Non-allelic homologous recombination driven translocation explains histidine-rich protein 3 deletion mechanism in *Plasmodium falciparum*

Nicholas J. Hathaway[1]*, Isaac E. Kim, Jr.[2,3]*, Neeva Wernsman Young[4], Sin Ting Hui[5], Rebecca DeFeo[5], David Giesbrecht[5], Emily Y. Liang[5], Christian P. Nixon[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 Email Addresses:
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 DeFeo: rebecca_defeo@brown.edu
David Giesbrecht: david_giesbrecht@brown.edu
Emily Y. Liang: emily_liang@brown.edu
Christian P. Nixon: christian_nixon@brown.edu
Jonathan J. Juliano: jonathan_juliano@med.unc.edu
Jonathan B. Parr: jonathan_parr@med.unc.edu
Jeffrey A. Bailey: jeffrey_bailey@brown.edu

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

ABSTRACT

Background: Significant progress has been made in the fight against *P. falciparum* malaria due in part to the widespread adoption of rapid diagnostic tests (RDTs) that detect histidine-rich protein 2 (PfHRP2) and its paralog PfHRP3 encoded by the *pfhrp2* and *pfhrp3* genes, respectively. Parasites without *pfhrp2* and *pfhrp3* genes are not detected by these RDTs. *Pfhrp3* loss appears to be more common in some geographical regions and has been observed *in vitro*. We sought to gain insight into geographic patterns of *pfhrp3* deletion and define the mechanism of gene loss.

Methods: Over 9,830 publicly available, whole-genome sequenced (WGS) *P. falciparum* field samples were analyzed for genotypes and coverage using `PathWeaver` for local assembly . DNA from two cultured isolates with *pfhrp3* deletion, HB3 and SD01, were sequenced with Oxford Nanopore Technologies (ONT) long-read sequencing, assembled with `Canu` and `Flye`, and genes annotated with `Companion`.

Results:

Two distinct *pfhrp3* deletion patterns were detected: 1) segmental deletion of chromosome 13 just centromeric to *pfhrp3* extending to the end of the chromosome, with co-occurring segmental duplication of the chromosome 11 subtelomeric region, and 2) segmental deletion of chromosome 13 starting at various locations centromeric to *pfhrp3* without chromosome 11 duplication. Pattern 1 was almost exclusively found in samples from Africa and South America, while pattern 2 was observed predominantly in Southeast Asia. The pattern 1 boundary fell within a 15kb nearly identical duplication on both chromosomes containing ribosomal genes. ONT assembly of HB3 and SD01 parasite lines revealed hybrid chromosomes and long-reads spanning the ribosomal duplication, consistent with recombination between non-homologous chromosomes.

Conclusion: Our findings demonstrate duplication-mediated non-homologous recombination creating a hybrid 13-11 chromosome that replaces *pfhrp3* and telomeric chromosome 13 with a translocated telomeric chromosome 11 sequence--essentially yielding a deletion of chromosome 13 sequence and interchromosomal duplication of chromosome 11 sequence. Given that existing ribosomal duplications likely predispose to the frequent occurrence of this translocation during meiosis, it suggests that subsequent selective forces are driving its presence or absence in different geographical regions. This mechanism appears to explain *pfhrp3* deletion in South America and Africa and may explain why *pfhrp3* deletions are more prevalent than *pfhrp2* deletions in many localities. However, the forces driving emergence of *pfhrp3*- parasites may be complex and encompass other genes affected by the recombination event. Further studies of the origins of *pfhrp2*- and *pfhrp3*-deleted strains and the selective pressures suppressing their occurrence or driving their expansion are needed.

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

INTRODUCTION

In 2020, almost half of the global population was at risk for malaria caused by *Plasmodium falciparum,* one of the most common causes of childhood mortality in Africa [1]. While significant efforts have been made to eradicate the disease, the latest report by the World Health Organization estimated that there were 241 million cases of malaria and 627,000 fatalities from malaria in 2020 alone [2]. Control efforts have mainly focused on vector control such as indoor residual spraying (IRS), more effective diagnosis including rapid diagnostic tests (RDTs), and treatments such as artemisinin-combination therapies [3]. While progress has been made against malaria, *P. falciparum* has remained difficult to control. One major reason is its ability of the parasite to repeatedly infect individuals which in part is due to the generation of antigenic diversity through gene conversion and non-allelic exchange between variant antigen-encoding genes that mainly reside in subtelomeric regions. [4].

Recombination between non-homologous chromosomes and non-allelic positions through misalignment occurs frequently in eukaryotic organisms. For example, Mikus and Petes reported recombination between genes on different chromosomes in *Saccharomyces cerevisiae*, resulting in both non-reciprocal and reciprocal translocations [5]. This recombination mechanism is critical in the evolution of these organisms and is widely recognized for recurrent, high-frequency rearrangements often driven by highly-identical segmental duplications through non-allelic homologous recombination [6–12]. This often leads to further duplications or deletions of the genome and is a major source of plasticity and copy number variations (CNVs) [13, 14]. Importantly, in contrast, some of these NAHR-induced CNVs at key sites can also result in phenotype and disease. Examples in humans include DiGeorge/velocardiofacial syndrome or 22q11.2 duplication syndrome, which depends on whether the deleted or duplicated chromosome is inherited at 22q11.2, respectively [15].

In many eukaryotes, duplications cluster near telomeres or centromeres. For example, studies reported segmental duplications of olfactory receptor genes near the telomeres of the human genome, possibly playing a key role in facilitating functional diversity of the olfactory receptor gene family. In *Plasmodium falciparum,* subtelomeric regions similarly contain gene families (*var*, *rifin*, *stevor*, etc) that are undergoing rapid evolution and frequent NAHR between chromosomes and generating antigenic diversity [16, 17]. While the exact mechanisms of these exchanges has not been studied in depth, analyses suggest that the parasite is likely unable to repair DNA double-strand breaks via traditional canonical non-homologous end joining and instead relies more on homologous recombination [18].

Large recombination events are more robustly studied using long-read sequencing that can better capture their full extent. In contrast to short-read sequencing, ONT's portable MinION sequencer yields sequences up to hundreds of kilobases (kb) in length [19–22]. Furthermore, while Illumina sequencing requires extensive human and capital resources and is typically performed in established laboratories [23], the MinION sequencer is small, cost-effective, and can be run on a standard desktop or laptop, making it an ideal option for real-time sequencing in the field, especially in resource-depleted areas [24, 25]. While ONT sequencing has significant limitations including its relatively high error rate of up to 10% [26], recent studies have used ONT sequencing to reconstruct repetitive sequences and discover structural variation not previously detected with Illumina sequencing [27].

The mainstay of malaria diagnosis across Africa is no longer microscopy but rapid diagnostic tests (RDTs) due to their simplicity and speed. The rapid adoption of these RDTs, coupled with artemisinin-combination therapies, has led to significant progress in the fight against malaria [28]. The predominant falciparum malaria RDTs detect the *P. falciparum* histidine-rich protein 2 (PfHRP2) and are cross-reactive with PfHRP3. The number of PfHRP2-based RDTs increased exponentially from less than 1 million sold in 2004 to more than 300 million in 2013 [29]. A number of studies, however, have reported laboratory and field

isolates with deletions of *pfhrp2* and *pfhrp3*, thereby enabling the parasite to evade diagnosis by PfHRP2-based RDTs [29–32]. These parasites appear to be spreading rapidly in some regions and have compromised existing test-and-treat programs, especially in the Horn of Africa [33–36].

Prevalence of parasites with *pfhrp2* and *pfhrp3* deletion varies markedly across continents in a manner not explained by RDT use alone. Parasites with these deletions are well established in areas where PfHRP2-based RDTs have not routinely been used such as parts of South America as well as those with low transmission [37]. Studies in Ethiopia, where false-negative RDTs owing to *pfhrp2 and pfhrp3* deletions are common, suggest that the *pfhrp3* deletion mutation occurred before that of *pfhrp2* given it is more frequent and shows a shorter haplotype [34]. The reason for *pfhrp3* deletions occurring first given the protein is less detectable by PfHRP2-based RDTs compared to pfhrp2 is puzzling. A 1994 study of a subclone of HB3 reported a *pfhrp3* deletion that was possibly caused by a subtelomeric pairing and exchange between chromosomes of 11 and 13 in mitosis [38]. To our knowledge, there have been no studies that further examined this recombination as a possible mechanism for *pfhrp3* deletion in current circulating parasites. The lack of such studies on *P. falciparum* is likely due to the difficulty of assembling the repetitive and paralogous sequences of its subtelomeres given the limited read length of traditional Illumina short-read sequencing. Understanding the mechanism of these deletions, however, could provide key insight into underlying pressures that allow them to arise and help us better combat their emergence. Thus, the objective of this study was to complement available Illumina genome sequencing data with ONT sequencing to elucidate the mechanism responsible for *pfhrp3* deletion in *P. falciparum* field isolates.

RESULTS

We examined all publicly available whole-genome sequences (WGS) from *P. falciparum* isolates from around the world, comprising 9,830 samples with Illumina data. We analyzed read depth across the genomic regions containing *pfhrp2* and *pfhrp3* after local assembly to detect deletions and duplications. We identified 23 isolates with *pfhrp2* deletion and 153 with *pfhrp3* deletion; global prevalence could not be determined in the absence of representative sampling. *Pfhrp3* deletions outnumbered *pfhrp2* deletions in all three geographic regions, particularly in South America and the Horn of Africa.

Two distinct pfhrp3 deletion patterns with geographical associations.

Analysis of the publicly available WGS data revealed two distinct *pfhrp3* deletion patterns based on read depth analysis: 1) segmental deletion of chromosome 13 just centromeric to *pfhrp3* extending to the end of the chromosome, with co-occurring segmental duplication of the chromosome 11 subtelomeric region, both proximal to a 15.2kb homologous region shared by chromosomes 11 and 13 (pattern 1), and 2) segmental deletion of chromosome 13 starting at various locations centromeric to *pfhrp3* and without duplicated chromosome 11 region (pattern 2) (**Figure 1**). Among the 153 field samples with *pfhrp3* deletion, 107 and 46 demonstrated patterns 1 and 2, respectively. Pattern 1 was exclusively found in samples from Africa and South America, while pattern 2 was observed almost exclusively in Southeast Asia (**Figure 1**).

Pattern 1 breakpoint resides in ribosomal-gene segmental duplication on chromosomes 11 and 13.

For pattern 1, the increased chromosome 11 coverage consistent with segmental duplication began within a 15.2-kb, highly-homologous (98.0% nucleotide identity), pre-existing segmental

duplication on chromosomes 11 and 13 (1,918,028 to 1,933,288 on 3D7 chromosome 11 and 2,792,021 to 2,807,295 on 3D7 chromosome 13). The duplicated regions are oriented similarly on both chromosomes and consist of a centromeric 7kb region encoding 2 undefined product proteins (98.9% identity) and a telomeric 8kb region containing one of the two S-type [16] ribosomal loci (99.7% identity) (**Figure 2**). *Pfhrp3* (2,840,236 to 2,842,840) is telomeric to the copy on chromosome 13. In an all-by-all comparison of the chromosomes of the 3D7 genome using `nucmer` [39] to perform unique k-mer sharing, this homologous region is the largest inter-chromosomal duplication in the core genome and second only to a 23.5 kb region between subtelomeric regions on the 5' ends of chromosomes 5 and 13, which includes the start of the telomere itself and a singular *var* gene. Together, the pattern of chromosome 11 duplicated segment suggested that this ribosomal segmental duplication could be mediating a translocation through non-allelic homologous recombination.

Pattern 1 is observed mainly in South America and Africa.

Pattern 1 was observed in 80 South American parasites, 38 African parasites (20 in Ethiopia, 6 in Senegal, 1 Mali, and 1 Ghana), and 2 Asian parasites (1 Cambodia, 1 Thailand) (**Figure 1**). Out of the 110 parasites with this deletion pattern, 89 parasites (77 of the 80 South American, 11 of the 28 African parasites, and 1 of the 2 Asian) had near-identical copies of the chromosome 11 duplicated region. Near-identical copies were defined as having >=99% variant sites (n=382) within the duplicated region identical (same variant allele between copies). The remaining 21 parasites had variation within this region; on average, 10.2% of variant sites differed between the copies (min 83.8% identity). The segmental duplication of chromosome 11 had several distinct haplotypes, with distinct haplotypes observed in South American and African strains. The segment of chromosome 11 found within parasites with the duplication can also be found within parasite with only a single copy of this segment of 11 and is not only associated with

duplication (**Supplemental Figure 5**, **Supplemental Figure 6**, **Supplemental Figure 7**, **Supplemental Figure 8**, **Supplemental Figure 9**, **Supplemental Figure 10**). Given the significant diversity between the haplotypes within the duplicated region of chromosome 11, there are either multiple translocation events and duplication of this region is ongoing, or there was a very distant duplication event that has since significantly diverged between strains.

Pattern 2 is predominantly found in Asia.

For pattern 2, there were several regions centromeric to *pfhrp3* where the deletion occurred. All deletions appeared to occur after a FIKK family protein gene PF3D7_1371700 and occurred between the 4 PHIST protein genes located between PF3D7_1371700 and *pfhrp3*. The majority of the breaks occurred before the first PHIST (PF3D7_1372100) protein centromeric to *pfhrp3* (**Figure 1**). This deletion type was seen primarily in Southeast Asia with 42 in Southeast Asia - East, 2 in Southeast Asia - West, 1 in East Africa and 1 in West Africa (**Figure 1**).

Previous PacBio assemblies

Given that the pattern of deletion and duplication and the breakpoints were suggestive of duplication-mediated recombination leading to translocation, we sought to examine high-quality PacBio genome assemblies of other strains containing the *pfhrp3* deletion. However, the gene annotations by `Companion` [40] of chromosome 11 of previous PacBio assemblies [41] showed that the assemblies of strains SD01 and HB3 lacked the sub-telomeric region on chromosome 11, instead this region was found only on chromosome 13 (**Supplemental Figure 1**, **Supplemental Figure 2**). The gene annotations of chromosome 13 of the assemblies found that the genes of a segment in chromosome 13 centromeric to *pfhrp3* corresponded to those of a subtelomeric region in chromosome 11, providing evidence of a possible segmental duplication (**Supplemental Figure 2**). Due to the absence of an assembled sub-telomeric

chromosome 11, however, these assemblies did not definitively resolve the mechanism behind the deletion of *pfhrp3*.

Representative strains from Southern America and Horn of Africa have long-reads spanning the homologous region of the normal chromosome 11 and hybrid chromosome 13-11

To confirm that pattern 1 involves a hybrid chromosome 13-11 chromosomal product consistent with a breakpoint within the homologous region described above, we further examined two parasite isolates with known *pfhrp3* deletion, HB3 from Honduras and SD01 from Sudan. We used the ONT MinION to sequence 7.35 gigabases of HB3 and 6.46 megabases of SD01 for an average coverage of 319x and 29.3x, respectively. We combined our ONT data with the publicly available PacBio sequencing data. To test for the presence of hybrid chromosomes we used a two prong approach 1) mapping raw long reads and 2) *de-novo* assembly of long reads into contigs.

We constructed 3D7 representations of hybrid chromosomes 13-11 and 11-13 by joining the fasta 3D7 chromosomes with breakpoints in the middle of the ribosomal duplication. We then mapped our Nanopore reads and the publicly available PacBio reads for HB3 and SD01 to 3D7's normal chromosome 11 and 13, hybrid 13-11, and hybrid 11-13 and identified reads that completely spanned the homologous region on these normal chromosomes and hybrid constructs. Reads were considered "spanning" if they extended at least 50 bp into the flanking unique regions. HB3 had 77 spanning reads across normal chromosome 11 and 91 spanning reads across hybrid chromosome 13-11, hereinafter referred to as 13-11 spanning reads. SD01 had 2 chromosome 11 spanning reads and 1 13-11 chromosome spanning read. Neither strain had PacBio or Nanopore spanning reads across normal chromosome 13 and hybrid 11-13 (**Figure 4**) that would be consistent with the reciprocal meiotic product. Although SD01 had a

small number of reads that spanned the full region, sub-analysis revealed 4 regions within this homologous region that had chromosome 11 and 13 specific nucleotide variation that was able to be used to further bridge across this region (**Supplemental Figure 14**, **Supplemental Figure 15**). The other 14 strains with *pfhrp3* intact from the PacBio dataset [41] had reads that spanned chromosome 11 and chromosome 13 across these regions but lacked any spanning reads across the hybrid 13-11 or 11-13 chromosomes.

ONT long-read assemblies

To further prove the presence of the hybridized 13-11 chromosome, *de novo* assemblies were created for the HB3 and SD01 lab strains. HB3 assembly yielded complete chromosomes with telomere-associated tandem repeat 1 (TARE-1) [17] detected on the ends of all chromosomes except for the 3' end of chromosome 7 and the 5' end of chromosome 5, indicating that close to entirety of the telomeric regions of HB3 was achieved. SD01, however, likely due to its lower quality input DNA and consequent lower sequencing coverage, had a much more disjointed assembly with 200 final contigs, N50 263,459 and L50 30. The HB3 assembly had a chromosome 11 that closely matched 3D7 chromosome 11 and a hybrid 13-11 that closely matched 3D7 chromosome 13 up to the ribosomal duplication region where it then subsequently best matched the 3D7 chromosome 11 (**Figure 3a**). Both of these chromosomes had TARE-1 [17] at their 5' and 3` end indicating that these were full length chromosome assemblies. The SD01 had two contigs - one that matched 3D7 chromosome 11 and the other that matched 3D7 chromosome 13 up to the homologous region and then subsequently matched 3D7 chromosome 11. These new assemblies were then annotated for genes by `Companion` [40]. The contig matching the hybrid chromosome 13-chromosome 11 for both strains had the genes normally found on chromosome 11 after the homologous region. The duplicated genes on chromosomes 11 and 13-11 that shared homology with 3D7 chromosome 11 included

Pf332(PF3D7_1149000), 2 ring erythrocyte surface antigens genes (PF3D7_1149200, PF3D7_1149500), 3 PHISTs genes, a FIKK family gene, and 2 hypothetical proteins and ends with a DnaJ gene (PF3D7_1149600) corresponding to 3D7 genes PF3D7_1148700 through PF3D7_1149600 (**Figure 3a**, **Figure 3b**). The number of DnaJ genes in the 3D7 genome is close to 27 but shared low homology. The closest DnaJ gene to PF3D7_1149600 shared 67% sequence identity. Homology between HB3 chromosome 11 and chromosome 13-chromosome 11 continued up through a *rifin*, then a *stevor* gene, and then the sequence completely diverged afterward with even a different gene family organization structure. The chromosome 13-11 SD01 contig reached the DNAJ protein and terminated (**Figure 3b**), likely because the assembly was unable to contend with the near complete identical sequence between chromosome 11 and chromosome 13-11. Illumina sequences mapped to the *var* regions on SD01 chromosome 11 where chromosome 13-11 terminated, and local assemblies revealed no variation but double coverage as compared to the rest of the genome, which would be consistent with this region being perfectly copied. Therefore, it's likely the SD01 clone has identical copies up and through the telomere of these two chromosomes. Analysis of the 11 other PacBio assemblies [41] with normal chromosome 11 showed that homology between strains also end at this DnaJ gene (PF3D7_1149600) with the genes immediately following being within the *stevor*, *rifin* and *var* gene families among other paralogous gene families. The genes on chromosome 13 but deleted in the hybrid chromosome 13-11 corresponded to 3D7 genes PF3D7_1371500 through PF3D7_1373500 including *pfhrp3*(PF3D7_1372200) and *EBL-1*(PF3D7_1371600).

Chr 11/13 homologous region found within all subgenus Plasmodium laverania.

The *P. laverania* genomes [42] were not fully assembled within their sub-telomeric regions. *P. praefalciparum*, *falciparum's* closest relative which diverged about 50,000 years ago, also

contained the similar S-type rRNA loci on chromosomes 11 and 13 and had a similar gene structure to *falciparum* in these regions as well as the region on chromosome 8 close to *pfhrp2* (**Supplemental Figure 16**, **Supplemental Figure 17**, **Supplemental Figure 18**). *P. praefalciparum* also contained the 15.2kb homologous region on both chromosomes 11 and 13 and was 96.7% similar to the 3D7 homologous region.

DISCUSSION

Here we use publicly available short-read and newly generated long-read sequencing data to identify mechanisms by which *pfhrp3* deletion commonly occurs. Our findings are consistent with a translocation event that appears to be the predominant mechanism in South America and the Horn of Africa. We confirm the presence of a hybrid 13-11 chromosome and a translocation breakpoint within a high-identity duplication containing ribosomal genes that resides on chromosomes 11 and 13, consistent with a mechanism of non-allelic homologous recombination. The presence of HB3 and SD01 spanning reads across the homologous region of chromosomes 11 and 13-11, as well as the assembled complete genome of HB3 and the partially assembled SD01 genome, indicate an interchromosomal segmental duplication of the subtelomeric portion of chromosome 11 onto that of chromosome 13 via non-homologous recombination. This recombination event creates a hybrid 13-11 chromosome and results in the deletion of *pfhrp3* and its surrounding genes from the homologous region and onwards (**Figures 3 and 4**). Such an event is consistent with the genomic coverage pattern we observed in publicly available data from 110 *pfhrp3*-deleted field samples (**Figure 1**). This non-homologous recombination mechanism resulting in the translocation and subsequent *pfhrp3*-deletion has significant biological and public health implications.

This duplication-mediated non-allelic homologous recombination event is likely to occur frequently during meiosis. In human disease, meiotic misalignment in NAHR can lead to high-frequency chromosomal aberrations (eg. 22q11 deletion syndrome due to misalignment of duplicated blocks on chromosome 22 occurs in 1 in 4000 births). In the parasite, the chromosome 11 and 13 translocation is likely a common occurrence within parasite populations around the world. These events probably result in decreased fitness, and as a result, are removed quickly from the population. 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.

Mok et al. reported on a segmental duplication of six functional genes unique to *P. falciparum* across the subtelomeric regions of chromosomes 1, 2, 3, 6, 7, 10, and 11 [43]. The authors were unable to determine the biological significance of these copy number variant segmental duplications, but given the conservation of these genes in otherwise highly varying subtelomeric regions, they conjectured that this segmental duplication had a functional role. In this study, the segmental duplication on chromosome 11 and 13-11 likely has multiple significant functional consequences given the number of genes affected. Not only do these regions contain members of multi-gene families of *rifins, stevors,* and *var* but also *pf332* on chromosome 11, which encodes for a large Duffy-binding-like domain containing a 27KDa protein [44–46]. *Pf332* has been found to be essential for the binding of the Maurer cleft to erythrocyte skeleton and is highly expressed in patients with cerebral malaria [47]. Lack of this protein likely leads to a large detriment to the survival of the parasite and may be the reason the reciprocal hybrid 11-13 is not observed. Other genes in this area include two ring-infected erythrocyte surface antigen (RESA) genes; there is only one other gene of this family which is located on chromosome 1. Little is known about these two genes. There is also a DnaJ gene within this duplicated region. There are about 30 genes in this family though this sequence is only 67% similar to the closest related DnaJ protein.

The benefit in South America and the Horn of Africa conferred by the segmental duplication of chromosome 11 and 13-11 may not only be due to duplications of certain genes but deletions of them, with the most obvious deleted gene being *pfhrp3*. Besides *pfhrp3, EBL-1*, which helps enable *P. falciparum* to invade erythrocytes, is lost, which may cause infected patients to have fewer symptoms and be less likely to seek treatment [34, 48]. Analysis of one of the closest relatives of *P. falciparum*, *P. reichenowi,* shows a lack of the EBL-1 protein, which suggests its presence might not be essential to parasite survival. *P. reichenowi* contains all the genes within the duplicated chromosome 11, suggesting that these genes might be more essential than the chromosome 13 genes. Thus, the loss of the *pfhrp3* and *EBL-1* genes may increase *P. falciparum*'s fitness [34]. Only parasites with a duplicated chromosome 11 and deleted chromosome 13 were detected in our analysis in the field samples; the complementary situation of segmental duplicated chromosome 13 and deleted chromosome 11 was never observed in any of the field samples. Only lab isolate FCR3 had any indication from coverage data that it had a duplicated chromosome 13 and a deleted chromosome 11. Given that the majority of the publicly available field samples were collected from studies using RDT-positive samples and that PfHRP3 encoded by duplicated *pfhrp3* would likely have been detected by RDT, samples should not have biased against detecting the reciprocal hybrid 11-13 (which would have 2 copies of *pfhrp3*). The apparent absence of parasites with this deletion/duplication pattern would be consistent with several scenarios: a benefit of duplicated chromosome 11, a significant detriment to an absence of chromosome 11, a benefit of lacking chromosome 13 (e.g. lack of *pfhrp3* and decreased RDT-based treatment,) or no benefit of a duplicated chromosome 13. The fact that there was not a single strain detected with this pattern seen in areas with large amounts of hybrid chromosome 13-11 suggests that evolutionary pressures are strong enough to keep the population of such strains at an undetectable level.

While Hinterberg et al. proposed that a general mechanism of non-homologous recombination of the subtelomeric regions may be responsible for the deletion of *pfhrp3*, our

findings indicate that this event involved the segmental duplication and translocation of chromosome 11 at bases 1,917,992 to 2,003,604 onto chromosome 13 from base 2,792,047 until the end [38]. The duplicated region of chromosome 11 has significant diversity between parasites with this duplication indicating that this duplication event is either an event that has been repeated several times or has an ancient origin. Pattern 2 involves a segmental deletion of chromosome 13 including *pfhrp3* without duplication of chromosome 11. Interestingly, there is a clear distinctive geographical pattern seen in the *pfhrp3* deletion patterns of the samples analyzed here. Pattern 1 is seen only in South America and Africa and Pattern 2 in South East Asia. It is not clear why a mechanistic deletion pattern would be geographically distributed in such a way.

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

METHODS

Genomic locations and read depth analysis

In order to study genomic deletions in the unstable and frequent recombinant subtelomeric regions surrounding *pfhrp2* and *pfhrp3*, stable regions surrounding these genes were determined. Stable genomic regions were created by selecting regions that were conserved between all strains without *pfhrp2/3* deletions and did not contain paralogous regions in order to best determine where possible deletions were occurring. This was accomplished by first marking the 3D7 genome with the program `tandem repeat finder` [49], then taking 200bp windows, stepping every 100bp between these tandem repeats, and blasting these regions against currently available chromosomal-level PacBio assembled genomes [41] that did not contain known deletions of *pfhrp2/3*. Regions were kept if they hit each genome only once, and overlapping regions were then merged. Regions from within the duplicated region on chromosome 11 and chromosome 13 were kept if they hit either chromosome 11 and/or 13 but not if they hit other chromosomes. This served to eliminate paralogous regions outside of 11 and 13, especially within the rRNA loci that were similar to the other S-type rRNAs on chromosome 1. Local haplotype reconstruction was performed on each region for 9,830 publicly available whole genome sequences (WGS) of *P. falciparum* field samples including 24 samples during a recent study in Ethiopia [34] where *pfhrp2-/3-* parasites are common. The called haplotypes were used to determine which subregions contained variation in order to type each chromosome (**Supplemental Figure 4**). Coverage was determined for each genomic region and then normalized to the coverage of the whole genome. Windows were examined on chromosomes 8, 11, and 13 starting from these positions to the end of each chromosome: chromosome 8 1,290,239, chromosome 11 1,897,151, and chromosome 13, 2,769,916. All coordinates in the manuscript are relative to *P. falciparum* 3D7 genome version 3 (version=2015-06-18).

Genomic structure

To investigate the genomic landscape around the *pfhrp* genes, an all-by-all comparison of each of 3D7's chromosomes 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. These larger regions were then grouped by taking regions that were located at least 1000bp within each other.

Long read sequences

All PacBio reads for the following 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) [41]. To supplement these reads and to improve upon previous assemblies that were unable to successfully assemble chromosomes 11 and 13, these strains were also sequenced with Oxford Nanopore Technologies' MinION device for long-read DNA and RNA sequencing [50–52]. 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` [53], an assembly pipeline for high-noise single-molecule sequencing, and `Flye` [54] using default settings. In order to assemble the low coverage and highly similar chromosome 11 and chromosome 13 segments of SD01, two assemblies were performed with `Flye` using chromosome 13 specific reads and chromosome 11 specific reads to get contigs

that represented the chromosome 11 and 13 segments. HB3 was assembled using the `Canu` assembler with default settings. The PacBio/Nanopore reads were mapped to reference genomes using `Minimap2`, a sequence alignment program [55]. Mappings were visualized using custom R scripts [56].

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 were examined [42]. The genomes of all *Laverania* have recently been sequenced and assembled using PacBio and Illumina data [42]. The assemblies were analyzed using their annotations and by using `LASTZ[57]` with 80% identity and 90% coverage of the genes in the surrounding regions on chromosomes 8, 11, and 13.

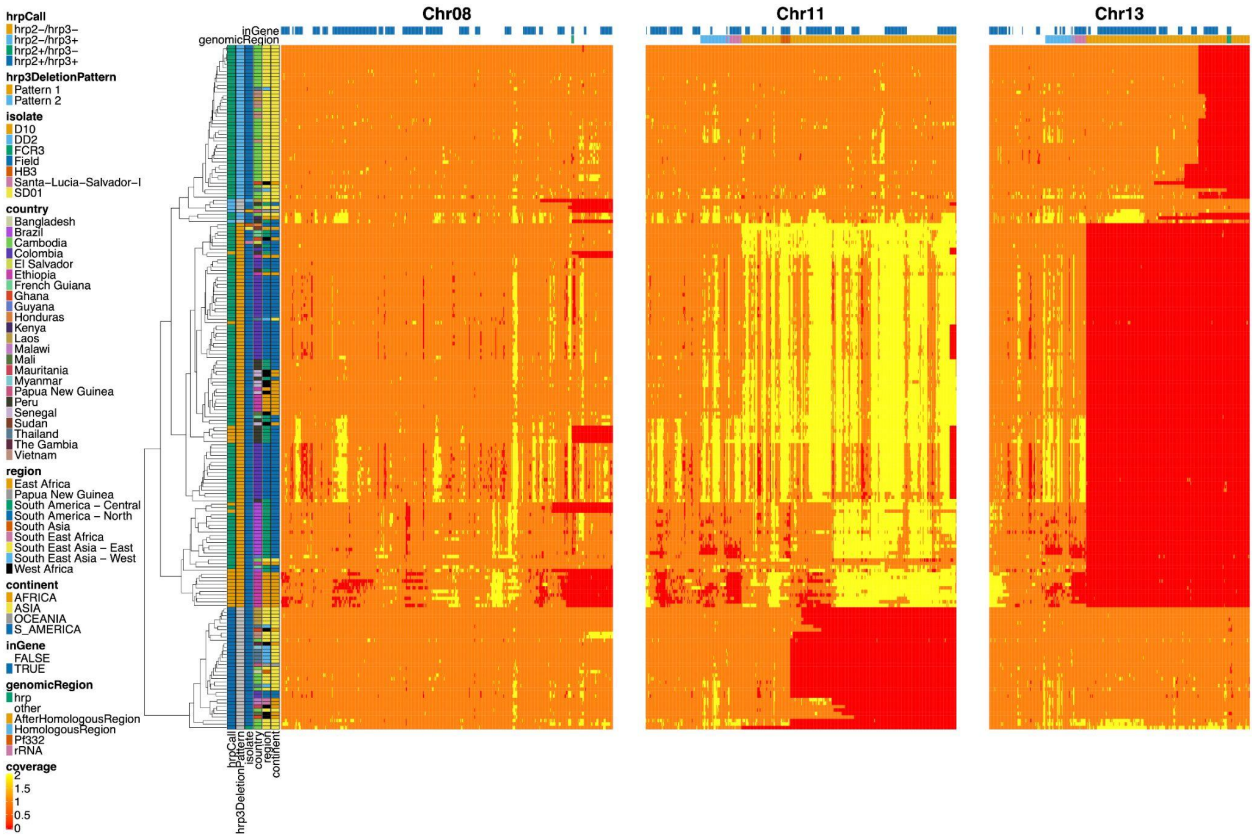


Figure 1. Sequence coverage across chromosomes 8, 11, and 13

The heatmap shows the relative normalized sequence coverage of stable regions, as defined by those conserved between all strains without *pfhrp2/3* deletions and not containing paralogous regions, within 3` telomeric regions of chromosomes 8, 11, and 13 across 9,830 publicly available global *P. falciparum* WGS samples subsetted to samples with possible genomic deletions of at least one of the following: *pfhrp2* deletion (chr. 8), *pfhrp3* deletion (chr. 11), and/or deletion of portion of chr 11. Each row is a sample and each column is a genomic region in genomic order. The top row annotation depicts whether the regions are within genes, while the second row delineates the locations of the homologous region, *pf332*, rRNA, and the *hrp* genes. The left annotation includes the country, region, and continent of sample origin; *pfhrp2/3* deletion calls; and identification of well-characterized strains. Two distinctive patterns can be seen for chr. 11 *pfhrp3* deletion with either a break centromeric to *pfhrp3* but occurring telomeric to the homologous region or a break at the homologous region with a complementary increase in coverage on chromosome 11 suggesting segmental deletion of chromosome 13 and co-occurring segmental duplication of 11. Various other coverage patterns are seen (e.g. partial zero coverage of *Pf332*), but this is likely an artifact of *P. falciparum* selective whole-genome amplification (sWGA) enrichment performed prior to sequencing.

Figure 2. Gene annotation of a large 15.2kb homologous region shared by chromosomes 11 and 13. The y-axis is the chromosome position of 3D7 chromosome 11, and the x-axis is the chromosome position of 3D7 chromosome 13. Black bars plotted diagonally are regions that are 100% conserved between chromosomes (>30bp segments). Conserved segments comprise 89.1% of the whole region, while the alignment identity between these two segments is up to 99.3%.

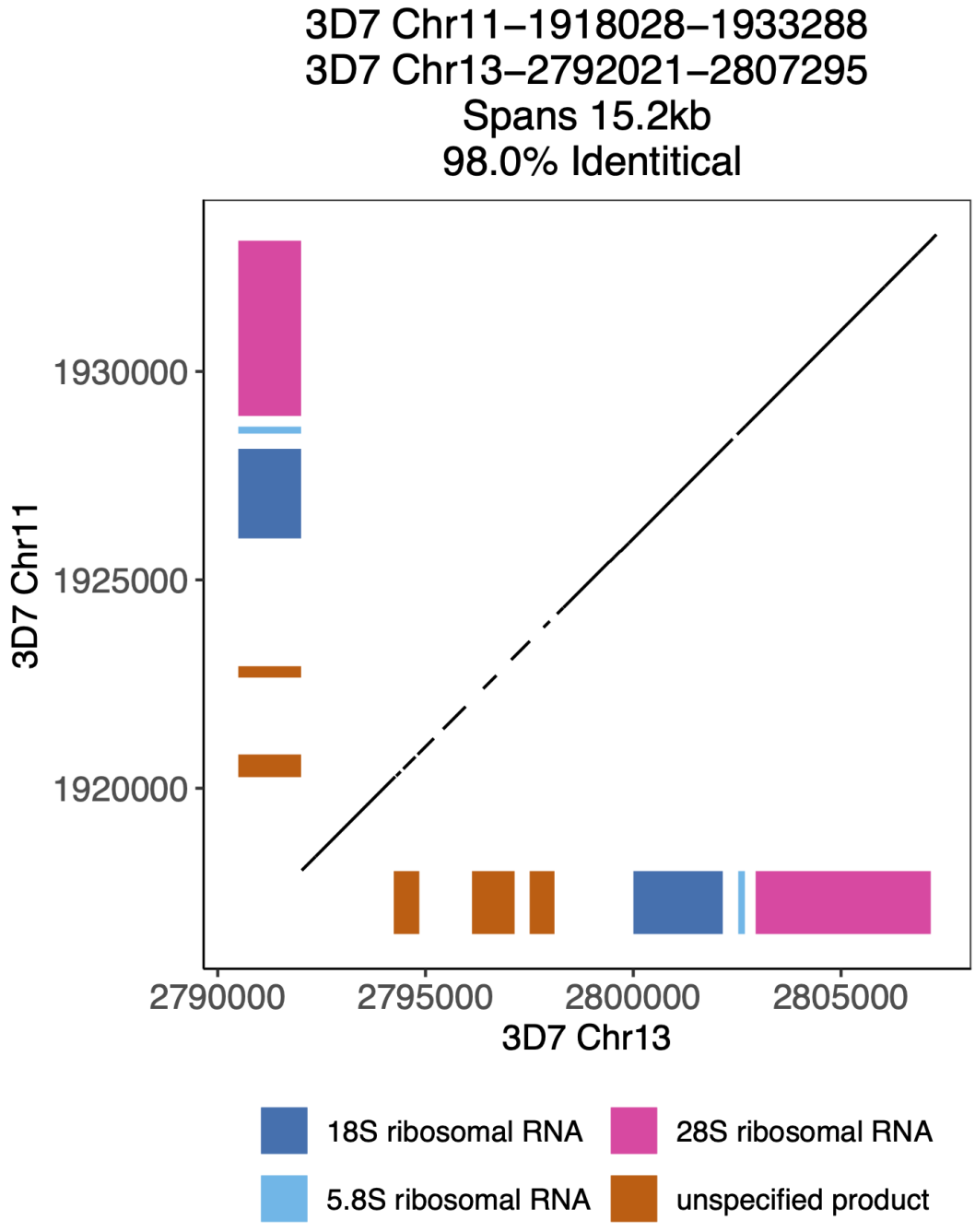


Figure 3a: Annotation of HB3 chromosomes 11 and 13. The ONT assembly of HB3 was annotated by Companion [40], and the ends of chromosomes 11 and 13 are shown above. The homologous region between chromosomes 11 and 13 is shown in blue under each chromosome, and the areas where HB3 chromosomes 11 and 13 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.

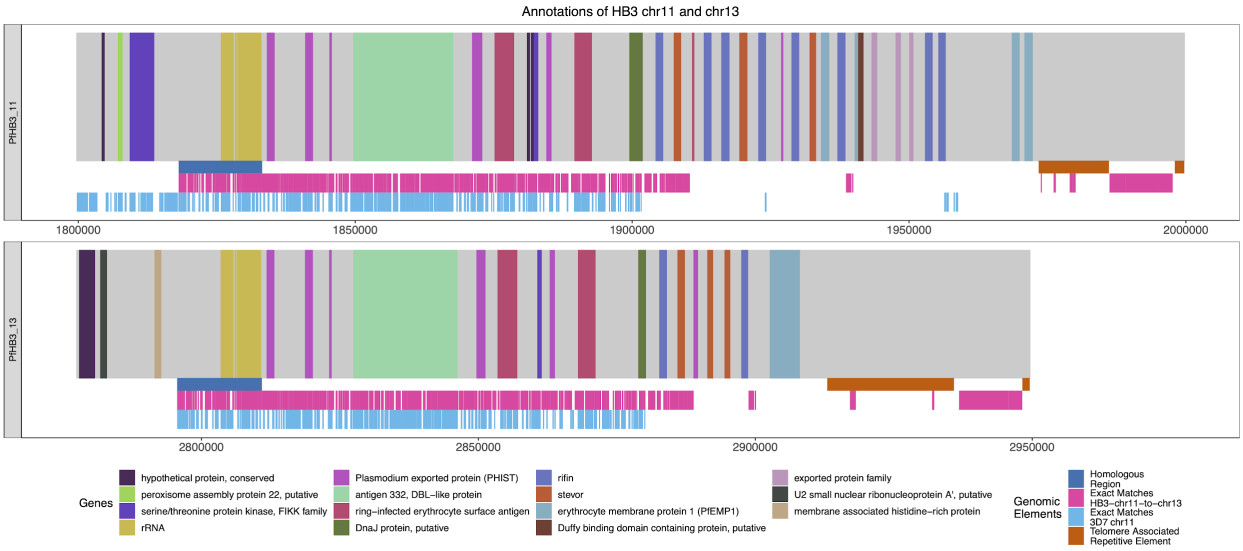


Figure 3b: Annotation of SD01 chromosomes 11 and 13. The Nanopore assembly of SD01 was annotated by Companion [40] and the ends of chromosomes 11 and 13 are shown above. The homologous region between chromosomes 11 and 13 is shown in green, and the areas where SD01 chromosomes 11 and 13 have exact matches 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 sample, 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 homologous region to the end of the 13 associated contig (98.4% similarity with only 1,428 difference over the 89,733 base region).

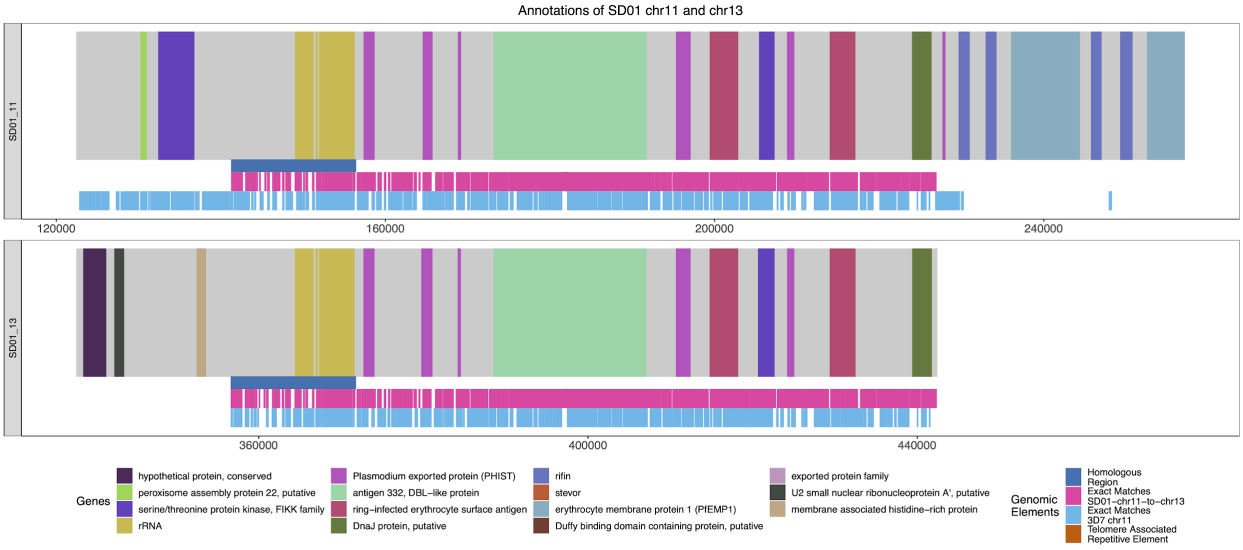
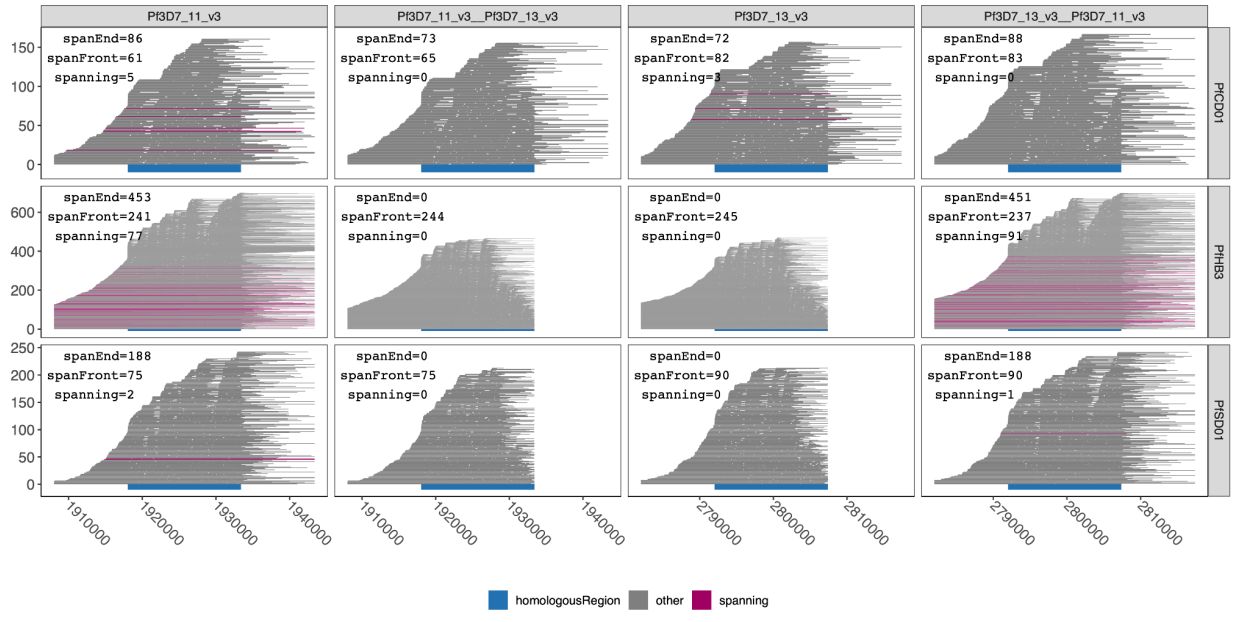
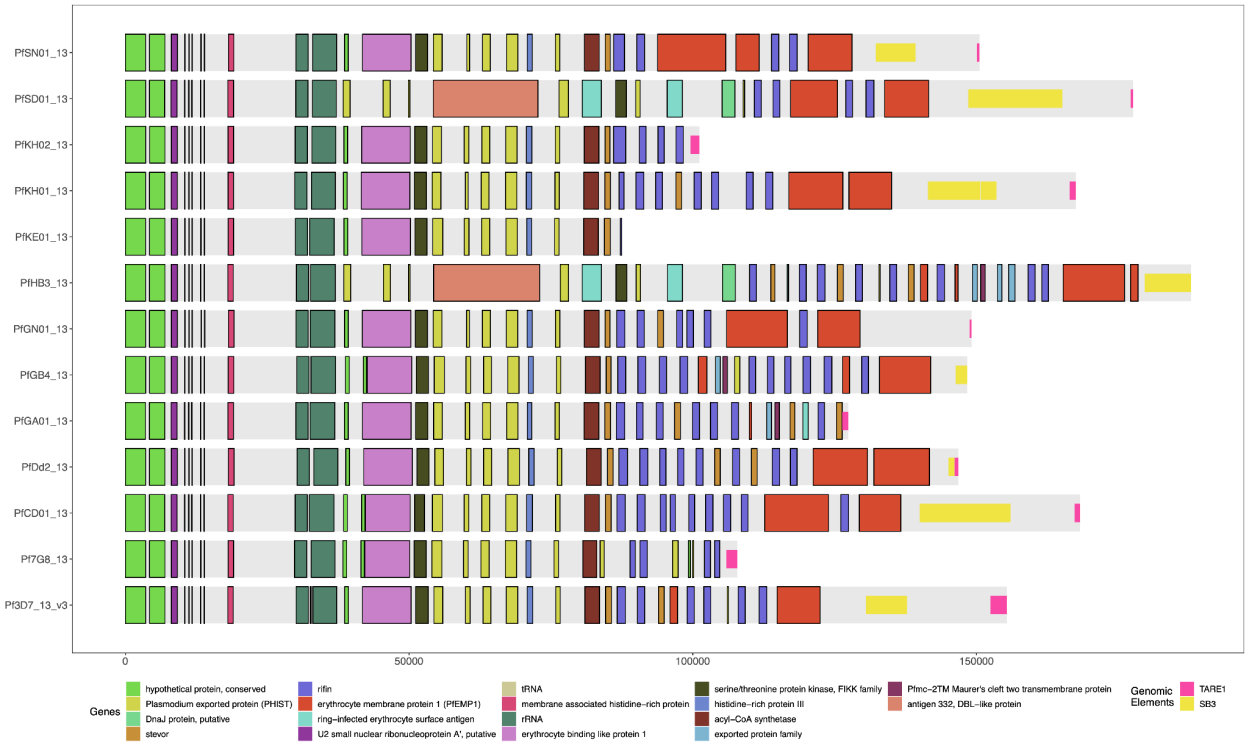


Figure 4. Spanning Reads for HB3, SD01, CD01. The PacBio and Nanopore reads for HB3, SD01, and CD01 overlapping the homologous region. Reads that completely span the region are shown in green. The reads are mapped to the 3D7 chromosomes 11 and 13 as well as artificially created hybrid chromosomes 11-13 and 13-11. The number of reads that span entirely as well as the number that span into and out of the region are shown. CD01 is being shown as an example of a strain with 3D7-consistent 11 and 13 with spanning reads as well as reads entering and leaving the homologous region on both 11 and 13. In contrast, SD01 and HB3 only have reads that span the homologous region on chromosome 11 but no reads that span out of the homologous region on chromosome 13. SD01 and HB3 also have spanning reads across the hybrid chromosome 13-11. No samples have spanning reads across 11 and 13. The other 13 strains were also mapped (not shown) and had a similar pattern to CD01.

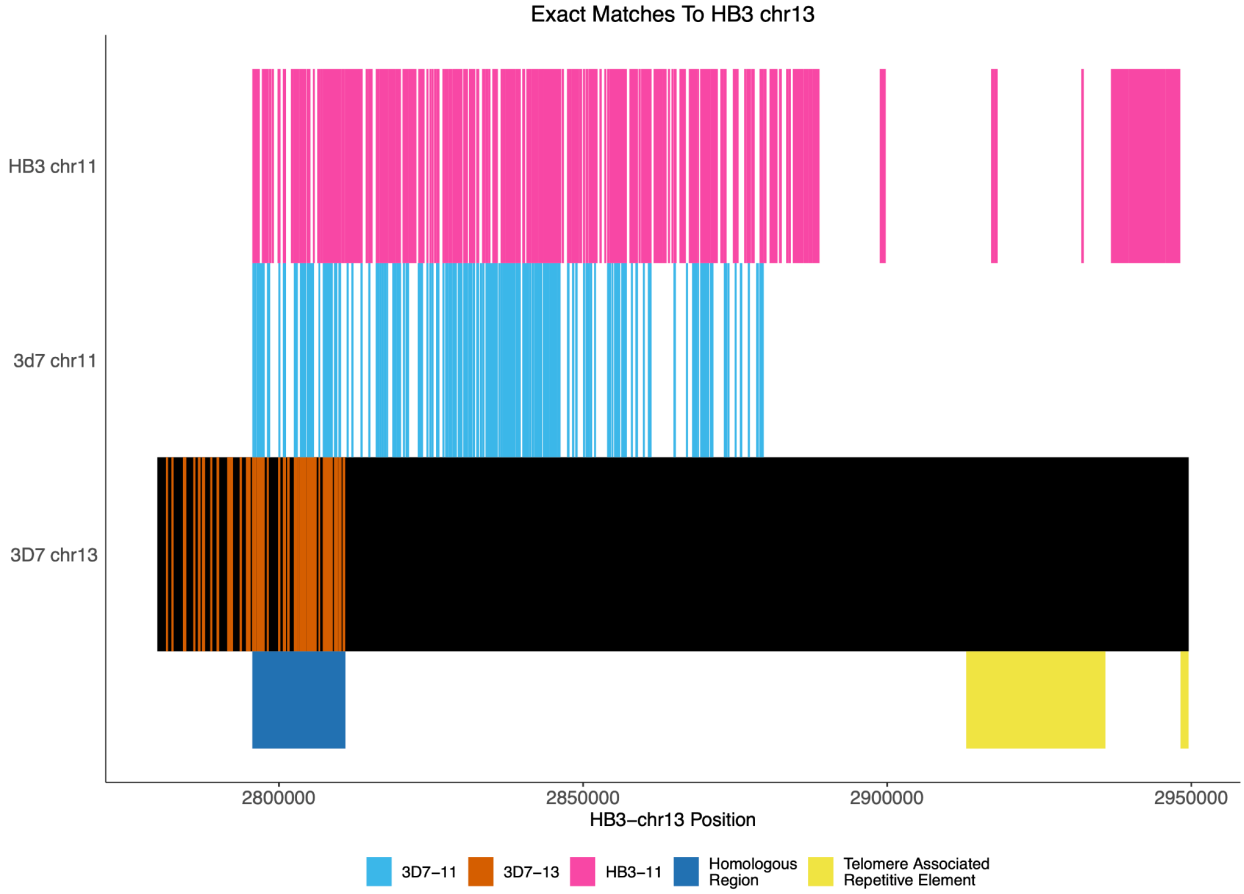


Supplemental Figure 1. Gene Annotations of Chromosome 11 of PacBio-assembled Genomes. The genomic annotations across the 3` telomeric regions of PacBio-assembled genomes[41] across chromosome 11 with the telomere repetitive elements (TAREs) are also shown if present. The presence of TAREs would suggest that 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 did not reach the telomeric regions for chromosome 11 and terminated in the homologous region between chromosome 11 and 13.

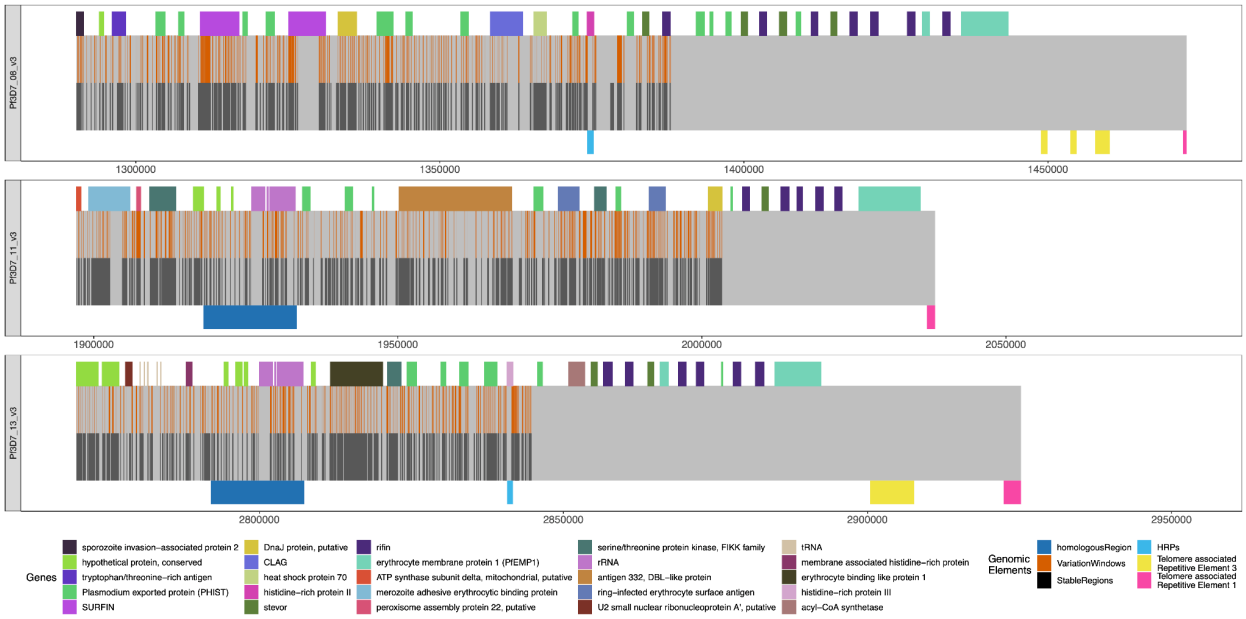
Supplemental Figure 2. Gene Annotations of Chromosome 13 of PacBio-assembled Genomes. The genomic annotations across the 3` telomeric regions of PacBio-assembled genomes[41] across chromosome 13 with the telomere repetitive elements (TAREs) are also shown if present. The presence of TAREs would suggest that 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 sequence beginning at the homologous region between chromosome 11 and 13.



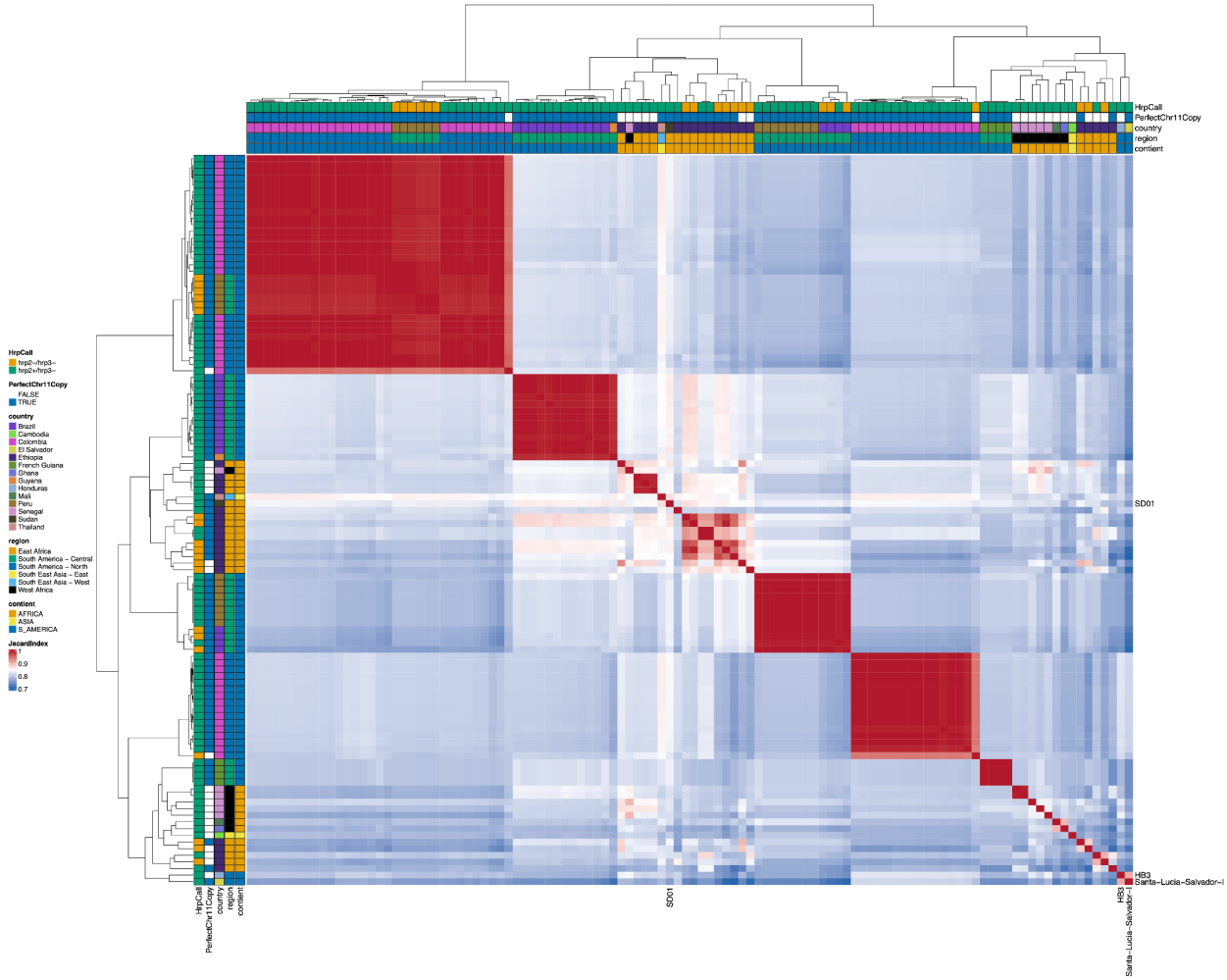
Supplemental Figure 3: Exact Matches between Nanopore-assembled HB3 chromosome 13 with HB3 chromosome 11, 3D7 chromosomes 11, 13. The locations of exact matches between the Nanopore-assembled HB3 chromosome 13 and between the assembled chromosome 11 as well as the chromosomes of 3D7 11 and 13. The green-shaded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



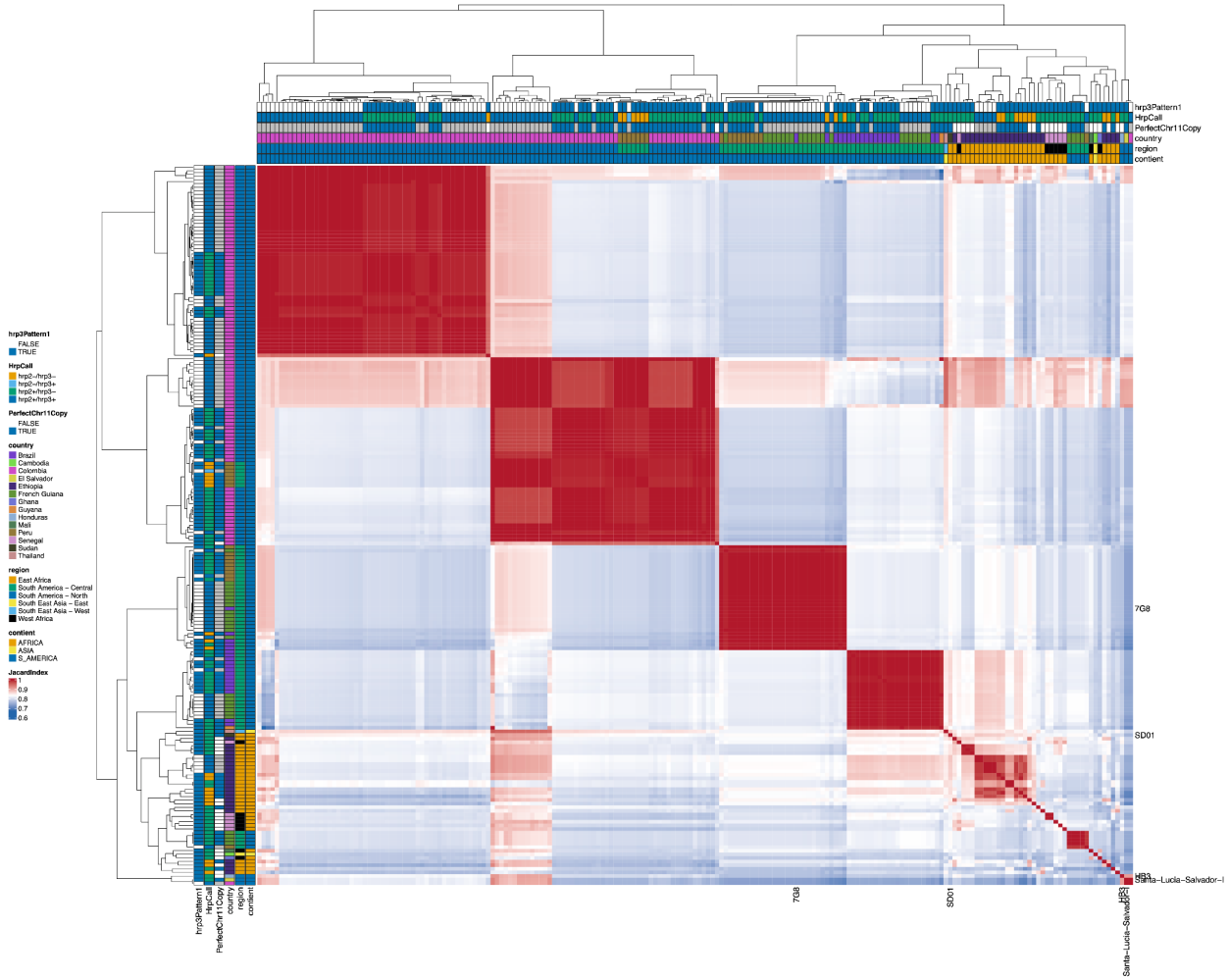
Supplemental Figure 4: Windows of interest. Windows passing filters as described in the Methods section are colored dark gray on each chromosome. The chromosomes are mapped from the beginning of the regions of interest to the chromosomes' ends with all genes/pseudogenes annotations shown. The orange bars on top of the dark gray bars are sub-regions of where there is variation that can be used to type the chromosomes. The homologous region between chromosome 11 and 13 is shown (green bars) as are the regions containing the HRP genes (blue bars)



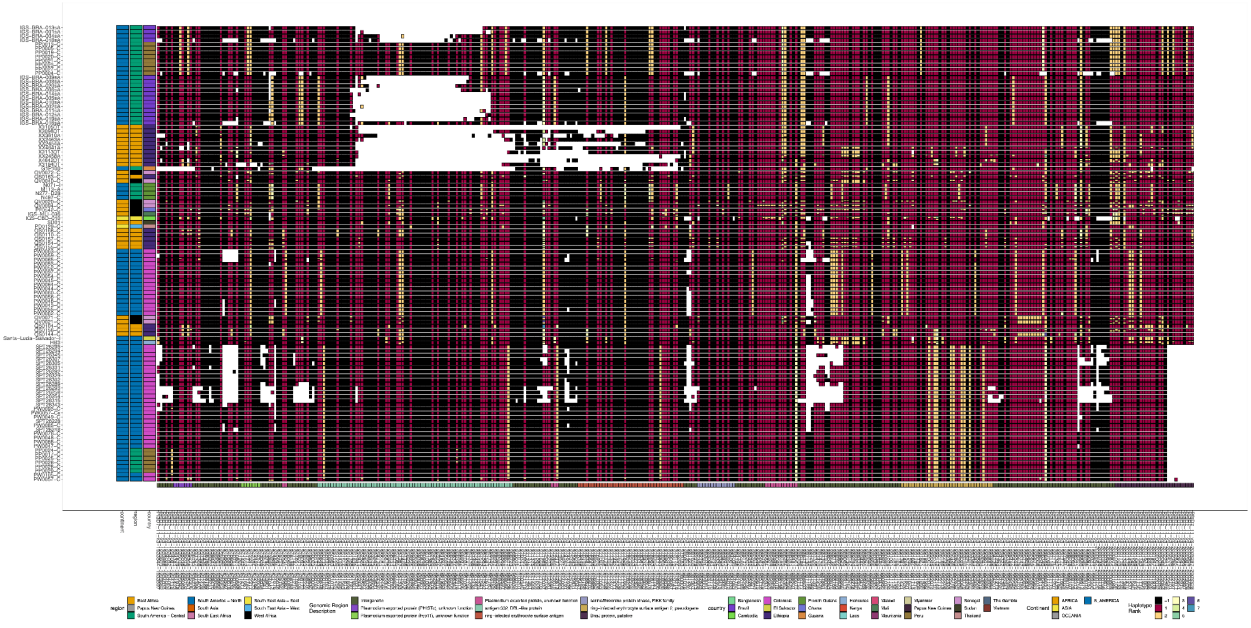
Supplemental Figure 5: **Chromosome 11 Duplicated Segment HRP3 deletion Pattern 1 samples Jaccard similarity between samples.** An all-by-all distance matrix showing Jaccard similarity for the chromosome 11 duplicated segment between all samples with HRP3 deletion pattern 1. The top triangle is identical to the bottom triangle. Samples' 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. 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. Haplotypes group strongly by geographical location.



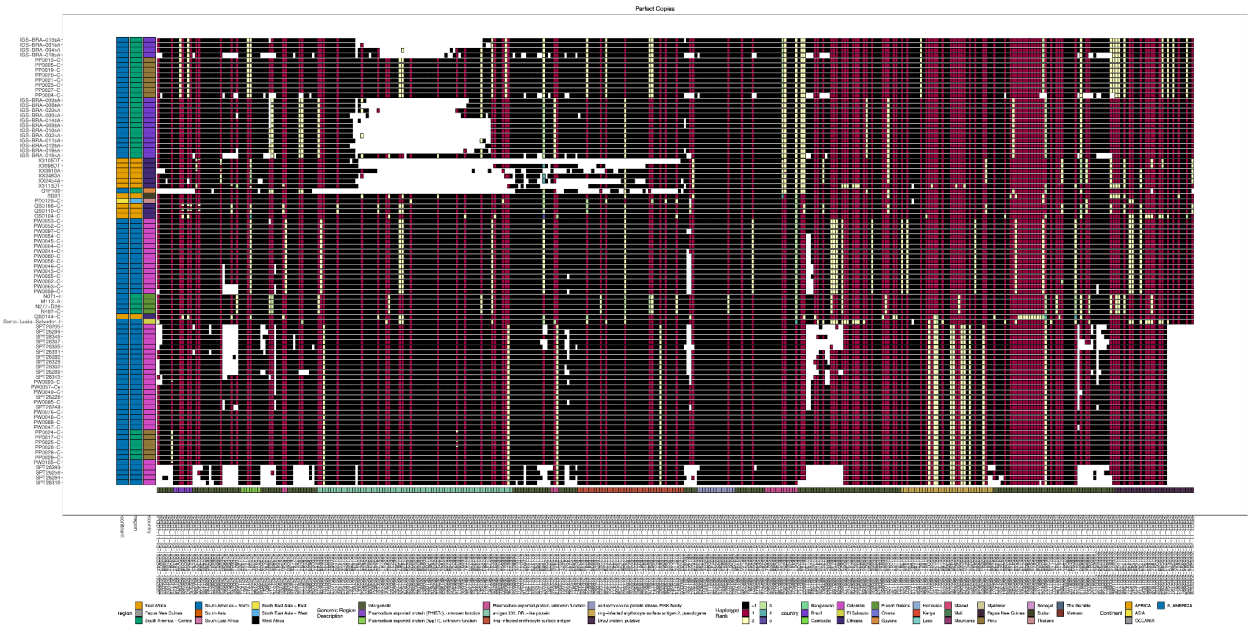
Supplemental Figure 6: **Chromosome 11 Duplicated Segment HRP3 deletion Pattern 1 samples Jaccard similarity between all samples.** Similar to **Supplemental Figure 5**, all-by-all heatmap of Jaccard similarity but this figure includes all samples with a similar chromosome segment to the samples with pattern 1 *pfhrp3* deletions. For the side and top annotation for the samples 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 samples 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 7: Chromosome 11 Duplicated Segment HRP3 deletion Pattern 1 samples. Plotted haplotype variation per sub-genomic regions across the duplicated chromosome 11 segment for the *pfhrp3* pattern 1 samples. 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 sample with pattern 1 of *pfhrp3* deletion where this segment of chromosome 11 is duplicated onto chromosome 13. The continent, region, and country are colored per sample 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 sample at a genomic location, the bar's height is the relative within-sample frequency of that haplotype for that sample. The samples are ordered in the same order as the heatmap dendrogram seen in **Supplemental Figure 6.** There are clear distinctive haplotypes for this duplicated region.

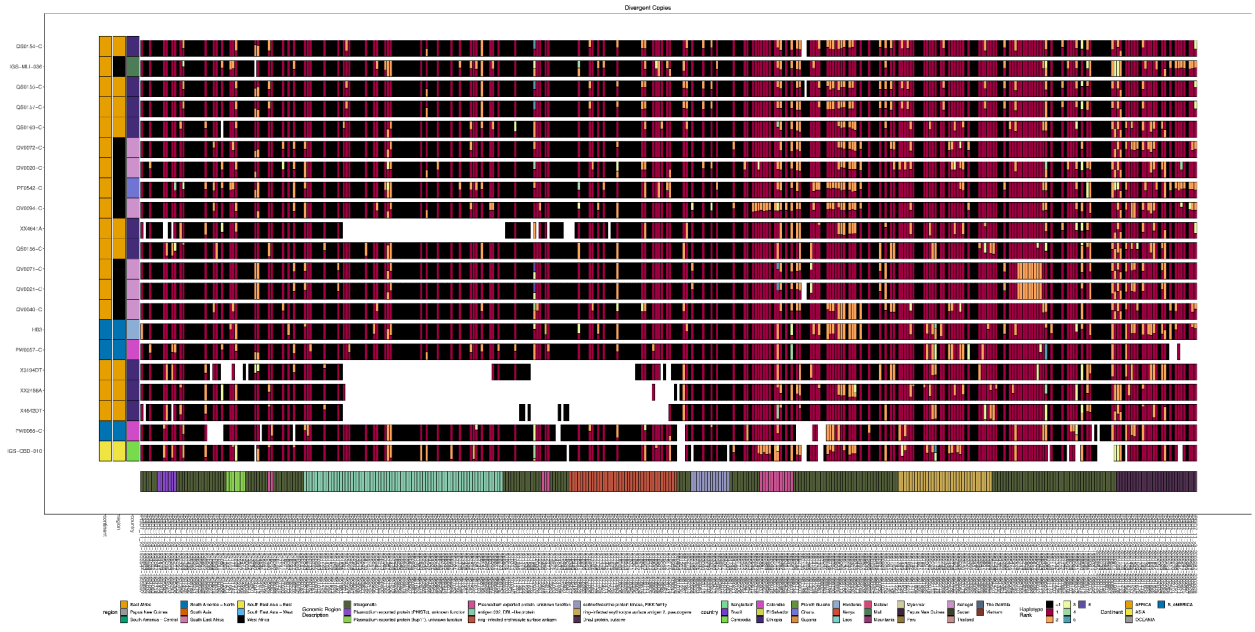


Supplemental Figure 8: **Chromosome 11 Duplicated Segment HRP3 deletion Pattern 1 samples with perfect copies.** Subset of the samples from **Supplemental Figure 7** for the samples that have a perfect duplication of the chromosome 11 segment. There are clearly very divergent haplotypes for the perfect 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.

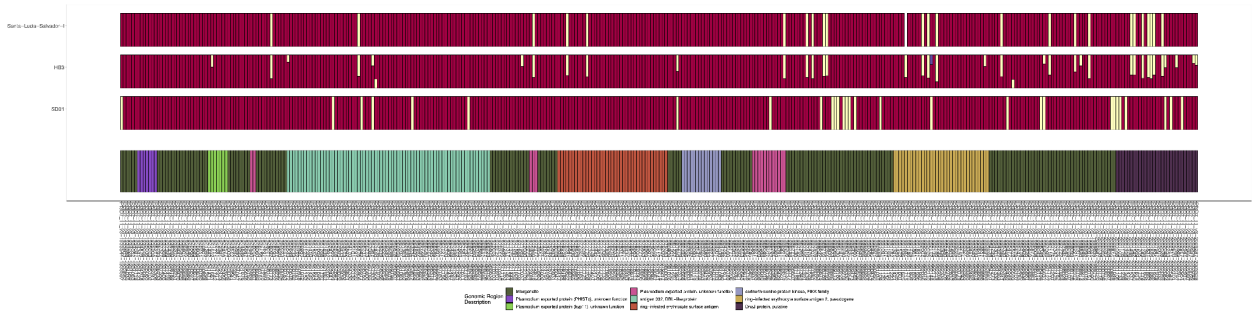


Supplemental Figure 9: **Chromosome 11 Duplicated Segment HRP3 deletion Pattern 1 samples with divergent copies.** Subset of the samples from **Supplemental Figure 7** for the samples that have divergent duplicates of the chromosome 11 segment. There are two pairs of

samples that have divergent chromosome 11 segments, but between samples, they share the same exact divergent copies, which would be consistent with the inheritance of the divergent copies. This could mean samples are inheriting from previous duplication events between divergent copies.

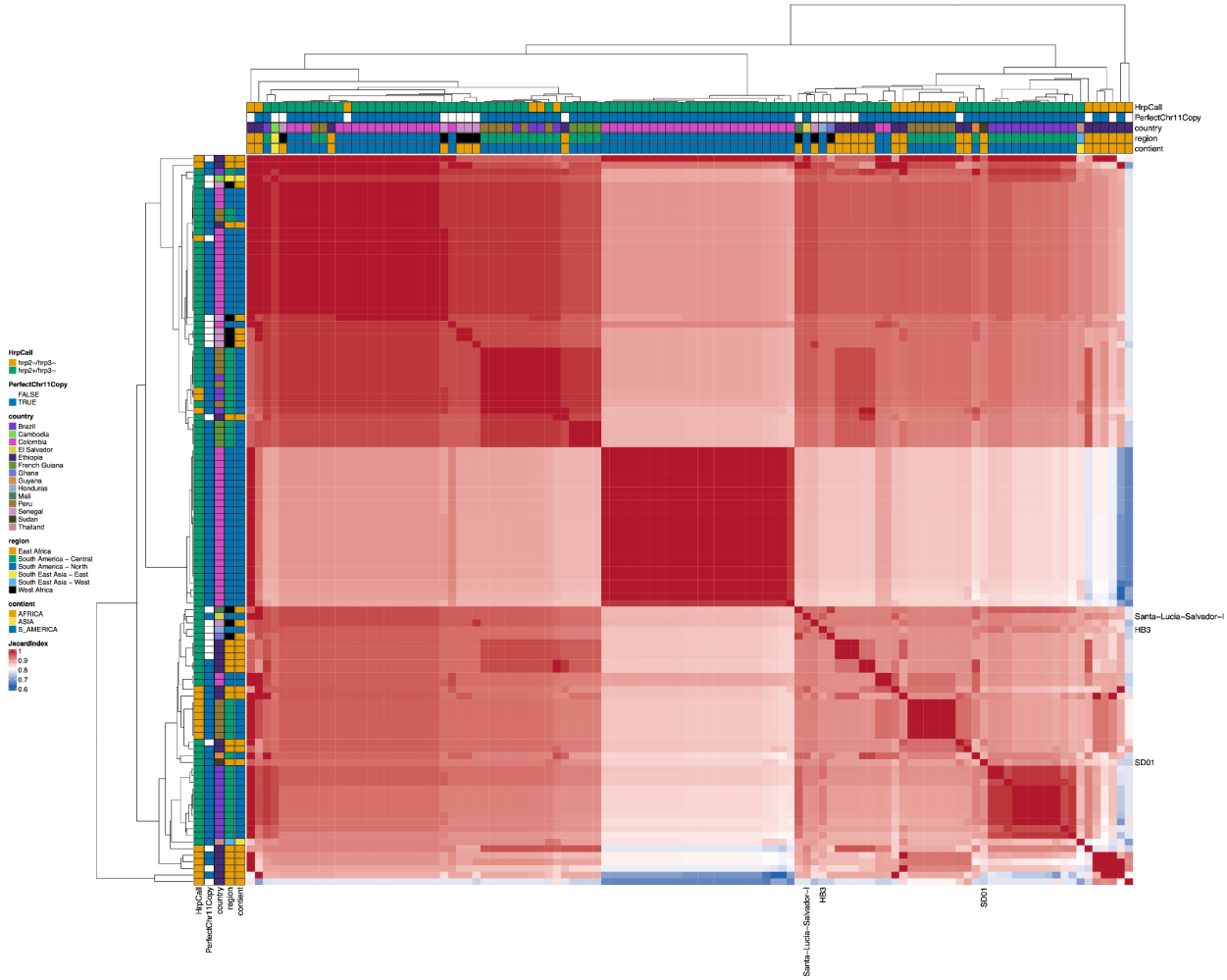


Supplemental Figure 10: Chromosome 11 Duplicated Segment HRP3 deletion Pattern 1 samples SD01, HB3, and Salvator 1. Subset of the samples from **Supplemental Figure 7** but for SD01 and HB3, which were sequenced in this paper and for Santa-Luca-Salvator-1, another lab isolate that shows similar *pfhrp3* deletion pattern 1. SD01 and Santa-Luca-Salvator-1 have perfect copies, while HB3 has divergent copies.

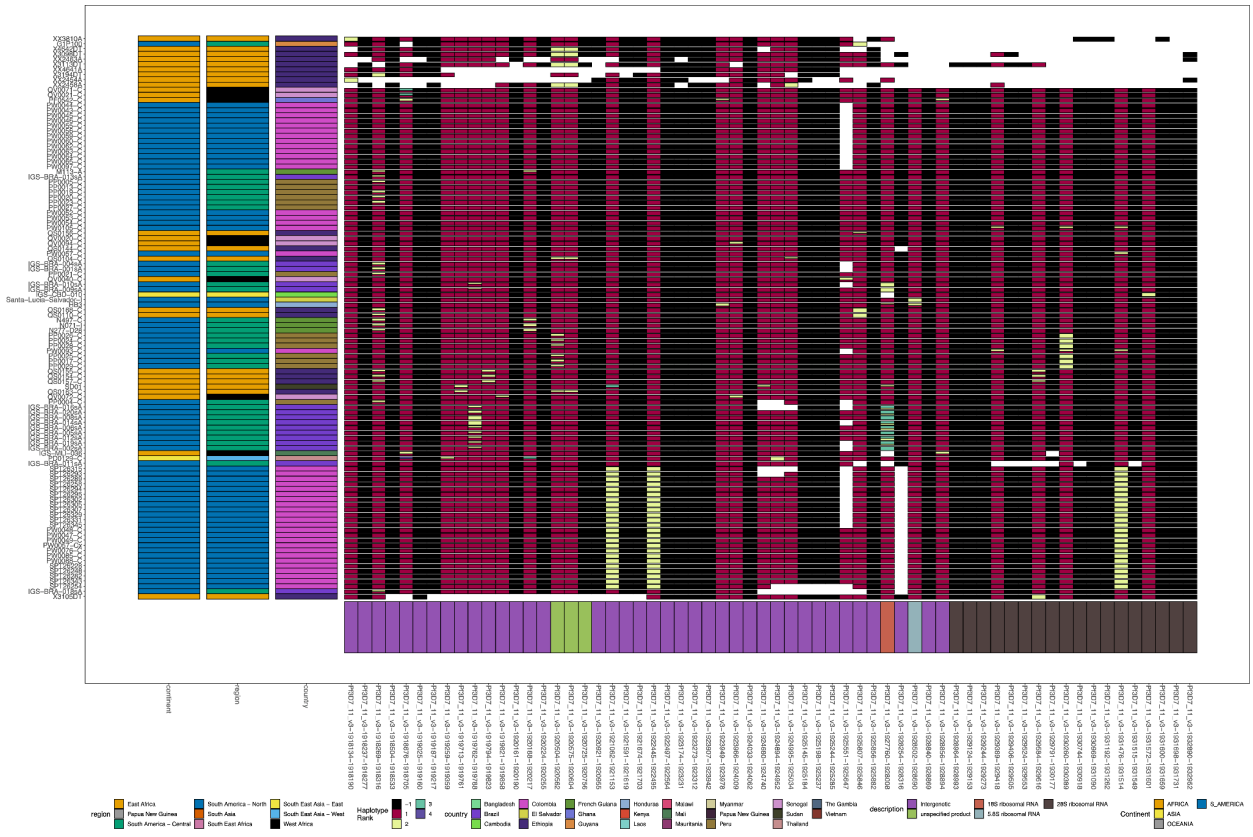


Supplemental Figure 11: Chromosome 11/13 15.2kb homologous region for *pfhrp3* deletion pattern 1 samples. An all-by-all distance matrix showing Jaccard similarity for the chromosome 11 and 13 homologous region between all the samples with *pfhrp3* deletion pattern 1. The top triangle is identical to the bottom triangle. Samples' 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. As is consistent with the fact that

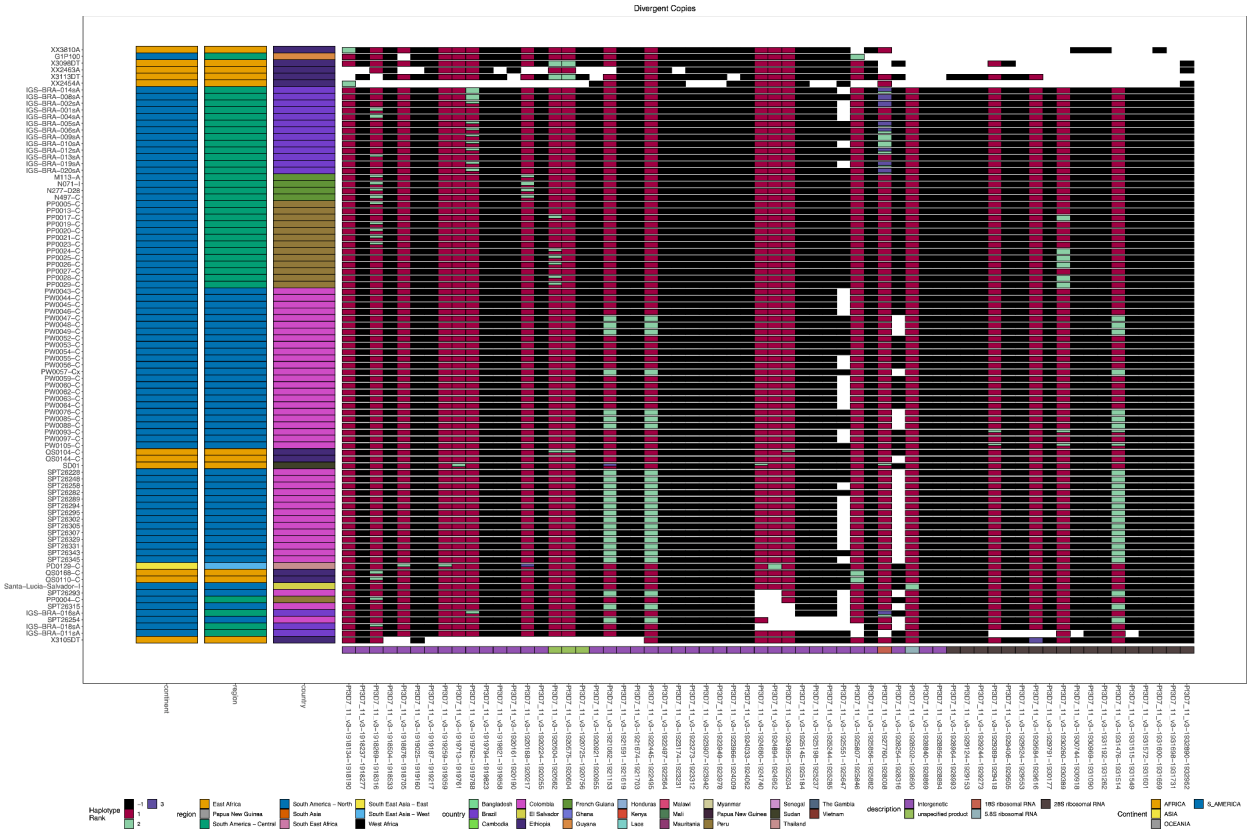
this region is highly similar between chromosomes 11 and 13 for all strains, there is high similarity between the samples here.



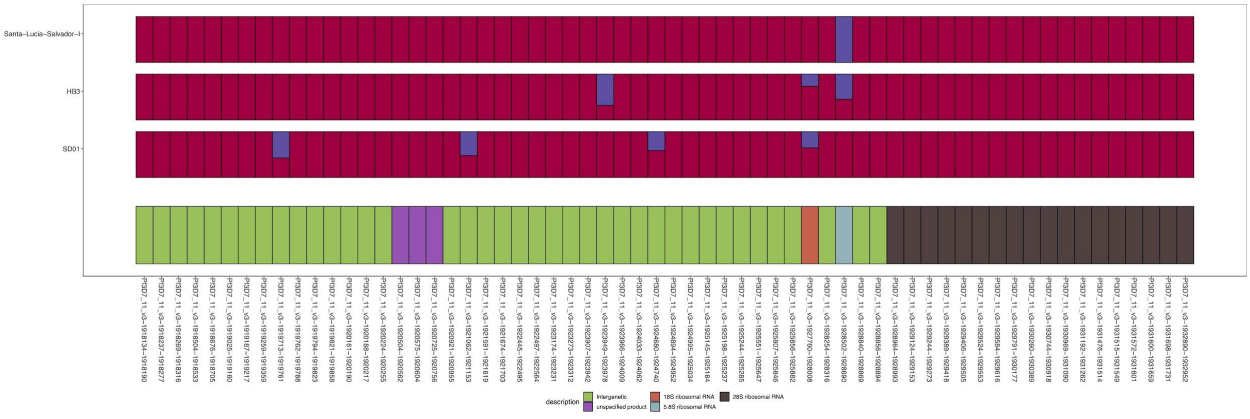
Supplemental Figure 12: **Chromosome 11/13 15.2kb homologous region for HRP3 deletion Pattern 1 samples haplotypes.** Plotted haplotype variation per subgenomic regions across the region shared between all chromosome 11 and 13 *pfhrp3* pattern samples. Across the x-axis are the genomic regions in genomic order, and the genomic region genes are colored on the bottom bar. Y-axis is each sample with pattern 1 of *pfhrp3* deletion where this segment of chromosome 11 is duplicated onto chromosome 13. The continent, region, and country are colored per sample on the left most of the plot. Each column contains the haplotypes for that genomic region colored by the haplotype rank at that window. If a 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 sample at a genomic location, the bar's height is the relative within-sample frequency of that haplotype for that sample. The samples are ordered in the same order as the heatmap dendrogram seen in **Supplemental Figure 11**.



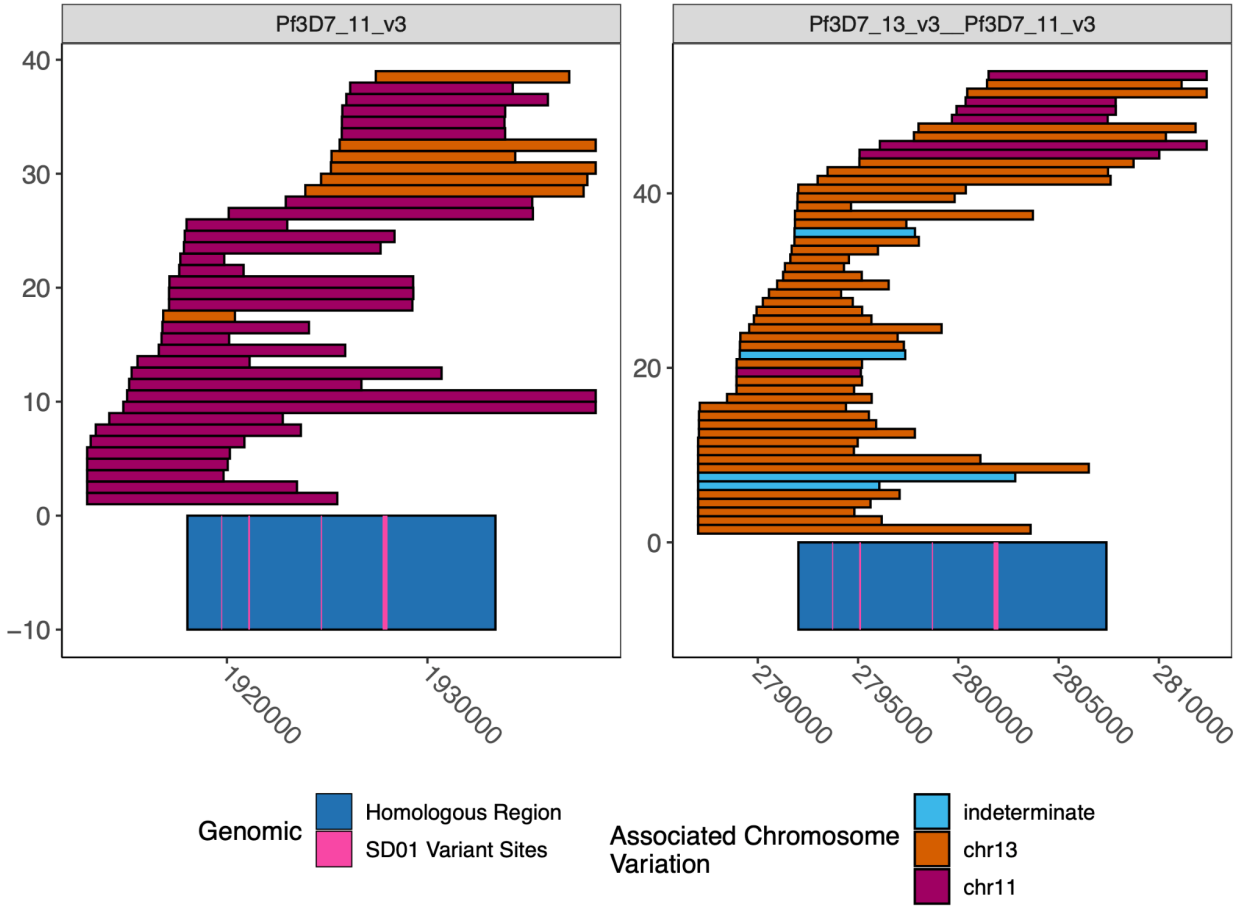
Supplemental Figure 13: Chromosome 11/13 15.2kb homologous region for HRP3 deletion Pattern 1 samples with perfect chromosome 11 segment copies haplotypes. Subset of samples from **Supplemental Figure 12** for the chromosome 11/13 homologous region for the samples with perfect chromosome 11 copies. There are several samples with divergent copies of this region despite the downstream chromosome 11 segment being a perfect copy. This would be consistent with the breakpoint for the duplication event being within this region itself.



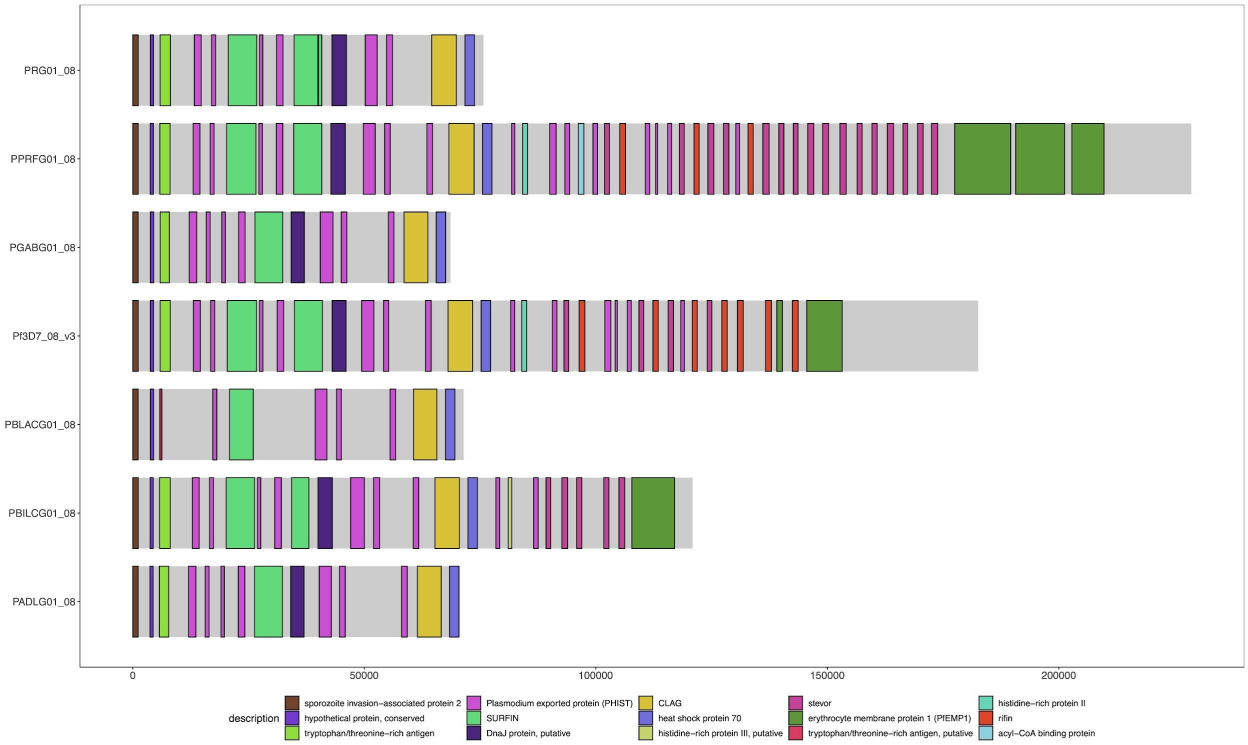
Supplemental Figure 14: Chromosome 11/13 15.2kb homologous region for samples SD01, HB3, and Salvator 1. Subset of the samples from **Supplemental Figure 12** but for SD01 and HB3, which were sequenced in this paper and for Santa-Luca-Salvator-1, another lab isolate that shows similar *pfhrp3* deletion pattern 1. SD01 and Santa-Luca-Salvator-1 have perfect copies, but while Santa-Luca-Salvator-1 has no variation within the homologous region, SD01 has variation at 4 loci within this region. HB3 has divergent copies of the duplicated chromosome 11 segment and also contains variation within this region.



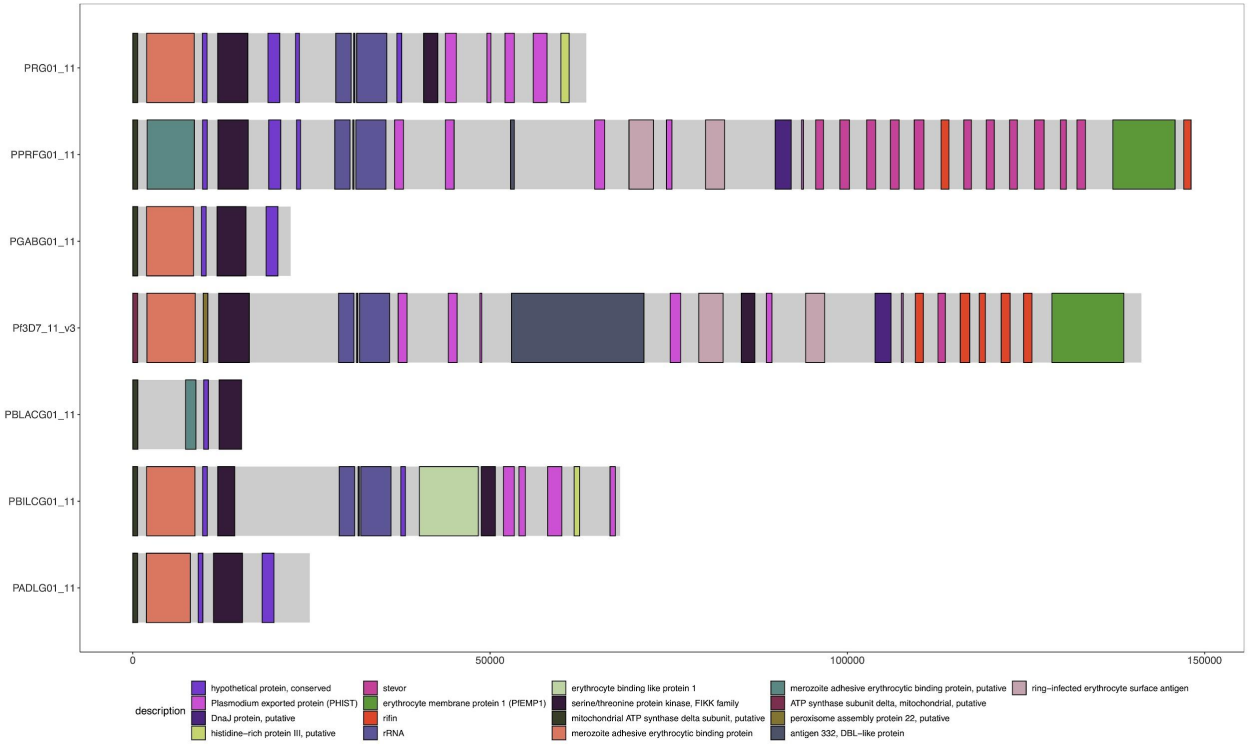
Supplemental Figure 15: Spanning PacBio and Nanopore Reads across the homologous region for SD01. The spanning Nanopore and PacBio reads across the chromosome 11 and 13 homologous regions. The visualization truncates the reads if they span outside of the range shown. The left panel is chromosome 11, and the right panel is the hybrid chromosome 13-11. The homologous region is colored on the bottom of the plot, and the 4 loci that the sample SD01 is found to be homologous are colored in blue over the region. 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 locus 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. The reads are plotted against 3D7, which does not contain the same variation as SD01, and with the hybrid chromosome being partially chromosome 11, the genome plotted against has a virtual identical sequence starting from the homologous region onwards. Therefore, any reads starting in this region on either plot match both chromosomes equally well, and what is important are the reads that span from before the homologous region to afterwards.



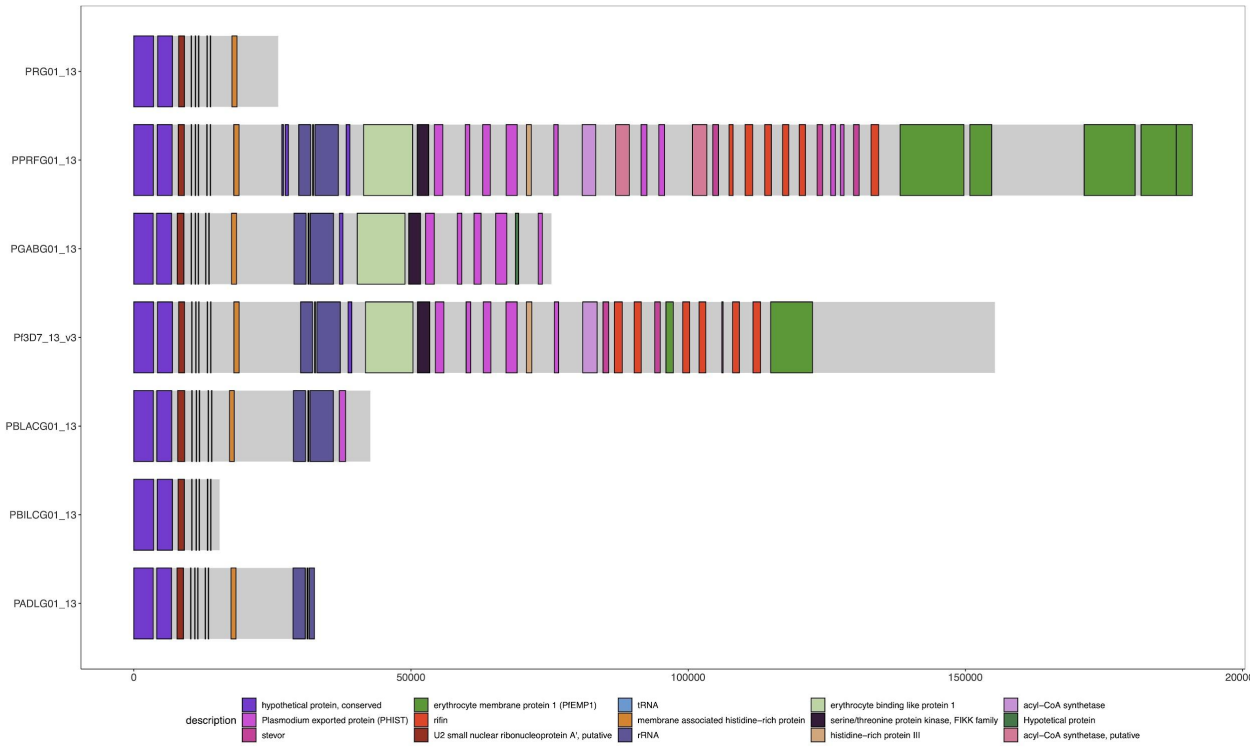
Supplemental Figure 16: Gene Annotations of Chromosome 8 of PacBio-assembled *P. Laverania* Genomes.



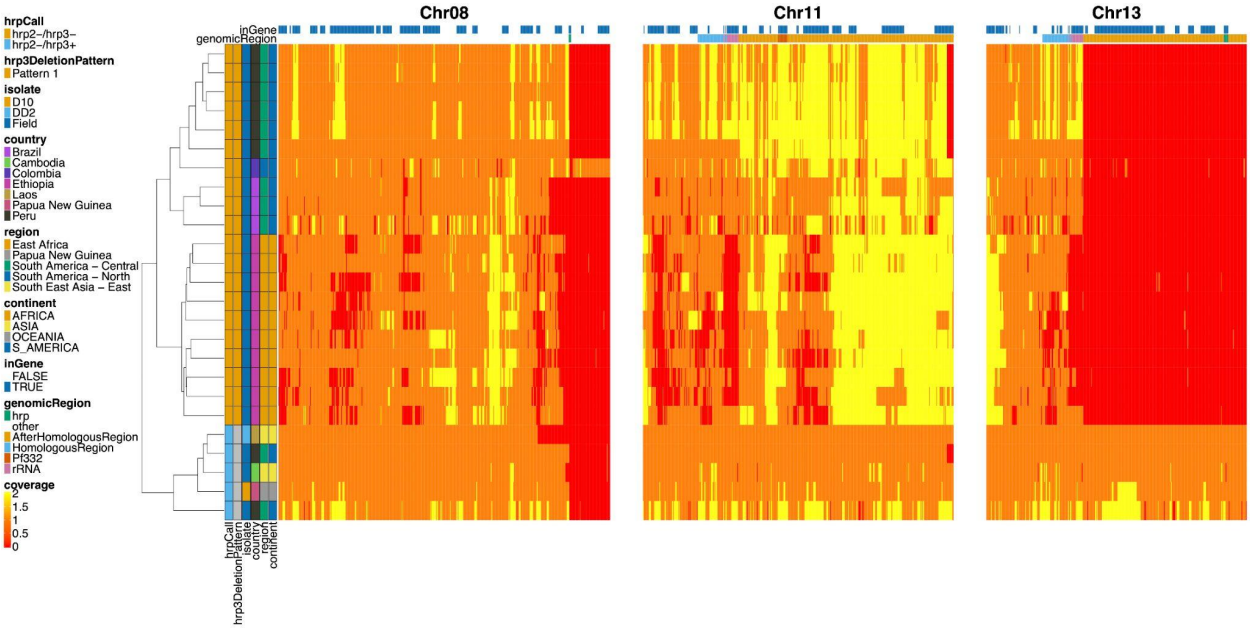
Supplemental Figure 17: Gene Annotations of Chromosome 11 of PacBio-assembled *P. Laverania* Genomes



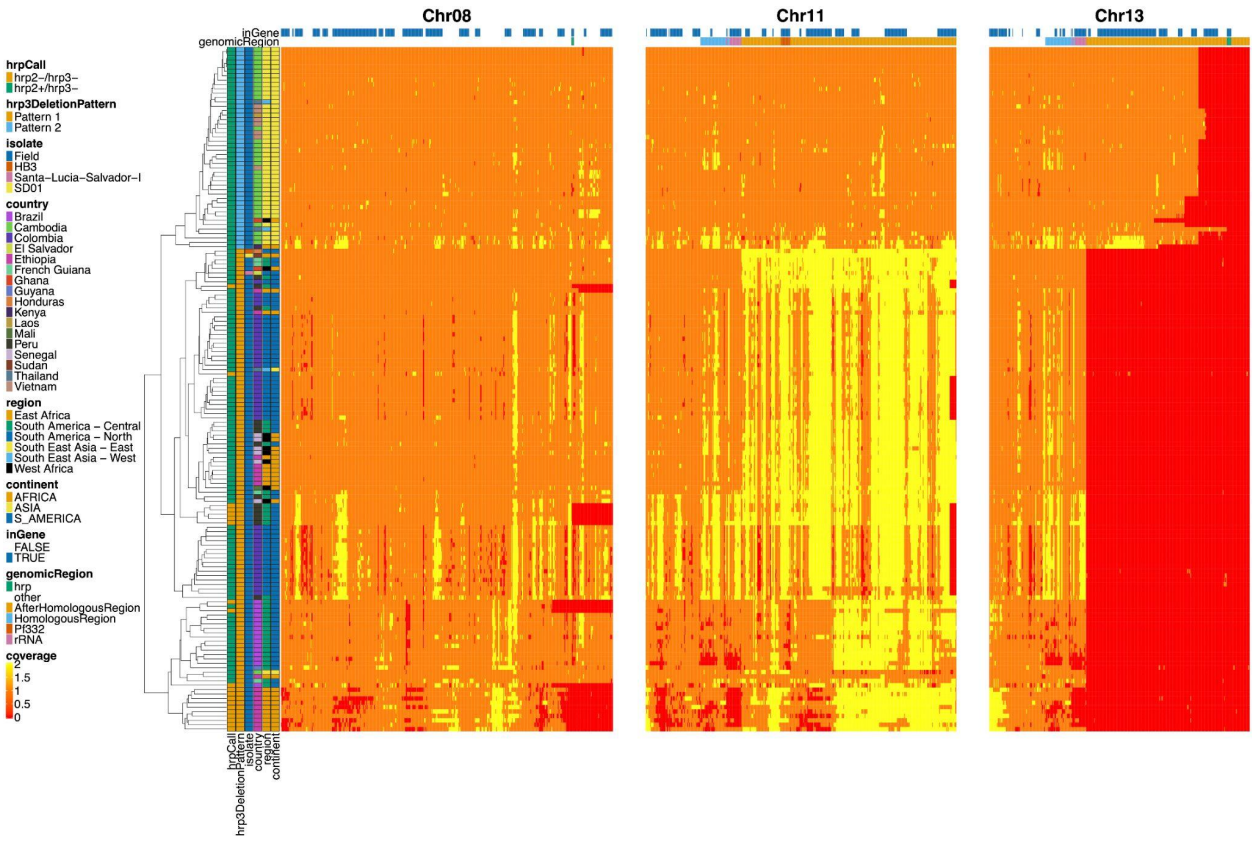
Supplemental Figure 18: Gene Annotations of Chromosome 13 of PacBio-assembled *P. Laverania* Genomes



Supplemental Figure 19: Genome coverage of possibly pfhrp2 deleted strains. Subset of the samples from Figure 1, showing only the strains that had possibly pfhrp2-deleted strains.



Supplemental Figure 20: Genome coverage of possibly pfhrp3 deleted strains. Subset of the samples from Figure 1, showing only the strains that had possibly pfhrp3-deleted strains.



DECLARATIONS

Ethics approval and consent to participate: This study is considered Not Human Subjects Research (NHSR) and thus the need for ethics approval and consent was waived.

Consent for publication: Not applicable.

Availability of data and materials: ONT data is available from the SRA (Project # pending)

Competing Interests: The authors declare that they have no competing interests.

Funding: 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 DeFeo: 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

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

References

1. Malaria. https://www.who.int/news-room/fact-sheets/detail/malaria. Accessed 13 Jan 2022
2. Organization WH, Others (2020) World malaria report 2020: 20 years of global progress and challenges. World malaria report 2020: 20 years of global progress and challenges
3. Benelli G, Beier JC (2017) Current vector control challenges in the fight against malaria. Acta Trop 174:91–96
4. Kirkman LA, Deitsch KW (2014) Recombination and Diversification of the Variant Antigen Encoding Genes in the Malaria Parasite Plasmodium falciparum. Microbiol Spectr. https://doi.org/10.1128/microbiolspec.MDNA3-0022-2014
5. Mikus MD, Petes TD (1982) Recombination between genes located on nonhomologous chromosomes in Saccharomyces cerevisiae. Genetics 101:369–404
6. Conrad DF, Pinto D, Redon R, et al (2010) Origins and functional impact of copy number variation in the human genome. Nature 464:704–712
7. Korbel JO, Urban AE, Affourtit JP, et al (2007) Paired-end mapping reveals extensive structural variation in the human genome. Science 318:420–426
8. Kidd JM, Graves T, Newman TL, Fulton R, Hayden HS, Malig M, Kallicki J, Kaul R, Wilson RK, Eichler EE (2010) A human genome structural variation sequencing resource reveals insights into mutational mechanisms. Cell 143:837–847
9. Consortium T 1000 GP, The 1000 Genomes Project Consortium (2012) An integrated map of genetic variation from 1,092 human genomes. Nature 491:56–65
10. Mills RE, Walter K, Stewart C, et al (2011) Mapping copy number variation by population-scale genome sequencing. Nature 470:59–65

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

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

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

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

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

16. Gardner MJ, Hall N, Fung E, et al (2002) Genome sequence of the human malaria parasite Plasmodium falciparum. Nature 419:498–511

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

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

19. Goldstein S, Beka L, Graf J, Klassen JL (2019) Evaluation of strategies for the assembly of diverse bacterial genomes using MinION long-read sequencing. BMC Genomics 20:23

20. Jansen HJ, Liem M, Jong-Raadsen SA, et al (2017) Rapid de novo assembly of the European eel genome from nanopore sequencing reads. Sci Rep 7:7213

21. Pennisi E (2012) Genome sequencing. Search for pore-fection. Science 336:534–537

22. Reiner JE, Balijepalli A, Robertson JWF, Campbell J, Suehle J, Kasianowicz JJ (2012) Disease detection and management via single nanopore-based sensors. Chem Rev 112:6431–6451

23. Quick J, Loman NJ, Duraffour S, et al (2016) Real-time, portable genome sequencing for Ebola surveillance. Nature 530:228–232

24. Brancaccio RN, Robitaille A, Dutta S, Rollison DE, Tommasino M, Gheit T (2021) MinION nanopore sequencing and assembly of a complete human papillomavirus genome. J Virol Methods 294:114180

25. McNaughton AL, Roberts HE, Bonsall D, et al (2019) Illumina and Nanopore methods for whole genome sequencing of hepatitis B virus (HBV). Sci Rep 9:7081

26. Chandak S, Neu J, Tatwawadi K, et al (2020) Overcoming High Nanopore Basecaller Error Rates for DNA Storage Via Basecaller-Decoder Integration and Convolutional Codes. bioRxiv 2019.12.20.871939

27. Jain M, Koren S, Miga KH, et al (2018) Nanopore sequencing and assembly of a human genome with ultra-long reads. Nat Biotechnol 36:338–345

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

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

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

31. Organization WH, Others (2011) Good practices for selecting and procuring rapid diagnostic tests for malaria.

32. Cheng Q, Gatton ML, Barnwell J, Chiodini P, McCarthy J, Bell D, Cunningham J (2014) Plasmodium falciparum parasites lacking histidine-rich protein 2 and 3: a review and recommendations for accurate reporting. Malar J 13:283

33. 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. Accessed 11 Aug 2021

34. Feleke SM, Reichert EN, Mohammed H, et al (2021) Plasmodium falciparum is evolving to escape malaria rapid diagnostic tests in Ethiopia. Nat Microbiol 6:1289–1299

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

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

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

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

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

40. Steinbiss S, Silva-Franco F, Brunk B, Foth B, Hertz-Fowler C, Berriman M, Otto TD (2016) Companion: a web server for annotation and analysis of parasite genomes. Nucleic Acids Res 44:W29–34

41. Otto TD, Böhme U, Sanders M, et al (2018) Long read assemblies of geographically dispersed Plasmodium falciparum isolates reveal highly structured subtelomeres. Wellcome Open Res 3:52

42. Otto TD, Gilabert A, Crellen T, et al (2018) Genomes of all known members of a Plasmodium subgenus reveal paths to virulent human malaria. Nat Microbiol 3:687–697

43. Mok BW, Ribacke U, Sherwood E, Wahlgren M (2008) A highly conserved segmental duplication in the subtelomeres of Plasmodium falciparum chromosomes varies in copy number. Malar J 7:46

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

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

46. Hodder AN, Maier AG, Rug M, et al (2009) Analysis of structure and function of the giant protein Pf332 in Plasmodium falciparum. Mol Microbiol 71:48–65

47. Bertin GI, Sabbagh A, Argy N, et al (2016) Proteomic analysis of Plasmodium falciparum parasites from patients with cerebral and uncomplicated malaria. Sci Rep 6:26773

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

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

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

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

52. Branton D, Deamer DW, Marziali A, et al (2008) The potential and challenges of nanopore sequencing. Nat Biotechnol 26:1146–1153

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

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

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

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

57. Harris RS (2007) Improved pairwise alignment of genomic DNA. The Pennsylvania State University