *pfhrp3* very different from the ones in the 3D7 reference genome. This investigation was, however, out of the scope of this study.

The genomic data of field samples could identify parasites lacking *pfhrp2* and *pfhrp3*. In contrast, a previous global analysis of these genes did not report any of these gene deletions using PCR amplification (Baker et al., 2010). This discrepancy in findings could be explained by the small sample sizes from each country in the Baker et al analysis, the challenge of identifying gene deletions hidden in polyclonal infections using PCR, and the possible problems associated with the molecular investigation of these gene deletions, as discussed in depth elsewhere (Cheng et al., 2014; Parr et al., 2018).

Data from the experimental crosses provided invaluable information on the segregation pattern of *pfhrp2* and *pfhrp3* deletions. In 3D7 x HB3 and HB3 x DD2, there was evidence for a lower frequency of *pfhrp3* gene deletion in the respective progeny than the one predicted by Mendelian segregation. A sampling bias towards self-fertilization of the parental line carrying the *pfhrp3* gene was clearly ruled out as the progeny of 3D7 x HB3 and HB3 x DD2 were mostly recombinant. In the past, it was hypothesized that parasites carrying *pfhrp3* gene have a competitive advantage in the laboratory conditions or in the chimpanzee host used in the genetic crosses

ACCEPTED MANUSCRIPT

(Wellems et al., 1987). A recent study has reported evidence for similar distributions of multiplication rates between 3D7 and HB3 in lab cultures (Murray et al., 2017). This study also suggested similar distributions of multiplication rates in heterologous cultures between DD2 and HB3 lines. Bringing these results together, the chance of finding rare *pfhrp3*-deleted progeny in 3D7 x HB3 and HB3 x DD2 is possibly due to be a better adaptation of the parasites carrying *pfhrp3* in the host instead of *in vitro* conditions. This tendency appears to explain the near-absence of *pfhrp3* deletions in whole-genome samples from African countries where malaria transmission intensity is higher than elsewhere. In general, transmission intensity is positively correlated with the frequency of polyclonal infections (Fola et al., 2017; Manjurano et al., 2011) and, therefore, populations from high transmission settings are likely to purge *pfhrp3*-deleted parasites in circulation by genetic crossing in the same host. If this interpretation is correct, these deletions can only be maintained in populations with low transmission intensities. This prediction from the genetic crosses is to some extent supported with data from field samples where *pfhrp3* deletions were only found across Southeast Asia and South America. In the same line of evidence, meta-analysis suggested similar posterior distribution of the mean proportion of *pfhrp3* deletions between African sites in relatively low intensity transmission areas (i.e., Gash Barka and Debub regions in Eritrea, Mbita in Kenya, Cape Coast and Accra in Ghana) and the South American sampling sites.

Data of HB3 x DD2 could not rule out a Mendelian segregation of the *pfhrp2* deletion in the respective progeny. Therefore, under experimental conditions, this genetic variant can be propagated through sexual recombination. In a natural environment, the segregation of this deletion is likely to be affected by factors related to parasite adaptation to the host, malaria transmission intensity as reflected by the number of polyclonal infections (e.g., multiplicity of infection) in the population, uncertain impact of loss of flanking loci and, finally, the selective pressure exerted by HRP2-based diagnostic testing. In the same genetic cross, *pfhrp3* deletion showed a lower frequency than the one expected by a Mendelian prediction. Similarly, no clones lacking *pfhrp3* were observed in the progeny of 3D7 x HB3. This unusual pattern of inheritance can be explained by a strong selective advantage of that progeny to the lab hosts (e.g., chimpanzees) and/or the experimental conditions of the *in vitro* cultures, as hypothesized previously (Wellems et al., 1987). Evidence for an independent segregation of *pfhrp2* and *pfhrp3* deletions in HB3 x DD2 suggests that

ACCEPTED MANUSCRIPT

the selective advantage to the genetic cross conditions is driven by *pfhrp3* alone, generating distinct inheritance patterns of *pfhrp2* and *pfhrp3* deletions.

From the malaria control standpoint, it was important to estimate the probability of generating a spontaneous *pfhrp2* deletion. According to the data from the 3D7 x HB3 and HB3 x DD2 crosses, this probability appears to be less than 6%. This estimate would, however, have benefited from larger sample sizes to reduce estimation uncertainty.

In contrast to the meta-analysis where the mean prevalence of *pfhrp2* deletions was globally estimated at ~17%, field samples harbouring these deletions were nearly absent, a finding which is in agreement with a previous global analysis of *pfhrp2* deletions (Baker et al., 2010). However, this apparent low prevalence of *pfhrp2* deletions in the field isolates should be interpreted with caution. These field samples have a significant selection bias, because they were collected under different conditions and for distinct scientific purposes.

The meta-analysis provided global estimates for the mean frequency of *pfhrp2* and *pfhrp3* deletions, respectively, though it has some caveats. Firstly, the studies considered different populations, such as asymptomatic children (Amoah et al., 2016; Beshir et al., 2017), children and adults (Menegon et al., 2017), or symptomatic patients (Abdallah et al., 2015; Dorado et al., 2016; Rachid Viana et al., 2017). Secondly, some studies referred to historical samples available from the respective populations (Trouvay et al., 2013) or to data passively collected from health centres or clinics (Akinyi Okoth et al., 2015). These data are intrinsically biased and do not reflect a true random sampling of the population. Thirdly, the studies were conducted in different years and the mean prevalence estimates of *pfhrp2* and *pfhrp3* deletions should be regarded as aggregated averages over the time period of the studies. Fourthly, some studies from Africa and Asia were excluded from this analysis, because they analysed only a subset of samples with negative HRP2-based RDT results.

An overall difficulty associated with the meta-analysis was related to the data quality of the reported *pfhrp2* and *pfhrp3* deletions. As mentioned earlier, some concerns about data quality were raised for the Ghanaian study where, for example, RDT positive samples lacked both *pfhrp2* and *pfhrp3* genes (Amoah et al., 2016). In theory, RDT positivity could be obtained from people without current infection, but with the HRP2 protein still circulating in the blood. Therefore, these samples should

ACCEPTED MANUSCRIPT

have been discarded as putative *pfhrp2* and *pfhrp3* deletions. This and other scientific inconsistencies found in the literature motivated a general discussion on how to carefully investigate and interpret evidence for these deletions (Cheng et al., 2014; Woodrow and Fanello, 2017; Parr et al., 2018). Current recommendations (Cheng et al., 2014) require that initial evidence should be gathered by testing if a sample is confirmed as either microscopy or PCR positive. Confirmatory evidence for gene deletion should contemplate the molecular amplification of different gene segments of *pfhrp2* and *pfhrp3* and the amplification of the respective flanking regions. An antigen analysis where the presence of the pfHRP2 protein is detected by ELISA or using a different non-pfHRP2-detecting RDT should also be performed. Unfortunately, all of these recommendations were not followed in the published data included in this study and, therefore, the meta-analysis might reflect the data quality of the original studies. Recently, a new testing algorithm was proposed for the investigation of *pfhrp2* and *pfhrp3* deletions (Parr et al., 2018). This algorithm appeared to produce a tenfold improvement in the underlying limit of detection, but remains to be tested in different populations and lab conditions. As a future alternative, the richness of genomic data together with a decrease in the sequence cost show promise to revolutionize the investigation not only of *pfhrp2* and *pfhrp3* deletions, but also of their relationship with other genetic variants important for malaria control and treatment.

In summary, the genomic data of the HB3 x DD2 cross showed evidence for a Mendelian segregation of *pfhrp2* deletion, thus, suggesting no fitness cost of this genetic variant in experimental settings. The same genomic data also suggested an independent segregation between deletions of *pfhrp2* and *pfhrp3*, as predicted by two unlinked genes in the absence of selection. The genomic data from field samples were inconclusive about the prevalence of *pfhrp2* deletion in natural populations due to convenience sampling. However, there was a strong evidence for a wide spread of *pfhrp3* deletion in Southeast Asia. The meta-analysis enabled to infer a positive association between these two gene deletions in the same parasite. Notwithstanding the problems and challenges highlighted in the investigation of these deletions in field studies, a positive association is consistent with a shared selective pressure of these two genes caused by HRP2-based diagnostic testing and subsequent treatment. Future studies will help to confirm or disprove this

ACCEPTED MANUSCRIPT

interpretation, thus, helping public health authorities to better decide possible changes in current RDT policies.

Acknowledgements

This work was supported by Fundação para a Ciência e Tecnologia, Portugal (NS; grant ref. UID/MAT/00006/2013) and the Medical Research Council UK (TGC; Grant no. MR/K000551/1 and MR/M01360X/1, MR/N010469/1, MC_PC_15103).

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

References

Abdallah, J.F., Okoth, S., Fontecha, G.A., Torres, R., Banegas, E.I., Matute, M., Bucheli, S., Goldman, I.F., de Oliveira, A., Barnwell, J.W., Udhayakumar, V., 2015. Prevalence of pfhrp2 and pfhrp3 gene deletions in Puerto Lempira, Honduras. Malar. J. 14, 19. https://doi.org/10.1186/s12936-014-0537-7

Akinyi, S., Hayden, T., Gamboa, D., Torres, K., Bendezu, J., Abdallah, J.F., Griffing, S.M., Quezada, W.M., Arrospide, N., De Oliveira, A.M., Lucas, C., Magill, A.J., Bacon, D.J., Barnwell, J.W., Udhayakumar, V., 2013. Multiple genetic origins of histidine-rich protein 2 gene deletion in Plasmodium falciparum parasites from Peru. Sci. Rep. 3, 2797. https://doi.org/10.1038/srep02797

Akinyi Okoth, S., Abdallah, J.F., Ceron, N., Adhin, M.R., Chandrabose, J., Krishnalall, K., Huber, C.S., Goldman, I.F., Macedo de Oliveira, A., Barnwell, J.W., Udhayakumar, V., 2015. Variation in Plasmodium falciparum Histidine-Rich Protein 2 (Pfhrp2) and Plasmodium falciparum Histidine-Rich Protein 3 (Pfhrp3) Gene Deletions in Guyana and Suriname. PLoS One 10, e0126805. https://doi.org/10.1371/journal.pone.0126805

Amoah, L.E., Abankwa, J., Oppong, A., 2016. Plasmodium falciparum histidine rich protein-2 diversity and the implications for PfHRP 2: based malaria rapid diagnostic tests in Ghana. Malar. J. 15, 101. https://doi.org/10.1186/s12936-016-1159-z

Baker, J., Ho, M.-F., Pelecanos, A., Gatton, M., Chen, N., Abdullah, S., Albertini, A., Ariey, F., Barnwell, J., Bell, D., Cunningham, J., Djalle, D., Echeverry, D.F., Gamboa, D., Hii, J., Kyaw, M.P., Luchavez, J., Membi, C., Menard, D., Murillo, C., Nhem, S., Ogutu, B., Onyor, P., Oyibo, W., Wang, S.Q., McCarthy, J., Cheng, Q., 2010. Global sequence variation in the histidine-rich proteins 2 and 3 of Plasmodium falciparum: implications for the performance of malaria rapid diagnostic tests. Malar. J. 9, 129. https://doi.org/10.1186/1475-2875-9-129

Berhane, A., Anderson, K., Mihreteab, S., Gresty, K., Rogier, E., Mohamed, S., Hagos, F., Embaye, G., Chinorumba, A., Zehaie, A., Dowd, S., Waters, N.C., Gatton, M.L., Udhayakumar, V., Cheng, Q., Cunningham, J., 2018. Major Threat to Malaria Control Programs by *Plasmodium falciparum* Lacking Histidine-Rich Protein 2, Eritrea. Emerg. Infect. Dis. 24, 462–470. https://doi.org/10.3201/eid2403.171723

Beshir, K.B., Sepúlveda, N., Bharmal, J., Robinson, A., Mwanguzi, J., Busula, A.O., de Boer, J.G., Sutherland, C., Cunningham, J., Hopkins, H., 2017. Plasmodium falciparum parasites with histidine-rich protein 2 (pfhrp2) and pfhrp3 gene deletions in two endemic regions of Kenya. Sci. Rep. 7, 14718. https://doi.org/10.1038/s41598-017-15031-2

Bharti, P.K., Chandel, H.S., Ahmad, A., Krishna, S., Udhayakumar, V., Singh, N., 2016. Prevalence of pfhrp2 and/or pfhrp3 Gene Deletion in Plasmodium falciparum Population in Eight Highly Endemic States in India. PLoS One 11, e0157949. https://doi.org/10.1371/journal.pone.0157949

Bhasin, V.K., Trager, A., 1984. Gametocyte-forming and non-gametocyte-forming clones of Plasmodium falciparum. Am. J. Trop. Med. Hyg 33, 534–537.

Boeva, V., Popova, T., Bleakley, K., Chiche, P., Cappo, J., Schleiermacher, G., Janoueix-Lerosey, I., Delattre, O., Barillot, E., 2012. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. Bioinformatics 28, 423–425. https://doi.org/10.1093/bioinformatics/btr670

ACCEPTED MANUSCRIPT

Bopp, S.E.R., Manary, M.J., Bright, A.T., Johnston, G.L., Dharia, N. V, Luna, F.L., McCormack, S., Plouffe, D., McNamara, C.W., Walker, J.R., Fidock, D.A., Denchi, E.L., Winzeler, E.A., 2013. Mitotic evolution of Plasmodium falciparum shows a stable core genome but recombination in antigen families. PLoS Genet. 9, e1003293. https://doi.org/10.1371/journal.pgen.1003293

Campino, S., Benavente, E.D., Assefa, S., Thompson, E., Drought, L.G., Taylor, C.J., Gorvett, Z., Carret, C.K., Flueck, C., Ivens, A.C., Kwiatkowski, D.P., Alano, P., Baker, D.A., Clark, T.G., 2016. Genomic variation in two gametocyte non-producing Plasmodium falciparum clonal lines. Malar. J. 15, 229. https://doi.org/10.1186/s12936-016-1254-1

Cheng, Q., Gatton, M.L., Barnwell, J., Chiodini, P., McCarthy, J., Bell, D., Cunningham, J., 2014. Plasmodium falciparum parasites lacking histidine-rich protein 2 and 3: a review and recommendations for accurate reporting. Malar. J. 13, 283. https://doi.org/10.1186/1475-2875-13-283

Das, S., Peck, R.B., Barney, R., Jang, I.K., Kahn, M., Zhu, M., Domingo, G.J., 2018. Performance of an ultra-sensitive Plasmodium falciparum HRP2-based rapid diagnostic test with recombinant HRP2, culture parasites, and archived whole blood samples. Malar. J. 17, 118. https://doi.org/10.1186/s12936-018-2268-7

Dolo, A., Masinde, G.L., Sagara, I., Diallo, M., Koita, O.A., Doumbo, O.K., Diakité, M., Krogstad, D.J., Konaré, A., Traoré, I., Tall, L.K., Tounkara, A., Ouattara, A., Doumbo, S.N., 2012. False-Negative Rapid Diagnostic Tests for Malaria and Deletion of the Histidine-Rich Repeat Region of the hrp2 Gene †. Am. J. Trop. Med. Hyg. 86, 194–198. https://doi.org/10.4269/ajtmh.2012.10-0665

Dorado, E.J., Okoth, S.A., Montenegro, L.M., Diaz, G., Barnwell, J.W., Udhayakumar, V., Murillo Solano, C., 2016. Genetic Characterisation of Plasmodium falciparum Isolates with Deletion of the pfhrp2 and/or pfhrp3 Genes in Colombia: The Amazon Region, a Challenge for Malaria Diagnosis and Control. PLoS One 11, e0163137. https://doi.org/10.1371/journal.pone.0163137

Fola, A.A., Harrison, G.L.A., Hazairin, M.H., Barnadas, C., Hetzel, M.W., Iga, J., Siba, P.M., Mueller, I., Barry, A.E., 2017. Higher Complexity of Infection and Genetic Diversity of Plasmodium vivax than Plasmodium falciparum across all Malaria Transmission Zones of Papua New Guinea. Am. J. Trop. Med. Hyg. Eur. Public Heal. Microbiol. Train. Program. Eur. Cent. Dis. Prev. Control 96, 630–641. https://doi.org/10.4269/ajtmh.16-0716

Gamboa, D., Ho, M.-F., Bendezu, J., Torres, K., Chiodini, P.L., Barnwell, J.W., Incardona, S., Perkins, M., Bell, D., McCarthy, J., Cheng, Q., 2010. A Large Proportion of P. falciparum Isolates in the Amazon Region of Peru Lack pfhrp2 and pfhrp3: Implications for Malaria Rapid Diagnostic Tests. PLoS One 5, e8091. https://doi.org/10.1371/journal.pone.0008091

Gomes, A.R., Ravenhall, M., Benavente, E.D., Talman, A., Sutherland, C., Roper, C., Clark, T.G., Campino, S., 2017. Genetic diversity of next generation antimalarial targets: A baseline for drug resistance surveillance programmes. Int. J. Parasitol. Drugs Drug Resist. 7, 174–180. https://doi.org/10.1016/j.ijpddr.2017.03.001

Gupta, H., Matambisso, G., Galatas, B., Cisteró, P., Nhamussua, L., Simone, W., Cunningham, J., Rabinovich, N.R., Alonso, P., Saute, F., Aide, P., Mayor, A., 2017. Molecular surveillance of pfhrp2 and pfhrp3 deletions in Plasmodium falciparum isolates from Mozambique. Malar. J. 16, 416. https://doi.org/10.1186/s12936-017-2061-z

Hayton, K., Gaur, D., Liu, A., Takahashi, J., Henschen, B., Singh, S., Lambert, L.,

ACCEPTED MANUSCRIPT

Furuya, T., Bouttenot, R., Doll, M., Nawaz, F., Mu, J., Jiang, L., Miller, L.H., Wellems, T.E., 2008. Erythrocyte Binding Protein PfRH5 Polymorphisms Determine Species-Specific Pathways of Plasmodium falciparum Invasion. Cell Host Microbe 4, 40–51. https://doi.org/10.1016/j.chom.2008.06.001

Houzé, S., Hubert, V., Le Pessec, G., Le Bras, J., Clain, J., 2011. Combined deletions of pfhrp2 and pfhrp3 genes result in Plasmodium falciparum malaria false-negative rapid diagnostic test. J. Clin. Microbiol. 49, 2694–6. https://doi.org/10.1128/JCM.00281-11

Johora, F.T., Hossain, M.E., Alam, M.S., Rahman, R., Kibria, M.G., Nima, M.K., Haque, R., Mohon, A.N., Hougard, T., 2017. Case Report: A Case of Plasmodium falciparum hrp2 and hrp3 Gene Mutation in Bangladesh. Am. J. Trop. Med. Hyg. 97, 1155–1158. https://doi.org/10.4269/ajtmh.16-0884

Kemp, D.J., Thompson, J.K., Walliker, D., Corcoran, L.M., 1987. Molecular karyotype of Plasmodium falciparum: conserved linkage groups and expendable histidine-rich protein genes. Proc. Natl. Acad. Sci. U. S. A. 84, 7672–6.

Kumar, N., Pande, V., Bhatt, R.M., Shah, N.K., Mishra, N., Srivastava, B., Valecha, N., Anvikar, A.R., 2013. Genetic deletion of HRP2 and HRP3 in Indian Plasmodium falciparum population and false negative malaria rapid diagnostic test. Acta Trop. 125, 119–121. https://doi.org/10.1016/j.actatropica.2012.09.015

Li, P., Xing, H., Zhao, Z., Yang, Z., Cao, Y., Li, W., Yan, G., Sattabongkot, J., Cui, L., Fan, Q., 2015. Genetic diversity of Plasmodium falciparum histidine-rich protein 2 in the China–Myanmar border area. Acta Trop. 152, 26–31. https://doi.org/10.1016/j.actatropica.2015.08.003

Manjurano, A., Okell, L., Lukindo, T., Reyburn, H., Olomi, R., Roper, C., Clark, T.G., Joseph, S., Riley, E.M., Drakeley, C., 2011. Association of sub-microscopic malaria parasite carriage with transmission intensity in north-eastern Tanzania. Malar. J. 10, 370. https://doi.org/10.1186/1475-2875-10-370

Manske, M., Miotto, O., Campino, S., Auburn, S., Almagro-Garcia, J., Maslen, G., O'Brien, J., Djimde, A., Doumbo, O., Zongo, I., Ouedraogo, J.-B., Michon, P., Mueller, I., Siba, P., Nzila, A., Borrmann, S., Kiara, S.M., Marsh, K., Jiang, H., Su, X.-Z., Amaratunga, C., Fairhurst, R., Socheat, D., Nosten, F., Imwong, M., White, N.J., Sanders, M., Anastasi, E., Alcock, D., Drury, E., Oyola, S., Quail, M.A., Turner, D.J., Ruano-Rubio, V., Jyothi, D., Amenga-Etego, L., Hubbart, C., Jeffreys, A., Rowlands, K., Sutherland, C., Roper, C., Mangano, V., Modiano, D., Tan, J.C., Ferdig, M.T., Amambua-Ngwa, A., Conway, D.J., Takala-Harrison, S., Plowe, C.V., Rayner, J.C., Rockett, K.A., Clark, T.G., Newbold, C.I., Berriman, M., MacInnis, B., Kwiatkowski, D.P., 2012. Analysis of Plasmodium falciparum diversity in natural infections by deep sequencing. Nature 487, 375–379. https://doi.org/10.1038/nature11174

Mendoza, N.M., Cucunubá, Z.M., Aponte, S., González, N.E., Bernal, S.D., 2013. [Field evaluation for diagnostic accuracy of the rapid test SD Bioline Malaria Antigen Pf/Pv® in Colombia]. Biomedica 33, 587–97.

Menegon, M., L'Episcopia, M., Nurahmed, A.M., Talha, A.A., Nour, B.Y.M., Severini, C., 2017. Identification of Plasmodium falciparum isolates lacking histidine-rich protein 2 and 3 in Eritrea. Infect. Genet. Evol. 55, 131–134. https://doi.org/10.1016/j.meegid.2017.09.004

Miles, A., Iqbal, Z., Vauterin, P., Pearson, R., Campino, S., Theron, M., Gould, K., Mead, D., Drury, E., O'Brien, J., Ruano Rubio, V., MacInnis, B., Mwangi, J., Samarakoon, U., Ranford-Cartwright, L., Ferdig, M., Hayton, K., Su, X., Wellems, T., Rayner, J., McVean, G., Kwiatkowski, D., 2016. Indels, structural

ACCEPTED MANUSCRIPT

variation, and recombination drive genomic diversity in *Plasmodium falciparum*. Genome Res. 26, 1288–1299. https://doi.org/10.1101/gr.203711.115

Miotto, O., Amato, R., Ashley, E.A., MacInnis, B., Almagro-Garcia, J., Amaratunga, C., Lim, P., Mead, D., Oyola, S.O., Dhorda, M., Imwong, M., Woodrow, C., Manske, M., Stalker, J., Drury, E., Campino, S., Amenga-Etego, L., Thanh, T.-N.N., Tran, H.T., Ringwald, P., Bethell, D., Nosten, F., Phyo, A.P., Pukrittayakamee, S., Chotivanich, K., Chuor, C.M., Nguon, C., Suon, S., Sreng, S., Newton, P.N., Mayxay, M., Khanthavong, M., Hongvanthong, B., Htut, Y., Han, K.T., Kyaw, M.P., Faiz, M.A., Fanello, C.I., Onyamboko, M., Mokuolu, O.A., Jacob, C.G., Takala-Harrison, S., Plowe, C. V, Day, N.P., Dondorp, A.M., Spencer, C.C.A., McVean, G., Fairhurst, R.M., White, N.J., Kwiatkowski, D.P., 2015. Genetic architecture of artemisinin-resistant Plasmodium falciparum. Nat. Genet. 47, 226–234. https://doi.org/10.1038/ng.3189

Murillo Solano, C., Akinyi Okoth, S., Abdallah, J.F., Pava, Z., Dorado, E., Incardona, S., Huber, C.S., Macedo de Oliveira, A., Bell, D., Udhayakumar, V., Barnwell, J.W., 2015. Deletion of Plasmodium falciparum Histidine-Rich Protein 2 (pfhrp2) and Histidine-Rich Protein 3 (pfhrp3) Genes in Colombian Parasites. PLoS One 10, e0131576. https://doi.org/10.1371/journal.pone.0131576

Murray, L., Stewart, L.B., Tarr, S.J., Ahouidi, A.D., Diakite, M., Amambua-Ngwa, A., Conway, D.J., 2017. Multiplication rate variation in the human malaria parasite Plasmodium falciparum. Sci. Rep. 7, 6436. https://doi.org/10.1038/s41598-017-06295-9

Parr, J.B., Anderson, O., Juliano, J.J., Meshnick, S.R., 2018. Streamlined, PCR-based testing for pfhrp2- and pfhrp3-negative Plasmodium falciparum. Malar. J. 17, 137. https://doi.org/10.1186/s12936-018-2287-4

Parr, J.B., Verity, R., Doctor, S.M., Janko, M., Carey-Ewend, K., Turman, B.J., Keeler, C., Slater, H.C., Whitesell, A.N., Mwandagalirwa, K., Ghani, A.C., Likwela, J.L., Tshefu, A.K., Emch, M., Juliano, J.J., Meshnick, S.R., 2016. *Pfhrp2* -deleted *Plasmodium falciparum* parasites in the Democratic Republic of Congo: A national cross-sectional survey. J. Infect. Dis. 216, jiw538. https://doi.org/10.1093/infdis/jiw538

Rachid Viana, G.M., Akinyi Okoth, S., Silva-Flannery, L., Lima Barbosa, D.R., Macedo de Oliveira, A., Goldman, I.F., Morton, L.C., Huber, C., Anez, A., Dantas Machado, R.L., Aranha Camargo, L.M., Costa Negreiros do Valle, S., Marins Póvoa, M., Udhayakumar, V., Barnwell, J.W., 2017. Histidine-rich protein 2 (pfhrp2) and pfhrp3 gene deletions in Plasmodium falciparum isolates from select sites in Brazil and Bolivia. PLoS One 12, e0171150. https://doi.org/10.1371/journal.pone.0171150

Ramutton, T., Hendriksen, I.C., Mwanga-Amumpaire, J., Mtove, G., Olaosebikan, R., Tshefu, A.K., Onyamboko, M.A., Karema, C., Maitland, K., Gomes, E., Gesase, S., Reyburn, H., Silamut, K., Chotivanich, K., Promnares, K., Fanello, C.I., von Seidlein, L., Day, N.P., White, N.J., Dondorp, A.M., Imwong, M., Woodrow, C.J., 2012. Sequence variation does not confound the measurement of plasma PfHRP2 concentration in African children presenting with severe malaria. Malar. J. 11, 276. https://doi.org/10.1186/1475-2875-11-276

Ravenhall, M., Benavente, E.D., Mipando, M., Jensen, A.T.R., Sutherland, C.J., Roper, C., Sepúlveda, N., Kwiatkowski, D.P., Montgomery, J., Phiri, K.S., Terlouw, A., Craig, A., Campino, S., Ocholla, H., Clark, T.G., 2016. Characterizing the impact of sustained sulfadoxine/pyrimethamine use upon the Plasmodium falciparum population in Malawi. Malar. J. 15, 575.

ACCEPTED MANUSCRIPT

https://doi.org/10.1186/s12936-016-1634-6
Sepúlveda, N., Campino, S.G., Assefa, S.A., Sutherland, C.J., Pain, A., Clark, T.G., 2013. A Poisson hierarchical modelling approach to detecting copy number variation in sequence coverage data. BMC Genomics 14, 128. https://doi.org/10.1186/1471-2164-14-128
Trouvay, M., Palazon, G., Berger, F., Volney, B., Blanchet, D., Faway, E., Donato, D., Legrand, E., Carme, B., Musset, L., 2013. High Performance of Histidine-Rich Protein 2 Based Rapid Diagnostic Tests in French Guiana are Explained by the Absence of pfhrp2 Gene Deletion in P. falciparum. PLoS One 8, e74269. https://doi.org/10.1371/journal.pone.0074269
Walliker, D., Quakyi, I.A., Wellems, T.E., McCutchan, T.F., Szarfman, A., London, W.T., Corcoran, L.M., Burkot, T.R., Carter, R., 1987. Genetic analysis of the human malaria parasite Plasmodium falciparum. Science 236, 1661–6.
Watson, O.J., Slater, H.C., Verity, R., Parr, J.B., Mwandagalirwa, M.K., Tshefu, A., Meshnick, S.R., Ghani, A.C., 2017. Modelling the drivers of the spread of Plasmodium falciparum hrp2 gene deletions in sub-Saharan Africa. Elife 6, e25008. https://doi.org/10.7554/eLife.25008
Wellems, T.E., Howard, R.J., 1986. Homologous genes encode two distinct histidine-rich proteins in a cloned isolate of Plasmodium falciparum. Proc. Natl. Acad. Sci. 83, 6065–6069.
Wellems, T.E., Panton, L.J., Gluzman, I.Y., do Rosario, V.E., Gwadz, R.W., Walker-Jonah, A., Krogstad, D.J., 1990. Chloroquine resistance not linked to mdr-like genes in a Plasmodium falciparum cross. Nature 345, 253–5. https://doi.org/10.1038/345253a0
Wellems, T.E., Walliker, D., Smith, C.L., do Rosario, V.E., Maloy, W.L., Howard, R.J., Carter, R., McCutchan, T.F., 1987. A histidine-rich protein gene marks a linkage group favored strongly in a genetic cross of Plasmodium falciparum. Cell 49, 633–42.
WHO | World Malaria Report 2015, 2016. , WHO. World Health Organization.
Woodrow, C.J., Fanello, C., 2017. Pfhrp2 Deletions in the Democratic Republic of Congo: Evidence of Absence, or Absence of Evidence? J. Infect. Dis. 216, 504–506. https://doi.org/10.1093/infdis/jix345
Wurtz, N., Fall, B., Bui, K., Pascual, A., Fall, M., Camara, C., Diatta, B., Fall, K.B., Mbaye, P.S., Diémé, Y., Bercion, R., Wade, B., Briolant, S., Pradines, B., 2013. Pfhrp2 and pfhrp3 polymorphisms in Plasmodium falciparum isolates from Dakar, Senegal: impact on rapid malaria diagnostic tests. Malar. J. 12, 34. https://doi.org/10.1186/1475-2875-12-34

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Figure 1: Coverage profiles (dark blue lines) of the lab strains 3D7 (A), HB3 (B), and DD2 (C) within *pfhrp2* and *pfhrp3* loci at chromosomes 8 and 13, respectively. The coding region of *pfhrp2/pfhrp3* genes and of other genes in the same locus (e.g., stevor, riffin and PHIST proteins) are shown in salmon and faded rose, respectively. Dashed blue lines represent the average coverage per position at the corresponding chromosome.

Figure 2: Examples of coverage profiles with a possible mixed genotype call (intact/deletion in the same isolate) for *pfhrp2*: (A) B1SD and CH3_61 isolates from HB3 x DD2 cross and (B) two field samples from Cambodia (CAM245) and Vietnam (VIE0101). Split-read analysis confirmed the existence of split reads in the left breakpoints of B1SD and CH3_61 (Supplementary Figure 3A). Similar analysis could not detect any split reads in the field samples with possible mixed genotype (Supplementary Figure 3B). See Figure 1 for further information about the coverage profiles.

Figure 3: Posterior densities of the mean probabilities of *pfhrp2* and *pfhrp3* deletions and the mean of the log odd ratio of the combined *pfhrp2* and *pfhrp3* genotypes (log Δ) (A) in separate meta-analyses including or not the Ghana study. Simulated values of the bivariate posterior distributions of the mean of the log odds of the HRP2 deletion and the mean of the log Δ in the meta-analysis of all studies, where R denotes the Pearson's correlation coefficient (B). These posterior inferences were based on a single MDD for the global analysis and on a different MDD for data each continent.

Supplementary Data 1: Meta-data from the 1970 field samples.

Supplementary Data 2: Model scripts used in meta-analysis.

Supplementary Data 3: Posterior samples obtained from the JAGS for the meta-analysis including or not the Ghanaian study.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Supplementary Figure 1: Sequence analysis of *pfhrp2* and pfhrp3 genes in the 3D7 reference genome. A. Sequence alignment using clustal multiple sequence alignment by muscle. B. Uniqueness of the *pfhrp2* and *pfhrp3* sequences. For each position, the uniqueness was calculated as follows. Using sliding window approach with a window size of 65 and a step size of 1 the sequences are divided into kmers. The number of times the each kmer appears in both gene sequences is stored (#kmer_hits). The uniqueness of each position is represented by 1/#kmers_hits.

Supplementary Figure 2: Coverage profiles (dark blue lines) of the 7G8 and GB8 lines within *pfhrp2* and *pfhrp3* loci at chromosomes 8 and 13, respectively, as explained in Figure 1.

Supplementary Figure 3: Split-read analysis of the *pfhrp2* gene for the potential polyclonal samples in Figure 2. Visualisation of reads from B1SD and CH3_61 clones of HB3 x DD2 (A), and CAM245 and VIE0101 from the field samples (B). Individual reads are represented by the horizontal coloured lines (red: clipped reads and blue: non-clipped reads). The vertical black line indicates the breakpoint of where the reads are clipped. The thick horizontal black line represents the gene overlapping the genomic region.

Supplementary Figure 4: Coverage profiles (dark blue lines) of the 34 field samples with deletions or possible mixed calls for *pfhrp2* and *pfhrp3* genes. See Figure 2 for figure legend.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 1: Summary of coverage for the different lab lines and corresponding progeny.

Samples	Sample size	Range of Mean Coverage		
		overall	*pfhrp2*	*pfhrp3*
3D7	1	122	99	82
HB3	2	80-100	51-88	0
DD2	1	122	0	72
7G8	1	87	98	91
GB4	1	104	59	90
3D7 x HB3	19	41-173	15-121	26-115
HB3 x DD2	35	22-637	0-450	0-329
7G8 x GB4	38	55-250	35-233	30-206

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 2: Frequency of clones from different lab strains and corresponding experimental crosses with no deletion (-), complete deletions (+) and a putative mixed genotype (+/-) of *pfhrp2* and *pfhrp3* and a locus including two PHIST-like proteins in the flanking region of *pfhrp2* (PF_0832200 and PF_0832300). The column 'line' refers to the genotype of each parental line in the genetic crosses.

pfhrp2	PHIST-like genes	*pfhrp3*	Parental line	3D7 x HB3	HB3 x DD2	7G8 x GB4
+	+	+	3D7, 7G8, GB4	3	0	38
+	-	+	N/A	15	13	0
+	-	-	HB3	0	3	0
-	-	+	DD2	0	15	0
-	-	-	N/A	0	2	0
+	+/-	+	N/A	1	0	0
+/-	+/-	+	N/A	0	2	0

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 3: Frequency of field samples with deletion (-) and possible mixed calls (+/-) for *pfhrp2* and *pfhrp3* genes using a flanking region of 1.0kb. Mixed calls could not be confirmed by split-read analysis.

Continent	Country	Sample size	*pfhrp2*		*pfhrp3*	
			-	+/-	-	+/-
Africa	Burkina Faso	47	0	0	0	0
	Congo (DR)	66	0	0	0	0
	Gambia	57	0	0	0	0
	Ghana	198	0	0	0	0
	Guinea	92	0	0	0	0
	Kenya	27	2	0	1	0
	Malawi	213	0	0	0	0
	Mali	55	0	0	0	0
	Nigeria	4	0	0	0	0
	Tanzania	19	0	0	0	0
Asia/Oceania	Bangladesh	50	0	0	0	0
	Cambodia	499	0	3	11	3
	India	2	0	0	1	0
	Laos	104	0	0	1	0
	Myanmar	97	0	0	0	0
	Papua New Guinea	16	0	0	0	0
	Thailand	219	0	1	1	0
	Vietnam	180	0	2	6	2
South America	Colombia	13	0	1	0	0
	Peru	12	0	0	2	0

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 4: Meta-analysis of the association between combined *pfhrp2/pfhrp3* genotype in data from 15 sites across the world, where *n* is the sample size, π_{HRP2} and π_{HRP3-} are the prevalences of *pfhrp2* and *pfhrp3* deletion, respectively, log Δ is the corresponding odds ratio in log scale, and 95% CI is the 95% credible interval. Bayesian estimates were obtained from fitting a different MDD to the data of each continent.

Continent	Country (Site)	*n*	Frequencies of *pfhrp2*/*pfhrp3* genotype				Posterior mean (95% CI)		
			+/+	+/-	-/+	-/-	π_{HRP2-}	π_{HPR3-}	log Δ
Africa	Eritrea (Gash Barka + Debub)	144	81	49	1	13	0.10 (0.06;0.16)	0.43 (0.35;0.51)	3.05 (1.37;5.46)
	Eritrea (Ghindae + Massawa)	50	9	10	0	31	0.58 (0.45;0.71)	0.77 (0.65;0.88)	5.82 (2.22;15.31)
	Ghana (Accra)	179	118	29	25	7	0.18 (0.13;0.24)	0.21 (0.15;0.27)	0.21 (-0.74;1.09)
	Ghana (Cape Coast)	109	46	19	14	30	0.40 (0.31;0.49)	0.45 (0.36;0.54)	1.67 (0.86;2.53)
	Kenya (Mbita)	91	69	4	14	4	0.20 (0.13;0.29)	0.11 (0.05;0.18)	1.57 (0.14;3.02)
South America	Bolivia (Beni)	25	7	17	1	0	0.05 (0.00;0.16)	0.64 (0.46;0.81)	-5.55 (-22.69;1.21)
	Brazil (Acre)	79	45	9	0	21	0.31 (0.21;0.41)	0.38 (0.28;0.48)	3.34 (2.12;4.74)
	Brazil (Pará)	59	29	30	0	0	0.01 (0.00:0.04)	0.50 (0.37;0.62)	2.56 (-16.44;29.37)
	Brazil (Rondonia)	60	49	9	0	2	0.04 (0.01;0.10)	0.19 (0.10;0.30)	10.01 (1.59;36.14)
	Colombia (Amazonas)	100	43	39	5	13	0.18 (0.11;0.26)	0.51 (0.42;0.61)	1.13 (0.03;2.32)
	Colombia (7 Departments)	253	147	91	0	15	0.06 (0.03;0.09)	0.42 (0.36;0.48)	10.86 (2.78;36.35)

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

	French Guiana (6 sites)	81	75	6	0	0		0.00 (0.00;0.03)	0.08 (0.03;0.15)	4.88 (-13.79;29.37)
	Guyana (Georgetown)	97	97	0	0	0		0.00 (0.00;0.02)	0.01 (0.00;0.04)	8.60 (-17.12;30.22)
	Honduras (Puerto Lempira)	68	38	30	0	0		0.01 (0.00;0.03)	0.44 (0.33;0.55)	2.80 (-17.12;30.22)
	Peru (Iquitos and others)	148	20	67	29	32		0.41 (0.33;0.49)	0.66 (0.59;0.74)	-1.05 (-1.74;-0.36)
	Suriname (Paramaribo)	78	66	1	9	2		0.14 (0.07;0.22)	0.05 (0.01;0.10)	2.38 (0.07;4.87)

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

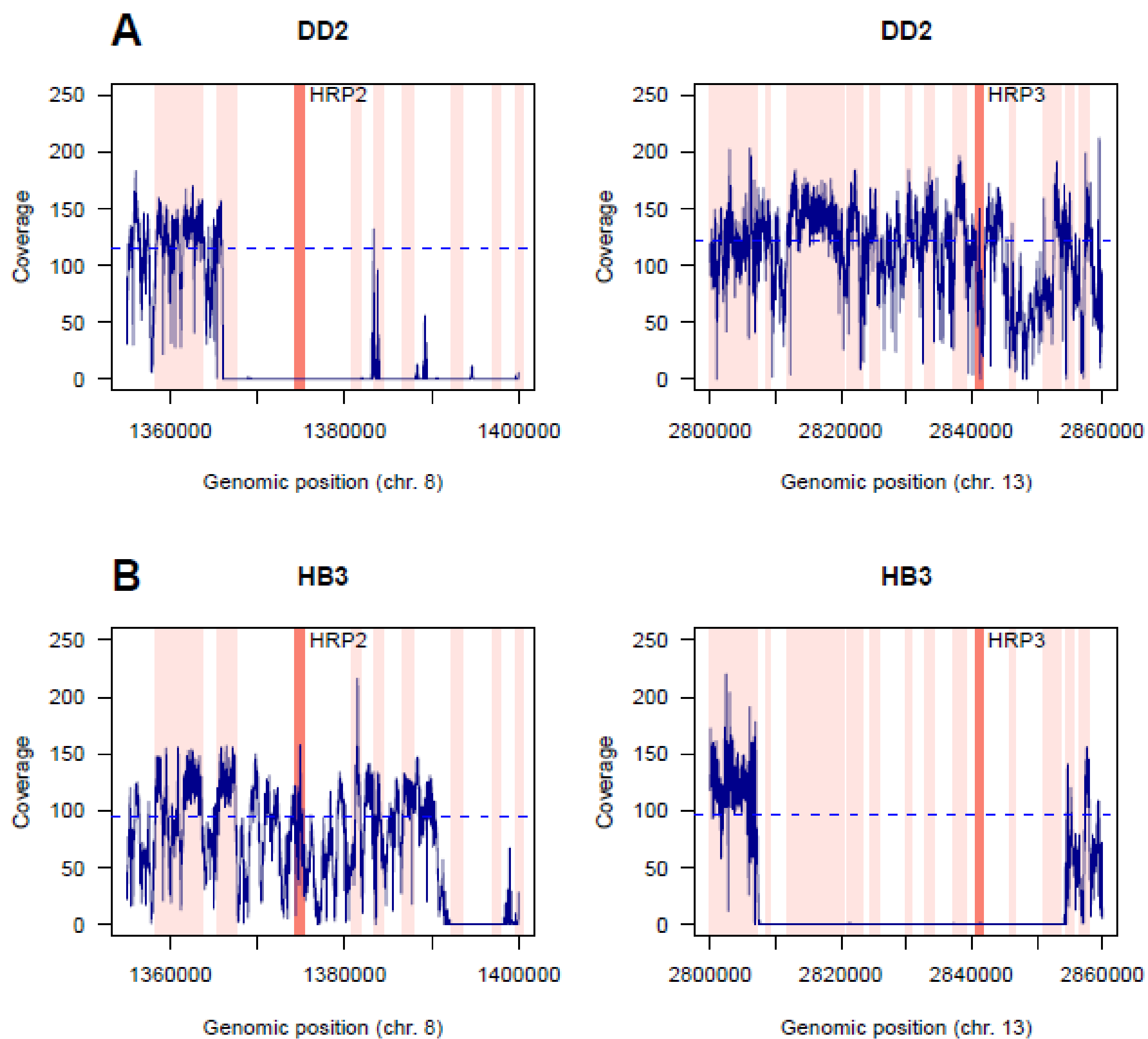

Fig. 1

ACCEPTED MANUSCRIPT

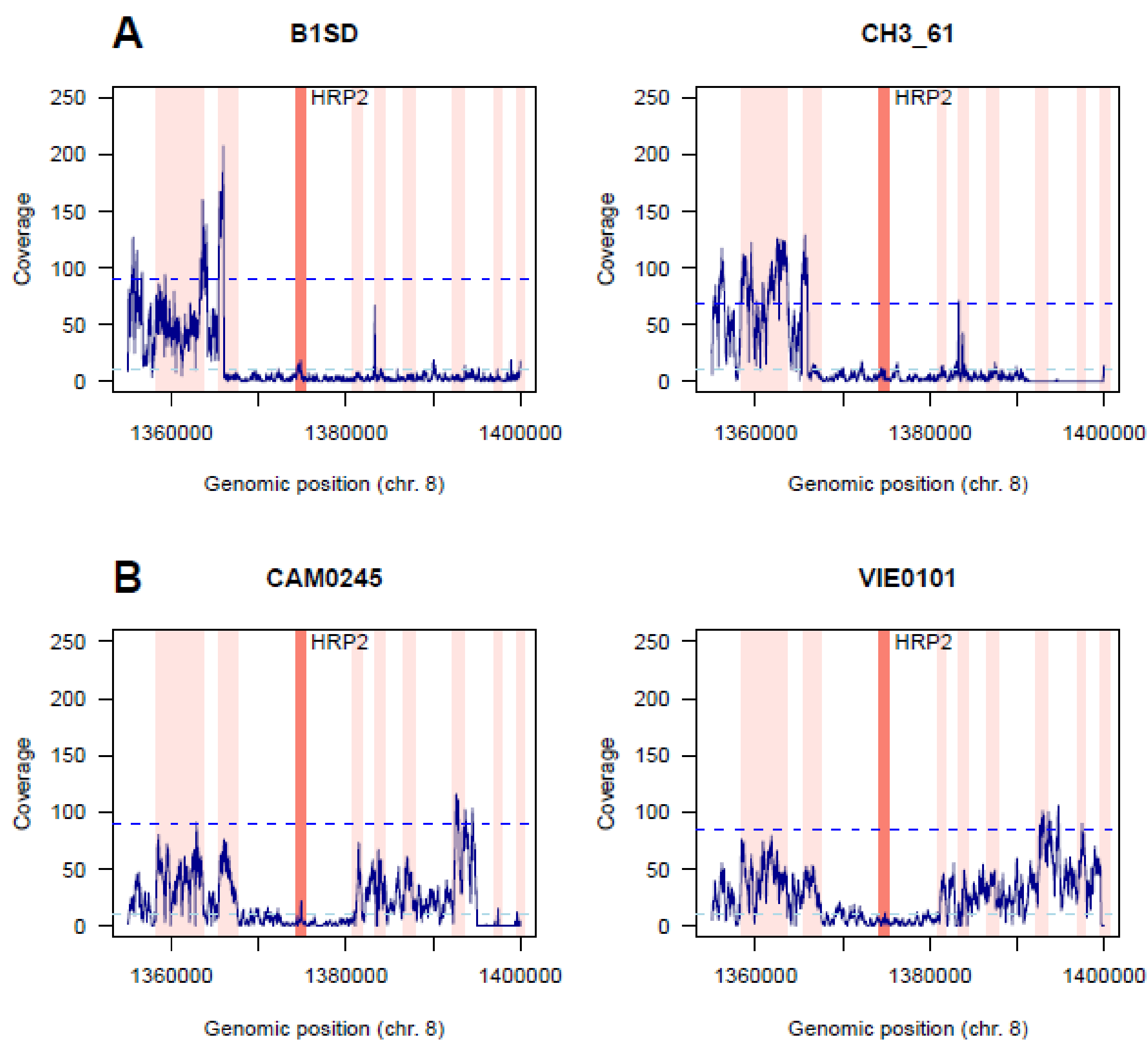

Fig. 2

ACCEP

ACCEPTED MANUSCRIPT

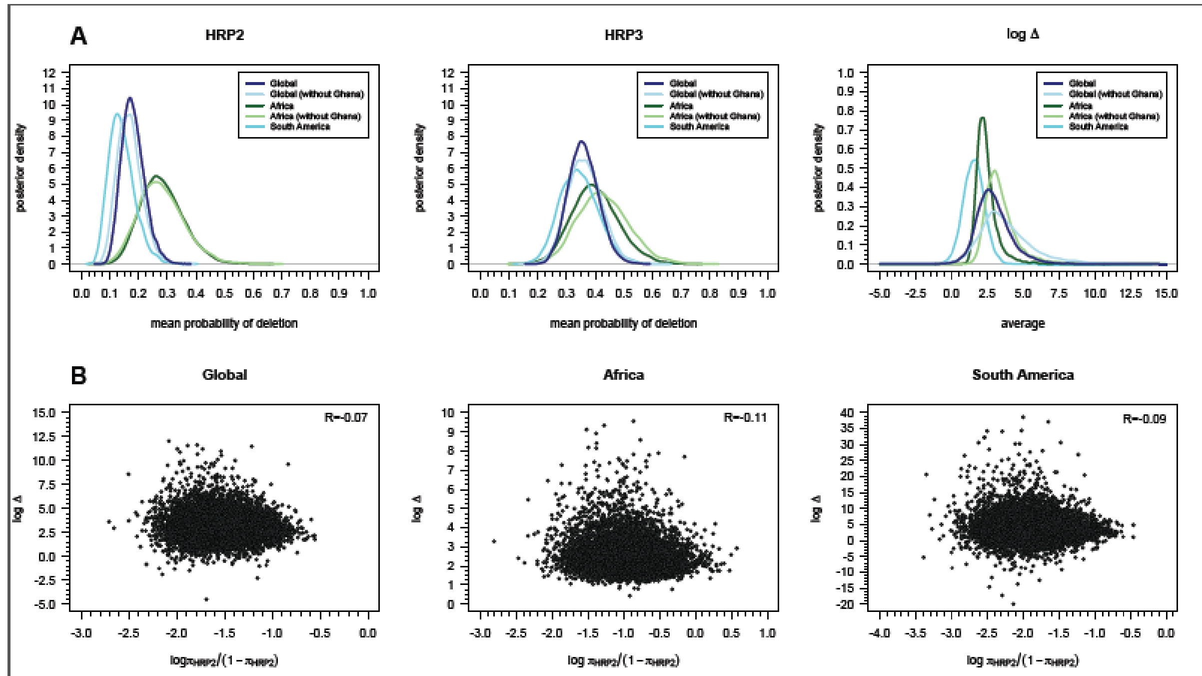

Fig. 3

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Highlights

- Deletions of *pfhrp2* and *pfhrp3* genes segregate independently of each other in the lab.
- Genomic data from experimental crosses identified gene deletions masked in polyclonal infections.
- Both *pfhrp2* and *pfhrp3* gene deletions are in low frequency worldwide.
- Deletions of *pfhrp3* gene are widespread in Southeast Asia.
- These gene deletions are under selection globally.

ACCEPTED MANUSCRIPT