Classification: Biological Sciences, Microbiology

Title: Normocyte binding protein required for human erythrocyte invasion by the zoonotic malaria parasite *Plasmodium knowlesi*

Short Title: Human erythrocyte invasion by *P. knowlesi*

Authors: Robert W. Moon[a,b], Hazem Sharaf[c], Claire H. Hastings[a], Yung Shwen Ho[c], Mridul B. Nair[c], Zineb Rchiad[c], Ellen Knuepfer[a], Abhinay Ramaprasad[c], Franziska Mohring[b], Amirah Amir[d], Noor A. Yusuf[a], Joanna Hall[e], Neil Almond[e], Yee Ling Lau[d], Arnab Pain[c,f], Michael J. Blackman[a,g] and Anthony A. Holder[a]

Author affiliations:

[a] The Francis Crick Institute, Mill Hill Laboratory, The Ridgeway, Mill Hill, London, NW7 1AA, UK

[b] Department of Immunology and Infection, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, WC1E 7HT, UK

[c] Pathogen Genomics Laboratory, Biological and Environmental Sciences and Engineering (BESE) Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Jeddah, 23955-6900 Kingdom of Saudi Arabia

[d] Department of Parasitology, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia

[e] Division of Retrovirology, National Institute for Biological Standards and Control, Health Protection Agency, South Mimms, Potters Bar, Hertfordshire EN6 3QG, UK

[f] Global Station for Zoonosis Control, Global Institution for Collaborative Research and Education (GI-CoRE); Hokkaido University; N20 W10 Kita-ku, Sapporo, 001-0020 Japan

[g] Department of Pathogen Molecular Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, WC1E 7HT, UK

Corresponding Authors: Anthony A. Holder, The Francis Crick Institute, Mill Hill Laboratory, The Ridgeway, Mill Hill, London, NW7 1AA, UK. Tel: +44 2088162175, Email: tony.holder@crick.ac.uk

Robert W. Moon, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK, Tel: +442079272743, Email: rob.moon@lshtm.ac.uk

Keywords: Invasion; Malaria parasite; Plasmodium knowlesi; normocyte binding protein; zoonotic malaria

Abstract

The dominant cause of malaria in Malaysia is now *Plasmodium knowlesi,* a zoonotic parasite of cynomolgus macaque monkeys found throughout South East Asia. Comparative genomic analysis of parasites adapted to in vitro growth in either cynomolgus or human red blood cells (RBCs) identified in parasites growing in cynomolgus RBCs but not human RBCs a genomic deletion that includes the gene encoding normocyte binding protein Xa (NBPXa). Experimental deletion of the *NBPXa* gene in parasites adapted to growth in human RBCs (which retain the ability to grow in cynomolgus RBCs) restricted them to cynomolgus RBCs, demonstrating that this gene is selectively required for parasite multiplication and growth in human RBCs. NBPXa-null parasites could bind to human RBCs but invasion of these cells was severely impaired. NBPXa is therefore identified as a key mediator of *P. knowlesi* human infection and may be a target for vaccine development against this emerging pathogen.

Significance statement

Plasmodium knowlesi is a parasite that naturally infects cynomolgus monkeys, but is also a major cause of severe zoonotic malaria in humans in South East Asia. Comparing the genomes of parasites restricted to growth in culture with cynomolgus red blood cells (RBCs) with those adapted to growth in human RBCs identified a gene specifically required for invasion of human red blood cells (RBCs), a process that is critical for parasite replication. This gene encodes NBPXa, a protein previously shown to bind human RBC and implicated in invasion. Disruption of this gene blocks invasion of human but

not cynomolgus RBCs, thus confirming a key mediator of human infection and a potential target for inclusion in vaccines to prevent human infection.

\body

Introduction

Plasmodium knowlesi, a malaria parasite of the cynomolgus macaque, *Macaca fascicularis*, is the causative agent of severe zoonotic disease in South East Asia (1, 2). In countries such as Malaysia where other malaria parasites are being controlled effectively, *P. knowlesi* infections are the dominant cause of human malaria (3, 4). In all malaria parasite species, disease is caused by cyclic parasite multiplication within red blood cells (RBCs), releasing invasive merozoites that invade new RBCs. Parasite adhesins transported onto the merozoite surface bind specific host cell receptors to facilitate RBC invasion (reviewed in (5)). The first adhesin-RBC interaction was identified in *P. knowlesi* and *P. vivax*, where it was shown that parasites must bind to the Duffy Antigen/Chemokine Receptor (DARC) to invade human RBCs (6). These adhesins are determinants of parasite virulence (7), host cell tropism (8), and potential vaccine candidates (9), and understanding the role of these adhesins may be critical to identifying the causes of increased risk of human infection and for the development of novel interventions.

Reticulocyte binding proteins (RBPs) were first described in *Plasmodium vivax* (10) and are the prototypical examples of the reticulocyte binding like/reticulocyte binding homologue (RBL/RH) proteins also characterized in other malaria parasite species including *P. cynomolgi* (11) and *P. yoelii* (12). *P. falciparum* normocyte binding proteins (NBPs) (13, 14) are members of this family that includes four conventional RBL/RH proteins and one unconventional short RBL/RH protein (RH5) in this species (reviewed in (5)). In *P. knowlesi* two proteins in the RBL family known as Normocyte Binding Protein (NBP)Xa and NBPXb have been identified (15), and recombinant fragments of these proteins bind RBCs (16). The RBL/RH adhesins and the Duffy binding protein/erythrocyte binding ligand (DBP/EBL) family of proteins, play key roles during merozoite invasion of RBCs (reviewed in (5,

9)). However, ascribing specific roles to these proteins has been hampered by functional redundancy and only RH5 may be essential for *P. falciparum* invasion (17). In *P. knowlesi,* DBPα is the parasite adhesin binding DARC, and two paralogues (DBPβ and DBPγ) have also been identified (18), but thus far only DBPα has been shown to be required for RBC invasion (19).

P. knowlesi can be readily grown in both cynomolgus and rhesus macaque (*Macaca mulatta*) RBCs in vitro (20-22), but requires extended adaptation for growth in human RBCs in vitro displaying insufficient invasion efficiency in human RBCs to support continuous culture. In previous work we overcame this impediment by extended adaptation in a mixture of human and cynomolgus RBCs, producing a parasite line that can be maintained in human RBCs alone (21). This previous analysis suggested that improved parasite invasion of human RBCs was essential for maintenance of parasite growth in vitro.

Here, we undertook a comparative analysis of the genomes of *P. knowlesi* lines adapted to human or cynomolgus RBCs in culture, in parallel with that of the pre-adapted parasites. The comparison highlighted a crucial role for NBPXa in invasion of human RBCs. Using the unique capacity to grow this parasite in two host cell types we were able to disrupt the gene encoding NBPXa and show it is required for invasion of human but not cynomolgus macaque RBCs. Human-adapted *P. knowlesi* parasites provide a robust experimental platform to unravel mechanisms of host cell invasion in both *P. knowlesi* and the related pathogen *Plasmodium vivax*.

Results

Culture adaptation of *P. knowlesi* is associated with distinct large genomic deletions and duplications. To identify genetic changes occurring during adaptation to in vitro culture we compared the parasite genomes at two critical points in the process: within 2 months of introducing the parasite to culture (A1-O), and after full adaptation to growth in human RBCs (A1-H) or in cynomolgus RBCs (A1-C) for a similar period (see Fig. 1A). Genome sequence analysis confirmed that the starting A1-O parasite is closely related to the H-strain (23, 24). Use of the PacBio

sequencing platform allowed the generation of high quality genome assemblies (Table S3) for clones of the adapted parasite lines (A1-H.1 and A1-C.2) to use as references to identify SNPs and INDELs (insertions/deletions) amongst both the uncloned lines and the other derived clones. Few nucleotide differences were identified (Fig. 1A and Table S1), with only 22 non-synonymous variants in protein coding genes throughout all lines sequenced. Only three of these differences were found exclusively in the human adapted parasite and none was present in all clones, so they were unlikely to be responsible for the adaptation phenotype (Table S1). Similarly there were no non-synonymous substitutions present in long-term culture adapted lines but absent from the A1-O parasite. In addition to these SNPs, four larger deletions (3.6-66.3 kb) and one duplication (7.7 kb) were identified. In contrast to the SNPs, these appeared to be specific to either human or cynomolgus adapted lines, resulting in deletion of 13 genes in A1-C derived lines, or deletion of 2 genes and duplication of 4 genes in A1-H derived lines (Table S2). The largest deletion was of ~66.3 kb at one end of chromosome 14 in the A1-C line (Fig. 1B and Fig S1). In A1-O and A1-H lines, this region encompasses 12 genes (Table S2), including PKNH_1472300 that encodes NBPXa (15). This observation was of particular interest as adaptation to human RBCs appeared to result from improved invasion efficiency rather than improved intracellular growth (21). The genomic comparisons identified no substantial differences in other known invasion gene loci, including *DBPα*, *DBPβ*, *DBPγ* and *NBPXb* genes. Similarly, whilst we detected a duplication in A1-H derived lines, which contained 4 genes including *GTP cyclohydrolase I*, we found no evidence of increased copy number for any of the DBP or RBL genes. However, using our data and a set of 48 *P. knowlesi* genomes from malaria infections in Malaysia (25), a high level of SNP diversity can be shown for the *NBPX and DBP* genes (Figure S3, Table S4).

Adaptation to human RBCs results in increased invasion efficiency for both human and cynomolgus RBCs. The genetic differences between cynomolgus and human RBC-adapted parasite lines, including the loss of *NBPXα* in the A1-C line, suggested that the A1-C and A1-H lines had been subjected to distinct selection processes during culture adaptation. Therefore growth of these lines

was compared with that of the A1-O line to try to correlate genetic and phenotypic differences (Fig. 2). To compare invasion efficiencies, schizonts purified from each line grown in cynomolgus RBCs were added to fresh human or cynomolgus RBCs, and then new intracellular parasites were quantified 24 h later. The invasion efficiency of the A1-O line in cynomolgus RBCs was around half that of both the A1-H.1 and A1-C.1 lines, consistent with selection for improved invasion following prolonged culture (Fig. 2A). Further examination of the cultures showed no evidence for defects in subsequent intracellular growth (Fig. 2B and C). In human RBCs the invasion efficiency of the A1-O line was approximately half that of the A1-H.1 line, but four times that of the A1-C.1 line. Thus, whilst the A1-H.1 line increased in invasion efficiency for both human and cynomolgus RBCs during more than a year in vitro, A1-C.1 achieved the same only in cynomolgus RBCs, with a reduction in invasion efficiency for human RBCs. Using the 24 hour invasion data to calculate a mean host RBC selectivity ratio for the different lines (cynomolgus RBC invasion efficiency/human RBC invasion efficiency) revealed no significant differences between host cell selectivity of the A1-O and A1-H.1 lines (cynomolgus selectivity [+/- 1 s.d.] = 1.53 +/-0.14 and 1.54 +/-0.49, respectively), clearly demonstrating that no specific increase in human RBC invasion efficiency had occurred in the A1-H.1 line. In contrast, the selectivity of the A1-C.1 line for cynomolgus RBC had significantly increased (cynomolgus selectivity = 16.1+/-8.1) compared to both the A1-H.1 and A1-O lines (1-way ANOVA, p=0.0044). The absence of *NBPXα* in the A1-C.1 line clearly did not reduce its invasion efficiency for cynomolgus cells. We therefore reasoned that its loss may have caused the shift in host cell selectivity through reduced invasion efficiency for human RBCs. The ability of the A1-H.1 line to grow interchangeably in the two host RBCs offered a unique opportunity to test this hypothesis. No other clear relationship between genotype and adaptation to culture was established.

The *NBPXα* gene can be disrupted in A1-H.1 parasites grown in cynomolgus but not human RBCs.

To analyse the role of *NBPXα*, A1-H.1 parasites were transfected with a gene disruption construct (Fig. 3A) and maintained under drug selection in either human or cynomolgus RBCs until resistant parasites emerged (8-15 days). Analysis of these parasites by PCR revealed successful disruption of

NBPXα in parasites from independent transfections maintained in cynomolgus RBCs (n=5) but not in parasites grown in human RBCs (n=3), despite successful integration of control plasmid (Fig. 3B). Thus, the *NBPXα* gene appeared to be refractory to disruption in parasites grown in human but not cynomolgus RBCs, indicating a host species-specific requirement for this gene. Two independent Δ*NBPXα* cloned lines were derived from A1-H.1 parasite transfections in cynomolgus RBCs and used for subsequent phenotypic analysis.

***NBPXα* is unnecessary for invasion of cynomolgus RBCs but crucial for invasion of human RBCs.** We examined the ability of the Δ*NBPXα* clones to invade and grow within either human or cynomolgus RBCs in the 24 hour invasion assay. Disruption of *NBPXα* had no effect on invasion and growth in cynomolgus RBCs, but invasion of human RBCs was significantly impeded, matching the phenotype of the A1-C.1 line (Fig. 4A and Fig. S2). Whilst the defect in human RBC invasion caused by loss of the *NBPXα* gene was severe and prevented multiplicative growth, a low frequency of successful invasions were detected for both the A1-C.1 and *ΔNBPXα* lines. To rule out the possibility that this was due to contamination of purified schizonts with cynomolgus RBCs, the parasites were allowed to invade RBCs pre-labelled with a fluorescent dye (26). The results showed that whilst contaminating cynomolgus RBCs contributed to a slight overestimation of invasion efficiency in human RBCs, a small fraction of both the *ΔNBPXα* and A1C.1 parasites was able to infect human RBCs. Parasitemia measurements at 48 h showed that the *ΔNBPXα* human RBC invasion phenotype was expressed in subsequent cycles, and the parasite could not undergo multiplicative growth in human RBCs (Fig. 4B and C). Extended culture of *ΔNBPXα* parasites in human RBCs resulted in a rapid reduction of parasitemia consistent with the severe defect. Although some parasites were detectable after 27 days culture these were revertants, which had deleted the *NBPXα* disruptant construct (Fig. S2B). These data demonstrate the severity of the *ΔNBPXα* phenotype, accounting for both the inability to disrupt the *NBPXα* gene and the severely reduced growth of the A1-C.1 line in human RBCs.

Further analysis showed that the *ΔNBPXα* block occurred within the first 7 h of the assay, consistent with a defect in invasion rather than in subsequent intracellular growth (Fig. 4B and C);

examination of Giemsa-stained blood smears taken at 7 h showed that whilst A1-H.1 parasites had successfully developed as ring stages in human RBCs, most of the *ΔNBPXa* parasites appeared to be extracellular merozoites that had attached to, but not invaded RBCs (Fig. 4D and E). To see whether or not this observation was consistent with a phenotype expressed subsequent to RBC binding, we used an immunofluorescence-based merozoite surface protein 1 (MSP1) processing assay (27) to distinguish merozoites that had successfully invaded and formed rings ($MSP1_{19}$ positive only) from those that had attached but failed to invade, ($MSP1_{33}$ and $MSP1_{19}$ double positive) (Fig.5A). Whilst 92% of A1-H.1 merozoites successfully formed rings in human RBCs, 76% of *ΔNBPXa* merozoites attached but failed to invade human RBCs (Fig 5B). In contrast, about 90% of merozoites from both lines successfully formed rings in cynomolgus RBCs (Fig 5B). These results show that *NBPXa* disruption results in a block in invasion following initial merozoite attachment.

To ensure that up (or down) regulation of another gene was not responsible for the observed phenotype, RNA-seq analysis of schizont stages of A1-H.1 and *ΔNBPXa* parasites was performed. This did not identify any significantly differentially expressed genes.

Deletion of NBPXa does not result in restriction to human reticulocytes. Previous work by others had suggested that the ability to invade and grow in human reticulocytes as well as normocytes is a key parasite adaptation to culture in human RBCs (22). Our observation that both the *ΔNBPXa* and A1-C.1 lines invaded a small percentage of human RBCs raised the possibility that loss of NBPXa function restricted the parasite to invasion of reticulocytes, which typically constitute 0.5-2% of total RBCs. To investigate this possibility, we used an adapted protocol (28, 29) to enrich reticulocytes to around 10% of human RBCs and used these cells for invasion assays. Reticulocyte enrichment resulted in a small increase in invasion efficiency over normal human RBCs for the A1-O, A1-H.1, A1-C.1 and *ΔNBPXa* lines, suggesting that younger RBCs are favoured by all lines (Fig. S2). However, these increases were insufficient to explain the *ΔNBPXa* phenotype as due to reticulocyte restriction. Furthermore, the few successfully invaded *ΔNBPXa* parasites were present in both reticulocytes and normocytes.

***NBPXa* is unnecessary for invasion of rhesus RBCs.** Disruption of *NBPXa* had no effect on invasion into rhesus RBCs (Fig. S2) although invasion efficiency was significantly higher for cynomolgus RBCs than rhesus RBCs for all lines tested (A1-H.1, A1-C.1, and *ΔNBPXa*), suggesting that the parasite has slightly different adhesin specificities even for these closely related hosts, although this difference may also be explained by phenotypic culture adaptation to cynomolgus RBCs.

Discussion

In this work we set out to identify parasite features that facilitate growth in human RBCs in vitro, exploiting *P. knowlesi* lines that we previously adapted to growth in either human or cynomolgus RBCs. Comparative genome analysis of the culture adapted lines revealed relatively few non-synonymous SNPs and none of these associated universally with either adaptation to human RBCs or long-term culture. In contrast, we detected several large genome deletions and a duplication, which were present in either human or cynomolgus adapted lines but not both and present in the uncloned lines as well as clones derived from them, suggesting that selection had led to fixation within the population. Further analysis of these deletions, either by individual gene deletion as we have demonstrated for *NBPXa*, or by gene complementation may be informative. The duplication in A1-H derived lines contains the gene coding for GTP cyclohydrolase, a rate limiting enzyme in folate metabolism, and duplications of this gene are associated with drug resistance in South East Asian strains of *P. falciparum* (30). Increased activity in this pathway and metabolic adaptation may provide a general selective advantage for growth in culture.

Correct assembly and read mapping for several key gene families, in particular the *DBP*s, was only possible with the use of PacBio sequencing, which produces longer read lengths. Using these data as a scaffold to map the higher coverage but shorter Illumina reads facilitated high quality de novo assemblies of both A1H.1 and A1-C.2 cloned lines, which complement the existing H-strain data and will benefit further work with the culture-adapted *P. knowlesi* lines.

Our previous comparison of cynomolgus- and human-adapted *P. knowlesi* parasite lines had suggested that an increase in human RBC invasion efficiency enabled the parasite to adapt to long term culture in human RBCs (21). However, the additional comparison with the invasion characteristics of the A1-O line, which had been in culture for less than 2 months, demonstrated that the host cell specificity of the A1-H lines had remained unchanged. Instead the overall invasion efficiency for both human and cynomolgus RBCs had improved, with invasion of human RBCs above the minimum threshold required to support in vitro multiplication and growth. In contrast, the A1-C line (and its derived clones) increased its invasion efficiency only for cynomolgus RBCs. This increased selectivity is likely the result of the loss of a section of one end of chromosome 14 containing the gene for NBPXa which we have demonstrated to be crucial for human RBC invasion. Given that the cynomolgus RBC invasion phenotype of both the A1-C.1 and A1-H.1 are indistinguishable, it is quite possible that if A1-C parasites had not lost *NBPXa*, invasion efficiency for human RBCs in culture would have risen to levels equivalent to that seen for A1-H.1 without prior exposure to human RBCs. In principle this could be investigated by genetic complementation of the lost *NBPXa* gene within the A1-C.1 line, but so far such efforts have been hampered by its large size (8.5 kb). The reasons for the different behaviour of the H-strain parasite (which was isolated from a patient (23)) in vivo and in vitro are unclear but our results suggest that for successful parasite multiplication and survival in vitro, prolonged physiological adaptation to culture is more important than adaptation to a specific host RBC. Our data identified clear differences in invasion efficiencies into RBCs from rhesus and cynomolgus macaques, which is surprising given that these two primate species are closely related (31) and that experimental infection of rhesus monkeys (not a natural host of *P. knowlesi*) results in hyperparasitemia (23) in contrast to the chronic low-level parasitemia characteristic of cynomolgus infection. Recombinant NBPXb expressed on cells binds cynomolgus RBCs four times more effectively than rhesus RBC (16), offering one possible mechanism for this difference in invasion efficiency. Independently established culture-adapted parasite strains developed using rhesus RBCs (20, 22) may have undergone quite different adaptive changes to those

observed for cynomolgus RBCs. It will also be of interest to determine the contribution of metabolic adaptation to growth in different RBCs in culture.

P. knowlesi is phylogenetically closely related to *P. vivax* (32), a malaria parasite which causes the majority of malaria cases outside of Africa and which currently cannot be cultured in vitro. Whilst both these parasites use an orthologous DBP to bind to the DARC receptor on host RBCs (6, 18, 33, 34), there are greater differences in the RBL family. *P. vivax* can only invade reticulocytes, a phenotype attributed to the presence of RBP1 and RBP2, which bind only to reticulocytes (10), and neither of which is conserved in *P. knowlesi*. Orthologues of the *P. knowlesi* RBLs are present in *P. vivax* (35, 36), but given that our data indicate the *P. knowlesi* NBPXa protein is functional in invasion of both normocytes and reticulocytes, it is tempting to speculate that the absence of this protein in *P. vivax* could account for the parasite's reticulocyte restriction. Intriguingly another closely related simian malaria parasite, *P. cynomolgi*, contains orthologues for both vivax and knowlesi specific RBL proteins (11, 37). Transferring orthologues from *P. vivax* or *P. cynomolgi* into *P. knowlesi* may offer opportunities to study both the requirements for and the evolution of *P. vivax* invasion.

The unique advantages of the *P. knowlesi* culture system have enabled us to demonstrate that NBPXa plays a crucial role in invasion of human RBCs. Its role appears to be after initial merozoite attachment, perhaps during establishment of the moving junction between RBC and parasite plasma membranes which is required for invasion. In other *Plasmodium* species, both the RBL and DBP protein families are highly redundant and ascribing precise roles has been difficult, although in *P. falciparum* they have been implicated in triggering signalling and release of secretory organelles (38). The ability to grow *P. knowlesi* interchangeably in different host cells with distinct adhesin requirements enables the generation of lines with gene knockouts for important parasite adhesins like NBPXa, providing a powerful and unique opportunity to investigate their precise role in red cell invasion.

In *P. knowlesi* the adhesin repertoire is smaller than in many malaria parasites (9, 15, 18), and loss of either *DBPa* (6, 19) or *NBPXa* creates a critical block in invasion of human RBCs, perhaps

because the paralogues cannot bind human RBCs; for example whilst PkNBPXa and DBP*a* proteins bind to both human and macaque RBCs, DBPβ, DBPγ and NBPXb bind only to macaque RBCs (16, 39). This apparent lack of adhesin redundancy for human RBC invasion suggests that *P. knowlesi* is vulnerable to perturbation of DBPα/NBPXa function. Although this lack of redundancy may encourage the use of these antigens in vaccine development their diversity in parasites responsible for human infection (25, 40) provides an important caveat. Vaccine development may be an important strategy for control of *P. knowlesi* malaria given the presence of a primate reservoir in Southeast Asia.

NBPXa polymorphisms are important determinants of high parasitemia and disease severity in *P. knowlesi* infection (7) and may impact on the potential for human to human transmission. . Given that we have shown a critical role for this protein in human RBC invasion, it is possible that increased virulence and multiplication rates of parasites containing particular alleles is due to an improved ability to bind and invade human RBCs. However even the human adapted *P. knowlesi* line has a higher efficiency of cynomolgus RBC invasion, suggesting that invasion efficiency may remain a bottleneck in human infections. Interestingly variation in NBPXb, which had been found not to bind to human RBCs, has also been linked to disease severity in humans (7, 16). The role of NBPXb and variants of both *P. knowl*esi NBPs can now be investigated in vitro. . Human-adapted *P. knowlesi* parasites provide a robust experimental platform to unravel mechanisms of host cell invasion and aid development of strategies to control malaria caused by *P. knowlesi* and the closely related pathogen *P. vivax*.

Methods

Macaque, Human and Reticulocyte-Enriched RBC. Cynomolgus and rhesus blood was collected by venous puncture. Animal work was approved by the United Kingdom Home Office as governed by United Kingdom law under the Animals (Scientific Procedures) Act 1986. Human blood (Groups A^+ and O^+; $Duffy^+$) was obtained from the UK National Blood Transfusion service. Reticulocytes were enriched using an adapted protocol (28, 29) and reticulocytemia assessed by microscopy.

Parasite lines and maintenance. *P. knowlesi* parasite cultures were derived from a frozen *in vivo* stabilate (A1, dated July 22 1975) supplied by Dr Louis Miller, National Institutes of Health and obtained following passage in vivo in rhesus macaque of the *P. knowlesi* H-strain derived from a clinical infection (23). A1-O is a non-clonal cultured line from the A1 stabilate propagated in cynomolgus RBCs in vitro for less than 2 months. A1-H is a non-clonal line derived from A1-O following maintenance in 80% human and 20% cynomolgus RBCs for 10 months and up to 7 months in human RBCs alone, with clones A1-H.1, A1-H.2 and A1-H.3 derived at this time. *P. knowlesi* A1-C is a non-clonal line derived from A1-O after maintenance in 100% cynomolgus RBCs for 15 months, with clones A1-C.1 and A1-C.2 derived at this time. Parasites were grown and synchronized as described previously (21).

***P. knowlesi* Genome Sequencing, Analysis and De Novo Assembly.** DNA for the A1-H.1 and A1-C.2 lines was used for preparation of libraries: HiSeq 2000 (Illumina Inc. USA) compatible genomic shotgun libraries (500 bp fragment size), PCR-free libraries sequenced on MiSeq (Illumina) and for genome assemblies, 20 kbp insert libraries sequenced on a PacBio RS-II instrument (Pacific Biosciences Inc.). Genome assembly was performed using PacBio's FALCON tools, and Genome Consensus suite. Annotation was transferred from the *P. knowlesi* H strain reference genome (24) and gene models were checked in Artemis BAM view and corrected using strand-specific RNA-seq datasets.

Antibody production. The mouse *P. knowlesi* $MSP1_{33}$-specific antibody has been described (27). A *P. knowlesi* $MSP1_{19}$-specific antibody was produced following rabbit immunization with recombinant $MSP1_{19}$. Rat anti-human RBC ghost antibody was prepared using A^+ human erythrocytes and the resulting antiserum used for immunolabeling both human and cynomolgus RBC due to cross reactivity.

Adhesion, invasion and growth assays. Purified schizonts were used to set up triplicate cultures with cynomolgus, rhesus, human or human reticulocyte-enriched RBCs, at a 2% hematocrit and a 1% parasitemia. All parasites (including A1-H.1) were grown in cynomolgus RBCs for at least one

month prior to growth assays, to control for any effect of prior host cell type on subsequent invasion and growth. Parasitemia was measured with a flow cytometry (FACS)-based assay (21, 26) at the start and following incubation at 37°C in a gassed chamber for 7 h, 24 h and 48 h. Samples were fixed , washed, permeabilized with Triton® X-100, and then washed again prior to RNAse treatment, staining with SYBR® Green I (Life Technologies) and FACS analysis. To control for invasion of RBCs carried over during schizont purification, target cells were pre-labelled with a far red fluorescence dye (26) using 7-hydroxy-9H-(1,3-dichloro-9,9-dimethylacridin-2-one) succinimidyl ester (DDAO-SE) (Life Technologies). Following washing the samples were analyzed on a CyAN ADP Analyzer (Beckman Coulter). Data were acquired and analysed using CellQuest (BD) or HyperCyt (Beckman Coulter) software.

To investigate invasion/binding phenotype and reticulocyte preference, samples taken at 7 h were incubated with reticulocyte stain solution (Sigma), then the cells were pelleted and Giemsa-stained thin blood films prepared. To attempt to readapt ΔNBPXa parasites to growth in human RBC, a culture at 1% parasitemia was set up and maintained for 1 month. After parasites were no longer detectable in blood smears (4 days), the culture was maintained by replacement with 30% fresh human RBCs and medium every 2-3 days. The (revertant) parasites that were detectable after 28 days were analysed by PCR together with A1-H.1 and ΔNBPXa controls.

The difference between adhesion and invasion of merozoites was also measured using an immunofluorescence assay. Synchronous *P. knowlesi* A1-H.1 or *ΔNBPXa* schizonts were added to triplicate cultures at 2% parasitemia with cynomolgus or human RBC and after 3 h thin blood films were prepared for immunofluorescence staining. The number of free merozoites ($MSP1_{19}$ positive, $MSP1_{33}$ positive, no contact with RBC), attached merozoites ($MSP1_{19}$ positive, $MSP1_{33}$ positive, in contact with RBCs), rings ($MSP1_{19}$ positive, $MSP1_{33}$ negative, within RBCs) and RBCs were counted. A minimum of 1000 RBCs was examined.

DNA constructs and PCR . The DNA construct (NBPXa_1XKO) for targeted disruption of *NBPXa* (GeneDB ID PKNH_1472300) via single cross over homologous recombination was adapted from a

precursor to the PkconGFP$_{ep}$ vector (21) and used a 1.2 kb region of the 5' end of the *NBPXa* gene to target the NBPXa locus, with linearization for transfection using the *BsiWI site* situated in the *NBPXa* fragment.

Transfection, cloning and genotyping. Parasites were transfected with NBPXa_1XKO or positive control plasmid PkconGFPp230p (21) and fresh human or cynomolgus RBCs added every 4 days until parasites were visible. Parasites were cloned by limiting dilution, a process requiring 9 or 14 days in either cynomolgus or human RBC respectively, and then expanded for genotypic and phenotypic analysis. Transfected parasites were analysed by diagnostic PCR with two different primer sets. The first was designed to amplify a 1,478 bp fragment only from the intact wild type *NBPXa* locus (ol1 and ol2) and the second was designed to amplify a 1,306 bp fragment after correct integration of NBPXa_1XKO (ol1 and ol3). PCR reactions were carried out using GoTaq Mastermix (Promega) using the following cycle conditions: 3 min at 96 °C, then 30 cycles of 25 s at 96 °C, 25 s at 52 °C, and 2 min at 64 °C, and a final extension of 5 min at 64 °C.

References

1. Lee K-S*, et al.* (2011) *Plasmodium knowlesi*: reservoir hosts and tracking the emergence in humans and macaques. *PLoS pathogens* 7(4):e1002015-e1002015.
2. Singh B*, et al.* (2004) A large focus of naturally acquired *Plasmodium knowlesi* infections in human beings. *The Lancet* 363(9414):1017-1024.
3. William T*, et al.* (2013) Increasing Incidence of *Plasmodium knowlesi* Malaria following Control of *P. falciparum* and *P. vivax* Malaria in Sabah, Malaysia. *PLoS Negl Trop Dis* 7(1):e2026-e2026.
4. Yusof R*, et al.* (2014) High proportion of knowlesi malaria in recent malaria cases in Malaysia. *Malaria Journal* 13(1):168.
5. Cowman AF & Crabb BS (2006) Invasion of Red Blood Cells by Malaria Parasites. *Cell* 124(4):755-766.

6. Miller LH, Mason SJ, Dvorak JA, McGinniss MH, & Rothman IK (1975) Erythrocyte receptors for (*Plasmodium knowlesi*) malaria: Duffy blood group determinants. *Science.* 189(4202):561-563.

7. Ahmed AM*, et al.* (2014) Disease Progression in *Plasmodium knowlesi* Malaria Is Linked to Variation in Invasion Gene Family Members. *PLoS Negl Trop Dis* 8(8):e3086-e3086.

8. Wanaguru M, Liu W, Hahn BH, Rayner JC, & Wright GJ (2013) RH5–Basigin interaction plays a major role in the host tropism of *Plasmodium falciparum*. *Proc Natl Acad Sci USA* 110(51):20735-20740.

9. Tham W-H, Healer J, & Cowman AF (2012) Erythrocyte and reticulocyte binding-like proteins of *Plasmodium falciparum*. *Trends in Parasitology* 28(1):23-30.

10. Galinski MR, Medina CC, Ingravallo P, & Barnwell JW (1992) A reticulocyte-binding protein complex of *Plasmodium vivax* merozoites. *Cell* 69:1213-1226.

11. Okenu DM*, et al.* (2005) The reticulocyte binding proteins of *Plasmodium cynomolgi*: a model system for studies of *P. vivax*. *Mol Biochem Parasitol* 143(1):116-120.

12. Ogun SA*, et al.* (2011) Targeted disruption of py235ebp-1: invasion of erythrocytes by *Plasmodium yoelii* using an alternative Py235 erythrocyte binding protein. *PLoS Pathog* 7(2):e1001288.

13. Rayner JC, Galinski MR, Ingravallo P, & Barnwell JW (2000) Two Plasmodium falciparum genes express merozoite proteins that are related to *Plasmodium vivax* and *Plasmodium yoelii* adhesive proteins involved in host cell selection and invasion. *Proc Natl Acad Sci U S A* 97(17):9648-9653.

14. Triglia T*, et al.* (2001) Identification of proteins from *Plasmodium falciparum* that are homologous to reticulocyte binding proteins in *Plasmodium vivax*. *Infect Immun* 69(2):1084-1092.

15. Meyer EVS*, et al.* (2009) The reticulocyte binding-like proteins of *P. knowlesi* locate to the micronemes of merozoites and define two new members of this invasion ligand family. *Molecular and Biochemical Parasitology* 165(2):111-121.

16. Semenya AA, Tran TM, Meyer EV, Barnwell JW, & Galinski MR (2012) Two functional reticulocyte binding-like (RBL) invasion ligands of zoonotic *Plasmodium knowlesi* exhibit differential adhesion to monkey and human erythrocytes. *Malar J* 11:228-228.

17. Baum J*, et al.* (2009) Reticulocyte-binding protein homologue 5 – An essential adhesin involved in invasion of human erythrocytes by *Plasmodium falciparum*. *International Journal for Parasitology* 39(3):371-380.

18. Adams JH*, et al.* (1990) The duffy receptor family of *Plasmodium knowlesi* is located within the micronemes of invasive malaria merozoites. *Cell* 63(1):141-153.

19. Singh AP*, et al.* (2005) Targeted deletion of *Plasmodium knowlesi* Duffy binding protein confirms its role in junction formation during invasion. *Molecular Microbiology* 55(6):1925-1934.

20. Kocken CHM*, et al.* (2002) *Plasmodium knowlesi* provides a rapid in vitro and in vivo transfection system that enables double-crossover gene knockout studies. *Infection and immunity* 70(2):655-660.

21. Moon RW*, et al.* (2013) Adaptation of the genetically tractable malaria pathogen *Plasmodium knowlesi* to continuous culture in human erythrocytes. *Proc Natl Acad Sci USA* 110(2):531-536.

22. Lim C*, et al.* (2013) Expansion of host cellular niche can drive adaptation of a zoonotic malaria parasite to humans. *Nature communications* 4:1638-1638.

23. Chin W, Contacos PG, Coatney GR, & Kimball HR (1965) A naturally acquired quotidian-type malaria in man transferable to monkeys. *Science (New York, N.Y.)* 149(3686):865-865.

24. Pain A*, et al.* (2008) The genome of the simian and human malaria parasite *Plasmodium knowlesi*. *Nature* 455:799-803.

25. Assefa S*, et al.* (2015) Population genomic structure and adaptation in the zoonotic malaria parasite Plasmodium knowlesi. *Proc Natl Acad Sci U S A* 112(42):13027-13032.

26. Theron M, Hesketh RL, Subramanian S, & Rayner JC (2010) An adaptable two-color flow cytometric assay to quantitate the invasion of erythrocytes by *Plasmodium falciparum* parasites. *Cytometry. Part A* 77(11):1067-1074.

27. Blackman MJ*, et al.* (1996) *Plasmodium knowlesi*: secondary processing of the malaria merozoite surface protein-1. *Exp Parasitol* 83(2):229-239.

28. Jun G*, et al.* (2009) Factors affecting reticulocyte enrichment by density gradient ultracentrifugation. *Acta haematologica* 121(4):207-211.

29. Sorette MP, Shiffer K, & Clark MR (1992) Improved isolation of normal human reticulocytes via exploitation of chloride-dependent potassium transport. *Blood* 80(1):249-254.

30. Nair S*, et al.* (2008) Adaptive Copy Number Evolution in Malaria Parasites. *PLoS Genet* 4(10):e1000243.

31. Satkoski Trask JA*, et al.* (2013) Single Nucleotide Polymorphisms Reveal Patterns of Allele Sharing across the Species Boundary between Rhesus (*Macaca mulatta*) and Cynomolgus (*M. fascicularis*) Macaques. *American journal of primatology* 75(2):135-144.

32. Carlton JM, Das A, & Escalante AA (2013) Genomics, population genetics and evolutionary history of *Plasmodium vivax*. *Adv Parasitol* 81:203-222.

33. Adams JH*, et al.* (1992) A family of erythrocyte binding proteins of malaria parasites. *Proc Natl Acad Sci USA* 89(15):7085-7089.

34. Miller LH, Mason SJ, Clyde DF, & McGinniss MH (1976) The Resistance Factor to *Plasmodium vivax* in Blacks. *New England Journal of Medicine* 295(6):302-304.

35. Carlton JM*, et al.* (2008) Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. *Nature* 455(7214):757-763.

36. Hester J*, et al.* (2013) De Novo Assembly of a Field Isolate Genome Reveals Novel *Plasmodium vivax* Erythrocyte Invasion Genes. *PLoS Negl Trop Dis* 7:e2569.

37. Tachibana S-I*, et al.* (2012) *Plasmodium cynomolgi* genome sequences provide insight into *Plasmodium vivax* and the monkey malaria clade. *Nat Genet* 44(9):1051-1055.

38. Singh S, Alam MM, Pal-Bhowmick I, Brzostowski JA, & Chitnis CE (2010) Distinct External Signals Trigger Sequential Release of Apical Organelles during Erythrocyte Invasion by Malaria Parasites. *PLoS Pathog* 6:e1000746.

39. Ranjan A & Chitnis CE (1999) Mapping regions containing binding residues within functional domains of *Plasmodium vivax* and *Plasmodium knowlesi* erythrocyte-binding proteins. *Proc Natl Acad Sci* 96(24):14067-14072.

40. Pinheiro MM*, USAet al.* (2015) *Plasmodium knowlesi* genome sequences from clinical isolates reveal extensive genomic dimorphism. *PLoS One* 10(4):e0121303.

41. Otto TD, Dillon GP, Degrave WS, & Berriman M (2011) RATT: Rapid Annotation Transfer Tool. *Nucleic acids research* 39(9):e57-e57.

42. Boeva V*, et al.* (2012) Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics* 28(3):423-425.

43. Abyzov A, Urban AE, Snyder M, & Gerstein M (2011) CNVnator: An approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Research* 21(6):974-984.

44. Smith SD, Kawash JK, & Grigoriev A (2014) GROM-RD: Resolving genomic biases to improve read depth detection of copy number variants. *PeerJ PrePrints* 2:e663v661.

Figure Legends

Figure 1: Genomic comparisons of adapted parasite lines. (A) Genomic differences across *Plasmodium knowlesi* lines and derived clones used in this study. Starting material was the A1 stabilate derived from the H strain grown in a rhesus macaque. Schematic shows SNPs (green) and small INDELs (blue) identified between the H strain, the original A1-O line and human RBC-adapted A1-H and cynomolgus RBC-adapted A1-C lines, and respective clones. Venn diagrams show the

relationship among A1-O, A1-H, and A1-C to the genome of a clone from each line. The presence of the known invasion genes in the clones (A1-H.1 to H.3, and A1-C.1 to C.2) is shown. **(B)** Mapping Illumina reads coverage of *P. knowlesi* A1-O (green), A1-C (red) and A1-H (blue) against chromosome 14 of *P. knowlesi* H viewed in Artemis . The highlighted area shows the lack of A1-C read coverage, corresponding to a 66.3 kbp deletion encompassing the invasion-related gene *NBPXa* (PKNH_1472300).

Figure 2: Phenotypic characterisation of the A1-O parasite line. (A) Growth of the A1-O line and the long-term culture adapted A1-C.1 and A1-H.1 clones in cynomolgus or human RBCs after 24 h (a mean of 4 independent experiments +/- 1 SD). Statistical significance determined using a two factor ANOVA with Sidak post hoc test to provide multiplicity adjusted p-values. Significant (*) and non-significant (ns) results are denoted. In cynomolgus RBCs both A1-C.1 (p=0.0040) and A1-H.1 (p=0.0021) lines grew significantly better than A1-O. In human RBCs only A1-H.1 (p=0.0285) showed increased growth rate relative to A1-O; in contrast A1-C.1 demonstrated a non-significant, but consistent, reduction in growth relative to A1-O in human RBCs. Parasitemia was determined after invasion (at 7 h) and after a cycle of intracellular growth (at 24 h) in either **(B)** cynomolgus or **(C)** human RBCs revealing that growth rates were determined by invasion efficiency and not through differences in subsequent intracellular growth (means of triplicate data points from a single experiment +/- 1 SD).

Figure 3: Disruption of *NBPXa* can be achieved when parasites are grown in macaque RBCs, but not in human RBCs. (A) Disruption of *NBPXa* by introduction of linearised pNBPXa_1XKO into the genome via integration at the 5' end of the *NBPXa* locus within a targeting region (hatched box). Primers for diagnostic analysis are indicated (thick black arrows). **(B)** PCR analysis of independently transfected parasites grown in human (T1 toT3) or cynomolgus (T1 toT5) RBCs, and two independent clones derived from the latter (T1.1 and T2.1), confirmed correct integration only for parasites

grown in cynomolgus RBCs. Panels show amplification of a product from wild type *NBPXα* (top panel, primers ol1 and ol2) and a species that can be produced only following integration (lower panel, primers ol1 and ol3).

Figure 4: Disruption of *NBPXα* prevents multiplicative growth in human RBCs through impaired invasion. (A) Growth of A1-C.1, A1-H.1 and Δ*NBPXα* in cynomolgus or human RBCs after 24 h (mean of 4 independent experiments +/- 1 SD). Statistical significance determined using a two factor ANOVA with Sidak post hoc test to provide multiplicity adjusted p-values. Significant (*) and non-significant (ns) results are shown. Compared to A1-H.1, Δ*NBPXα* showed no significant difference in growth in cynomolgus, but a significant reduction in human RBCs (p=0.0481). There was no significant growth difference of Δ*NBPXα* and A1-C.1 in either human or cynomolgus RBCs. Parasite growth and invasion over two cycles in either **(B)** cynomolgus or **(C)** human RBCs revealed a growth defect at or around invasion (means of triplicate data points from a single experiment +/- 1 SD). Light microscopic images of Giemsa-stained thin blood films from **(D)** A1-H.1 and **(E)** Δ*NBPXα* parasites in human RBCs at 7 h: newly formed rings (black arrowheads); attached but uninvaded merozoites (red arrowheads). Scale bar, 5µm.

Figure 5: Δ*NBPXα* parasites attach to human RBC but fail to invade. (A) A1-H.1 and ΔNBPXa schizonts were purified and released merozoites were allowed to invade human or cynomolgus RBCs for 3 h. Panels show blood smears stained with anti-$MSP1_{33}$ (green), anti-$MSP1_{19}$ (red) and a merge of both MSP1 antibodies with anti-human RBC (magenta) and DAPI (blue). MSP1 is processed during invasion with $MSP1_{33}$ being shed and $MSP1_{19}$ retained, distinguishing external attached merozoites ($MSP1_{33}$ and $MSP1_{19}$ positive) from those that had successfully invaded to form rings ($MSP1_{19}$ only positive). Whilst A1-H.1 merozoites (top panel) were able to invade human RBC and form rings (white arrows), most ΔNBPXa merozoites (bottom panel) attached but did not invade human RBC (white arrowheads). Scale bar, 10µm. **(B)** Parasite stages present in the immunofluorescence assays

were quantified and graph shows the number of free merozoites, RBC-attached merozoites and rings stages expressed as a percentage of total merozoite-derived forms counted (means of triplicate cultures from a single experiment +/- 1 SD). The number in brackets indicates the average ring stage parasitemia for each of the parasite/blood combinations.