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

Main Figures

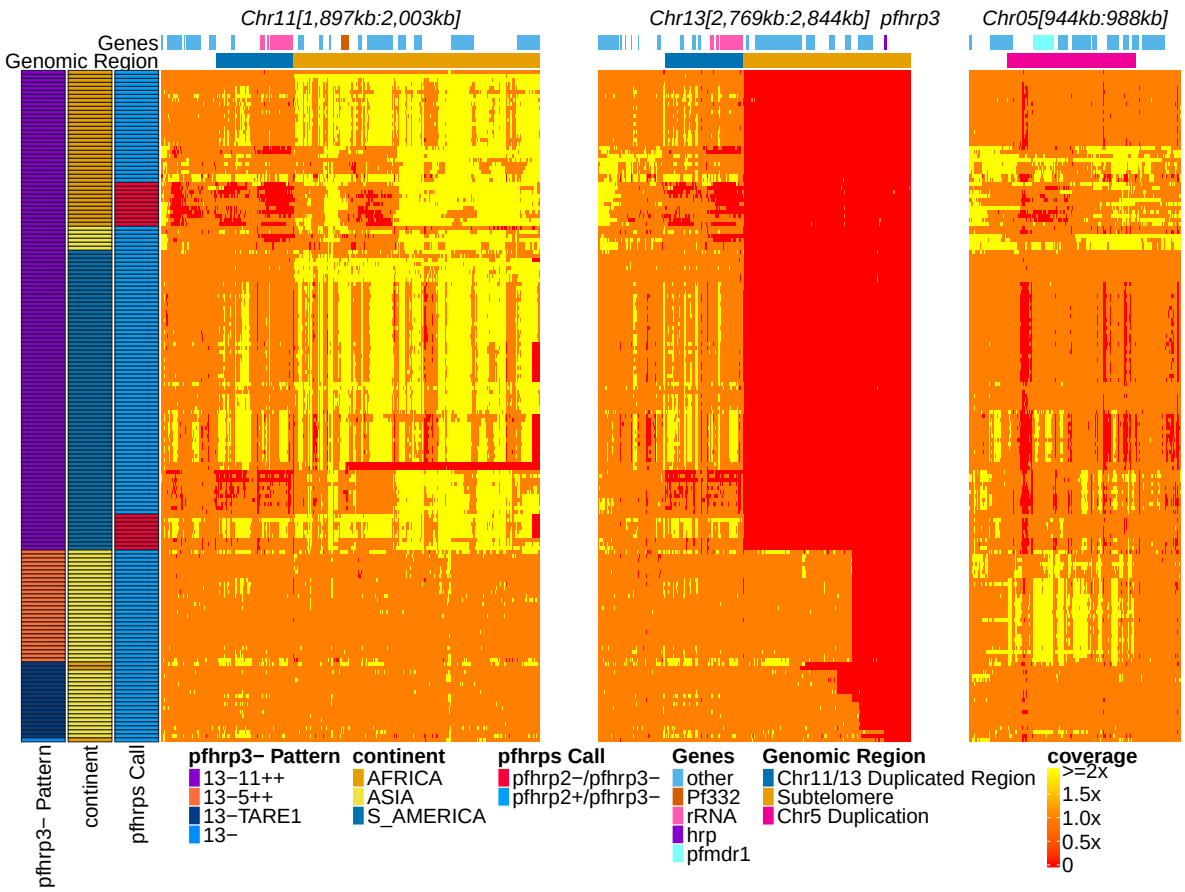

Figure 1 Pfhrp2/3 deleted parasites with altered sequence coverage across regions of chromosomes 11, 13, and 5.

Sequence coverage heatmap of *pfhrp3* deletion associated regions of chromosomes 11 (1,897,151 - 2,003,328 bp), 13 (2,769,916 - 2,844,785 bp), and 5 (944,389 - 988,747 bp) in the 168 samples with evidence of *pfhrp3* deletion out of the 19,313 publicly available samples. 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* involved in its duplication event. Each row is a WGS sample, and each column is normalized coverage. The top annotation along chromosomes depicts the location of genes with relevant genes colored: rRNA (light 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 before sequencing.

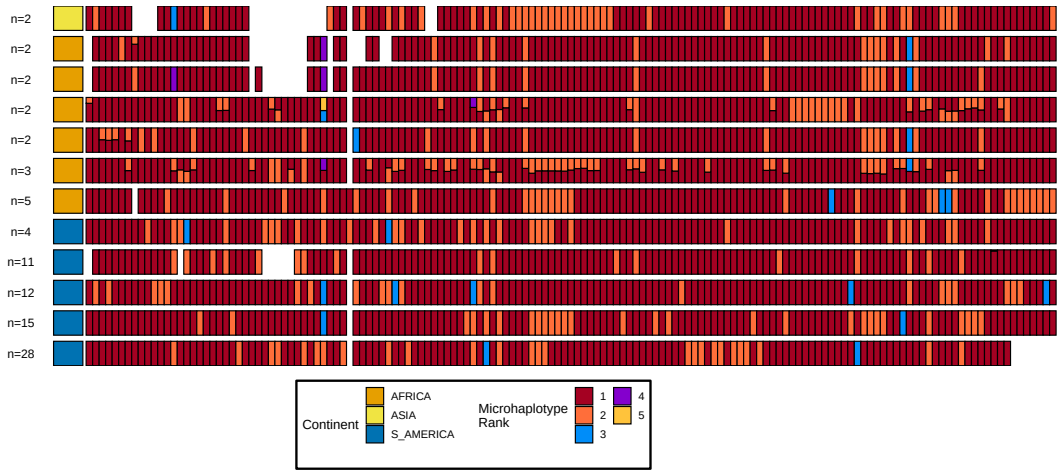

Figure 2 Microhaplotype patterns for the duplicated portion of chromosome 11 in $13^{-}11^{++}$ parasites form 11 distinct haplotype groups with a geographic distinction between Africa and the Americas.

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 (named Microhaplotype Rank), with 1=red being the most prevalent, 2=orange being the 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). Most 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, consistent with separate translocations emerging and spreading independently.

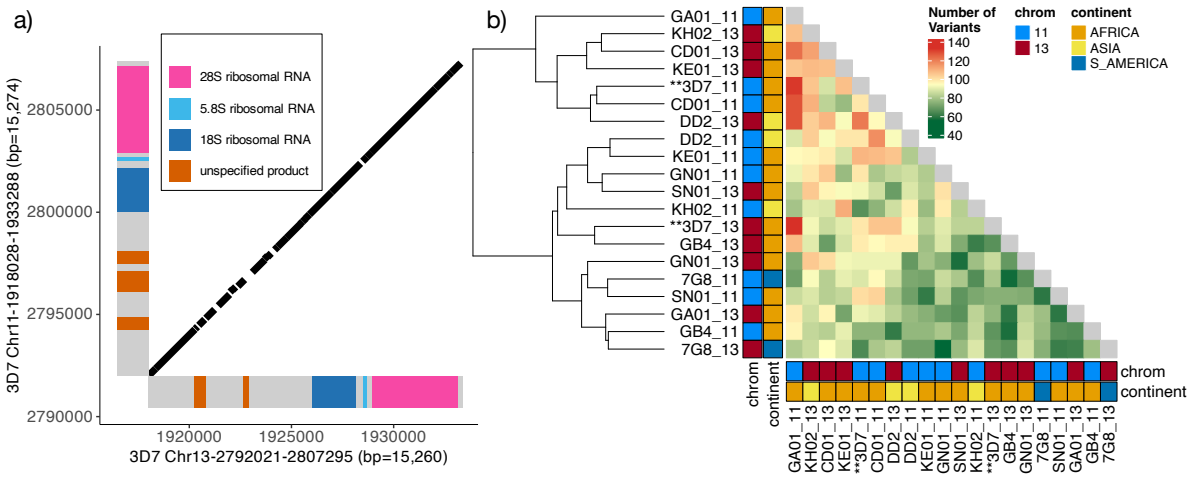

Figure 3 Characterization of the 15.2 kb segmental duplication containing ribosomal genes on chromosomes 11 and 13

(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(Otto et al., 2018a) 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, with the number of variants between segments ranging from 55 to 133 with no clear separation by continent or chromosome.

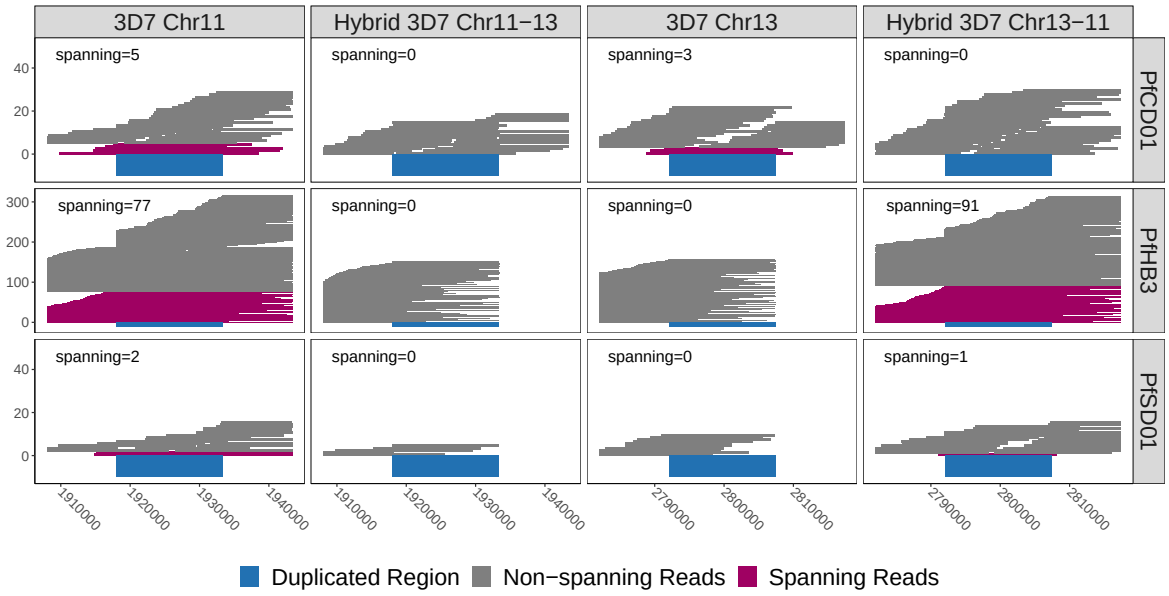

Figure 4 Long reads spanning the 15kb duplicated region confirm the presence of translocated chromosome 13-11 in *pfhrp3*-deleted HB3 (Americas) and SD01 (Africa) but not *pfhrp3*- intact chromosomes.

PacBio and Nanopore reads >15kb for HB3, SD01, and CD01 are shown aligned to normal chromosomes 11 and 13 and 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. Spanning reads mapped only to this one location, whereas the non-spanning reads mapped to both the hybrid or normal chromosomes as these chromosome segments are identical. 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.

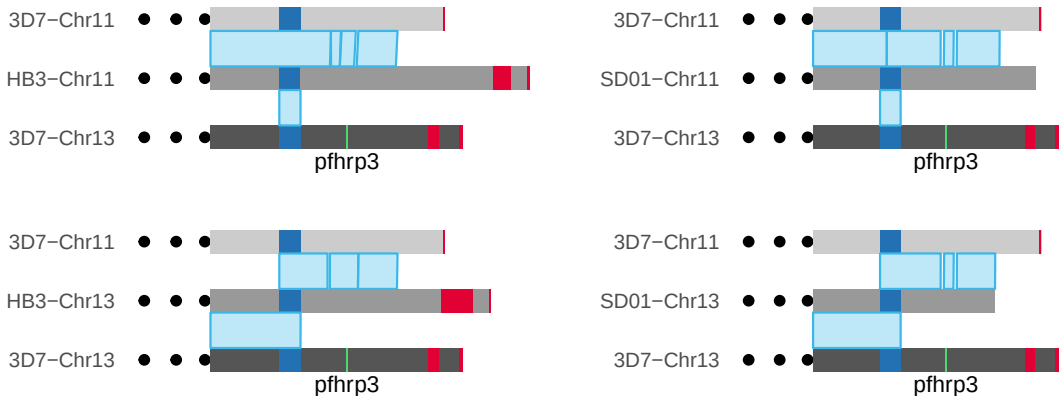

Figure 5 A comparison of long-read assemblies of chromosomes 11 and 13 of HB3 and SD01 with the reference genome 3D7 confirms hybridized chromosomes 13-11.

On top, chromosome 11 of HB3 and SD01 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 segmentally 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) sequence. 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).

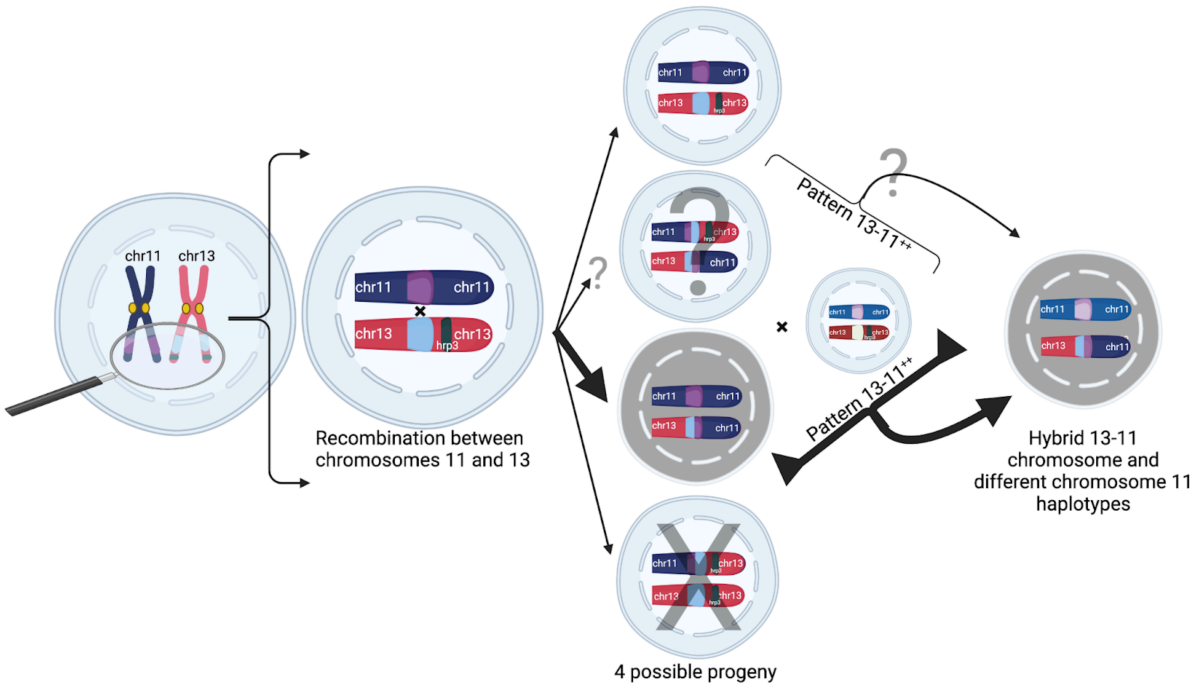

Figure 6 Proposed model of duplication-mediated non-allelic homologous recombination during intrastrain meiotic recombination yielding $13^{-}11^{++}$ parasites.

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). Figure created using Biorender.