Interchromosomal segmental duplication drives translocation and loss of *P. falciparum* histidine-rich protein 3

Nicholas J. Hathaway[1]*, Isaac E. Kim, Jr.[2,3]*, Neeva Wernsman Young[4], Sin Ting Hui[5], Rebecca Crudale[5], Emily Y. Liang[5], Christian P. Nixon[5], David Giesbrecht[5], Jonathan J. Juliano[6,7,8], Jonathan B. Parr[7,8], Jeffrey A. Bailey[2,3,5]

*Nicholas J. Hathaway and Isaac E. Kim, Jr. have equal contributions and are co-first authors

[1]Department of Medicine, University of Massachusetts Chan Medical School, Worcester, MA, USA
[2]Center for Computational Molecular Biology, Brown University, Providence, RI, USA
[3]Warren Alpert Medical School,, Brown University, Providence, RI, USA
[4]Department of Molecular Pharmacology, Physiology and Biotechnology, Brown University Providence, RI, USA
[5]Department of Pathology and Laboratory Medicine, Brown University, Providence, RI
[6]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, USA
[7]Division of Infectious Diseases, Department of Medicine, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[8]Curriculum in Genetics and Molecular Biology, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Author Contact Information:
Nicholas J. Hathaway: nicholas.hathaway@umassmed.edu
Isaac E. Kim, Jr.: isaac_kim@brown.edu
Neeva Wernsman Young: neeva_wernsman_young@brown.edu
Sin Ting Hui: sin_ting_hui@alumni.brown.edu
Rebecca Crudale: rebecca_crudale@brown.edu
David Giesbrecht: david_giesbrecht@brown.edu
Emily Y. Liang: emily_liang@brown.edu
Christian P. Nixon: christian_nixon@brown.edu
Jonathan J. Juliano: jonathan_juliano@med.unc.edu
Jonathan B. Parr: jonathan_parr@med.unc.edu
Jeffrey A. Bailey: jeffrey_bailey@brown.edu

Address all correspondence to:
Jeffrey A. Bailey, MD, PhD
Mencoff Family Associate Professor of Translational Research, Associate Professor of Pathology and Laboratory Medicine
Box G-E5, Providence, RI 02912, USA
Tel: 401-444-5160
Fax: 401-444-4377
Email: jeffrey_bailey@brown.edu

Preprint Server: Manuscript was deposited as a preprint to BioRxiv and is available at https://doi.org/10.1101/2022.12.07.519189. Licensing information states that the authors of this

manuscript are the copyright holders and have granted bioRxiv a license to display the preprint under a CC-BY-NC-ND 4.0 International license.

Keywords: non-allelic homologous recombination, segmental duplication, *Plasmodium falciparum*, rapid diagnostic tests (RDTs), Nanopore sequencing

ABSTRACT

Most malaria rapid diagnostic tests (RDTs) detect histidine-rich protein 2 (PfHRP2) and PfHRP3. *Pfhrp2* and *pfhrp3* gene deletions make parasites undetectable by RDTs. However, deletions have emerged where RDT use is uncommon and can occur independently of each other with *pfhrp3* deletion more frequent in many parasite populations. To improve our understanding of geographic patterns and mechanisms of *pfhrp3* deletion, we analyze 8,765 publicly available, whole-genome-sequenced *P. falciparum* field samples. Telomere healing is the only *pfhrp2* deletion mechanism observed among the few available *pfhrp2*-deleted genomes. Three *pfhrp3* deletion patterns predominate, all having chromosome 13 loss centromeric to *pfhrp3* extending to the chromosome end: 1) chromosome 13 deletion at varied locations centromeric to *pfhrp3* with signs of telomere healing (Asia; Pattern 13^{-}); 2) congruent duplication of a chromosome 5 segment which includes multidrug-resistant 1 (*pfmdr1*) gene (Asia; Pattern $13^{-}5^{++}$); and 3) congruent duplication of chromosome 11's subtelomeric region (Americas/Africa; Pattern $13^{-}11^{++}$). The most frequently observed pattern $13^{-}11^{++}$'s centromeric boundary falls within the largest (15kb) known segmental duplication in the core genome. Long-read sequencing of *pfhrp3*-deleted isolates from Honduras (HB3) and Sudan (SD01) confirms 13-11 hybrid chromosomes. Our findings demonstrate that *pfhrp3* deletion in the Americas and Africa often results from duplication-mediated non-allelic homologous recombination, likely a common occurrence in meiosis, while in Asia, deletion of *pfhrp3* may be

related to benefits of duplicating *pfmdr1*. Furthermore, this suggests that specific conditions including low transmission are needed for emergence of these deletions, a finding that can inform refined surveillance strategies.

INTRODUCTION

Malaria caused by *Plasmodium falciparum* remains one of the most common causes of childhood mortality in Africa despite significant efforts to eradicate the disease (WHO). The latest report by the World Health Organization estimated 247 million cases of malaria and 619,000 fatalities in 2021 alone with the vast majority of deaths occurring in Africa (WHO 2022). Control activities focus on control of the mosquito vector using insecticide-treated bednets and indoor residual spraying, point-of-care diagnosis using rapid diagnostic tests (RDTs), and treatment with effective antimalarial drugs (Benelli and Beier 2017). Adding to the problems of insecticide and drug resistance, the emergence of “diagnostic-resistant” *P. falciparum* strains with deletion mutations in their subtelomeric regions represents a new threat to malaria control and elimination efforts (WHO 2021).

The mainstay of malaria diagnosis across Africa is no longer microscopy but RDTs due to their simplicity and speed. The rapid adoption of RDTs coupled with artemisinin-combination therapies has led to significant progress in the fight against malaria (Chiodini 2014). The predominant falciparum malaria RDTs detect *P. falciparum* histidine-rich protein 2 (PfHRP2) and are cross-reactive with its paralogue PfHRP3. PfHRP3 remains clinically relevant because RDTs can be positive for *pfhrp2*-deleted parasites if *pfhrp3* remains intact (Kong et al. 2021). The number of PfHRP2-based RDTs increased exponentially from less than 1 million sold in 2004 to more than 300 million in 2013 (Poti et al. 2020). A growing number of studies, however, have reported laboratory and field isolates with deletions of *pfhrp2* and *pfhrp3* in the subtelomeric regions of chromosomes 8 and 13, respectively, which allow the parasite to escape diagnosis by

widely used PfHRP2-based RDTs (Baker et al. 2005; Organization and Others 2011; Cheng et al. 2014; Poti et al. 2020; Thomson et al. 2020). These parasites appear to be spreading rapidly in some regions and have compromised existing test-and-treat programs, especially in the Horn of Africa (Berhane et al. 2018; Verma et al. 2018; Feleke et al. 2021; WHO 2021).

The prevalence of parasites with *pfhrp2* and *pfhrp3* deletion varies markedly across continents in a manner not explained by RDT use alone. Parasites with these deletions are well-established in areas where PfHRP2-based RDTs have never been used routinely such as parts of South America (Thomson et al. 2020). Studies in Ethiopia, where false-negative RDTs owing to *pfhrp2 and pfhrp3* deletions are common, suggest that the *pfhrp3* deletion mutation occurred before that of *pfhrp2* given it is more frequent and shows a shorter shared haplotype (Feleke et al. 2021). The reason why *pfhrp3* deletion occurred first remains unclear. A 1994 study of the HB3 laboratory strain reports a *pfhrp3* deletion thought to be caused by a subtelomeric pairing and exchange between chromosomes 11 and 13 in meiosis (Hinterberg et al. 1994). Further examination of recombination as a possible mechanism for *pfhrp3* deletion in contemporary, circulating parasites is needed.

Studies of *P. falciparum* have been limited by the difficulty of assembling the repetitive and paralogous sequences of its subtelomeres using short-read (e.g. Illumina) sequencing. In *P. falciparum*, low complexity AT-rich intergenic sequence is challenging and within these sub-telomeric regions, repeats and highly similar gene families can be found in high density. Subtelomeric regions represent roughly 5% of the genome, are unstable, and contain rapidly diversifying gene families (e.g., *var*, *rifin*, *stevor*) that undergo frequent conversion between chromosomes mediated by a system of double-stranded breakage (DSB), telomere healing, and non-allelic homologous recombination (NAHR) (Vernick and McCutchan 1988; Mattei and Scherf 1994; Gardner et al. 2002; Lee et al. 2014; Calhoun et al. 2017; Zhang et al. 2019; Reed et al. 2021). Subtelomeric/centromeric exchange importantly allows for unbalanced progeny without large-scale ramifications of altering a larger proportion of a chromosome. Once

duplications are present, there can be a predisposition to further duplications or other rearrangements through NAHR between highly identical paralogous regions, which can potentiate the rapid expansion of gene families and their spread across subtelomeric regions. (Korbel et al. 2007; Conrad et al. 2010; Kidd et al. 2010; Mills et al. 2011; Consortium and The 1000 Genomes Project Consortium 2012; Didelot et al. 2012; Parks et al. 2015). These duplication-rich regions can rapidly gain or lose sequence and lead to differences in phenotype, including a possible advantage of a *P. falciparum* adhesion receptor in pregnancy-associated malaria and increased diversity of major surface antigens on *P. falciparum*-infected red blood cells (Bailey and Eichler 2006; Sander et al. 2009; Liu et al. 2011; Zhang et al. 2019). Notably, *pfhrp2* and *pfhrp3* are adjacent to but not considered part of the subtelomeric regions because recombination of *var* genes does not necessarily result in the deletion of *pfhrp2* and *pfhrp3* (Otto, Böhme, et al. 2018; Feleke et al. 2021).

Telomere healing is another common mechanism of subtelomeric deletion in *P. falciparum* that involves chromosomal breakage followed by de novo telomere addition via telomerase activity that serves to stabilize the end of the chromosome with complete loss of the downstream coding regions from the breakage. Deletion of the *P. falciparum* knob-associated histidine-rich protein *(KAHRP* or *pfhrp1)* and *pfhrp2* genes via this mechanism was first reported to occur in laboratory isolates (Pologe and Ravetch 1988). Since then, studies have further elucidated telomere healing, including the critical role of telomerase in *P. falciparum* and several occurrences affecting a number of genes including *pfhrp1, Pf332,* and *Pf87* in laboratory isolates, (Scherf and Mattei 1992; Mattei and Scherf 1994; Bottius et al. 1998). Scherf and Mattei observed that this mechanism of deletion for *pfhrp1* and *pfhrp2* only occurred in laboratory isolates but not in clinical isolates, suggesting it may occasionally occur in mitosis but is corrected for *in vivo* due to the genes' essential role (Scherf and Mattei 1992).

Similarly, the duplicative transposition of a subtelomeric region of one chromosome onto another chromosome frequently occurs in *P. falciparum*. Specifically, prior studies have found

that duplicative transposition results in the formation of a chimeric variant antigen encoding gene, duplications of *var2csa*, alternating expression of *P. falciparum*-erythrocyte membrane protein 1 (*PfEMP1*), mutations of cytochrome b, amplification of *P. falciparum* multidrug resistance-associated protein *(mrp1),* and *var* sequence diversity (Sander et al. 2009; Bopp et al. 2013; Claessens et al. 2014; Sander et al. 2014; Zhang et al. 2019).

Meanwhile, the other 95% of the *P. falciparum* genome (termed the core genome) is highly conserved with strong synteny between species (Bopp et al. 2013; Miles et al. 2016; Bruske et al. 2018; Otto, Böhme, et al. 2018). Duplication of genes within the core genome can also benefit the parasite such as when exposed to antimalarial drugs - examples include duplication of multidrug resistance protein 1 (PF3D7_0523000 *pfmdr1*) leading to mefloquine resistance (Price et al. 2004) and duplication of plasmepsin 2 and 3 (PF3D7_148000 pm2, PF3D7_148100 pm3) that correlate with piperaquine resistance (Amato et al. 2017; Witkowski et al. 2017). The primary mechanism proposed for these duplications is large segmental tandem duplication mediated by monomeric As/Ts (Nair et al. 2007; Isabel Veiga et al. 2016; Huckaby et al. 2019). It is known that parasites are unable to repair DNA double-strand breaks via traditional canonical non-homologous end joining due to the lack of appropriate protein machinery and instead rely more on homologous recombination (Barry et al. 2007; Sander et al. 2009; Rask et al. 2010; Chen et al. 2011; Kirkman et al. 2014; Lee et al. 2014; Tessema et al. 2015; Day et al. 2017). However, the exact mechanisms of exchanges in the subtelomeric regions involving the loss of *pfhrp2* and *pfhrp3* have not been studied in depth.

An improved understanding of the mechanism of these deletions can provide new insights into how frequently they occur and the evolutionary pressures driving their emergence and help inform strategies to address them. Though available whole-genome sequences include few *pfhrp2* deletions, we find that parasites with *pfhrp3* deletion available from multiple countries have geographically distinct deletion mechanisms. Therefore, we seek to elucidate the mechanism responsible for *pfhrp3* deletion using publicly available genomic sequencing data

complemented by confirmatory Oxford Nanopore Technologies (Nanopore) sequencing. Our findings shed light on geographical differences in *pfhrp3* deletion patterns, a major underlying mechanism, and how they likely emerged, providing key information for improved surveillance.

RESULTS

Pfhrp2 and pfhrp3 deletions in the global P. falciparum genomic dataset

We examine all publicly available Illumina whole-genome sequencing (WGS) data from *P. falciparum* isolates from around the world as of July 2022, comprising 8,765 field samples (**Table S1**). We analyze read depth across the genomic regions containing *pfhrp2* and *pfhrp3* to detect nucleotide and copy number variations (e.g. deletions and duplications) (**Table S2, Table S3, Table S4)**. We identify 23 isolates with *pfhrp2* deletion and 153 with *pfhrp3* deletion; the majority are from the Americas and Africa although sampling is not globally representative of their prevalence (**Supplemental Figure 1**, **Supplemental Figure 2**). Parasites with *pfhrp3* deletion outnumber those with *pfhrp2* deletion in all three geographic regions, particularly in the Americas and the Horn of Africa.

Pfhrp2 deletions

Within the samples analyzed, there are 25 total samples with genomic coverage evidence of *pfhrp2* deletion including 2 lab strains and 23 field samples: 10 in East Africa, 1 in Papua New Guinea (isolate D10), 12 in South America, and 2 in South East Asia (1 being isolate DD2) (**Supplemental Figure 1**). Of these samples, 12 show evidence of breakage and telomeric healing as evidenced by the presence of telomere-associated tandem repeat 1 (TARE-1) (Vernick and McCutchan 1988) contiguous with the genomic sequence at locations where coverage drops to zero (**Supplemental Figure 25**). These breaks are detected at various locations with 8 of the South American isolates and the D10 isolate falling within *pfhrp2* itself,

isolate DD2 at 1,365,960 within heat-shock protein 70, 1 Brazilian isolate at 1,370,074, and 1 Cambodian isolate at 1,373,785 (**Supplemental Figure 25**). The other samples do not have detectable TARE1 or evidence of genomic rearrangement but are also amplified with sWGA, limiting the ability to detect the TARE1 sequence.

Chromosome 11 deletions

There are 33 isolates with evidence of sub-telomeric chromosome 11 deletions. There is evidence of telomere healing on 31 out of these 33 isolates with the presence of TARE1 sequence contiguous with chromosome 11 sequence at the genomic location where coverage drops to zero. The leftmost position of these breaks falls within Pf332 at 1,952,807 with 14 total isolates with telomere healing of Pf332. Only one parasite (lab isolate FCR3) has deletion up and through Pf332 to the duplicated region with subsequent duplication of chromosome 13 that would be consistent with the reciprocal of $13^{-}11^{++}$. No field samples have this pattern. The related clone of FCR3, IT, does not contain this pattern and indicates that FCR3 duplicates this segment of chromosome 13 via translocation within culture and not in the field.

Three distinct pfhrp3 deletion patterns with geographical associations

Analysis of the publicly available WGS data reveals three distinct *pfhrp3* deletion patterns based on read depth analysis: 1) deletion of chromosome 13 starting at various locations centromeric to *pfhrp3* and unassociated with other copy number changes (**pattern 13^{-}**); 2) deletion of chromosome 13 from position 2,835,587 to the end of the chromosome and congruent duplication of chromosome 5 from position 952,668 to 979,203 which includes *pfmdr1* (**pattern $13^{-}5^{++}$**); and 3) loss of a portion of chromosome 13 (starting just centromeric to *pfhrp3* and extending to the end of the chromosome) and the gain of a copy of the chromosome 11 subtelomeric region – suggestive of a translocation replacing the end of chromosome 13 with chromosome 11 (**pattern $13^{-}11^{++}$**) (**Figure 1**). Among the 153 field samples with *pfhrp3*

deletion, 19 (12.4%) have pattern 13^{-}, 27 (17.6%) have pattern $13^{-}5^{++}$, and 107 (69.9%) demonstrate pattern $13^{-}11^{++}$. All deletions encompass *pfhrp3* at 2,840,236 to 2,842,840. Pattern 13^{-} shows deletions of the core genome on average of 19kb (ranging from 13kb to 31kb) encompassing the *pfhrp3* gene. Of these 13^{-} deletions, 17 out of 19 have a TARE1 sequence adjacent to the breakpoint consistent with telomere healing (Vernick and McCutchan 1988). Relative to the reference genome, pattern $13^{-}5^{++}$ is missing 17,895 base pairs from 2,835,587 to 2,853,482, with a gain of 26,535 bases of chromosome 5 including *pfmdr1*. Pattern $13^{-}11^{++}$ is missing 46,323 base pairs from chromosome 13's core genome (from 2,807,159 to 2,853,482) with a gain of 70,175 bases of chromosome 11's core genome (from 1,933,138 to 2,003,313). Based on genomes that are assembled to the end of their telomeres (Otto, Böhme, et al. 2018), an additional 17-84kb is deleted from the paralogous sub-telomeric region on chromosome 13, and an additional 15-87kb of the paralogous sub-telomeric region on chromosome 11 is duplicated. Pattern $13^{-}11^{++}$ is almost exclusively found in isolates from Africa and the Americas, while $13^{-}5^{++}$ is exclusively in Asia, and 13^{-} telomere healing is observed predominantly in Southeast Asian isolates (n=15), with only 2 from Africa (Kenya and Ghana) (**Figure 1**).

Pattern 13^{-} deletion pattern and telomere healing

The third pattern, 13^{-}, unassociated with other gains or losses, is found in 17 isolates. Analysis of the breakpoint sequences shows a transition to the TARE1 sequence. Deletions are at the following locations on chromosome 13: 2,837,093 (n=5), 2,830,893 (n=4), 2,837,343 (n=2), 2,822,443 (n=1), 2,823,593 (n=1), 2,830,793 (n=1), 2,834,543 (n=1), 2,835,493 (n=1), 2,840,993 (n=1) (**Supplemental Figure 3**).

Pattern $13^{-}5^{++}$: pfhrp3 deletion and pfmdr1 duplication

Pattern $13^{-}5^{++}$ is seen in 27 samples. These isolates have evidence of a genomic rearrangement that involves a 26bp AT di-nucleotide repeat at 2,835,587 on chromosome 13

and a 20bp AT di-nucleotide repeat at 979,203 on chromosome 5, which is supported by paired-end reads with discordant mapping with one read mapping to chromosome 13 and the other mapping to chromosome 5. When reads are extracted from these regions and assembled, a contig traverses from chromosome 13 to chromosome 5 unique sequence (from 2,835,587 on chromosome 13 and then into the reverse strand on chromosome 5 at 979,203). This incorporates the duplication of a 26kb region of chromosome 5 supported by a two-fold increase in coverage from 979,203 to 952,668 in the reverse, where it truncates in the TARE1 sequence, again consistent with telomere healing. This TARE1 transition occurs within the gene PF3D7_0522900 (a zinc finger gene). This 25kb duplication contains the following genes: PF3D7_0523600, PF3D7_0523500, PF3D7_0523400, PF3D7_0523300, PF3D7_0523200, PF3D7_0523100, PF3D7_0523000 (*pfmdr1*) (**Supplemental Figure 4**, **Supplemental Figure 5**). Thus, this rearrangement involves the deletion of *pfhrp3* and duplication of *pfmdr1*.

In addition to the transposed *pfmdr1* duplication, three $13^{-}5^{++}$ isolates also have evidence of tandem duplication of the core *pfmdr1* region based on discordant read pairs and increased coverage to 3x of the rest of the genome, indicating that these isolates have three copies of the *pfmdr1* gene, two in tandem on chromosome 5, and one on the transposed segment on chromosome 13. The samples have breakpoints in A-rich stretches with two samples from Cambodia sharing the same breakpoints, Ax22 at chromosome 5 946,684 and Ax36 at 964,504 for both samples from Cambodia; and Ax19 at 947,967 and Ax18 at 970,043 for the one sample from Laos. Such tandem duplication breakpoints in monomeric stretches are consistent with previous findings for *pfmdr1* (Nair et al. 2007).

All isolates with evidence for $13^{-}5^{++}$ have only wild-type *pfmdr1* or are a mix of wild-type and 184F. The three isolates with three copies of *pfmdr1* have 2x coverage of 184F and 1x coverage of wild-type *pfmdr1*, which is consistent with the duplicated *pfmdr1* on chromosome 13 being wild-type *pfmdr1*. All isolates with this deletion pattern are found within Asia.

Pattern 13^{-}11^{++} breakpoint resides in a segmental duplication containing ribosomal genes on chromosomes 11 and 13

Pattern 13^{-}11^{++} is found predominantly among parasites from the Americas and Africa with a centromeric breakpoint appearing to fall within an interchromosomal segmental duplication. In an all-by-all unique k-mer comparison of the chromosomes of the 3D7 genome using nucmer (Kurtz et al. 2004), this segmental duplication is the largest inter-chromosomal duplication in the core genome and is second only to a 23.5 kb region between subtelomeric regions on the 5' ends of chromosomes 5 and 13 that encompasses a singular *var* gene and the start of the telomere itself **(Supplemental Figure 6)**. We further investigate the nature of the segmental duplication. The two copies on chromosome 11 (1,918,028-1,933,288) and chromosome 13 (2,792,021-2,807,295) are nearly identical (99.0% similarity) (**Figure 2**) and oriented similarly. Each copy contains a centromeric 7kb region encoding 2-3 undefined product proteins (98.9% identity) and a telomeric 8kb region containing one of the two S-type (Gardner et al. 2002) ribosomal loci (99.7% identity) (**Figure 2**, **Supplemental Figure 7**, **Supplemental Figure 8**). Pairwise alignment of the duplicated segments of chromosomes 11 and 13 from non-deleted strains shows more than 99.0% similarity and no distinct separation of the paralogs **(Figure 2**). Together, this pattern suggests that this segmental duplication is likely undergoing frequent exchanges and maintaining high identity through gene conversion and/or translocations through non-allelic homologous recombination.

Previous PacBio assemblies do not fully resolve chromosome 11 and 13 subtelomeres

Given that the pattern of deletion and duplication and the breakpoints are suggestive of duplication-mediated recombination leading to translocation, we seek to examine high-quality PacBio genome assemblies of other strains containing the *pfhrp3* deletion. However, the gene annotations by Companion (Steinbiss et al. 2016) of chromosome 11 of previous PacBio assemblies (Otto, Böhme, et al. 2018) show that the assemblies of strains SD01 and HB3 are

unable to completely assemble the sub-telomeric regions on chromosomes 11 and 13. For both assemblies, the sub-telomeric region commonly found on chromosome 11 is found only on chromosome 13 (**Supplemental Figure 9 and Supplemental Figure 10**) with the chromosome 11 sub-telomeric region missing from segmental duplication onwards. The absence of an assembled sub-telomeric region on chromosome 11 prevents detailed analysis of the mechanism behind the deletion of *pfhrp3* and is likely a result of the inability of the assembler and/or underlying PacBio reads to unambiguously traverse the segmental duplication and separate the duplicated chromosome 11 subtelomeric region sequence into two copies.

Combined analysis of Nanopore and PacBio reads confirms a segmental duplicated region of the normal chromosome 11 and hybrid chromosome 13-11

To better examine and confirm that pattern $13^{-}11^{++}$ involves a hybrid chromosome 13-11 chromosomal product consistent with a breakpoint within the segmental duplication, we further examine two representative parasite isolates with known *pfhrp3* deletion and coverage suggestive of pattern $13^{-}11^{++}$: HB3 from Honduras and SD01 from Sudan. We use Oxford Nanopore Technologies' MinION to sequence 7.35 gigabases of HB3 and 645.93 megabases of SD01 for an average coverage of 319x and 29.3x, respectively. We combine our Nanopore data with the publicly available PacBio sequencing data and test for the presence of hybrid chromosomes using a two-pronged approach: 1) mapping raw long reads directly to normal and hybrid constructs and 2) more optimized *de-novo* assembly of the higher quality Nanopore long reads into contigs.

For direct read mapping, we construct 3D7-based representations of hybrid chromosomes 13-11 and 11-13 by joining the 3D7 chromosomal sequences at breakpoints in the middle of the segmental duplication (Methods). We then map all of the strains with PacBio sequencing along with our additional Nanopore reads for HB3 and SD01 to the normal and hybrid constructs to detect reads that completely span the duplicated region (**Figure 4**). Reads

are considered “spanning” if they extend at least 50 bp into the flanking unique regions. HB3 has 77 spanning reads across normal chromosome 11 and 91 spanning reads across hybrid chromosome 13-11, hereinafter referred to as 13-11 spanning reads. SD01 has two chromosome 11 spanning reads and one 13-11 chromosome spanning read. Neither strain has PacBio or Nanopore spanning reads across the normal chromosome 13 and hybrid 11-13 (**Figure 4**), which would have been consistent with the reciprocal meiotic product. Due to a lack of sufficient sample, SD01 has a small number of reads that span the full region. Sub-analysis reveals 4 regions within this duplicated region that have chromosome 11 and 13-specific nucleotide variation that is able to be used to further bridge across this region for additional confirmation (**Supplemental Figure 11** and **Supplemental Figure 12**). The other 14 strains with intact *pfhrp3* from the PacBio dataset (Otto, Böhme, et al. 2018) have reads that span chromosome 11 and chromosome 13 but lack any spanning reads across the hybrid 13-11 or 11-13 chromosomes. For example, the *pfhrp3*-intact strain CD01 has five and three spanning reads for normal chromosome 11 and 13, respectively, but no spanning reads for hybrid 11-13 and hybrid 13-11 (**Figure 4**). In summary, the presence of long reads for HB3 and SD01 spanning across normal chromosome 11 and hybrid chromosome 13-11 construct confirms a segmental duplicated region of these chromosomes.

Genomic Localization of breakpoint for $13^{-}11^{++}$

The analysis of the variation within the 15.2kb duplicated region between strains that contain a near-perfect chromosome 11 duplication does not reveal one distinctive breakpoint. Though the telomeric region of chromosome 11 is identical, the variation sites within the 15.2kb duplicated region reveal sites of variability that would represent the two different paralogs between the telomeric chromosome 11 copies, and these sites of variability can be seen up to near the end of the duplicated region (**Supplemental Figure 8**). This would indicate that there is either a shared breakpoint at the very end of the duplicated region or that there are various

breakpoints within the duplicated region that can then subsequently result in the NAHR resulting in the $13^{-}11^{++}$ pattern.

De novo long-read assemblies of pfhrp3-deleted strains further confirm hybridized 13-11 chromosome

To further ensure the presence of the hybridized 13-11 chromosome and not some more complicated structural alterations involving other regions of the genome, *de novo* whole-genome assemblies are created for the HB3 and SD01 lab strains from Nanopore long reads. HB3 assembly yields 16 contigs representing complete chromosomes. TARE-1 (Vernick and McCutchan 1988) is detected on the ends of all chromosomes except for the 3' end of chromosome 7 and the 5' end of chromosome 5, indicating that close to the entirety of the telomeric regions of HB3 is achieved, N50 1,595,798 and L50 5. SD01, however, likely due to its lower sequencing coverage, has a more disjointed assembly with 200 final contigs, N50 263,459, and L50 30. The HB3 assembly has a chromosome 11 that closely matches 3D7 chromosome 11 and a hybrid 13-11 that closely matches 3D7 chromosome 13 up to the ribosomal duplication region where it then subsequently best matches the 3D7 chromosome 11 (**Figure 3**, **Supplemental Figure 13**). Both of these chromosomes have TARE-1 (Vernick and McCutchan 1988) at their 5' and 3` end, indicating that these are full-length chromosome assemblies. The SD01 assembly has two contigs – one that matches 3D7 chromosome 11 and the other that matches 3D7 chromosome 13 up to the duplicated region and then subsequently matches 3D7 chromosome 11.

These new assemblies are further annotated for genes by Companion (Steinbiss et al. 2016). The contig matching the hybrid chromosome 13-chromosome 11 for both strains have the genes found on 3D7 chromosome 11 after the duplicated region. The duplicated genes on chromosomes 11 and 13-11 that share homology with 3D7 chromosome 11 subtelomere include *pf332* (PF3D7_1149000), two ring erythrocyte surface antigens genes (PF3D7_1149200,

PF3D7_1149500), three PHISTs genes, a FIKK family gene, and two hypothetical proteins and ends with a DnaJ gene (PF3D7_1149600) corresponding to 3D7 genes PF3D7_1148700 through PF3D7_1149600 (**Supplemental Figure 14 and Supplemental Figure 15**). Homology between HB3 chromosome 11 and chromosome 13-11 continues up through a *rifin*, then a *stevor* gene, and then the sequence completely diverges afterward with a different gene family organization structure but both consisting of *stevor*, *rifin,* and *var* gene families among other paralogous gene families **(Supplemental Figure 14)**. The chromosome 13-11 SD01 contig reaches a DNAJ protein (PF3D7_1149600) and terminates (**Supplemental Figure 15**), likely because the assembly is unable to contend with the near complete identical sequence between chromosome 11 and chromosome 13-11. Illumina sequences map to the *var* regions on SD01 chromosome 11 where chromosome 13-11 terminates, and local assemblies reveal no variation but double coverage as compared to the rest of the genome, which is consistent with this region being perfectly copied. Therefore, it's likely the SD01 clone has identical copies up and through the telomere of these two chromosomes.

Analysis of the 11 other PacBio assemblies (Otto, Böhme, et al. 2018) with normal chromosome 11 shows that homology between strains also ends at this DnaJ gene (PF3D7_1149600) with the genes immediately following being within the *stevor*, *rifin,* and *var* gene families among other paralogous gene families. The genes on chromosome 13 but deleted in the hybrid chromosome 13-11 correspond to 3D7 genes PF3D7_1371500 through PF3D7_1373500 including *pfhrp3* (PF3D7_1372200) and *EBL-1* (PF3D7_1371600). The *de-novo* long-read assemblies of HB3 and SD01 further confirm the presence of a normal chromosome 11 and hybrid chromosome 13-11 that share a duplicated region.

Ribosomal gene segmental duplication is also observed in closely related P. praefalciparum

To look at the conservation of the segmental duplication containing the ribosomal genes, we examine genomes of closely related *Plasmodium* parasites in the *Laverania* subgenus, which comprises *P. falciparum* and other primate *Plasmodium* spp. The *Plasmodium praefalciparum* genome*,* which is *P. falciparum's* closest relative having diverged about 50,000 years ago (Otto, Gilabert, et al. 2018), also contains the similar S-type rRNA loci on chromosomes 11 and 13 and has a similar gene synteny to *P. falciparum* in these regions as well as the region on chromosome 8 close to *pfhrp2* (**Supplemental Figure 16**, **Supplemental Figure 17**, **and Supplemental Figure 18**). *P. praefalciparum* contains the 15.2kb duplicated region on both chromosomes 11 and 13 and is 96.7% similar to the 3D7 duplicated region. Other *Laverania* genomes (Otto, Gilabert, et al. 2018) are not fully assembled within their subtelomeric regions.

Pattern $13^{-}11^{++}$ predominates in the Americas and Africa

Pattern $13^{-}11^{++}$ is observed in 80 American parasites, 38 African parasites (20 in Ethiopia, 6 in Senegal, 1 Mali, and 1 Ghana), and 2 Asian parasites (1 Cambodia, 1 Thailand) (**Figure 1**). Out of the 107 parasites with this deletion pattern, 89 parasites (77 of the 80 American, 11 of the 28 African parasites, and 1 of the 2 Asian) have near-identical copies of the chromosome 11 duplicated region. Near-identical copies are defined as having ≥99% identity across 382 variant sites within the duplicated region identical (same variant allele between copies) (**Table S4**). These 89 parasites containing identical copies do not all share the same haplotypes but rather show five major groups (**Figure 5)**. The remaining 21 parasites have variation within this region; on average, 10.2% of variant sites differ between the copies (min 83.8% identity). The five haplotype groups show separation by geography with distinct haplotypes observed in American and African strains. The haplotypes for the segment of

chromosome 11 found within $13^{-}11^{++}$ parasites can also be found within normal parasites lacking the translocation (**Supplemental Figure 19**, **Supplemental Figure 20**, **Supplemental Figure 21**, **Supplemental Figure 22**, **Supplemental Figure 23**, **and Supplemental Figure 24**). The multiple distinct haplotypes within the duplicated segment $13^{-}11^{++}$ parasites suggest multiple origins due to separate translocation events, likely originating from intrastrain (self) recombination followed by clonal propagation.

DISCUSSION

Here, we use publicly available short-read and newly generated long-read sequencing data to identify *pfhrp2* and *pfhrp3* deletions and their mechanisms in field *P. falciparum* samples collected across the globe. The limited number of *pfhrp2* deleted strains shows chromosome 8 breakpoints with evidence of telomere healing, a common repair mechanism in *P. falciparum* (Lee et al. 2014; Calhoun et al. 2017). In the current dataset, *pfhrp3* deletions occur through three different mechanisms. The least common mechanism is telomere healing (representing the 13^{-} pattern). The second most common deletion pattern is likely the result of NAHR incorporating a 26,535bp region of chromosome 5 containing *pfmdr1* onto 13, utilizing short block homology between 28bp di-nucleotide AT repeats on chromosome 5 and 13 and deleting *pfhrp3* while duplicating *pfmdr1* (representing the $13^{-}5^{++}$ pattern). There appears to be one origin of $13^{-}5^{++}$, which is only observed in the Asia population, and its continued presence is likely driven by the added benefit of *pfmdr1* duplication and the presence of mefloquine in Asia. The most common pattern $13^{-}11^{++}$ predominates in the Americas and Africa and is the result of NAHR between chromosome 11 and 13 at a large 15.2kb highly-identical homologous block, 8kb of which is an rRNA locus, resulting in duplication of 70,175bp of core chromosome 11 plus 15-87kb paralogous sub-telomeric region. While Hinterberg *et al.* proposes that a general mechanism of non-homologous recombination of the subtelomeric regions may be responsible

for the deletion of *pfhrp3* (Hinterberg et al. 1994), our analysis reveals the exact genomic regions involved. The predominant $13^{-}11^{++}$ pattern involves the segmental duplication and translocation of chromosome 11 at bases 1,917,992 to 2,003,604 onto chromosome 13 from base 2,792,047 until the end. Furthermore, we provide the first detailed analysis of this deletion mechanism in field samples, not just a consequence of culture adaptation of HB3. Importantly, there is evidence for multiple origins and localities for $13^{-}11^{++}$ due to intrastrain recombination, suggesting this may be a common meiotic event that normally is selected against except in certain populations.

Pattern $13^{-}5^{++}$ with *pfmdr1* duplication is only seen in South East Asia, which is also where tandem duplication of *pfmdr1* related to mefloquine resistance has been primarily observed (Isabel Veiga et al. 2016). We confirm this pattern based on short reads with evidence of discordant mates between chromosomes 13 and 5, the local assembly connecting these two interchromosomal regions, and the coverage observed within the proposed duplicated region on chromosome 5 with evidence of TARE1 on chromosome 5 contigs at locations of coverage change. Studies of the dynamics of *pfmdr1* duplication have shown that these duplications are unstable, with both increases and decreases in copy numbers frequently occurring (Samarakoon et al. 2011). Similar to *var* gene recombination events where double-stranded DNA is displaced (Zhang et al. 2019), which creates a highly recombinogenic piece of DNA, a de-amplification (decrease in copy number) event of *pfmdr1* can create a free DNA fragment containing *pfmdr1*. This recombinogenic fragment can then interact with chromosome 13, subsequently preserving the number of *pfmdr1* copies and deleting *pfhrp3*. After recombining with chromosome 13, there can then be a telomerase healing event (Calhoun et al. 2017), sealing the end of the fragment downstream from *pfmdr1* and stabilization of the 13-5 hybrid chromosome. Given the relative randomness of telomere healing observed previously (Calhoun et al. 2017) and the fact that all isolates with *pfmdr1* duplication have the same TARE1 transition point, it is highly likely that these isolates represent a clonal expansion potentially due to the

benefit of the extra *pfmdr1* copy on chromosome 13 and loss of *pfhrp3*. Given all isolates with evidence of this duplication either have only wild-type *pfmdr1* or are mixed, the 13-5 chromosome copy most likely has wild-type *pfmdr1*. In 20 out of 27 *pfmdr1* duplication cases with a mixed genotype of *pfmdr1* (**Supplemental Figure 5**), the core genome *pfmdr1* has the Y184F mutation with no other mutations detected within the *pfmdr1* gene. Isolates containing only Y184F in *pfmdr1* have been shown to be outgrown by wild-type *pfmdr1* (Duvalsaint et al. 2021), which would mean having the wild-type *pfmdr1* duplication on chromosome 13 might confer a survival advantage beyond just increased resistance to mefloquine.

To confirm the NAHR event between 11 and 13 leading to loss of *pfhrp3* observed in our analysis of short-read data, we long-read sequence *pfhrp3*-deleted isolates to generate reads spanning the 15kb duplicated region, showing support for a normal chromosome 11 and a hybrid 13-11 in both HB3 and SD01. These findings support an NAHR event between the two 15kb duplicated regions causing this interchromosomal exchange and leading to progeny with a hybrid 13-11 chromosome lacking *pfhrp3* and its surrounding genes from the 15kb duplicated region and onwards (**Figure 3** and **Figure 4**). This is consistent with the genomic coverage pattern we observe in publicly available data from 107 *pfhrp3*-deleted field samples (**Figure 1**). Such translocation patterns have been described and also confirmed by long-read sequencing, but have generally involved multigene families such as *var* genes within the subtelomeres (Calhoun et al. 2017; Zhang et al. 2019). The event described here represents a much larger section of loss/duplication of 70kb of the core genome in addition to the subtelomere.

We propose a mechanistic model in which homology misalignment and recombination between chromosomes 11 and 13 initially occurs in an oocyst from identical parasites, resulting in four potential progeny including one with normal chromosomes 11 and 13 and three with translocations **(Figure 6)**. This can account for the identical haplotypes observed in the two copies of the chromosome 11 segment. Based on the identical haplotypes observed in the majority of samples, the most direct and likely mechanism involves progeny with two copies of

chromosome 11 recombining with an unrelated strain to yield unrelated chromosome 11 haplotypes. This duplication-mediated NAHR event is likely to occur frequently during meiosis and can explain the frequent rearrangements seen between chromosomes 11 and 13 in the previous experimental cross of HB3 x DD2 (Hinterberg et al. 1994). Meiotic misalignment and subsequent NAHR is a common cause of high-frequency chromosomal rearrangements including human disease (eg. in humans, 22q11 deletion syndrome due to misalignment of duplicated blocks on chromosome 22 occurs in 1 in 4000 births) (Vergés et al. 2014). Thus, in parasite populations, the chromosome 11 and 13 translocation is likely a common occurrence during meiosis. This can explain why *pfhrp3*-deleted isolates are more common in many populations (Góes et al. 2020; Feleke et al. 2021; Vera-Arias et al. 2022) given we see lower levels of telomere healing in *pfhrp3*, the apparent main mechanism for *pfhrp2* deletions. However, more extensive sequencing of RDT-negative *P. falciparum* samples is needed to confirm that there are no other deletion mechanisms responsible for *pfhrp2* loss.

The lack of hybrid chromosome 13-11 worldwide suggests such events are normally quickly removed from the population due to fitness costs, an idea supported by recent *in vitro* competition studies in culture showing decreased fitness of *pfhrp2/3*-deleted parasites (Nair et al. 2022). This decreased fitness of parasites with *pfhrp2/3* deletions also argues against a mitotic origin as deletions arising after meiosis would have to compete against more numerous and more fit intact parasites. The fact that abundant *pfhrp3* deletions have only been observed in low-transmission areas where within-infection competition is rare is consistent with this hypothesis of within-infection competition suppressing emergence. In the setting of RDT use, existing *pfhrp3* deletions in such a low-transmission environment may provide a genetic background on which less frequent *pfhrp2* deletion events can lead to a fully RDT-negative parasite. This is supported by evidence that *pfhrp3* deletion appears to predate *pfhrp2* deletions in the Horn of Africa (Feleke et al. 2021).

The biological effects of *pfhrp2* and *pfhrp3* loss are complicated due to other genes lost and gained due to the extent of the rearrangements. For 13-11++*, EBL-1* on chromosome 13 which helps enable *P. falciparum* to invade erythrocytes, is missing on the hybrid 13-11 chromosome. The loss of parasite *EBL-1* may cause infected patients to have fewer symptoms and be less likely to seek treatment (Jaskiewicz et al. 2019; Feleke et al. 2021). Analysis of one of the closest relatives of *P. falciparum*, *P. reichenowi,* shows a lack of the EBL-1 protein, which suggests its presence might not be essential to parasite survival. However, *P. reichenowi* has all the genes within the duplicated chromosome 11, suggesting that these genes are more essential than the chromosome 13 genes. Thus, the loss of the *pfhrp3* and *EBL-1* genes may increase *P. falciparum*'s fitness (Feleke et al. 2021). Other affected genes include multi-gene families of *rifins, stevors,* DnaJ genes, and ring-infected erythrocyte surface antigen (RESA) genes but also the single copy *pf332* on chromosome 11, which encodes a large Duffy-binding-like domain containing a 27KDa protein (Hodder et al. 2009; Nilsson et al. 2011; Rug et al. 2014). Increased copies could be beneficial, as *pf332* has been found to be essential for the binding of the Maurer cleft to the erythrocyte skeleton and is highly expressed in patients with cerebral malaria (Bertin et al. 2016). Conversely, lack of this protein is likely detrimental to survival and may be the reason the reciprocal hybrid 11-13 is not observed in field isolates. Only lab isolate FCR3 has any indication from coverage data that it has a duplicated chromosome 13 and a deleted chromosome 11. Given that the majority of the publicly available field samples are collected from studies using RDT-positive samples and that PfHRP3 encoded by duplicated *pfhrp3* will likely have been detected by RDT, sampling should not be biased against detecting the reciprocal hybrid 11-13 (which would have two copies of *pfhrp3*). The lack of any field 11-13 rearrangements suggests that the selective disadvantage of the lost and gained genes is strong enough to prevent its emergence in the natural parasite population

While further studies are needed to determine the reasons for these geographical patterns of *pfhrp2/3* deletions, our results provide an improved understanding of the mechanism

of structural variation underlying *pfhrp3* deletion. They also suggest general constraints against emergence in high-transmission regions due to within-host competition and specific requirements for emergence in low-transmission settings. If selective constraints of *pfhrp2* and *pfhrp3* deletions are similar, the high frequency of the NAHR-mediated loss and the additional drug pressure from duplication of *pfmdr1* may explain why *pfhrp3* loss precedes *pfhrp2* loss despite RDT pressure presumably exerting stronger survival advantage with loss of *pfhrp2* vs *pfhrp3*. However, given we still have a limited understanding of their biological roles, there may be situations where selective forces may favor the loss of *pfhrp2* relative to *pfhrp3*. Overall, our findings are clinically important, because the continued loss of these genes without timely intervention may result in a rapid decrease in the sensitivity of HRP2-based RDTs. Future studies focused on *pfhrp2/3* deletions that employ representative sampling are needed to determine the prevalence, interactions, and impacts of *pfhrp2* and *pfhrp3* deletions, as well as to examine the selective pressures and complex biology underlying them.

MATERIALS AND METHODS

Genomic locations and read-depth analysis

In order to study genomic deletions in the non-conserved and frequent recombinant subtelomeric regions surrounding *pfhrp2* and *pfhrp3*, conserved regions surrounding these genes were determined. Conserved genomic regions were determined by selecting regions that were present between all strains without *pfhrp2/3* deletions and lacked paralogy within 3D7. This was accomplished by first marking the 3D7 genome with the program tandem repeat finder (Benson 1999), then taking 200bp windows, stepping every 100bp between these tandem repeats, and using LASTZ (Harris 2007) to align these regions against currently available chromosomal-level PacBio assembled genomes (Otto, Böhme, et al. 2018) that did not contain known deletions of *pfhrp2/3*. Regions were kept if they hit (matched/aligned >70% identity) each

genome only once, and overlapping regions were then merged. Regions from within the duplicated region on chromosome 11 and chromosome 13 were kept if they hit either chromosome 11 or 13 but not if they hit other chromosomes. This served to eliminate paralogous regions outside of 11 and 13, particularly within the rRNA loci that were similar to the other S-type partial rRNA loci on chromosome 1. Local haplotype reconstruction was performed on each region for 8,765 publicly available whole genome sequences (WGS) of *P. falciparum* field samples including 24 samples from a recent study in Ethiopia (Feleke et al. 2021) where *pfhrp2-/3-* parasites are common. The called haplotypes were compared to determine which subregions contained variation in order to genotype each chromosome (**Supplemental Figure 26**). Coverage was determined for each genomic region by dividing the read depth of a region by the median read depth across the whole genome within each sample. Windows were examined on chromosomes 5, 8, 11, and 13 starting from chromosome 5 944,389, chromosome 8 1,290,239, chromosome 11 1,897,151, and chromosome 13 2,769,916 to their respective telomeres. All coordinates in the manuscript are zero-based positioning and are relative to *P. falciparum* 3D7 genome version 3 (version=2015-06-18) **(Supplemental Figure 26).**

Tandem repeat associated element 1 (TARE1) analysis and telomere healing determination

The 7bp pattern of TT[CT]AGGG(Vernick and McCutchan 1988) was used to determine the presence of TARE1, the presence of this pattern at least twice in tandem was required. To search for the presence of TARE1 within the short read Illumina WGS datasets, reads from the entire regions of interest were pulled down across chromosomes 5, 8, 11, and 13, and the above TARE1 pattern was searched for in each read. Regions that had TARE1 detected in their reads were then assembled to ensure the TARE1 sequence was contiguous with the genomic region from which the reads were pulled. The regions that had the presence of TARE1

contiguous with genomic regions were then compared to the coverage pattern within the area, and a sample was marked as having evidence of telomere healing if TARE1 was detected in the regions where coverage then dropped to 0, or down to the genomic coverage of the rest of the genome in the case of the chromosome 5 duplication.

Chromosome 5 Breakpoint Determination

The breakpoints of recombination on chromosome 5 were determined by looking for discordant read pairs, mate mapping to different chromosomal positions, around areas of interest, then assembling the discordant pairs, and mapping back to the assembled contig. Breakpoints were then determined by looking at the coordinates where a contig switches from one chromosomal region to the next.

Homologous Genomic Structure

To investigate the genomic landscape of recent segmental duplications across the genome and around the *pfhrp* genes, an all-by-all comparison of 3D7 reference genome was performed by first finding kmers of size 31 unique within each chromosome and then determining the locations of these kmers in the other chromosomes. If kmers were found in adjacent positions in both chromosomes, they were then merged into larger regions.

Comparisons within Laverania

To investigate the origins of this region shared between chromosomes 11 and 13, the six closest relatives of *Plasmodium falciparum* within the *Laverania* subgenus with available assembled genomes were examined (Otto, Gilabert, et al. 2018). The genomes of all *Laverania* have recently been sequenced and assembled using PacBio and Illumina data (Otto, Gilabert, et al. 2018). The assemblies were analyzed using their annotations and by using LASTZ (Harris

2007) with 80% identity and 90% coverage of the genes in the surrounding regions on chromosomes 8, 11, and 13.

Long-read sequences

All PacBio reads for strains with known or suspected *pfhrp3* deletions were obtained by SRA accession numbers from the National Center for Biotechnology Information (NCBI): HB3/Honduras (ERS712858) and SD01/Sudan (ERS746009) (Otto, Böhme, et al. 2018). To supplement these reads and to improve upon previous assemblies that were unable to fully assemble chromosomes 11 and 13, we further sequenced these strains using Oxford Nanopore Technologies' MinION device (Kasianowicz et al. 1996; Branton et al. 2008; Jain et al. 2016). The *P. falciparum* lab-isolate HB3/Honduras (MRA-155) was obtained from the National Institute of Allergy and Infectious Diseases' BEI Resources Repository, while the field strain SD01/Sudan was obtained from the Department of Cellular and Applied Infection Biology at Rheinisch-Westfälische Technische Hochschule (RWTH) Aachen University in Germany. Nanopore base-calling was done with Guppy version 5.0.7. Genome assemblies were performed with Canu (Koren et al. 2017), an assembly pipeline for high-noise single-molecule sequencing, and Flye (Kolmogorov et al. 2019) using default settings. In order to assemble the low coverage and highly similar chromosome 11 and chromosome 13 segments of SD01, two assemblies were performed with Flye using chromosome 13-specific reads and chromosome 11-specific reads to get contigs that represented the chromosome 11 and 13 segments. HB3 was assembled using the Canu assembler with default settings. Note that SD01 had a more disjointed assembly likely due to coming from the last remaining cryopreserved vial that was low parasitemia and nonviable and subsequent lower amount of input DNA. The PacBio/Nanopore reads were mapped to reference genomes using Minimap2, a sequence alignment program (Li 2018). Mappings were visualized using custom R scripts (Gu 2022; R Core Team 2022; Wickham et al. 2022).

DATA AND RESOURCE AVAILABILITY

Nanopore data is available from the SRA (Project # pending). The datasets generated and/or analyzed during the current study are available in SeekDeep, https://seekdeep.brown.edu/Windows_Surrounding_HRP2_3_deletions/, while the code for analyzing Nanopore reads can be found in the Github repository https://github.com/bailey-lab/hrp3.

ACKNOWLEDGEMENTS AND FUNDING SOURCES

Acknowledgements: We thank Drs. Ngwa Julius Che, Matthias Frank, and Gabriele Pradel from Rheinisch-Westfälische Technische Hochschule (RWTH) Aachen University for generously providing a residual SD01 sample. The following reagent was obtained through BEI Resources, NIAID, NIH: *Plasmodium falciparum*, Strain HB3, MRA-155, contributed by Thomas E. Wellems.

Funding Sources: We thank the National Institutes of Allergy and Infectious Diseases (NIAID) for their support via the grants R01AI132547 (JJJ JBP and JAB) and K24AI134990 (JJJ)

Authors Contributions:

Nicholas J. Hathaway: Conception and design of the study, data acquisition and interpretation, and drafting and editing the article

Isaac E. Kim, Jr.: Conception and design of the study, data acquisition and interpretation, and drafting and editing the article

Neeva Wernsman Young: data acquisition, and drafting and editing the article

Sin Ting Hui: data acquisition, and drafting and editing the article

Rebecca Crudale: data acquisition, and drafting and editing the article

David Giesbrecht: data acquisition and interpretation, and drafting and editing the article

Emily Y. Liang: data acquisition, and drafting and editing the article

Christian P. Nixon: data acquisition, and drafting and editing the article

Jonathan J. Juliano: Design of the study, data interpretation, and drafting and editing the article

Jonathan B. Parr: Design of the study, data interpretation, and drafting and editing the article

Jeffrey A. Bailey: Supervision of the project, conception and design of the study, data acquisition and interpretation, and drafting and editing the article

Competing Interests: JBP reports past research support from the World Health Organization focused on *pfhrp2* and *pfhrp3* deletions; as well as research support from Gilead Sciences, non-financial support from Abbott Laboratories, and consulting for Zymeron Corporation outside the scope of this manuscript. All other authors declare that they have no competing interests.

FIGURES

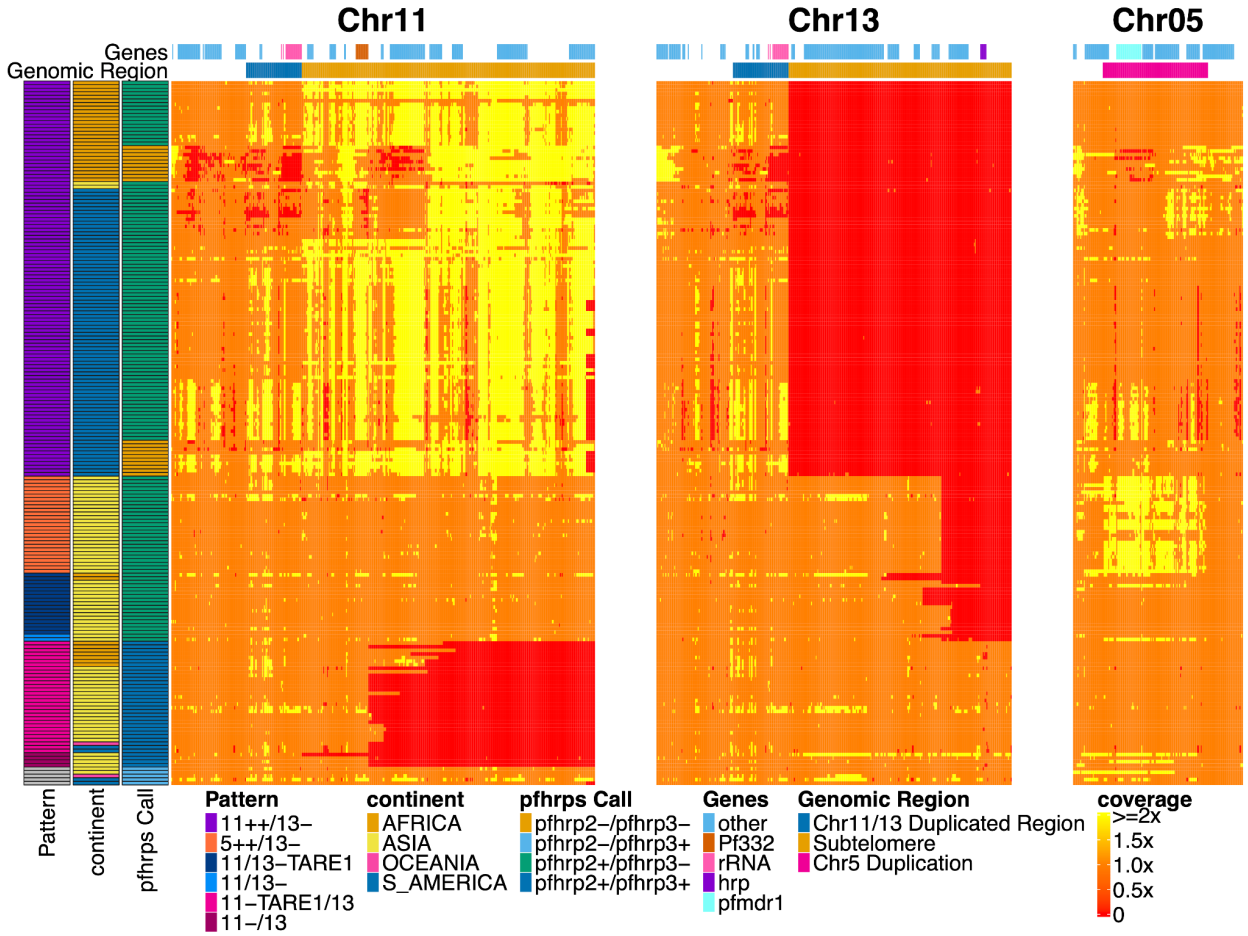

Figure 1. Parasites with altered sequence coverage across regions on chromosomes 11, 13, and 5. Sequence coverage heatmap showing the regions of chromosomes 11 (bp:1,897,157 - 2,003,328), 13 (bp:2,769,916 - 2,844,777), and 5 (bp:944,389 - 988,747) in the subset of the 8,765 samples showing evidence of deletion of either *pfhrp2*, *pfhrp3*, and/or chromosome 11 and control lab isolates. Each row is an isolate. The top annotation along chromosomes depicts the location of genes with relevant genes colored: rRNA (pink), *pf332* (red-orange), *pfhrp3* (purple), *pfmdr1* (electric-blue) and all other genes are colored light-blue. The second row delineates significant genomic regions: The chromosome 11/13 duplicated region (dark blue), the subtelomere regions of chr11/13 (orange), and the chromosome 5 duplicated region (Fuchsia). The left sample annotation includes the genomic rearrangement/deletion pattern (patterns with -TARE1 have evidence of TARE1 sequence at the site of genomic deletion), the continent of origin, and *pfhrp2/3* deletion calls. Variations and biases in coverage correlate with selective whole-genome amplification (sWGA), which does not uniformly amplify the whole plasmodium genome and are likely an artifact of this enrichment process of *P. falciparum*.

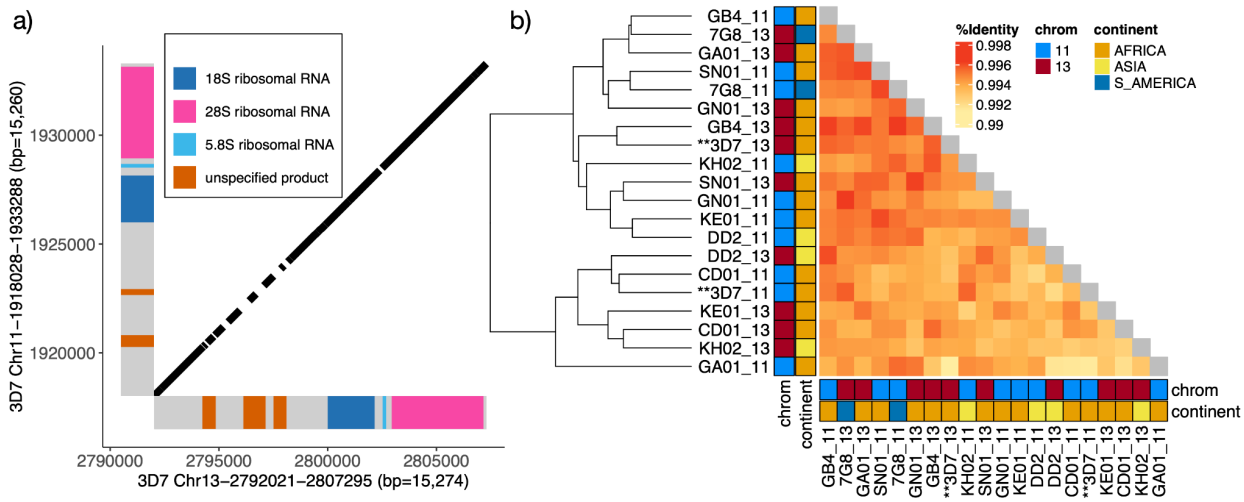

Figure 2. Characterization of the 15.2 kb segmental duplication containing ribosomal genes on chromosomes 11 and 13. (a) Shows the alignment of the 3D7 reference genome copies with an overall identity of 99.3% with diagonal black bars showing 100% conserved regions (>30 bp) representing 89.1% of the alignment. Gene annotation is colored. (b) Comparison by global pairwise alignments of the duplicated copies from non-deleted strains assemblies (n=10) does not show a discrete separation of the paralogs with copies intermixed (chromosome 11 in blue and 13 in red). All copies are ≥99.0% similar to each other with no clear separation by continent or chromosome.

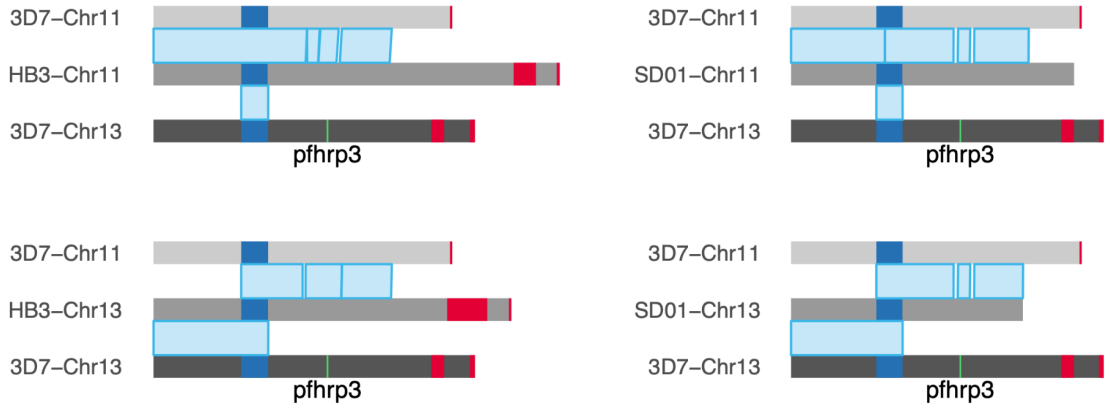

Figure 3. Comparison of long-read assemblies of chromosomes 11 and 13 of HB3 and SD01 to the reference genome 3D7 confirms hybridized chromosome 13-11. Chromosome 11 of HB3 and SD01 mapped entirely to the reference chromosome 11 of 3D7 except for the segmental duplication, which is shared between chromosomes 11 and 13 and denoted in dark blue. The assembly of chromosome 13 of HB3 and SD01 maps to the reference chromosome 13 of 3D7 before the duplicated region, but after the duplication (where *pfhrp3* (green) should be found), it maps to chromosome 11 of 3D7 instead of chromosome 13. Red blocks mark telomere-specific repeats.

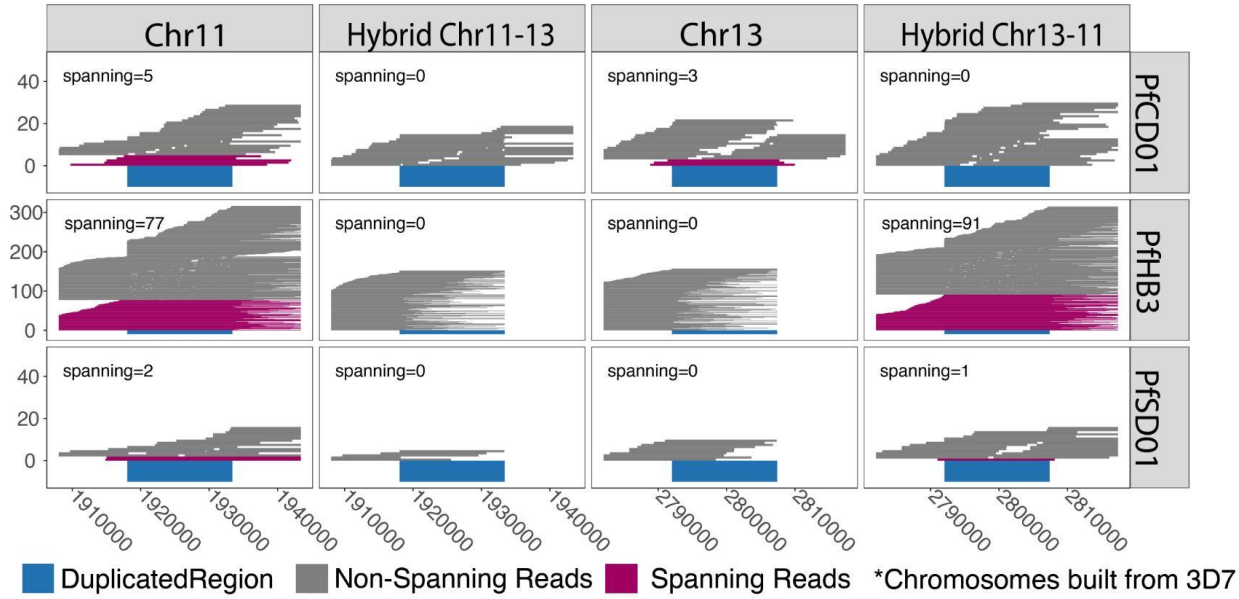

Figure 4. Long reads spanning the 15kb duplicated region confirm presence of translocated chr 13-11 in *pfhrp3*-deleted HB3 (Americas) and SD01 (Africa) but not *pfhrp3*-intact CD01 strains. The best mapping to the segmental duplication region of PacBio and Nanopore reads >15kb for HB3, SD01, and CD01 region are shown against normal chromosomes 11 and 13 as well as hybrid chromosomes 11-13 and 13-11 constructed from 3D7 sequence. Reads that completely span the duplication anchoring in the unique flanking sequence are shown in maroon. Long-reads from CD01 show a control strain where spanning reads are only observed for the normal chromosomes. In contrast, SD01 and HB3 only have reads that span the duplicated region on chromosome 11 but no reads that span out of the duplicated region on chromosome 13. Instead, SD01 and HB3 have spanning reads across the hybrid chromosome 13-11. No samples have spanning reads across the hybrid 11-13 chromosome. The other 13 non-deleted long-read sequenced strains (not shown) had a similar mapping pattern to CD01.

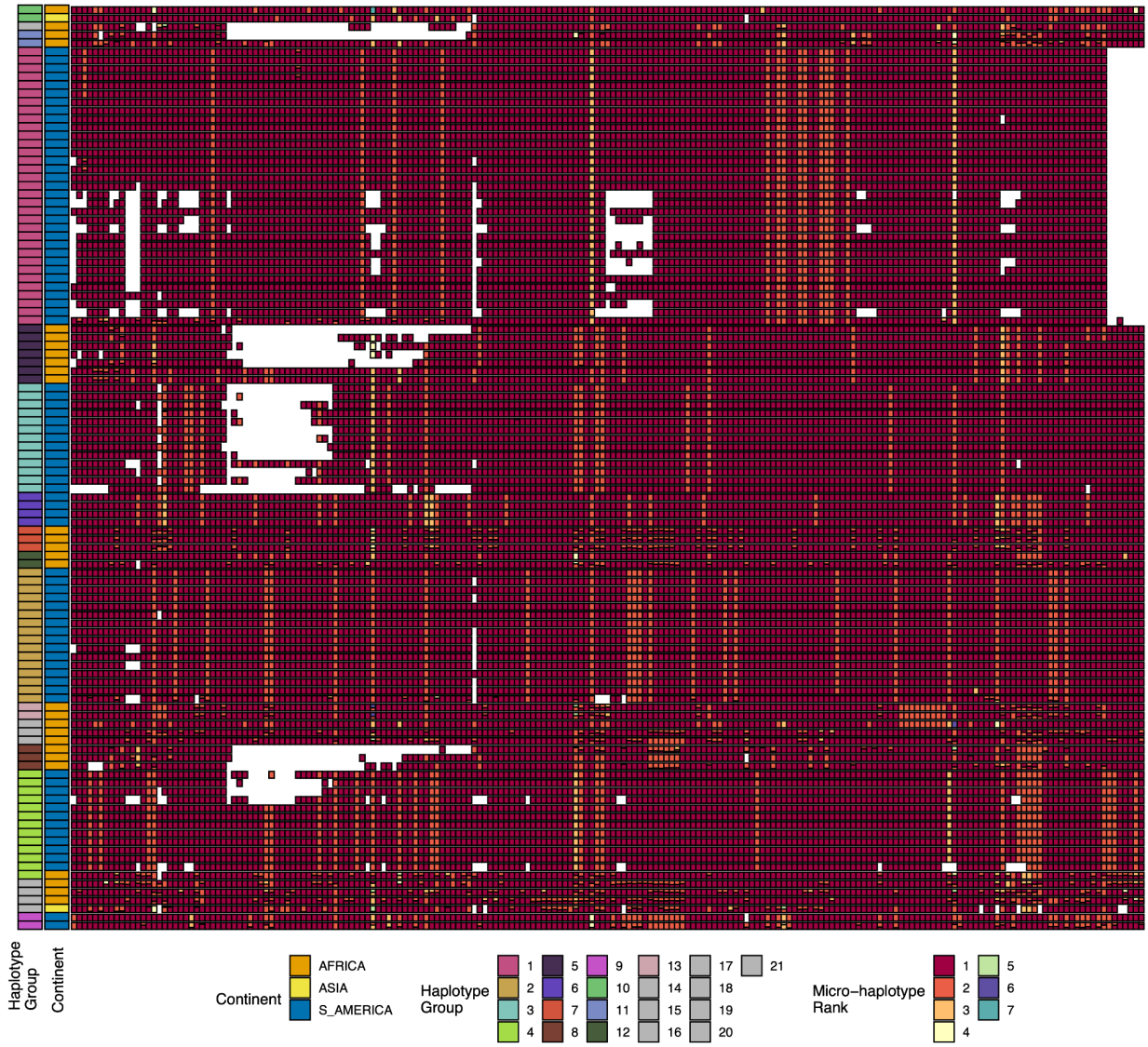

Figure 5. Microhaplotype patterns for the duplicated portion of chromosome 11 in 11++13- parasites indicate 5 major haplotype groups, largely separated by geography with distinct haplotypes observed in American and African strains. Most strains show singular haplotypes at variant positions despite two copies consistent with identical haplotypes. Each row is a 13-11++ isolate, and each column is a different genomic region for which the microhaplotype has been color coded across the duplicated portion of chromosome 11. Only sites with microhaplotype variation are shown (n=202). In each column, the isolate is typed by microhaplotype (colored by the prevalence of each microhaplotype, 1=red being most prevalent, 2=orange second most prevalent, etc). The far leftmost column is colored by haplotype group, determined by hierarchical clustering of the isolates based on the microhaplotype variation. Clusters containing only one isolate are colored gray. The other column shows from which continent the isolate originates.

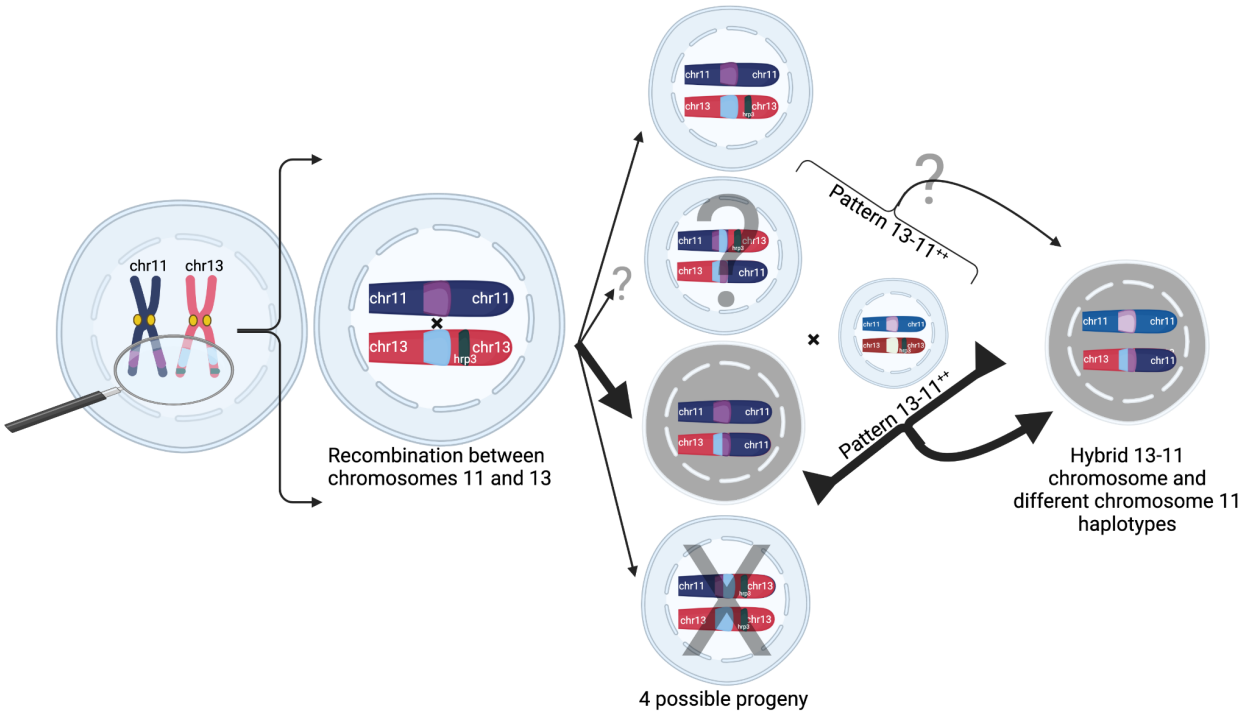

Figure 6. Proposed model of duplication-mediated non-allelic homologous recombination driving translocation during intrastrain meiotic recombination. Homology misalignment and recombination between chromosomes 11 and 13 first occur in an oocyst formed from identical parasites (intrastrain), which can then segregate, resulting in potential progeny (3 translocated progeny). Bold lines show the most direct path to 13-11 hybrid loss of *pfhrp3* and with two copies of chromosome 11 (identical haplotypes in the majority of samples; deletion pattern 13-11^{++}). Subsequent recombination with an unrelated strain yields unrelated chromosome 11 haplotypes. However, there is potential for balanced products, occurring with subsequent recombination events leading to *pfhrp3* loss and either identical haplotypes (intrastrain) or different haplotypes (unrelated strain). The chromosome 11-13 hybrid product was observed once in a lab strain, confirming it can exist and supporting that in vivo selective constraints prevent its emergence. Figure created using Biorender.

References

Amato R, Lim P, Miotto O, Amaratunga C, Dek D, Pearson RD, Almagro-Garcia J, Neal AT, Sreng S, Suon S, et al. 2017. Genetic markers associated with dihydroartemisinin-piperaquine failure in Plasmodium falciparum malaria in Cambodia: a genotype-phenotype association study. *Lancet Infect. Dis.* 17:164–173.

Bailey JA, Eichler EE. 2006. Primate segmental duplications: crucibles of evolution, diversity and disease. *Nat. Rev. Genet.* 7:552–564.

Baker J, McCarthy J, Gatton M, Kyle DE, Belizario V, Luchavez J, Bell D, Cheng Q. 2005. Genetic diversity of Plasmodium falciparum histidine-rich protein 2 (PfHRP2) and its effect on the performance of PfHRP2-based rapid diagnostic tests. *J. Infect. Dis.* 192:870–877.

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

Benelli G, Beier JC. 2017. Current vector control challenges in the fight against malaria. *Acta Trop.* 174:91–96.

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27:573–580.

Berhane A, Anderson K, Mihreteab S, Gresty K, Rogier E, Mohamed S, Hagos F, Embaye G, Chinorumba A, Zehaie A, et al. 2018. Major Threat to Malaria Control Programs by Plasmodium falciparum Lacking Histidine-Rich Protein 2, Eritrea. *Emerg. Infect. Dis.* 24:462–470.

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

Bopp SER, Manary MJ, Bright AT, Johnston GL, Dharia NV, Luna FL, McCormack S, Plouffe D, McNamara CW, Walker JR, et al. 2013. Mitotic evolution of Plasmodium falciparum shows a stable core genome but recombination in antigen families. *PLoS Genet.* 9:e1003293.

Bottius E, Bakhsis N, Scherf A. 1998. Plasmodium falciparum telomerase: de novo telomere addition to telomeric and nontelomeric sequences and role in chromosome healing. *Mol. Cell. Biol.* 18:919–925.

Branton D, Deamer DW, Marziali A, Bayley H, Benner SA, Butler T, Di Ventra M, Garaj S, Hibbs A, Huang X, et al. 2008. The potential and challenges of nanopore sequencing. *Nat. Biotechnol.* 26:1146–1153.

Bruske E, Otto TD, Frank M. 2018. Whole genome sequencing and microsatellite analysis of the Plasmodium falciparum E5 NF54 strain show that the var, rifin and stevor gene families follow Mendelian inheritance. *Malar. J.* 17:376.

Calhoun SF, Reed J, Alexander N, Mason CE, Deitsch KW, Kirkman LA. 2017. Chromosome End Repair and Genome Stability in Plasmodium falciparum. *MBio* [Internet] 8. Available from: http://dx.doi.org/10.1128/mBio.00547-17

Chen DS, Barry AE, Leliwa-Sytek A, Smith T-A, Peterson I, Brown SM, Migot-Nabias F, Deloron P, Kortok MM, Marsh K, et al. 2011. A molecular epidemiological study of var gene diversity to characterize the reservoir of Plasmodium falciparum in humans in Africa. *PLoS One* 6:e16629.

Cheng Q, Gatton ML, 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.

Chiodini PL. 2014. Malaria diagnostics: now and the future. *Parasitology* 141:1873–1879.

Claessens A, Hamilton WL, Kekre M, Otto TD, Faizullabhoy A, Rayner JC, Kwiatkowski D. 2014. Generation of antigenic diversity in Plasmodium falciparum by structured

rearrangement of Var genes during mitosis. *PLoS Genet.* 10:e1004812.

Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P, et al. 2010. Origins and functional impact of copy number variation in the human genome. *Nature* 464:704–712.

Consortium T 1000 GP, The 1000 Genomes Project Consortium. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* [Internet] 491:56–65. Available from: http://dx.doi.org/10.1038/nature11632

Day KP, Artzy-Randrup Y, Tiedje KE, Rougeron V, Chen DS, Rask TS, Rorick MM, Migot-Nabias F, Deloron P, Luty AJF, et al. 2017. Evidence of strain structure in *Plasmodium falciparum var* gene repertoires in children from Gabon, West Africa. *Proc. Natl. Acad. Sci. U. S. A.* 114:E4103–E4111.

Didelot X, Méric G, Falush D, Darling AE. 2012. Impact of homologous and non-homologous recombination in the genomic evolution of Escherichia coli. *BMC Genomics* 13:256.

Duvalsaint M, Conrad MD, Tukwasibwe S, Tumwebaze PK, Legac J, Cooper RA, Rosenthal PJ. 2021. Balanced impacts of fitness and drug pressure on the evolution of PfMDR1 polymorphisms in Plasmodium falciparum. *Malar. J.* 20:292.

Feleke SM, Reichert EN, Mohammed H, Brhane BG, Mekete K, Mamo H, Petros B, Solomon H, Abate E, Hennelly C, et al. 2021. Plasmodium falciparum is evolving to escape malaria rapid diagnostic tests in Ethiopia. *Nat Microbiol* 6:1289–1299.

Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, et al. 2002. Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature* 419:498–511.

Góes L, Chamma-Siqueira N, Peres JM, Nascimento JM, Valle S, Arcanjo AR, Lacerda M, Blume L, Póvoa M, Viana G. 2020. Evaluation of Histidine-Rich Proteins 2 and 3 Gene Deletions in Plasmodium falciparum in Endemic Areas of the Brazilian Amazon. *Int. J. Environ. Res. Public Health* [Internet] 18. Available from: http://dx.doi.org/10.3390/ijerph18010123

Gu Z. 2022. ComplexHeatmap: Make Complex Heatmaps.

Harris RS. 2007. Improved pairwise alignment of genomic DNA. Available from: http://www.bx.psu.edu/~rsharris/rsharris_phd_thesis_2007.pdf

Hinterberg K, Mattei D, Wellems TE, Scherf A. 1994. Interchromosomal exchange of a large subtelomeric segment in a Plasmodium falciparum cross. *EMBO J.* 13:4174–4180.

Hodder AN, Maier AG, Rug M, Brown M, Hommel M, Pantic I, Puig-de-Morales-Marinkovic M, Smith B, Triglia T, Beeson J, et al. 2009. Analysis of structure and function of the giant protein Pf332 in Plasmodium falciparum. *Mol. Microbiol.* 71:48–65.

Huckaby AC, Granum CS, Carey MA, Szlachta K, Al-Barghouthi B, Wang Y-H, Guler JL. 2019. Complex DNA structures trigger copy number variation across the Plasmodium falciparum genome. *Nucleic Acids Res.* 47:1615–1627.

Isabel Veiga M, Dhingra SK, Henrich PP, Straimer J, Gnädig N, Uhlemann A-C, Martin RE,

Lehane AM, Fidock DA. 2016. Globally prevalent PfMDR1 mutations modulate Plasmodium falciparum susceptibility to artemisinin-based combination therapies. *Nat. Commun.* 7:11553.

Jain M, Olsen HE, Paten B, Akeson M. 2016. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol.* 17:239.

Jaskiewicz E, Jodłowska M, Kaczmarek R, Zerka A. 2019. Erythrocyte glycophorins as receptors for Plasmodium merozoites. *Parasit. Vectors* 12:317.

Kasianowicz JJ, Brandin E, Branton D, Deamer DW. 1996. Characterization of individual polynucleotide molecules using a membrane channel. *Proc. Natl. Acad. Sci. U. S. A.* 93:13770–13773.

Kidd JM, Graves T, Newman TL, Fulton R, Hayden HS, Malig M, Kallicki J, Kaul R, Wilson RK, Eichler EE. 2010. A human genome structural variation sequencing resource reveals insights into mutational mechanisms. *Cell* 143:837–847.

Kirkman LA, Lawrence EA, Deitsch KW. 2014. Malaria parasites utilize both homologous recombination and alternative end joining pathways to maintain genome integrity. *Nucleic Acids Res.* 42:370–379.

Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* 37:540–546.

Kong A, Wilson SA, Ah Y, Nace D, Rogier E, Aidoo M. 2021. HRP2 and HRP3 cross-reactivity and implications for HRP2-based RDT use in regions with Plasmodium falciparum hrp2 gene deletions. *Malar. J.* 20:207.

Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, Simons JF, Kim PM, Palejev D, Carriero NJ, Du L, et al. 2007. Paired-end mapping reveals extensive structural variation in the human genome. *Science* 318:420–426.

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27:722–736.

Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. *Genome Biol.* 5:R12.

Lee AH, Symington LS, Fidock DA. 2014. DNA repair mechanisms and their biological roles in the malaria parasite Plasmodium falciparum. *Microbiol. Mol. Biol. Rev.* 78:469–486.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100.

Liu P, Lacaria M, Zhang F, Withers M, Hastings PJ, Lupski JR. 2011. Frequency of nonallelic homologous recombination is correlated with length of homology: evidence that ectopic synapsis precedes ectopic crossing-over. *Am. J. Hum. Genet.* 89:580–588.

Mattei D, Scherf A. 1994. Subtelomeric chromosome instability in Plasmodium falciparum: short telomere-like sequence motifs found frequently at healed chromosome breakpoints. *Mutat. Res.* 324:115–120.

Miles A, Iqbal Z, Vauterin P, Pearson R, Campino S, Theron M, Gould K, Mead D, Drury E, O'Brien J, et al. 2016. Indels, structural variation, and recombination drive genomic diversity in Plasmodium falciparum. *Genome Res.* 26:1288–1299.

Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, Abyzov A, Yoon SC, Ye K, Cheetham RK, et al. 2011. Mapping copy number variation by population-scale genome sequencing. *Nature* 470:59–65.

Nair S, Li X, Nkhoma SC, Anderson T. 2022. Fitness Costs of pfhrp2 and pfhrp3 Deletions Underlying Diagnostic Evasion in Malaria Parasites. *J. Infect. Dis.* 226:1637–1645.

Nair S, Nash D, Sudimack D, Jaidee A, Barends M, Uhlemann A-C, Krishna S, Nosten F, Anderson TJC. 2007. Recurrent gene amplification and soft selective sweeps during evolution of multidrug resistance in malaria parasites. *Mol. Biol. Evol.* 24:562–573.

Nilsson S, Moll K, Angeletti D, Albrecht L, Kursula I, Jiang N, Sun X, Berzins K, Wahlgren M, Chen Q. 2011. Characterization of the Duffy-Binding-Like Domain of Plasmodium falciparum Blood-Stage Antigen 332. *Malar. Res. Treat.* 2011:671439.

Organization WH, Others. 2011. Good practices for selecting and procuring rapid diagnostic tests for malaria. Available from: https://apps.who.int/iris/bitstream/handle/10665/44530/9789241501125_eng.pdf

Otto TD, Böhme U, Sanders M, Reid A, Bruske EI, Duffy CW, Bull PC, Pearson RD, Abdi A, Dimonte S, et al. 2018. Long read assemblies of geographically dispersed Plasmodium falciparum isolates reveal highly structured subtelomeres. *Wellcome Open Res* 3:52.

Otto TD, Gilabert A, Crellen T, Böhme U, Arnathau C, Sanders M, Oyola SO, Okouga AP, Boundenga L, Willaume E, et al. 2018. Genomes of all known members of a Plasmodium subgenus reveal paths to virulent human malaria. *Nat Microbiol* 3:687–697.

Parks MM, Lawrence CE, Raphael BJ. 2015. Detecting non-allelic homologous recombination from high-throughput sequencing data. *Genome Biol.* 16:72.

Pologe LG, Ravetch JV. 1988. Large deletions result from breakage and healing of P. falciparum chromosomes. *Cell* 55:869–874.

Poti KE, Sullivan DJ, Dondorp AM, Woodrow CJ. 2020. HRP2: Transforming Malaria Diagnosis, but with Caveats. *Trends Parasitol.* 36:112–126.

Price RN, Uhlemann A-C, Brockman A, McGready R, Ashley E, Phaipun L, Patel R, Laing K, Looareesuwan S, White NJ, et al. 2004. Mefloquine resistance in Plasmodium falciparum and increased pfmdr1 gene copy number. *Lancet* 364:438–447.

Rask TS, Hansen DA, Theander TG, Gorm Pedersen A, Lavstsen T. 2010. Plasmodium falciparum erythrocyte membrane protein 1 diversity in seven genomes--divide and conquer. *PLoS Comput. Biol.* [Internet] 6. Available from: http://dx.doi.org/10.1371/journal.pcbi.1000933

R Core Team. 2022. R: A Language and Environment for Statistical Computing. Available from: https://www.R-project.org/

Reed J, Kirkman LA, Kafsack BF, Mason CE, Deitsch KW. 2021. Telomere length dynamics in

response to DNA damage in malaria parasites. *iScience* 24:102082.

Rug M, Cyrklaff M, Mikkonen A, Lemgruber L, Kuelzer S, Sanchez CP, Thompson J, Hanssen E, O'Neill M, Langer C, et al. 2014. Export of virulence proteins by malaria-infected erythrocytes involves remodeling of host actin cytoskeleton. *Blood* 124:3459–3468.

Samarakoon U, Gonzales JM, Patel JJ, Tan A, Checkley L, Ferdig MT. 2011. The landscape of inherited and de novo copy number variants in a Plasmodium falciparum genetic cross. *BMC Genomics* 12:457.

Sander AF, Lavstsen T, Rask TS, Lisby M, Salanti A, Fordyce SL, Jespersen JS, Carter R, Deitsch KW, Theander TG, et al. 2014. DNA secondary structures are associated with recombination in major Plasmodium falciparum variable surface antigen gene families. *Nucleic Acids Res.* 42:2270–2281.

Sander AF, Salanti A, Lavstsen T, Nielsen MA, Magistrado P, Lusingu J, Ndam NT, Arnot DE. 2009. Multiple var2csa-type PfEMP1 genes located at different chromosomal loci occur in many Plasmodium falciparum isolates. *PLoS One* 4:e6667.

Scherf A, Mattei D. 1992. Cloning and characterization of chromosome breakpoints of Plasmodium falciparum: breakage and new telomere formation occurs frequently and randomly in subtelomeric genes. *Nucleic Acids Res.* 20:1491–1496.

Steinbiss S, Silva-Franco F, Brunk B, Foth B, Hertz-Fowler C, Berriman M, Otto TD. 2016. Companion: a web server for annotation and analysis of parasite genomes. *Nucleic Acids Res.* 44:W29–W34.

Tessema SK, Monk SL, Schultz MB, Tavul L, Reeder JC, Siba PM, Mueller I, Barry AE. 2015. Phylogeography of var gene repertoires reveals fine-scale geospatial clustering of Plasmodium falciparum populations in a highly endemic area. *Mol. Ecol.* 24:484–497.

Thomson R, Parr JB, Cheng Q, Chenet S, Perkins M, Cunningham J. 2020. Prevalence of Plasmodium falciparum lacking histidine-rich proteins 2 and 3: a systematic review. *Bull. World Health Organ.* 98:558–568F.

Vera-Arias CA, Holzschuh A, Oduma CO, Badu K, Abdul-Hakim M, Yukich J, Hetzel MW, Fakih BS, Ali A, Ferreira MU, et al. 2022. High-throughput Plasmodium falciparum hrp2 and hrp3 gene deletion typing by digital PCR to monitor malaria rapid diagnostic test efficacy. *Elife* [Internet] 11. Available from: http://dx.doi.org/10.7554/eLife.72083

Vergés L, Molina Ò, Geán E, Vidal F, Blanco J. 2014. Deletions and duplications of the 22q11.2 region in spermatozoa from DiGeorge/velocardiofacial fathers. *Molecular Cytogenetics* [Internet] 7. Available from: http://dx.doi.org/10.1186/s13039-014-0086-3

Verma AK, Bharti PK, Das A. 2018. HRP-2 deletion: a hole in the ship of malaria elimination. *Lancet Infect. Dis.* 18:826–827.

Vernick KD, McCutchan TF. 1988. Sequence and structure of a Plasmodium falciparum telomere. *Mol. Biochem. Parasitol.* 28:85–94.

WHO. 2021. Statement by the Malaria Policy Advisory Group on the urgent need to address the high prevalence of pfhrp2/3 gene deletions in the Horn of Africa and beyond. Available

from: https://www.who.int/news/item/28-05-2021-statement-by-the-malaria-policy-advisory-group-on-the-urgent-need-to-address-the-high-prevalence-of-pfhrp2-3-gene-deletions-in-the-horn-of-africa-and-beyond

WHO. 2022. World malaria report 2022. Available from: https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2022

WHO. Malaria. Available from: https://www.who.int/news-room/fact-sheets/detail/malaria

Wickham H, Chang W, Henry L, Pedersen TL, Takahashi K, Wilke C, Woo K, Yutani H, Dunnington D. 2022. ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics. Available from: https://CRAN.R-project.org/package=ggplot2

Witkowski B, Duru V, Khim N, Ross LS, Saintpierre B, Beghain J, Chy S, Kim S, Ke S, Kloeung N, et al. 2017. A surrogate marker of piperaquine-resistant Plasmodium falciparum malaria: a phenotype-genotype association study. *Lancet Infect. Dis.* 17:174–183.

Zhang X, Alexander N, Leonardi I, Mason C, Kirkman LA, Deitsch KW. 2019. Rapid antigen diversification through mitotic recombination in the human malaria parasite Plasmodium falciparum. *PLoS Biol.* 17:e3000271.