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
Emily Y. Liang: emily_liang@brown.edu
Christian P. Nixon: christian_nixon@brown.edu
David Giesbrecht: david_giesbrecht@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 *Plasmodium falciparum* histidine-rich protein 2 (PfHRP2) and PfHRP3, but deletions of *pfhrp2* and *phfrp3* genes make parasites undetectable by RDTs. To better understand these deletions, we analyzed 19,289 public whole-genome-sequenced *P. falciparum* field samples. *Pfhrp2* deletion only occurred by chromosomal breakage with subsequent telomere healing. *Pfhrp3* deletions involved loss from *pfhrp3* to the telomere and showed 3 patterns: no other associated rearrangement with evidence of telomere healing at breakpoint (Asia; Pattern 13^{-}); associated with duplication of a chromosome 5 segment containing multidrug-resistant-1 gene (Asia; Pattern $13^{-}5^{++}$); and most commonly, associated with duplication of a chromosome 11 segment (Americas/Africa; Pattern $13^{-}11^{++}$). We confirmed a 13-11 hybrid chromosome with long-read sequencing, consistent with a translocation product arising from recombination between large interchromosomal ribosome-containing segmental duplications. Within most $13^{-}11^{++}$ parasites, the duplicated chromosome 11 segments were identical to each other. Across parasites, multiple distinct haplotype groupings were consistent with emergence due to clonal expansion of progeny from intrastrain meiotic recombination. Together, these observations suggest negative selection normally removes $13^{-}11^{++}$ *pfhrp3 deletions*, and specific conditions are needed for their emergence and spread including low transmission, findings that can help refine surveillance strategies.

INTRODUCTION

P. falciparum malaria remains a leading cause of childhood mortality in Africa

Plasmodium falciparum remains one of the most common causes of malaria and 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[1].

The mainstay of malaria diagnosis across Africa is no longer microscopy but rapid diagnostic tests (RDTs) due to their simplicity and speed. Their swift adoption, now totaling hundreds of millions a year, coupled with effective artemisinin-based combination therapies (ACTs) has led to significant progress in malaria control[2,3]. The predominant and most sensitive falciparum malaria RDTs detect *P. falciparum* histidine-rich protein 2 (PfHRP2) and, to a lesser extent, its paralog PfHRP3 due to cross-reactivity.

Increasing numbers of pfhrp2 and pfhrp3-deleted parasites escaping diagnosis by RDTs

Unfortunately, a growing number of studies have reported laboratory and field isolates with deletions of *pfhrp2* (PF3D7_0831800) and *pfhrp3* (PF3D7_1372200) in the subtelomeric regions of chromosomes 8 and 13, respectively. The resulting lack of these proteins allows the parasite to fully escape diagnosis by PfHRP2-based RDTs[3–7]. Deleted parasites appear to be spreading rapidly in some regions and have compromised existing test-and-treat programs, especially in the Horn of Africa[8–11]. The prevalence of parasites with *pfhrp2* and *pfhrp3* deletion varies markedly across continents and regions 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[7]. Studies in Ethiopia, where false-negative RDTs owing to *pfhrp2 and pfhrp3* deletions are common, suggest that the *pfhrp3* deletion arose first given it is more prevalent and shows a shorter shared haplotype[9]. The reason why *pfhrp3* deletion occurred prior to *pfhrp2* remains unclear. A 1994 study of the HB3

laboratory strain reported frequent meiotic translocation of a *pfhrp3* deletion from chromosomes 13 to 11[12]. Explanation of this mechanism, whether it might occur in natural populations, and how it relates to initial loss of *pfhrp3* has not been fully explored.

Precise pfhfp2 and pfhrp3 deletion mechanisms remain unknown

Studies of *P. falciparum* structural rearrangements are challenging and *pfhrp2* and *pfhrp3* deletions particularly so due to their position in complex subtelomeric regions. 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 non-allelic homologous recombination (NAHR) and double-stranded breakage (DSB) and telomere healing[13–19]. Subtelomeric exchange importantly allows for unbalanced progeny without the usual deleterious ramifications of altering a larger proportion of a chromosome. Of note, newly formed duplications predispose to further duplications or other rearrangements through NAHR between highly identical paralogous regions. Together, this potentiates the rapid expansion of gene families and their spread across subtelomeric regions[20–25]. The duplicative transposition of a subtelomeric region of one chromosome onto another chromosome frequently occurs in *P. falciparum*. Specifically, prior studies have found duplicative transposition events involving several genes including *var2csa* and *cytochrome b*[17,26–29]. Notably, *pfhrp2* and *pfhrp3* are adjacent to but not considered part of the subtelomeric regions, and recombination of *var* genes does not result in the deletion of *pfhrp2* and *pfhrp3*[9,30].

Telomere healing, de novo telomere addition via telomerase activity, is associated with subtelomeric deletion events in *P. falciparum* that involve chromosomal breakage and loss of all downstream genes. Healing serves to stabilize the end of the chromosome. 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[31]. Since then, studies have defined the critical role of telomerase in *P. falciparum* and additional occurrences affecting a number of genes including *pfhrp1, Pf332,* and *Pf87* in laboratory isolates[15,32,33]. For *pfhrp1* and *pfhrp2,* this

mechanism of deletion only occurred in laboratory isolates but not in clinical samples, suggesting the genes have important infections in normal infections and their loss is selected against[32].

An improved understanding of the patterns and mechanisms of *pfhrp2* and *pfhrp3* deletions can provide important insights into how frequently they occur and the evolutionary pressures driving their emergence and help inform control strategies. Here, using available whole-genome sequences and additional long-read sequencing, we examined the pattern and nature of *pfhrp2* and *pfhrp3* deletions. Our findings shed light on geographical differences in *pfhrp3* deletion patterns, their mechanisms, and how they likely emerged, providing key information 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 global *P. falciparum* isolates as of January 2023, comprising 19,289 field samples and lab isolates (**Table S1**). We analyzed the genomic regions containing *pfhrp2* on chromosome 8 and *pfhrp3* on chromosome 13 to detect nucleotide and copy number variation (e.g., deletions and duplications) using local haplotype assembly and sequencing depth. Regions on chromosomes 5 and 11 associated with these duplicates were also analyzed (**Table S2, Table S4, Methods)**. We identified 27 parasites with *pfhrp2* deletion, 172 with *pfhrp3* deletion, and 21 with both *pfhrp2* and *pfhrp3* deletions. Across all regions, *pfhrp3* deletions were more common than *pfhrp2* deletions; specifically, *pfhrp3* deletions and *pfhrp2* deletions were present in Africa in 43 and 12, Asia in 53 and 4, and South America in 76 and 11 parasites. It should be noted that these numbers are not accurate measures of prevalence given that most WGS specimens have been collected based on RDT positivity.

Pfhrp2 deletion associated with variable breakpoints and telomeric healing

We further examined the breakpoints of 27 parasites (25 patient parasites and 2 lab isolates) (**Supplemental Figure 1**). Twelve parasites showed evidence of breakage and telomeric healing as suggested by telomere-associated tandem repeat 1 (TARE-1)[14] sequence contiguous with the genomic sequence at locations where coverage drops to zero (**Supplemental Figure 2**). The majority of breakpoints occur within *pfhrp2*, found in 9 South American parasites and lab isolate D10 (**Supplemental Figure 2**). The other *pfhrp2*-deleted parasites did not have detectable TARE1 or evidence of genomic rearrangement but were amplified with sWGA, limiting the ability to detect the TARE1 sequence. Thus, *pfhrp2* deletion likely occurs solely through breakage events with subsequent telomeric healing.

Three distinct pfhrp3 deletion patterns with geographical associations

Exploration of read depth revealed three distinct deletion copy number patterns associated with *pfhrp3* deletion (chromosome 13: 2,840,236-2,842,840): first, sole deletion of chromosome 13 starting at various locations centromeric to *pfhrp3* to the end of the chromosome with detectable TARE1 telomere healing and unassociated with other rearrangements (**pattern 13^{-}**); second, deletion of chromosome 13 from position 2,835,587 to the end of the chromosome and associated with duplication of a chromosome 5 segment from position 952,668 to 979,203, which includes *pfmdr1* (**pattern $13^{-}5^{++}$**); and third, deletion of chromosome 13 commencing just centromeric to *pfhrp3* and extending to the end of the chromosome and associated with duplication of the chromosome 11 subtelomeric region (**pattern $13^{-}11^{++}$**) (**Figure 1**). Among the 172 parasites with *pfhrp3* deletion, 21 (12.2%) were pattern 13^{-}, 29 (16.9%) were pattern $13^{-}5^{++}$, and the majority with 122 (70.9%) demonstrated pattern $13^{-}11^{++}$. Pattern $13^{-}11^{++}$ was almost exclusively found in parasites from Africa and the Americas, while $13^{-}5^{++}$ was only observed in Asia (**Figure 1**).

Pattern 13^{-} associated with telomere healing

The 21 parasites with pattern 13^{-} had deletions of the core genome averaging 19kb (range: 11-31kb). Of these 13^{-} deletions, 20 out of 21 had detectable TARE1 adjacent to the breakpoint, consistent with telomere healing (**Supplemental Figure 3**).

Pattern 13^{-}5^{++} associated with NAHR-mediated pfmdr1 duplication and subsequent telomere healing

The 29 parasites with deletion pattern 13^{-}5^{++} had a consistent loss of 17.9kb of chromosome 13 and a gain of 25kb from chromosome 5. 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. Analysis revealed paired-end reads with discordant mapping with one read mapping to chromosome 13 and the other mapping to chromosome 5. Reads assembled from these regions form a contig of normally unique sequence that connects chromosome 13 (position 2,835,587) to chromosome 5 (position 979,203) in reverse orientation. Read depth coverage analysis revealed more than a two-fold increase on chromosome 5 from 979,203 to 952,668 with TARE1 sequence contiguously extending from 952,668, consistent with telomere healing. This 25kb duplication contained several genes including intact PF3D7_0523000 (*pfmdr1*) (**Supplemental Figure 4 and Supplemental Figure 5**), and TARE1 transition occurred within the gene PF3D7_0522900 (a zinc finger gene). Further read depth, discordant read, and assembly analysis revealed four 13^{-}5^{++} parasites that, in addition to the chromosome 5 segment duplication on chromosome 13, had the described *pfmdr1* tandem duplication on chromosome 5 associated with drug resistance, resulting in overall 3-fold read depth across *pfmdr1 gene* (**Supplemental Figure 4**)[34].

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

Pattern $13^{-}11^{++}$ was observed in 74 American parasites, 39 African parasites, and 6 Asian parasites (**Figure 1**). Of the 122 parasites with this *pfhrp3* deletion pattern, 98 parasites (73 of the 74 American, 20 of 40 African parasites, and 5 of 6 Asian) had near-identical copies of the chromosome 11 duplicated region. Near-identical copies were defined as having ≥99% identity (same variant microhaplotype between copies) across 382 variant microhaplotypes within the duplicated region far less than normal between parasite allelic differences (**Table S4**). These 98 parasites containing identical copies did not all share the same overall haplotypes but rather showed 11 major haplotype groups (**Figure 5)**. The remaining 24 parasites had variation within this region; on average, 10.2% of variant sites differed between the copies (min 83.8% identity). The overall 11 haplotype groups showed geographical separation with distinct haplotypes observed in American and African strains. The overall haplotypes for the segment of chromosome 11 found within $13^{-}11^{++}$ parasites could also be found within the parasites lacking the $13^{-}11^{++}$ translocation (**Supplemental Figure 6**, **Supplemental Figure 7**, **Supplemental Figure 8**, **Supplemental Figure 9**, **Supplemental Figure 10**, **and Supplemental Figure 11**).

Pattern $13^{-}11^{++}$ breakpoint occurs in a segmental duplication of ribosomal genes on chromosomes 11 and 13

Pattern $13^{-}11^{++}$ has a centromeric breakpoint consistently occurring within a 15kb interchromosomal segmental duplication on chromosome 11 and 13. It was the largest duplication in the core genome based on an all-by-all unique k-mer comparison of the genome using nucmer[35] **(Supplemental Figure 12)**. The two copies on chromosome 11 and chromosome 13 in the reference genome were 99.0% identical (**Figure 2**) and oriented similarly. Each copy contained a centromeric 7kb region encoding 2-3 undefined protein products (98.9%

identity) and a telomeric 8kb nearly identical region (99.7% identity) containing one of the two S-type[13] ribosomal genes (S=sporozoite), which are primary expressed during life cycle stages in mosquito vector (**Figure 2**, **Supplemental Figure 13 and Supplemental Figure 14).** Pairwise alignment of the chromosomes 11 and 13 paralogs showed similar levels of allelic and paralogous identity and no consistent nucleotide differences were found between the paralogs leading to no distinct separation between copies when clustered **(Figure 2**). This suggests ongoing interchromosomal exchanges or conversion events maintaining paralog homogeneity.

Ribosomal gene segmental duplication exists in closely related P. praefalciparum

To look at the conservation of the segmental duplication containing the ribosomal genes, we examined genomes of closely related *Plasmodium* parasites in the *Laverania* subgenus, which comprised *P. falciparum* and other *Plasmodium* spp found in African apes. The *Plasmodium praefalciparum* genome*,* which is *P. falciparum's* closest relative having diverged about 50,000 years ago[36], also contained similar S-type rRNA loci on chromosomes 11 and 13 and had a similar gene synteny to *P. falciparum* in these regions and the region on chromosome 8 neighboring *pfhrp2* (**Supplemental Figure 15**, **Supplemental Figure 16**, **and Supplemental Figure 17**). *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[36] were not fully assembled within their subtelomeric regions.

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

Given pattern $13^{-}11^{++}$ was suggestive of duplication-mediated recombination leading to translocation, we examined high-quality PacBio genome assemblies of SD01 from Sudan and HB3 from Honduras, both containing the *pfhrp3* deletion. However, the Companion[37] gene annotations of chromosome 11[30] showed that these strains were not fully assembled in the relevant regions (**Supplemental Figure 18 and Supplemental Figure 19**).

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

To better examine the genome structure of pattern $13^{-}11^{++}$, we whole-genome sequenced the $13^{-}11^{++}$ isolates HB3 and SD01 with long-read Nanopore technology. We generated 7,350 Mb and 645 Mb of data representing an average coverage of 319x and 29.3x for HB3 and SD01, respectively. We combined our Nanopore data with the publicly available PacBio sequencing data and tested for the presence of hybrid chromosomes using a two-pronged approach: 1) mapping the long reads directly to normal and hybrid chromosome 11/13 constructs and 2) optimized *de-novo* assembly of the higher quality Nanopore long reads.

To directly map reads, 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 aligned all PacBio and Nanopore reads for each isolate to the normal and hybrid constructs to detect reads completely spanning the duplicated region extending at least 50bp into flanking unique regions (**Figure 4**). HB3 had 77 spanning reads across normal chromosome 11 and 91 spanning reads across hybrid chromosome 13-11. SD01 had two chromosome 11 spanning reads and one 13-11 chromosome spanning read. SD01 had a small number of spanning reads due to lower overall Nanopore reads secondary to insufficient input sample. Further analysis on SD01 revealed 4 regions within this duplicated region that had chromosome 11 and 13-specific nucleotide variation, which was leveraged to further bridge across this region for additional confirmation given SD01's low coverage (**Supplemental Figure 20** and **Supplemental Figure 21**). Neither isolate had long-reads spanning normal chromosome 13 or hybrid 11-13, which represented the reciprocal translocation product (**Figure 4**). Importantly, the other 12 isolates with intact *pfhrp3* from the PacBio dataset[30] all had reads consistent with normal chromosomes -- reads spanning chromosome 11 and chromosome 13 and no reads spanning the hybrid 13-11 or 11-13

chromosomes (**Figure 4**). Thus, long reads for HB3 and SD01 confirmed the presence of a hybrid 13-11 chromosome.

De novo long-read assemblies of pfhrp3-deleted strains further confirmed hybrid 13-11 chromosome

To further examine the parasites with hybrid 13-11 chromosomes and exclude potentially 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 (N50 1,5985,898 and L50 5). TARE-1[14] 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 telomere-to-telomere coverage had been achieved. SD01, however, with lower sequencing coverage, had a more disjointed assembly with 200 final contigs (N50 263,459, and L50 30). The HB3 and SD01 assemblies both had a chromosome 11 that closely matched normal 3D7 chromosome 11 and a separate hybrid 13-11 that closely matched 3D7 chromosome 13 until the ribosomal duplication region where it then subsequently best-matched chromosome 11 (**Figure 3**, **Supplemental Figure 22**). HB3's 11 and hybrid 13-11 chromosomes had TARE-1 at their ends[14], indicating that these chromosomes were complete. These new assemblies were further annotated for genes by Companion[37]. The contig matching the hybrid 13-11 for both strains essentially contained a duplicated portion of chromosome 11 telomeric to the ribosomal duplication. The duplicated genes within this segment 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 ended with a DnaJ gene (PF3D7_1149600) corresponding to 3D7 genes PF3D7_1148700 through PF3D7_1149600 (**Supplemental Figure 23 and Supplemental Figure 24**). Homology between HB3 chromosomes 11 and 13-11 continued up through a *rifin*, then a *stevor* gene, and then the sequence completely diverged in the most

telomeric region with a different gene family organization structure but both consisting of *stevor*, *rifin,* and *var* gene families along with other paralogous gene families **(Supplemental Figure 23)**. The chromosome 13-11 SD01 contig reached the DNAJ protein (PF3D7_1149600) and terminated (**Supplemental Figure 24**), while normal 11 continued through 2 *var* genes and 4 rifin genes, likely because the assembly was unable to contend with the near complete identical sequence between the two chromosomes. Examination of the longer normal 11 portion revealed two-fold coverage and no variation. Therefore, it is likely the SD01 has identical 11 segments intact to the telomere of each chromosome.

Analysis of the 11 other PacBio assemblies[30] with normal chromosome 11 showed that homology between strains also ended 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 deleted in the hybrid chromosome 13-11 corresponded to 3D7 genes PF3D7_1371500 through PF3D7_1373500 and include notably *pfhrp3* 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 without other structural alterations.

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

To better define the breakpoint, we examined microhaplotypes within the 15.2kb ribosomal duplication for the 98 $13^{-}11^{++}$ strains containing near-perfect chromosome 11 segments (**Supplemental Figure 8**). Within each strain, the microhaplotypes in the telomeric region are identical, consistent with a continuation of the adjacent chromosome 11 duplication. However, for nearly all strains, as the region traverses towards the centromere within the ribosomal duplication, there is an abrupt transition where the haplotypes begin to differ. These transition points vary but are shared within specific groupings correlating with the chromosome 11 microhaplotypes (**Supplemental Figure 8**). These transition points likely represent NAHR

exchange breakpoints, and their varied locations further support that multiple intrastrain translocation events have given rise to $13^{-}11^{++}$ parasites in the population.

DISCUSSION

Here, we used publicly available short-read and long-read from parasites across the world and newly generated long-read sequencing data to identify *pfhrp2* and *pfhrp3* deletions and their mechanisms in field *P. falciparum* parasites. The limited number of *pfhrp2*-deleted strains showed chromosome 8 breakpoints predominantly in the gene with evidence of telomere healing, a common repair mechanism in *P. falciparum*[16,18]. We found that *pfhrp3* deletions occurred through three different mechanisms. The least common mechanism was the simplest involving simple breakage loss of chromosome 13 from pfhrp3 to the telomere, followed by telomere healing (13^{-} pattern). The second most common pattern $13^{-}5^{++}$ was likely the result of NAHR, within 20-28bp di-nucleotide AT repeats translocating a 26,535bp region of chromosome 5 containing *pfmdr1* onto chromosome 13, thereby duplicating *pfmdr1* and deleting *pfhrp3*. There appeared to be one origin of $13^{-}5^{++}$, which was only observed in the Asia population, and its continued presence was potentially driven by the added benefit of *pfmdr1* duplication in the presence of mefloquine. The most common pattern, $13^{-}11^{++}$, predominated in the Americas and Africa and was the result of NAHR between chromosome 11 and 13 within the large 15.2kb highly-identical ribosomal duplication, translocating and thereby duplicating 70,175bp of core chromosome 11 plus 15-87kb of paralogous sub-telomeric region replacing the chromosomal region on chromosome 13 that contained *pfhrp3*. Importantly, NAHR-mediated translocations resulting in deletion have repeatedly occurred based on evidence of multiple breakpoints and chromosome 11 haplotypes with identical copies in parasites. These findings combined with identical copies of the shared chromosome 11 segment suggest that these parasites represent

multiple instances of intrastrain (self) NAHR-mediated translocation followed by clonal propagation of $13^{-}11^{++}$ progeny. While Hinterberg *et al.* proposed that a general mechanism of non-homologous recombination of the subtelomeric regions may be responsible for translocating the already existing deletion of *pfhrp3*[12], our analysis would suggest ribosomal duplication-mediated NAHR is the likely cause of the *pfhrp3* deletion itself. The high frequency of the meiotic translocation in the laboratory cross further supports the hypothesis that these NAHR-mediated translocations are occurring at a high frequency in meiosis in natural populations. Consequently, this suggests that progeny must be strongly selected against in natural populations apart from where specific conditions exist, allowing for *pfhrp3* deletion to emerge and expand (e.g. South America and the Horn of Africa).

Positive selection due to drug resistance may underlie pattern $13^{-}5^{++}$ translocation that duplicates *pfmdr1* onto chromosome 13. In South East Asia, the only place containing pattern $13^{-}5^{++}$, existing tandem duplications of *pfmdr1* exist that provide mefloquine resistance, and mefloquine has been used extensively as an artemisinin partner drug, unlike Africa[38]. Discordant reads, local assembly, and TARE1 identification support NAHR-mediated translocation of *pfmdr1* followed by telomeric healing to create a functional chromosome. All strains showed the same exact NAHR breakpoint and TARE1 localization consistent with a single origin event giving rise to all 13-5++ parasites. Interestingly, *pfmdr1* duplications have been shown to be unstable with both increases and decreases in copy numbers frequently occurring[39]. During de-amplification, a free fragment of DNA containing a *pfmdr1* copy may have been the substrate that integrated into chromosome 13 by NAHR, followed by telomerase healing to stabilize the 13-5 hybrid chromosome, analogous to *var* gene recombination events where double-stranded DNA is displaced, becoming highly recombinogenic[17]. A clonal expansion of $13^{-}5^{++}$ parasites could be due to the benefit of the extra *pfmdr1* copy on chromosome 13, the loss of *pfhrp3*, or both. Its expansion in SEA would be consistent with selection due to copy-number-associated mefloquine resistance given mefloquine's extensive use as an individual and artemisinin partner

drug in the region. Furthermore, given all isolates with evidence of this duplication either had only wild-type *pfmdr1* or were mixed, the 13-5 chromosome copy most likely had wild-type *pfmdr1*. In 20 out of 27 *pfmdr1* duplication cases with a mixed genotype of *pfmdr1* (**Supplemental Figure 5**), the core genome *pfmdr1* had the Y184F mutation with no other mutations detected within the *pfmdr1* gene. Isolates containing only Y184F in *pfmdr1* were shown to be outgrown by wild-type *pfmdr1*[40], which would mean having the wild-type *pfmdr1* duplication on chromosome 13 might confer a stable (non-tandem) "heterozygous" survival advantage beyond just increased copy number-mediated 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 sequenced *pfhrp3* deleted lab isolates, HB3 and SD01, to generate reads spanning the 15kb duplicated region, showing support for a normal chromosome 11 and a hybrid 13-11 in both isolates. These findings supported 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 was consistent with the genomic coverage pattern we observed in publicly available data from 122 *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[16,17]. The event described here represented 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 predominately in low transmission settings, resulting in four potential progeny including one with normal chromosomes 11 and 13 and three with translocations **(Figure 6)**. This could 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 parasites, 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 occurs frequently during meiosis and can explain the frequent rearrangements seen between chromosomes 11 and 13 in the previous experimental cross of HB3 x DD2[12]. Meiotic misalignment and subsequent NAHR is a common cause of high-frequency chromosomal rearrangements including in human disease (eg. in humans, 22q11 deletion syndrome due to misalignment of duplicated blocks on chromosome 22 occurs in 1 in 4000 births)[41]. This high frequency could explain why *pfhrp3*-deleted isolates are more common in many populations relative to pfhrp2[9,42,43], which likely requires infrequent random breaks along with rescue by telomere healing. In the future, more extensive sequencing of RDT-negative *P. falciparum* parasites 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[44]. 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. Additionally, *pfhrp3* deletions arising in culture have not been observed. 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 occur, leading to a fully RDT-negative parasite. This is supported by evidence that *pfhrp3* deletion appears to predate *pfhrp2* deletions in the Horn of Africa[9].

The biological effects of *pfhrp2* and *pfhrp3* loss and potential selective forces are complicated due to other genes lost and gained and the extent of the rearrangements.

Increased copies of genes on chromosome 11 could be beneficial, as *pf332* on the chromosome 11 duplicated segment was found to be essential for the binding of the Maurer cleft to the erythrocyte skeleton and is highly expressed in patients with cerebral malaria[45]. Conversely, lack of this protein is likely detrimental to survival and may be the reason the reciprocal hybrid 11-13 was not observed in field isolates. 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 RDT would have likely detected the increased PfHRP3 encoded by duplicated pfhrp3, sampling should not be biased against detecting parasites with this reciprocal hybrid 11-13. Thus, the lack of 11-13 rearrangement in field isolates suggests that the selective disadvantage of the lost and gained genes was 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 *pfhrp3* 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 that there are likely further 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* versus *pfhrp3*. However, given we still have a limited understanding of their biological roles, there may be situations where selective forces may favor loss of *pfhrp2* relative to *pfhrp3*. Overall, our findings are clinically important, because 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 these deletions including representative sampling are needed to determine

the prevalence, interactions, and impacts of *pfhrp2* and *pfhrp3* deletions and to examine the selective pressures and complex biology underlying them.

MATERIALS AND METHODS

Genomic locations and read-depth analysis

Conserved non-paralogous genomic regions surrounding *pfhrp2* and *pfhrp3* were determined to study the genomic deletions encompassing these genes. This was accomplished by first marking the 3D7 genome with the program tandem repeat finder[46], then taking 200bp windows, stepping every 100bp between these tandem repeats, and using LASTZ[47](version 1.04.22) to align these regions against the reference genome 3D7 (version 3, 2015-06-18) and 10 currently available chromosomal-level PacBio assembled genomes[30] that lacked *pfhrp2* and *pfhrp3 deletions*. Regions that aligned at >70% identity in each genome only once were kept, and overlapping regions were then merged. Regions from within the duplicated region on chromosome 11 and chromosome 13 were kept if they aligned to either chromosome 11 or 13 but not to other chromosomes. Local haplotype reconstruction was performed on each region for 19,289 publicly available whole genome sequences (WGS) of *P. falciparum* field samples, including 24 samples from a recent study in Ethiopia[9] 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 25**). 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 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. The analyzed region on chromosome 5 that included *pfmdr1* spanned from 929,384 to 988,747. 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 25, Supplemental Figure 26).**

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

The 7bp pattern of TT[CT]AGGG[14] was used to determine the presence of TARE1, of which the presence of this pattern was required to occur at least twice in tandem. 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 parasite 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 pfmdr1 Duplication 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 *pfhrp2 and pfhrp3*, 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[36]. The genomes of all *Laverania* have recently been sequenced and assembled using PacBio and Illumina data[36]. The assemblies were analyzed using their annotations and by using LASTZ[47] with 80% identity and 90% coverage of the genes in the surrounding regions on chromosomes 5, 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)[30]. 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[48–50]. 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[51], an assembly pipeline for high-noise single-molecule sequencing, and Flye[52] using default settings. In order to assemble the low coverage and highly similar chromosome 11 and 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 [53]. Mappings were visualized using custom R scripts[54–56].

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 at https://seekdeep.brown.edu/Analysis_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

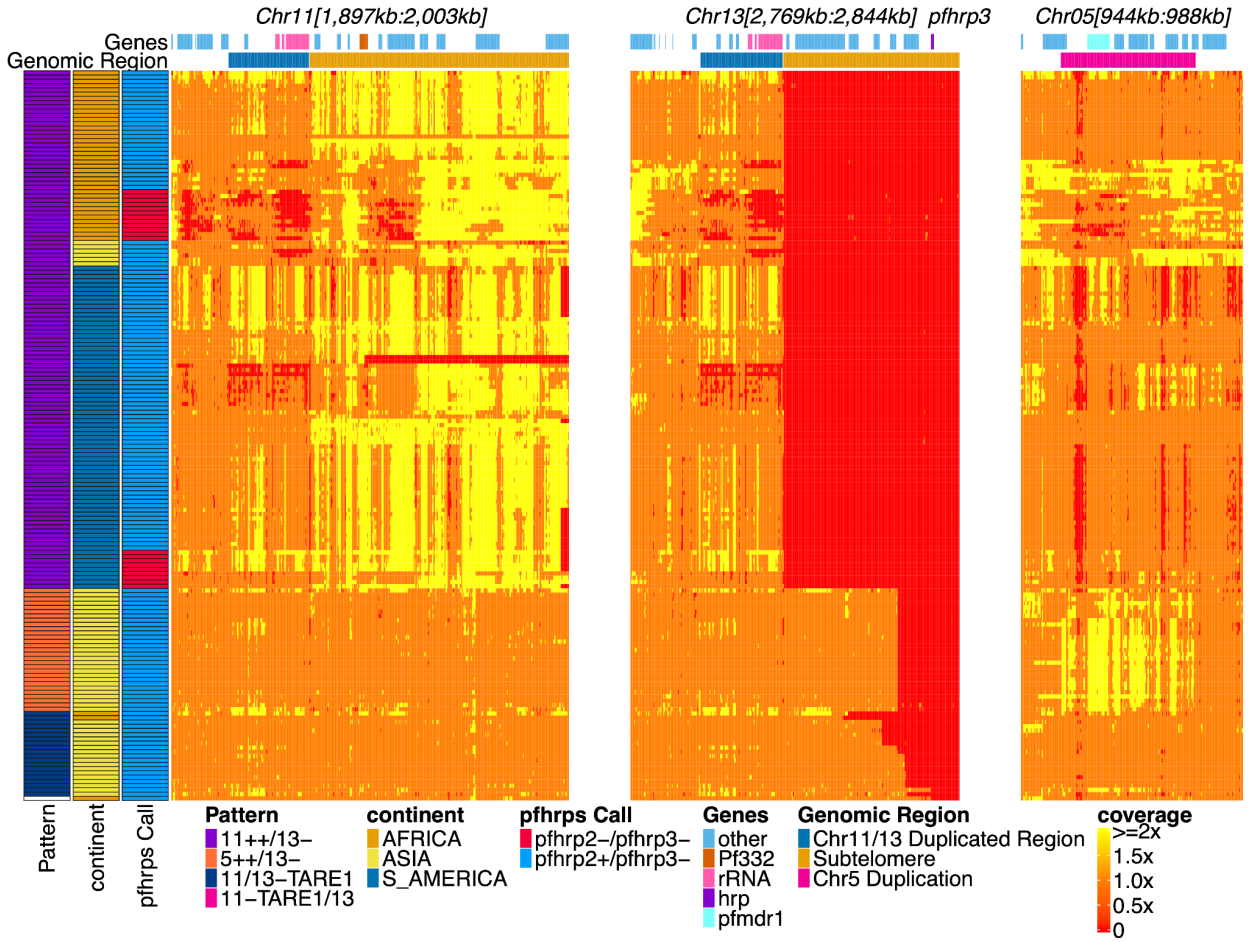


Figure 1. Pfhrp2/3 deleted parasites with altered sequence coverage across regions of chromosomes 11, 13, and 5. Sequence coverage heatmap of *pfhrp3* deletion associated regions of chromosomes 11 (1,897,157 - 2,003,328 bp), 13 (2,769,916 - 2,844,777 bp), and 5 (944,389 - 988,747 bp) in the the 172 samples out of the 19,289 samples with evidence of *pfhrp3* deletion. The regions from chromosomes 11 and 13 are to the end of their core region, while the region from chromosome 5 is the region around *pfmdr1* that is involved in its duplication event. Each row is a WGS sample and each colun is normalized coverage. 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 annotation for samples includes the genomic rearrangement/deletion pattern (patterns with -TARE1 have evidence of TARE1 addition following deletion), the continent of origin, and *pfhrp2/3* deletion calls. Increased variation and biases in coverage correlate with *P. falciparum* selective whole-genome amplification (sWGA), which adds variance and biases to the sequence coverage prior to sequencing.

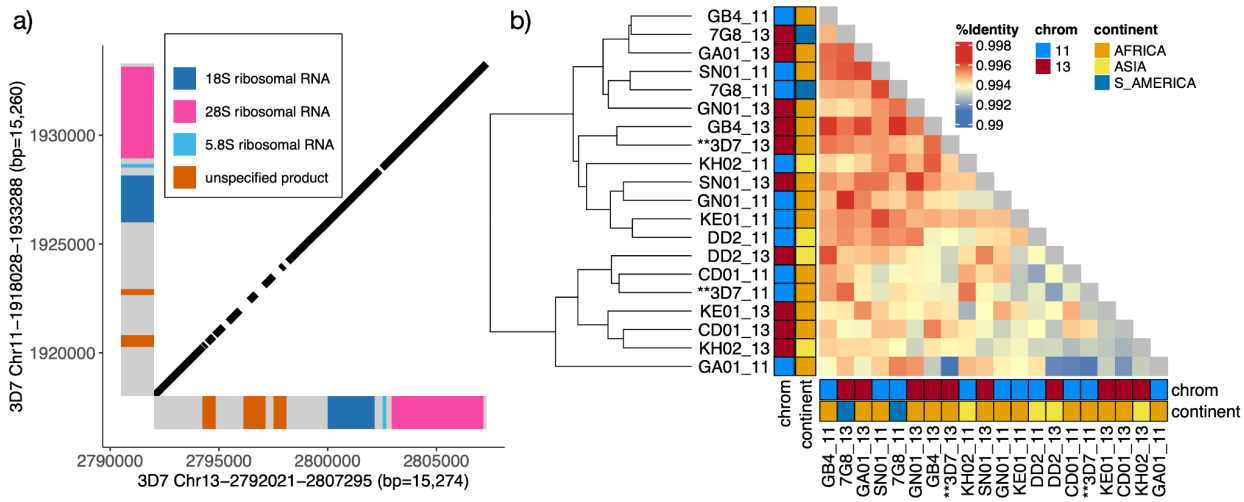


Figure 2. Characterization of the 15.2 kb segmental duplication containing ribosomal genes on chromosomes 11 and 13. (a) Alignment of 3D7 reference genome copies on chromosome 11 (1,918,028-1,933,288 bp) and chromosome 13 (2,792,021-2,807,295 bp). These two regions are 99.3% identical. The diagonal black bars show 100% conserved regions of at least 30 bp in length representing 89.1% of the alignment. Gene annotation is colored. (b) Comparison by pairwise alignments of the duplicated copies from non-*pfhrp3* deleted strains[30] 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.

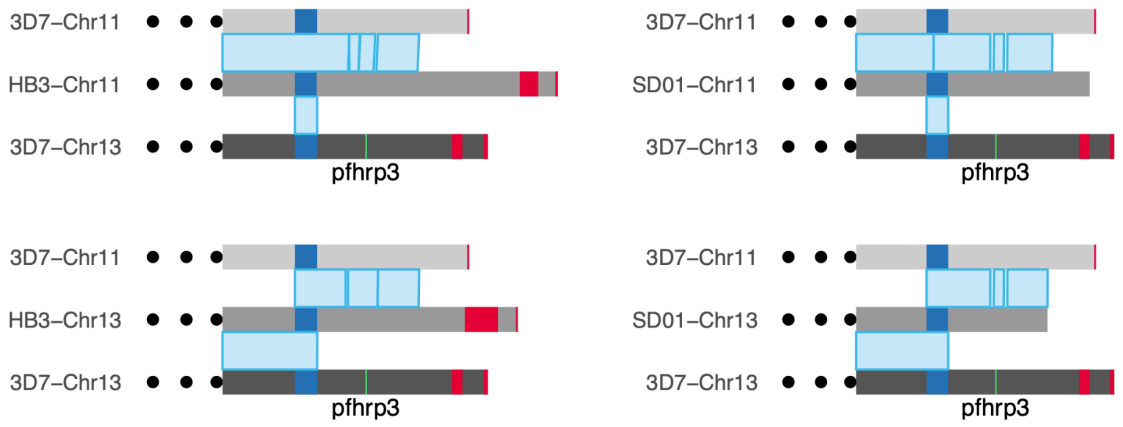


Figure 3. Comparison of long-read assemblies of chromosomes 11 and 13 of HB3 and SD01 to the reference genome 3D7 confirms hybridized chromosome 13-11. Chromosome 11 of HB3 and SD01, on top, mapped entirely to the reference chromosome 11 of 3D7 with the segmental duplication region in dark blue mapped to both 11 and 13. The assembly of chromosome 13 of HB3 and SD01 maps to the reference chromosome 13 of 3D7 up through the segmental 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

associated repetitive elements (TARE) repeats. Displaying only from 50kb upstream from the duplicated region to the end of the chromosomes. Chromosome 11 on 3D7 spans 1,918,029 - 2,038,340 (120,311bp in length) and chromosome 13 on 3D7 spans 2,792,022 - 2,925,236 (133,214bp in length).

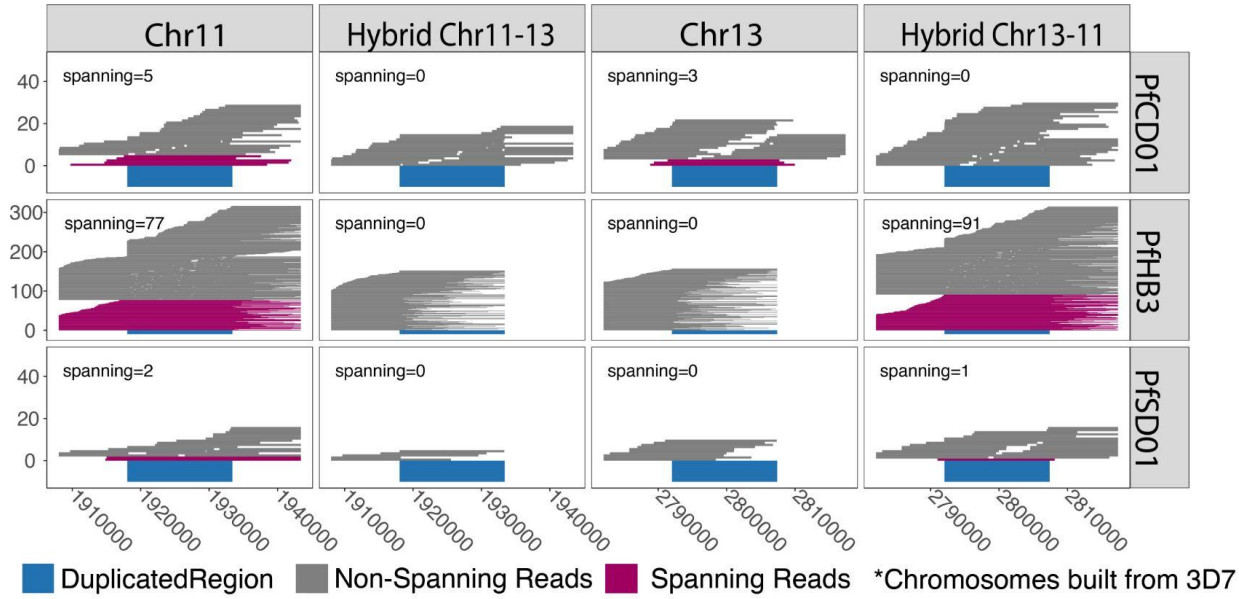


Figure 4. Long reads spanning the 15kb duplicated region confirm presence of translocated chromosome 13-11 in *pfhrp3*-deleted HB3 (Americas) and SD01 (Africa) but not *pfhrp3*-intact chromosomes. PacBio and Nanopore reads >15kb for HB3, SD01, and CD01 are shown aligned to normal chromosomes 11 and 13 as well as hybrid chromosomes 11-13 and 13-11 constructed from 3D7 sequence. Reads that completely span the segmental duplication (dark blue) anchoring in the unique flanking sequence are shown in maroon. 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. Other non-deleted isolates had spanning reads mapped solely to normal chromosomes exemplified by CD01 (top row). No isolates had spanning reads across the hybrid 11-13 chromosome.

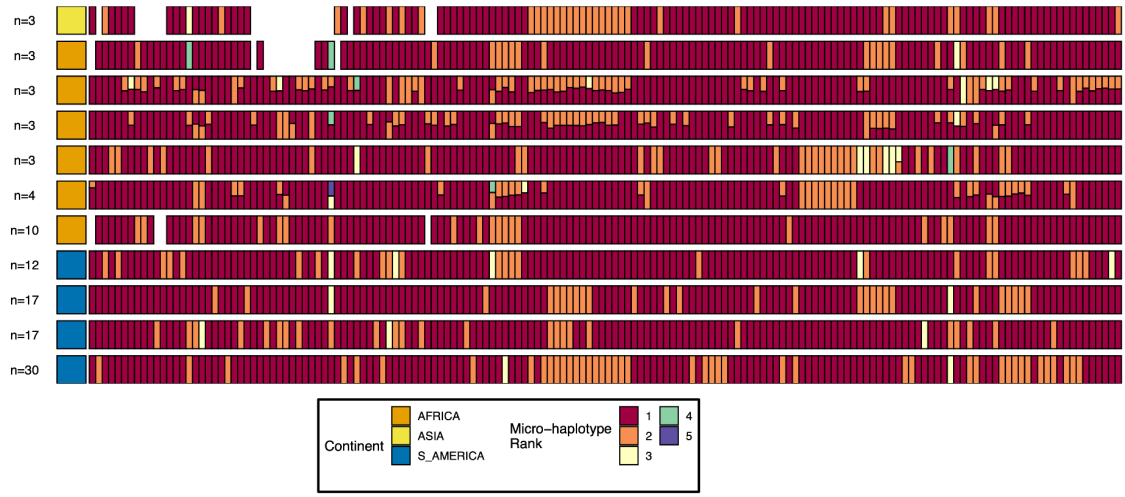


Figure 5. Microhaplotype patterns for the duplicated portion of chromosome 11 in $13^{-}11^{++}$ parasites form 11 distinct haplotype groups with a geographic distinction between Africa and the Americas. Each row represents a group of $13^{-}11^{++}$ parasites based on shared haplotypes on the chromosome 11 duplicated segment. The number of parasites and continent of origin are on the left for each group. Each column is a different genomic region across the duplicated portion of chromosome 11. In each column, the microhaplotype is colored by the prevalence of each microhaplotype, 1=red being most prevalent, 2=orange second most prevalent, and so forth. If more than one microhaplotype for a parasite is present at a genomic location its height is relative to within-parasite frequency. Only sites with microhaplotype variation are shown (n=202). The majority of parasites show singular haplotypes at variant positions despite two copies consistent with identical haplotypes in the group and when there are multiple microhaplotypes the relative frequencies are 50/50 consistent with two divergent copies. Overall haplotype groups are markedly different, which is consistent with separate translocations emerging and spreading independently.

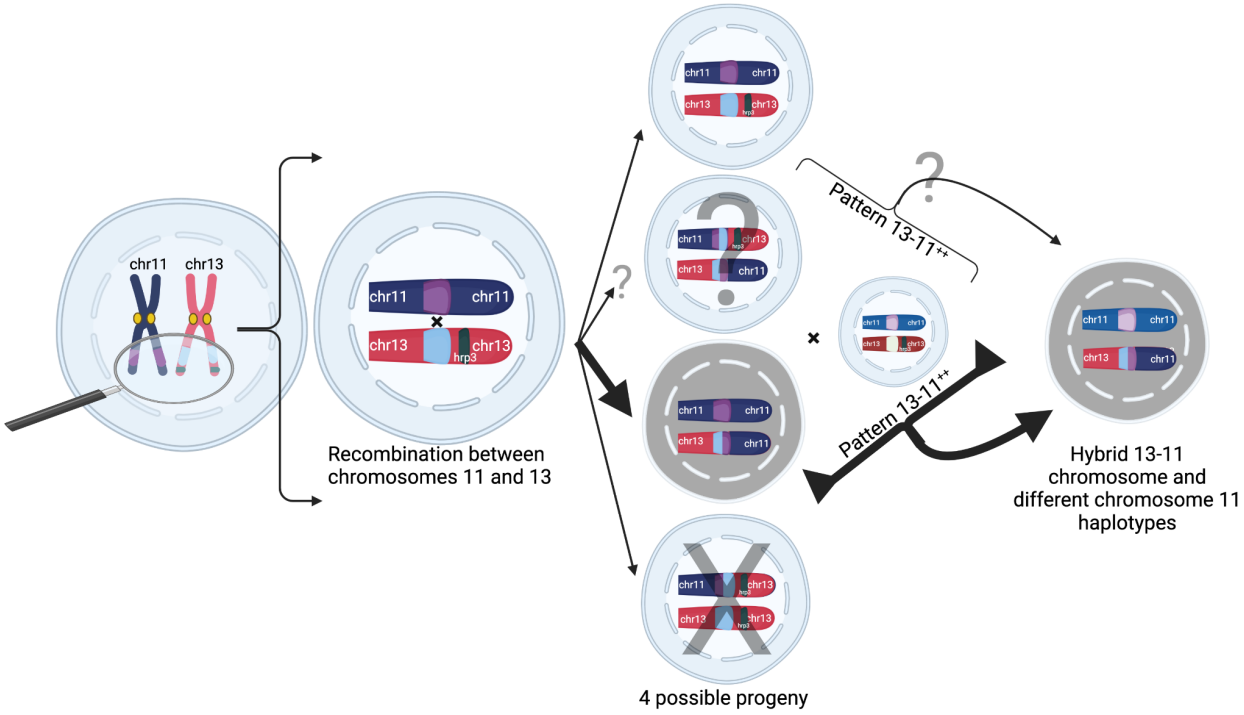
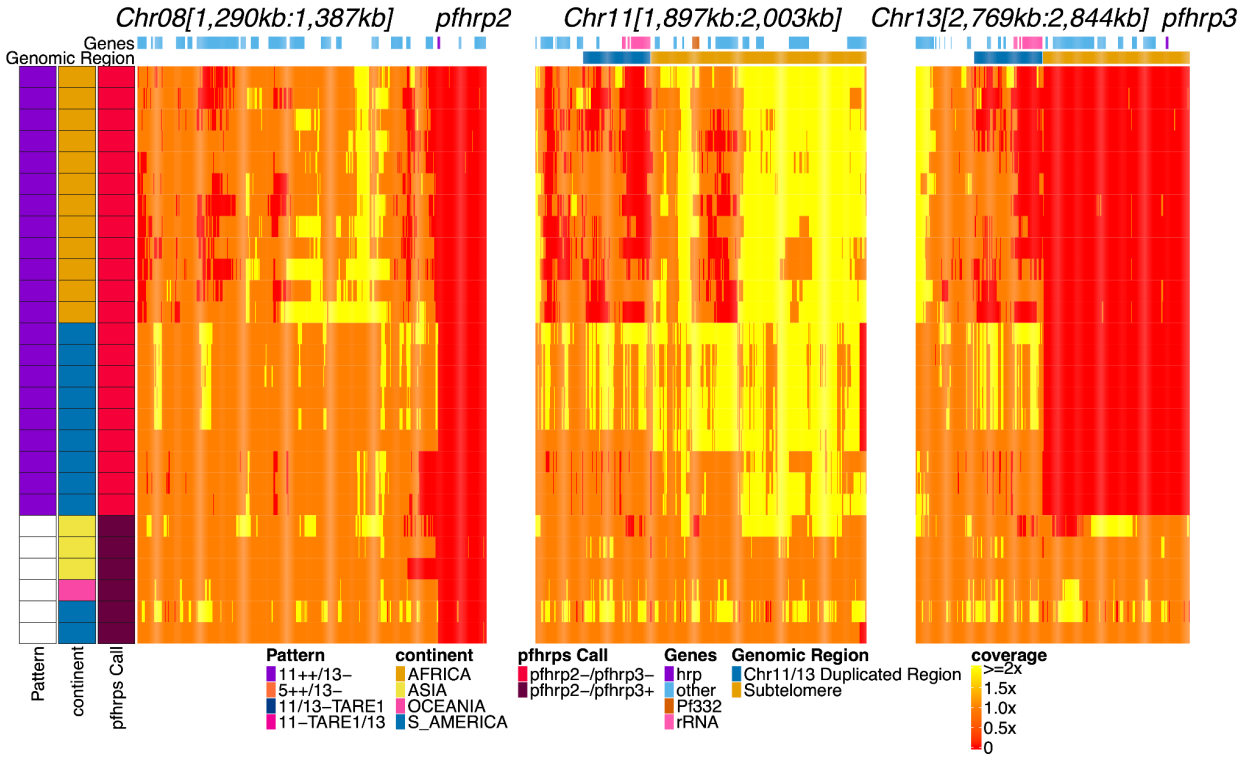


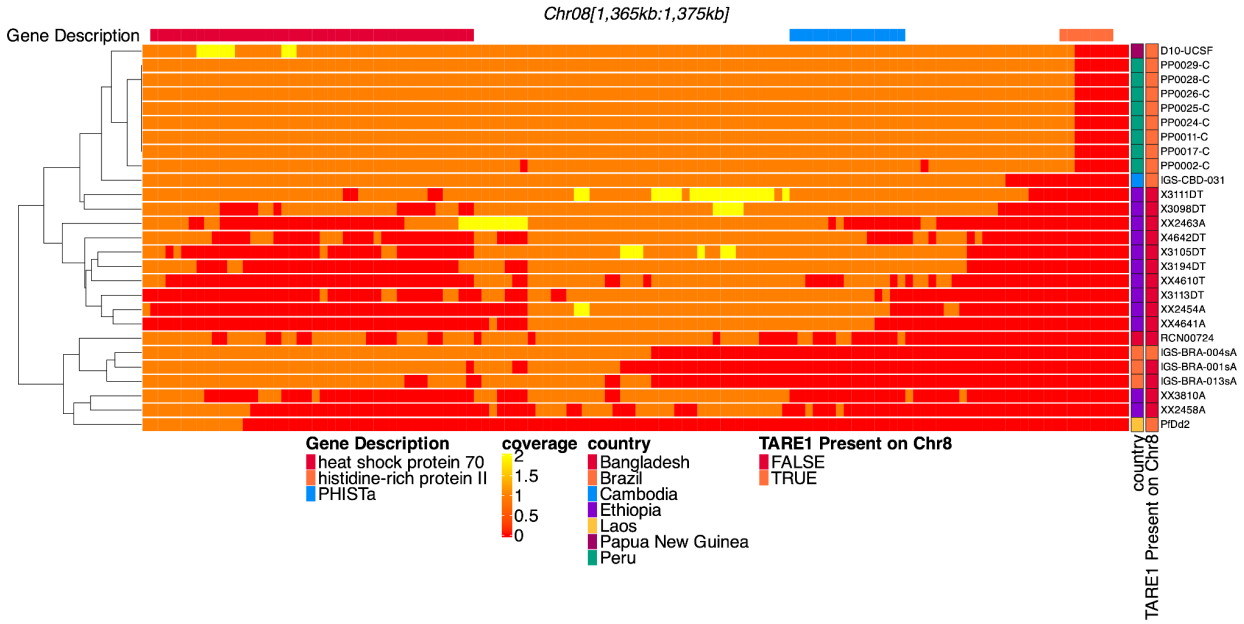
Figure 6. Proposed model of duplication-mediated non-allelic homologous recombination during intrastrain meiotic recombination yielding $13^{-}11^{++}$ parasites. Homology misalignment and NAHR between chromosomes 11 and 13 first occur in an oocyst formed from identical parasite gametes (intrastrain), which can then segregate, resulting in potential progeny (normal and 3 translocated progeny). Bold lines show the most direct path to a $13^{-}11^{++}$ parasite containing a 13-11 hybrid chromosome lacking *pfhrp3* and two identical copies of duplicated chromosome 11 segment seen predominantly. Subsequent recombination with an unrelated strain yields parasites with differing chromosome 11 duplication haplotypes but this can occur with subsequent interstrain meioses. Additionally, 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 $11^{-}13^{++}$ coverage pattern consistent with an 11-13 hybrid chromosome was observed once in a lab strain (FCR3), confirming it can form and thereby supporting that in vivo selective constraints prevent its emergence. Figure created using Biorender.

SUPPORTING INFORMATION



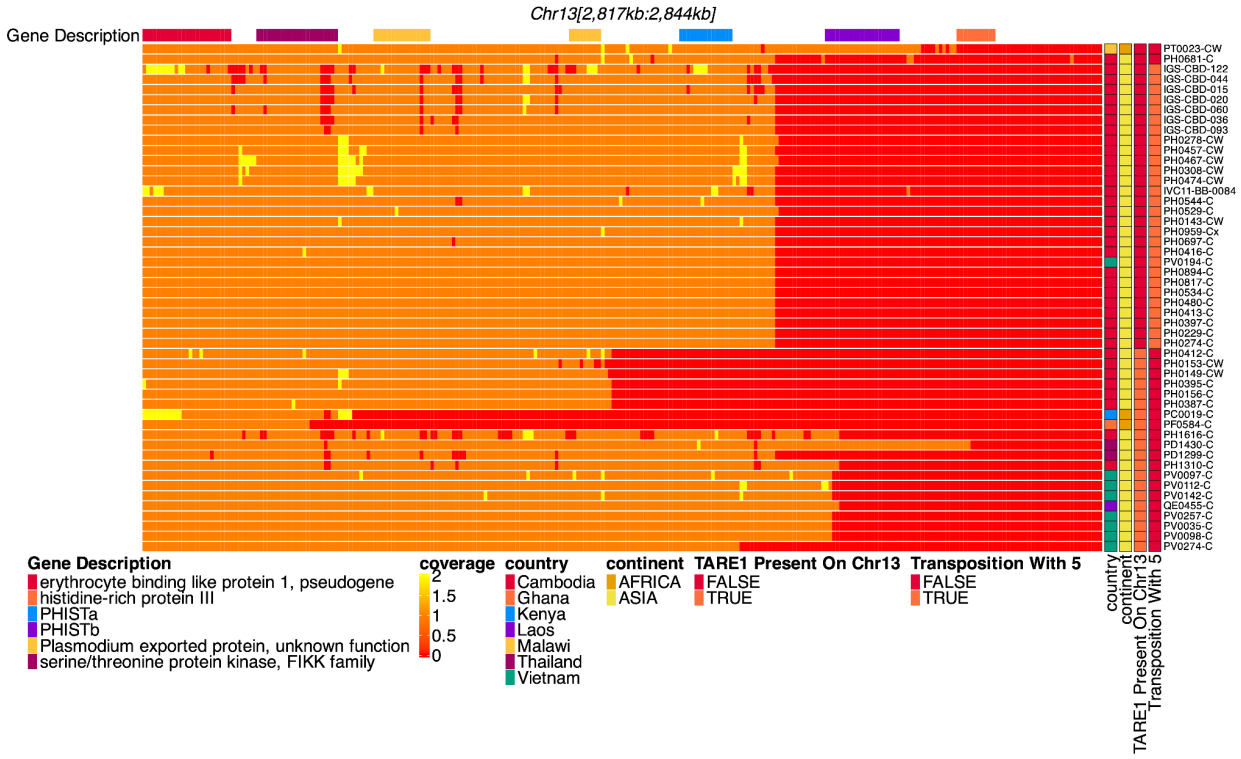
Supplemental Figure 1: Genome coverage of isolates with evidence of *pfhrp2* deletion. Sequence coverage heatmap of chromosomes 8 (1,290,365 - 1,387,982 bp), 11 (1,897,157 - 2,003,328 bp), 13 (2,769,916 - 2,844,777 bp). Displaying the 27 parasites out of the 19,289 total samples that have signs of possible *pfhrp2* deletions. Each row is a parasite. The top annotation along chromosomes depicts the location of genes, and the second row delineates the duplicated region (dark blue) and subtelomere region (orange). The left parasite annotation includes the deletion pattern, continent of origin, and *pfhrp2/3* deletion calls. The 21 parasites

that also have evidence of HRP3 deletion were only found within South America and Africa and had evidence of the $13^{-}11^{++}$ deletion HRP3 deletion pattern. Of the 6 parasites without HRP3 deletion, 2 were from South America, 3 from Asia, and 1 from Oceania.



Supplemental Figure 2: **Coverage of sub-telomeric region of chromosome 8 before *pfhrp2* of parasites with *pfhrp2* deletion.**

Heatmap coverage normalized to genomic coverage of the sub-telomeric region of chromosome 8 (spanning 1,365,360-1,375,435 bp, 10,075bp in length) for the 27 parasites with *pfhrp2* genomic deletion. Each row is a parasite, and each column is a genomic location. The top annotates which gene the region falls within. The right side annotation shows the country of origin and which parasites have evidence of TARE1 at the location where genomic coverage drops to zero within this region. The majority of parasites without evidence of TARE1 or other genomic rearrangement are sWGA parasites and may lack the coverage to detect such events.



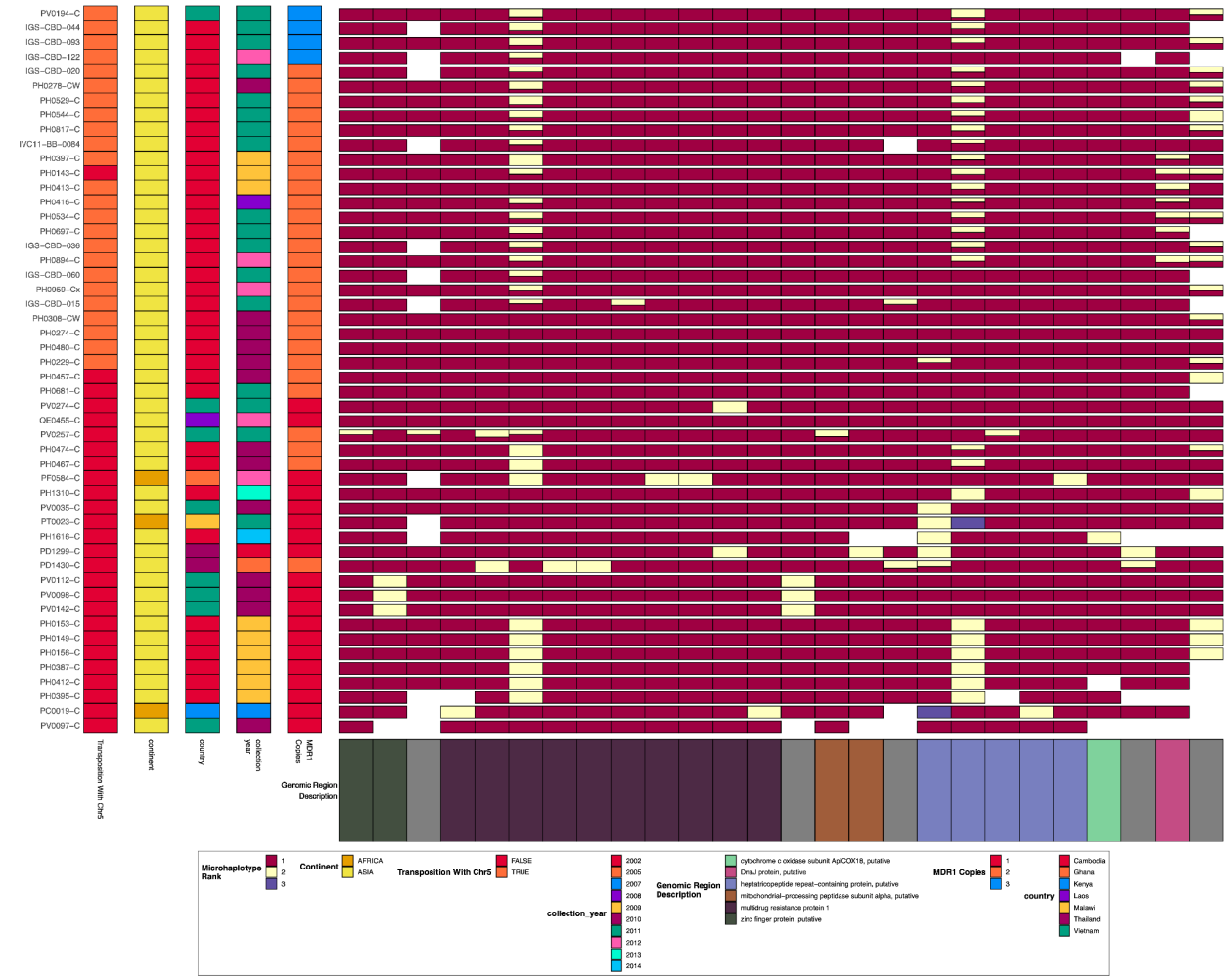
Supplemental Figure 3: **Coverage of chromosome 13 for parasites with *pfhrp3* deletion pattern 13^{-}.**

Heatmap coverage normalized to genomic coverage of the sub-telomeric region of chromosome 13 (spanning 2,817,793 - 2,844,785 bp, 26,992bp in length) for the 50 parasites with *pfhrp3* deletion pattern 13^{-}. Each row is a parasite, and each column is a genomic location. The top annotates which gene the region falls within. The right side annotation shows which parasites have evidence of TARE1 at the location where genomic coverage drops to zero within this region (n=20) and which parasites have evidence of genomic rearrangement with chromosome 5 from discordant paired-end reads, which results in duplication of *pfmdr1* (n=28). The top 2 parasites lack evidence of either deletion type. The next 28 parasites have evidence of rearrangement with chromosome 5 with discordant reads with mates mapping to chromosome 13 and other mates mapping to chromosome 5. The next 20 parasites have evidence of TARE1 contiguous with chromosome 13 sequence on various locations consistent with breakage and telomere healing. These breaks occur on chromosome 13 at 2,836,793 (n=9), 2,830,793 (n=4), 2,829,793 (n=2), 2,821,793 (n=1), 2,822,793 (n=1), 2,833,793 (n=1), 2,834,793 (n=1), 2,836,793 (n=1), and 2,840,793 (n=1).

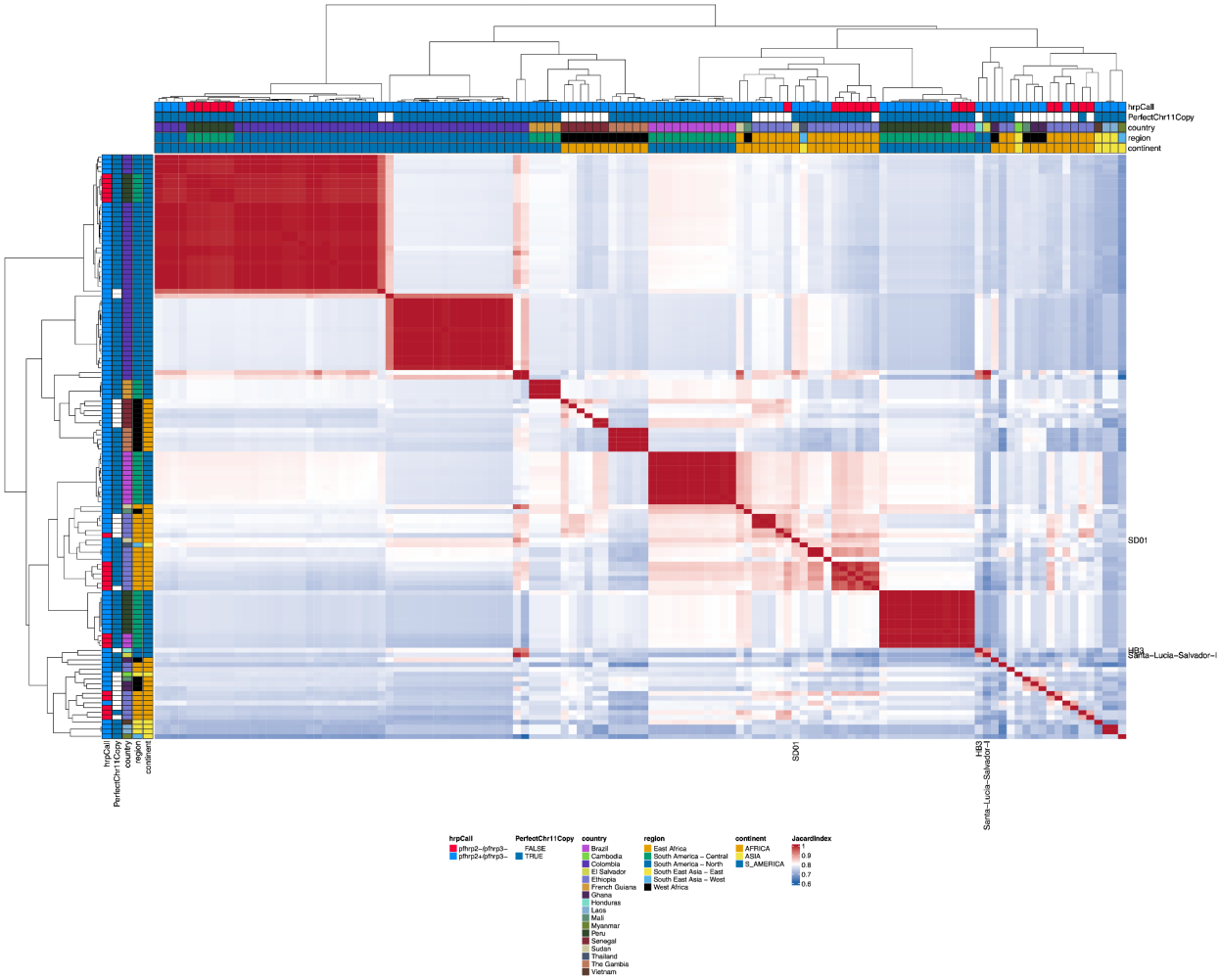
Supplemental Figure 4: **Coverage of chromosome 5 for parasites with *pfhrp3* deletion pattern 13^{-} and 13^{-}5^{++}.**

Heatmap coverage normalized to genomic coverage of a region of chromosome 5 (spanning 944,534 - 988,747 bp, 44,213bp in length) for parasites with *pfhrp3* deletion pattern 13^{-}. Each row is a parasite, and each column is a genomic location. The top annotates which gene the region falls within. The right side annotates the country of origin of the parasite. The 3rd annotation bar on the right side shows which parasites have TARE1 detected on chromosome 13. The parasites with green have TARE1 detected on chromosome 13 that are clustered on the bottom of the graph and have normal coverage across this region of chromosome 5, while the parasites on top have evidence of re-arrangement between chromosome 13 (position 2,835,587) and chromosome 5 (position 979,203) and show increased coverage across chromosome 5 up to the point where TARE1 sequence is detected on the reverse strand. The second annotation pink bar from the top indicates the duplicated region. The beginning of this bar is where the TARE1 sequence on the reverse strand is detected. The above would be consistent with a genomic rearrangement between chromosome 13 at position 2,835,587 and chromosome 5 at position 979,203, which results in the deletion of chromosome 13 from 2,835,587 onwards and the duplication of a 26KB region of the reverse strand of chromosome 5 from position 979,203 to 952,668 resulting in the duplication of *pfmdr1* and the deletion of *pfhrp3*. The parasites that had transposition evidence between 13 and 5 but no TARE1 detected on 5 were all sWGA parasites, which likely limited the ability to detect the TARE1 sequence. The top 4 parasites appear coverage-wise to have 3 copies of *pfmdr1*. The 4 top parasites, in addition to the likely copy on chromosome 13, has a tandem duplication on chromosome 5. There were 2 different tandem duplications detected. The 1st parasite is tandemly duplicated from a monomeric stretch of Ax19 at 947,967 and Ax18 at 970,043 and is from Laos. The 2nd, 3rd, and 4th parasites, all from Cambodia, are tandemly duplicated from Ax22 at chromosome 5 946,684 and Ax36 at 964,504. All parasites with evidence for 13^{-}5^{++} have only wild-type *pfmdr1* or are a mix of wild-type and 184F. The four parasites 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 parasites with this deletion pattern are

found within Asia. Of note, parasite PV0257-C, 6th from the bottom, has a COI of 2 and 2 different copies of MDR1 but normal genomic coverage.

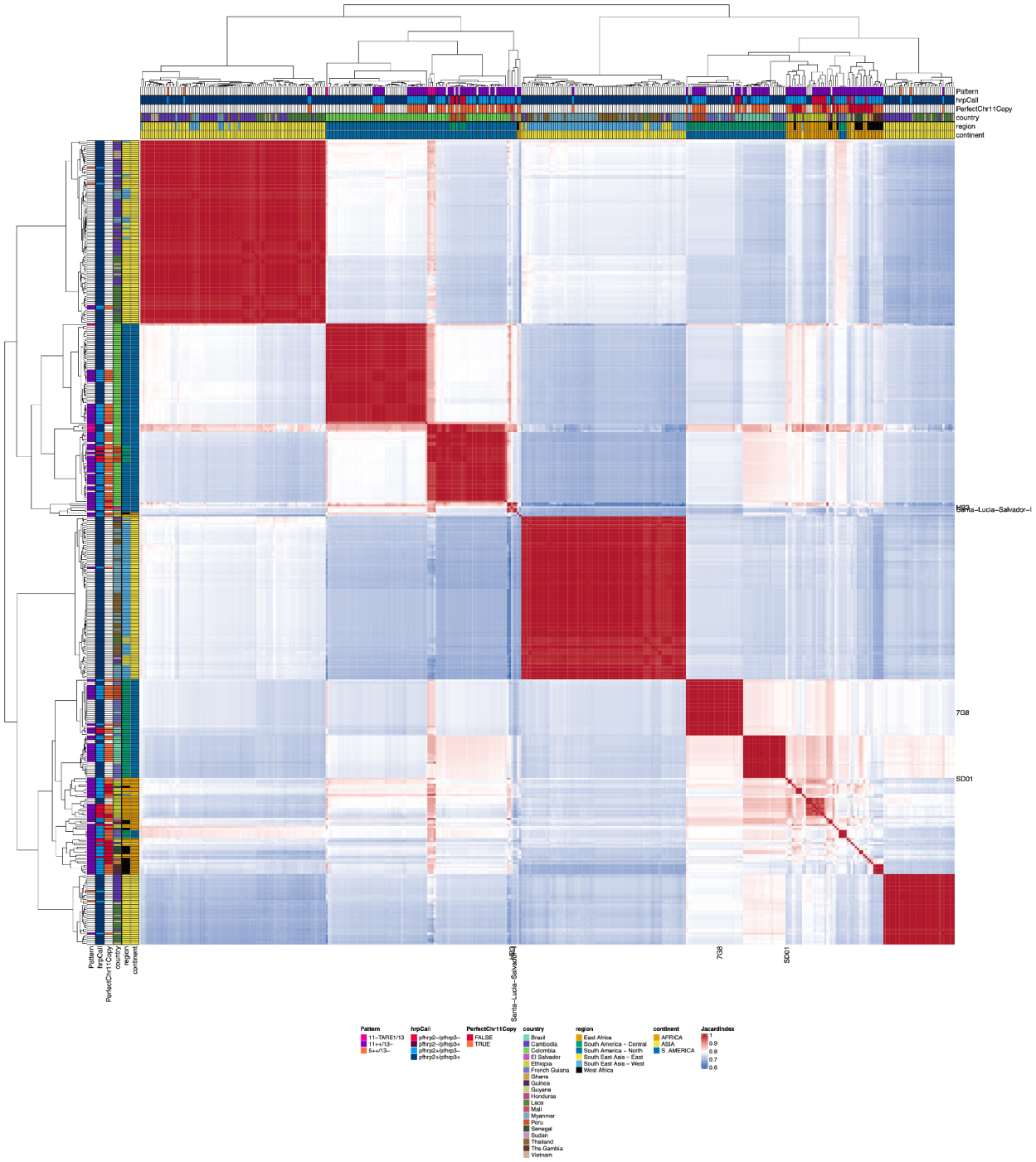


Supplemental Figure 5: **Chromosome 5 duplicated region microhaplotypes.**
The 26 microhaplotype regions across the duplicated portion of chromosome 5 952,668 to 979,203 for the 46 isolates with chromosome 13 deletion without chromosome 11 duplication. Each row is an isolate. In each column, the isolate is typed by microhaplotype (colored by the prevalence of each microhaplotype 1=red being most prevalent, 2=yellow second most prevalent, 3=purple least prevalent). This color coding system is specific to each column, and the same color across columns does not indicate the same haplotype, just the prevalence in the population for that column. Associated metadata for each isolate can be seen on the left after the isolate's name. The majority (n=28) of these show evidence of a complex recombination with chromosome 5 at 952,668 and 2,835,587 on chromosome 13, which results in deletion of *pfhrp3* and duplication of *pfmdr1*. Only 6 isolates with MDR duplication have no variation within *pfmdr1*. The other isolates have a wild type Y184 (yellow) on one copy and 184F (red) on the other copy.



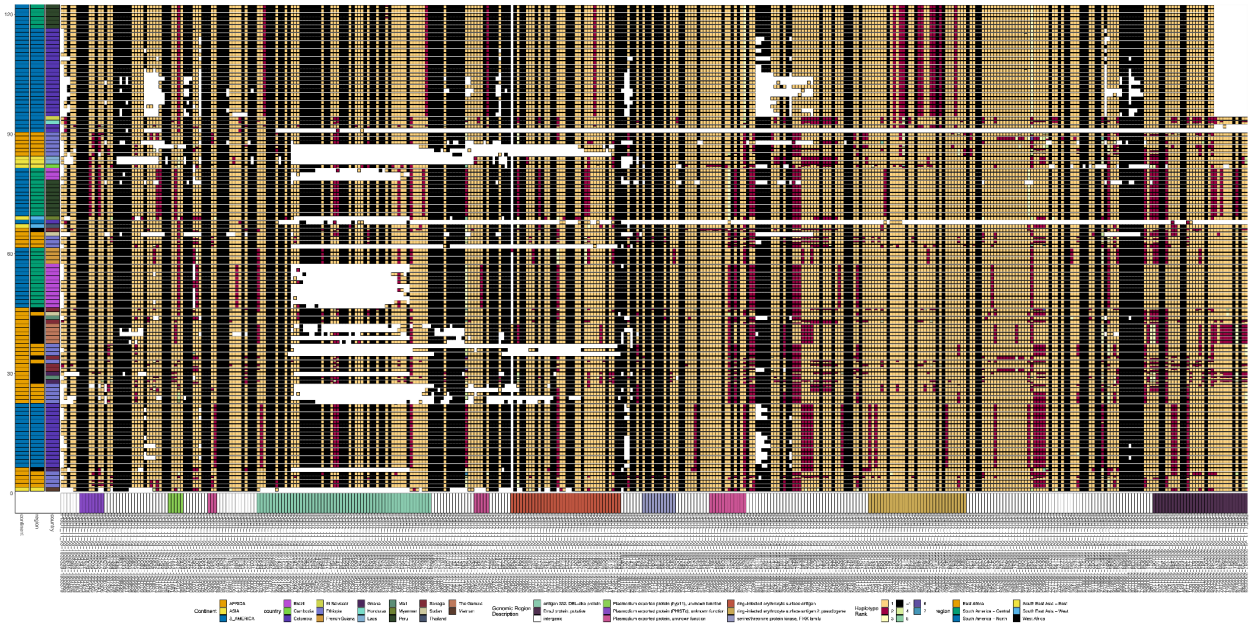
Supplemental Figure 6: **Jaccard similarity between parasites for chromosome 11 duplicated segment for *pfhrp3* deletion pattern $13^{-}11^{++}$ parasites.**

An all-by-all distance matrix showing Jaccard similarity for the duplicated chromosome 11 segments between all parasites with *pfhrp3* deletion pattern $13^{-}11^{++}$. The parasites' continent, region, and country are annotated on the sides of the heatmap as well as the *pfhrp2/3* deletion calls and whether the chromosome 11 duplicated segment shows an identical haplotype across the chromosome 11 duplicated segment. There are clearly several different haplotypes within the duplicated chromosome 11 segment, and there does not appear to be one specific haplotype associated with the duplication. Parasites group strongly by geographical location.



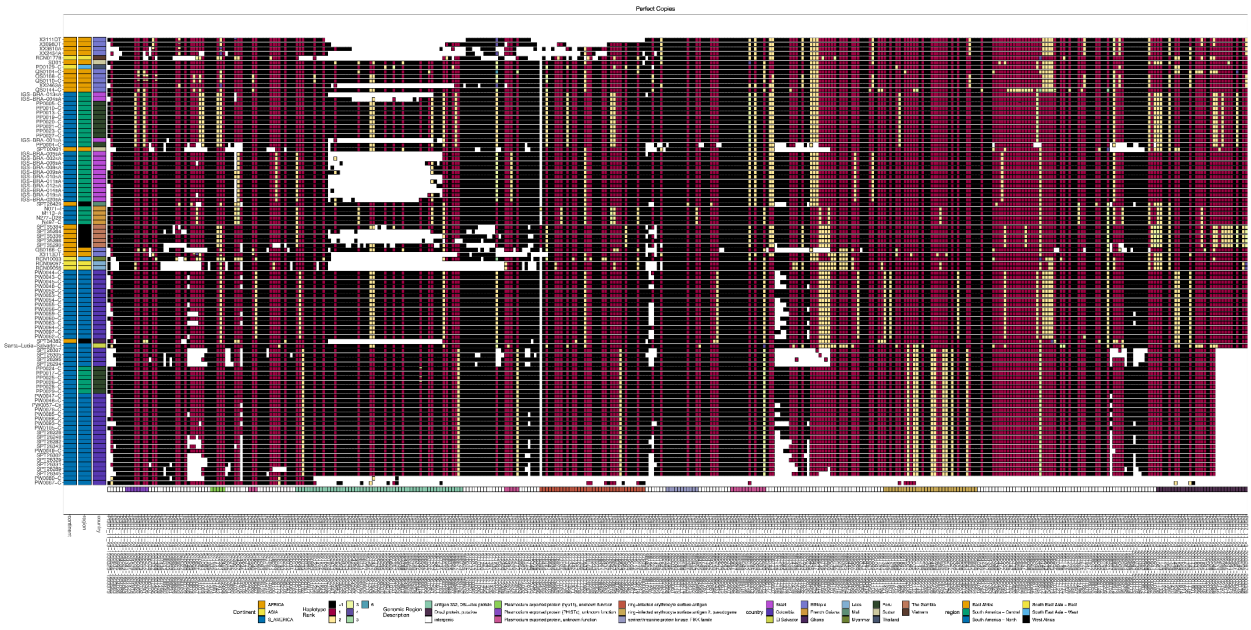
Supplemental Figure 7: **Jaccard similarity for chromosome 11 duplicated segment**
All parasites with micohaplotypes similar to the duplicated chromosome 11 microhaplotypes for the *pfhrp3* deletion Pattern $13^{-}11^{++}$ parasites**.** While similar to **Supplemental Figure 6**, this all-by-all heatmap of Jaccard similarity includes all parasites with a similar chromosome segment to the parasites with pattern $13^{-}11^{++}$ *pfhrp3* deletions. For the side and top annotation for the parasites that do not have chromosome 11 duplication, there is a gray bar for whether or not they have perfect chromosome 11 duplication. There are many parasites with closely related chromosome 11 segments to the duplicated chromosome 11 segments, indicating that the

duplicated chromosome 11 segments are also circulating within the population in strains with normal chromosome 11 and 13 arrangements.

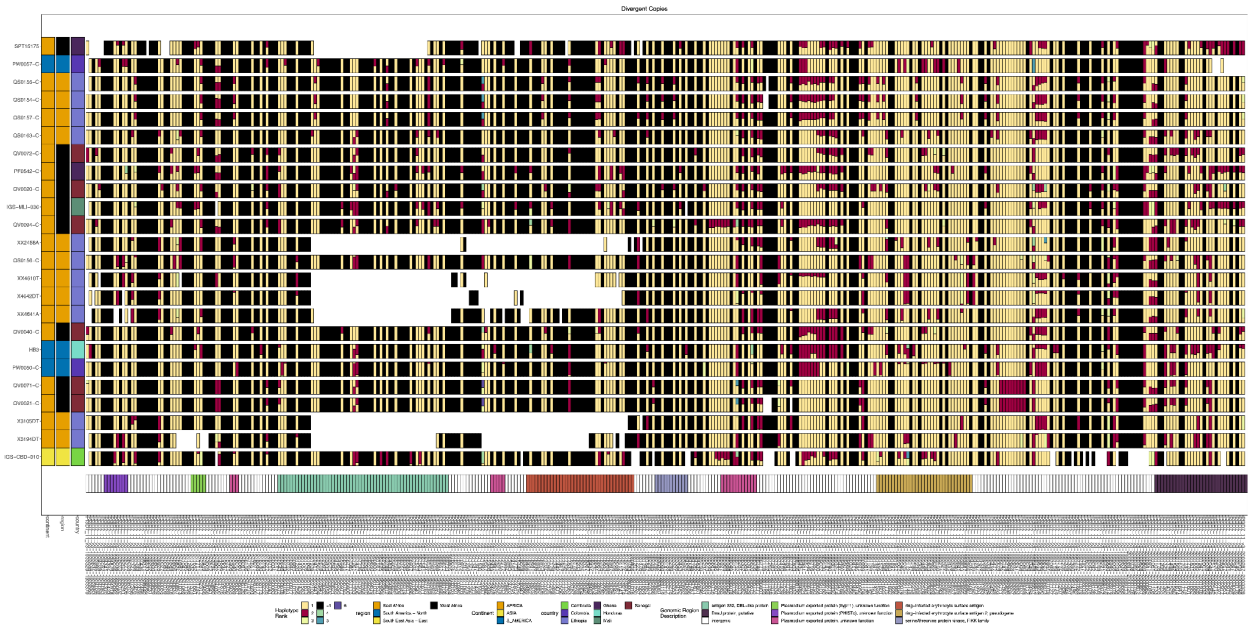


Supplemental Figure 8: Chromosome 11 Duplicated Segment *pfhrp3* deletion Pattern $13^{-}11^{++}$ parasites.

Plotted haplotype variation per sub-genomic regions across the duplicated chromosome 11 segment for the *pfhrp3* pattern $13^{-}11^{++}$ parasites. Across the x-axis are the genomic regions in genomic order, and the genomic region genes are colored on the bottom bar. Y-axis is each parasite with pattern $13^{-}11^{++}$ of *pfhrp3* deletion where this segment of chromosome 11 is duplicated onto chromosome 13. The continent, region, and country are colored per parasite on the leftmost of the plot. Each column contains the haplotypes for that genomic region colored by the haplotype rank at that window. If the column is black, there is no variation at that genomic window. Colors are done by the frequency rank of the haplotypes, and shared colors between columns do not mean they are the same haplotype. If there is more than one variant for a parasite at a genomic location, the bar's height is the relative within-parasite frequency of that haplotype for that parasite. The parasites are ordered in the same order as the heatmap dendrogram seen in **Supplemental Figure 6**. There are clear distinctive haplotypes for this duplicated region.

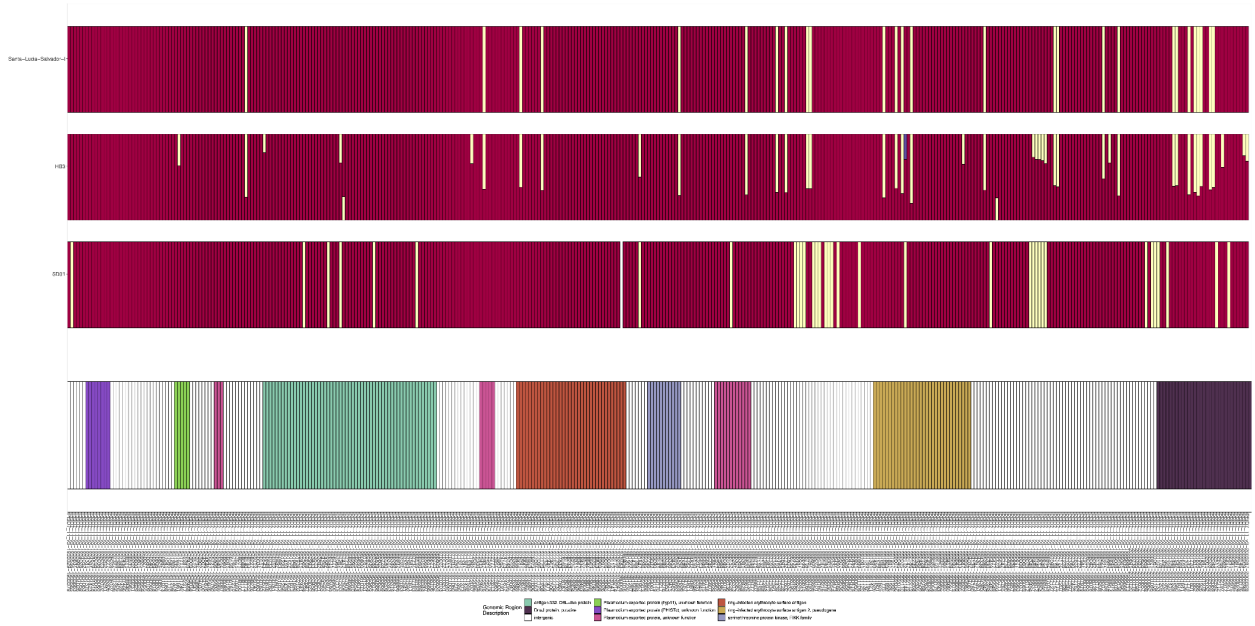


Supplemental Figure 9: **Chromosome 11 Duplicated Segment *pfhrp3* deletion Pattern $13^{-}11^{++}$ parasites with perfect copies.**
Subset of the parasites from **Supplemental Figure 8** for the parasites that have a perfect duplication of the chromosome 11 segment. There are clearly very divergent haplotypes for the perfect duplications, which would indicate that the duplication event is happening multiple times and is not stemming from a single event that all parasites are descended from.

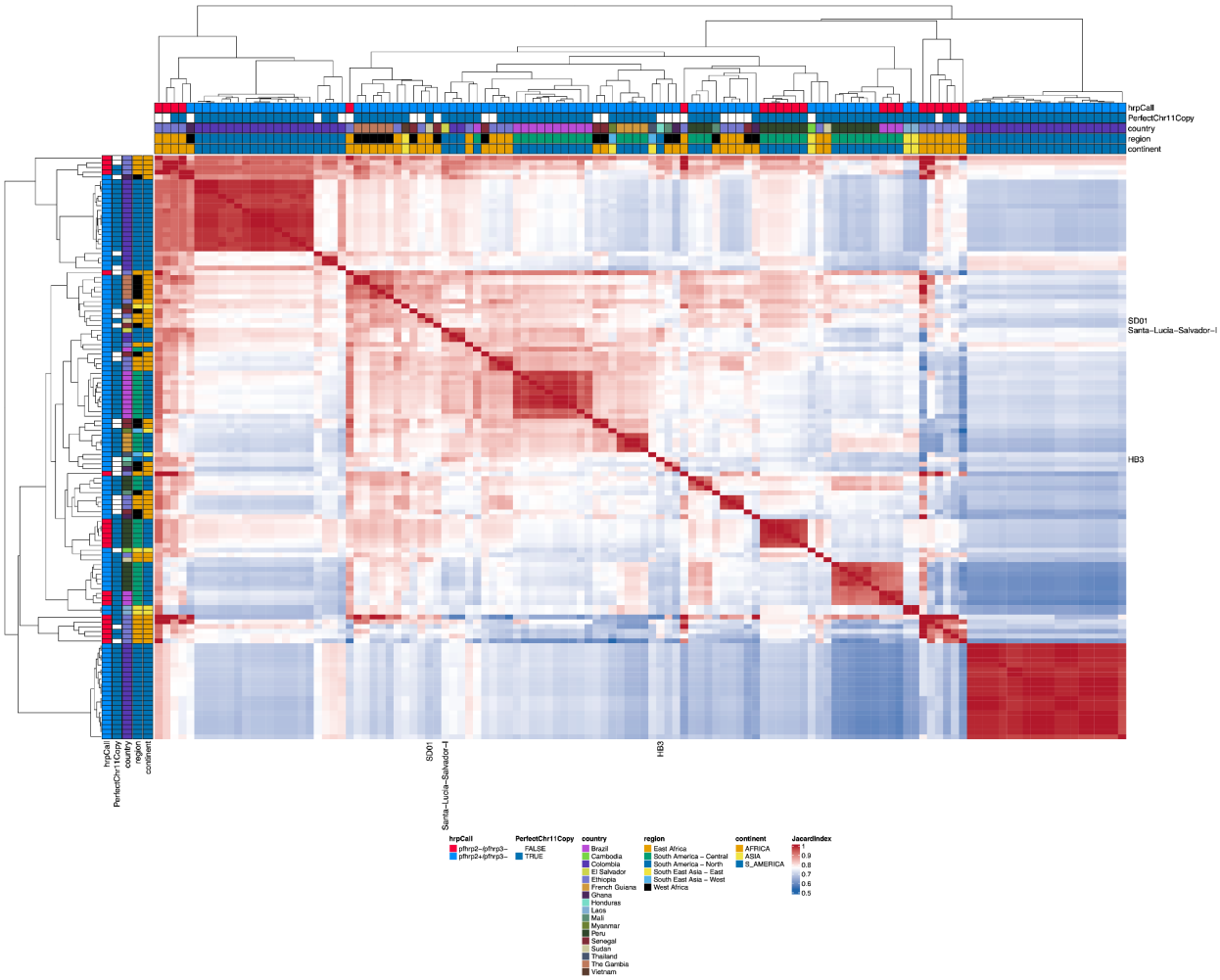


Supplemental Figure 10: **Chromosome 11 Duplicated Segment *pfhrp3* deletion Pattern $13^{-}11^{++}$ parasites with divergent chromosome 11 copies.**

Subset of the parasites from **Supplemental Figure 8** for the parasites that have divergent duplicates of the chromosome 11 segment. There are several parasites that have divergent chromosome 11 segments, but they share the same exact divergent copies with other parasites, which would be consistent with the coinheritance of the two divergent copies simultaneously. This could be consistent with parasites inheriting from previous duplication events involving divergent copies or meiotic recombination between parasites with two separate duplication events of disparate chromosome 11 segments, inheriting one chromosome 11 segment on chromosome 13 from parent 1 and a different chromosome 11 segment on chromosome 11 from parent 2.

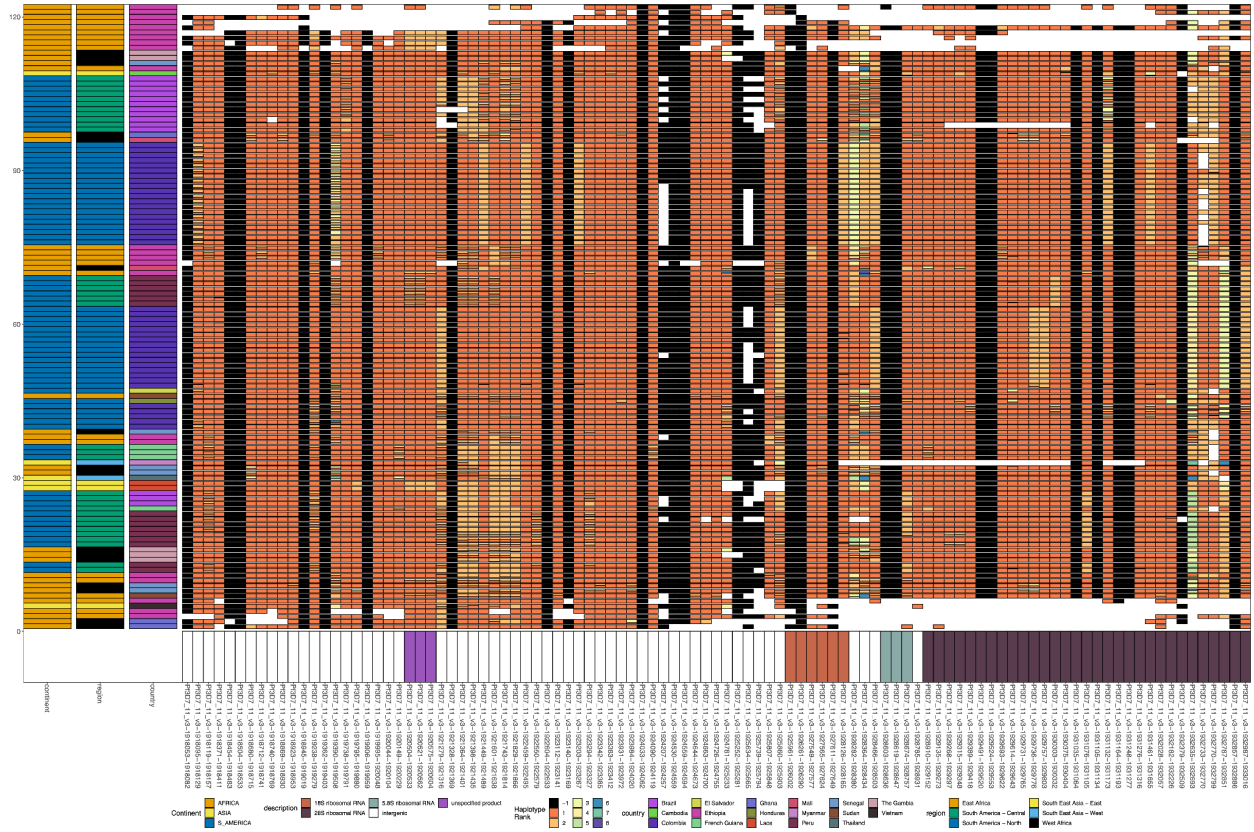


Supplemental Figure 11: Chromosome 11 Duplicated Segment coverage for *pfhrp3* deletion Pattern $13^{-}11^{++}$ parasites SD01, HB3, and Salvador 1.
Subset of the parasites from **Supplemental Figure 8** but for SD01 and HB3, which were sequenced in this paper, and for Santa-Luca-Salvador-1, another lab isolate that shows similar *pfhrp3* deletion pattern $13^{-}11^{++}$. SD01 and Santa-Luca-Salvador-1 have perfect copies, while HB3 has divergent copies.



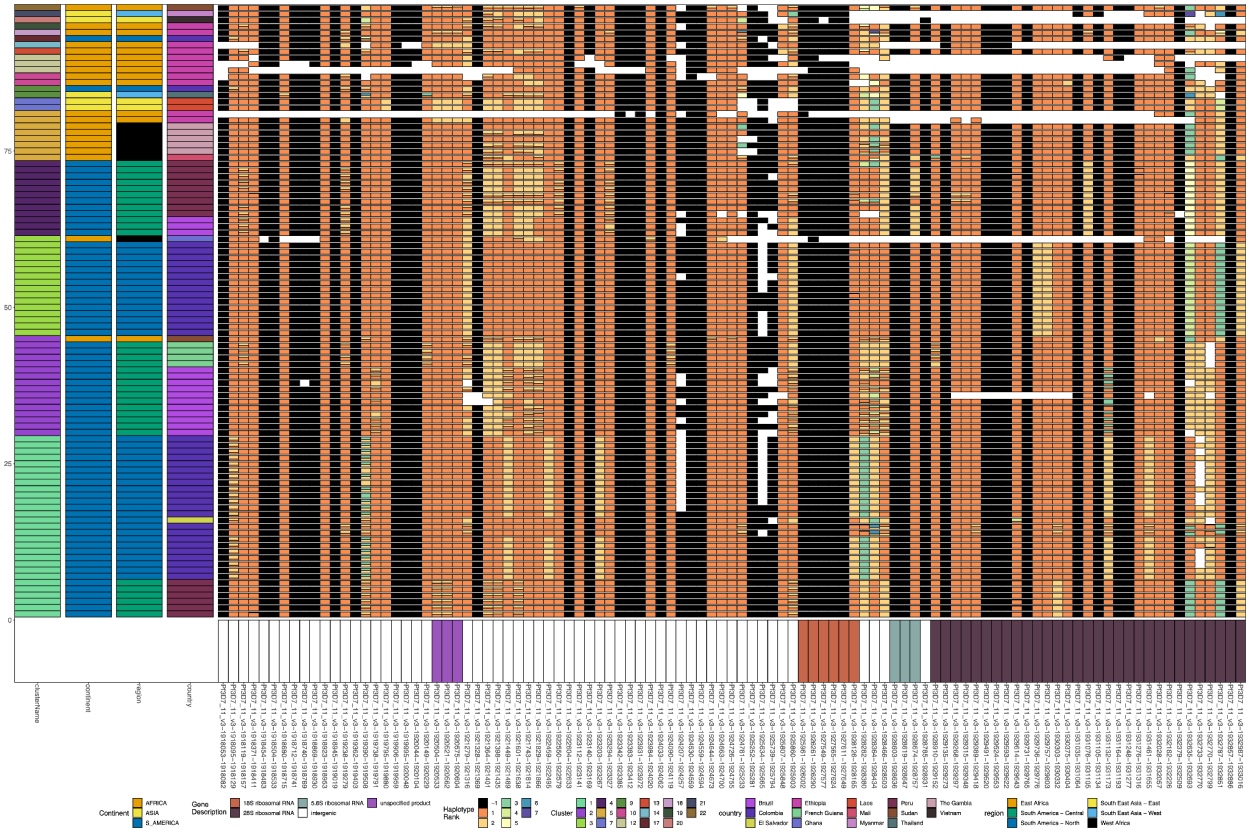
Supplemental Figure 12: Chromosome 11/13 15.2kb duplicated region for *pfhrp3* deletion pattern $13^{-}11^{++}$ parasites.

An all-by-all distance matrix showing Jaccard similarity for the chromosome 11 and 13 duplicated region between all the parasites with *pfhrp3* deletion pattern $13^{-}11^{++}$. The top triangle is identical to the bottom triangle. Parasites' continent, region, and country are annotated on the sides of the heatmap as well as the *pfhrp2/3* deletion calls and whether the chromosome 11 duplicated segment is a perfect copy or not. Sequences tend to cluster per geographic region with similar sequences being from the same country though parasites are not as strong separately by continent as they were for the duplicated chromosome 11 segment. Despite all parasites having duplicated chromosome 11 via this region, there are clear different haplotype groups, which is consistent with multiple different origins of this duplication event.

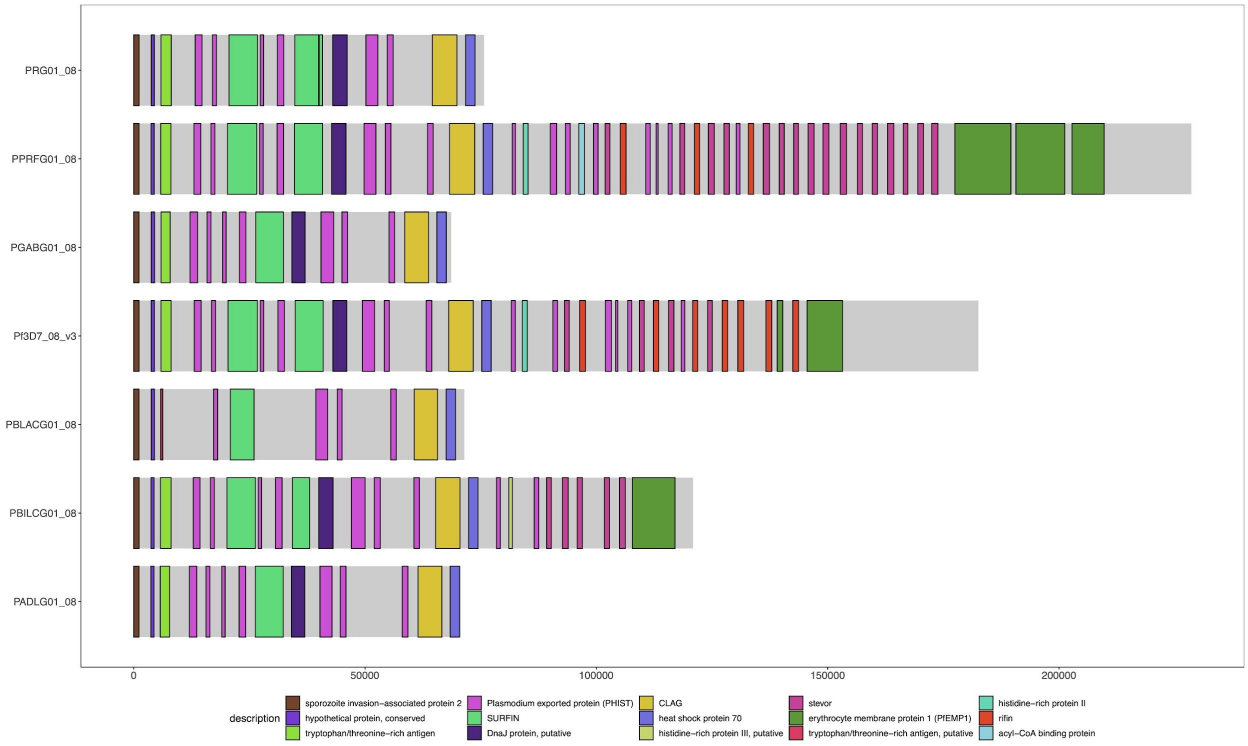


Supplemental Figure 13: **Chromosome 11/13 15.2kb duplicated region for *pfhrp3* deletion pattern $13^{-}11^{++}$ parasites.**

Plotted microhaplotype variation per subgenomic regions across the region shared between all chromosomes 11 and 13, . Across the x-axis are the genomic regions in genomic order, and the genomic region genes are colored on the bottom bar. Y-axis is each parasite with pattern $13^{-}11^{++}$ of *pfhrp3* deletion where this segment of chromosome 11 is duplicated onto chromosome 13. The continent, region, and country are colored per parasite on the left most of the plot. Each column contains the microhaplotypes for that genomic region colored by the microhaplotype rank at that window. If a column is black, there is no variation at that genomic window. Colors are by the frequency rank of the microhaplotypes, and shared colors between columns do not mean they are the same microhaplotype. If there is more than one variant for a parasite at a genomic location, the bar's height is the relative within-parasite frequency of that microhaplotype for that parasite. Pattern $13^{-}11^{++}$ is missing 46,323 bases from chromosome 13 (2,807,159 to 2,853,482) with a gain of 70,175 bases of chromosome 11 (1,933,138 to 2,003,313). Based on genomes that are assembled to the end of their telomeres[30], 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.

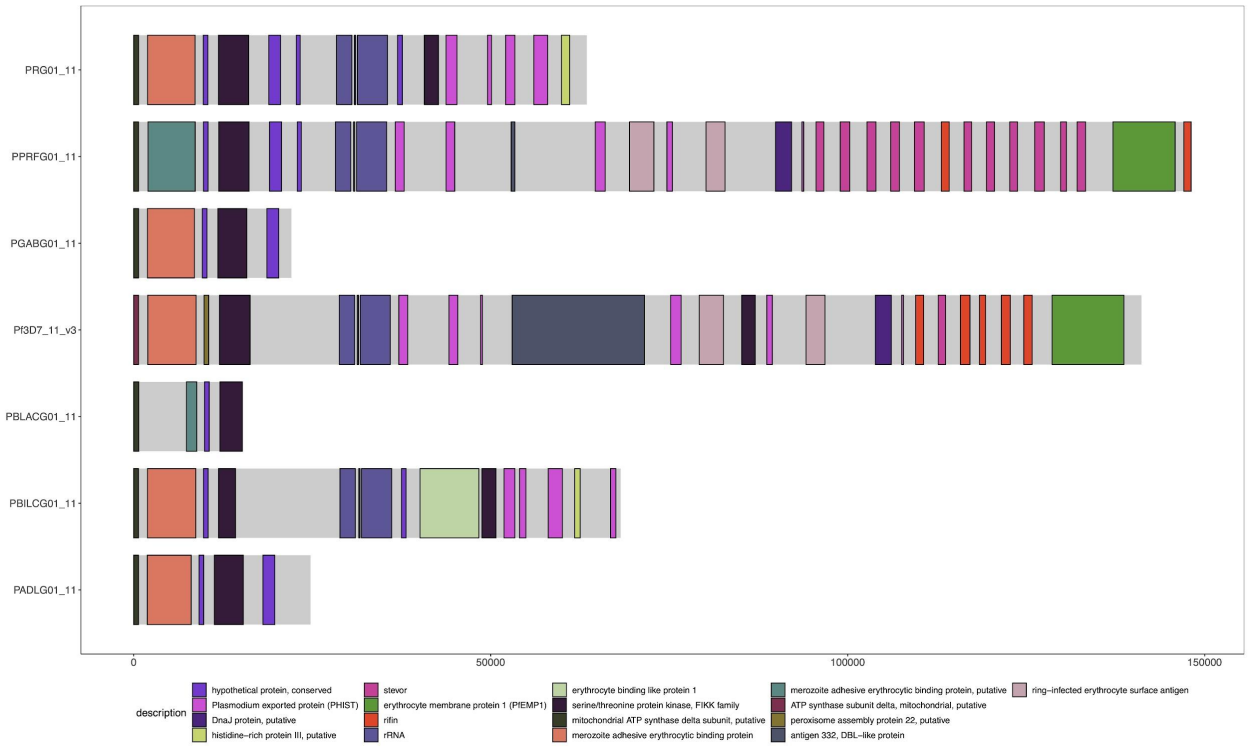


Supplemental Figure 14: Chromosome 11/13 15.2kb duplicated region for *pfhrp3* deletion pattern $13^{-}11^{++}$ parasites with identical chromosome 11 segment haplotypes.
Subset of parasites from **Supplemental Figure 13** for the chromosome 11/13 duplicated region for the parasites with identical chromosome 11 segments based on their microhaplotypes. The left most column contains the groupings based on the microhaplotypes on chromosome 11. There are several parasites with divergent copies of the 15.2kb duplicated region despite the downstream chromosome 11 segments being a perfect copy. This would be consistent with the breakpoint for the duplication event being within this region itself where recombination occurred between nonidentical copies.



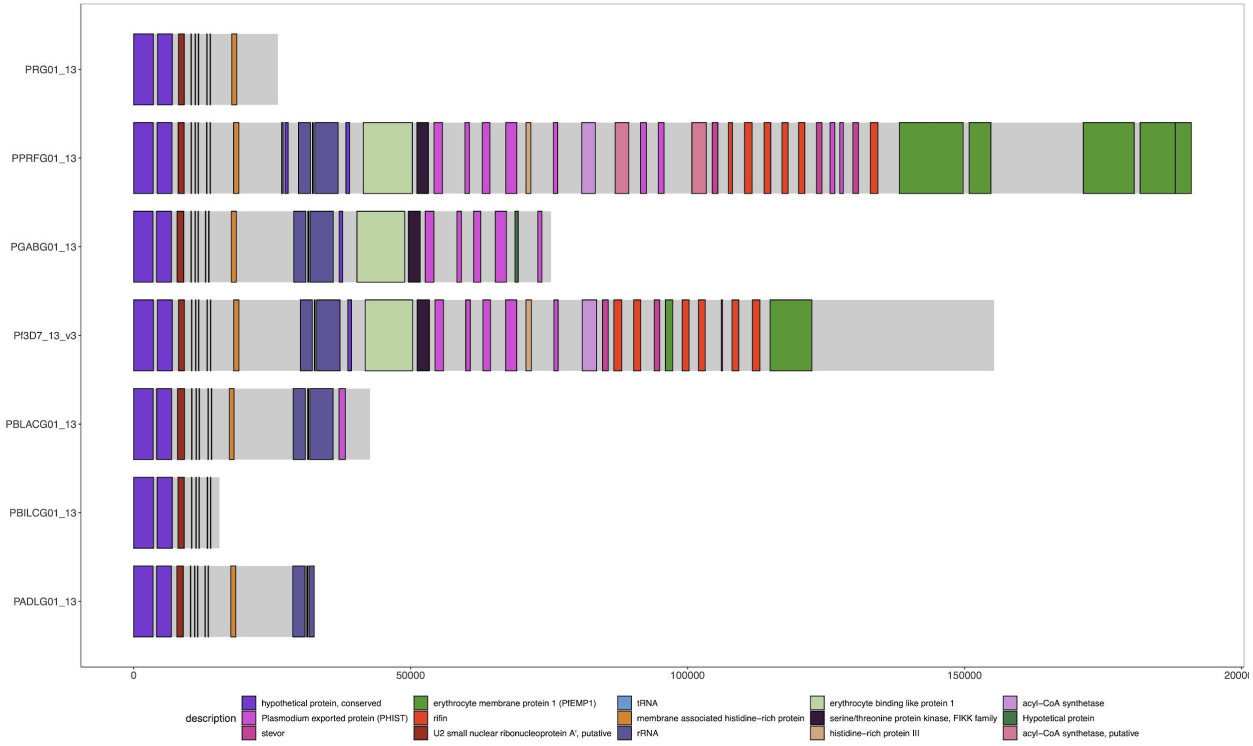
Supplemental Figure 15: Gene Annotations of Chromosome 8 of PacBio-assembled *P. Laverania* Genomes.

Plots of the peri-telomere regions of chromosome 8 across all sequenced *Laverania* Genomes[36]. Assembly of this region is incomplete for the majority of strains, and only *Pf3d7* and the closest relative to falciparum, *PPRFG01*, contain *hrp2*, but they are in similar locations.



Supplemental Figure 16: Gene Annotations of Chromosome 11 of PacBio-assembled *P. Laverania* Genomes

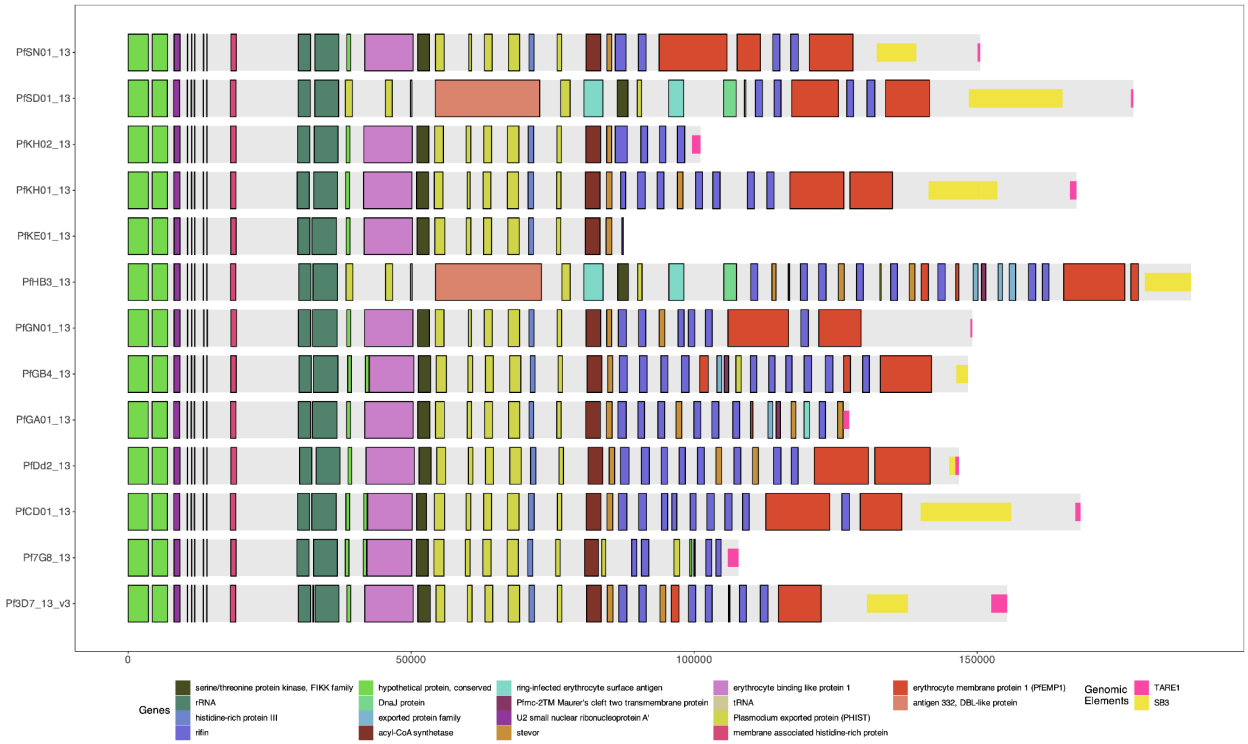
Plots of the peri-telomere regions of chromosome 11 across all sequenced *Laverania* Genomes[36]. Assembly of this region is incomplete for the majority of strains. Plots begin 25kb before the rRNA loci on this region where the duplicated region between chromosomes 11 and 13 is. All strains assemblies that contain this region have this region shared between species and between chromosomes 11 and 13.



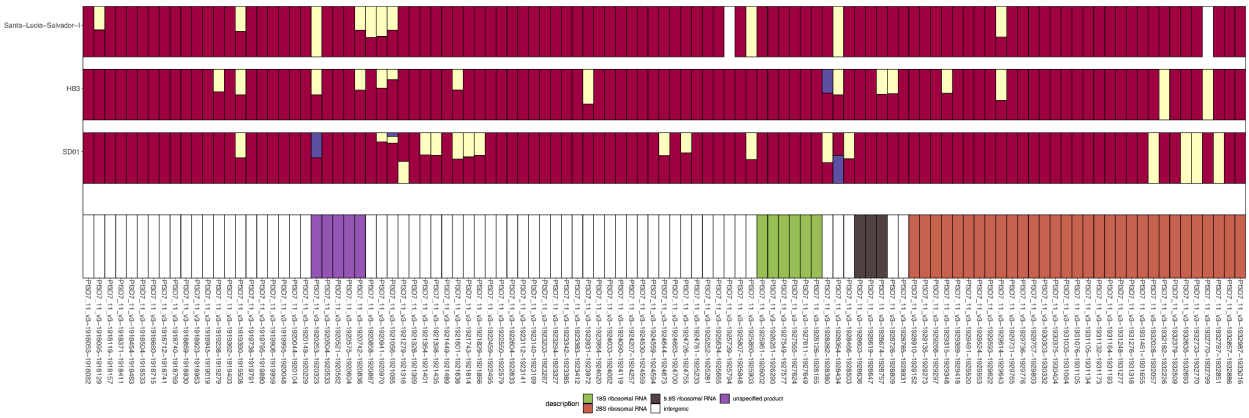
Supplemental Figure 17: Gene Annotations of Chromosome 13 of PacBio-assembled *P. Laverania* Genomes

Plots of the peri-telomere regions of chromosome 13 across all sequenced *Laverania* Genomes[36]. Assembly of this region is incomplete for the majority of strains. Plots begin 25kb before the rRNA loci on this region where the duplicated region between chromosomes 11 and 13 is. All strains assemblies that contain this region have this region shared between species and between chromosomes 11 and 13.

Supplemental Figure 18. Gene Annotations of Chromosome 11 of PacBio-assembled Genomes. The genomic annotations across the 3` telomeric regions of PacBio-assembled genomes[30] across chromosome 11 with the telomere repetitive elements (TAREs) are also shown if present. The presence of TAREs suggests that the assembly has made its way to the sub-telomeric region (end) of the chromosome. The previously published PacBio assembled genomes for SD01 and HB3 did not reach the TAREs for chromosome 11 and terminated in the segmental duplication. 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.

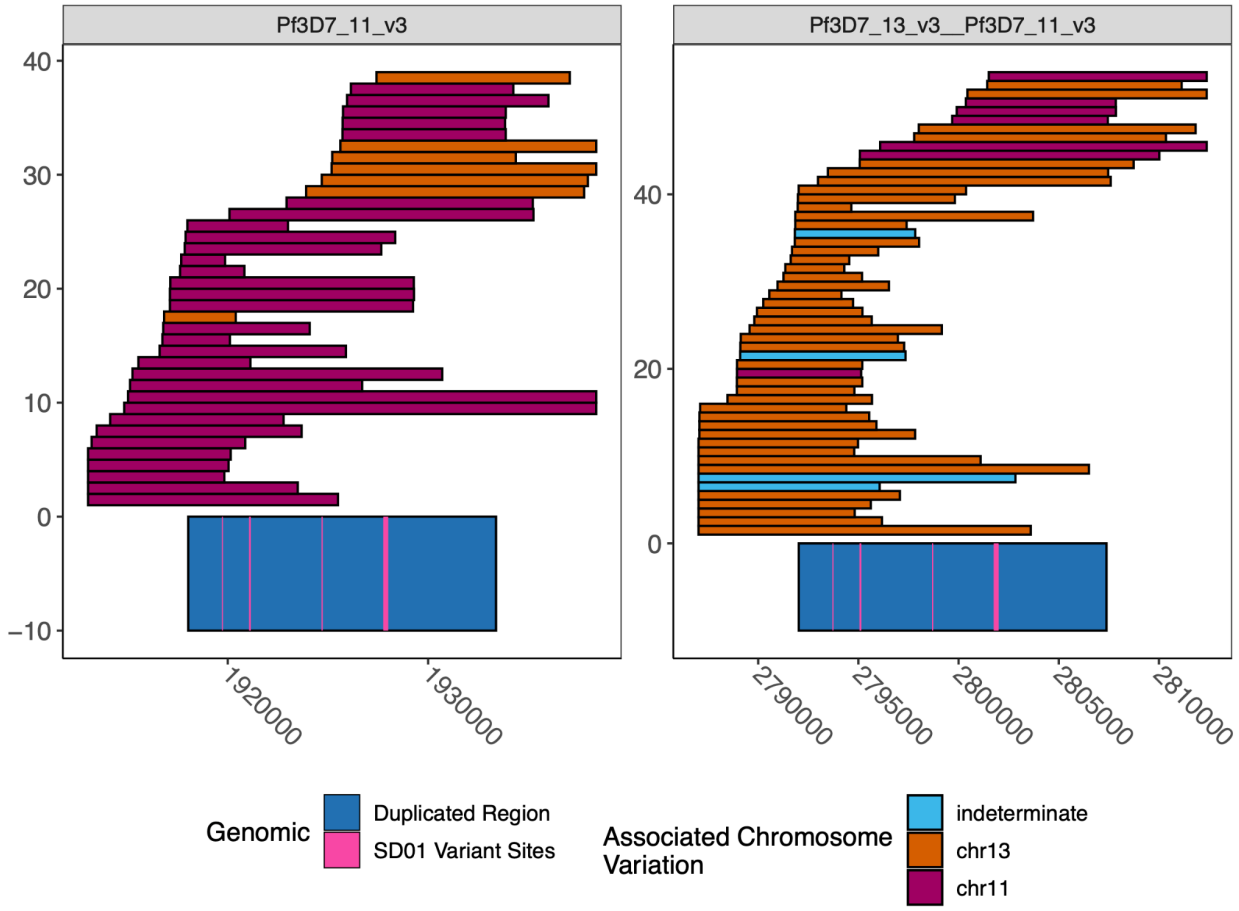


Supplemental Figure 19. Gene Annotations of Chromosome 13 of PacBio-assembled Genomes. The genomic annotations across the 3` telomeric regions of PacBio-assembled genomes[30] across chromosome 13 with the telomere repetitive elements (TAREs) are also shown if present. The presence of TAREs would suggest that the assembly has made its way all the way through the sub-telomeric region for the chromosome. The previously published PacBio-assembled genomes for SD01 and HB3 have sub-telomeric chromosome 11 sequences beginning after the segmental duplication, which is suggestive of a translocation but given the incompleteness of the chromosome 11 assembly in **Supplemental Figure 18** it cannot be determined if this is simply a misassembly or a true translocation.

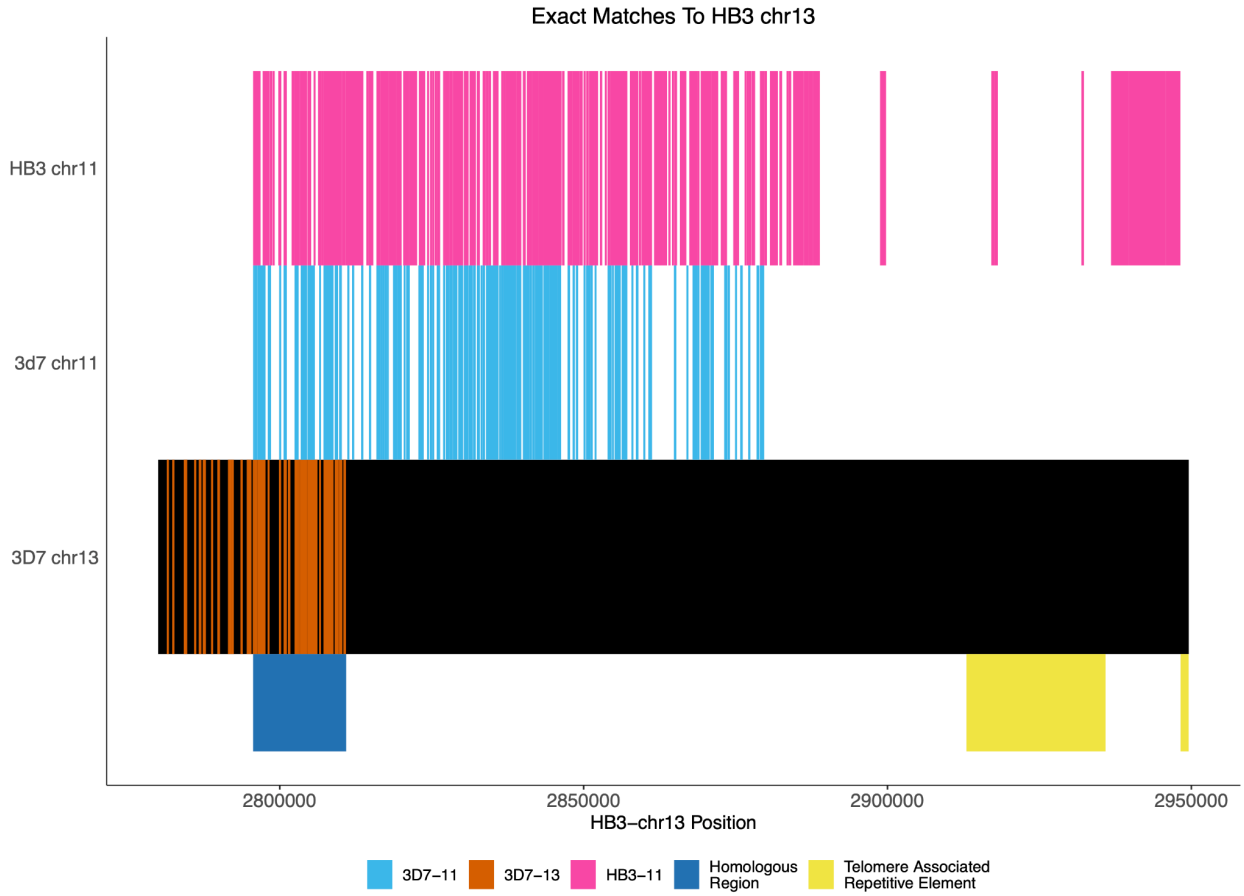


Supplemental Figure 20: Chromosome 11/13 15.2kb duplicated region for parasites SD01, HB3, and Salvador 1. Subset of the parasites from **Supplemental Figure 13** but for SD01 and HB3, which were sequenced in this paper, and for Santa-Luca-Salvador-1, another lab isolate that shows similar *pfhrp3* deletion pattern $13^{-}11^{++}$. SD01 and Santa-Luca-Salvador-1 have

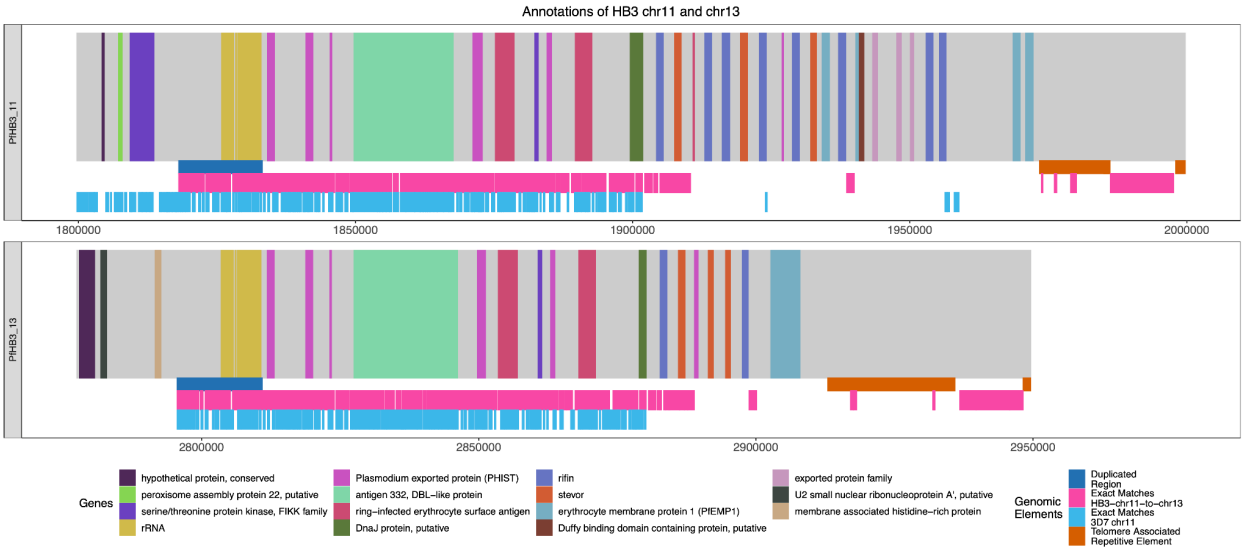
perfect copies, but Santa-Luca-Salvador-1 has variation at 7 loci within the duplicated region, and SD01 has variation at 16 loci within this region. HB3 has divergent copies of the duplicated chromosome 11 segment and also contains variation within this region.



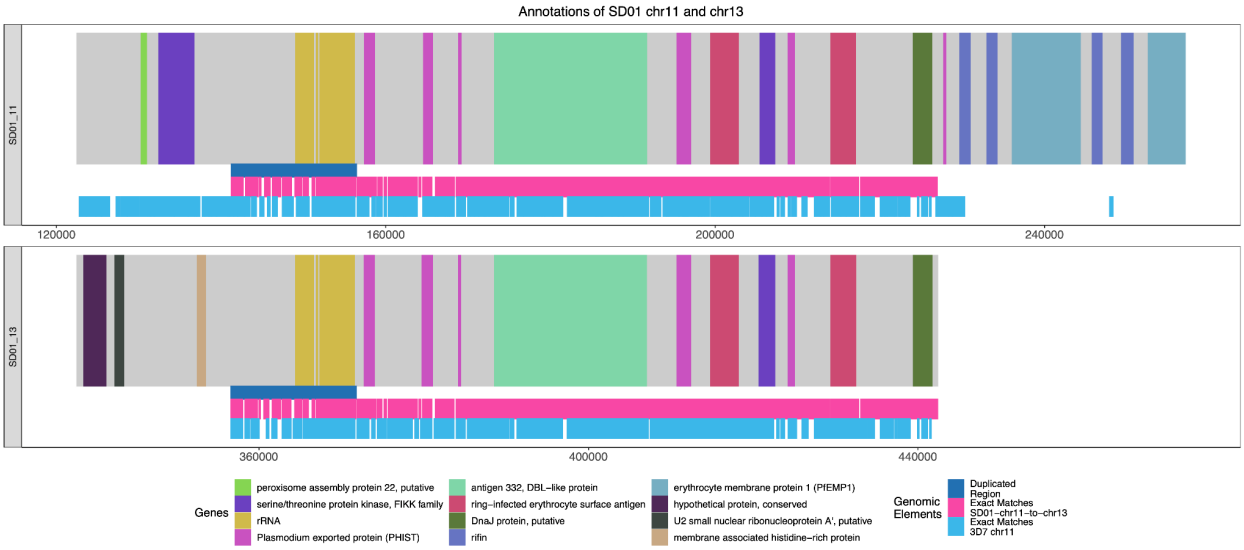
Supplemental Figure 21: Spanning PacBio and Nanopore Reads across the duplicated region for SD01. The spanning Nanopore and PacBio reads across the chromosome 11 and 13 duplicated regions for isolate SD01. The visualization truncates the reads if they span outside of the range shown. The left panel is chromosome 11, and the right panel is the hybrid chromosome 13-11. The chr11/13 duplicated region is colored in dark blue on the bottom of the plot, and the 4 loci where the isolate SD01 has key variation within this region, which can be used to optimize bridging across this duplicated region are colored pink. The reads are colored by the chromosome associated with the variation seen in each read. The association was made by linking the variation found within each of the 4 loci and looking at the reads spanning from each chromosome to see which variants were associated with which chromosome. Each locus had 2 variants and had a strong association with each chromosome.



Supplemental Figure 22: Exact Matches between Nanopore-assembled HB3 chromosome 13 with HB3 chromosome 11, 3D7 chromosomes 11, 13. The locations of exact matches between the Nanopore-assembled HB3 chromosome 13 and between the assembled chromosome 11 as well as the chromosomes of 3D7 11 and 13. The dark blue shaded region shows the location of the duplicated region between chromosomes 11 and 13. The assembled chromosome 13 matches the 3D7 chromosome 13 until this duplicated region and then more closely matches 3D7 chromosome 11 as well as its own chromosome 11. The figure begins 50,000bp before duplicated region, but the new HB3 chromosome 11 matches 3D7 chromosome 11 for the rest of the beginning of the contig.

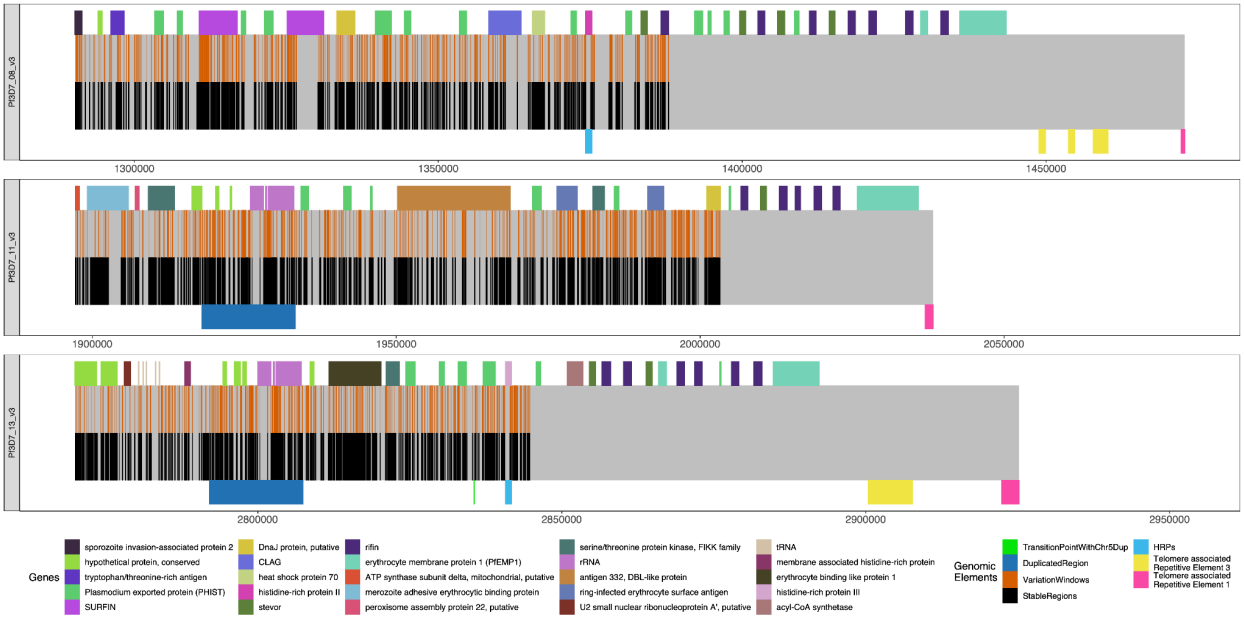


Supplemental Figure 23: Annotation of HB3 chromosomes 11 and 13-11. The new Nanopore assembly of HB3 was annotated by Companion[37], and the ends of chromosomes 11 and 13 are shown above. The duplicated region between chromosomes 11 and 13-11 is shown in blue under each chromosome, and the areas where HB3 chromosomes 11 and 13 have exact matches of at least 31bp are labeled in red underneath. Exact matches of at least 31bp to 3D7 chromosome 11 are shown in green. Both chromosomes end with telomere-associated repetitive elements (TARE), and both end with TARE1, which indicates that both assembled chromosomes reached the end of the telomere.

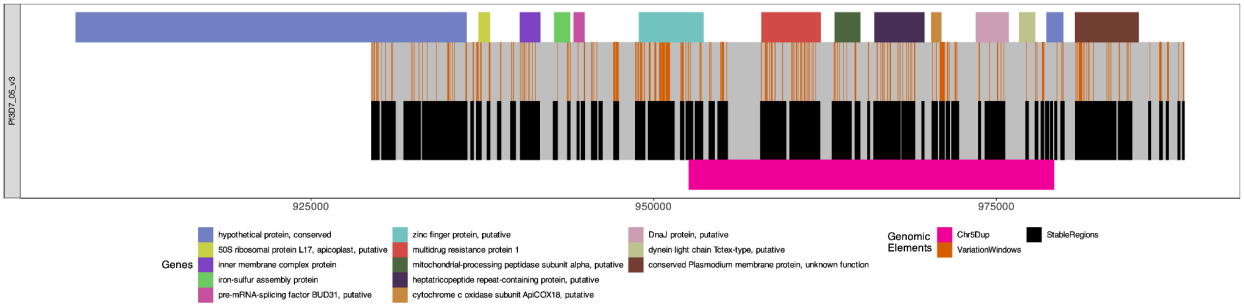


Supplemental Figure 24: Annotation of SD01 chromosomes 11 and 13. The Nanopore assembly of SD01 was annotated by Companion[37], and the ends of chromosomes 11 and 13 are shown above. The duplicated region between chromosomes 11 and 13 is shown in green, and the areas where SD01 chromosomes 11 and 13 have exact matches of at least 31bp are marked out in red underneath. Exact matches of at least 31bp to 3D7 chromosome 11 are shown in green. Due to the low quality of the input DNA of the SD01 parasite, the assembly of

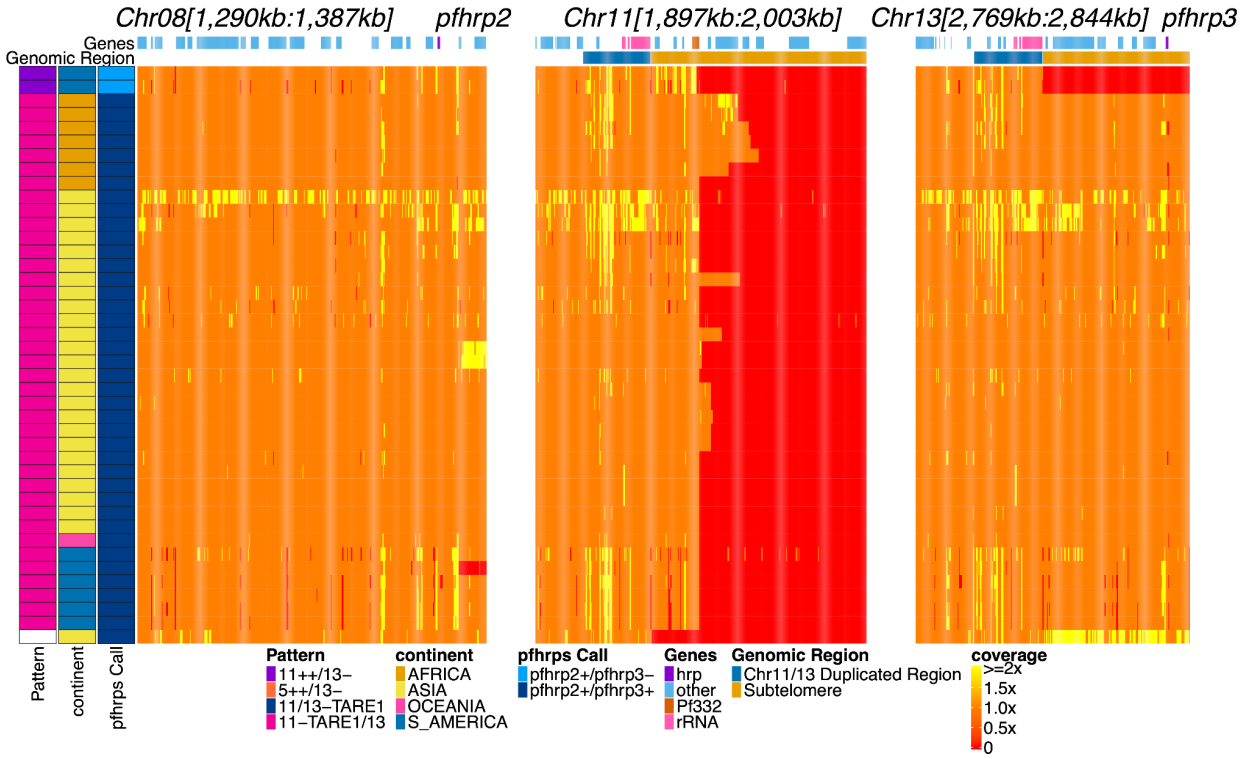
these chromosomes did not reach the end of the telomere given the fact that these assembled contigs did not contain TARE. The assembly of these two chromosomes shows a high degree of similarity from the duplicated region to the end of the 13 associated contig (98.4% similarity with only 1,428 difference over the 89,733 base region).



Supplemental Figure 25: Windows of interest chromosomes 8, 11, 13. The chromosomes are mapped from the beginning of the regions of interest to the chromosomes' ends with all genes/pseudogenes annotations shown colored on top of the gray bars representing the chromosomes. From top to bottom, the regions are from chromosomes 8 (1290239-1387982), 11 (1897151-2003328), 13 (2769916-2844785). The black bars on the bottom half of each chromosome are non-paralogous regions present in all strains, as described in the Methods section. The last black bar to the end of the gray bar represent the sub-telomeric regions. The orange bars on top of the black bars are sub-regions where there is variation that can be used to type the chromosomes. The duplicated region between chromosomes 11 and 13 is shown (dark blue bars below chromosomes 11 and 13) as are the regions containing the *pfhrp* genes (lighter blue bars below chromosomes 11 and 13). The yellow and pink bars on the bottom of the chromosomes represent the telomere-associated tandem repeats found at the end of chromosomes.



Supplemental Figure 26: Windows of interest chromosome 05 around *pfmdr1*
The windows used to investigate the duplication around *pfmdr1* on chromosome 5 associated with the deletion of *pfhrp3*. All genes/pseudogenes annotations are shown on top of the gray bars representing the chromosome region investigated (929384-988747). The black bars on the bottom half of each chromosome are non-paralogous regions present in all strains, as described in the Methods section. The orange bars on top of the black bars are sub-regions where there is variation that can be used to type the chromosome. The pink bar shows the region that is duplicated in pattern $5^{++}13^{-}$.



Supplemental Figure 27: Genome coverage chromosome 8, 11, and 13 regions of isolates with subtelomere deletion of chromosome 11. Sequence coverage heatmap showing involved regions of chromosomes 11 (1,897,157 - 2,003,328 bp), 13 (2,769,916 - 2,844,777 bp), and 5 (944,389 - 988,747 bp) in the subset of the 19,289 parasites along with key lab isolates showing evidence of deletion of putative chromosome 11 subtelomere deletions. There are 42 parasites with evidence of *pfhrp3* deletions. Each row is a parasite. The top annotation along chromosomes depicts the location of genes, and the second row delineates the duplicated region (dark blue) and subtelomere region (orange). The left parasite annotation includes the deletion pattern, continent of origin, and *pfhrp2/3* deletion calls. There were 42 parasites with evidence of sub-telomeric chromosome 11 deletions, 39 of which contained TARE1 sequence where coverage drops to zero, which would be consistent with telomere healing. Only one parasite (lab isolate FCR3) had deletion up and through *pf332* to the ribosomal duplicated region with subsequent duplication of chromosome 13 that would be consistent with the reciprocal of $13^{-}11^{++}$. No field parasites had this pattern. The related clone of FCR3, IT, did not

contain this pattern and would suggest that FCR3 duplicated this segment of chromosome 13 via translocation within culture and not in the field.

References

1. WHO. World malaria report 2022. https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2022 (2022).
2. Chiodini, P. L. Malaria diagnostics: now and the future. *Parasitology* **141**, 1873–1879 (2014).
3. Poti, K. E., Sullivan, D. J., Dondorp, A. M. & Woodrow, C. J. HRP2: Transforming Malaria Diagnosis, but with Caveats. *Trends Parasitol.* **36**, 112–126 (2020).
4. Baker, J. *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).
5. Organization, W. H. & Others. Good practices for selecting and procuring rapid diagnostic tests for malaria. (2011).
6. Cheng, Q. *et al.* Plasmodium falciparum parasites lacking histidine-rich protein 2 and 3: a review and recommendations for accurate reporting. *Malar. J.* **13**, 283 (2014).
7. Thomson, R. *et al.* Prevalence of Plasmodium falciparum lacking histidine-rich proteins 2 and 3: a systematic review. *Bull. World Health Organ.* **98**, 558–568F (2020).
8. WHO. 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. 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 (2021).
9. Feleke, S. M. *et al.* Plasmodium falciparum is evolving to escape malaria rapid diagnostic tests in Ethiopia. *Nat Microbiol* **6**, 1289–1299 (2021).

10. Verma, A. K., Bharti, P. K. & Das, A. HRP-2 deletion: a hole in the ship of malaria elimination. *Lancet Infect. Dis.* **18**, 826–827 (2018).
11. Berhane, A. *et al.* Major Threat to Malaria Control Programs by Plasmodium falciparum Lacking Histidine-Rich Protein 2, Eritrea. *Emerg. Infect. Dis.* **24**, 462–470 (2018).
12. Hinterberg, K., Mattei, D., Wellems, T. E. & Scherf, A. Interchromosomal exchange of a large subtelomeric segment in a Plasmodium falciparum cross. *EMBO J.* **13**, 4174–4180 (1994).
13. Gardner, M. J. *et al.* Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature* **419**, 498–511 (2002).
14. Vernick, K. D. & McCutchan, T. F. Sequence and structure of a Plasmodium falciparum telomere. *Mol. Biochem. Parasitol.* **28**, 85–94 (1988).
15. Mattei, D. & Scherf, A. Subtelomeric chromosome instability in Plasmodium falciparum: short telomere-like sequence motifs found frequently at healed chromosome breakpoints. *Mutat. Res.* **324**, 115–120 (1994).
16. Calhoun, S. F. *et al.* Chromosome End Repair and Genome Stability in Plasmodium falciparum. *MBio* **8**, (2017).
17. Zhang, X. *et al.* Rapid antigen diversification through mitotic recombination in the human malaria parasite Plasmodium falciparum. *PLoS Biol.* **17**, e3000271 (2019).
18. Lee, A. H., Symington, L. S. & Fidock, D. A. DNA repair mechanisms and their biological roles in the malaria parasite Plasmodium falciparum. *Microbiol. Mol. Biol. Rev.* **78**, 469–486 (2014).
19. Reed, J., Kirkman, L. A., Kafsack, B. F., Mason, C. E. & Deitsch, K. W. Telomere length dynamics in response to DNA damage in malaria parasites. *iScience* **24**, 102082 (2021).
20. Conrad, D. F. *et al.* Origins and functional impact of copy number variation in the human genome. *Nature* **464**, 704–712 (2010).
21. Korbel, J. O. *et al.* Paired-end mapping reveals extensive structural variation in the human

genome. *Science* **318**, 420–426 (2007).

22. Kidd, J. M. *et al.* A human genome structural variation sequencing resource reveals insights into mutational mechanisms. *Cell* **143**, 837–847 (2010).
23. Mills, R. E. *et al.* Mapping copy number variation by population-scale genome sequencing. *Nature* **470**, 59–65 (2011).
24. Parks, M. M., Lawrence, C. E. & Raphael, B. J. Detecting non-allelic homologous recombination from high-throughput sequencing data. *Genome Biol.* **16**, 72 (2015).
25. Didelot, X., Méric, G., Falush, D. & Darling, A. E. Impact of homologous and non-homologous recombination in the genomic evolution of Escherichia coli. *BMC Genomics* **13**, 256 (2012).
26. Sander, A. F. *et al.* Multiple var2csa-type PfEMP1 genes located at different chromosomal loci occur in many Plasmodium falciparum isolates. *PLoS One* **4**, e6667 (2009).
27. Sander, A. F. *et al.* DNA secondary structures are associated with recombination in major Plasmodium falciparum variable surface antigen gene families. *Nucleic Acids Res.* **42**, 2270–2281 (2014).
28. Bopp, S. E. R. *et al.* Mitotic evolution of Plasmodium falciparum shows a stable core genome but recombination in antigen families. *PLoS Genet.* **9**, e1003293 (2013).
29. Claessens, A. *et al.* Generation of antigenic diversity in Plasmodium falciparum by structured rearrangement of Var genes during mitosis. *PLoS Genet.* **10**, e1004812 (2014).
30. Otto, T. D. *et al.* Long read assemblies of geographically dispersed Plasmodium falciparum isolates reveal highly structured subtelomeres. *Wellcome Open Res* **3**, 52 (2018).
31. Pologe, L. G. & Ravetch, J. V. Large deletions result from breakage and healing of P. falciparum chromosomes. *Cell* **55**, 869–874 (1988).
32. Scherf, A. & Mattei, D. 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 (1992).

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

34. Nair, S. *et al.* Recurrent gene amplification and soft selective sweeps during evolution of multidrug resistance in malaria parasites. *Mol. Biol. Evol.* **24**, 562–573 (2007).

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

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

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

38. Isabel Veiga, M. *et al.* Globally prevalent PfMDR1 mutations modulate Plasmodium falciparum susceptibility to artemisinin-based combination therapies. *Nat. Commun.* **7**, 11553 (2016).

39. Samarakoon, U. *et al.* The landscape of inherited and de novo copy number variants in a Plasmodium falciparum genetic cross. *BMC Genomics* **12**, 457 (2011).

40. Duvalsaint, M. *et al.* Balanced impacts of fitness and drug pressure on the evolution of PfMDR1 polymorphisms in Plasmodium falciparum. *Malar. J.* **20**, 292 (2021).

41. Vergés, L., Molina, Ò., Geán, E., Vidal, F. & Blanco, J. Deletions and duplications of the 22q11.2 region in spermatozoa from DiGeorge/velocardiofacial fathers. *Molecular Cytogenetics* vol. 7 Preprint at https://doi.org/10.1186/s13039-014-0086-3 (2014).

42. Vera-Arias, C. A. *et al.* High-throughput Plasmodium falciparum hrp2 and hrp3 gene deletion typing by digital PCR to monitor malaria rapid diagnostic test efficacy. *Elife* **11**, (2022).

43. Góes, L. *et al.* 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*

18, (2020).

44. Nair, S., Li, X., Nkhoma, S. C. & Anderson, T. Fitness Costs of pfhrp2 and pfhrp3 Deletions Underlying Diagnostic Evasion in Malaria Parasites. *J. Infect. Dis.* **226**, 1637–1645 (2022).
45. Bertin, G. I. *et al.* Proteomic analysis of Plasmodium falciparum parasites from patients with cerebral and uncomplicated malaria. *Sci. Rep.* **6**, 26773 (2016).
46. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
47. Harris, R. S. Improved pairwise alignment of genomic DNA. (The Pennsylvania State University, 2007).
48. Kasianowicz, J. J., Brandin, E., Branton, D. & Deamer, D. W. Characterization of individual polynucleotide molecules using a membrane channel. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 13770–13773 (1996).
49. Jain, M., Olsen, H. E., Paten, B. & Akeson, M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol.* **17**, 239 (2016).
50. Branton, D. *et al.* The potential and challenges of nanopore sequencing. *Nat. Biotechnol.* **26**, 1146–1153 (2008).
51. Koren, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
52. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019).
53. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
54. R Core Team. R: A Language and Environment for Statistical Computing. Preprint at https://www.R-project.org/ (2022).
55. Wickham, H. *et al. ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. https://CRAN.R-project.org/package=ggplot2 (2022).

56. Gu, Z. *ComplexHeatmap: Make Complex Heatmaps*. (2022).