

**Title:**

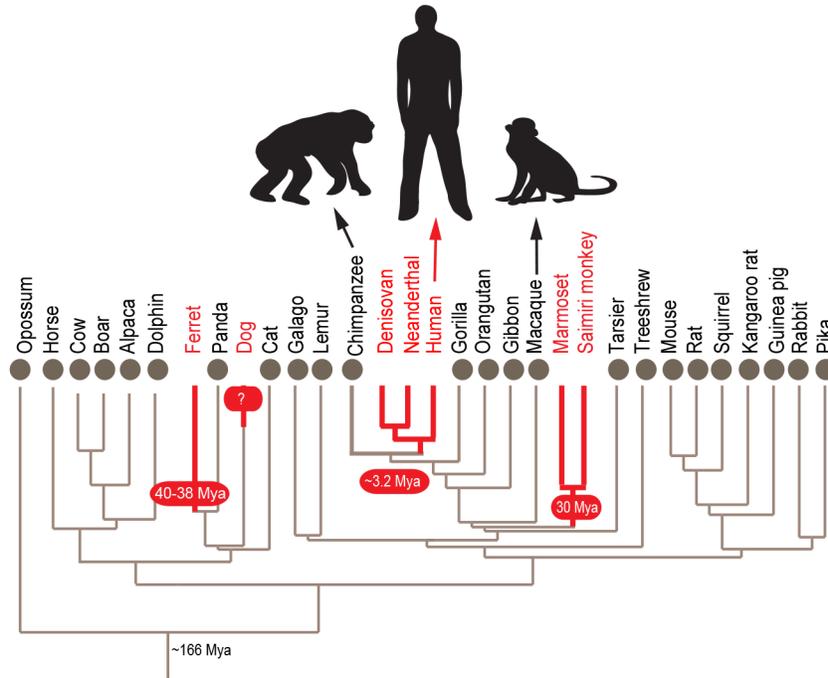
Ancient human sialic acid variant restricts an emerging zoonotic malaria parasite

**Authors:**

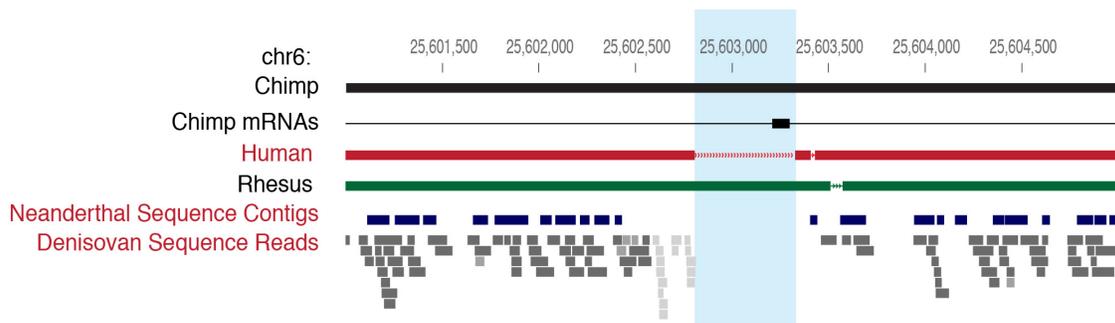
Selasi Dankwa<sup>1</sup>, Caeul Lim<sup>1</sup>, Amy K. Bei<sup>1</sup>, Rays H.Y. Jiang<sup>1</sup>, James R. Abshire<sup>2</sup>, Saurabh D. Patel<sup>1,3</sup>, Jonathan M. Goldberg<sup>1</sup>, Yovany Moreno<sup>1</sup>, Maya Kono<sup>1</sup>, Jacquin C. Niles<sup>2</sup>, Manoj T. Duraisingh<sup>1\*</sup>

\*Corresponding author: Manoj T. Duraisingh, mduraisi@hsph.harvard.edu

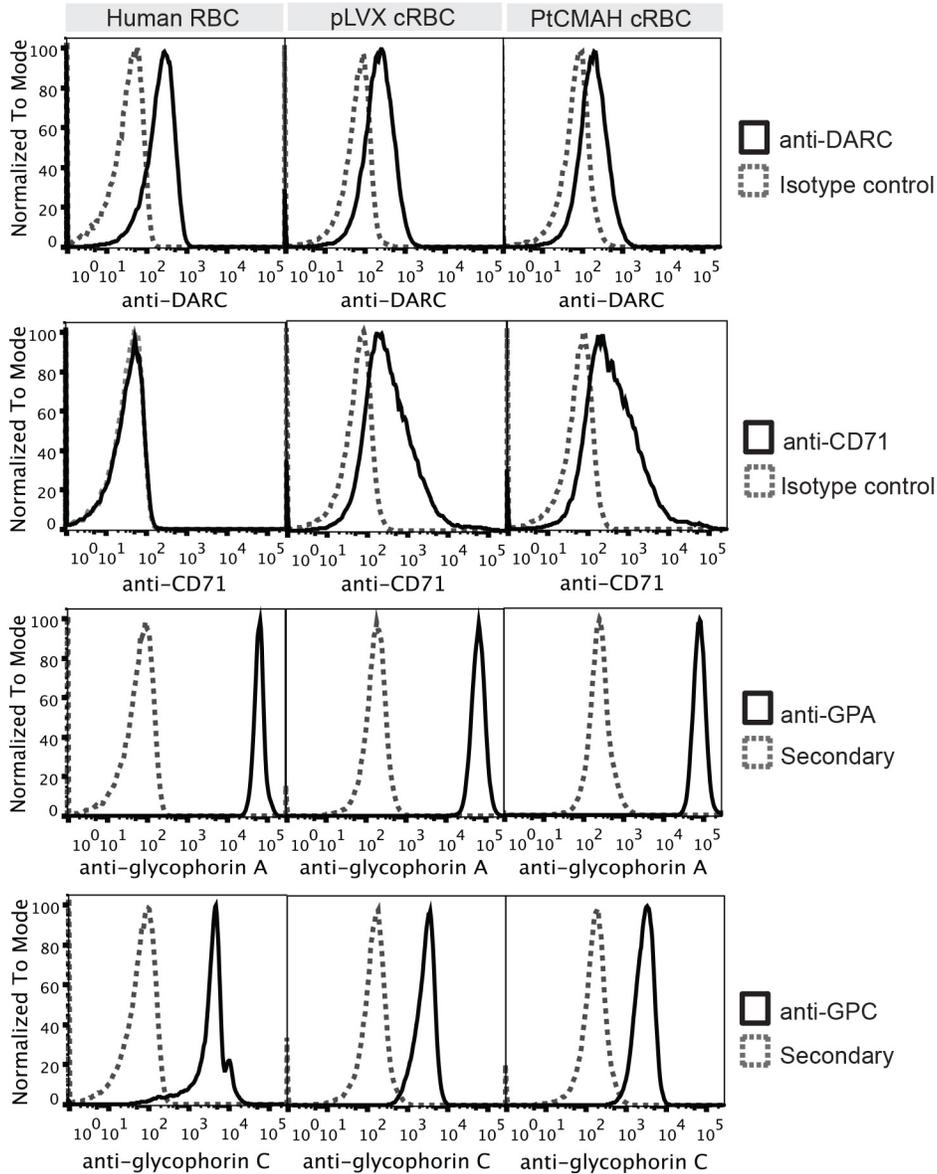
**Supplementary Information**



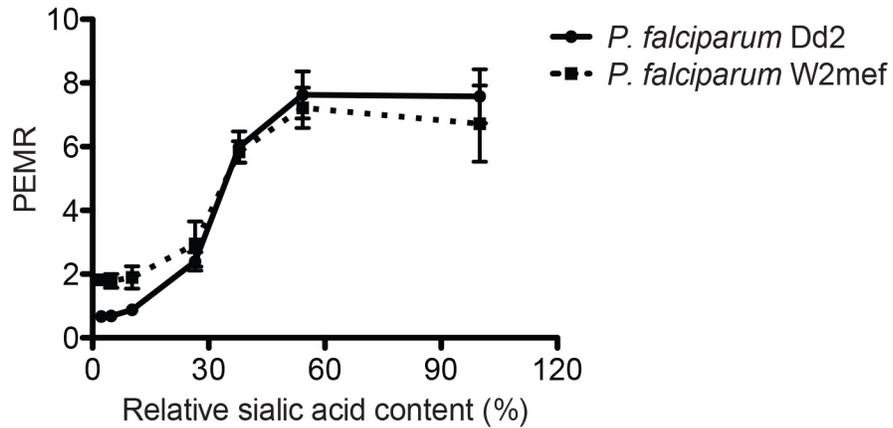
**Supplementary Figure 1. Loss of *CMAH* has occurred independently in a few mammalian lineages including the human lineage.** The phylogenetic tree for a range of mammalian species indicates the presence or absence of *CMAH* as determined by genome alignments and gene loss analysis (CAFE). Loss of *CMAH* is highlighted in red and dating of the loss, where known, is indicated. Mya – million years ago.



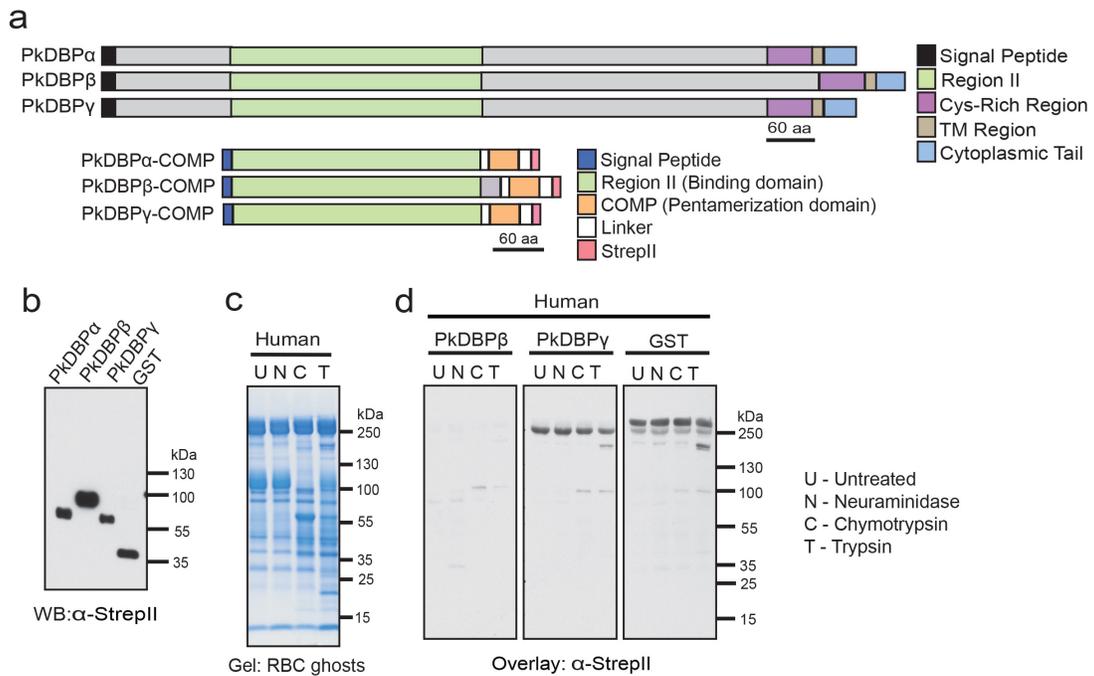
**Supplementary Figure 2. Humans, Neanderthals and Denisovans share the *CMAH* exon deletion.** The *CMAH* gene locus in the chimpanzee genome was used as the reference for alignments. Sequence reads of the Neanderthal genome and the Denisovan genome were aligned to chimpanzee genome PanTro3 at the *CMAH* gene locus with the UCSC genome browser. For the Neanderthal genome, sequence contigs were plotted. For Denisovan sequence reads, darker shades of grey indicate higher level of sequence similarity. All hominin sequences show a similar gap in the alignments. Chr6 – chromosome 6.



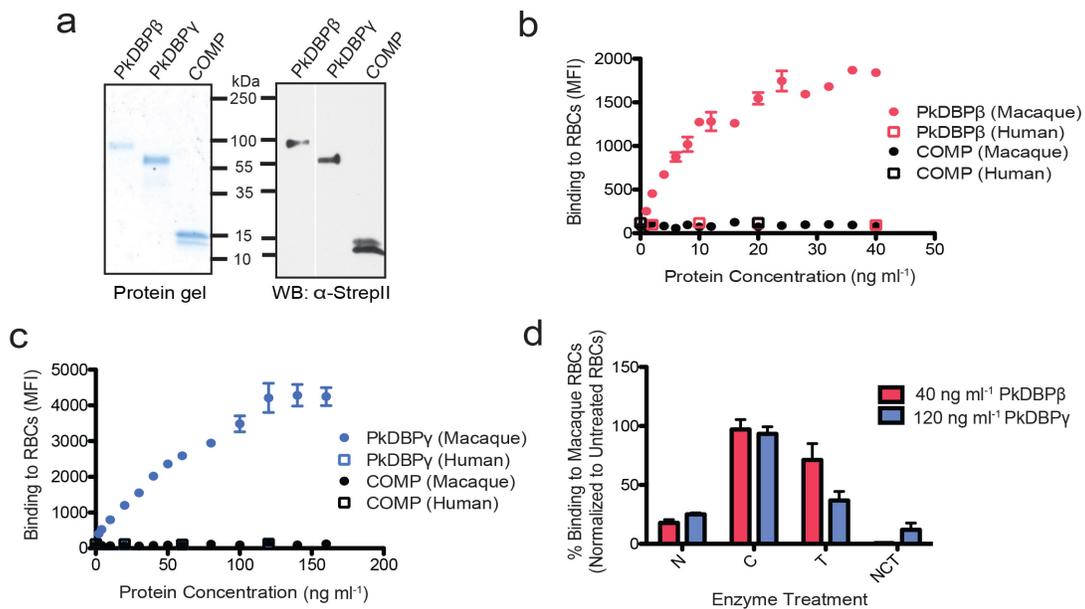
**Supplementary Figure 3. Similar maturation of PtCMAH cRBCs and pLVX cRBCs.** Flow cytometry plots showing expression of DARC, CD71, glycophorin A (GPA) and glycophorin C (GPC) on human RBCs and PtCMAH and pLVX cRBCs. Negative control samples were either stained with an anti-mouse IgG2a-PE antibody (isotype control) or an anti-mouse IgG-Alexa Fluor 488 antibody (secondary). Normalized to mode – normalization to the modal fluorescence value.



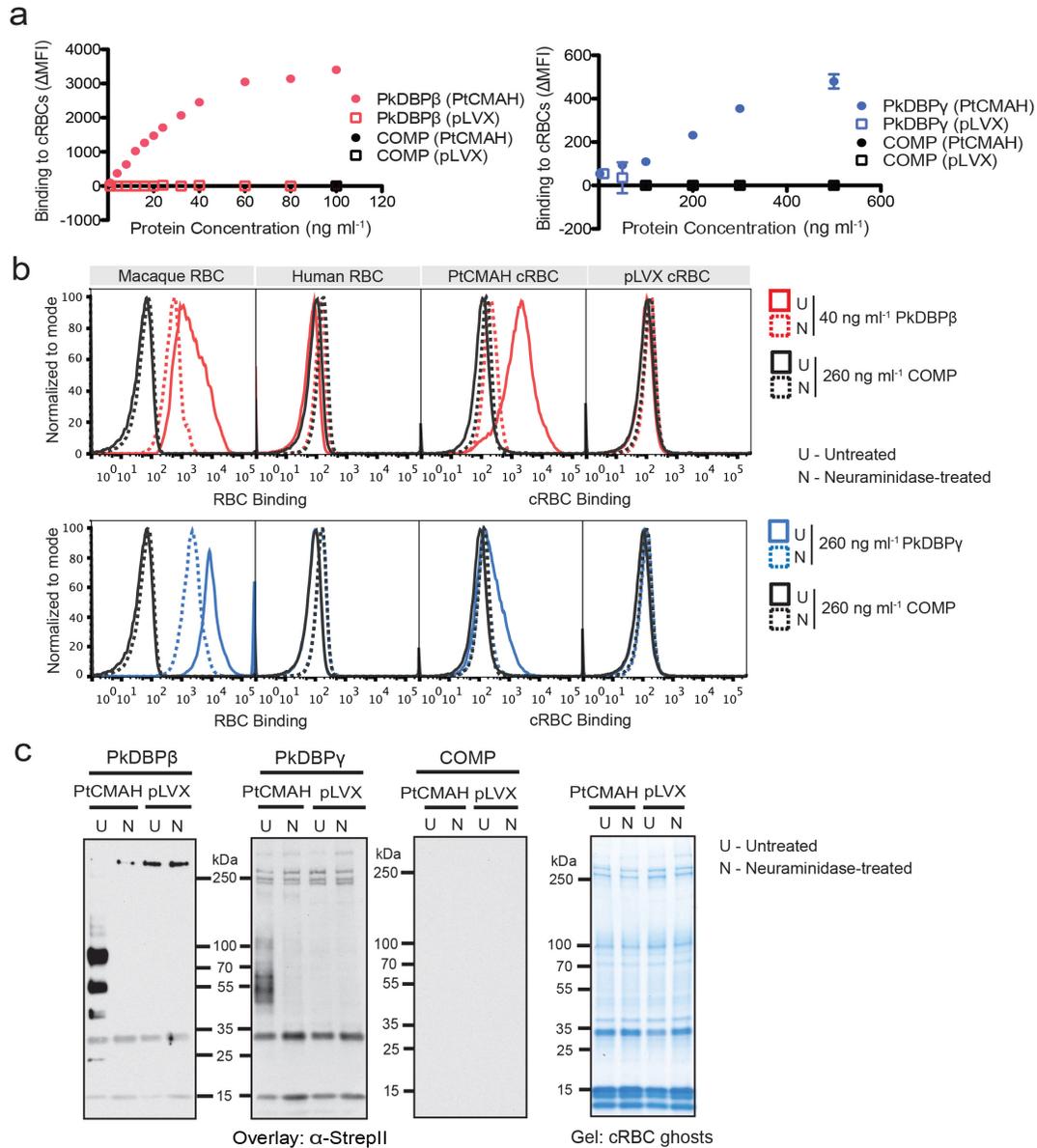
**Supplementary Figure 4. Sialic acid dependence in *P. falciparum* invasion.** Shown is the Parasitized Erythrocyte Multiplication Rate (PEMR) of *P. falciparum* sialic acid-dependent laboratory strains, Dd2 and W2mef, into human RBCs treated with decreasing concentrations of neuraminidase. The sialic acid remaining on RBCs after neuraminidase treatment relative to untreated RBCs is plotted. The data for one biological replicate performed in triplicate is shown. Error bars represent the s.d.



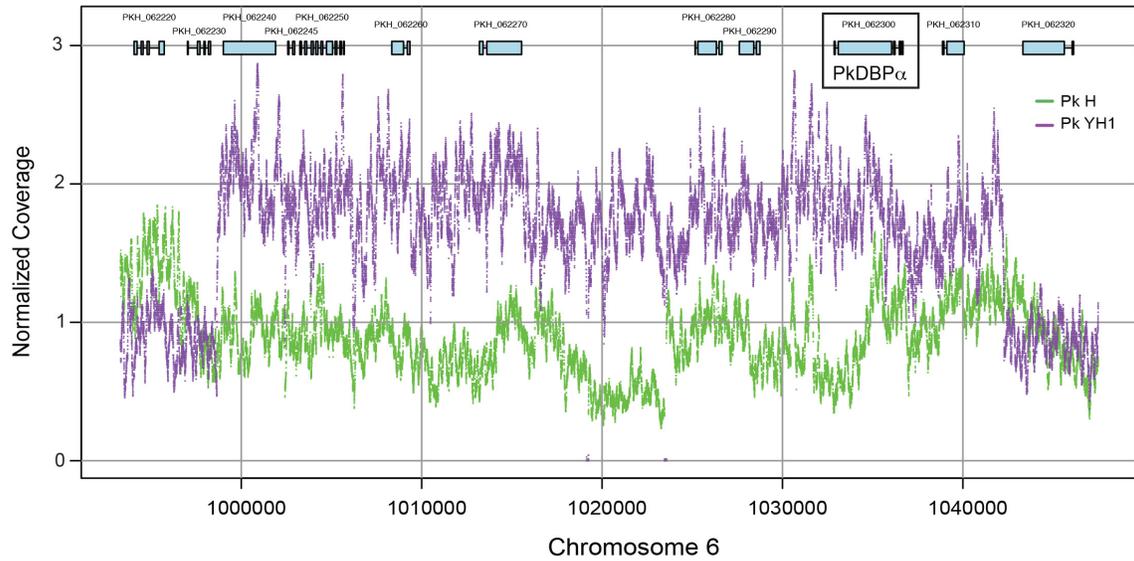
**Supplementary Figure 5. Domain structures, expression of PkDBP $\alpha$ , PkDBP $\beta$  and PkDBP $\gamma$  and lack of binding of PkDBP $\beta$  and PkDBP $\gamma$  to human RBCs. (a) Domain structures of native PkDBP $\alpha$ , PkDBP $\beta$  and PkDBP $\gamma$  and expression constructs of the Duffy binding domain (Region II) of each protein. (b) Western blot showing expression of strepII-tagged PkDBP $\alpha$ -COMP, PkDBP $\beta$ -COMP, PkDBP $\gamma$ -COMP and GST-COMP protein. (c) Coomassie-stained protein gel of enzyme-treated and untreated human RBC ghosts. U – untreated, N – neuraminidase-treated, C – chymotrypsin-treated, T - trypsin-treated. (d) PkDBP $\beta$ -COMP, PkDBP $\gamma$ -COMP and GST-COMP control protein do not bind to untreated or enzyme-treated human RBC ghosts. Data are representative of at least 4 independent experiments. (See Fig. 3b for binding to rhesus macaque RBC ghosts.)**



**Supplementary Figure 6. Dose-dependent and saturable binding of PkDBPβ and PkDBPγ to rhesus macaque RBCs.** (a) Coomassie-stained protein gel (Left), and α-StrepII Western blot (Right) of purified PkDBPβ-COMP, PkDBPγ-COMP and COMP protein. (b,c) Binding of PkDBPβ-COMP (b; red circles) and PkDBPγ-COMP (c; blue circles) to macaque RBCs in solution increases to the point of saturation with increasing concentration of the respective ligand. COMP protein (black) does not bind to RBCs and neither PkDBPβ nor PkDBPγ binds to human RBCs (red, blue open squares). The flow cytometry-based assay, representative of two biological replicates, was seeded in triplicate; error bars represent the s.d. (d) Binding of saturating concentrations of PkDBPβ-COMP and PkDBPγ-COMP to untreated and enzyme-treated macaque RBCs. The mean fluorescence intensity (MFI) was determined for binding of PkDBPβ-COMP and PkDBPγ-COMP to enzyme-treated macaque RBCs, then normalized to ‘no protein’ controls and further normalized to untreated macaque RBCs. Shown is the average of three biological replicates, performed in triplicate. Error bars represent the s.e.m. N – neuraminidase, C – chymotrypsin, T – trypsin, NCT – neuraminidase, chymotrypsin, trypsin.



**Supplementary Figure 7. PkDBP $\beta$  and PkDBP $\gamma$  bind to Neu5Gc-sialylated receptors.** (a) Dose-dependent binding of PkDBP $\beta$ -COMP (red; left panel) and PkDBP $\gamma$ -COMP (blue; right panel) to PtCMAH cRBCs (circles) but not pLVX cRBCs (squares). The assay was seeded in duplicate; error bars indicate the range. (b) PkDBP $\beta$ -COMP (red; top panel) and PkDBP $\gamma$ -COMP (blue; bottom panel) bind to untreated (U; solid trace) PtCMAH cRBCs and rhesus macaque RBCs, but not to pLVX cRBCs or human RBCs. Binding to neuraminidase-treated (N; dashed trace) PtCMAH cRBCs and macaque RBCs is greatly decreased. COMP protein (black) does not bind to any cell type. Each binding condition was set up in triplicate; a representative trace is shown. The binding assay was done in 2 biological replicates (see Fig. 3c). (c) Left – Protein overlays showing binding of PkDBP $\beta$ -COMP and PkDBP $\gamma$ -COMP to PtCMAH, but not pLVX cRBC ghosts or neuraminidase-treated PtCMAH cRBC ghosts. Right – Coomassie-stained protein gel showing protein from cRBC ghost samples used in protein overlays.



**Supplementary Figure 8. Duplication of a region of chromosome 6 containing PkDBP $\alpha$  in the Pk YH1 human-adapted strain.** Shown is the sequence read coverage normalized to the average genome-wide read coverage for Pk YH1 and the parental Pk H strain aligned against the *P. knowlesi* H reference genome. The normalized read coverage for the duplicated region is 2. The genes within the depicted region are shown above the corresponding genomic sequence.

**Supplementary Table 1. Sialic acid content of red blood cell membranes**

Experiment <sup>a</sup>	RBC Sample	Sialic Acid <sup>b</sup>				
		% Neu5Gc	% Neu5Ac	Neu5Gc (pmol)	Neu5Ac (pmol)	Total (pmol)
1	PtCMAH	61.9	38.1	130.6	80.4	211.0
2	PtCMAH	80.8	19.2	225.8	53.7	279.5
3	PtCMAH	74.2	25.8	372.0	129.1	501.0
1	pLVX	ND <sup>c</sup>	100.0	ND	159.8	159.8
2	pLVX	ND	100.0	ND	175.4	175.4
3	pLVX	ND	100.0	ND	319.6	319.6
1	Human	ND	100.0	ND	262.3	262.3
2	Human	ND	100.0	ND	170.7	170.7
3	Human	ND	100.0	ND	271.9	271.9
1	Macaque	98.9	1.1	155.1	1.7	156.8
2	Macaque	100.0	ND	235.7	ND	235.7
3	Macaque	93.2	6.8	325.1	23.7	348.8

<sup>a</sup>Biological Replicate

<sup>b</sup>Sialic acid was released from red blood cell ghosts prepared from  $1 \times 10^7$  cells by mild acid hydrolysis, derivatized by reaction with the fluorogenic agent, 1,2-Diamino-4,5-methylenedioxybenzene, and analysed by HPLC. Sialic acid quantity was estimated from Neu5Ac and Neu5Gc standards.

<sup>c</sup>ND – Not detected by HPLC

**Supplementary Table 2. Genes in the duplicated region of chromosome 6 in Pk YH1**

Gene ID <sup>a</sup>	Gene Name <sup>a</sup>	Signal Peptide? <sup>b</sup>	Transmembrane Prediction <sup>c</sup>
PKH_062230	antigen UB05, putative	No	2 (15-37, 58-80)
PKH_062240	GDP dissociation inhibitor, putative	No	0
PKH_062245	conserved Plasmodium protein, unknown function	No	0
PKH_062250	conserved Plasmodium protein, unknown function	No	0
PKH_062260	Plasmodium exported protein, unknown function, pseudogene	No	1 (35-54)
PKH_062270 <sup>d</sup>	erythrocyte binding protein, putative	No	1 (40-59)
PKH_062280	KIR-like protein	No	1 (333-355)
PKH_062290	Plasmodium exported protein, unknown function	No	1 (44-66)
PKH_062300	erythrocyte binding protein (alpha) (PkDBP $\alpha$ )	Yes	1 (1009-1031)
PKH_062310	Plasmodium exported protein, unknown function	Yes	3 (4-21, 276-295, 300-322)

<sup>a</sup>Obtained from PlasmoDB

<sup>b</sup>Signal peptide prediction using locally installed SignalP 4.1.

<sup>c</sup>Transmembrane prediction using locally installed TMHMM 2.0. The amino acids spanning each transmembrane region are in parentheses.

<sup>d</sup>Of the seven predicted membrane proteins, apart from PkDBP $\alpha$ , only PKH\_062270 is annotated as an erythrocyte binding protein. This protein, however, has a PEXEL motif, designating it as an exported protein and making it unlikely to play a role in invasion.

**Supplementary Table 3. Primer sequences**

Gene	Gene ID <sup>a</sup>	Primer	Sequence
CMAH	AF074481.1	Fwd <sup>b</sup>	ggaggGGATCCCAAGCTTGAATTCGCCACCA TGG
		Rev <sup>c</sup>	ggaggTCTAGAATCGATTTCGAATCATCACA GGTCCTC
Pk ssuD	PKH_050112	Fwd	gcgcGGATCCCTTGTCTCAAAGATTAAGCCA TGC
		Rev	gcgcGTCGACTTCACCTACGGAAACCTTGTT ACG
PkDBP $\alpha$	PkH_062300	Fwd	ggaggGCTAGCGTTATTAATCAAACCTTTTCTT CAAACAATGTAATGGACAAGTGTAATGA TAAGAGAAAACG
		Rev	ggaggGCGGCCGCACCTGATTGAGAACCTGG ATCAGCTTCATTTGTGCTAGACTTACCTTC ACCTTTG
PkDBP $\beta$	PkH_000490	Fwd	ggaggGCTAGCGTTATTAATCAAACCTTTTCTT CAGAAGAATGTAATGAGGAGG
		Rev	ggaggGCGGCCGCGAATTTATACGTATCAAC ATTTTTGGCAGCC
PkDBP $\gamma$	PkH_134580	Fwd	ggaggGCTAGCGCTATTAATCAAATTTTTCTT CAAACAATGTAATGG
		Rev	ggaggGCGGCCGCACCTGGTTGGGAACCTGG
GST		Fwd	ggaggGCTAGCTCCCCTATACTAGGTTATTG GAAAATTAAG
		Rev	ggaggGCGGCCGCATCCGATTTTGGAGGATG GTC

<sup>a</sup>Obtained from GenBank (CMAH) or PlasmDB (*P. knowlesi* genes)

<sup>b</sup>Forward

<sup>c</sup>Reverse