Bulk Segregant Approaches to Nutritional Genomics in *Plasmodium falciparum*

Short title: BSA and Malaria Parasite Nutritional Genomics

Sudhir Kumar[1†], Xue Li[2†], Marina McDew-White[2], Ann Reyes[2], Abeer Sayeed[2], Meseret T. Haile[1], Spencer Y. Kennedy[1], Nelly Camargo[1], Lisa A. Checkley[3], Katelyn M. Vendrely[3], Katrina A. Button-Simons[3], Ian H. Cheeseman[4], Stefan H. Kappe[1,5], François Nosten,[6,7], Michael T. Ferdig[3], Ashley M. Vaughan[1,5*] and Tim J. C. Anderson[2*]

[1] Center for Global Infectious Disease Research, Seattle Children's Research Institute, Seattle, Washington, USA.

[2] Program in Disease Intervention and Prevention, Texas Biomedical Research Institute, San Antonio, Texas, USA.

[3] Eck Institute for Global Health, Department of Biological Sciences, University of Notre Dame, Notre Dame, Indiana, USA.

[4] Program in Host Pathogen Interactions, Texas Biomedical Research Institute, San Antonio, Texas, USA.

[5] Department of Global Health, University of Washington, Seattle, Washington, USA.

[6] Shoklo Malaria Research Unit, Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Mae Sot, Thailand.

[7] Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine Research building, University of Oxford Old Road campus, Oxford, UK.

[†] Sudhir Kumar and Xue Li contributed equally to this work.

[*]Co-corresponding authors: ashley.vaughan@seattlechildrens.org (AMV) and tanderso@TxBiomed.org (TJCA).

Abstract

Nutrient acquisition/metabolism pathways provide potent targets for drug design. We conducted crosses between African (NF54) and Asian (NHP4026) malaria parasites, and compared genome-wide allele frequency changes in independent progeny populations grown in human serum or AlbuMAX, a commercial bovine serum formulation. We detected three QTLs linked with differential growth that contained strong candidate genes: aspartate transaminase *AST (chromosome 2), cysteine protease ATG4 (chr. 13)* and *EBA-140 (chr. 14)*. Alleles inherited from NF54 (chr. 2 and 14) and from NHP4026 (chr. 13) were positively selected in AlbuMAX, while the same alleles were selected against in serum. Selection driving differential growth was strong (s = 0.10 – 0.23 per 48-hour lifecycle) and observed in all biological replicates. These results demonstrate the effectiveness of bulk segregant approaches for revealing nutritional polymorphisms in *Plasmodium falciparum*. This approach will allow systematic dissection of nutrient acquisition/metabolism pathways that are potential targets for intervention against *P. falciparum*.

Introduction

Nutrient acquisition and metabolism pathways are promising targets for antimalarial development. Among existing drugs, artemisinin - the frontline drug against malaria - is activated by hemoglobin digestion (*1*), while chloroquine interferes with haem polymerization into non-toxic haemozoin (*2*), and antifolate drugs (pyrimethamine and sulfadoxine) are competitive inhibitors that interrupt the folate biosynthesis pathway (*3*). The mutations conferring resistance to these drugs are also involved with parasite nutrition transport/metabolism pathways. For example, resistance to artemisinin (ART), is mediated by mutations in *ketch13*, which is required for hemoglobin endocytosis (*1*). Mutations in the chloroquine resistance transporter (*pfCRT*), which normally functions as a transport channel for ions and peptides (*2*) mediate resistance to a variety of drugs including chloroquine (CQ) and piperaquine. Mutations in dihydrofolate reductase and dihydropteroate synthase, components of the folate synthesis pathway (*3*) confer resistance to pyrimethamine and sulfadoxine. Given the importance of nutrient acquisition/metabolism, effective methods for locating genes and pathways involved in these processes are urgently needed.

Exploiting natural variation in parasite nutrient acquisition and metabolism pathways provides one promising approach. Nguitragool *et al.* used a *P. falciparum* genetic crosses conducted in chimpanzee hosts to identify an important channel (plasmodial surface anion channel, PSAC) involved in ion transport (*4*). Similarly, Wang *et al.* have used a comparable linkage mapping approach to investigate the ability of parasites to utilize exogenous folate (*5*) as this is important for determining the success of drugs that target the parasite folate synthesis pathway. These key discoveries demonstrate how differences in metabolism or nutrient acquisition between parasites can be effectively exploited to better understand the genetic underpinning of metabolic pathways and transport systems, which in turn can highlight potential targets for intervention. However, traditional linkage mapping is laborious and expensive because individual parasite progeny must be cloned, phenotyped and sequenced.

Continuous *in vitro* culture of asexual erythrocytic stages of the malaria parasite *P. falciparum* requires human erythrocytes, buffered RPMI 1640 medium and human serum, with a low oxygen atmosphere at 37°C (*6*). RPMI 1640 medium is the main resource for sugar (glucose), salts, essential amino acids and multiple vitamins (*7*). Human hemoglobin can supply amino acids other

than isoleucine (*8*), while human serum provides all the other nutrients needed for parasite growth, such as inorganic and organic cations. Lipid-enriched bovine albumin (AlbuMAX) is the most widely used human serum substitute. Human serum typically contains more phospholipid and cholesterol, and less fatty acid than AlbuMAX (*9*). AlbuMAX has several advantages over human serum for culture of *P. falciparum*, due to its low cost, compatibility with any blood type, and lower batch-to-batch variation. AlbuMAX supplemented culture medium has made significant contributions to malaria research, facilitating *in vitro* drug sensitivity assays for screening and monitoring of antimalarial drugs, parasite growth competition assays to measure fitness costs, and research on parasite molecular biology and immunology. However, several studies have found that parasite growth rate differs between AlbuMAX- and human serum-based cultures and can impact drug susceptibility. For example, AlbuMAX supported parasite growth less well than human serum for clinical isolates from Cameroonian patients (*10*), and for long-term lab culture adapted parasites (*11*). Furthermore, the 50% inhibitory concentrations (IC_{50}s) of multiple antimalarial drugs obtained with AlbuMAX, including CQ, amodiaquine, quinine and artemisinin, were almost twice the corresponding values obtained with human non-immune serum (*12*).

Our central aim was to evaluate the efficacy of genetic crosses and a rapid linkage mapping method - bulk segregant analysis (BSA) - for understanding the genetic basis of nutrition-related phenotypes in *P. falciparum* using differential growth and fitness in serum- or AlbuMAX-based *in vitro* culture as a test system. Carrying out genetic crosses in *Anopheles* mosquitoes and human hepatocyte-liver chimeric mice (FRG huHep mice) (*13*) now allows us to routinely generate large pools of recombinant *P. falciparum* progeny (**Fig. 1A**) without the need for chimpanzee hosts. The FRG huHep mouse supports liver stage development of *P. falciparum* and the transition to asexual blood stage. We have previously generated four independent *P. falciparum* genetic crosses with large numbers of unique recombinant progeny using this approach (*13-15*).

BSA provides a simple and fast approach to identify loci that contribute to complex traits (*16*) that typically identified using traditional QTL mapping. Using pooled sequencing of progeny populations, BSA measures changes in allele frequency following application of different selection pressures (here asexual blood stage growth in serum or AlbuMAX, **Fig. 1B**). BSA, also referred to as linkage group selection (LGS), has been extensively used for genetic crosses of rodent malaria parasites to map genes determining blood stage multiplication rate, virulence and immunity in

Plasmodium yoelii, mutations conferring ART resistance and strain-specific immunity in *Plasmodium chabaudi* (reviewed in (*17*)).

In this study, we generated replicated genetic crosses between a long-term lab adapted parasite (NF54, Africa) and a newly cloned clinical parasite (NHP4026, Southeast Asia). We measured changes of allele frequency in independent progeny populations during parallel asexual blood stage growth with serum and AlbuMAX. We detected three repeatable quantitative trait loci (QTLs) regions, on chromosomes 2, 13 and 14, that were associated with parasite growth rates in the different culture media.

Results

Genetic crosses and segregant pools

We generated the crosses using *Anopheles stephensi* mosquitoes and FRG huHep mice as described in Vaughan *et al* (*13*) and Li et al (*15*). *P. falciparum* parasites NF54 and NHP4026 were cloned by limiting dilution and used as parents for this cross. NF54 is a parasite of Africa origin that has been maintained in the lab for decades, in a variety of conditions that not limited to different serum/AlbuMAX, human erythrocytes and laboratories. NHP4026 was isolated from a patient visiting the Shoklo Malaria Research Unit (SMRU) clinic on the Thailand-Myanmar border, 2007. NHP4026 has only been cultured in the lab with AlbuMAX for a limited period of time (in total of 80 days). There are total of 13,195 single-nucleotide polymorphisms (SNPs, approximately 1 SNP per 1.6 kb, **Table S1**) between the two parental parasites within the 21 Mb core genome (see supplementary Materials and Methods for details).

To generate parallel recombinant pools, we mixed gametocyte from both parents at ~1:1 ratio to infect ~450 mosquitoes (three separate cages). Recombinants are generated after gametes fuse to form zygotes in the mosquito midgut (**Fig. 1A**). Replication of the four meiotic products ultimately leads to the generation of thousands of haploid sporozoites within each oocyst. For this cross, the oocyst prevalence was 93% (range: 86-100%), with an average burden of 14 oocysts per mosquito midgut (range: 2-61), giving an estimate of 56 (14×4) recombinant genotypes per infected mosquito. We generated three independent recombinant pools from this experiment. To ensure that recombinants from each pool are independent, we injected pooled sporozoites into each

of three FRG huHep mice from a different cage of mosquitoes (~100). This gave us ~5600 (56 × ~100) unique recombinants per pool. The allele frequencies of NF54 in recombinant pools of parasites emerging from the liver were 0.518±0.001, 0.503±0.002 and 0.549±0.002 for the three mice. Whole genome sequencing (WGS) data of cloned progeny from a mixture of all pools revealed low numbers of selfed progeny (2/55) in this cross with very little redundancy among recombinants (**Table S2**) (see also (*14*)).

We cultured each recombinant pool with O-positive non-immune serum and AlbuMAX for 34 days in parallel, with two technical replicates in each media type for each of the three biological replicates. We collected samples for BSA every four days and used WGS to analyze all segregant pools to high (191 ± 40 read depth) genome coverage (**Fig. 1B**). To pinpoint the loci that determine parasite fitness in different culture media, we plotted the allele frequencies and calculated G' value (see supplementary Materials and Methods for details) to measure the significance of allelic skews throughout the culture process (**Fig. 2**), and to determine the strength of selection acting on different genome regions in the two different culture conditions we calculated the selection coefficient (*s*, slope of the linear model between the natural log of the allele ratio [freq (NF54)/freq (NHP4026)] against time, **Fig. S1**) across the whole genome.

QTLs for differential growth in Serum and AlbuMAX

By comparing allele frequency changes in serum and AlbuMAX culture over 34 days, we detected three QTLs situated at the beginning of chr. 2, the end of chr. 13 and the first half of chr. 14 (**Fig. 3, Fig. S1**). For each detected QTL (G' > 20, **Fig. 3**), we calculated 95% confidence intervals to narrow down the size of the genome regions and thus the list of genes that could be driving selection (**Fig. 4**). The list of genes inside the QTL regions is summarized in **Table S3**. We prioritized genes within these genome regions by the following criteria: i) if the gene is expressed in blood stages (**Table S3**); ii) gene annotations and related metabolic pathways; iii) we also inspected the SNPs and indels that differentiated the two parents, but not as a determinant factor as gene function might be altered through epigenetic regulations (**Tables S3 and S4**). Those deemed most probable candidate genes driving these QTLs are listed in **Table S5**.

For the chr. 2 and chr. 13 QTL, the NF54 allele frequency increased over time in AlbuMAX and decreased in serum. The patterns of divergent selection observed were consistent across all

biological and technical replicates (**Fig. S1, Table S6**). We observed $s = 0.04 \pm 0.01$ in serum and $s = -0.06 \pm 0.01$ in AlbuMAX for the chr. 2 QTL region; $s = -0.08 \pm 0.02$ in serum and $s = 0.15 \pm 0.01$ in AlbuMAX for chr. 13. The two QTL regions (chr. 2 and chr. 13 QTLs) are located at the beginning of a chromosome. It is difficult to define QTL boundaries to a narrow confidence interval due to limited recombination and difficulties of sequence alignment in these regions. For the chr. 2 QTL, we inspected genes located from the beginning of chr. 2 (which has a length of 947 kb) to 220 kb (**Fig. 4**). This region contained 53 genes, 14 were not expressed in blood stage parasites and were excluded, and 13 of the remaining 39 are high priority candidates. Among these, the aspartate transaminase gene (*AST*, PF3D7_0204500, also known as aspartate aminotransferase, *AspAT*), is a critical enzyme for amino acid metabolism. There were no non-synonymous mutations in the coding region between *AST* alleles from the two parents, but we found multiple differences within the 5' UTR and gene expression regulatory regions, which include two SNPs (coding region (c.)-18C>T and c.-29A>C) and three microsatellites (**Table S4**). The chr. 13 QTL contained 33 genes (23 expressed in blood stage, 8 high priority candidates) and spanned 163 kb at the beginning of chr. 13. Among the genes in this QTL, we identified the erythrocyte binding antigen-140 (*EBA-140*, PF3D7_1301600) inside the first 100 kb of chr. 13. EBA-140 mediates the *P. falciparum* erythrocyte invasion by binding to the red blood cell receptor glycophorin C and initiating merozoite entry (*18*). Interestingly, there was one non-synonymous mutation (Leu112Phe) between NF54 and NHP4026 at the *EBA-140* gene locus. This SNP is located before the first Duffy-binding-like (DBL) domain (*19*) and is common in malaria populations (**Fig. S2**). The gene expression levels of *EBA-140* are also variable in parasites from different clinical isolates (*20*). The *EBA-140* allele from NHP4026 was preferentially selected for in AlbuMAX (**Fig. 4**).

In the chr. 14.1 QTL region, the NF54 allele frequency did not change over time during growth in serum, but increased significantly during growth in AlbuMAX (**Fig. 2)**. The selection coefficient $s = 0.02 \pm 0.02$ in serum and $s = -0.10 \pm 0.03$ in AlbuMAX for chr. 14 (**Fig. 5**). This QTL located in the first half of chr. 14 (630 kb – 813 kb, numbered 14.1 in **Table S5**, spanned 183 kb and contained 38 genes. Of these genes, 37 are expressed in blood stage, and 11 are high priority candidates (**Table S5**). Interestingly, NHP4026 carried a single amino acid deletion (Asn226del) and three non-synonymous mutations (Ser329Pro, Asn503Lys and Val556Ile) in the cysteine protease *autophagy-related 4* gene (*ATG4*, PF3D7_1417300).

Systematic Skews observed in both Serum and AlbuMAX

We observed three genome regions that showed strong distortions in allele frequency in each independent replicate cross in both human serum and AlbuMAX cultures (**Fig. 2A**), consistent with our earlier report (*14, 15*). We used G' values to measure the significance of these allelic skews (**Fig. 3**) and identified three QTLs with G' > 20: on chr. 7, in the middle of chr. 12 and on the second half of chr. 14. We observed strong selection against NHP4026 alleles on both the chr. 7 and chr. 14 QTL regions (s = 0.29±0.04 in serum and 0.26±0.04 in AlbuMAX for chr. 7, s = 0.12±0.02 in both serum and AlbuMAX for chr. 14), while NF54 alleles were selected against in the chr. 12 QTL region, with s = 0.30±0.05 in serum and 0.25±0.06 in AlbuMAX (**Fig. 5, Fig. S1**). The skews were consistent among all three biological replicates.

The chr. 7 QTL spanned from 340,864 kb to 476,223 kb (135kb) and contained 33 genes (**Table S3**). The *pfCRT* (PF3D7_0709000), which is known to carry high fitness cost with CQ resistant alleles (*21*), is located at the peak of the chr. 7 QTL (**Fig. 4**). Here, NHP4026 carries the CQ resistant *pfCRT* allele while NF54 is CQ sensitive (**Table S5**). In both serum and AlbuMAX cultures, the resistant CQ allele carries extremely high fitness cost (s = 0.29 ±0.04 in serum and s = 0.26±0.04 in AlbuMAX, **Fig. 5, Table S6**), and is evident on day 4 of in vitro culture **Fig. 2 and 5)**.

The QTL on chr. 12 spanned from 1,141 kb to 1,283 kb (142 kb, 32 genes, all expressed in blood stage, and 9 high priority candidates) and the QTL on chr. 14 (numbered 14.2 in **Table S5**) spanned from 2,356 kb to 2,485 kb (129 kb, 23 genes, 10 high priority candidates). These same regions carried high fitness costs in a BSA analysis from an independent genetic cross between two different parental parasites – ART-S MKK2835 and ART-R NHP1337 (*15*). The multidrug resistance-associated protein 2 (*MRP2*, PF3D7_1229100) and apicoplast ribosomal protein S10 (*ARPS10*, PF3D7_1460900), are both located inside of the chr. 12 and chr. 14 QTL regions respectively. There are total of four non-synonymous mutations and six indels within the *MRP2* locus, and all four parental parasites carry different MRP2 alleles. Those from NF54 and NHP1337 carried high fitness costs. There were two non-synonymous mutations (Val127Met and Asp128His) in the ARPS10 locus (**Table S5**) on chr. 14. *ARPS10* alleles with these two mutations (NHP4026 in this study and NHP1337 in previous study by Li et al. (*15*)) carry a high fitness cost during *in*

vitro culture with serum. The Val127Met mutation is one of the genetic background SNPs for *kelch13* alleles on which artemisinin resistance emerged in Cambodia (*22*).

Discussion

Locus specific selection between serum and AlbuMAX cultures

We found three QTL regions where allele frequencies of recombinant progeny parasites showed dramatic divergence, depending on whether they were grown in serum or in AlbuMAX (**Fig. 4**). The QTLs for differential selection in serum and AlbuMAX were observed in each of the recombinant pools and across technical replicates for each pool. Furthermore, selection driving change in allele frequency in these three QTL regions is strong ($s = 0.10 \pm 0.01$ [chr. 2], $s = 0.23 \pm 0.02$ [chr. 13] and $s = 0.12 \pm 0.02$ [chr. 14], **Fig. 5, Table S6**). AlbuMAX is a lipid-loaded bovine serum albumin (BSA), while the composition of serum is more complex, containing a variety of proteins and peptides (albumins, globulins, lipoproteins, enzymes and hormones), nutrients (carbohydrates, lipids and amino acids), electrolytes, and small organic molecules. Serum also contains more phospholipid and cholesterol and less fatty acid than AlbuMAX (*9*). It has been previously reported that AlbuMAX didn't support malaria parasite growth as well as serum (*10-12*). Our analysis suggests there are multiple loci across the genome that determine parasite growth rate and in serum and AlbuMAX, suggesting that these QTL regions contain genes involved in nutrient uptake or metabolism. For example, in AlbuMAX, the chr. 2 and 14 QTL regions skew toward NF54 alleles, while the chr. 13 QTL skews toward the NHP4026 allele. However, in serum culture the opposite patterns are seen. By inspection of the genes under these QTL peaks, we identified three genes – *AST* (chr. 2), EBA-140 (chr. 13) and cysteine protease ATG4 (chr. 14.1) – as the strongest candidates (full gene listed in **Table S3**).

Chr. 2: *P. falciparum* acquires nutrients from the host through catabolism of hemoglobin in RBCs. During this process, plasmodial AST plays a critical role in the classical tricarboxylic acid (TCA) cycle, and also functions to maintain homeostasis of carbohydrate metabolic pathways (reviewed in (*23*)). In our study, the AST allele from NHP4026 shows lower fitness than the AST allele from NF54 during competitive AlbuMAX culture, and the trend is reversed during serum

culture. Furthermore, AST is essential (*24*), which highlights this gene as a potential bottleneck in energy metabolism and as a target for the design of novel therapeutic strategies. We speculate that selection acts on the regulation of *AST* gene expression levels: we detected no non-synonymous mutations in the coding region, but we found multiple variants in the 5' UTR and regulatory regions (**Table S4**). However, we cannot exclude that other neighboring loci may drive the observed allele frequency changes.

Chr. 13: During the asexual erythrocytic stages, malaria parasites rapidly grow and replicate every 48 hr resulting in the release of RBC-infectious merozoites. Erythrocyte binding-like (EBL) ligands which bind to red blood cell receptors, have been identified as major determinants of erythrocyte invasion by merozoites (reviewed in (*25*)). To date, four EBL ligands have been characterized in *P. falciparum*: EBA-175, EBA-181, EBL-1 and EBA-140. In our study, the end of chr. 13, which contains EBA-140 shows strong skews depending on whether parasites are cultured in serum or AlbuMAX (**Fig. 2, Fig. 5**). Interestingly, *P. falciparum* infected RBCs show decreased cytoadherence to multiple cell surface receptors when maintained in AlbuMAX rather than serum (*9*) and multiple studies have showed that mutations in *EBA-140* influence parasite binding to RBC surface receptors (*18, 19*). In our study, only one amino acid change (Leu112Phe) distinguished the *EBA-140* from NF54 and NHP4026. We hypothesize that this mutation influences merozoite binding to the RBC surface.

Chr. 14: Cysteine protease ATG4 is essential for autophagy in both yeast and mammals, as well as in parasites such as *Leishmania major* and *Trypanosoma cruzi* (reviewed in (*26*)). A number of cysteine proteases have been described in malaria parasites with functions including hemoglobin hydrolysis, erythrocyte rupture, and erythrocyte invasion (*27*) and two cysteine proteases (SERA 3 and SERA4) show expression increases >2-fold in AlbuMAX supplemented media (*28*). Autophagy involves vesicular trafficking and is important for protein and organelle degradation during cellular differentiation (*29*). Conditional knock down of *ATG4* in the apicomplexan *Toxoplasma* leads to severe growth defects (*30*). Autophagy can also be triggered by starvation (*29*). For example, in response to isoleucine amino acid starvation, *Plasmodium* parasites enter a quiescent state, and autophagy of parasite proteins could provide a source of isoleucine in such conditions (*31*). In our study, the NF54 *ATG4* allele had higher fitness than NHP4026 during AlbuMAX cultures, while the allele frequencies didn't change in serum (**Fig. 5**).

Compared to serum, AlbuMAX contains fewer components (*19*). We speculate that parasites may require higher levels of autophagy to compensate for specific nutrients that are missing in AlbuMAX. Since the *ATG4* allele from NHP4026 differs from NF54 in having one amino acid deletion and three missense variants we hypothesize that NHP4026 *ATG4* is less able to maintain high levels of autophagy required for efficient growth in AlbuMAX.

Replication of fitness-related QTLs in independent genetic crosses

We detected two additional regions of the genome (chr. 12 and chr. 14.2) that show extreme skews in allele frequencies in both serum and AlbuMAX cultures. We have previously observed skews in these two regions during *in vitro* culture of progeny from an independent genetic cross between parasites MKK2835 and NHP1337 (*15*). Like NHP4026, MKK2835 and NHP1337 are newly cloned parasites isolated from patients on the Thailand-Myanmar border (**Table S7**), while NF54 has been long-term lab adapted. These observations suggest that these skews result from adaptations to laboratory culture.

We observed extremely strong selection against NF54 at the chr. 12 QTL, with selection coefficients (s) of 0.25 per 48 hr asexual cycle in serum and (s) of 0.30 per 48 hr asexual cycle in AlbuMAX (**Fig. 5**). The strong and repeatable selection observed in independent crosses against the NF54 allele indicates that these skews are not artificial effects of the *in vitro* culture system but are determined by the parasite genetic background. In both our previous genetic cross using different parental parasites (*15*) and in the current study, multidrug resistance-associated protein 2 (MRP2) was located at the peak of the chr. 12 QTL. *MRP2* belongs to the C-family of ATP binding cassette (ABC) transport proteins that are well known for their role in multidrug resistance. MRP2 mediates the export of drugs, toxins, and endogenous and xenobiotic organic anions (*32*), which can thus lead to resistance to multiple drugs. *MRP2* may also contribute to the detoxification of antimalarial drugs. Further experiments are needed to directly determine the function of *MRP2* in parasite fitness during *in vitro* culture.

Mutations in the *Kelch13* gene underlie artemisinin resistance and mutations in the *arps10* gene provide a permissive genetic background for emergence of artemisinin resistance, but are not thought to directly contribute to drug resistance (*22*). We observed selection against the mutant *arps10* allele at the chr. 14 QTL locus (**Table S7**, NHP4026 and NHP1337) in both this study and

our previous cross (*15*): we measured (*s*) of 0.12 per 48 hr asexual generation using both serum and AlbuMAX for the cross between NF54 and NHP4026 (this study); and (*s*) of 0.18 per 48 hr asexual cycle for the cross between MKK2835 and NHP1337 (*15*). Among the four parental parasites, only NHP1337 carries the mutant *kelch13* allele (C580Y). Further studies are required to determine whether the mutant *arps10* allele might compensate for fitness costs associated with mutant *Kelch13* through epistatic interactions.

Fitness costs and compensation at the *PfCRT* locus on chr. 7

We observed strong selection against the *PfCRT* allele (chr. 7 QTL) conferring CQ resistance (CQR) in both serum and AlbuMAX cultures by comparing the initial allele frequencies and those after 30 days of culture. This same skew is also is also observed in clones isolated from our previous crosses between the same parental parasites (*14*). The rapid change in allele frequency of *PfCRT* alleles equates to selection coefficients (*s*) of 0.29 per asexual life cycle in serum and 0.26 in AlbuMAX (Fig. 5A). These fitness cost estimates are strikingly close to those calculated from laboratory competition experiments that measured fitness costs of *Dd2*-type *PfCRT* allele to be ~0.3 per asexual life cycle relative to a wild type *PfCRT* alleles (reviewed in (*21*)). The long-term decline in frequency of parasites with CQR *PfCRT* observed in Malawi, Kenya and China following the withdrawal of CQ use also reveals high fitness costs of CQR *PfCRT* alleles, with selection coefficients of 0.12) (Malawi), 0.05 (Kenya), and 0.01 (China) per generation (*21*). These estimates assuming three parasite generations (from mosquito to mosquito) per year).

The prevalence of *PfCRT* CQR alleles is currently > 90% in SE Asia (*21*) and has been maintained at this level for over 20 years after CQ was abandoned as first-line treatment for *P. falciparum* infections. The reasons for this could include (i) reduced opportunity to compete with wild type parasites due to the low rate of polyclonal infections, (ii) selection resulting from treatment of co-infecting *P. vivax* cases with CQ (*33*), or (iv) the presence of compensatory loci in the NHP4026 genetic background of CQ resistant parasites. Our data suggest the existence of loci that may compensate for costly CQR *PfCRT* alleles. Interestingly, NHP4026, the parent carrying the CQR *PfCRT* allele, shows high competitive fitness in the laboratory, ranking above other SE Asia clinical isolates (*34*), and even outcompetes the long-term lab adapted NF54 bearing the CQ sensitive (CQS) *PfCRT* allele. We speculate that the chr. 12 QTL region (**Fig. 5**, see discussion

below), which brings high growth advantage to NHP4026, may compensate for the deleterious effects of the CQR *PfCRT* to restore parasite fitness.

Pros and cons of bulk segregant approaches for genetic analysis

Our study uses a BSA strategy to systemically identify genes involved in competitive nutrient uptake and metabolism. The highly repeatable results stem from use of independent recombinant pools and demonstrate the power of the BSA approach. In contrast, conducting these analyses in a traditional framework, using isolation of individual progeny, measurement of growth phenotypes of each progeny in different media, and genomic characterization of progeny and parents, as required for previous nutritional genetics experiments with *P. falciparum* (*4, 35, 36*) is laborious, time consuming and expensive. A particular advantage of BSA is that growth phenotypes of all progeny are determined in a single culture, removing batch effects and experimental variation resulting from conducting parallel measures with individual progeny clones.

BSA is well suited to examine the impact of multiple different nutritional components in parallel. There are multiple different nutrient components that can be explored using this approach. For example, we can remove, or reduce concentrations of one nutrient (e.g. glucose, salts, lipid, amino acids and vitamins) at a time (*37-39*), and compare progeny populations grown in normal versus depleted medium, to understand the genetic basis for acquisition and metabolism of specific nutrients. In this particular experiment, we compared media containing different lipid sources. However, because human serum and AlbuMAX differ in multiple components, more precisely constructed media that differ in single lipid components will be needed to identify key nutrients involved in driving the differential selection. We also note that other aspects of parasite culture are amenable to genetic dissection using BSA. For example, we can examine selection at different temperatures, which range from normal to those mimicking fever within infections (*40*) or we can examine impacts of pH, which may vary in infected patients due to acidosis (*41*).

However, BSA approaches allow QTL location, but not functional validation of candidate loci. Furthermore, BSA approaches cannot effectively detect epistatic interactions between loci. Fortunately, *P. falciparum* can be cloned by dilution, allowing isolation of single progeny. We can use cloned progeny, with different conformations of alleles from the parents in the QTL regions, or with CRISPR/Cas9 modifications of candidate loci, to confirm the influences of different alleles

on specific parasite metabolic pathways or to examine interactions between loci. In this study, the chemical differences between serum and AlbuMAX are complex, so it will be difficult to identify the precise media components driving differential selection. However, we anticipate that verification of candidate loci determining nutrient metabolism will be possible in future studies, in which single components of media are varied.

Implications for malaria parasite biology and control

Malaria parasites isolated from different regions of the world differ in ease with which they can be grown in culture using serum and AlbuMAX. For example, several studies suggest that African parasites grow better in human serum than in AlbuMAX (*10*), while Thai parasites grow better in AlbuMAX. We examined the frequencies of the alleles at chr. 2, 13 and 14 QTLs in worldwide genomic database for *P. falciparum* (MalariaGEN, https://www.malariagen.net/, **Fig. S2**). For two (chr. 2: *AST*, chr. 14: *ATG4*) of the three candidate genes, we found large differences in allele frequency in parasites from Africa and Asia. *AST* (chr. 2) showed minimal variation in Asia with only one common SNP at frequencies ranging from 0.05 – 0.19, while in Africa there were five common SNPs with frequencies ranging from 0.14 – 0.40. Similarly, for *ATG* (chr. 14), one SNP is at a high frequency (0.80 – 0.86) in African locations, and at a low frequency in Asian populations. Allele frequencies of these loci in global collections are consistent with phenotypic difference in ability to culture parasites in different culture media, but do not demonstrate any causative association. Parasites do not encounter AlbuMAX in human infections, but it is possible that regional difference in human diet and nutrition select for geographical allele frequency differences observed in the chr. 2, 13 and 14 QTLs. Indeed, there is evidence that indicates that the nutrient status of host will influence parasite replication and virulence (*42*).

Use of selective markers is widely used for efficient recovery of transfectants. Drug resistance markers are commonly used for both negative and positive selection of malaria parasites carrying plasmids (*43*). However, the selective growth medium technique which is widely used in bacterial and yeast genetics, has not yet been effectively utilized for parasitic (*44*) studies. We envisage that the understanding of key genes underpinning parasite metabolic pathways will allow the development of novel selection systems for parasite genetic engineering.

Essential genes underlying nutrient acquisition or metabolism provide excellent loci for drug development. Pioneering work on minimal media in *P. falciparum* has demonstrated that isoleucine alone is required to support *Plasmodium* growth in culture (*31*), and subsequent work validated the isoleucine metabolism pathway as a possible target for intervention (*45, 46*). In this study we show that BSA provides an additional powerful tool for dissecting the genetic basis of other nutrient related pathways in *P. falciparum* utilizing natural genetic variation segregating in crosses.

Material and Methods

Ethics approval and consent to participate

The study was performed in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health (NIH), USA. To this end, the Seattle Children's Research Institute (SCRI) has an Assurance from the Public Health Service (PHS) through the Office of Laboratory Animal Welfare (OLAW) for work approved by its Institutional Animal Care and Use Committee (IACUC). All of the work carried out in this study was specifically reviewed and approved by the SCRI IACUC.

Culture media with serum and AlbuMAX

Contents and manufacturers of the culture media used in this study for *P. falciparum* are as listed in **Table S8**. In summary, we used RPMI 1640 as basal medium. We added 2 mM L-glutamine as amino acid supplement, 25mM HEPES for maintaining culture pH and 50 µM hypoxanthine as a nutrient additive which helps in cell growth. To prevent the growth of fungi and bacteria, we used 50 IU/ml penicillin, 50 µg/ml streptomycin and 50 µg/ml vancomycin for the first two asexual life cycles (4 days). The basal medium was supplemented either with 10% O+ human serum or with 0.5 % AlbuMAX II. O+ erythrocytes were added every two days into the culture media to support amplification of the parasite population. We maintained all the cultures at 37ºC with 5% O_2, 5% CO_2, and 90% N_2, with 2% hematocrit. Only one batch of reagents were used through the whole experiment.

Preparation of genetic cross

We generated the cross using FRG NOD huHep mice with human chimeric livers and *A. stephensi* mosquitoes as described in Vaughan *et al* (*13*) and Li et al (*15*) (**Fig. 1A**). We used NF54 (lab adapted Africa parasite) and NHP4026 (newly cloned clinical isolate from the Thai-Myanmar border) in this study. Gametocytes from both the parasite strains were diluted to 0.5% gametocytemia using human serum erythrocyte mix, to generate infectious blood meals (IBMs). IBMs from each parent was mixed at equal ratio and fed to ~450 mosquitos (3 cages of 150 mosquitoes, of which ~45 were mosquitoes were sacrificed for prevalence test).

We examined the mosquito infection rate and oocyst number per infected mosquito 7 days post-feeding. Fifteen mosquitoes were randomly picked from each cage and dissected under microscopy. Sporozoites are isolated from infected mosquito salivary glands and 2-4 Million sporozoites from each cage of mosquitoes were injected into three FRG huHep mice (one cage per mouse), intravenously. To allow the liver stage-to-blood stage transition, mice are infused with human erythrocytes six and seven days after sporozoite injection. Four hours after the second infusion, the mice are euthanized and exsanguinated to isolate the circulating ring stage *P. falciparum*-infected human erythrocytes. The parasites from each mouse constitute the initial recombinant pools for further segregation experiment. All initial recombinant pools were maintained using AlbuMAX culture for 24hr to stabilize the newly transitioned ring stage parasites. We prepared three recombinant pools in this study.

Sample collection and sequencing

We aliquoted each initial recombinant pool into four cultures and maintained two of them with serum medium and the other two with AlbuMAX medium (**Fig. 1B**). There were total of 12 cultures (3 mice as biological replicates × [serum + AlbuMAX II] × 2 technical replicates). We maintained all the cultures with standard six-well plate for 45 days. Freshly packed huRBCs were added every 2 days to each replicate, meanwhile parasites cultures were diluted to 1% parasitemia to avoid stressing. 70ul packed red blood cells (RBCs) were collected and frozen down every 2-4 days.

We extracted and purified genomic DNA using the Qiagen DNA mini kit, and quantified amounts using Qubit. We constructed next generation sequencing libraries using 50-100 ng DNA or sWGA product following the KAPA HyperPlus Kit protocol with 3-cycle of PCR. All libraries were sequenced at 150bp pair-end to a minimum coverage of 100× using Illumina Novaseq S4 or Hiseq X sequencers.

Bulk segregant analysis

We performed the Bulk segregant analysis as described in (*15*). In short, we first mapped and genotyped the whole-genome sequencing reads both from parental parasites and progeny pools against the NF54 genome. To filter out the low-quality genotypes, recalibrated variant quality scores (VQSR) were calculated by comparing the raw variant distribution with the known and verified *Plasmodium* variant dataset, and loci with VQSR less than 1 were removed from further analysis. Only loci with coverage > 30× and distinct in two parents were used for further analysis. Allele frequencies of NF54 were plotted across the genome. G' values are calculated to detect extreme-QTLs. Once a QTL was detected (G' > 20), we calculated and approximate 95% confidence interval to localize causative genes. We also calculated selection coefficient (*s*) to measure the fitness cost at each mutation. See **Supplementary Materials and Methods** for details in Bulk segregant analysis.

References

1. J. Birnbaum, S. Scharf, S. Schmidt, E. Jonscher, W. A. M. Hoeijmakers, S. Flemming, C. G. Toenhake, M. Schmitt, R. Sabitzki, B. J. S. Bergmann, A Kelch13-defined endocytosis pathway mediates artemisinin resistance in malaria parasites. **367**, 51-59 (2020).
2. S. H. Shafik, S. A. Cobbold, K. Barkat, S. N. Richards, N. S. Lancaster, M. Llinás, S. J. Hogg, R. L. Summers, M. J. McConville, R. E. J. N. C. Martin, The natural function of the malaria parasite's chloroquine resistance transporter. **11**, 1-16 (2020).
3. A. Gregson, C. V. J. P. r. Plowe, Mechanisms of resistance of malaria parasites to antifolates. **57**, 117-145 (2005).
4. W. Nguitragool, A. A. Bokhari, A. D. Pillai, K. Rayavara, P. Sharma, B. Turpin, L. Aravind, S. A. Desai, Malaria parasite clag3 genes determine channel-mediated nutrient uptake by infected red blood cells. *Cell* **145**, 665-677 (2011).
5. P. Wang, R. K. Brobey, T. Horii, P. F. Sims, J. E. Hyde, Utilization of exogenous folate in the human malaria parasite Plasmodium falciparum and its critical role in antifolate drug synergy. *Molecular microbiology* **32**, 1254-1262 (1999).
6. W. Trager, J. B. Jensen, Human malaria parasites in continuous culture. *Science* **193**, 673-675 (1976).
7. G. E. Moore, R. E. Gerner, H. A. J. J. Franklin, Culture of normal human leukocytes. **199**, 519-524 (1967).
8. I. J. B. o. t. W. H. O. Sherman, Transport of amino acids and nucleic acid precursors in malarial parasites. **55**, 211 (1977).
9. S. Frankland, S. R. Elliott, F. Yosaatmadja, J. G. Beeson, S. J. Rogerson, A. Adisa, L. Tilley, Serum lipoproteins promote efficient presentation of the malaria virulence protein PfEMP1 at the erythrocyte surface. *Eukaryotic cell* **6**, 1584-1594 (2007).
10. L. K. Basco, Molecular epidemiology of malaria in cameroon. XX. Experimental studies on various factors of in vitro drug sensitivity assays using fresh isolates of Plasmodium falciparum. *The American Journal of Tropical Medicine and Hygiene* **70**, 474-480 (2004).
11. C. Dohutia, P. K. Mohapatra, D. R. Bhattacharyya, K. Gogoi, K. Bora, B. K. Goswami, In vitro adaptability of Plasmodium falciparum to different fresh serum alternatives. *Journal of Parasitic Diseases* **41**, 371-374 (2017).

12. P. Ringwald, F. S. Meche, J. Bickii, L. K. Basco, In Vitro Culture and Drug Sensitivity Assay ofPlasmodium falciparum with Nonserum Substitute and Acute-Phase Sera. *Journal of clinical microbiology* **37**, 700-705 (1999).
13. A. M. Vaughan, R. S. Pinapati, I. H. Cheeseman, N. Camargo, M. Fishbaugher, L. A. Checkley, S. Nair, C. A. Hutyra, F. H. Nosten, T. J. Anderson, Plasmodium falciparum genetic crosses in a humanized mouse model. *Nature methods* **12**, 631 (2015).
14. K. A. Button-Simons, S. Kumar, N. Carmago, M. T. Haile, C. Jett, L. A. Checkley, S. Y. Kennedy, R. S. Pinapati, D. A. Shou, M. McDew-White, X. Li, F. H. Nosten, S. H. Kappe, T. J. C. Anderson, J. Romero-Severson, M. T. Ferdig, S. J. Emrich, A. M. Vaughan, I. H. Cheeseman, Surprising variation in the outcome of two malaria genetic crosses using humanized mice: implications for genetic mapping and malaria biology. *bioRxiv*, 2019.2012.2013.871830 (2019).
15. X. Li, S. Kumar, M. McDew-White, M. Haile, I. H. Cheeseman, S. Emrich, K. Button-Simons, F. Nosten, S. H. I. Kappe, M. T. Ferdig, T. J. C. Anderson, A. M. Vaughan, Genetic mapping of fitness determinants across the malaria parasite Plasmodium falciparum life cycle. *PLoS Genet* **15**, e1008453 (2019).
16. I. M. Ehrenreich, N. Torabi, Y. Jia, J. Kent, S. Martis, J. A. Shapiro, D. Gresham, A. A. Caudy, L. Kruglyak, Dissection of genetically complex traits with extremely large pools of yeast segregants. *Nature* **464**, 1039 (2010).
17. K. M. Vendrely, S. Kumar, X. Li, A. M. Vaughan, Humanized Mice and the Rebirth of Malaria Genetic Crosses. *Trends in Parasitology*, (2020).
18. J. H. Adams, B. Sim, S. A. Dolan, X. Fang, D. C. Kaslow, L. H. Miller, A family of erythrocyte binding proteins of malaria parasites. *Proceedings of the National Academy of Sciences* **89**, 7085-7089 (1992).
19. D. H. Lin, B. M. Malpede, J. D. Batchelor, N. H. Tolia, Crystal and solution structures of Plasmodium falciparum erythrocyte-binding antigen 140 reveal determinants of receptor specificity during erythrocyte invasion. *Journal of Biological Chemistry* **287**, 36830-36836 (2012).
20. N. Gomez-Escobar, A. Amambua-Ngwa, M. Walther, J. Okebe, A. Ebonyi, D. J. J. T. J. o. i. d. Conway, Erythrocyte invasion and merozoite ligand gene expression in severe and mild Plasmodium falciparum malaria. **201**, 444-452 (2010).
21. A. Ecker, A. M. Lehane, J. Clain, D. A. Fidock, PfCRT and its role in antimalarial drug resistance. *Trends in parasitology* **28**, 504-514 (2012).
22. O. Miotto, R. Amato, E. A. Ashley, B. MacInnis, J. Almagro-Garcia, C. Amaratunga, P. Lim, D. Mead, S. O. Oyola, M. Dhorda, Genetic architecture of artemisinin-resistant Plasmodium falciparum. *Nature genetics* **47**, 226 (2015).
23. C. Wrenger, I. B Muller, A. M Silber, R. Jordanova, V. S Lamzin, M. R Groves, Aspartate Aminotransferase-Bridging Carbohydrate and Energy Metabolism in Plasmodium Falciparum. *Current drug metabolism* **13**, 332-336 (2012).
24. L.-M. Ting, W. Shi, A. Lewandowicz, V. Singh, A. Mwakingwe, M. R. Birck, E. A. T. Ringia, G. Bench, D. C. Madrid, P. C. Tyler, Targeting a novel Plasmodium falciparum purine recycling pathway with specific immucillins. *Journal of Biological Chemistry* **280**, 9547-9554 (2005).
25. E. Jaskiewicz, M. Jodłowska, R. Kaczmarek, A. Zerka, Erythrocyte glycophorins as receptors for Plasmodium merozoites. *Parasites vectors* **12**, 317 (2019).
26. J. L. Siqueira-Neto, A. Debnath, L.-I. McCall, J. A. Bernatchez, M. Ndao, S. L. Reed, P. J. J. P. n. t. d. Rosenthal, Cysteine proteases in protozoan parasites. **12**, e0006512 (2018).
27. P. J. Rosenthal, Cysteine proteases of malaria parasites. *International journal for parasitology* **34**, 1489-1499 (2004).
28. K. Singh, A. Agarwal, S. I. Khan, L. A. Walker, B. L. Tekwani, Growth, drug susceptibility, and gene expression profiling of Plasmodium falciparum cultured in medium supplemented with human serum. *Journal of biomolecular screening* **12**, 1109-1114 (2007).
29. B. Levine, D. J. Klionsky, Development by self-digestion: molecular mechanisms and biological functions of autophagy. *Developmental cell* **6**, 463-477 (2004).
30. M. A. Kong-Hap, A. Mouammine, W. Daher, L. Berry, M. Lebrun, J.-F. Dubremetz, S. Besteiro, Regulation of ATG8 membrane association by ATG4 in the parasitic protist Toxoplasma gondii. *Autophagy* **9**, 1334-1348 (2013).

31. S. E. Babbitt, L. Altenhofen, S. A. Cobbold, E. S. Istvan, C. Fennell, C. Doerig, M. Llinás, D. E. Goldberg, Plasmodium falciparum responds to amino acid starvation by entering into a hibernatory state. *Proceedings of the National Academy of Sciences* **109**, E3278-E3287 (2012).
32. A. T. Nies, D. Keppler, The apical conjugate efflux pump ABCC2 (MRP2). *Pflügers Archiv-European Journal of Physiology* **453**, 643-659 (2007).
33. L. S. Ross, S. K. Dhingra, S. Mok, T. Yeo, K. J. Wicht, K. Kümpornsin, S. Takala-Harrison, B. Witkowski, R. M. Fairhurst, F. Ariey, Emerging Southeast Asian PfCRT mutations confer Plasmodium falciparum resistance to the first-line antimalarial piperaquine. *Nature communications* **9**, 1-13 (2018).
34. A. R. Tirrell, K. M. Vendrely, L. A. Checkley, S. Z. Davis, M. McDew-White, I. H. Cheeseman, A. M. Vaughan, F. H. Nosten, T. J. Anderson, M. T. Ferdig, Pairwise growth competitions identify relative fitness relationships among artemisinin resistant Plasmodium falciparum field isolates. *Malaria journal* **18**, 295 (2019).
35. A. Gupta, A. A. Bokhari, A. D. Pillai, A. K. Crater, J. Gezelle, G. Saggu, A. S. Nasamu, S. M. Ganesan, J. C. Niles, S. A. Desai, Complex nutrient channel phenotypes despite Mendelian inheritance in a Plasmodium falciparum genetic cross. *PLoS pathogens* **16**, e1008363 (2020).
36. A. Gupta, P. Balabaskaran-Nina, W. Nguitragool, G. S. Saggu, M. A. Schureck, S. A. Desai, CLAG3 self-associates in malaria parasites and quantitatively determines nutrient uptake channels at the host membrane. *MBio* **9**, e02293-02217 (2018).
37. S. A. Desai, Insights gained from P. falciparum cultivation in modified media. *The Scientific World Journal* **2013**, (2013).
38. A. A. Divo, T. G. Geary, N. L. Davis, J. B. Jensen, Nutritional requirements of Plasmodium falciparum in culture. I. Exogenously supplied dialyzable components necessary for continuous growth. *The Journal of protozoology* **32**, 59-64 (1985).
39. T. G. Geary, A. A. Divo, L. C. Bonanni, J. B. Jensen, Nutritional Requirements of Plasmodium falciparum in Culture. III. Further Observations on Essential Nutrients and Antimetabolites 1: METABOLISM AND ANTIMETABOLITES IN P. FALCIPARUM. *The Journal of protozoology* **32**, 608-613 (1985).
40. M. Gravenor, D. Kwiatkowski, An analysis of the temperature effects of fever on the intra-host population dynamics of Plasmodium falciparum. *Parasitology* **117**, 97-105 (1998).
41. K. Plewes, H. W. Kingston, A. Ghose, R. J. Maude, M. T. Herdman, S. J. Leopold, H. Ishioka, M. M. U. Hasan, M. S. Haider, S. Alam, Cell-free hemoglobin mediated oxidative stress is associated with acute kidney injury and renal replacement therapy in severe falciparum malaria: an observational study. *BMC infectious diseases* **17**, 313 (2017).
42. L. Mancio-Silva, K. Slavic, M. T. G. Ruivo, A. R. Grosso, K. K. Modrzynska, I. M. Vera, J. Sales-Dias, A. R. Gomes, C. R. MacPherson, P. Crozet, Nutrient sensing modulates malaria parasite virulence. *Nature* **547**, 213-216 (2017).
43. T. Anderson, S. Nkhoma, A. Ecker, D. Fidock, How can we identify parasite genes that underlie antimalarial drug resistance? *Pharmacogenomics* **12**, 59-85 (2011).
44. V. Zuzarte-Luís, M. M. Mota, Parasite sensing of host nutrients and environmental cues. *Cell Host Microbe* **23**, 749-758 (2018).
45. E. S. Istvan, N. V. Dharia, S. E. Bopp, I. Gluzman, E. A. Winzeler, D. E. Goldberg, Validation of isoleucine utilization targets in Plasmodium falciparum. *Proceedings of the National Academy of Sciences* **108**, 1627-1632 (2011).
46. J. Liu, E. S. Istvan, I. Y. Gluzman, J. Gross, D. E. Goldberg, Plasmodium falciparum ensures its amino acid supply with multiple acquisition pathways and redundant proteolytic enzyme systems. *Proceedings of the National Academy of Sciences* **103**, 8840-8845 (2006).
47. A. Miles, Z. Iqbal, P. Vauterin, R. Pearson, S. Campino, M. Theron, K. Gould, D. Mead, E. Drury, J. O'Brien, Indels, structural variation, and recombination drive genomic diversity in Plasmodium falciparum. *Genome research* **26**, 1288-1299 (2016).
48 M. McDew-White, X. Li, S. C. Nkhoma, S. Nair, I. Cheeseman, T. J. Anderson, Mode and tempo of microsatellite length change in a malaria parasite mutation accumulation experiment. *Genome Biology and Evolution* **11**, 1971-1985 (2019).
49 L. Davies, U. Gather, The identification of multiple outliers. *Journal of the American Statistical Association* **88**, 782-792 (1993).

50 B. N. Mansfeld, R. Grumet, QTLseqr: An R package for bulk segregant analysis with next-generation sequencing. *The Plant Genome* **11**, 180006 (2018).
51. P. M. Magwene, J. H. Willis, J. K. Kelly, The statistics of bulk segregant analysis using next generation sequencing. *PLoS computational biology* **7**, e1002255 (2011).
52. H. Li, A quick method to calculate QTL confidence interval. *Journal of genetics* **90**, 355-360 (2011).
53. D. Dykhuizen, D. L. Hartl, Selective neutrality of 6PGD allozymes in E. coli and the effects of genetic background. *Genetics* **96**, 801-817 (1980).
54. E. A. Nadaraya, On estimating regression. *Theory of Probability & Its Applications* **9**, 141-142 (1964).
55. G. S. Watson, Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, 359-372 (1964).

Acknowledgments

General: SMRU is part of the Mahidol Oxford University Research Unit supported by the Wellcome Trust of Great Britain.

Funding: This work was supported by National Institutes of Health (NIH) program project grant P01 AI127338 (to MF), by NIH grant R37 AI048071 (to TJCA) and NIH grant R21 AI133369 (to AMV). Work at Texas Biomedical Research Institute was conducted in facilities constructed with support from Research Facilities Improvement Program grant C06 RR013556 from the National Center for Research Resources.

Author contributions: S.K., X.L. and T.J.C.A. designed the experiments. S.K., M.T.H, S.Y.K and N.C. prepared the crosses and collected samples. M.M.W, A.R. and A.S. prepared the genomic DNA libraries. F.N. provided the parental parasite from Thailand. L.A.C cloned the parental parasites. X.L. performed all the NGS analysis and data curation. S.K. and X.L. and T.J.C.A. wrote the original manuscript. K.M.V, K.A.B, I.H.C, S.H.K, F.N., M.T.F and A.M.V reviewed and edited the manuscript.

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

Data and materials availability: All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. All raw sequencing data have been submitted to the NABI Sequence Read Archive (SRA, https://www.ncbi.nlm.nih.gov/sra) under the project number of PRJNA524855. Additional data related to this paper may be requested from the authors.

Figures

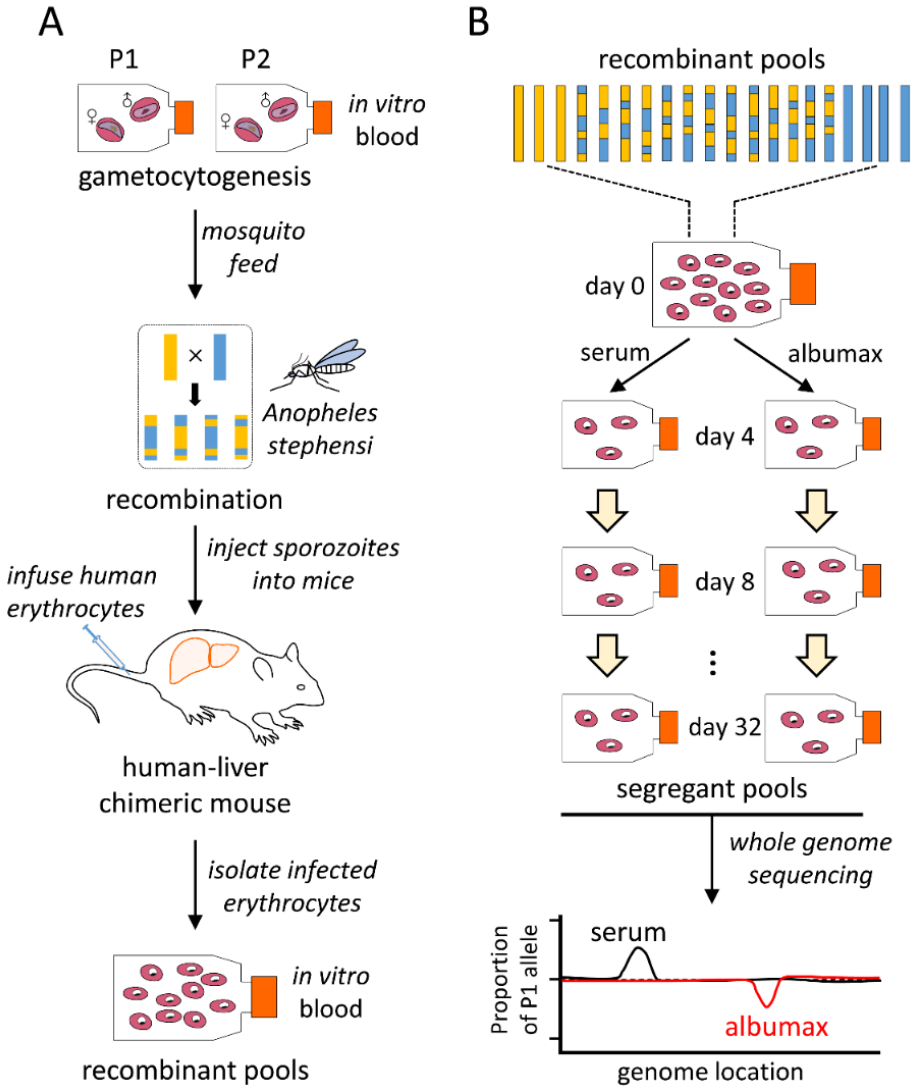

Fig. 1. Mapping parasite fitness under different culture conditions. **(A),** Recombinant progeny pool generation. Genetic crosses are generated using female *Anopheles stephensi* mosquitoes and FRG huHep mice as described by Vaughan et al (*13*). Recombination of parasite genomes occurs during meiosis in the mosquito midgut. Recombinant pools are collected from infected mice and maintained through *in vitro* blood cultures. **(B),** Bulk segregant analysis. Pools of progeny are cultured in parallel with serum or AlbuMAX and samples are collected every 4 days. Each segregant pool is then whole-genome sequenced and genotyped in bulk. Differences in allele frequency among different groups are used to identify QTLs.

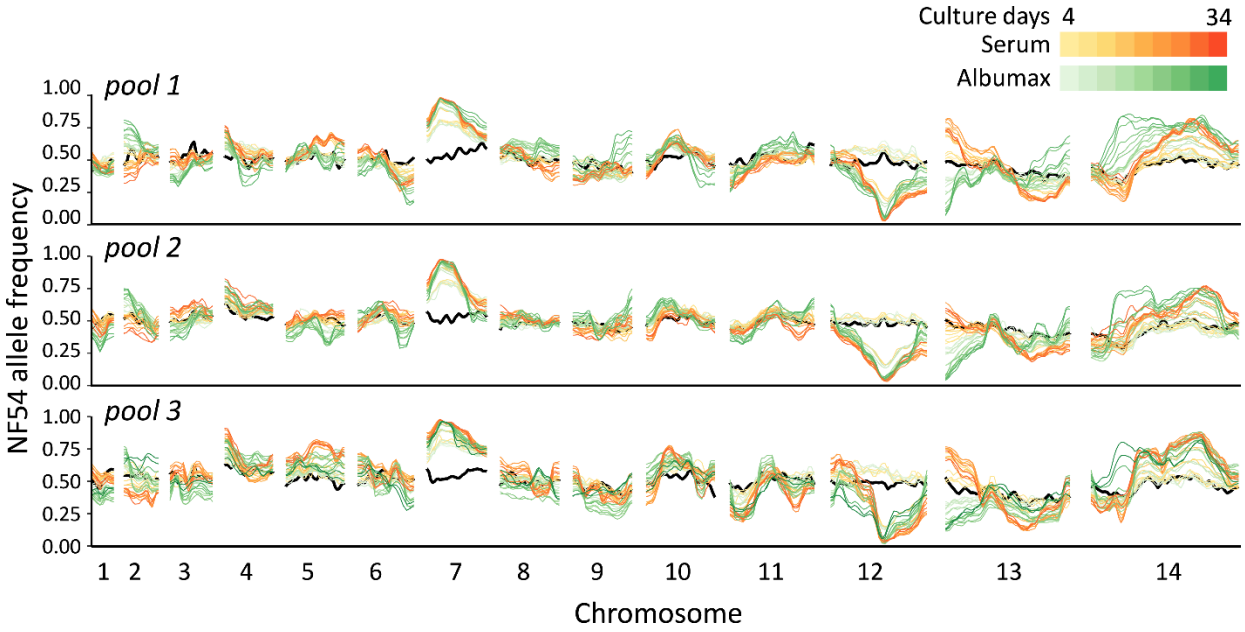

Fig. 2. Change in frequency across the genome in different culture conditions. The black lines show allele frequencies from the initial recombinant pools, while red and green lines indicate allele frequency changes during serum or AlbuMAX cultures. The three plots show allele frequencies from different recombinant pools (biological replicates). The lines with same color in each panel show values for the two technical replicates.

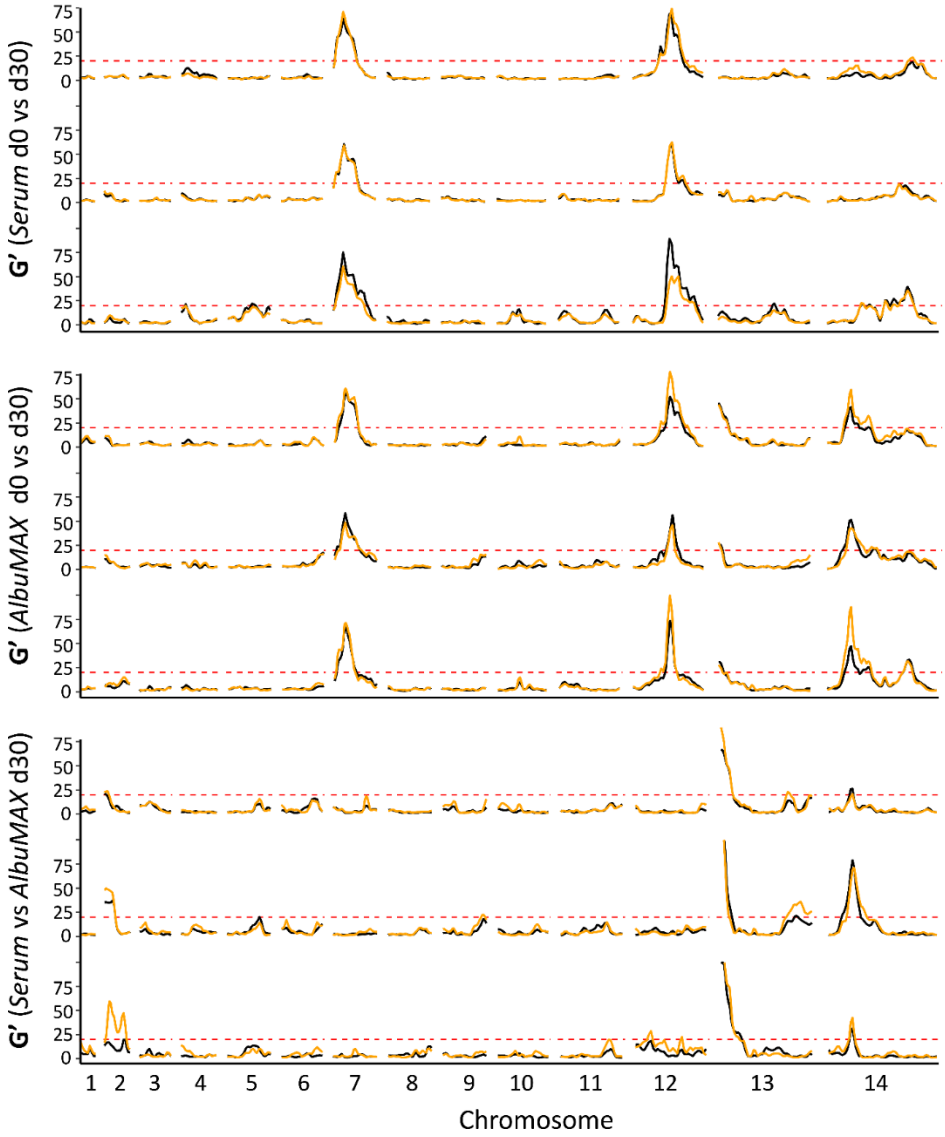

Fig. 3. QTLs defined with G' approach. The top and middle panels show QTLs detected through comparing allele frequencies from the initial recombinant pools and pools after 30 days of serum (top) or AlbuMAX (middle) culture. The bottom panel indicates QTLs at day 30 between serum and AlbuMAX cultures. There are three plots in each panel, which are from different recombinant pools (biological replicates). Orange and black lines are technical replicates in each experiment. We used a threshold (G' > 20) to determine significant QTLs.

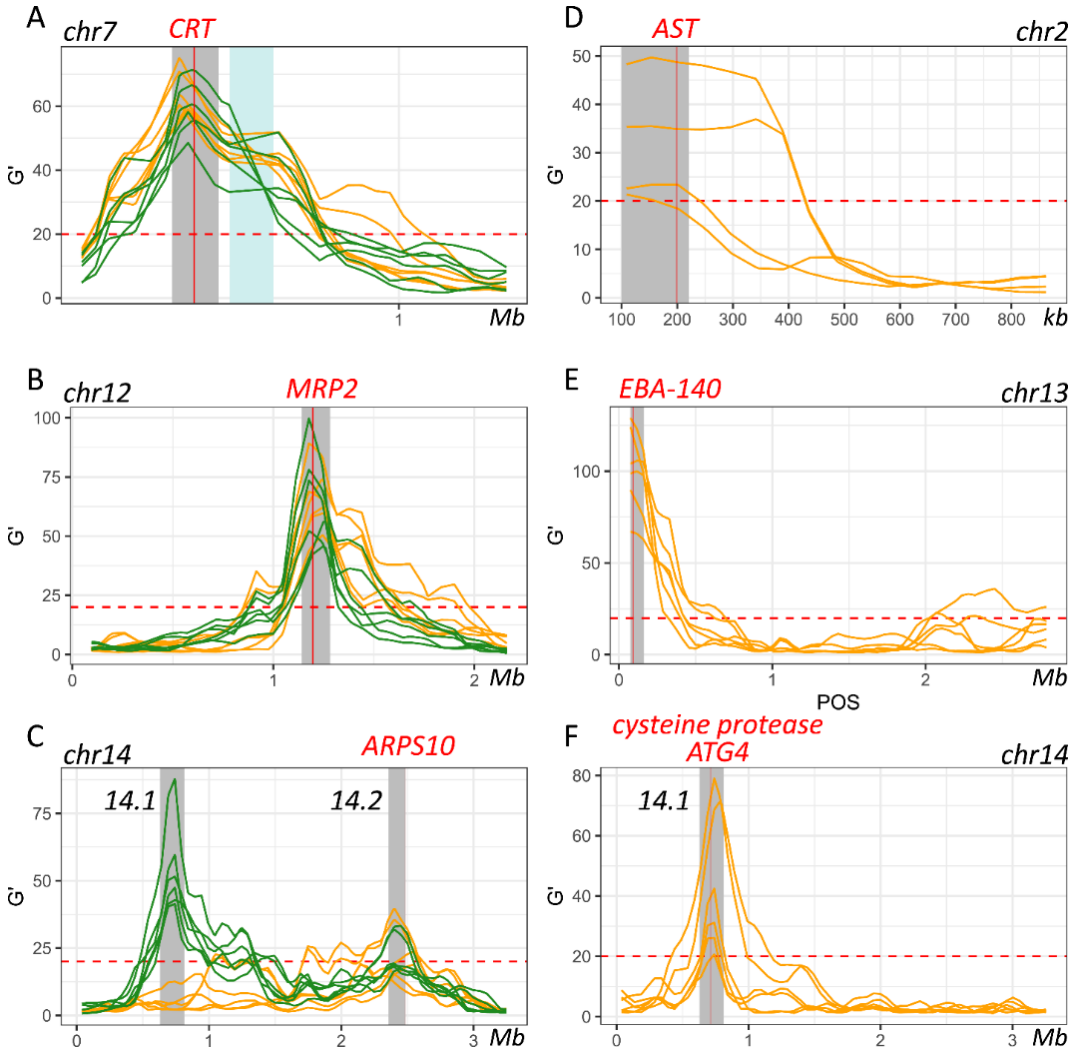

Fig. 4. Genes inside QTL regions. (**A**, **B** & **C**), QTLs detected through comparing allele frequencies from the initial recombinant pools and pools after 30 days of serum (orange) or AlbuMAX (green) culture. (**D**, **E** & **F**), QTLs by comparing serum and AlbuMAX after 30 days of culture. Each line is one comparison. Grey shadows indicate boundaries of the merged 95% confidential intervals (CIs) of all the QTLs. The light cyan region from chr. 7 is a high variable genome region in which SNPs could not be called.

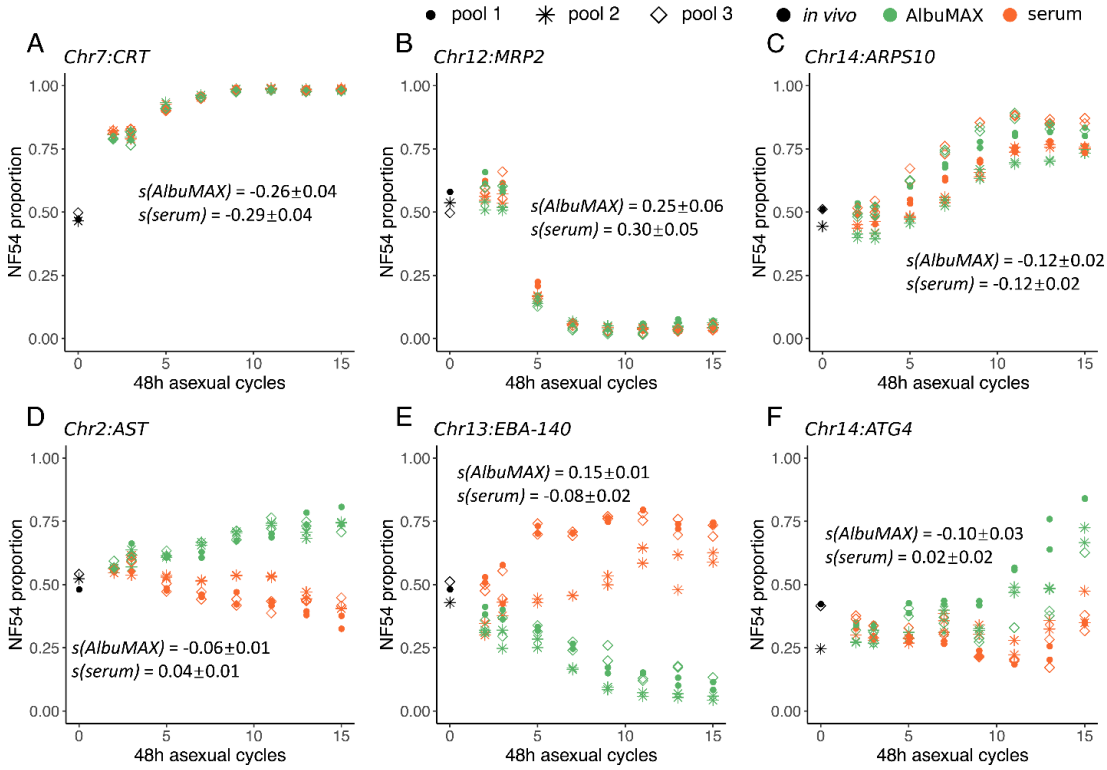

Fig. 5. Estimation of selection coefficients from the changes in allele frequencies in candidate gene regions. **(A-F)**, selection coefficients for gene regions of *CRT*, *MRP2*, *ARPS10*, *AST*, *EBA-10* and *ATG*, separately.

Supplemental Materials

Supplementary Methods and Materials

Fig. S1. Selection coefficients (*s*) across the genome. Estimation of *s* was based on the changes of allele frequency from day1 to day30 of cultures. Positive values of *s* indicate a disadvantage for alleles inherited from NHP4026. Red and green lines indicate cultures by serum and AlbuMAX.

Fig. S2. NF54 allele frequency at candidate gene regions in world-wide malaria parasite populations. WAF: west Africa, EAF: east Africa, CAF: central Africa, SAM: south America, ESEA: east Southeast (SE) Asia, SAS: south Asia, WSEA: west SE Asia, OCE: Pacific Ocean.

Table S1. SNPs between NF54 and NHP4026 parental parasites.

Table S2. Summary of cloned progeny from cross between NF54 and NHP4026.

Table S3. Genes inside the QTL regions.

Table S4. SNPs and indels inside of the QTL regions.

Table S5. Top candidate genes located inside of the QTL regions. A full list of genes found within QTL regions is shown in Table S3.

Table S6. Summary of selection coefficients (*s*) at candidate gene regions.

Table S7. Genotypes of parental parasites.

Table S8. Summary of media components used in this study.

677 **Supplemental Methods and Materials**

678

679 **Genotype calling**

680 We genotyped the parental strains and bulk populations as described in (*15*). We first generated
681 a "mock" genome according to the genotype of parent NF54. We mapped the whole-genome
682 sequencing reads both from parental parasites and progeny against this mock genome using BWA
683 mem (http://bio-bwa.sourceforge.net/) under the default parameters. We excluded the high variable
684 genome regions (subtelomeric repeats, hypervariable regions and centromeres) and only
685 performed genotype calling in the 21 Mb core genome (defined in (*47*)). The resulting alignments
686 were then converted to SAM format, sorted to BAM format, and deduplicated using picard tools
687 v2.0.1 (http://broadinstitute.github.io/picard/). We used Genome Analysis Toolkit GATK v3.7
688 (https://software.broadinstitute.org/gatk/) to recalibrate the base quality score based on a set of
689 verified known variants (*47*). We called variants using HaplotypeCaller and then merged using
690 GenotypeGVCFs with default parameters except for sample ploidy 1.

691 We only applied filters to the GATK genotypes of parental parasites, using standard filter
692 methods described by McDew-White et al (*48*). The recalibrated variant quality scores (VQSR)
693 were calculated by comparing the raw variant distribution with the known and verified
694 *Plasmodium* variant dataset, and loci with VQSR less than 1 were removed from further analysis.
695 After filtration, we selected SNP loci that are distinct in two parents, and only used those for further
696 bulk segregant analysis.

697

698 **Bulk segregant analysis**

699 Only loci with coverage > 30× were used for bulk segregant analysis. We counted reads with
700 genotypes of each parent and calculated allele frequencies at each variable locus. Allele
701 frequencies of NF54 were plotted across the genome, and outliers were removed following
702 Hampel's rule (*49*) with a window size of 100 loci. We performed the BSA analyses using the R
703 package QTLseqr (*50*). Extreme-QTLs were defined as regions with G' > 20 (*51*). Once a QTL
704 was detected, we calculated and approximate 95% confidence interval using Li's method (*52*) to
705 localize causative genes. We also measured the fitness cost at each mutation by fitting a linear
706 model between the natural log of the allele ratio (freq[allele1]/freq[allele2]) against time
707 (measured in 48hr parasite asexual cycles). The slope provides a measure of the selection

708 coefficient (*s*) driving each mutation (*53*). The raw *s* values were tricube-smoothed with a window
709 size of 100 kb to remove noise (*54, 55*).
710
711

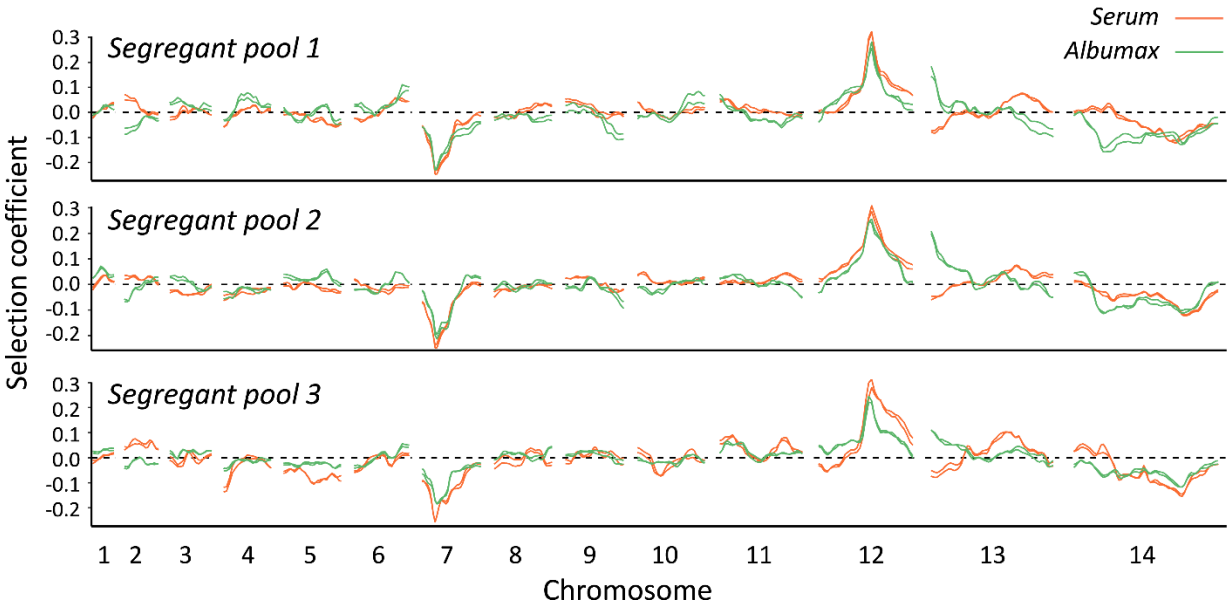

Fig. S1. Selection coefficients (*s*) across the genome. Estimation of *s* was based on the changes of allele frequency from day1 to day30 of cultures. Positive values of *s* indicate a disadvantage for alleles inherited from NHP4026. Red and green lines indicate cultures by serum and AlbuMAX.

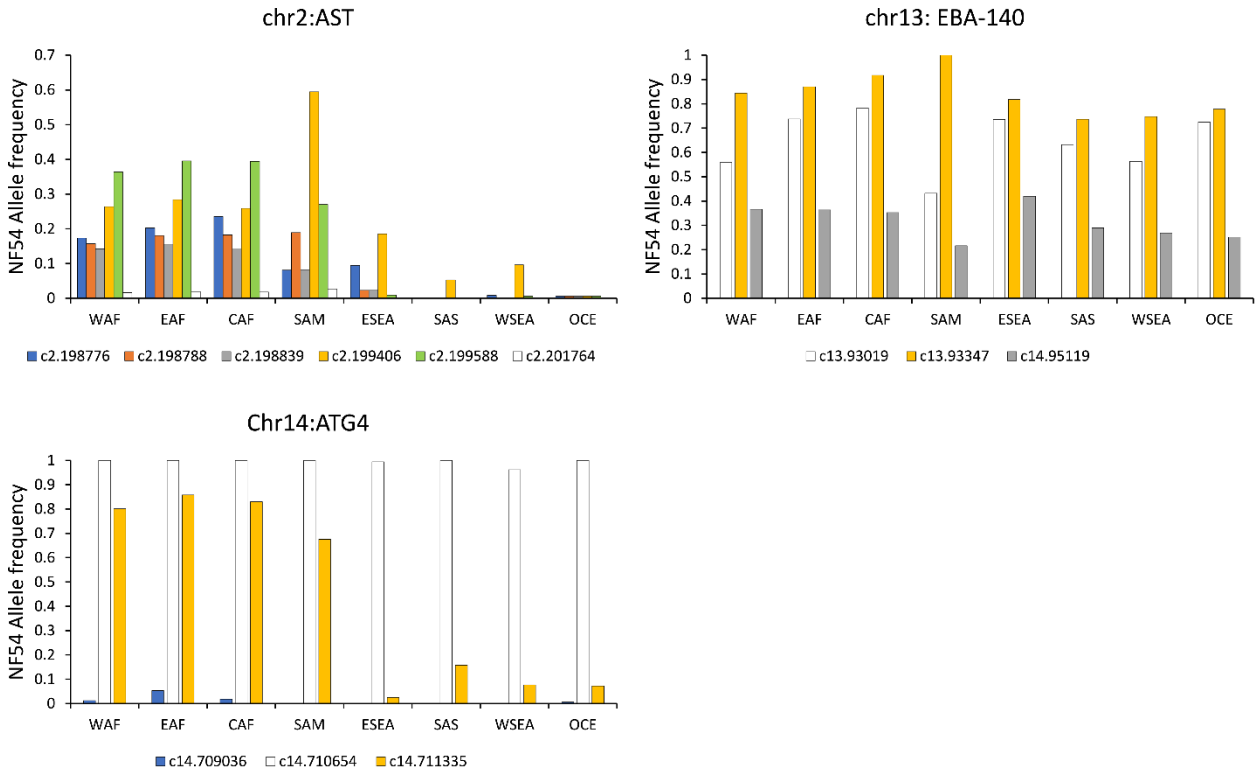

Fig. S2. NF54 allele frequency at candidate gene regions in world-wide malaria parasite populations. WAF: west Africa, EAF: east Africa, CAF: central Africa, SAM: south America, ESEA: east Southeast (SE) Asia, SAS: south Asia, WSEA: west SE Asia, OCE: Pacific Ocean.