RTS,S/AS01 malaria vaccine mismatch observed among *Plasmodium falciparum* isolates from southern and central Africa and globally

Julia C. Pringle1, Giovanna Carpi1, Jacob Almagro-Garcia2,3,4, Sha Joe Zhu2, Tamaki Kobayashi5, Modest Mulenga6, Thierry Bobanga7, Mike Chaponda6, William J. Moss1,5 & Douglas E. Norris1

The RTS,S/AS01 malaria vaccine encompasses the central repeats and C-terminal of *Plasmodium falciparum* circumsporozoite protein (PfCSP). Although no Phase II clinical trial studies observed evidence of strain-specific immunity, recent studies show a decrease in vaccine efficacy against non-vaccine strain parasites. In light of goals to reduce malaria morbidity, anticipating the effectiveness of RTS,S/AS01 is critical to planning widespread vaccine introduction. We deep sequenced C-terminal Pfcs
d from 77 individuals living along the international border in Luapula Province, Zambia and Haut-Katanga Province, the Democratic Republic of the Congo (DRC) and compared translated amino acid haplotypes to the 3D7 vaccine strain. Only 5.2% of the 193 PfCSP sequences from the Zambia-DRC border region matched 3D7 at all 84 amino acids. To further contextualize the genetic diversity sampled in this study with global PfCSP diversity, we analyzed an additional 3,809 Pfcs
d sequences from the Pf3k database and constructed a haplotype network representing 15 countries from Africa and Asia. The diversity observed in our samples was similar to the diversity observed in the global haplotype network. These observations underscore the need for additional research assessing genetic diversity in *P. falciparum* and the impact of PfCSP diversity on RTS,S/AS01 efficacy.

Although indoor residual spraying (IRS) and insecticide treated bednets (ITNs) have dramatically decreased malaria transmission, the global impact of malaria remains high with an estimated 216 million cases reported in 20161–2. Sub-Saharan Africa experiences a disproportionately high burden of *Plasmodium falciparum* malaria, even in regions with high coverage of IRS and ITNs1,3. Recent World Health Organization (WHO) goals aim to reduce both malaria mortality and case incidence by 90% of 2015 levels by 20304. Given the inadequacy of IRS and ITNs to eliminate malaria in all transmission settings, additional tools are necessary5. Of particular interest is an effective vaccine which might enhance control efforts and reduce malaria associated morbidity and mortality, particularly in regions refractory to current interventions.

Circumsporozoite protein (CSP), the dominant surface protein coating infectious stage sporozoites, has been a focus of vaccine development since the observation that bites from irradiated, infectious mosquitoes induce protective immune responses5. *P. falciparum* CSP (PICSP) has three distinct regions: the conserved amino (N)-terminal region, the central repeat region (CRR) comprised of 37–42 NAMP repeats, and a polymorphic

---

1W. Harry Feinstone Department of Molecular Microbiology and Immunology, Johns Hopkins Malaria Research Institute, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. 2Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, UK. 3Medical Research Council (MRC) Centre for Genomics and Global Health, University of Oxford, Oxford, UK. 4The Wellcome Trust Sanger Institute, Hinxton, UK. 5Department of Epidemiology, Johns Hopkins Malaria Research Institute, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. 6Tropical Diseases Research Centre, Ndola, Zambia. 7Université Protestante au Congo and University of Kinshasa, Kinshasa, Democratic Republic of the Congo. Correspondence and requests for materials should be addressed to J.C.P. (email: jpringl3@jhu.edu)
carboxyl (C) -terminal containing two sub-regions of high diversity known as Th2R and Th3R that elicit T-cell responses. The CRR contains the most immunogenic B-cell sporozoite epitopes and anti-CSP antibodies induced by exposure to irradiated, live sporozoites prevent infection by binding at the CRR in animal models. Malaria vaccine development efforts have spanned multiple decades and recently culminated in the licensure of the RTS,S/AS01 vaccine by GlaxoSmithKline (GSK) in 2015. The RTS,S/AS01 vaccine is a recombinant protein vaccine containing a portion of the NANN repeats (B-cell epitopes) and the C-terminal region (B-cell and T-cell epitopes) of the PfCSP fused with hepatitis B surface antigen (HBsAg) and is administered with a novel adjuvant, AS01. The vaccine construct is based on the P. falciparum 3D7 clone, which was derived from the NF54 strain isolated from a patient living near Schipol Airport in Amsterdam. Phase III clinical trials carried out at 11 sites across seven countries in sub-Saharan Africa demonstrated an overall vaccine efficacy (estimated using negative binomial regression) against clinical malaria from month zero to study end (children: median 48 months until study end; infants: median 38 months until study end) of 36.3% in children aged 5–17 months who received 3 primary doses of RTS,S plus a booster at 20 months. In 2015, the European Medicines Agent (EMA) approved the use of RTS,S/AS01. The Malaria Vaccine Implementation Programme (MVIP) led by WHO will begin RTS,S/AS01 implementation in three high transmission regions of Ghana, Kenya, and Malawi in 2018 with the goals of continued evaluation of the vaccine's impact on mortality, evaluating the feasibility of deploying the four dose vaccine series, and continued monitoring of vaccine safety.

Because the gene encoding P. falciparum CSP (PfcsP) is globally diverse, multiple studies were conducted during the RTS,S/AS01 Phase II clinical trials to monitor PfcsP haplotypes from vaccine and placebo recipients for signals of allele-specific vaccine-induced immunity. Four studies conducted in The Gambia, Kenya, and Mozambique found no evidence of allele-specific vaccine-induced immunity. These genetic surveillance analyses relied either on Sanger sequencing or oligonucleotide hybridization assays to assign genotypes to P. falciparum isolates. While state of the art assays at the time, the advent of affordable and scalable next generation sequencing technologies with the capacity to rapidly analyze larger sample sets has rendered both methods outdated.

Following the Phase III clinical trials, researchers used Illumina MiSeq and PacBio next generation sequencing technologies respectively to sequence both the C-terminal and CRR regions of PfcsP from parasites collected from individuals vaccinated with RTS,S/AS01 or placebo at 11 Phase III trial sites. Cumulative vaccine efficacy was reduced from 50.3% for parasites with a perfect PfcsP C-terminal sequence match to only 33.4% for parasites with any amino acid mismatch in this region. Although previous studies did not provide evidence to support the allele dependent nature of RTS,S/AS01 vaccine efficacy, this recent analysis using technologically advanced methods and a larger sample size suggests allele-specific immunity is important in eliciting protection. Further, this observation offers a potential explanation into the wide range of RTS,S/AS01 efficacies observed during Phase III clinical trials across the 11 sites (range: 22.0–74.6%).

In Zambia, malaria risk is heterogeneous with regions targeted for malaria elimination in the south and districts in which prevalence is greater than 50% throughout the year in the north. Nchelenge District is located in northern Zambia in Luapula Province, the province with the highest malaria prevalence in children younger than five years of age (>50% by malaria rapid diagnostic test in 2014). Despite a decade of malaria control interventions in Nchelenge District, including implementation of ITNs and IRS, malaria transmission remains holoendemic, with prevalence greater than 50% through 2017 (unpublished data). While Zambia scales up to meet its 2021 malaria elimination goal, it is important to consider the utility of introducing RTS,S/AS01 into the current arsenal of tools to reduce malaria morbidity and mortality in this region. Towards this goal, the genetic diversity of the C-terminal PfcsP was characterized with respect to the vaccine strain in parasites collected from the border of northern Zambia and the Democratic Republic of the Congo (DRC). Further, the genetic diversity of the C-terminal PfcsP sequences in this study was contextualized within global PfcsP diversity from the PFk database.

**Methods**

**Sample Collection, DNA Extraction, and Quantification.** Dried blood spot (DBS) samples were collected at one time point from consenting individuals living in randomly selected households during June and July 2016 in Nchelenge District, Zambia as well as two villages, Kilwa and Kashobwe, located directly across the Zambian border in the DRC. One hundred and three DBS samples from unique individuals (64 Zambian participants and 39 DRC participants) ranging in age from 8 months to 72 years (mean 17.7 years) and identified to be P. falciparum positive through qPCR screening of the DBS in Zambia were shipped to the Johns Hopkins Bloomberg School of Public Health, extracted using 20% Chelex, and quantified using P. falciparum lactose dehydrogenase (Pfdh) qPCR.

**Amplicon Generation and Sequencing.** A 300-bp amplicon containing the C-terminal PfcsP (839-1,139-bp, clone 3D7 0304600.1, PlasmoDB) was amplified from P. falciparum positive samples using the forward primer GACAAGGTCACAATATGGCACA and reverse primer ACATTAAACACACTGGAACATTTTTC fused with Illumina MiSeq adapter sequences for library indexing during PCR. PCR amplification reaction components included: 10μL DNA template, 12.5μL KAPA HiFi HotStart ReadyMix (Kapa Biosystems, Wilmington, Massachusetts), 0.25μL each of 20μM forward and reverse primers containing Illumina adapters, and 2μL of 25μM magnesium chloride. PCR cycling conditions were 95°C for 5 minutes followed by 30 cycles of 98°C for 20 seconds, 61°C for 30 seconds, 72°C for 1 minute, 72°C for 5 minutes, and a holding step at 4°C. Amplicon size (300-bp) was verified using TapeStation (Agilent 4200, Santa Clara). Seventy-five percent of the 103 samples successfully generated 300-bp PfcsP amplicons. Among samples with >50 p/μL by qPCR, the amplicon generation success rate was 100%; for samples with <50 p/μL by qPCR, the success rate was 7%. The arithmetic mean
of parasite copy number for samples that failed to generate amplicons was 11.1 p/μL (range: 0.5 p/μL–44.5 p/μL) compared with arithmetic mean 8,666 p/μL (range: 4.0 p/μL–81,196.4 p/μL) for samples for which amplicon generation was successful.

Amplicons were uniquely barcoded in a subsequent PCR reaction containing Nextera (Illumina, San Diego, California) indexes as described by Illumina. Indexed amplicon sizes were verified using TapeStation, purified using AMPure beads (Beckman Coulter, Brea, California), and quantified using PicoGreen (ThermoFisher Scientific, Waltham, Massachusetts). The Pfcs sequences were normalized and combined into a single pool for 300-bp paired end sequencing on a MiSeq at the Sequencing and Synthesis Facility at the Johns Hopkins School of Medicine.

Bioinformatic Processing and Analysis. Forward and reverse reads were merged using FLASH, trimmed for quality (sliding window = 50-bp, step size = 5-bp, quality threshold = 20) and collapsed by haplotype using SeekDeep’s default Illumina settings and allowing for one high quality mismatch within individuals. Samples included in the final analysis were supported by high read coverage, with an average of 29,439 reads. Haplotypes found to represent at least 1% of a sample were considered in the final analysis in order to minimize the inclusion of false positive haplotypes. The number of genetically distinct parasite haplotypes per individual, or complexity of infection (COI), was determined by the number of unique haplotype clusters per individual, as estimated by SeekDeep. The Pfcs sequences were translated to amino acid sequences and aligned to the 3D7 vaccine reference strain (3D7 0304600.1, PlasmoDB) in Geneious (version 9.1.5). The number of amino acid differences was calculated for each sequence and the 3D7 reference sequence for the 84 amino acids in the C-terminal amplicon (amino acids 288–371).

Data Availability Statement. The unique DNA sequences obtained from this study have been deposited in GenBank (accession numbers MG715504-MG715555).

Pf3k Sequence Acquisition and Global Diversity Analysis. Global Pfcs diversity was examined by mining the MalariaGEN Pf3k Project (release 5) which includes 2,512 Plasmodium falciparum full genomes from 14 countries worldwide. We retrieved all genetic variants on chromosome 3 available in Pf3k in variant call format (VCF) from release 5.1. Variant calls were made using GATK best practice haplotypeCaller with variants in the C-terminal Pfcs region (nucleotides 866–1,113) were extracted for the 1,147 monoclonal infections from Africa and Asia. The individual Pfcs haplotype sequences for the 1,365 multi-clonal samples were reconstructed using DEploid with appropriate reference panels of mono-clonal samples from the Pf3k dataset. We constructed a network based on the method by Templeton, Cranndall, and Sing (TCS) using PopArt to assess genealogical relationships between the global Pfcs haplotypes found in Pf3k and the haplotypes from Zambia and the DRC. Genetic diversity metrics were calculated using DnaSP (version 6.10.01). We compared sequences between African and Asian countries in terms of diversity and for evidence of population differentiation by calculating FST. Similarly, we compared samples from east and west African countries and calculated FST for signatures of population structure. For the purposes of comparing east and west Africa, we grouped samples in the DRC with east African samples.

Ethics Approval and Informed Consent. This research was approved by the Institutional Review Board at the Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland, USA, the Ethics Review Committee for the Tropical Diseases Research Centre in Ndola, Zambia, and by Le Comité d’éthique de l’Université Protestante au Congo in Kinshasa, the Democratic Republic of the Congo. All studies were conducted in accordance with the ethical guidelines set forth by the aforementioned review boards. All adults who participated in these studies gave informed consent. All child participants gave assent and had parental consent for participation.

Results

Of the 103 DBS samples extracted from our collections in Zambia and the DRC, 77 yielded suitable Pfcs amplicons for sequencing. Fifty five of the 77 samples sequenced passed quality filtering steps implemented by FLASH and SeekDeep. Two samples were excluded from the analysis for lacking full length Pfcs sequences. Overall, 193 PICSP haplotype sequences from 53 individuals were characterized, corresponding to 52 unique haplotypes (Table 1) ranging in population frequency from 1 to 22 (mean 3.7) observations across the 53 individuals. The 53 individuals were infected with a mean of 3.6 genetically distinct parasite haplotypes (range: 1–10). Of the 193 parasite sequences characterized from the 53 human samples, only ten matched the 3D7 haplotype at all 84 amino acid positions (5.2%). The median number of amino acid differences across all 193 parasite haplotypes was seven. Of the 193 3D7-type parasites, eight were found in individuals from Zambia and two were found in individuals from the DRC. The proportion of individuals harboring a 3D7-type parasite was similar between countries (Zambia: 25.8%, DRC: 9.1%, Pearson’s chi-squared test of equal proportions: p = 0.24) and between sampling sites (Kilwa: 0%, Kashobwe: 13.3%, Nchelenge lakeside: 31.8% Nchelenge inland: 11.1%, Pearson’s chi-squared test of equal proportions: p = 0.20). The frequency of 3D7 (vaccine) matched parasites was lower than previously reported in other African countries. Furthermore, all 10 3D7-type parasites were found in the context of polyclonal infections (range 3–7 haplotypes). The 3D7-type was not the major haplotype in nine of these ten (90%) polyclonal infections (median relative abundance of 3D7-type haplotypes in an individual = 5%; range = 1–24%). Divergent regions of the PICSP amplicon from 3D7 were visualized by plotting the percentage of samples sharing the 3D7 reference amino acid at each of the 84 amino acid positions (Fig. 1). Twenty-two polymorphic amino acid positions were identified, eight of which were within the Th2R region (amino acids 311–327) and six were contained in Th3R (amino acids 352–363).
The Pfcsn nucleotide diversity observed within Nchelenge District, Zambia, and Kilwa and Kashobwe in the DRC appeared to be representative of African Pfcsn in the TCS haplotype network constructed using samples from this study in addition to 3,809 sequences from the Pf3k database. In total, we identified 393 unique Pfcsn haplotypes, of which seven account for 51.3% of all the 4,002 sequences analyzed in this study. No clear population structure was identified between east and west African isolates ($F_{ST} = 0.008$), although signatures of moderate population differentiation were observed between African and Asian samples (Fig. 2), with $F_{ST} = 0.163$. We observed higher nucleotide diversity among African isolates than in Asian isolates, but no difference between east and west African isolates (Table 2). Among all African isolates analyzed, the 3D7 haplotype represented only 5.3% of the African 2,635 sequences. Including both Asian and African isolates, the 3D7 haplotype represented only 3.6% of 4,002 sequences. Among the Asian isolates, only three sequences matched the 3D7 haplotype (0.2%), all reported from Bangladesh.

### Table 1. Amino Acid Haplotype Frequencies

| Rank order of unique haplotypes | Frequency | Number of AA Differences |
|---------------------------------|-----------|--------------------------|
| 1                               | 22        | 8                        |
| 2                               | 15        | 5                        |
| 3                               | 13        | 8                        |
| 4                               | 11        | 7                        |
| 5                               | 10        | 0                        |
| 6                               | 10        | 6                        |
| 7                               | 8         | 7                        |
| 8                               | 7         | 7                        |
| 9                               | 7         | 8                        |
| 10                              | 7         | 8                        |
| 11                              | 5         | 7                        |
| 12–15                           | 4         | 5–7                      |
| 16–21                           | 3         | 4–8                      |
| 22–33                           | 2         | 2–9                      |
| 34–52                           | 1         | 5–10                     |

### Figure 1. RTS,S Amino Acid Changes and Positions

The 84 amino acids (positions 288–371) comprising the C-terminal amplicon are represented by columns in the bar-chart. The percentage of samples sharing the 3D7 amino acid are represented in pale yellow. Non-3D7 amino acid alternatives are represented in descending order of frequency in dark blue, red, light blue, or orange. Below the bar-chart, the 3D7 amino acid sequence is shown, with positions corresponding to the coordinates above. The substitutions at each of the 84 positions are enumerated below the 3D7 sequence.
**Discussion**

RTS,S vaccine efficacy has been shown to decline from 50.3% for vaccine matched parasite CSP haplotypes to 33.4% for unmatched haplotypes. In our study sites, only 10 parasite sequences of the 193 recovered (5.2%) were an exact match to the amino acid sequence in the C-terminal CSP region of the vaccine strain. The proportion of

---

**Table 2. Pfcsp Global Diversity Statistics.** Summarizes the samples included in the Pfcsp network analysis, including 193 samples sequenced in this study and 3,809 from Pf3k database. n = number of sequences, h = number of unique haplotypes, S = number of segregating sites (out of total possible 35), K = average number of pairwise nucleotide differences, Hd = haplotype diversity, π = nucleotide diversity.
individuals harboring 3D7-type parasites was similar between Zambia and the DRC. The frequency of vaccine matched parasites observed in this study is lower than previously reported in other African countries. It was previously demonstrated that RTS,S vaccine efficacy declined substantially for parasites not matching 3D7 in the C-terminal region of Pfcsp, and that vaccine efficacy declines as the number of amino acid differences increases. Therefore, the implication of parasites along the Zambia-DRC border differing from 3D7 at a median of seven amino acids is potentially significant. All 10 3D7-type parasites identified in our study occurred in the context of polyclonal infections (range 3–7 haplotypes). While Neafsey et al. evaluated the proportion of infections containing a 3D7 matched parasite haplotype as a function of COI, how vaccine efficacy differs between monoclonal 3D7-type infections, infections where 3D7 is the major of multiple haplotypes, and infections where 3D7 is the minor of multiple haplotypes has not yet been studied. Additional studies aimed at clarifying the effect of polyclonal infections on vaccine efficacy are warranted.

RTS,S/AS01, the only currently licensed malaria vaccine, is based on the sequence of just one parasite clone, 3D7, of African origin. Given that the vaccine is based on an African parasite clone, considering to what extent circulating parasites from Asia differ relative to those in Africa can provide us with insight into how well RTS,S/AS01 may perform if implemented in Asian countries. In this study, we observe moderate population differentiation between Asian and African isolates (FST = 0.163). Previous studies that aimed to assess global Pfcsp diversity identified population structure between isolates from Africa and Asia, consistent with our observations. Barry et al. used Pfcsp sequences from GenBank (n = 604) and characterized global diversity in an approximately similar C-terminal cspp region (nucleotides 909–1140) to this study (nucleotides 866–1113). Although the Pfcsp network created by Barry et al. included only five of the sixteen countries represented in this study, the overall pattern of high global Pfcsp diversity was consistent across the two studies, strengthening the conclusions presented in both analyses. Notably, previous analyses of global Pfcsp diversity have not included isolates from multiple east African countries. Here, we include 910 Pfcsp sequences from three east African countries, providing, to our knowledge, the first large scale characterization of P. falciparum genetic diversity in relation to the RTS,S/AS01 vaccine across multiple countries in this historically understudied area. Further, previous research has focused on describing Pfcsp diversity from monoclonal malaria isolations which may underestimate true population genetic diversity. This study characterizes Pfcsp haplotypes from multiple clones present in polyclonal infections, providing a more complete analysis of population diversity.

Ideally, an effective malaria vaccine would provide protection against the majority of circulating parasites across multiple geographic regions. Our data provide evidence that Pfcsp exhibits high genetic diversity both locally and globally. Interestingly, the prevalence of the 3D7-type parasite strains in our study area (5.2%) is the same as that across all of the African countries included in the Pfhk dataset (5.3%, n = 140/2635). Among Asian isolates, the 3D7 haplotype is even less frequently observed (n = 3/1367) at only 0.2% prevalence. In fact, Bangladesh is the only Asian country in the Pfhk dataset in which the 3D7 haplotype was observed. These data support previous observations that C-terminal Pfcsp is diverse globally, and that 3D7-type parasites are more frequently found in African countries than Asian countries. The high degree of global genetic Pfcsp diversity may potentially reduce RTS,S/AS01 vaccine effectiveness, particularly in Asian countries where 3D7 was not or only rarely observed. Monitoring differential vaccine efficacy by Pfcsp haplotype during possible future RTS,S/AS01 implementation programs will be valuable.

Publically available sequence databases provide unparalleled opportunities to understand global pathogen population genetics. However, it is important to acknowledge the limitations of drawing inferences from non-randomly sampled sequences. Notably, countries, as well as regions within countries, have unequal rates of sample deposition into databases, leading to a geographically biased set of sequences which over-represent genotypes from a small number of geographic foci while under-representing large swaths of the globe. Further, the conclusions drawn from sequences obtained from any given sequence repository are subject to change as sample sizes and geographic distributions are continually updated and expanded. Finally, the sequences obtained from Pfhk represent a multitude of sampling strategies, time periods, and sequencing technologies, which prevent samples from various regions from being optimally comparable. These samples come from patients across a spectrum of ages rather than specifically from children who are the recipients of the RTS,S/AS01 vaccine. However, despite these inherent limitations, sequence databases are powerful tools capable of elucidating global patterns in pathogen population genetic diversity. These resources, coupled with functional laboratory studies as well as observational field research, have a critical role to play in vaccine development efforts against a repertoire of global pathogens, including malaria. In the context of this study and with these limitations, we recognize that we have likely underestimated the true global diversity of Pfcsp.

The data presented here highlight the diversity of C-terminal Pfcsp, particularly in sub-Saharan Africa which is a key target region for malaria vaccination programs. This study underscores the importance of incorporating population genetic studies into future malaria vaccine design, laboratory and clinical evaluation. Most importantly, assessing the diversity of C-terminal Pfcsp should be a component of the RTS,S/AS01 Malaria Vaccine Implementation Programme and in further refining our understanding of how genetic diversity affects RTS,S/AS01 efficacy.

References
1. World Health Organization. World Malaria Report (2017).
2. Bhatt, S. et al. The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. Nature. 526, 207–211 (2015).
3. Mukonka, V. M. et al. High burden of malaria following scale-up of control interventions in Nchelenge District, Luapula Province, Zambia. Malar J. 13, 153 (2014).
4. Clyde, D. F., Most, H., McCarthy, V. C. & Vanderberg, J. P. Immunization of man against sporozoite-induced falciparum malaria. Am J Med Sci. 266, 169–177 (1973).
5. Malik, A., Egan, J. E., Houghten, R. A., Sadoff, J. C. & Hoffman, S. L. Human cytotoxic T lymphocytes against the Plasmodium falciparum circumsporozoite protein. Med Sci. 88, 3300–3304 (1991).
6. Good, M. F. et al. Human T-cell recognition of the circumsporozoite protein of Plasmodium falciparum: Immunodominant T-cell domains map to the polymorphic regions of the molecule. Immunology, 85, 1199–1203 (1988).
7. Zavala, F., Cochrane, A. H., Nardin, E. H., Nussenzweig, R. S. & Nussenzweig, V. Circumsporozoite proteins of malaria parasites contain a single immunodominant region with two or more identical epitopes. J Exp Med. 157, 1947–1957 (1983).
8. Casares, S., Brumme, T.-D. & Rice, T.L. The RTS,S malaria vaccine. Vaccine. 28, 4880–4894 (2010).
9. Gordon, D. M. et al. Safety, immunogenicity, and efficacy of a recombinantly produced Plasmodium falciparum circumsporozoite protein-hepatitis B surface antigen subunit vaccine. J Infect Dis. 171, 1576–1585 (1995).
10. Walliker, D. et al. Genetic Analysis of the Human Malaria Parasite Plasmodium falciparum. Source Sci New Ser. 236, 1661–1666 (1987).
11. Clinical Trials Partnership. Efficacy and safety of RTS,S/AS01 malaria vaccine with or without a booster dose in infants and children in Africa: final results of a phase 3, individually randomised, controlled trial. Lancet. 361, 31–45 (2015).
12. World Health Organization. Q&A on the malaria vaccine implementation programme (MVIP). http://www.who.int/malaria/media/malaria-vaccine-implementation/en/ (2018).
13. Jalloh, A., Jalloh, M. & Matsuoka, H. T-cell epitope polymorphisms of the Plasmodium falciparum circumsporozoite protein field isolates from Sierra Leone: age-dependent haplotype distribution? Malar J. 8, 120 (2009).
14. Zeeshan, M. et al. Genetic Variation in the Plasmodium falciparum Circumsporozoite Protein in India and Its Relevance to RTS,S Malaria Vaccine. Gruener, A. C., ed. PLoS One. 7, e3430 (2012).
15. Barry, A. E., Schultz, L., Buckee, C. O., Reeder, J. C. Contrasting Population Structures of the Genes Encoding Ten Leading Vaccine-Candidate Antigens of the Human Malaria Parasite, Plasmodium falciparum. Rönis, L., ed. PLoS One. 4, e8497 (2009).
16. Wainmbi, J. N. et al. Impact of RTS,S/AS02A and RTS,S/AS01B on genotypes of P. falciparum in adults participating in a malaria vaccine clinical trial. PLoS One. 4, 1–8 (2009).
17. Alloouche, A. et al. Protective Efficacy of the RTS,S/AS02 Plasmodium falciparum Malaria Vaccine is Not Strain Specific. Am J Trop Med Hyg. 68, 97–101 (2003).
18. Enosse, M. et al. RTS,S/AS02A malaria vaccine does not induce parasite CSP T cell epitope selection and reduces multiplicity of infection. PLoS Clin Trials. 1, e5 (2006).
19. Bojang, K. A. et al. Efficacy of RTS, S/AS02 malaria vaccine against Plasmodium falciparum infection in semi-immune adult men in The Gambia: a randomised trial. Lancet. 358, 1927–1934 (2001).
20. Neafsey, D. E. et al. Genetic Diversity and Protective Efficacy of the RTS,S/AS01 Malaria Vaccine. N Engl J Med. 373, 2025–2037 (2015).
21. Moss, W. J. et al. JW. Malaria Epidemiology and Control within the International Centers of Excellence for Malaria Research. Am J Trop Med Hyg. 93, 5–15 (2015).
22. Kamuliro, M. et al. The changing burden of malaria and association with vector control interventions in Zambia using district-level surveillance data, 2006–2011. Malar J. 12, 437 (2013).
23. Pazz, J. B. et al. Estimation of Plasmodium falciparum Transmission Intensity in Lilongwe, Malawi, by Microscopy, Rapid Diagnostics Testing, and Nucleic Acid Detection. Am J Trop Med Hyg. 95, 373–377 (2016).
24. Aurrecoechea, C. et al. PlasmoDB: a functional genomic database for malaria parasites. Nucleic Acids Res. 37, D539–D543 (2009).
25. Illumina. 16S Metagenomic Sequencing Library Preparation. http://www.illumina.com. 1–28 (2013).
26. Mag, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics. 27, 2957–2963 (2011).
27. Hathaway NJ, Paroek CM, Juliano JJ, Bailey JA. SeekDeep: single-base resolution de novo clustering for amplicon deep sequencing. Nucleic Acids Res. November 2017.
28. The Pf3K Project: pilot data release 5. http://www.malarialogen.net/data/pf3k-5 (2016).
29. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303 (2010).
30. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 43, 491–498 (2011).
31. Zhu, S. J., Almagro-Garcia, J. & McVean, G. Deconvolution of multiple infections in Plasmodium falciparum from high throughput sequencing data. Bioinformatics. 34, 9–15 (2017).
32. Leigh, J. W. & Bryant, D. popart: full-feature software for haplotype network construction. In: S. Nakagawa ed.. Methods Ecol Evol. 6, (1110–1116 (2015).
33. Clement, M., Posada, D. & Crandall, K. A. TCS: a computer program to estimate genealogies. Mol Ecol. 9, 1657–1659 (2000).
34. Templeton, A. R., Crandall, K. A. & Sing, C. F. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics. 132, 619–33 (1992).
35. Preston, M. D. et al. A barcode of organellar genome polymorphisms identifies the geographic origin of Plasmodium falciparum strains. Nat Commun. 5, 1–7 (2014).
36. Bailey, I. A. et al. Use of Massively Parallel Pyrosequencing to Evaluate the Diversity of and Selection on Plasmodium falciparum csr T-Cell Epitopes in Lilongwe, Malawi. J Infect Dis. 206, 580–587 (2012).

Acknowledgements

We would like to thank the Southern and Central Africa International Centers of Excellence in Malaria Research (ICEMR) contributors and collaborators as well as the study participants and the communities of Nchelenge, Zambia and Haut-Katanga, DRC. This research is supported financially by The Molecular and Cellular Basis of Infectious Diseases (MCBID)T32 Training Grant (5T32AI07417-22) to JCP in the Department of Molecular Microbiology and Immunology (MMI) at the Johns Hopkins Bloomberg School of Public Health (JHSPH), the Bloomberg Philanthropies, a post-doctoral fellowship award to GC from the Johns Hopkins Malaria Research Institute (JHMR), and the NIH-funded Southern and Central Africa International Centers of Excellence in Malaria Research (ICEMR) 2U19AI089680. SJZ is funded by the Wellcome Trust grant (100956/Z/13/Z) to Gil McVean.

Author Contributions

J.C.P., G.C., W.J.M., D.E.N. conceived and designed the study; J.C.P. conducted the literature review; J.C.P., G.C., J.A.G., S.J.Z., T.K., M.M., T.B. collected and generated data; J.C.P., G.C., J.A.G., S.J.Z., D.E.N. analyzed and interpreted the data; W.J.M., D.E.N., J.A.G., S.J.Z., M.C. contributed reagents/materials/analysis tools; J.C.P., G.C., J.A.G., S.J.Z. wrote the manuscript; all authors read, edited and approved the manuscript for publication.
