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

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_defeo@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-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 encoded by the *pfhrp2* and *pfhrp3* genes. Deletions of both *pfhrp2* and *pfhrp3* make parasites undetectable by RDTs. However, deletions have emerged in settings where RDT use is uncommon and can occur independent 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 analyzed 8,765 publicly available, whole-genome-sequenced *P. falciparum* field samples. Two *pfhrp3* deletion patterns predominated: 1) chromosome 13 loss centromeric to *pfhrp3* extending to the chromosome end with duplication of chromosome 11's subtelomeric region (Americas/Africa; Pattern $13^{-}11^{++}$) and 2) deletion of chromosome 13 starting at varied locations centromeric to *pfhrp3* (Asia; Pattern 13^{-}). Pattern $13^{-}11^{++}$'s centromeric boundary fell within the largest (15kb) segmental duplication in the core genome containing ribosomal genes. Long-read sequencing of *pfhrp3*-deleted isolates from Honduras (HB3) and Sudan (SD01) confirmed 13-11 hybrid chromosomes. The majority of isolates had the same haplotype (lack of variation between copies) along their chr. 11 and chr. 13-11 hybrid shared regions, suggesting intrastrain (self) meiotic recombination as the primary mechanism generating $13^{-}11^{++}$. Together, our findings demonstrate that *pfhrp3* deletion in the Americas and Africa often results from duplication-mediated non-homologous recombination, which drives translocation and creation of a hybrid 13-11 chromosome lacking *pfhrp3*. Furthermore, this suggests specific conditions including low transmission are needed for emergence, a finding that can inform refined surveillance strategies.

SIGNIFICANCE STATEMENT

Most cases of *Plasmodium falciparum* malaria are diagnosed using rapid diagnostic tests (RDTs) that detect histidine-rich protein 2 (HRP2) and HRP3. RDT effectiveness is threatened by the emergence of parasites that escape detection owing to deletion of the *pfhrp2* and *pfhrp3* genes. However, these deletions have been difficult to detect and study, and we lack understanding of their causes. Using whole-genome short and long-read sequencing, we found that duplication-mediated non-homologous recombination leading to translocation and a hybrid chromosome may be responsible for *pfhrp3* deletion in the Americas and Africa, suggesting that the deletion occurs frequently and can only emerge under specific conditions. Defining these conditions will provide avenues for improved control and surveillance.

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 (1). 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 (2). 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 (3). Adding to the problems of insecticide and drug resistance, the emergence of "diagnostic-resistant" *P. falciparum* strains with deletion mutations in their subtelomeric regions represent a new threat to malaria control and elimination efforts (4).

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 (5). The predominant falciparum malaria RDTs detect *P. falciparum* histidine-rich protein 2 (PfHRP2) and are cross-reactive with its paralogue PfHRP3. The number of PfHRP2-based RDTs increased exponentially from less than 1 million sold in 2004 to more than 300 million in 2013 (6). 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 (6–10). These parasites appear to be spreading rapidly in some regions and have compromised existing test-and-treat programs, especially in the Horn of Africa (4, 11–13).

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 (10). 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 (11). The reason why *pfhrp3* deletion occurred first remains unclear. A 1994 study of the HB3 laboratory strain reported a *pfhrp3* deletion that was possibly caused by a subtelomeric pairing and exchange between chromosomes 11 and 13 in meiosis (14). Further examination of recombination as a possible mechanism for *pfhrp3* deletion in current 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,* AT-rich intergenic sequence is challenging, and repeats and highly similar gene families abound. Subtelomeric regions representing roughly 5% of the genome contain rapidly evolving gene families (*var*, *rifin*, *stevor*, etc) that undergo frequent non-allelic homologous recombination (NAHR) and conversion between chromosomes, thereby generating antigenic breadth (15, 16). Subtelomeric/centromeric exchange importantly allows for unbalanced progeny without large-scale ramifications of altering a larger proportion of a chromosome. Once duplications are present, then more duplications can be generated through NAHR between highly identical paralogs, and combined gene families can spread across subtelomeric regions. Once duplications are present, then these regions are predisposed to further duplication through NAHR between highly identical paralogs, resulting in the spread of combined gene families across subtelomeric regions (17–23). These duplication-rich regions can rapidly gain or lose sequence and lead to phenotype and disease (24, 25). The other 95% of *Plasmodium falciparum* genome (termed the core genome) is highly conserved with strong synteny between species (26–29). The exact mechanisms of exchanges in the subtelomeric regions involving *pfhrp2* and *pfhrp3* have not been studied in depth. However, parasites are likely unable to repair DNA double-strand breaks via traditional canonical non-homologous end joining and instead rely more on homologous recombination (30–36).

Improved understanding of the mechanism of these deletions could provide new insights into the evolutionary pressures driving their emergence and help inform strategies to address them. Though available whole-genome sequences include few *pfhrp2* deletions, we found that parasites with *pfhrp3* deletion available from multiple countries had geographically distinct copy number changes. We therefore sought 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 and how they emerged, providing opportunities for improved surveillance.

RESULTS

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

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

Two distinct pfhrp3 deletion patterns with geographical associations

Analysis of the publicly available WGS data revealed two distinct *pfhrp3* deletion patterns based on read depth analysis: 1) 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 chr13 with chromosome 11 (**pattern $13^{-}11^{++}$**); and 2) deletion of chromosome 13 starting at various locations centromeric to *pfhrp3* and unassociated with other copy number changes (**pattern 13^{-}**) (**Figure 1**). Relative to the reference genome, pattern $13^{-}11^{++}$ is missing 46,323 bases from the core genome of chromosome 13 (from 2,807,159 to 2,853,482 and encompassing *pfhrp3* at 2,840,236 to 2,842,840) and has gained 70,175 bases of chromosome 11's core genome (from 1,933,138 to 2,003,313). Based on genomes that were assembled to the end of their telomeres (29), an additional 17kb to 84kb is deleted from the paralogous sub-telomeric region on chromosome 13, and an additional 15kb to 87kb of the paralogous sub-telomeric region on chromosome 11 is duplicated. Pattern 13^{-} showed deletions of the core genome on average of 19kb (ranging from 13kb to 31kb) encompassing the *pfhrp3* gene. Among the 153 field samples with *pfhrp3* deletion, 107 and 46 demonstrated patterns $13^{-}11^{++}$ and 13^{-}, respectively. Pattern $13^{-}11^{++}$ was almost exclusively found in samples from Africa and the Americas, while pattern 13^{-} was observed predominantly in Southeast Asia (**Figure 1**).

Pattern 13^{-} predominantly falls within FIKK genes

Pattern 13^{-}'s centromeric breakpoint appears variable, with multiple deletion breakpoints centromeric to *pfhrp3*. All deletion breakpoints appeared to occur within a region containing 4 PHIST protein genes located between PF3D7_1371700, a FIKK family protein gene, and *pfhrp3*. The majority of the breaks occurred before the first PHIST (PF3D7_1372100) protein (**Figure 1**). This deletion type was seen predominantly in Southeast Asia, with 42 parasites identified in Southeast Asia - East and 2 in Southeast Asia - West, while none were observed in the Americas and only 1 in East Africa and 1 in West Africa (**Figure 1**).

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

Pattern 13^{-}11^{++} was common among parasites from the Americas and Africa, with a centromeric breakpoint that appears 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 (37), this segmental duplication was the largest inter-chromosomal duplication in the core genome and was 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 **(Figure S3)**. We further investigated 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), were nearly identical (99.0% similarity) (**Figure 2**) and oriented similarly. Each copy contained 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 (15) ribosomal loci (99.7% identity) (**Figure 2**, **Figure S4**, **Figure S5**). Pairwise alignment of the duplicated segments of chromosomes 11 and 13 from non-deleted strains showed 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 did not fully resolve chromosome 11 and 13 subtelomeres

Given that the pattern of deletion and duplication and the breakpoints were suggestive of duplication-mediated recombination leading to translocation, we sought to examine high-quality PacBio genome assemblies of other strains containing the *pfhrp3* deletion. However, the gene annotations by Companion (38) of chromosome 11 of previous PacBio assemblies (29) showed that the assemblies of strains SD01 and HB3 were 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 was found only on chromosome 13 (**Figure S6 and Figure S7**), 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 chr 11 subtelomeric region sequence into two copies.

Combined analysis of Nanopore and PacBio reads confirmed 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 examined two representative parasite isolates with known *pfhrp3* deletion: HB3 from Honduras and SD01 from Sudan. We used Nanopore's 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 combined our Nanopore data with the publicly available PacBio sequencing data and tested for the presence of hybrid chromosomes using a two-prong 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 constructed 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 mapped 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 spanned the duplicated region (**Figure 4**). Reads were considered "spanning" if they extended at least 50 bp into the flanking unique regions. HB3 had 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 had 2 chromosome 11 spanning reads and 1 13-11 chromosome spanning read. Neither strain had PacBio or Nanopore spanning reads across the normal chromosome 13 and hybrid 11-13 (**Figure 4**) which would be consistent with the reciprocal meiotic product. Due to lack of sufficient sample, SD01 had a small number of reads that spanned the full region. Sub-analysis revealed 4 regions within this duplicated region that had chromosome 11 and 13 specific nucleotide variation that was able to be used to further bridge across this region for additional confirmation (**Figure S8** and **Figure S9**). The other 14 strains with intact *pfhrp3* from the PacBio dataset (29) had reads that spanned chromosome 11 and chromosome 13 but lacked any spanning reads across the hybrid 13-11 or 11-13 chromosomes. For example, the *pfhrp3*-intact strain CD01 had 5 and 3 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 confirmed a segmental duplicated region of these chromosomes.

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 were created for the HB3 and SD01 lab strains from Nanopore long reads. HB3 assembly yielded 16 contigs representing complete chromosomes. Telomere-associated tandem repeat 1 (TARE-1) (16) was 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 was achieved, N50 1,595,798 and L50 5. SD01, however, likely due to its lower sequencing coverage, had a more disjointed assembly with 200 final contigs, N50 263,459, and L50 30. The HB3 assembly had a chromosome 11 that closely matched 3D7

chromosome 11 and a hybrid 13-11 that closely matched 3D7 chromosome 13 up to the ribosomal duplication region where it then subsequently best matched the 3D7 chromosome 11 (**Figure 3**, **Figure S10**). Both of these chromosomes had TARE-1 (16) at their 5' and 3` end indicating that these were full-length chromosome assemblies. The SD01 assembly had two contigs – one that matched 3D7 chromosome 11 and the other that matched 3D7 chromosome 13 up to the duplicated region and then subsequently matched 3D7 chromosome 11.

These new assemblies were further annotated for genes by Companion (38). The contig matching the hybrid chromosome 13-chromosome 11 for both strains had the genes found on 3D7 chromosome 11 after the duplicated region. The duplicated genes on chromosomes 11 and 13-11 that shared homology with 3D7 chromosome 11 subtelomere included *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 (**Figure S11 and Figure S12**). The number of DnaJ genes in the 3D7 genome is approximately 27 but are not highly similar with <67% nucleotide identity. Homology between HB3 chromosome 11 and chromosome 13-11 continued up through a *rifin*, then a *stevor* gene, and then the sequence completely diverged afterward with even a different gene family organization structure consisting of *stevor*, *rifin,* and *var* gene families among other paralogous gene families **(Figure S11)**. The chromosome 13-11 SD01 contig reached the DNAJ protein and terminated (**Figure S12**), likely because the assembly was unable to contend with the near complete identical sequence between chromosome 11 and chromosome 13-11. Illumina sequences mapped to the *var* regions on SD01 chromosome 11 where chromosome 13-11 terminated, and local assemblies revealed no variation but double coverage as compared to the rest of the genome, which would be 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 (29) with normal chromosome 11 showed that homology between strains also end 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 corresponded 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 confirmed the presence of a normal chromosome 11 and hybrid chromosome 13-11 that shared a duplicated region.

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

To look at conservation of the segmental duplication containing the ribosomal genes, we examined genomes of closely related *Plasmodium* parasites in the *Laverania* subgenus, which comprises *P. falciparum* and ape *Plasmodium* spp. We found evidence of the segmental duplication in *Plasmodium praefalciparum,* which is *P. falciparum's* closest relative having diverged about 50,000 years ago. Its genome also contained the similar S-type rRNA loci on chromosomes 11 and 13 and had a similar gene synteny to *P. falciparum* in these regions, as well as the region on chromosome 8 close to *pfhrp2* (**Figure S13**, **Figure S14**, **and Figure S15**). *P. praefalciparum* contained the 15.2kb duplicated region on both chromosomes 11 and 13 and was 96.7% similar to the 3D7 duplicated region. Other *Laverania* genomes (39) were not fully assembled within their subtelomeric regions.

Genomic localization of breakpoint for $13^{-}11^{++}$ does not show one distinctive breakpoint

The analysis of the variation within the 15.2kb duplicated region between strains that contained 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 which 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 (**Figure S5**). 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.

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

Pattern $13^{-}11^{++}$ was 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 110 parasites with this deletion pattern, 89 parasites (77 of the 80 American, 11 of the 28 African parasites, and 1 of the 2 Asian) had near-identical copies of the chromosome 11 duplicated region. Near-identical copies were defined as having ≥99% identity across 382 variant sites within the duplicated region identical (same variant allele between copies). These 89 parasites containing identical copies did not all share the same haplotypes but rather show 5 major groups (**Figure 5)**. The remaining 21 parasites had variation within this region; on average, 10.2% of variant sites differed between the copies (min 83.8% identity). The 5 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 (**Figure S16**, **Figure S17**, **Figure S18**, **Figure S19**, **Figure S20**, **and Figure S21**). 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.

Deletion of chromosome 11 with duplication of chromosome 13 observed in one lab isolate but no field isolates

There are 33 isolates with evidence of sub-telomeric chromosome 11 deletions. The majority of these deletions begin immediately after Pf332. Only one parasite has deletion up and through Pf332 to the duplicated region with subsequent duplication of chromosome 13 that would be consistent with the mirror of $13^{-}11^{++}$, which is lab isolate FCR3. There are no field samples with this pattern. As a related clone of FCR3, IT does not contain this pattern and would indicate that FCR3 obtained this duplication of 13 via translocation within culture and not in the field.

DISCUSSION

Here, we used publicly available short-read and newly generated long-read sequencing data to identify mechanisms by which *pfhrp3* deletion commonly occurs. Our findings are consistent with a translocation event, which appears to be the predominant mechanism in the Americas and the Horn of Africa. We then long-read sequenced SD01 and HB3 and confirmed the presence of a hybrid 13-11 chromosome and its translocation breakpoint within a high-identity segmental duplication containing ribosomal genes that resides on chromosomes 11 and 13, consistent with a mechanism of non-allelic homologous recombination. The presence of HB3 and SD01 spanning reads across the duplicated region of chromosomes 11 and 13-11, as well as the assembled complete genome of HB3 and the partially assembled SD01 genome, indicate an interchromosomal exchange of the subtelomeric portion of chromosome 11 onto that of chromosome 13 via non-homologous recombination. Thus, this recombination event creates a hybrid 13-11 chromosome and results in the deletion of *pfhrp3* and its surrounding genes from

the duplicated region and onwards (**Figure 3** and **Figure 4**). Such an event is consistent with the genomic coverage pattern we observed in publicly available data from 110 *pfhrp3*-deleted field samples (**Figure 1**).

Based on our findings, 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 4 potential progeny including 1 with normal chromosomes 11 and 13 and 3 with translocations. This would account for the identical haplotypes observed in the 2 copies of the chromosome 11 segment. Based on the identical haplotypes observed in the majority of samples, the most direct and likely mechanism involves a progeny with 2 copies of chromosome 11 recombining with an unrelated strain to yield unrelated chromosome 11 haplotypes. Another possible albeit less direct mechanism may involve a progeny with balanced but hybrid 13-11 and 11-13 chromosomes recombining with the same strain or unrelated strain to yield identical or different haplotypes, respectively **(Figure 6)**. This non-homologous recombination mechanism resulting in the translocation and subsequent *pfhrp3* deletion may provide significant biologic insight and have public health implications.

This duplication-mediated non-allelic homologous recombination event is likely to occur frequently during meiosis. Meiotic misalignment in NAHR can lead to high-frequency chromosomal aberrations (eg. in humans, 22q11 deletion syndrome due to misalignment of duplicated blocks on chromosome 22 occurs in 1 in 4000 births) (40). In the parasite, the chromosome 11 and 13 translocation is likely a common occurrence within parasite populations around the world. The lack of hybrid chromosome 13-11 elsewhere suggests such events are quickly removed from the population due to fitness costs, an idea supported by recent *in vitro* competition studies (41). 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. However, if this is limited to within host competition, this fitness costs may disappear

in low transmission settings where other selective forces including RDTs and unknown biological factors can come into play.

The benefit in America and the Horn of Africa conferred by the loss of *pfhrp3* and *pfhrp2* is unclear and complicated due to the involvement of other genes. First in and of itself, the loss of *pfhrp2* and *pfhrp3* likely confer a survival advantage related to their specific biology in the Americas given that the two genes are on different loci and that both are lost in areas without RDTs. Nair *et al.* proposed that the fitness cost potentially associated with these two gene deletions may be outweighed by two factors: 1) low effective population sizes of the parasite in the Americas may decrease the selection against these alleles and 2) high rates of linkage disequilibrium between genes on different chromosomes in parasites in the Americas (41–44). Besides *pfhrp2* and *pfhrp3, EBL-1*, which helps enable *P. falciparum* to invade erythrocytes, is lost, which may cause infected patients to have fewer symptoms and be less likely to seek treatment (11, 45). 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. *P. reichenowi* contains all the genes within the duplicated chromosome 11, suggesting that these genes might be more essential than the chromosome 13 genes. Thus, the loss of the *pfhrp3* and *EBL-1* genes may increase *P. falciparum*'s fitness (11).

Additionally, the duplicated segments of chromosomes 11 and 13 from non-deleted strains demonstrated very high similarity and intermixed paralogs, suggesting ongoing gene conversions and translocations. Only parasites with a duplicated chromosome 11 and deleted chromosome 13 were detected in our analysis in the field samples; the complementary situation of segmental duplicated chromosome 13 and deleted chromosome 11 was never observed in any of the field samples. Only lab isolate FCR3 had any indication from coverage data that it had a duplicated chromosome 13 and a deleted chromosome 11. Given that the majority of the publicly available field samples were collected from studies using RDT-positive samples and that PfHRP3 encoded by duplicated *pfhrp3* would likely have been detected by RDT, samples

should not have biased against detecting the reciprocal hybrid 11-13 (which would have two copies of *pfhrp3*). The apparent absence of parasites with this deletion/duplication pattern would be consistent with several scenarios: a benefit of duplicated chromosome 11, a significant detriment to an absence of chromosome 11, a benefit of lacking chromosome 13 (e.g. lack of *pfhrp3* and decreased RDT-based treatment when *pfhrp2* deletion is also present), or no benefit of a duplicated chromosome 13. The fact that there was not a single clinical isolate detected with the reciprocal hybrid 11-13 pattern seen in areas with large amounts of hybrid chromosome 13-11 suggests that evolutionary pressures are strong enough to prevent its emergence in the natural parasite population.

There are likely additional biological effects of the 13-11 translocation in low transmission areas beyond those of RDT detection, HRP2, and HRP3. In this chromosome 13-11 translocation, chromosome 13 genes are lost and chromosome 11 genes are gained. Not only do these regions contain members of multi-gene families of *rifins, stevors,* and *var* but also *pf332* on chromosome 11, which encodes a large Duffy-binding-like domain containing a 27KDa protein (46–48). *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 (49). Lack of this protein likely leads to a large detriment to the survival of the parasite and may be the reason the reciprocal hybrid 11-13 is not observed. Other genes in this area include two ring-infected erythrocyte surface antigen (RESA) genes; there is only one other gene of this family, which is located on chromosome 1. Little is known about these two genes. There is also a *DnaJ* gene within this duplicated region. There are about 30 genes in this family though this sequence is only 67% similar to the closest related DnaJ protein.

While Hinterberg *et al.* proposed that a general mechanism of non-homologous recombination of the subtelomeric regions may be responsible for the deletion of *pfhrp3*, our findings indicate that this event involved 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 (14). The duplicated region of chromosome 11 has significant diversity between parasites within the duplication, indicating that this duplication event has either been repeated several times or has an ancient origin. Pattern 13^{-} involves a segmental deletion of chromosome 13 including *pfhrp3* without duplication of chromosome 11. Interestingly, we observed a distinct geographical *pfhrp3* deletion pattern in our analysis. Pattern $13^{-}11^{++}$ is seen only in the Americas and Africa and Pattern 13^{-} in South East Asia. It is not clear why a mechanistic deletion pattern would be geographically distributed in such a way.

Our findings indicate that the deletion of *pfhrp3* in Africa and the Americas is caused by a translocation driven by duplication-mediated NAHR leading to segmental duplication of chromosome 11 and deletion of chromosome 13. While further studies are needed to determine the reasons for these geographical patterns, our results provide an improved understanding of the mechanism of structural variation underlying *pfhrp3* deletion. They also suggest constraints against emergence in high-transmission regions 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 may explain why *pfhrp3* loss precedes *pfhrp2* loss, despite RDT pressure presumably exerting stronger survival advantage with loss of *pfhrp2* vs *pfhrp3*. Future studies involving a larger number of strains with *pfhrp2* deletion are needed to investigate its mechanism and the interactions between the deletions of *pfhrp2* and *pfhrp3,* as well as to examine the selective pressures behind these patterns.

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 (50), then taking 200bp windows, stepping every 100bp between these tandem repeats, and using LASTZ (51) to align these regions against currently available chromosomal-level PacBio assembled genomes (29) 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 and/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 rRNAs 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 (11) where *pfhrp2-/3-* parasites are common. The called haplotypes were compared to determine which subregions contained variation in order to type each chromosome (**Figure S22**). Coverage was determined for each genomic region by normalizing based on the coverage of the whole genome. Windows were examined on chromosomes 8, 11, and 13 starting from 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 relative to *P. falciparum* 3D7 genome version 3 (version=2015-06-18) **(Figure S22).**

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 (39). The genomes of all *Laverania* have recently been sequenced and assembled using PacBio and Illumina data (39). The assemblies were analyzed using their annotations and by using LASTZ (51) 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) (29). 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 Nanopore's MinION device (52–54). 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 (55), an assembly pipeline for high-noise single-molecule sequencing, and Flye (56) 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 (57). Mappings were visualized using custom R scripts (58–60).

AVAILABILITY OF DATA AND MATERIALS

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

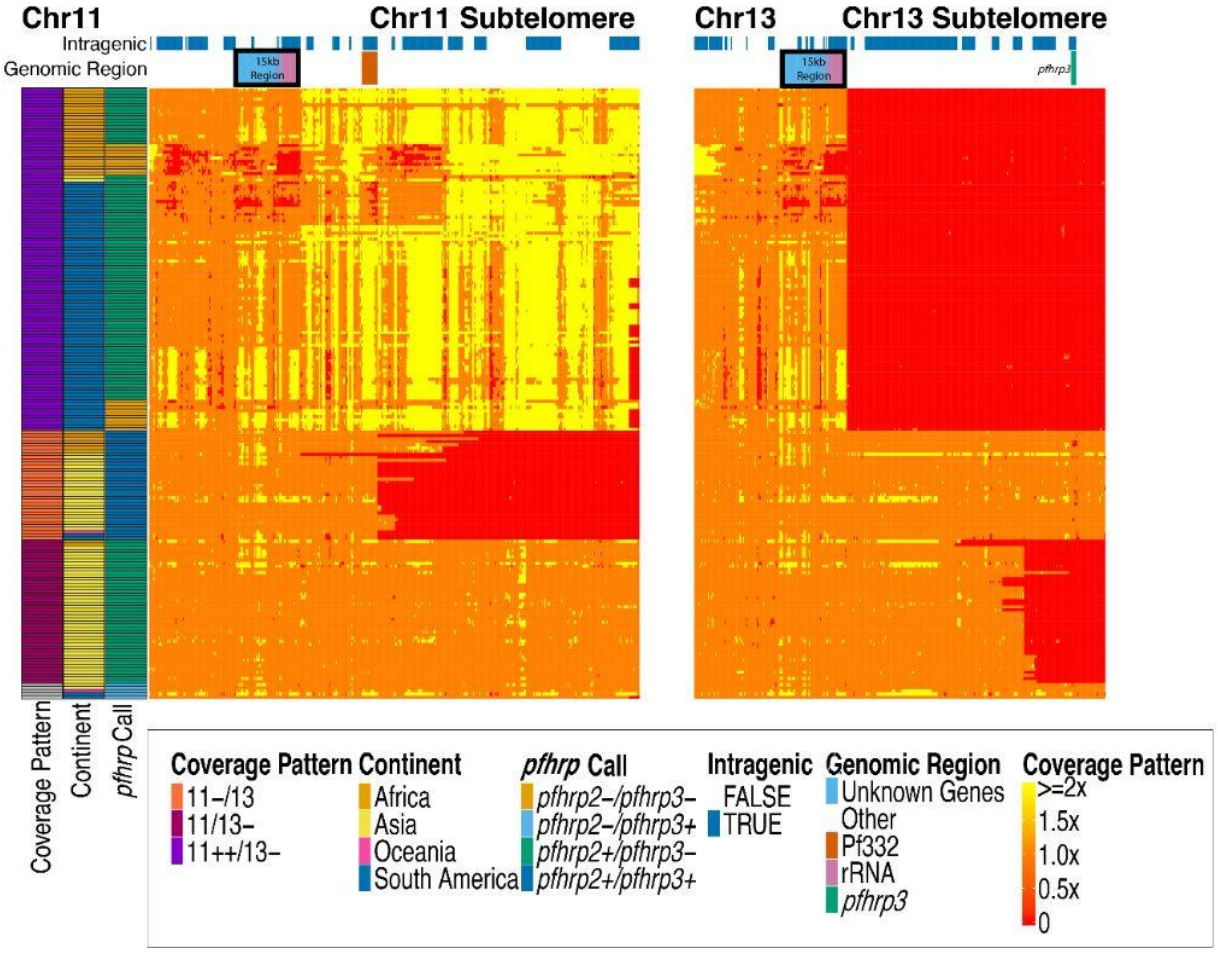

Figure 1. Parasites with altered sequence coverage across chromosomes 11 and 13 subtelomeres. Sequence coverage heatmap showing the sub-telomeric regions of chromosomes 11 and 13 in the subset of the 8,765 samples along with key lab isolates showing evidence of deletion of either *pfhrp2*, *pfhrp3*, and/or chromosome 11. Each row is a sample. The top annotation along chromosomes depicts the location of genes, and the second row delineates the duplicated region (light blue), the rRNA region within the duplicated region (light purple), *pf332* (red-orange), rRNA (purple), and the *pfhrp3* gene (green). The left sample annotation includes the deletion pattern, continent of origin, and *pfhrp2/3* deletion calls,3. 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*.

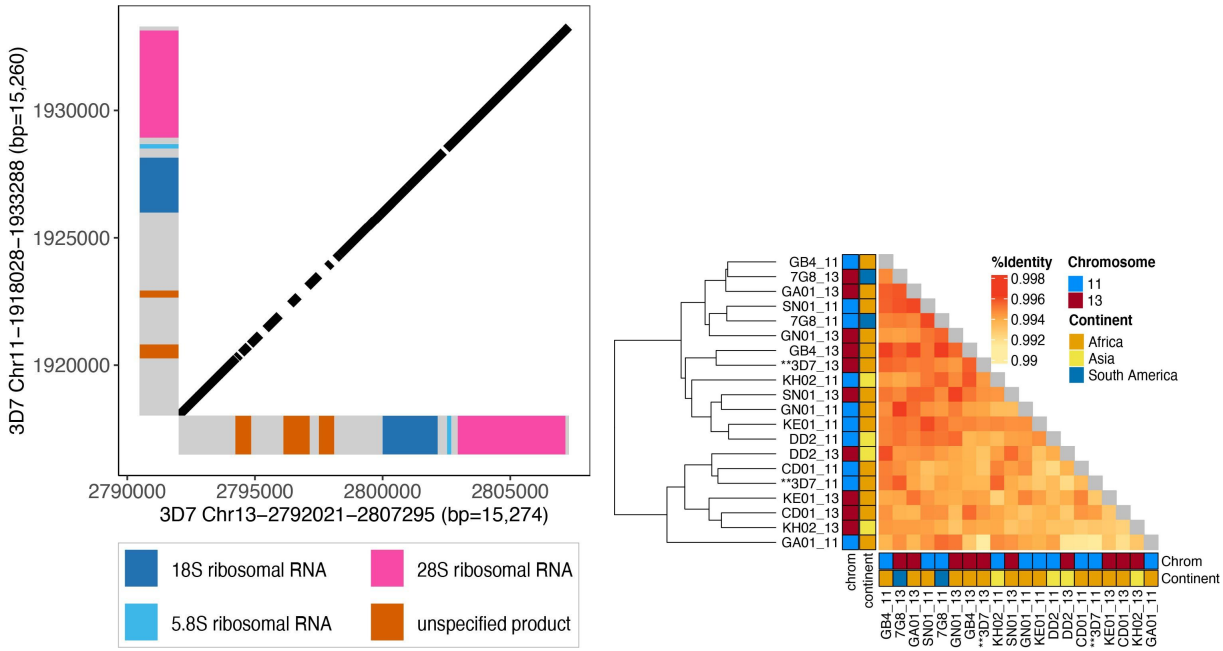

Figure 2. Characterization of the 15.2 kb segmental duplication containing ribosomal genes on chromosomes 11 and 13. (a) Shows the alignment of the 3D7 reference genome copies with diagonal black bars showing 100% conserved regions (>30 bp). Perfectly conserved segments comprise 89.1% of the whole region with an overall identity of 99.3% by direct alignment comparison. 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.

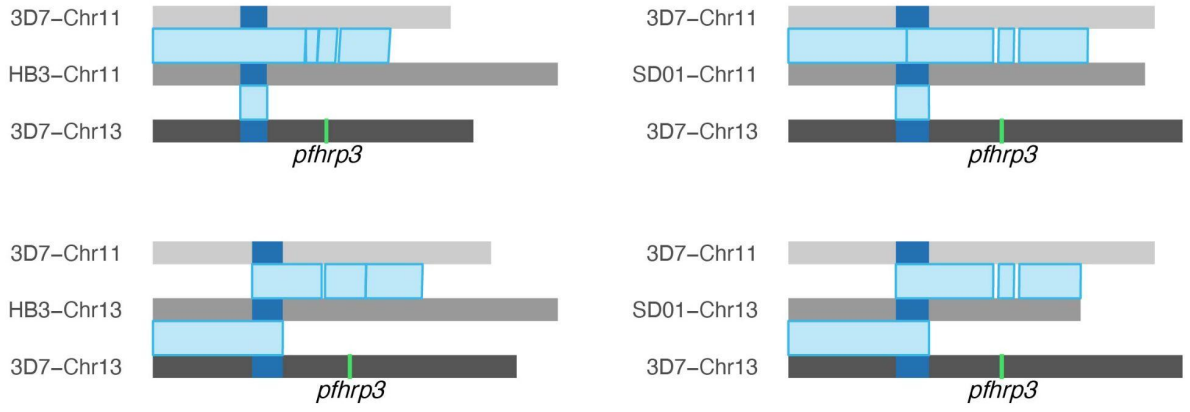

Figure 3. Comparison of long-read assemblies of chromosomes 11 and 13 of HB3 and SD01 to the reference genome 3D7 confirms hybridized chromosome 13-11. Chromosome 11 of HB3 and SD01 mapped entirely to the reference chromosome 11 of 3D7 except for the segmental duplication which is shared between chromosomes 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* should be found), it maps to chromosome 11 of 3D7 instead of chromosome 13.

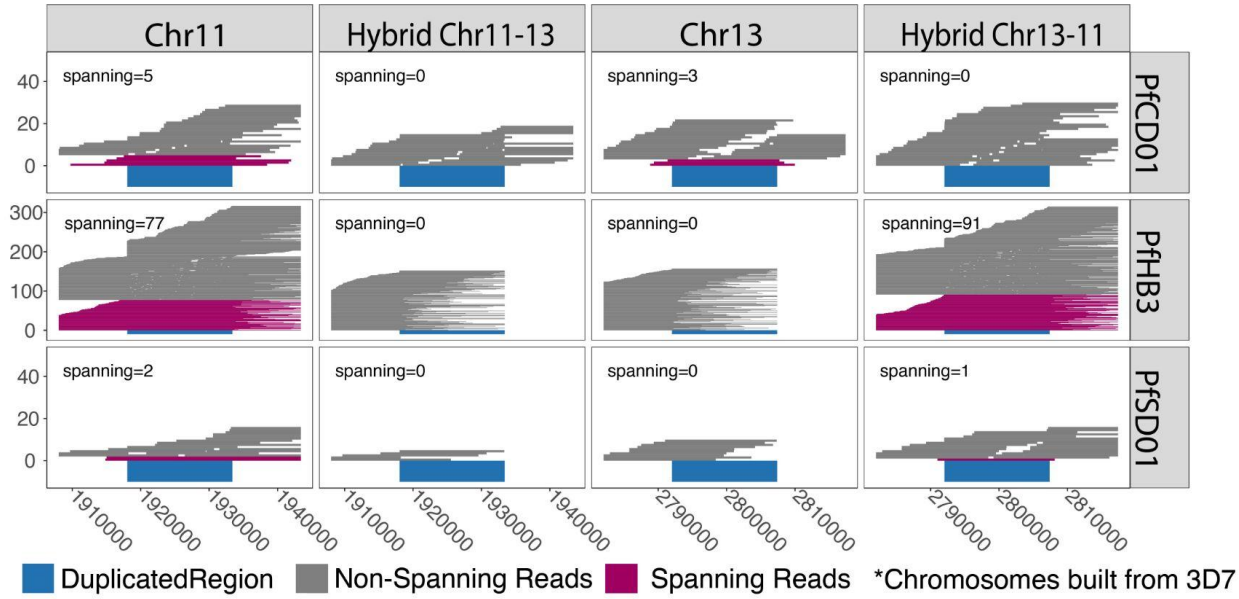

Figure 4. Long reads spanning the 15kb duplicated region confirm presence of translocated chr 13-11 in *pfhrp3*-deleted HB3 (Americas) and SD01 (Africa) but not *pfhrp3*-intact CD01 strains. The best mapping to the segmental duplication region of PacBio and Nanopore reads >15kb for HB3, SD01, and CD01 region are shown against normal chromosome 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.

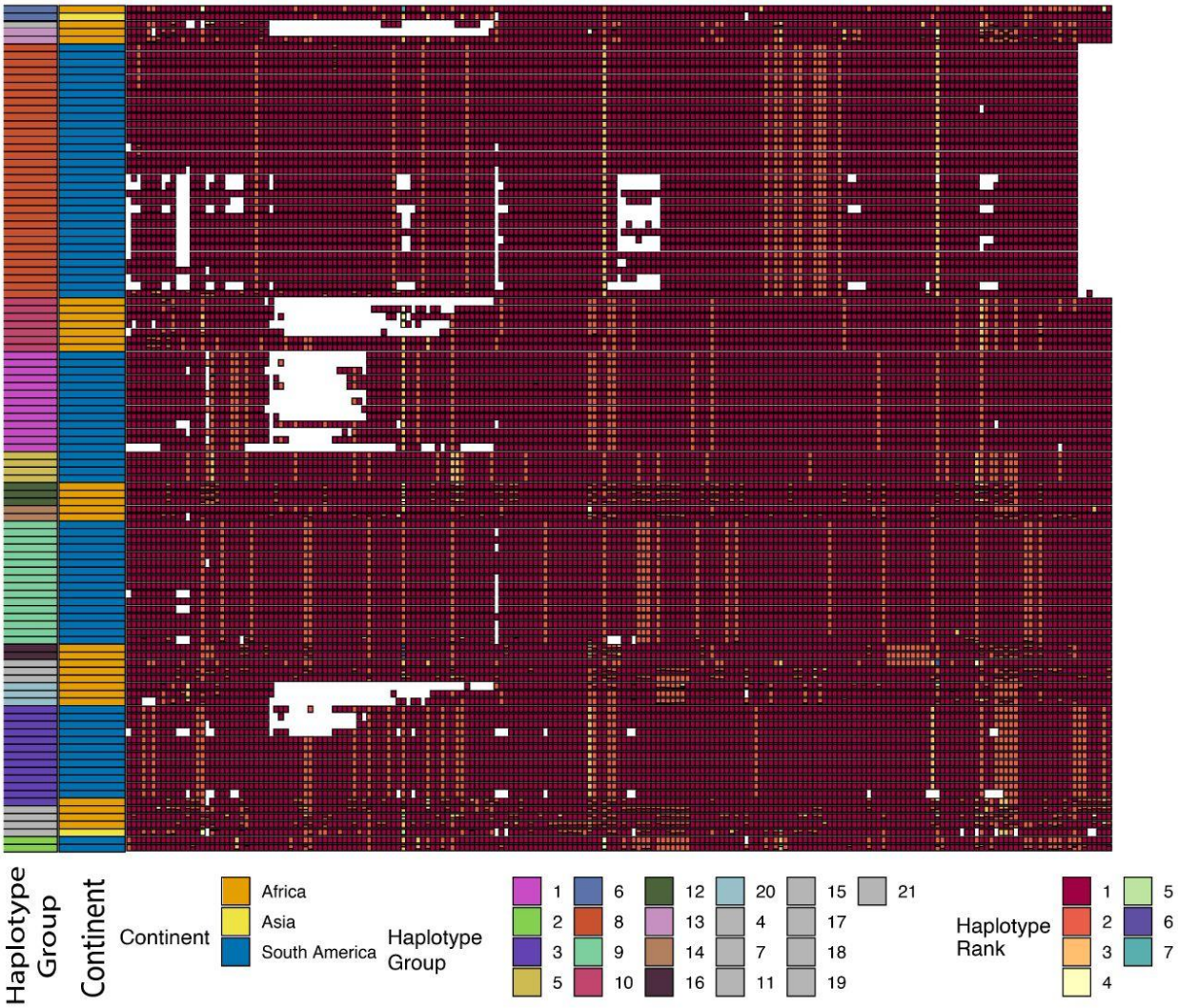

Figure 5. Microhaplotype patterns for duplicated portion of chromosome 11 in 11++13- parasites indicate 5 major haplotype groups, largely separated by geography with distinct haplotypes observed in American and African strains. Most strains show singular haplotypes at variant positions despite two copies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

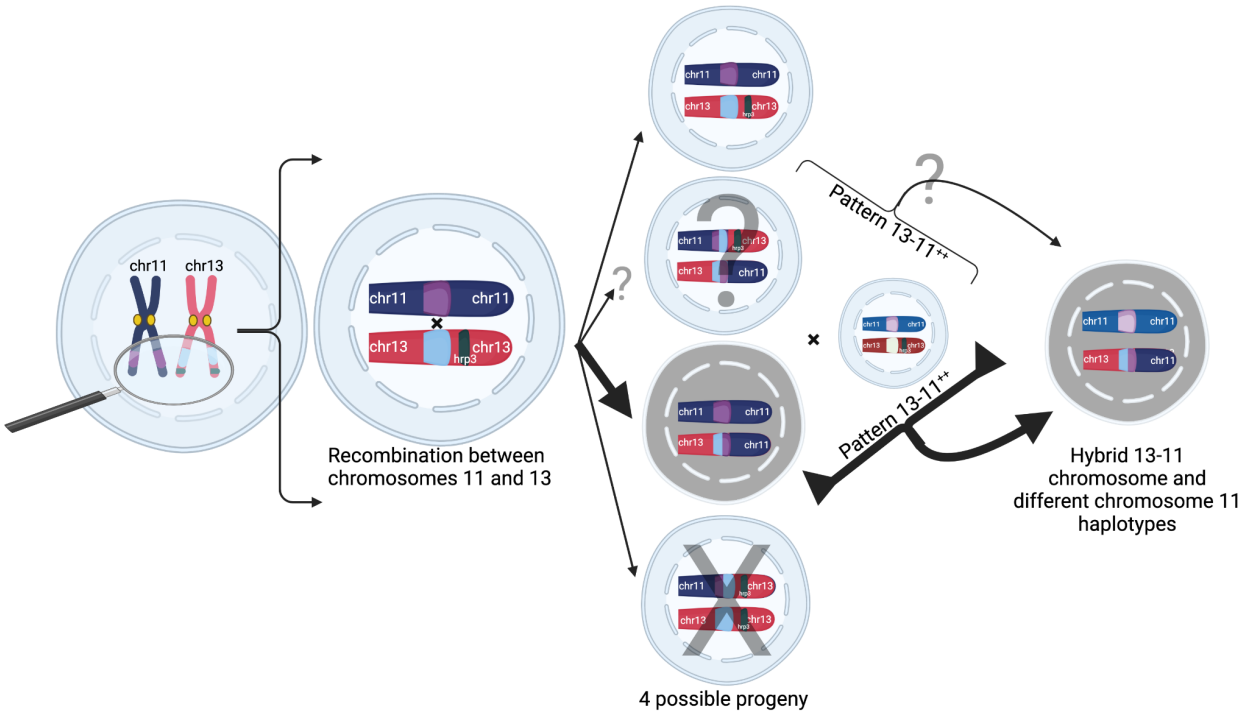

Figure 6. Proposed model of duplication-mediated non-allelic homologous recombination driving translocation during intrastrain meiotic recombination. Homology misalignment and recombination between chromosomes 11 and 13 first occurs in an oocyst formed from identical parasites (intrastrain), which can 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 majority of samples; deletion pattern 13-11^{++}). Subsequent recombination with an unrelated strain yields unrelated chromosome11 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

1. Malaria (January 13, 2022).

2. World malaria report 2022 (January 27, 2023).

3. G. Benelli, J. C. Beier, Current vector control challenges in the fight against malaria. *Acta Trop.* **174**, 91–96 (2017).

4. 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 (August 11, 2021).

5. P. L. Chiodini, Malaria diagnostics: now and the future. *Parasitology* **141**, 1873–1879 (2014).

6. K. E. Poti, D. J. Sullivan, A. M. Dondorp, C. J. Woodrow, HRP2: Transforming Malaria Diagnosis, but with Caveats. *Trends Parasitol.* **36**, 112–126 (2020).

7. J. Baker, *et al.*, 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 (2005).

8. W. H. Organization, Others, Good practices for selecting and procuring rapid diagnostic tests for malaria (2011).

9. Q. Cheng, *et al.*, Plasmodium falciparum parasites lacking histidine-rich protein 2 and 3: a review and recommendations for accurate reporting. *Malar. J.* **13**, 283 (2014).

10. R. Thomson, *et al.*, Prevalence of Plasmodium falciparum lacking histidine-rich proteins 2 and 3: a systematic review. *Bull. World Health Organ.* **98**, 558–568F (2020).

11. S. M. Feleke, *et al.*, Plasmodium falciparum is evolving to escape malaria rapid diagnostic tests in Ethiopia. *Nat Microbiol* **6**, 1289–1299 (2021).

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

13. A. Berhane, *et al.*, Major Threat to Malaria Control Programs by Plasmodium falciparum Lacking Histidine-Rich Protein 2, Eritrea. *Emerg. Infect. Dis.* **24**, 462–470 (2018).

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

15. M. J. Gardner, *et al.*, Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature* **419**, 498–511 (2002).

16. K. D. Vernick, T. F. McCutchan, Sequence and structure of a Plasmodium falciparum telomere. *Mol. Biochem. Parasitol.* **28**, 85–94 (1988).

17. D. F. Conrad, *et al.*, Origins and functional impact of copy number variation in the human genome. *Nature* **464**, 704–712 (2010).

18. J. O. Korbel, *et al.*, Paired-end mapping reveals extensive structural variation in the human genome. *Science* **318**, 420–426 (2007).

19. J. M. Kidd, *et al.*, A human genome structural variation sequencing resource reveals insights into mutational mechanisms. *Cell* **143**, 837–847 (2010).

20. T. 1000 G. P. Consortium, The 1000 Genomes Project Consortium, An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).

21. R. E. Mills, *et al.*, Mapping copy number variation by population-scale genome sequencing. *Nature* **470**, 59–65 (2011).

22. M. M. Parks, C. E. Lawrence, B. J. Raphael, Detecting non-allelic homologous recombination from high-throughput sequencing data. *Genome Biol.* **16**, 72 (2015).

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

24. P. Liu, *et al.*, 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 (2011).

25. J. A. Bailey, E. E. Eichler, Primate segmental duplications: crucibles of evolution, diversity and disease. *Nat. Rev. Genet.* **7**, 552–564 (2006).

26. S. E. R. Bopp, *et al.*, Mitotic evolution of Plasmodium falciparum shows a stable core genome but recombination in antigen families. *PLoS Genet.* **9**, e1003293 (2013).

27. E. Bruske, T. D. Otto, M. Frank, 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 (2018).

28. A. Miles, *et al.*, Indels, structural variation, and recombination drive genomic diversity in Plasmodium falciparum. *Genome Res.* **26**, 1288–1299 (2016).

29. T. D. Otto, *et al.*, Long read assemblies of geographically dispersed Plasmodium falciparum isolates reveal highly structured subtelomeres. *Wellcome Open Res* **3**, 52 (2018).

30. K. P. Day, *et al.*, 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 (2017).

31. T. S. Rask, D. A. Hansen, T. G. Theander, A. Gorm Pedersen, T. Lavstsen, Plasmodium falciparum erythrocyte membrane protein 1 diversity in seven genomes--divide and conquer. *PLoS Comput. Biol.* **6** (2010).

32. A. E. Barry, *et al.*, Population genomics of the immune evasion (var) genes of Plasmodium falciparum. *PLoS Pathog.* **3**, e34 (2007).

33. D. S. Chen, *et al.*, A molecular epidemiological study of var gene diversity to characterize the reservoir of Plasmodium falciparum in humans in Africa. *PLoS One* **6**, e16629 (2011).

34. S. K. Tessema, *et al.*, Phylogeography of var gene repertoires reveals fine-scale geospatial

clustering of Plasmodium falciparum populations in a highly endemic area. *Mol. Ecol.* **24**, 484–497 (2015).

35. A. F. Sander, *et al.*, Multiple var2csa-type PfEMP1 genes located at different chromosomal loci occur in many Plasmodium falciparum isolates. *PLoS One* **4**, e6667 (2009).

36. L. A. Kirkman, E. A. Lawrence, K. W. Deitsch, Malaria parasites utilize both homologous recombination and alternative end joining pathways to maintain genome integrity. *Nucleic Acids Res.* **42**, 370–379 (2014).

37. S. Kurtz, *et al.*, Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12 (2004).

38. S. Steinbiss, *et al.*, Companion: a web server for annotation and analysis of parasite genomes. *Nucleic Acids Res.* **44**, W29–34 (2016).

39. T. D. Otto, *et al.*, Genomes of all known members of a Plasmodium subgenus reveal paths to virulent human malaria. *Nat Microbiol* **3**, 687–697 (2018).

40. L. Vergés, Ò. Molina, E. Geán, F. Vidal, J. Blanco, Deletions and duplications of the 22q11.2 region in spermatozoa from DiGeorge/velocardiofacial fathers. *Molecular Cytogenetics* **7** (2014).

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

42. I. H. Cheeseman, *et al.*, Population Structure Shapes Copy Number Variation in Malaria Parasites. *Mol. Biol. Evol.* **33**, 603–620 (2016).

43. A. R. Taylor, D. F. Echeverry, T. J. C. Anderson, D. E. Neafsey, C. O. Buckee, Identity-by-descent with uncertainty characterises connectivity of Plasmodium falciparum populations on the Colombian-Pacific coast. *PLoS Genet.* **16**, e1009101 (2020).

44. S. M. Griffing, *et al.*, South American Plasmodium falciparum after the malaria eradication era: clonal population expansion and survival of the fittest hybrids. *PLoS One* **6**, e23486 (2011).

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

46. S. Nilsson, *et al.*, Characterization of the Duffy-Binding-Like Domain of Plasmodium falciparum Blood-Stage Antigen 332. *Malar. Res. Treat.* **2011**, 671439 (2011).

47. M. Rug, *et al.*, Export of virulence proteins by malaria-infected erythrocytes involves remodeling of host actin cytoskeleton. *Blood* **124**, 3459–3468 (2014).

48. A. N. Hodder, *et al.*, Analysis of structure and function of the giant protein Pf332 in Plasmodium falciparum. *Mol. Microbiol.* **71**, 48–65 (2009).

49. G. I. Bertin, *et al.*, Proteomic analysis of Plasmodium falciparum parasites from patients with cerebral and uncomplicated malaria. *Sci. Rep.* **6**, 26773 (2016).

50. G. Benson, Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids*

Res. **27**, 573–580 (1999).

51. R. S. Harris, “Improved pairwise alignment of genomic DNA,” The Pennsylvania State University. (2007).

52. J. J. Kasianowicz, E. Brandin, D. Branton, D. W. Deamer, Characterization of individual polynucleotide molecules using a membrane channel. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 13770–13773 (1996).

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

54. D. Branton, *et al.*, The potential and challenges of nanopore sequencing. *Nat. Biotechnol.* **26**, 1146–1153 (2008).

55. S. Koren, *et al.*, Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).

56. M. Kolmogorov, J. Yuan, Y. Lin, P. A. Pevzner, Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019).

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

58. R Core Team, R: A Language and Environment for Statistical Computing (2022).

59. H. Wickham, *et al.*, “ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics” (2022).

60. Z. Gu, “ComplexHeatmap: Make Complex Heatmaps” (2022).