Curriculum Vitae
Matthew Adams

Degree and Date: PhD, July 2025

Education

1999/08 – 2002/05	No-degree, Biology, St. Mary's College of Maryland
2002/08 – 2003/12	B.S., Neurobiology and Physiology, University of Maryland College Park
2018/09 – 2025/07	PhD Candidate, University of Maryland School of Medicine Program: Epidemiology and Human Genetics Track: Molecular Epidemiology Dissertation Committee Chair: Shannon Takala Harrison, PhD Dissertation Title: Identifying Correlates of Protective Immunity and Antigenic Escape in *Plasmodium falciparum* Malaria

Employment History

2004/06 – 2005/07	Biological Laboratory Technician, Henry Jackson Foundation
2005/07 – 2007/06	Biological Laboratory Technician, U.S. Department of Defense
2007/07 – 2008/10	Research Assistant, University of Maryland School of Medicine
2008/10 – 2010/05	Research Specialist, University of Maryland School of Medicine
2010/05 – 2014/06	Research Specialist, Howard Hughes Medical Institute
2014/06 – 2016/05	Laboratory Supervisor, Howard Hughes Medical Institute
2016/05 – 2018/07	Laboratory Supervisor, University of Maryland School of Medicine

Honors and Awards

2021	Patricia Langenberg Endowment in Women's Health and Epidemiology Award, University of Maryland Baltimore

Patents, Inventions and Copyrights

2016	PCT/US2016/058371, "Methods for Extracting Nucleic Acids from Dried Blood and Uses Thereof"

National Service

2015 - 2018	Ad Hoc Reviewer: PLOS One (1x/yr)
2015 - 2018	Ad Hoc Reviewer: American Journal of Tropical Medicine and Hygiene (1x/yr)

Professional Society Membership

2015 - 2016	American Society for Tropical Medicine and Hygiene

Institutional Service

2022 – 2024	Organizer, Graduate Student Forum of Malaria Research Program

Graduate Teaching Service

2019	Teaching Assistantship: Application of Legal and Regulatory Issues in Clinical Research, 9-12 graduate students (5 hours/week)
2020	Teaching Assistantship: Clinical and Translational Research Project Design and Implementation, 9-12 graduate students (5 hours/week)
2020	Teaching Assistantship: Molecular Epidemiology, 5-8 graduate students (5 hours/week)
2020	Guest Lecturer, Molecular Epidemiology (PREV780), 9 graduate students (2 lectures, 6 hours total)
2021	Teaching Assistantship: Foundations of Public Health, 30-35 graduate students (4 hours/week)
2021	Teaching Assistantship: Application of Legal and Regulatory Issues in Clinical Research, 9-15 graduate students (5 hours/week)
2021	Guest Lecturer, Foundations of Public Health (PH608) 20 graduate students (1 lecture, 2 hours)
2022	Guest Lecturer, Molecular Epidemiology (PREV780), 9 graduate students (2 lectures, 6 hours total)
2024	Guest Lecturer, Molecular Epidemiology (PREV780), 9 graduate students (2 lectures, 6 hours total)

Peer-reviewed Journal Articles (ORCID = 0000-0001-7062-903X)

1. Lyke KE, Laurens M, **Adams M**, Billingsley PF, Richman A, Loyevsky M, Chakravarty S, Plowe CV, Sim BK, Edelman R, Hoffman SL. *Plasmodium falciparum* Malaria Challenge by the Bite of Aseptic *Anopheles stephensi* Mosquitoes: Results of a Randomized Infectivity Trial. PLoS One. 2010 Oct 21;5(10):e13490. PubMed PMID: 21042404; PubMed Central PMCID: PMC2958836.

2. Takala-Harrison S, Clark TG, Jacob CG, Cummings MP, Miotto O, Dondorp AM, Fukuda MM, Nosten F, Noedl H, Imwong M, Bethell D, Se Y, Lon C, Tyner SD, Saunders DL, Socheat D, Ariey F, Phyo AP, Starzengruber P, Fuehrer HP, Swoboda P, Stepniewska K, Flegg J, Arze C, Cerqueira GC, Silva JC, Ricklefs SM, Porcella SF, Stephens RM, **Adams M**, Kenefic LJ, Campino S, Auburn S, MacInnis B, Kwiatkowski DP, Su XZ, White NJ, Ringwald P, Plowe CV. Genetic Loci Associated With Delayed Clearance of

Plasmodium falciparum Following Artemisinin Treatment in Southeast Asia. Proc Natl Acad Sci U S A. 2013 Jan 2;110(1):240-5. Epub 2012 Dec 17. PubMed PMID: 23248304; PubMed Central PMCID: PMC3538248.

3. Laurens MB, Billingsley P, Richman A, Eappen AG, **Adams M**, Li T, Chakravarty S, Gunasekera A, Jacob CG, Sim BK, Edelman R, Plowe CV, Hoffman SL, Lyke KE. Successful Human Infection with *P. falciparum* Using Three Aseptic *Anopheles stephensi* Mosquitoes: A New Model for Controlled Human Malaria Infection. PLoS One. 2013;8(7):e68969. Print 2013. PubMed PMID: 23874828; PubMed Central PMCID: PMC3712927.

4. Seder RA, Chang LJ, Enama ME, Zephir KL, Sarwar UN, Gordon IJ, Holman LA, James ER, Billingsley PF, Gunasekera A, Richman A, Chakravarty S, Manoj A, Velmurugan S, Li M, Ruben AJ, Li T, Eappen AG, Stafford RE, Plummer SH, Hendel CS, Novik L, Costner PJ, Mendoza FH, Saunders JG, Nason MC, Richardson JH, Murphy J, Davidson SA, Richie TL, Sedegah M, Sutamihardja A, Fahle GA, Lyke KE, Laurens MB, Roederer M, Tewari K, Epstein JE, Sim BK, Ledgerwood JE, Graham BS, Hoffman SL, VRC 312 Study Team. Protection Against Malaria by Intravenous Immunization with a Nonreplicating Sporozoite Vaccine. Science. 2013 Sep 20;341(6152):1359-65. Epub 2013 Aug 8. PubMed PMID: 23929949 (Member of VRC 312 Study Team).

5. Villapol S, Wang Y, **Adams M**, Symes AJ. Smad3 Deficiency Increases Cortical and Hippocampal Neuronal Loss Following Traumatic Brain Injury. Exp Neurol. 2013 Oct;250:353-65. Epub 2013 Oct 9. PubMed PMID: 24120438.

6. Murphy SC, Hermsen CC, Douglas AD, Edwards NJ, Petersen I, Fahle GA, **Adams M**, Berry AA, Billman ZP, Gilbert SC, Laurens MB, Leroy O, Lyke KE, Plowe CV, Seilie AM, Strauss KA, Teelen K, Hill AV, Sauerwein RW. External Quality Assurance of Malaria Nucleic Acid Testing for Clinical Trials and Eradication Surveillance. PLoS One. 2014;9(5):e97398. eCollection 2014. PubMed PMID: 24838112; PubMed Central PMCID: PMC4023973.

7. Takala-Harrison S, Jacob CG, Arze C, Cummings MP, Silva JC, Dondorp AM, Fukuda MM, Hien TT, Mayxay M, Noedl H, Nosten F, Kyaw MP, Nhien NT, Imwong M, Bethell D, Se Y, Lon C, Tyner SD, Saunders DL, Ariey F, Mercereau-Puijalon O, Menard D, Newton PN, Khanthavong M, Hongvanthong B, Starzengruber P, Fuehrer HP, Swoboda P, Khan WA, Phyo AP, Nyunt MM, Nyunt MH, Brown TS, **Adams M**, Pepin CS, Bailey J, Tan JC, Ferdig MT, Clark TG, Miotto O, MacInnis B, Kwiatkowski DP, White NJ, Ringwald P, Plowe CV. Independent Emergence Of Artemisinin Resistance Mutations Among *Plasmodium falciparum* In Southeast Asia. J Infect Dis. 2015 Mar 1;211(5):670-9. Epub 2014 Sep 1. PubMed PMID: 25180241; PubMed Central PMCID: PMC4334802.

8. Ouattara A, Kone A, **Adams M**, Fofana B, Maiga AW, Hampton S, Coulibaly D, Thera MA, Diallo N, Dara A, Sagara I, Gil JP, Bjorkman A, Takala-Harrison S, Doumbo OK, Plowe CV, Djimde AA. Polymorphisms In the K13-Propeller Gene in Artemisinin-Susceptible *Plasmodium falciparum* Parasites from Bougoula-Hameau and Bandiagara, Mali. Am J Trop Med Hyg. 2015 Jun;92(6):1202-6. Epub 2015 Apr 27. PubMed PMID: 25918205; PubMed Central PMCID: PMC4458826.

9. Travassos MA, Coulibaly D, Bailey JA, Niangaly A, **Adams M**, Nyunt MM, Ouattara A, Lyke KE, Laurens MB, Pablo J, Jasinskas A, Nakajima R, Berry AA, Takala-Harrison S, Kone AK, Kouriba B, Rowe JA, Doumbo OK, Thera MA, Laufer MK, Felgner PL, Plowe CV. Differential Recognition of Terminal Extracellular *Plasmodium falciparum* VAR2CSA Domains by Sera from Multigravid, Malaria-exposed Malian Women. Am J Trop Med Hyg. 2015 Jun;92(6):1190-1194. Epub 2015 Apr 27. PubMed PMID: 25918203; PubMed Central PMCID: PMC4458824.

10. Huang F, Takala-Harrison S, Jacob CG, Liu H, Sun X, Yang H, Nyunt MM, **Adams M**, Zhou S, Xia Z, Ringwald P, Bustos MD, Tang L, Plowe CV. A Single Mutation in K13 Predominates in Southern China and is Associated with Delayed Clearance of *Plasmodium falciparum* Following Artemisinin Treatment. J Infect Dis. 2015 Nov 15;212(10):1629-35. Epub 2015 Apr 24. PubMed PMID: 25910630; PubMed Central PMCID: PMC4621243.

11. Travassos MA, Coulibaly D, Laurens MB, Dembélé A, Tolo Y, Koné AK, Traoré K, Niangaly A, Guindo A, Wu Y, Berry AA, Jacob CG, Takala-Harrison S, **Adams M**, Shrestha B, Mu AZ, Kouriba B, Lyke KE, Diallo DA, Doumbo OK, Plowe CV, Thera MA. Hemoglobin C Trait Provides Protection from Clinical Falciparum Malaria in Malian Children. J Infect Dis. 2015 Dec 1;212(11):1778-86. Epub 2015 May 27. PubMed PMID: 26019283; PubMed Central PMCID: PMC4633765.

12. Lyke KE, Laurens MB, Strauss K, **Adams M**, Billingsley PF, James E, Manoj A, Chakravarty S, Plowe CV, Li ML, Ruben A, Edelman R, Green M, Dube TJ, Sim BKL, Hoffman SL. Optimizing Intradermal Administration of Cryopreserved *Plasmodium falciparum* Sporozoites in Controlled Human Malaria Infection. Am J Trop Med Hyg. 2015 Dec;93(6):1274-1284. Epub 2015 Sep 28. PubMed PMID: 26416102; PubMed Central PMCID: PMC4674246.

13. **Adams M**, Joshi SN, Mbambo G, Mu AZ, Roemmich SM, Shrestha B, Strauss KA, Johnson NE, Oo KZ, Hlaing TM, Han ZY, Han KT, Thura S, Richards AK, Huang F, Nyunt MM, Plowe CV. An Ultrasensitive Reverse Transcription Polymerase Chain Reaction Assay to Detect Asymptomatic Low-Density *Plasmodium falciparum* and *Plasmodium vivax* Infections in Small Volume Blood Samples. Malar J. 2015 Dec 23;14:520. PubMed PMID: 26701778; PubMed Central PMCID: PMC4690410.

14. Cohee LM, Kalilani-Phiri L, Mawindo P, Joshi S, **Adams M**, Kenefic L, Jacob CG, Taylor TE, Laufer MK. Parasite Dynamics in the Peripheral Blood and the Placenta During Pregnancy-Associated Malaria Infection. Malar J. 2016 Sep 21;15(1):483. PubMed PMID: 27653788; PubMed Central PMCID: PMC5031282.

15. Dara A, Travassos MA, **Adams M**, Schaffer DeRoo S, Drábek EF, Agrawal S, Laufer MK, Plowe CV, Silva JC. A New Method for Sequencing the Hypervariable *Plasmodium falciparum* Gene Var2csa from Clinical Samples. Malar J. 2017 Aug 17;16(1):343. PubMed PMID: 28818101; PubMed Central PMCID: PMC5561619.

16. Niangaly A, Karthikeyan Gunalan, Amed Ouattara, Coulibaly D, Sá JM, **Adams M**, Travassos MA, Ferrero J, Laurens MB, Kone AK, Thera MA, Plowe CV, Miller LH, Doumbo OK. *Plasmodium vivax* Infections Over 3 Years in Duffy Blood Group Negative

Malians in Bandiagara, Mali. Am J Trop Med Hyg. 2017 Sep;97(3):744-752. Epub 2017 Jul 27. PubMed PMID: 28749772; PubMed Central PMCID: PMC5590610.

17. Zainabadi K, **Adams M**, Han ZY, Lwin HW, Han KT, Ouattara A, Thura S, Plowe CV, Nyunt MM. A Novel Method for Extracting Nucleic Acids from Dried Blood Spots for Ultrasensitive Detection of Low-Density *Plasmodium falciparum* and *Plasmodium vivax* Infections. Malar J. 2017 Sep 18;16(1):377. PubMed PMID: 28923054; PubMed Central PMCID: PMC5604154.

18. Boudová S, Walldorf JA, Bailey JA, Divala T, Mungwira R, Mawindo P, Pablo J, Jasinskas A, Nakajima R, Ouattara A, **Adams M**, Felgner PL, Plowe CV, Travassos MA, Laufer MK. Mother-Newborn Pairs in Malawi Have Similar Antibody Repertoires to Diverse Malaria Antigens. Clin Vaccine Immunol. 2017 Oct;24(10). Print 2017 Oct. PubMed PMID: 28835359; PubMed Central PMCID: PMC5629668.

19. Huang F, Takala-Harrison S, Liu H, Xu JW, Yang HL, **Adams M**, Shrestha B, Mbambo G, Rybock D, Zhou SS, Xia ZG, Zhou XN, Plowe CV, Nyunt MM. Prevalence of Clinical and Subclinical *Plasmodium falciparum* and *Plasmodium vivax* Malaria in Two Remote Rural Communities on the Myanmar-China Border. Am J Trop Med Hyg. 2017 Nov;97(5):1524-1531. Epub 2017 Oct 10. PubMed PMID: 29016341; PubMed Central PMCID: PMC5817756.

20. Olotu A, Urbano V, Hamad A, Eka M, Chemba M, Nyakarungu E, Raso J, Eburi E, Mandumbi DO, Hergott D, Maas CD, Ayekaba MO, Milang DN, Rivas MR, Schindler T, Embon OM, Ruben AJ, Saverino E, Abebe Y, Kc N, James ER, Murshedkar T, Manoj A, Chakravarty S, Li M, **Adams M**, Schwabe C, Segura JL, Daubenberger C, Tanner M, Richie TL, Billingsley PF, Lee Sim BK, Abdulla S, Hoffman SL. Advancing Global Health through Development and Clinical Trials Partnerships: A Randomized, Placebo-Controlled, Double-Blind Assessment of Safety, Tolerability, and Immunogenicity of PfSPZ Vaccine for Malaria in Healthy Equatoguinean Men. Am J Trop Med Hyg. 2018 Jan;98(1):308-318. Epub 2018 Jan 1. PubMed PMID: 29141739; PubMed Central PMCID: PMC5928718.

21. Travassos MA, Niangaly A, Bailey JA, Ouattara A, Coulibaly D, Lyke KE, Laurens MB, Pablo J, Jasinskas A, Nakajima R, Berry AA, **Adams M**, Jacob CG, Pike A, Takala-Harrison S, Liang L, Kouriba B, Kone AK, Rowe JA, Moulds J, Diallo DA, Doumbo OK, Thera MA, Felgner PL, Plowe CV. Children With Cerebral Malaria Or Severe Malarial Anaemia Lack Immunity To Distinct Variant Surface Antigen Subsets. Sci Rep. 2018 Apr 19;8(1):6281. PubMed PMID: 29674705; PubMed Central PMCID: PMC5908851.

22. Ouattara A, Tran TM, Doumbo S, **Adams M**, Agrawal S, Niangaly A, Nelson-Owens S, Doumtabé D, Tolo Y, Ongoiba A, Takala-Harrison S, Traoré B, Silva JC, Crompton PD, Doumbo OK, Plowe CV. Extent and Dynamics of Polymorphism in the Malaria Vaccine Candidate *Plasmodium falciparum* Reticulocyte-Binding Protein Homologue-5 in Kalifabougou, Mali. Am J Trop Med Hyg. 2018 Jul;99(1):43-50. Epub 2018 May 24. PubMed PMID: 29848401; PubMed Central PMCID: PMC6085788.

23. Yaqoob A, Khattak AA, Nadeem MF, Fatima H, Mbambo G, Ouattara A, **Adams M**, Zeeshan N, Takala-Harrison S. Prevalence of Molecular Markers of Sulfadoxine-pyrimethamine and Artemisinin Resistance in *Plasmodium falciparum* from

Pakistan. Malar J. 2018 Dec 17;17(1):471. PubMed PMID: 30558587; PubMed Central PMCID: PMC6296135.

24. Friedman-Klabanoff DJ, Laurens MB, Berry AA, Travassos MA, **Adams M**, Strauss KA, Shrestha B, Levine MM, Edelman R, Lyke KE. The Controlled Human Malaria Infection Experience at the University of Maryland. Am J Trop Med Hyg. 2019 Mar;100(3):556-565. PubMed PMID: 30675854; PubMed Central PMCID: PMC6402913.

25. Zhou AE, Berry AA, Bailey JA, Pike A, Dara A, Agrawal S, Stucke EM, Ouattara A, Coulibaly D, Lyke KE, Laurens MB, **Adams M**, Takala-Harrison S, Pablo J, Jasinskas A, Nakajima R, Niangaly A, Kouriba B, Kone AK, Rowe JA, Doumbo OK, Thera MA, Patel JJ, Tan JC, Felgner PL, Plowe CV, Travassos MA. Antibodies to Peptides in Semiconserved Domains of RIFINs and STEVORs Correlate with Malaria Exposure. mSphere. 2019 Mar 20;4(2). PubMed PMID: 30894432; PubMed Central PMCID: PMC6429043.

26. Seilie AM, Chang M, Hanron AE, Billman ZP, Stone BC, Zhou K, Olsen TM, Daza G, Ortega J, Cruz KR, Smith N, Healy SA, Neal J, Wallis CK, Shelton L, Mankowski TV, Wong-Madden S, Mikolajczak SA, Vaughan AM, Kappe SHI, Fishbaugher M, Betz W, Kennedy M, Hume JCC, Talley AK, Hoffman SL, Chakravarty S, Sim BKL, Richie TL, Wald A, Plowe CV, Lyke KE, **Adams M**, Fahle GA, Cowan EP, Duffy PE, Kublin JG, Murphy SC. Beyond Blood Smears: Qualification of *Plasmodium* 18S rRNA as a Biomarker for Controlled Human Malaria Infections. Am J Trop Med Hyg. 2019 Jun;100(6):1466-1476. PubMed PMID: 31017084; PubMed Central PMCID: PMC6553913.

27. Stucke EM, Niangaly A, Berry AA, Bailey JA, Coulibaly D, Ouattara A, Lyke KE, Laurens MB, Dara A, **Adams M**, Pablo J, Jasinskas A, Nakajima R, Zhou AE, Agrawal S, Friedman-Klabanoff DJ, Takala-Harrison S, Kouriba B, Kone AK, Rowe JA, Doumbo OK, Felgner PL, Thera MA, Plowe CV, Travassos MA. Serologic Responses to the Pfemp1 DBL-CIDR Head Structure May Be a Better Indicator of Malaria Exposure Than Those to the DBL-A Tag. Malar J. 2019 Aug 13;18(1):273. PubMed PMID: 31409360; PubMed Central PMCID: PMC6692945.

28. Laurens MB, Berry AA, Travassos MA, Strauss K, **Adams M**, Shrestha B, Li T, Eappen A, Manoj A, Abebe Y, Murshedkar T, Gunasekera A, Richie TL, Lyke KE, Plowe CV, Kennedy JK, Potter GE, Deye GA, Sim BKL, Hoffman SL. Dose-Dependent Infectivity of Aseptic, Purified, Cryopreserved *Plasmodium falciparum* 7G8 Sporozoites in Malaria-Naive Adults. J Infect Dis. 2019 Nov 6;220(12):1962-1966. PubMed PMID: 31419294; PubMed Central PMCID: PMC6834064.

29. Moser KA, Drábek EF, Dwivedi A, Stucke EM, Crabtree J, Dara A, Shah Z, **Adams M**, Li T, Rodrigues PT, Koren S, Phillippy AM, Munro JB, Ouattara A, Sparklin BC, Dunning Hotopp JC, Lyke KE, Sadzewicz L, Tallon LJ, Spring MD, Jongsakul K, Lon C, Saunders DL, Ferreira MU, Nyunt MM, Laufer MK, Travassos MA, Sauerwein RW, Takala-Harrison S, Fraser CM, Sim BKL, Hoffman SL, Plowe CV, Silva JC. Strains used in whole organism *Plasmodium falciparum* vaccine trials differ in genome structure, sequence, and immunogenic potential. Genome Med. 2020 Jan 8;12(1):6. PubMed PMID: 31915075; PubMed Central PMCID: PMC6950926.

30. Bailey JA, Berry AA, Travassos MA, Ouattara A, Boudova S, Dotsey EY, Pike A, Jacob CG, **Adams M**, Tan JC, Bannen RM, Patel JJ, Pablo J, Nakajima R, Jasinskas A, Dutta S, Takala-Harrison S, Lyke KE, Laurens MB, Niangaly A, Coulibaly D, Kouriba B, Doumbo OK, Thera MA, Felgner PL, Plowe CV. Microarray Analyses Reveal Strain-Specific Antibody Responses to *Plasmodium falciparum* Apical Membrane Antigen 1 Variants Following Natural Infection and Vaccination. Sci Rep. 2020 Mar 3;10(1):3952. PubMed PMID: 32127565; PubMed Central PMCID: PMC7054363.

31. Shah Z, **Adams M**, Moser KA, Shrestha B, Stucke EM, Laufer MK, Serre D, Silva JC, Takala-Harrison S. Optimization of Parasite DNA Enrichment Approaches to Generate Whole Genome Sequencing Data for *Plasmodium falciparum* from Low Parasitaemia Samples. Malar J. 2020 Mar 30;19(1):135. PubMed PMID: 32228559; PubMed Central PMCID: PMC7106660.

32. Ouattara A, Niangaly A, **Adams M**, Coulibaly D, Kone AK, Traore K, Laurens MB, Tolo Y, Kouriba B, Diallo DA, Doumbo OK, Plowe CV, Djimdé A, Thera MA, Laufer MK, Takala-Harrison S, Silva JC. Epitope-Based Sieve Analysis of *Plasmodium falciparum* Sequences from a FMP2.1/AS02A Vaccine Trial is Consistent with Differential Vaccine Efficacy Against Immunologically Relevant AMA1 Variants. Vaccine. 2020 Jul 31;38(35):5700-5706. Epub 2020 Jun 20. PubMed PMID: 32571720; PubMed Central PMCID: PMC7375901.

33. Huang F, Jacob CG, Takala-Harrison S, **Adams M**, Yang HL, Liu H, Xia ZG, Zhou SS, Tang LH, Plowe CV. Genomic Epidemiology of Antimalarial Drug Resistance in *Plasmodium falciparum* in Southern China. Front Cell Infect Microbiol. 2020;10:610985. eCollection 2020. PubMed PMID: 33489939; PubMed Central PMCID: PMC7820777.

34. Shah Z, Naung MT, Moser KA, **Adams M**, Buchwald AG, Dwivedi A, Ouattara A, Seydel KB, Mathanga DP, Barry AE, Serre D, Laufer MK, Silva JC, Takala-Harrison S. Whole-Genome Analysis of Malawian *Plasmodium falciparum* Isolates Identifies Possible Targets of Allele-Specific Immunity to Clinical Malaria. PLoS Genet. 2021 May;17(5):e1009576. eCollection 2021 May. PubMed PMID: 34033654; PubMed Central PMCID: PMC8184011.

35. Friedman-Klabanoff DJ, Travassos MA, Ifeonu OO, Agrawal S, Ouattara A, Pike A, Bailey JA, **Adams M**, Coulibaly D, Lyke KE, Laurens MB, Takala-Harrison S, Kouriba B, Kone AK, Doumbo OK, Patel JJ, Thera MA, Felgner PL, Tan JC, Plowe CV, Berry AA. Epitope-Specific Antibody Responses to a *Plasmodium falciparum* Subunit Vaccine Target in a Malaria-Endemic Population. J Infect Dis. 2021 Jun 4;223(11):1943-1947. PubMed PMID: 32992328; PubMed Central PMCID: PMC8176638.

36. Ventimiglia NT, Stucke EM, Coulibaly D, Berry AA, Lyke KE, Laurens MB, Bailey JA, **Adams M**, Niangaly A, Kone AK, Takala-Harrison S, Kouriba B, Doumbo OK, Felgner PL, Plowe CV, Thera MA, Travassos MA. Malian Adults Maintain Serologic Responses to Virulent Pfemp1s Amid Seasonal Patterns of Fluctuation. Sci Rep. 2021 Jul 13;11(1):14401. PubMed PMID: 34257318; PubMed Central PMCID: PMC8277812.

37. Berry AA, Obiero JM, Travassos MA, Ouattara A, Coulibaly D, **Adams M**, de Assis RR, Jain A, Taghavian O, Sy A, Nakajima R, Jasinskas A, Laurens MB, Takala-Harrison S, Kouriba B, Kone AK, Doumbo OK, Sim BKL, Hoffman SL, Plowe CV, Thera

MA, Felgner PL, Lyke KE. Immunoprofiles Associated with Controlled Human Malaria Infection and Naturally Acquired Immunity Identify a Shared Iga Pre-Erythrocytic Immunoproteome. NPJ Vaccines. 2021 Sep 13;6(1):115. PubMed PMID: 34518543; PubMed Central PMCID: PMC8438027.

38. Shrestha B, Shah Z, Morgan AP, Saingam P, Chaisatit C, Chaorattanakawee S, Praditpol C, Boonyalai N, Lertsethtakarn P, Wojnarski M, Deutsch-Feldman M, **Adams M**, Sea D, Chann S, Tyner SD, Lanteri CA, Spring MD, Saunders DL, Smith PL, Lon C, Gosi P, Sok S, Satharath P, Rekol H, Lek D, Vesely BA, Lin JT, Waters NC, Takala-Harrison S. Distribution and Temporal Dynamics of *Plasmodium falciparum* Chloroquine Resistance Transporter Mutations Associated With Piperaquine Resistance in Northern Cambodia. J Infect Dis. 2021 Sep 17;224(6):1077-1085. PubMed PMID: 33528566; PubMed Central PMCID: PMC8448439.

39. Ouattara A, Dwivedi A, **Adams M**, Niangaly A, Laurens MB, Nyunt MM, Plowe CV, Djimde A, Takala-Harrison S, Silva JC. An In Silico Analysis of Malaria Pre-Erythrocytic-Stage Antigens Interpreting Worldwide Genetic Data to Suggest Vaccine Candidate Variants and Epitopes. Microorganisms. 2022 May 25;10(6). PubMed PMID: 35744609; PubMed Central PMCID: PMC9231253.

40. Sobota RS, Goron AR, Berry AA, Bailey JA, Coulibaly D, **Adams M**, Kone AK, Kouriba B, Doumbo OK, Sztein MB, Felgner PL, Plowe CV, Lyke KE, Thera MA, Travassos MA. Serologic and Cytokine Profiles of Children with Concurrent Cerebral Malaria and Severe Malarial Anemia Are Distinct from Other Subtypes of Severe Malaria. Am J Trop Med Hyg. 2022 Jun 13. PubMed PMID: 35895583; PubMed Central PMCID: PMC9393435.

41. Mungwira RG, Laurens MB, Nyangulu W, Divala TH, Nampota-Nkomba N, Buchwald AG, Nyirenda OM, Mwinjiwa E, Kanjala M, Galileya LT, Earland DE, **Adams M**, Plowe CV, Taylor TE, Mallewa J, van Oosterhout JJ, Laufer MK. High Burden of Malaria Among Malawian Adults on Antiretroviral Therapy After Discontinuing Prophylaxis. AIDS. 2022Oct1;36(12):1675-1682.PubMed PMID: 35848575; PubMed Central PMCID: PMC9444947.

42. Nyangulu W, Mungwira RG, Divala TH, Nampota-Nkomba N, Nyirenda OM, Buchwald AG, Miller J, Earland DE, **Adams M**, Plowe CV, Taylor TE, Mallewa JE, van Oosterhout JJ, Parikh S, Laurens MB, Laufer MK. Artemether-Lumefantrine Efficacy Among Adults on Antiretroviral Therapy in Malawi. Malar J. 2023 Jan 27;22(1):32. PubMed PMID: 36707795; PubMed Central PMCID: PMC9881508.

43. Lyke KE, Berry AA, Mason K, Idris AH, O'Callahan M, Happe M, Strom L, Berkowitz NM, Guech M, Hu Z, Castro M, Basappa M, Wang L, Low K, Holman LA, Mendoza F, Gordon IJ, Plummer SH, Trofymenko O, Strauss KS, Joshi S, Shrestha B, **Adams M**, Chagas AC, Murphy JR, Stein J, Hickman S, McDougal A, Lin B, Narpala SR, Vazquez S, Serebryannyy L, McDermott A, Gaudinski MR, Capparelli EV, Coates EE, Wu RL, Ledgerwood JE, Dropulic LK, Seder RA. Low-Dose Intravenous and Subcutaneous CIS43LS Monoclonal Antibody For Protection Against Malaria (VRC 612 Part C): A Phase 1, Adaptive Trial. Lancet Infect Dis. 2023 May;23(5):578-588. Epub 2023 Jan 25. PubMed PMID: 36708738; PubMed Central PMCID: PMC10121890.

44. Tewey MA, Coulibaly D, Lawton JG, Stucke EM, Zhou AE, Berry AA, Bailey JA, Pike A, Dara A, Ouattara A, Lyke KE, Ifeonu O, Laurens MB, **Adams M**, Takala-Harrison S, Niangaly A, Kouriba B, Koné AK, Rowe JA, Doumbo OK, Patel JJ, Tan JC, Felgner PL, Plowe CV, Thera MA, Travassos MA. Natural Immunity to Malaria Preferentially Targets the Endothelial Protein C Receptor-Binding Regions of PFEMP1s. mSphere. 2023 Oct 24;8(5):e0045123. Epub 2023 Oct 4. PubMed PMID: 37791774; PubMed Central PMCID: PMC10597466.

45. Aung PP, Han KT, Groot W, Biesma R, Thein ZW, Htay T, Lin Z, Aye KH, **Adams M**, Pavlova M. Heterogeneity in the Prevalence of Subclinical Malaria, Other Co-infections and Anemia Among Pregnant Women in Rural Areas of Myanmar: A Community-Based Longitudinal Study. Trop Med Health. 2024 Mar 8;52(1):22. PubMed PMID: 38459581; PubMed Central PMCID: PMC10921590.

46. Vareta J, Horstman NA, **Adams M**, Seydel KB, McCann RS, Cohee LM, Laufer MK, Takala-Harrison S. Genotyping *Plasmodium falciparum* gametocytes using amplicon deep sequencing. Malar J. 2024 Apr 6;23(1):96. PubMed PMID: 38582837; PubMed Central PMCID: PMC10999092.

47. Chou RT, Ouattara A, **Adams M**, Berry AA, Takala-Harrison S, Cummings MP. Positive-unlabeled Learning Identifies Vaccine Candidate Antigens in the Malaria Parasite *Plasmodium falciparum*. NPJ Syst Biol Appl. 2024 Apr 27;10(1):44. PubMed PMID: 38678051; PubMed Central PMCID: PMC11055854.

48. Bailey RD, Lawton JG, Niangaly A, Stucke EM, Bailey JA, Berry AA, Ouattara A, Coulibaly D, Lyke KE, Laurens MB, Zhou AE, Pablo J, Jasinskas A, Nakajima R, **Adams M**, Takala-Harrison S, Kouriba B, Kone AK, Guindo A, Rowe JA, Diallo DA, Doumbo OK, Felgner PL, Plowe CV, Thera MA, Travassos MA. Children with Hemoglobin C Or S Trait Have Low Serologic Responses to a Subset of Malaria Variant Surface Antigens. J Infect. 2024 Oct;89(4):106257. Epub 2024 Aug 30. PubMed PMID: 39216830; PubMed Central PMCID: PMC11546587.

49. Ouattara A, Andronescu LR, **Adams M**, Sharma A, Nakajima R, Jain A, Taghavian O, Jasinskas A, Felgner PL, Mathanga DP, Chinkhumba J, Laufer MK, Travassos MA. Choice of Drug for Malaria Prevention During Pregnancy Does Not Affect Infant Serologic Responses to *Plasmodium falciparum* Erythrocyte Membrane Proteins 1. Open Forum Infect Dis. 2025 Feb;12(2):ofaf037. eCollection 2025 Feb. PubMed PMID: 39963702; PubMed Central PMCID: PMC11832042.

Abstract

Title of Dissertation: Identifying Correlates of Protective Immunity and Antigenic Escape in *Plasmodium falciparum* Malaria

Matthew Adams, Doctor of Philosophy, 2025

Dissertation Directed by: Dr. Shannon Takala Harrison, PhD, Professor, University of Maryland, Baltimore School of Medicine

Malaria remains a leading cause of illness and death worldwide, particularly in sub-Saharan Africa, which bears the majority of malaria-related morbidity and mortality. While first-generation malaria vaccines represent a significant advance, their modest and time-limited efficacy underscores the need for improved vaccine strategies. Progress is impeded by the complex immunological landscape governing protection against *Plasmodium falciparum* infection and disease, as well as the parasite's extensive antigenic diversity which has evolved to evade host immunity through antigenic escape. In this study, I applied high-throughput serological profiling to identify correlates of protection following whole-organism vaccination and implemented temporally informed population genetic analyses to detect immune-driven antigenic escape.

Proteome-wide IgG reactivity (>3.9 million 16-mers) was measured using sera collected before and after PfSPZ vaccination from individuals later challenged by controlled human malaria infection. Although vaccination induced broad serological responses across more than 4,000 proteins, few peptide-level responses were associated with protection. Most differentially reactive peptides were elevated in unprotected individuals both before and after vaccination. Pre-vaccination differences may reflect preexisting cross-reactive antibodies. After vaccination, protection-associated targets were enriched for membrane-localized proteins, while peptides elevated in unprotected

individuals were linked to nuclear processes. These findings suggest that linear peptide reactivity alone may be insufficient to define correlates of protection and that additional immune mechanisms are likely to contribute.

To investigate antigenic escape, I analyzed whole-genome sequence data from clinical isolates collected over 15 years from symptomatic infections in southern Malawi. Using Tajima's D and Hudson's F_{ST} across sliding genomic windows, I identified 21 genes showing consistent signals of balancing selection and temporal allele frequency shifts. One identified gene, PF3D7_0710200, encoding a conserved protein of unknown function, underwent further analysis using both supervised and unsupervised machine learning to identify key amino acid residues and haplotype clusters. Random forest analysis identified nine informative amino acid residues that delineated four major haplotypes that remained stable over time, with three accounting for over 90% of circulating diversity.

These findings support the integration of immunological and evolutionary approaches to inform vaccine design and identify promising targets for broadly protective malaria vaccines.

Identifying Correlates of Protective Immunity and Antigenic Escape
in *Plasmodium falciparum* Malaria

by
Matthew Adams

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, Baltimore in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2025

To my mother and father

You instilled in me the habits of listening with care, questioning without fear, and reasoning with depth. From your example, I learned to place thoughtfulness above ease, and inquiry above certainty.

To my wife

When I was depleted, you helped me stay upright. Your steadiness turned endurance into progress when progress felt impossible.

To Amed

As a longtime friend and one of the few people with whom I could have in-depth conversations about science, I'm grateful for the honesty of our inquiry and the clarity it brought.

Acknowledgements

I would like to thank my dissertation committee for their time, expertise, and guidance throughout this process. Dr. Shannon Takala Harrison, my chair, provided patient oversight during each stage of the work. I am also grateful to Dr. Andrea Berry, Dr. Michael Cummings, Dr. Miriam Laufer, Dr. Helen Powell, and Dr. Joana Carneiro da Silva for their thoughtful feedback and contributions. Their perspectives helped shape this dissertation and its direction, and I especially appreciated moments of intellectual curiosity, clarity, and open dialogue.

To my graduate school cohort, Dr. Richard Smith, Dr. Kareshma Mohanty, and Dr. Sarah Robbins, thank you for finishing before me and reminding me that everyone moves at their own pace. Your friendship and example meant more than you probably realized. Thanks as well to the Graduate Student Forum, past and present, whose collective wisdom, generosity, and camaraderie supported early ideas, stubborn problems, and final presentations.

This work benefited from substantial technical guidance and collaborative problem solving that shaped both the analysis and the thinking behind it. I am especially grateful to Mr. Roland Laboulaye and Mr. Jean-Paul Courneya, whose programming support and many conversations helped me grow as both an analyst and a thinker. Many individuals were responsible for the fieldwork and laboratory efforts behind the studies analyzed here. Their contributions formed the foundation of this research, and I am deeply indebted to their work.

Finally, I acknowledge and thank the study participants and their families, without whom this work could not have been conducted.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AnAPN1	Anopheline Alanyl Aminopeptidase N 1
AUC-ROC	Area Under the Receiver Operator Characteristic Curve
BP	Biological Process
CC study	Chloroquine Combination Randomized Clinical Trial
CC (Gene Ontology)	Cellular Components
CDF	Cumulative Density Function
CHMI	Controlled Human Malaria Infection
cM	Centimorgan
DNA	Deoxyribonucleic Acid
EBA	Erythrocyte-binding Antigen
FDR	False Discovery Rate
FI	Fluorescence Intensity
F_{ST}	Fixation Index
F_{WS}	Within-host Diversity Statistic
GATK	Genome Analysis Toolkit
GO	Gene Ontology
GPI	Glycosylphosphatidylinositol
GVCF	Genomic Variant Call Format
ICEMR	International Center for Excellence in Malaria Research
IFNγ	Interferon Gamma
Ig(A/D/E/G/M)	Immunoglobulin Types A, D, E, G, or M
IL-12	Interleukin-12

LARC	Late-arresting Replication-competent
MCS	Mfera Cohort Study
MF	Molecular Function
MOI	Multiplicity of Infection
nsSNP	Nonsynonymous Single Nucleotide Polymorphism
OOB	Out-of-bag
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
PfAMA1	*Pf* Apical Membrane Antigen 1
pfcrt	*Pf* Chloroquine Resistance Transporter gene
PfCSP	*Pf* Circumsporozoite Protein
PfEBA-175	*Pf* Erythrocyte Binding Antigen-175
PfEMP1	*Pf* Erythrocyte Membrane Protein 1
PfGAP3KO vaccine	*Pf* Genetically Attenuated Parasite with Three Knockouts vaccine
PfGARP	*Pf* Glutamic-acid-rich Protein
PfGLURP	*Pf* Glutamate-rich Protein
PfLARC vaccine	*Pf* Late-Arresting Replication-Competent vaccine
pfmdr1	*Pf* Multidrug Resistance Protein1gene
PfMSP1	*Pf* Merozoite Surface Protein 1
PfMSP3	*Pf* Merozoite Surface Protein 3
PfNUP390	*Pf* Nucleoporin NUP390, putative
PfRH3	*Pf* Reticulocyte-binding Protein Homolog 3
PfRH5	*Pf* Reticulocyte-binding Protein Homolog 5

PfRON2	*Pf* Rhoptery Neck Protein 2
PfRON3	*Pf* Rhoptry Neck Protein 3
Pfs230	*Pf* 6-cysteine Protein P230
Pfs25	*Pf* Ookinete Surface Protein P25
Pfs28	*Pf* Ookinete Surface Protein P28
Pfs48/45	*Pf* 6-cysteine Protein P48/45
PfSEA1	*Pf* Schizont Egress Antigen-1
PfSERA5	*Pf* Serine Repeat Antigen 5
PfSPZ	*Pf* Radiation-attenuated Sporozoite Vaccine
PfSPZ-CVAC	*Pf* Chemoprophylaxis-attenuated Sporozoite Vaccine
PfSPZ-GA1	*Pf* Genetically Attenuated Sporozoite 1 vaccine
PfSPZ-LARC2	*Pf* Late-Arresting Replication-Competent Sporozoite 2 Vaccine
PfTRAP	Thrombospondin-related Anonymous Protein
qPCR	Quantitative Polymerase Chain Reaction
Rh	Reticulocyte-binding Homologue
RNA	Ribonucleic Acid
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SD	Standard Deviation
SNP	Single Nucleotide Polymorphism
SPZ	Sporozoites
sWGA	Selective Whole Genome Amplification
TNF-α	Tumor Necrosis Factor-α
TSR	Thrombospondin-like Type I Repeat

UMAP	Uniform Manifold Approximation and Projection
WGS	Whole Genome Sequencing
WHO	World Health Organization

Chapter 1

Introduction and Specific Aims

1.1 **Background and Introduction**

Malaria is a parasitic disease that, each year, infects hundreds of millions of people across the world and, in 2023 alone, resulted in an estimated 597,000 deaths [1]. Africa carries most of the malaria burden, including 94% of malaria cases and 95% of malaria deaths worldwide. Through strengthened control and elimination efforts, malaria mortality has decreased over the past two decades. However, progress has stalled since 2015, leaving hundreds of thousands of dead each year. This plateau in malaria cases and deaths highlights the need for additional tools to control, eliminate, and eventually eradicate malaria.

A vaccine against *Plasmodium falciparum*, the deadliest human malaria species, would prevent deaths and accelerate eradication efforts [2]. Since the first report of human protection from malaria by vaccination in 1973 [3, 4], decades of research and development have culminated in the recent approval of two vaccines by the World Health Organization (WHO): RTS,S/AS01 and R21/Matrix-M. Both are subunit vaccines targeting the Circumsporozoite Protein (PfCSP) [1]. While the approval of these vaccines represents a major milestone, their efficacy against clinical malaria and asymptomatic infection is modest [5-7]. Therefore, improving upon these first-generation vaccines remains a critical objective to further reduce malaria morbidity and mortality and to advance toward eradication. To date, malaria subunit vaccine development efforts have concentrated on a small number of well-characterized proteins [8]. With over 5,300 genes in the *P. falciparum* genome [9], with different genes expressed in each of multiple life cycle stages, identification of vaccine candidate antigens has been challenging and has been further complicated by extensive genetic diversity of antigens, driven in part by selection pressures from host immune responses [10].

In malaria endemic regions, repeated infection with malaria parasites leads to the development of non-sterilizing immunity characterized by the presence of circulating parasites without clinically apparent symptoms i.e., asymptomatic or subclinical infection [11, 12]. *P. falciparum* antigens exhibit high genetic diversity, resulting in multiple antigenic variants that allow the parasite to escape recognition by the host immune system [13]. Over the course of repeated infections, individuals acquire a repertoire of anti-parasite immune responses, some of which are directed against specific variants of a given protein [14]. After many exposures, an accumulation of these variant-specific immune responses against multiple antigens is thought to contribute to protection against clinical malaria [15]. As variant-specific immune responses are not equally protective against heterologous variants, the polymorphism of parasite antigens allows for antigenic escape, a phenomenon that has been observed in the context of naturally-acquired immunity, and which has limited the efficacy of several promising vaccine candidates in field trials [16-19]. As such, understanding the potential for antigenic escape may aid in the development of a broadly protective malaria subunit vaccine.

The primary objectives of this research were to identify correlates of vaccine-derived protection against *P. falciparum* malaria and to identify parasite loci potentially involved in antigenic escape by detecting evolutionary signatures of balancing selection, a genetic signature strongly associated with immune-driven maintenance of antigenic diversity, using samples collected during three studies conducted in southern Malawi between 2007 and 2022.

1.1.1 **Malaria Epidemiology**

Derived from the Italian phrase *mala aria*, meaning "bad air," malaria is not caused from breathing noxious air from marshes as originally suspected in the 1800s, but rather is caused by infection with a mosquito-borne protozoan of the genus [20]. There are five species of *Plasmodium* that infect humans, namely, *Plasmodium falciparum*, *Plasmodium knowlesi*, *Plasmodium malariae*, *Plasmodium ovale*, and *Plasmodium vivax.* Malaria is endemic in sub-Saharan Africa, Central and South America, Asia, and Oceania [21]. The World Health Organization estimated that in 2023, 263 million cases of malaria occurred globally resulting in an estimated 597,000 deaths. Sub-Saharan Africa was disproportionately burdened with 246 million cases and 569,000 deaths, with 73.7% of malaria deaths befell children aged under five years [1]. In sub-Saharan Africa, *P. falciparum* is the dominant species of human-infecting *Plasmodium*. Malaria manifests clinically as flu-like symptoms, including paroxysmal fever, headache, body aches, malaise, and fatigue [22]. The cyclic symptoms tend to coincide with acute spikes in parasitemia that occur as a new generation of parasites is released into the bloodstream [22]. Individuals develop partial immunity after numerous *P. falciparum* infections and exposure to a large spectrum of parasite antigens, which reduces the severity of clinical symptoms but does not protect against future infection [23]. Over time, with frequent and repeated exposure to a large range of *P. falciparum* antigens, acquired immunity may prevent clinical malaria entirely resulting in asymptomatic or subclinical infection [24].

1.1.2 **Lifecycle**

For *P. falciparum*, the causative agent of the most lethal form of malaria [1], the complex life cycle requires a female *Anopheles* mosquito, the definitive host, and humans, the intermediate host [25]. Through the bite of an infective *Anopheles* mosquito, haploid sporozoites are inoculated into the human skin layer, where they quickly migrate to enter the blood stream [26]. Sporozoites that are able to reach the bloodstream then travel to the liver invading hepatocytes where the exoerythrocytic amplification cycle begins [27]. During this 6-to-14-day period, sporozoites mature into hepatic schizonts that contain thousands of merozoites. Upon maturation, the schizonts rupture and release merozoites into the bloodstream. These merozoites invade red blood cells, beginning the erythrocytic cycle. In the red cell, the parasite develops into a ring stage, then a mature trophozoite, and finally into a schizont that contains multiple merozoites. The schizont ruptures and releases merozoites into the blood stream and this process of asexual amplification repeats, greatly increasing the number of parasites in the blood. A subset of parasites will commit to differentiate from asexual forms to gametocytes which are male and female sexual forms. Mature gametocytes in the human blood stream are ingested by a female *Anopheles* mosquito during a blood meal [27]. In the midgut of the mosquito, male gametocytes undergo exflagellation to form eight male gametes, which are then available to fertilize a female gamete [28]. The male and female gametes fuse and generate a zygote which develops into the ookinete, a motile and elongated diploid form that invades the mosquito's mid-gut wall and develops into an oocyst, which produces thousands of haploid sporozoites that migrate to the salivary glands of the mosquito where they can infect another human host during a blood meal [27].

1.1.3 **Malaria Vaccinology**

Effective antimalarial treatment, vector control and diagnostics have been historic pillars of malaria control [29]. While great strides have been made to control malaria, diagnostic-resistant [30, 31] and multidrug-resistant [32] malaria parasites, along with insecticide-resistant mosquitoes [33], emphasize the need to approach elimination from multiple fronts. An effective vaccine would be a valuable and sustainable component of this multi-pronged approach [34]; indeed, to date, no infectious disease has been eradicated in the absence of a vaccine [2]. However, the development of an effective malaria vaccine faces multiple challenges. These challenges include the complex life cycle of *P. falciparum* with multiple life stages, each presenting different antigens, an incomplete understanding of protective immunity, and the extensive antigenic diversity that is thought to have evolved to avoid host immune defenses [10, 35, 36]. This diversity has implications for malaria interventions, including development of a broadly efficacious vaccine.

Pre-erythrocytic vaccines target the early stages of *P. falciparum* infection in the human host [37]. These vaccines induce immune responses against sporozoites and liver-stage parasites. The historic goal of pre-erythrocytic vaccines was to prevent the establishment of infection by blocking sporozoite passage to the liver and/or hepatocyte invasion and development. This 'sterilizing' immunity has the added benefit of preventing disease and transmission, making this approach attractive both to prevent deaths, but also to reduce the burden of infection and disease. However, the two World Health Organization (WHO)-approved vaccines, both of which are pre-erythrocytic vaccines, have not been shown to induce sterilizing immunity [1, 4]. In October 2021, the WHO recommended RTS,S/AS01 for the use in children aged 5 months and above living in regions of moderate to high transmission for the prevention of *P. falciparum* malaria and the reduction of malaria burden [38]. However, protective efficacy against clinical malaria as measured in a large Phase 3 trial in children aged 5-17 months ranged from 11-32% and in children aged 6-12 weeks was ranged from 23-41% [7], underscoring the need for more effective and durable vaccines [39]. Another promising pre-erythrocytic subunit vaccine, R21/Matrix-M™, was approved by WHO and is similar to RTS,S/AS01 in that both vaccines are based on the C-terminus and central repeats of the *P. falciparum* circumsporozoite protein fused with either a single moiety of hepatitis B antigen (R21/Matrix-M) or four moieties of hepatitis B antigens (RTS,S/AS01) [40, 41]. A randomized controlled trial of R21/Matrix-M in Burkinabe children aged 5-17 months, reported an efficacy of 77% (95% CI: 67-84) [6]. Although this estimate is higher than that observed in RTS,S trials, differences in trial design and transmission intensity of the trial sites make direct comparison challenging. Importantly, both vaccines provide only partial protection, and efficacy wanes over time. Additionally, incomplete blockage of sporozoite development by either vaccine could lead to active blood stage replication that might partially negate the sought-after infection- and transmission-blocking effects of pre-erythrocytic vaccines. Nonetheless, the potential of these pre-erythrocytic vaccines to reduce the burden of malaria disease and death is both groundbreaking and cause for optimism; novel vaccines with higher and longer-lasting efficacy against clinical malaria, ideally in tandem with infection- and transmission-blocking efficacy, may be required to achieve the goal of malaria eradication [42].

By combining antigens that protect against multiple parasite life-stages, the likelihood of vaccine escape may be decreased [43]. Erythrocytic or blood stage vaccines, as the name suggests, target the parasite after release from the liver into the blood stream. As malaria symptoms develop during the erythrocytic stage coinciding with cyclic increases in parasitemia, vaccines could attempt to ameliorate symptoms by reducing parasitemia [44]. Antibodies against erythrocytic parasite antigens may function in multiple ways, such as inhibition of merozoite invasion in the red blood cell, and inhibition of parasite egress from the red blood cell, among others. A well-studied erythrocytic antigen that is involved in merozoite invasion, Apical Membrane Antigen 1 (PfAMA1), has been incorporated as a component of multiple subunit vaccines. In early clinical trials of PfAMA1-based vaccines, efficacy was not observed [45] or was limited and variant-specific [46]. However, novel vaccines incorporating PfAMA1 in a complex, usually with Rhoptry Neck Protein 2 (PfRON2), have shown promise in mouse models [47]. Another promising erythrocytic antigen, Reticulocyte-binding Protein Homolog 5 (PfRH5), has been reported to be essential for erythrocyte invasion into the blood cell [48], and a vaccine containing PfRH5 has been shown to be safe, immunogenic, and delay symptoms after blood-stage challenge [49]. Other antigens considered for erythrocytic vaccines inhibit non-invasion cellular mechanisms with a similar goal to decrease parasite proliferation. For example, the inhibition of Schizont Egress Antigen-1 (PfSEA1) prevents merozoite egress during schizont rupture resulting in decreased parasite replication [50, 51]. Similarly, in vitro antibody inhibition of Glutamic-acid-rich Protein (PfGARP) induces programmed cell death [50]. A subset of erythrocytic candidate vaccines aim to protect against specific manifestations of clinical malaria. One such vaccine is based on the Erythrocyte Membrane Protein 1 (PfEMP1) variant surface antigen, VAR2CSA, which aims to reduce virulence by preventing sequestration in the placenta during pregnancy [52].

Unlike pre-erythrocytic and erythrocytic vaccines, transmission blocking vaccines do not directly prevent disease [53]. Rather, transmission blocking vaccines interrupt the transmission of gametocytes to mosquitoes or their development into sporozoites, preventing spread to community members. In 1976, Gwadz et al. demonstrated that infectivity of the *Plasmodium gallinaceum* was reduced in malaria-infected chickens that had previously received three weekly doses of formalin-treated or irradiated parasite-infected blood by intravenous injection [54]. Later that same year, Carter and Chen improved on these results by enriching for gametocytes prior to irradiation resulting in a 99.9% decrease in infectivity compared to unvaccinated chickens [55]. Recent approaches to develop transmission blocking vaccines have moved away from whole organism approaches towards subunit vaccines with leading subunit candidate antigens such as Ookinete Surface Protein P25 (Pfs25) [56], Ookinete Surface Protein P28 (Pfs28) [57], 6-cysteine Protein P230 (Pfs230) [58], 6-cysteine Protein P48/45 (Pfs48/45) [59], and the *Anopheles* protein Anopheline Alanyl Aminopeptidase N 1 (AnAPN1) [60]. One advantage of transmission blocking vaccines is that candidate antigens tend to be less polymorphic than erythrocytic or pre-erythrocytic antigens, reducing the likelihood of antigenic escape [61]. Like erythrocytic vaccines, transmission blocking vaccines target forms of the parasite that exist in a chokepoint of the lifecycle when the number of parasites is relatively small [61]. However, evaluating efficacy of transmission blocking vaccines is challenging due to the requirement to assess a community level outcome, resulting in the use of surrogate assays that show transmission inhibition via *ex vivo* membrane feeding assays [62]. The extent to which these assays reflect community-level transmission reduction in endemic settings is unknown.

Attenuated whole-organism vaccines have demonstrated protective efficacy in challenge studies in malaria-naïve individuals [63]. One method of attenuation, irradiation, was based on early experiments by Nussenzweig et al. that reported protection in a rodent malaria model after administration of sporozoites from irradiated *Anopheles stephensi* mosquitos infected with *Plasmodium berghei* [64]. Further proof-of-concept of the protective efficacy of irradiated sporozoites was performed by Clyde et al. [3, 65-68] and Rieckmann et al. [69, 70] demonstrating protection from infectious mosquito challenge after more than one thousand irradiated mosquito bites [63]. An updated version of radiation attenuated sporozoites, developed by the U.S.-based biotechnology company Sanaria, obviates the need for mosquito-bite administration by isolating sporozoites from irradiated mosquitoes by dissection and cryopreserving these sporozoites for future vaccination [63, 71]. The Sanaria® PfSPZ Vaccine was evaluated in Phase 1 clinical trials by intravenous administration to volunteers resulting in 100% protection against homologous malaria infection in the group receiving the highest does of irradiated sporozoites [72]. Additional trials in naïve volunteers confirmed this protective efficacy enabling progression to Phase 2 clinical trials in endemic regions [73, 74]. However, the same high level of protection in malaria-naïve individuals was not maintained in clinical trials performed in endemic areas of Mali [75] and Tanzania [76]. Despite limited efficacy, a Phase 3 trial of the PFSPZ Vaccine was scheduled to occur in 2021 but was delayed by the SARS-CoV-2 pandemic. Other promising whole organism vaccines based on attenuated parasites include chemoprophylaxis-attenuated sporozoites (PfSPZ-CVac) [77], early-arresting genetically attenuated sporozoites (PfSPZ-GA1 and PfGAP3KO) [78, 79], late-stage genetically attenuated sporozoites (PfLARC and PfSPZ-LARC2) [80, 81], and chemically attenuated asexual blood-stage parasites [82, 83]. Radiation-attenuated and chemoprophylaxis-attenuated sporozoites have shown promise when administered to malaria naïve volunteers, with each vaccine generating protection after challenge with infective mosquitoes or direct venous injection of sporozoites [72-74, 77]. In endemic settings, the vaccine efficacy of both the radiation-attenuated and chemo-attenuated PfSPZ vaccines has been modest in natural [75] and controlled challenge trials [84-86]. However, in a study of Malian women who became pregnant within 24 weeks after vaccination, the PfSPZ Vaccine demonstrated up to 65% efficacy against infection [87]. The two early-stage genetically attenuated vaccines were evaluated in humans with PfSPZ-GA1 protecting only 3 of 25 malaria-naïve volunteers from infection after challenge [78] and PfGAP3KO protecting 10 of 10 malaria-naïve volunteers after challenge, though without infectivity or placebo controls [88]. Late-arresting Replication-competent (LARC) genetically attenuated parasites, such as PfLARC, PfSPZ-LARC2, and PfGA2-MB (PfΔmei2) are designed to arrest development late in the liver stage, thereby exposing the host to a broader repertoire of parasite antigens. In a Phase 1 clinical trial evaluating the efficacy of the PfGA2-MB vaccine delivered by mosquito bite, 90% of vaccinated volunteers were protected following controlled human malaria infection, compared to 0% protection in the placebo group [89].

A novel whole organism blood stage vaccine employs chemical attenuation using tafuramycin-A to prevent asexual replication and produce protective immunity in rodent models [82]. Additional approaches have used whole blood stage parasites with mannose presenting liposomes to present to immune cells, abrogating the potential for self-recognition of parasitized red blood cells [83].

PfCSP is the primary antigen used in both currently approved malaria vaccines, RTS,S/AS01 and R21/Matrix-M [90]. PfCSP is one of the most abundant surface proteins on *P. falciparum* sporozoites [91], playing important roles in the parasite's life cycle within the mosquito and human hosts [92, 93]. Structurally, PfCSP is composed of three distinct regions: an N-terminus domain containing the signal sequence and a conserved KLKQP motif, a central repeat region characterized by NPNA-repeating tetrapeptides, and a C-terminus domain containing a thrombospondin-like type I repeat (TSR) domain and the glycosylphosphatidylinositol (GPI)-anchor [94, 95]. In mosquitoes, PfCSP functions include sporozoite development, release of sporozoites from the oocyst, and subsequent sporozoite migration to the mosquito salivary glands [96]. The N-terminus of PfCSP is necessary for mosquito salivary gland localization, with its cleavage reducing sporozoite infectivity [92]. In humans, PfCSP is involved in sporozoite motility, host cell traversal, and the subsequent invasion into liver cells, establishing the liver-stage of infection [92, 97]. During migration from the site of infection, the full-length form of PfCSP conceals the C-terminus domain, which maintains the sporozoite in a migratory state [92]. The proteolytic cleavage of PfCSP releases the C-terminus domain and exposes the adhesive N-terminus domain, enabling hepatocyte invasion and establishing the liver-stage of infection [98]. PfCSP is highly immunogenic [99]. Monoclonal antibodies that target the central repeat region and the junctional region of the N-terminus with the central repeat region have shown high protective efficacy in clinical trials [100-103]. Though the NPNA-repeat motif is a constant epitope targeted by many of these monoclonals, several have dual-binding capacity in which they recognize junctional region motifs as well [104-106]. In fact, CIS43 monoclonal antibodies preferentially bind the DPNA-containing motifs, while L9 antibodies preferentially bind NPNV-containing motifs with higher affinity than NPNA-repeats alone [107]. The C-terminal TSR domain also contains epitopes that are recognized by antibodies and T cells, although these are less accessible due to the structural conformation of PfCSP, which can mask these sites from immune detection [92]. This structural diversity has implications for vaccine design and the development of monoclonal antibodies for passive immunity.

While PfCSP is a major target of both subunit vaccines and monoclonal antibodies, its utility as a vaccine antigen is complicated by extensive genetic diversity. For instance, an analysis of global *pfcsp* sequences revealed 138 different haplotypes in the C-terminal region, with only a small fraction matching the 3D7 strain used in the RTS,S and R21-Matrix-M vaccines [108]. This high level of polymorphism, particularly in immunogenic regions, suggests that vaccine-elicited immune responses may be strain-specific and less effective against heterologous parasites [109]. Indeed, allele-specific efficacy has been observed for RTS,S, with protection biased toward parasites carrying *pfcsp* sequences matching the vaccine strain [18]. These findings underscore the potential for immune escape and highlight the need to identify additional antigens under balancing selection or subject to vaccine-induced selection pressure.

1.1.4 **Controlled Human Malaria Infection (CHMI)**

Early clinical evaluation of novel vaccines or antimalarial drugs involved the intentional inoculation of study volunteers with malaria parasites, either by injecting live sporozoites [110] or malaria infected erythrocytes [111], or allowing infective mosquitoes to feed on volunteers [112], followed by monitoring and curative treatment. CHMI studies allow for a well-controlled, early evaluation of vaccine safety, immunogenicity, and efficacy that permits promising vaccines to continue to field trials in endemic regions [111]. Due to their established safety in non-endemic settings, CHMI studies are increasingly being conducted in low- and middle-income countries, where they allow for the evaluation of interventions in populations naturally exposed to malaria [113]. The general process of evaluating a vaccine in a Phase 1 clinical trial with controlled human malaria infection (CHMI) begins with the recruitment of malaria-naïve volunteers. Administration of the vaccine initiates the follow-up period for monitoring adverse events, after which participants undergo malaria challenge via exposure to infective mosquitoes or injection with replicating parasites. Additionally, volunteers who were not administered the vaccine are exposed to the infective mosquito/replicating parasites to confirm the infectivity of the challenge. All challenged volunteers are monitored closely for 3 - 4 weeks to detect the development of symptoms or the presence of parasites in the blood, either by microscopy or molecular methods [114]. Volunteers who fail to develop detectable parasitemia are considered protected, while volunteers who develop parasitemia are administered curative drug therapy and monitored for safety.

1.1.5 **Immunology of Malaria Infections**

The human immune system mounts a complex and multifaceted response to infection with *Plasmodium parasites*. The effectors of the immune system against malaria can be broadly categorized into two interrelated components: innate and adaptive, with the adaptive response further (imperfectly) classified as cellular and humoral immune effector responses [115, 116]. The innate immune system acts as the first line of defense, rapidly recognizing and responding to *Plasmodium* infection [117]. When sporozoites of *P. falciparum* enter hepatocytes, they induce a Type I interferon response driven by the detection of parasitic RNA, which attracts natural killer and natural killer T cells to the liver [118]. Liver-stage innate immune responses to *Plasmodium* are limited, primarily involving cytosolic sensors in hepatocytes that detect parasite RNA and trigger type I interferon production [117]. In contrast, blood-stage infection elicits a more robust innate response, characterized by rapid production of pro-inflammatory cytokines (including type I interferons, TNF-α, and IL-12) by dendritic cells and macrophages [118]. This blood-stage response also involves activation of pattern recognition receptors and increased phagocytic activity, collectively aiming to control parasite replication and shape subsequent adaptive immunity [118]. The liver-stage cellular immune response against *P. falciparum* involves CD8+ T cells, including liver-resident CD8+ T cells, which patrol hepatic sinuses and contribute to protective immunity against pre-erythrocytic stages [119, 120]. The blood-stage cellular response engages various T cell subsets, with CD4+ T cells playing a role in parasite control and providing help for antibody production [121]. The balance and timing of these innate and cellular responses, along with their interactions with humoral immunity, influence the course of infection and the development of protective immunity against malaria.

Humoral immunity, characterized by antibody responses, is important for limiting blood-stage parasite infections and reducing disease severity. Antibodies or immunoglobulins (Ig) are crucial components of the adaptive immune system and are classified into five main isotypes: IgG, IgA, IgM, IgD, and IgE [122]. IgG is the most abundant isotype in serum and is important for neutralization [123, 124], opsonic phagocytosis [125, 126], and complement activation [127] responses against the parasite. It is the only antibody capable of crossing the placenta during a healthy pregnancy, providing passive immunity to the fetus [128]. Antibody isotypes can be further classified into subclasses or subtypes that exhibit distinct structural and functional characteristics. IgG subtypes include IgG1, IgG2, IgG3, and IgG4, which differ in their ability to activate complement, bind to Fc receptors, and mediate effector function [129, 130]. IgG1 and IgG3 participate in complement activation and antibody-dependent cell-mediated cytotoxicity [131-133]. Increased risk for malaria has been correlated with IgG2 [129, 134] and IgG4 [135] and both subtypes have been correlated with severe malaria [136]. Though, both IgG2 [135] and IgG4 [129] have also been associated with protection. The variation in distribution and effect of malaria antibody subclasses may stem from differences in human individuals, parasite life stages, or specific antigens, but the precise factors determining this variation remain unclear [137, 138].

1.1.6 **Correlates of Protection**

Stanley Plotkin, a distinguished vaccinologist, defines a "correlate of protection" as an immune response that both causes and is statistically associated with protection [139]. He further clarifies this concept by differentiating it from related but distinct terms. An "absolute correlate" refers to a threshold level of immune response highly correlated with protection. A "relative correlate" describes an immune response with a variable level of correlation to protection. "Co-correlates" are multiple factors that may interact in alternative, additive, or synergistic ways to correlate with protection. Lastly, a "surrogate" represents an immune response that substitutes for the actual immunologic correlate of protection, which may remain unidentified [139]. Identifying immune correlates of protection would accelerate vaccine discovery and could also identify individuals or populations who, after vaccination, developed insufficient protective responses. Although, correlates of protection against malaria have been identified [140, 141], they remain unvalidated and thus robustness of these correlates of protection for purposes such a vaccine testing remains limited [141, 142]. One reason for the difficulty in establishing correlates of protection is the lack of a well-defined protective phenotype. Multiple phenotypes of protection have been established including protection against symptomatic malaria (but not infection), protection against infection, and transmission-blocking protection [15]. Adding to the complexity is that vaccine-induced immune responses are often multifactorial, combining innate, cell-mediated, and humoral components of the immune system to generate protection [139] and even then, protective responses may be individual dependent [34, 143]. As the PfSPZ vaccine demonstrated high levels of sterile protection in malaria-naïve North Americans after challenge, this clear and well-defined phenotype may enable detection of a correlate of protection that could be useful in evaluating efficacy in future clinical trials. Identifying correlates of protection also presents an opportunity to discover potential vaccine-induced responses or even pre-immunization immunological conditions that are associated with protection. For example, some hemoglobinopathies have been associated with adaptive immune responses and protection from malaria [144, 145] and therefore might have a detectable impact on the humoral response to vaccination. By identifying the antibody responses associated with protection, the mechanisms underlying that response can be interrogated and may help to guide the design of more effective vaccines.

1.1.7 **PfSPZ Vaccine Mechanisms of Protection**

Although the mechanism of protection conferred by the PfSPZ vaccine has been well studied, much remains unknown, especially in humans. Pre-erythrocytic immunity has traditionally been thought to require T cell mediated protection via liver resident T cells that eliminate *Plasmodium*-infected hepatocytes [146]. Mouse models have established a role in pre-erythrocytic immunity for several cellular immunological components, specifically, CD8+, IFNγ-producing T cells, and γδT cells [147]. Studies in non-human primates have shown that protective immunity was established after administration of irradiated *Plasmodium knowlesi* sporozoites; however, this protective effect was abrogated after administering an antibody against CD8 lymphocytes and protection was returned after the waning of anti-CD8 antibodies, suggesting that CD8 lymphocytes were necessary for protection [148]. The extent to which these non-human malaria models represent human immunological protection and the extent to which other immunological effectors are necessary for protection in humans remains unclear. In addition to cellular responses, serological responses may be important for protection in the liver. Recently, by screening sera from a volunteer who was protected after vaccination with radiation-attenuated sporozoites [72], a monoclonal antibody, CIS43, was isolated and found to confer a high-level of sterile protection in mouse models [149]. CIS43 binds to the junction of the N-terminus and central repeat region of the circumsporozoite protein, a site not previously thought to be involved in protection [102]. Hence, CIS43 was optimized to increase antibody half-life while maintaining epitope recognition and binding, resulting in CIS43LS [150]. CIS43LS was evaluated in a Phase 1 clinical trial wherein 9 out of 9 volunteers administered with the monoclonal antibody by direct venous injection were protected after challenge [102]. During the same screening, another monoclonal antibody, L9, was isolated from another protected volunteer who received irradiated sporozoites [107]. In mouse studies, L9 was shown to have similar or even improved reduction of infected hepatocyte burden to that of CIS43 suggesting it may have more inhibitory capacity than CIS43 [107]. As both novel monoclonal antibodies were isolated from a screen of serum from individuals exposed to irradiated sporozoites and protected after subsequent challenge, the extent to which these and other antibodies contributed to protection is unknown. Additionally, while CIS43- and L9-like antibodies were not identified in most individuals whose sera were screened, the limited sampling prevents conclusions about their overall frequency in the vaccinated population, underscoring potential heterogeneity in vaccine-induced responses.

1.1.8 **Population Genetics and Antigenic Escape**

The *P. falciparum* genome exhibits substantial genetic diversity [151, 152], which plays a pivotal role in the parasite's ability to adapt and persist against a wide range of environmental pressures. This genetic diversity enables the parasite to evade host immune responses [153], resist anti-malarial drugs [154], and adapt to novel interventions such as vaccines [155] and diagnostics [156]. Understanding the evolutionary forces shaping this diversity, such as natural selection and genetic drift, is crucial for developing effective strategies to control and eliminate malaria. The following sections delve into these interconnected topics, exploring how antigenic diversity aids immune evasion, how population genetics informs our understanding of vaccine development, and how evolutionary processes like selection contribute to the resilience and adaptability of *P. falciparum* [157, 158]. These insights provide a framework for understanding the diversity dynamics underlying the persistence of *P. falciparum* and the implications for malaria control, elimination, and eradication.

1.1.9 **Population Structure**

The population genetics of *P. falciparum* has been extensively studied to understand patterns of genetic variation across geographic regions. A central concept in these studies is population structure, which refers to the non-random distribution of alleles among subpopulations. Population structure can arise from a variety of evolutionary and ecological processes, including mutation, limited gene flow, local adaptation, genetic drift, and historical demographic events, among other factors. These processes contribute to differences in allele frequencies between groups [159, 160]. Population differentiation quantifies the extent of these differences, providing a measure of how genetically distinct subpopulations are from one another. It is commonly assessed using *F*-statistics such as F_{ST}, which estimates the proportion of total genetic variance that is attributable to between-population variation relative to total genetic variation [161, 162].

The *P. falciparum* genome consists of more than 5,300 genes, some of which exhibit substantial genetic diversity. This diversity is critical to the parasite's adaptability, including its evasion of host immune responses and resistance to antimalarial drugs [163]. Geographically, *P. falciparum* populations exhibit differences in genetic diversity and population structure. For example, populations in sub-Saharan Africa typically display higher levels of genetic diversity compared to those in Southeast Asia, South America, and the Pacific Islands [164-166]. These patterns are influenced by multiple factors, such as variation in transmission intensity and human movement, as well as other ecological and epidemiological dynamics that affect parasite population structure. For example, in high-transmission settings such as much of sub-Saharan Africa, genetic diversity is maintained through frequent recombination events among genetically distinct clones [167]. In contrast, low-transmission settings in Southeast Asia, South America, the Pacific Islands, and some region in sub-Saharan Africa, such as Eswatini, may sustain diversity through the importation of parasites from higher transmission areas [168]. Global studies also reveal that high-transmission settings are associated with a greater multiplicity of infection (MOI), the number of genetically distinct clones within an infection, and lower population differentiation, while low-transmission settings typically show reduced MOI and stronger genetic differentiation [158, 169-172]. In the context of vaccine development, analyzing genetic diversity, including population structure and variation over time, is necessary for identifying key polymorphisms associated with immune escape. Such insights help design vaccines that effectively cover antigenic diversity and mitigate the parasite's adaptive capabilities [13, 173, 174].

1.1.10 **Antigenic Escape**

Antigenic escape refers to the failure of the host immune system to recognize novel variants of pathogen antigens, allowing the pathogen to evade variant-specific immune responses [175, 176]. This phenomenon has been observed in the context of naturally acquired immunity to malaria and has limited the efficacy of several promising vaccine candidates in field trials [16-19]. Variation in parasite antigens at immunologically relevant loci prevents the recognition and elimination by highly specific immune responses enabling recurrent clinical episodes and the reduction of vaccine efficacy. To distinguish escape-driven variation, molecular epidemiological approaches have been developed to detect antigenic escape by examining the within-host dynamics of antigen alleles in individuals followed longitudinally [16, 19, 46, 177-179]. Using this approach, protection against symptomatic malaria has been associated with the acquisition of novel alleles at several antigen loci, indicating that variant-specific immune recognition may contribute to immune evasion. Although several antigen genes exhibit patterns consistent with immune selection and antigenic escape [180-182], the full spectrum of genetic polymorphisms contributing to this process in *P. falciparum* remains incompletely characterized.

1.1.11 **Genetic Diversity**

The genetic diversity of *P. falciparum* is shaped by several factors, including mutation, recombination, inbreeding, genetic drift, and migration. While mutation serves as the fundamental source of genetic variation, high rates of meiotic and mitotic recombination in *P. falciparum* generate novel genetic combinations, contributing to within-population diversity [183-192]. However, in regions with low transmission intensity, high levels of inbreeding can result in reduced genetic diversity and increased linkage disequilibrium [164, 193]. Similarly, genetic drift (change in the allele frequency in a population due to random chance) can also reduce genetic diversity. Geographic barriers and human movement patterns influence gene flow, leading to distinct population structures across different regions [194, 195]. For example, population structure has been observed in regions of Southeast Asia and South America where transmission is low and populations are more fragmented [194, 196]. Population size fluctuations also affect genetic diversity. In general, areas with high transmission rates have increased multiplicity of infection and large effective population sizes that facilitate frequent sexual recombination events, maintaining high genetic diversity [197, 198], although diversity can be maintained in low transmission settings, as well [151, 168]. Population bottlenecks, whether resulting from intensified control efforts or reductions in transmission, can reduce genetic diversity. In the case of intervention-driven bottlenecks, selective pressures may favor certain alleles while purging others, leading to both reduced diversity and potential shifts in allele frequency [199, 200]. Understanding these dynamics is essential for designing effective malaria control strategies that account for the complex and evolving landscape of *P. falciparum* genetic diversity.

1.1.12 **Balancing and Positive Directional Selection**

Balancing selection is a process that maintains genetic diversity within a population by preserving alleles at higher frequencies than expected from genetic drift alone [201]. Though multiple mechanisms can lead to higher-than-expected allele frequencies, negative frequency-dependent selection is particularly relevant in the interaction between pathogens and the host immune response [202]. Negative frequency-dependent selection favors rare phenotypes contributing to the parasite's ability to evade host immunity. Hosts mount immune responses that target the specific antigenic variants they have previously encountered, reducing the fitness of more common variants while providing a selective advantage to rarer variants. This dynamic creates a constant turnover of antigenic variants, maintaining high levels of genetic diversity within parasite populations [203]. Such diversity is evident in the erythrocyte invasion proteins of *P. falciparum*, which are critical for the parasite's ability to invade host red blood cells and are subject to strong immune selection pressures [204]. Key invasion ligands include members of the *reticulocyte-binding protein homolog* (*Rh*) gene family and the *erythrocyte-binding antigen* (*EBA*) gene family, both of which tend to exhibit significant genetic diversity [205, 206]. Over time, this process helps stabilize genetic diversity, facilitating the parasite's adaptability and complicating vaccine development efforts.

Positive directional selection, unlike balancing selection, tends to reduce genetic diversity by favoring alleles that confer a selective advantage under specific environmental pressures [207]. When selection is strong, this process can lead to a rapid increase in the frequency of beneficial alleles, potentially resulting in their fixation and leading to diminishing genetic variation in the affected region of the genome for a period of time. In the context of *P. falciparum*, positive directional selection often occurs when drug resistance emerges. For example, mutations in genes encoding drug targets or transporters, such as the *P. falciparum* chloroquine resistance transporter gene (*pfcrt*) for chloroquine resistance or *P. falciparum* multidrug resistance protein 1 gene (*pfmdr1*) for drug resistance against multiple drugs, are selected for under drug pressure [208-211]. These beneficial alleles increase in frequency within the population, leaving a genomic signature known as a selective sweep. This signature is characterized by reduced genetic diversity and increased linkage disequilibrium in the region of the genome linked to the selected allele [212]. The contrasting genomic signatures and mechanisms of balancing and positive directional selection highlight the interplay between evolutionary forces in shaping the *P. falciparum* genome, driving both immune escape and the anti-malarial resistance.

1.2 **Specific Aims**

1.2.1 **Overarching Research Question**

To facilitate the rational design of a protective malaria vaccine targeting a parasite with more than 5,300 genes, it is essential to determine which parasite components, including antigens and epitopes, are most strongly associated with protective immunity. One question is which antibody targets distinguish vaccine recipients who are protected from those who are not. Addressing this question may include a broad and unbiased assessment of the humoral response directed against each component of the entire *P. falciparum* proteome. By systematically profiling antibody reactivity using a high-throughput peptide array and comparing serological profiles between protected and non-protected individuals, it becomes possible to identify both known and previously uncharacterized antigens that contribute to vaccine-mediated protection. Integrating these data with machine learning and statistical modeling enables the discovery of immunological signatures predictive of protection, offering a path toward the identification and prioritization of antigenic regions most relevant to protective immunity. Additionally, considering the extensive diversity exhibited by candidate vaccine antigens, delineating the patterns of antigen evolution under immune- and intervention-driven pressures is crucial for vaccine developers. Understanding the historical and temporal patterns of genetic diversity and selection in candidate vaccine antigen genes, shaped by natural immunity and control interventions, can inform vaccine development and help minimize vaccine-induced selection. This study will address these questions by analyzing longitudinal parasite genomic data to detect signatures of selection at antigen loci. These analyses will help identify antigens that are under strong immune-driven selection that remains relatively stable over time, thereby supporting the prioritization of vaccine targets with greater potential for durable and broadly effective malaria vaccines.

1.2.2 **Specific Aim 1**

Identify humoral immune responses directed against *P. falciparum* antigens that are associated with protection following malaria challenge in individuals immunized with a whole organism, irradiated sporozoite vaccine.

Rationale and Approach: Despite progress in reducing malaria burden through vector control, chemoprevention, treatment strategies, and malaria vaccines [213], more efficacious and durable vaccines are still needed to reduce transmission and infection. Whole-organism vaccines, such as the irradiated *P. falciparum* sporozoite vaccine, have shown promising levels of protection, even against infection, in malaria naïve volunteers; however, the mechanisms underlying this protection remain incompletely understood [72-74]. Identifying the specific immune responses associated with protection will advance our understanding of vaccine-induced immunity and inform the selection of antigenic targets for future vaccine development. These insights may also help to distinguish immunologically relevant responses from non-specific or non-protective antibody responses, thereby refining correlates of protection and providing a foundation for improving both antigen design and vaccine trial efficiency [214].

To address this aim, a high-throughput peptide array covering the entire *P. falciparum* proteome was used to probe serum samples collected from vaccine recipients before and after immunization. Following vaccination, participants were challenged with *P. falciparum*, and outcomes were recorded to distinguish those who were protected from those who were not, enabling a systematic comparative analysis of peptide-specific antibody responses. By linking fine-scale immunological responses to vaccine outcomes, this work provides insight on the immunological mechanisms underpinning vaccine efficacy and the prioritization of key antigenic regions for further investigation, ultimately informing the development of improved malaria vaccines that elicit robust and durable protection.

Hypothesis: Vaccine recipients protected from infection upon challenge will have a repertoire of antibody responses that are distinctive from the immune responses of non-protected vaccine recipients.

1.2.3 **Specific Aim 2**

Identify signatures of antigenic escape in *P. falciparum* genes by investigating how naturally acquired immunity has shaped antigen evolution over time in southern Malawi. This aim seeks to identify changes in allele frequencies and patterns of genetic variation, including genetic signatures of balancing selection and population differentiation that are consistent with immune selection, providing insight into the evolutionary stability of vaccine targets and informing the development of broadly protective malaria vaccines.

Rationale and Approach: *Plasmodium falciparum* exhibits remarkable genetic diversity in antigenic loci that complicates vaccine development [17, 18, 215]. This diversity allows for antigenic escape whereby immune responses directed against specific antigen variants are evaded through the emergence and selection of new variants, enabling the parasite to avoid recognition and clearance by the host immune system [13, 174]. Temporal analyses of genetic diversity in vaccine candidate genes provide insights into the parasite's evolutionary response to natural immunity, offering key information for prioritizing durable vaccine targets that minimize the risk of immune escape.

Balancing selection, driven by allele-specific host immune responses, maintains polymorphisms in key antigenic regions that allow the parasite to evade immune detection [216, 217]. In this context, selection is frequency-dependent, whereby parasites with rare alleles have an advantage over parasites with common alleles, because they are less likely to be recognized by responses elicited by a prior infection. This aim focuses on identifying population genetic signatures across the genome that are consistent with antigenic escape, including signals of balancing selection and temporal fluctuations in allele frequencies that reflect selection against common alleles and for rare alleles. As balancing selection can maintain antigenic variants with essential functional roles, such antigens may represent constrained, immunologically relevant targets that are attractive for vaccine design.

To detect patterns of immune-driven selection and temporal changes in parasite genetic diversity, blood samples from symptomatic malaria cases were collected during three longitudinal studies conducted in Malawi: cohort 1 (2007–2009), cohort 2 (2014–2017), and cohort 3 (2019–2022). Genomic sequencing was then performed on these samples and the resulting genomic sequences were analyzed to identify signatures of balancing selection within each cohort, measure population differentiation between cohorts, and for loci identified in both analyses, temporal changes in major allele frequencies were examined to detect shifts consistent with frequency-dependent immune selection. Specifically, within each cohort, population genetic metrics such as nucleotide diversity (π), the number of segregating sites (S), and Tajima's D were used to assess deviations from neutrality and identify loci under balancing selection. Genetic differentiation between time points was then assessed using pairwise F_{ST}. Finally, for loci identified in either analysis, temporal changes in major allele frequencies were examined to detect shifts consistent with frequency-dependent immune selection. Temporal changes in major allele frequencies were examined to detect shifts in allele prevalence at polymorphic loci. These complementary analyses support the identification of immunologically relevant genes and polymorphisms, particularly those under selective pressure, and inform the prioritization of vaccine targets that may represent functionally constrained regions consistently recognized by host immunity. This approach facilitates the design of vaccines that reflect the naturally circulating antigenic diversity in endemic populations, enhancing their relevance and potential efficacy across genetically diverse host and parasite backgrounds.

Hypothesis: Antigenic loci involved in immune escape are expected to exhibit signals of balancing selection that reflect immune-mediated maintenance of diversity. These patterns will distinguish loci under immune pressure from those evolving under neutral processes or directional selection, thereby identifying genes most likely to contribute to antigenic escape.

Chapter 2

Research Design and Methods

2.1 **Aim 1 Methods**

2.1.1 **PfSPZ Vaccination, Sera Collection, and CHMI**

This investigation utilized serum samples from three independent clinical trials aimed at assessing the protective efficacy of the PfSPZ Vaccine in malaria-naïve adult volunteers [72-74]. Blood samples were drawn before and after immunization, and efficacy was evaluated by exposing participants to CHMI three weeks following the final vaccine dose. Malaria infection was closely monitored after exposure. Serum was obtained by centrifuging whole blood and stored at -80°C for subsequent immunological evaluations. Participants chosen for microarray profiling met three specific conditions: (1) they underwent homologous CHMI three weeks post-final dose, (2) the vaccine was delivered via direct-venous inoculation (DVI), and (3) their assigned vaccine dose and regimen groups included both individuals who were protected and those who were not post-CHMI. For peptide microarray analysis, sera from 42 individuals were tested both before the first dose and three weeks after the final immunization. Analyses were conducted on baseline and post-vaccination data separately, as well as on the difference between the two time points.

2.1.2 **Peptide Array Design and Probing**

To quantify antibody responses, a comprehensive peptide microarray was employed that spanned the entire proteome of the *P. falciparum* NF54 strain. The array featured overlapping 16-mer peptides with a 15 amino acid overlap, representing all 5,542 predicted protein-coding genes. Each unique peptide appeared only once on the array.

Peptides were synthesized in situ using a light-directed, maskless array synthesis system. The arrays were imaged using the Roche NimbleGen MS200 scanner at 635 nm with a 2 μm resolution and a 25% gain setting. Image alignment was handled using NimbleScan 2 software. To correct for spatial signal variation, a 2D loess normalization was applied, and background correction was carried out using the R package preprocessCore [218]. All statistical and machine learning analyses were performed on the background normalized raw fluorescence data from 42 subjects across two time points, totaling 3,918,353 unique peptides.

2.1.3 **Random Forests**

Random forest classification was employed due to its robustness in modeling high-dimensional immunological data with complex, non-linear relationships among features. As an ensemble method, random forests construct numerous tree-based statistical models using bootstrap sampling of the observations and random subsampling of variables, which collectively reduce overfitting and improve model stability [219]. In this study, class imbalance between protected and unprotected vaccinees was mitigated by under-sampling during the bagging phase. The model's predictive performance was assessed using the out-of-bag (OOB) error estimate, a form of internal cross-validation that leverages the unused observations in each bootstrap sample to test model accuracy. OOB error provides an unbiased estimate of generalization performance and is particularly valuable when external test sets are limited or unavailable [220]. With the construction of 100,000 trees, the model ensured convergence and stability of OOB error rates while maintaining computational efficiency in R.

2.1.4 **Elastic Net**

Elastic net regularization was used to predict protection status due to its efficacy in handling datasets where the number of predictors exceeds the number of observations ($p > n$), and where multicollinearity among predictors is substantial, as in the case of tiled peptide arrays [221]. The elastic net combines L1 (lasso) and L2 (ridge) penalties, allowing for both feature selection and coefficient shrinkage, improving model interpretability and generalizability [221]. This approach was implemented using the SGDClassifier in scikit-learn with a nested cross-validation framework to optimizes hyperparameters and avoid information leakage. The inner loop performed a grid search across a range of alpha values and L1 ratios using the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) as the selection criterion, while the outer loop evaluated the model's performance on held-out folds. This hierarchical validation design provides an unbiased estimate of model performance and reduces the risk of overfitting common in high-dimensional omics datasets [222]. Prior to modeling, features were log2-transformed and standardized, and class imbalance was managed via under-sampling, further ensuring robustness.

2.1.5 **Differential Seroreactivity**

The Wilcoxon signed-rank test and the Brunner-Munzel test were chosen to evaluate peptide-level immune responses due to their non-parametric robustness, especially under small sample sizes [223-225]. The Brunner-Munzel test is particularly robust against imbalanced classes which relevant to the imbalance between the 13 unprotected and 29 unprotected vaccine recipients in this study. The Wilcoxon signed-rank test, a paired non-parametric alternative to the t-test, assesses whether the ranks order of differences in peptide reactivity before and after vaccination are significantly different, effectively capturing vaccine-induced shifts in immune response without assuming normality [224]. In contrast, the Brunner-Munzel test extends this approach to unpaired data and specifically addresses violations of the assumption of equal variances and distributional shapes between groups, providing a robust alternative to the Mann-Whitney U test under these conditions [225]. This makes it suitable for immunological studies where differences in response magnitude and variability may confound simpler tests. The use of the Benjamini-Hochberg procedure [226] was employed to control the false discovery rate due to the high-dimensional nature of the peptide dataset.

2.2 **Aim 2 Methods**

2.2.1 **Chloroquine Combination Randomized Clinical Trial**

A randomized, open-label clinical trial was conducted in Ndirande, a malaria-endemic township in Blantyre, Malawi, to evaluate the effects of repeated antimalarial treatment over a 12-month period [227]. Children aged 6 months to 5 years presenting with symptoms consistent with uncomplicated malaria and *P. falciparum* mono-infection were enrolled if they met inclusion criteria and provided informed consent. Participants were randomized equally to receive chloroquine alone or in combination with artesunate, azithromycin, or atovaquone-proguanil at each malaria episode. Routine follow-up occurred every four weeks, with additional unscheduled visits for illness. Malaria diagnosis was confirmed by microscopy, and treatment was administered under direct observation. At each malaria episode and follow-up visit, blood samples were collected for microscopy, clinical measurements, and molecular analysis. Collected whole blood was cryopreserved before freezing and shipping to the University of Maryland. DNA was extracted via the QIAamp DNA Blood Mini Kit (Qiagen, CA) according to the manufacturer’s instructions.

To enhance the amount of parasite DNA relative to host DNA, selective whole genome amplification (sWGA) was performed using a protocol specifically optimized for clinical samples [228].

2.2.2 Mfera Cohort Study

A longitudinal study was conducted in the Chikwawa district in southern Malawi to following symptomatic and asymptomatic *P. falciparum* infections over time. The study, previously described in detail [229, 230], involved enrolling 120 participants from the Chikwawa district who initially presented with symptomatic, non-severe malaria at the Mfera Health Centre between mid-2014 and early 2015. Following initial diagnosis and antimalarial treatment, participants were longitudinally observed through both routine monthly assessments and unscheduled illness-driven visits for a duration of up to two years. At each encounter, a sample of whole blood was collected and screened for parasitemia using both microscopy and molecular assays to ensure sensitive detection of *P. falciparum*. For this genomic analysis, only samples corresponding to confirmed symptomatic infections during passive follow-up were included. Sampled whole blood was centrifuged, plasma was removed, and the red blood cell pellets frozen and stored before shipping and subsequent DNA extraction and whole genome sequencing. Nucleic acids were extracted using a high-sensitivity protocol validated for low-density parasitemia [231]. To increase parasite DNA yield relative to host DNA, samples underwent selective whole genome amplification (sWGA) employing a protocol refined for clinical isolates [228].

2.2.3 **Malawi ICEMR Household Study**

The International Centers of Excellence for Malaria Research (ICEMR) carried out a household-based cohort study within the catchment areas of Namanolo Health Centre in Balaka District and Ntaja Health Centre in Machinga District [232]. Starting in April 2019, households were randomly chosen with the objective of enrolling all residents, regardless of age. At the Namanolo site, 449 individuals were enrolled, and 537 were enrolled at the Ntaja site. Participants were monitored for approximately one year, with follow-up concluding in June 2021 for the Balaka site and in March 2022 for the Machinga site. Uncomplicated malaria cases were identified based on malaria-like symptoms and confirmed by rapid diagnostic tests and PCR. Blood samples were collected from participants who tested positive for uncomplicated malaria and preserved in RNA Protect before freezing and shipping to the University of Maryland, Baltimore for DNA extraction, sample processing, and parasite genome sequencing was conducted. DNA was extracted using the QIAamp DNA Blood Mini Kit (Qiagen, CA) following the manufacturer's instructions. Purified DNA was amplified using selective primer to preferentially amplify parasites sequences over human sequences following previously describe protocols [233].

2.2.4 **Whole Genome Sequencing and Data Processing**

Subsequent sequencing libraries were prepared from fragmented DNA using a modified workflow based on the KAPA Library Preparation Kit, incorporating bead-based purification and quantification via AMPure XP beads and qPCR, respectively [177, 234]. Equimolar pooling was followed by high-throughput Illumina sequencing to produce 150 bp paired-end reads suitable for whole-genome analysis.

2.2.5 **Balancing Selection and Population Differentiation**

To detect genetic signatures consistent with antigenic escape, I employed two population genetic metrics: Tajima's D [235] and the F-statistic, F_{ST} [159, 236]. Tajima's D was used to identify loci exhibiting deviations from neutrality suggestive of balancing selection, which is expected at antigenic sites maintained by allele-specific host immune responses [203, 237]. A positive Tajima's D indicates an excess of intermediate-frequency alleles, consistent with frequency-dependent selection favoring low-frequency variants that escape immune recognition. Tajima's D was calculated within each cohort across sliding genomic windows to identify regions under putative balancing selection.

To evaluate temporal population differentiation, Hudson's F_{ST} was calculated between pairs of time-separated cohorts (cohorts 1 vs. 2; 2 vs. 3; 1 vs. 3) [238, 239]. F_{ST} measures genetic differentiation between subpopulations relative to the total population, where values near 0 indicate no differentiation and values approaching 1 reflect nearly complete divergence [240]. Hudson's estimator was selected over the Weir–Cockerham method due to its reduced sensitivity to unequal sample sizes [241]. To assess empirical significance of F_{ST} for each window, a permutation procedure was performed in which cohort labels were randomly shuffled, F_{ST} was recalculated for each window, and this process was repeated 10,000 times to generate a null distribution of F_{ST} values. Observed F_{ST} values were then compared to this null distribution to calculate empirical P values, and loci with F_{ST} $P < 0.05$ were considered significantly differentiated across time.

Genes exhibiting both elevated Tajima's D and significant F_{ST} were prioritized as candidates under immune selection, as this combination of metrics highlights loci with maintained polymorphisms and shifting allele frequencies. For these loci, I additionally assessed temporal shifts in major allele frequencies to detect patterns consistent with selection against high-frequency alleles and for low-frequency alleles. Together, these analyses provide a robust framework for identifying immunologically relevant targets shaped by natural immunity.

Chapter 3

Peptide Microarray Profiling of PfSPZ Vaccine Responses that Correlate with Protection

3.1 **Abstract**

Identifying serological correlates of protection remains a critical barrier to the development of effective malaria vaccines. To identify peptides and proteins associated with protection, peptide-level antibody responses comparing serological responses of protected and unprotected malaria-naïve adults who received the PfSPZ Vaccine and subsequently underwent controlled human malaria infection (CHMI). Using a high-density microarray tiling 3.9 million overlapping 16-mer peptides across the *P. falciparum* proteome, I compared pre- and post-vaccination IgG reactivity profiles between protected and unprotected individuals. Vaccination induced broad serological responses, with significant IgG increases observed for 76,695 peptides from 4,292 proteins. However, few peptide responses were significantly associated with protection, and most differential signals were more reactive in unprotected participants. Predictive models using random forests and elastic net logistic regression failed to reliably classify protection status, suggesting that IgG responses to linear peptides alone are insufficient to predict vaccine efficacy. Bivariate differential reactivity analysis revealed a consistent pattern in which most differential peptides were inversely associated with protection, while relatively few peptides were associated with protection. This pattern was consistent when comparing pre, post, and post minus pre (change) reactivities between protected and unprotected groups. Gene ontology enrichment analysis linked protection-associated responses to membrane-localized proteins, while inverse associations were enriched for nuclear-related functions.

These findings highlight the complexity of vaccine-induced immunity and underscore the need for integrated multi-omic approaches to capture the complex immunological results leading to protection enabling the identification of robust correlates of protection in malaria vaccine trials.

3.2 **Introduction**

The development of effective and durable malaria vaccines will represent an important milestone in the fight against this devastating disease that killed 597,000 individuals in 2023 [1]. Two malaria vaccines, RTS,S/AS01 [38] and R21/Matrix-M [242], have been approved by the World Health Organization for widespread use. Both demonstrate moderate efficacy, ranging from 30% to over 70% protection against clinical malaria in field trials in Africa [5, 6, 243]. Despite their impact, both vaccines share multiple challenges including the need for improved and sustained efficacy and ongoing surveillance for parasite resistance driven by antigenic variation [18, 174]. Although these vaccines represent progress against malaria, their limitations underline the ongoing need for continued research and development of novel malaria vaccines with enhanced efficacy and durability of protection [244].

Attenuated whole-organism vaccines have shown promising protective efficacy against malaria in controlled human malaria infection (CHMI) studies involving malaria-naïve individuals [63]. The concept of irradiation-based attenuation originated from groundbreaking experiments by Nussenzweig et al., who demonstrated protection in a rodent malaria model using sporozoites from irradiated *Anopheles stephensi* mosquitos infected with *P. berghei* [64]. Clyde et al. [3, 65, 66] and Rieckmann et al. [69, 70] further validated this approach in humans, showing protection against infectious mosquito challenge after exposure to over a thousand bites from irradiated mosquitoes. Building on this foundation, Sanaria, a U.S.-based biotechnology company, developed a method to isolate and cryopreserve sporozoites from irradiated mosquitoes, eliminating the need for mosquito-bite administration [63, 71]. Sanaria PfSPZ Vaccine was evaluated in clinical trials by intravenous administration to volunteers resulting in 100% protection against homologous malaria infection in the group receiving the highest does of irradiated sporozoites [72]. Additional trials on naïve volunteers confirmed this protective efficacy enabling progression to heterologous [245, 246] and natural challenge trials in endemic regions [73, 74]. However, trials in these malaria endemic settings resulting in limited efficacy compared to the high protection levels observed in malaria-naïve individuals.

Identifying a correlate of protection will accelerate clinical trials and identify individuals or populations who develop insufficient protective responses after vaccination. A robust correlate of protection for malaria has long been pursued but remains elusive. One reason for the difficulty in establishing a correlate of protection is that a protective phenotype has been equally challenging to establish. Multiple definitions of protection have been employed including partial acquired immunity which is protective against symptomatic or severe malaria (but not infection) and sterile immunity which protects against infection, symptoms, and transmission [15]. The complexity of the parasite life cycle, the large number of proteins expressed by the parasite, and the extensive antigenic diversity of parasitic proteins compound the difficulty of identifying a robust correlate of protection. As the PfSPZ Vaccine demonstrated high levels of sterile protection in malaria-naïve North Americans, this clear and well-defined phenotype may enable detection of a correlate of protection that could be useful in evaluating efficacy in future clinical trials.

Additionally, identifying the correlates of protection in a pre-erythrocytic malaria vaccine trial presents an opportunity to discover potential vaccine-induced responses that are associated with protection from infection and may help to guide the rational design of more effective vaccines. Hence, I aimed to use peptide array data generated during several PfSPZ Phase 1 CHMI trials to identify correlates of protection and proteins that are critical in clinical malaria. I hypothesized that our approach would facilitate the identification of proteins or peptides with potential utility as components of future malaria vaccines.

3.3 **Results**

3.3.1 **Peptide-array**

To identify serological correlates of protection, reactivities against *P. falciparum* peptides were measured from 42 malaria-naïve adults enrolled in three Phase I clinical trials of the PfSPZ Vaccine. Blood was collected both before vaccination and three weeks after the final vaccination with subsequent challenge after which protection from infection was assessed, resulting in 29 protected and 13 unprotected vaccinees. Antibody responses from before and after the vaccination (but before the challenge) were measured using a high-density peptide microarray containing 16-amino acid peptides with 15-amino acid overlap representing the entire *P. falciparum* proteome derived from the NF54 strain, resulting in serological responses against 3,918,353 unique 16-mer peptides. Global seroreactivities from all peptides from all vaccinees were used to assess reactive versus unreactive peptides after which 420,214 peptides were deemed reactive while 3,498,139 peptides were considered unreactive and excluded from further analysis. Differential seroreactivity by bivariate statistical testing and machine learning methods were performed to identify peptides associated with or predictive of protection after challenge. For exploratory and predictive analyses additional covariates were included describing the vaccinee's originating study, sporozoite dosages, timing and number of doses received, and the processing batch. Parallel exploration of pre-vaccination, post-vaccination, and the absolute change between post- and pre-vaccination responses ensured that baseline antibody profiles, vaccine-induced responses, and vaccination-associated changes were each evaluated when comparing protected versus unprotected volunteers and when training machine-learning models.

3.3.2 Global Seroreactivity Patterns and Targets of PfSPZ Vaccination

Uniform Manifold Approximation and Projection (UMAP) of the spatial- and background-normalized peptide intensities revealed that, for most volunteers, pre- and post-vaccination samples clustered closely together indicating that differences in antibody responses between different participants were more pronounced than the changes observed within the same participant over time. A Wilcoxon signed-rank test was conducted to compare pre-vaccination and post-vaccination seroreactivity to identify antibody responses that increased after vaccination. A total of 76,695 peptides corresponding to 4,292 proteins exhibited significantly increased IgG reactivity two weeks post-vaccination (Benjamini–Hochberg FDR $\alpha = 0.05$), corresponding to 1.96% of the total 3,918,353 peptides assayed and 18.2% of the 420,214 reactive peptides (Fig. 3.1a). Only 2 peptides exhibited decreasing seroreactivity after vaccination, though both were from proteins that also contained peptides that significantly increased after vaccination. The circumsporozoite protein (PfCSP) emerged as the most immunogenic antigen with 77 of its 386 tiled peptides (19.9%) demonstrating significant increases after vaccination (Table 3.1, Figure 3.1b, and Figure 3.2).

Table 3.1: Table of five proteins with the highest proportion of peptides with increased reactivity after vaccination.

Protein Description	Gene ID	Reactive Peptides	Total Peptides	Proportion Reactive Peptides
Circumsporozoite Protein	PF3D7_0304600	77	386	0.20
Splicing Factor 3B Subunit 4, putative	PF3D7_1420000	75	473	0.16
Conserved *Plasmodium* Protein, unknown function	PF3D7_1119700	48	341	0.14
Conserved Protein, unknown function	PF3D7_1368900	55	439	0.13
Conserved *Plasmodium* Protein, unknown function	PF3D7_0711300	13	108	0.12

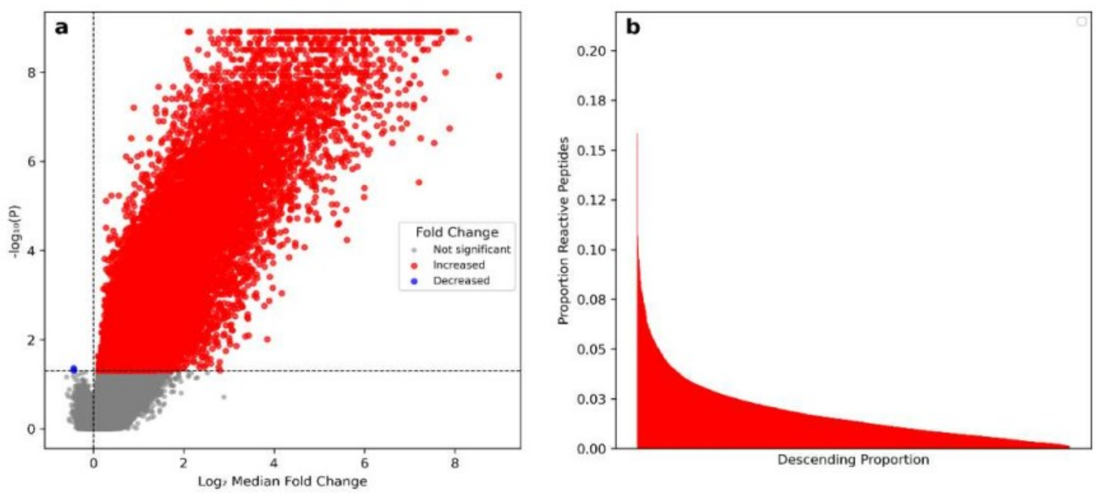


Figure 3.1: Differential seroreactivity after vaccination. (a) Volcano plot of differential pre-vaccination and post-vaccination serological responses for each peptide. The x-axis represents median of the fold change (log2(pre/post)) from each participant for a given peptide. Each spot represents one peptide. The y-axis is the -log10(FDR-corrected P value). (b) Bar plot of the sorted proportion of peptides with positive fold change in reactivity after vaccination divided by the total number of peptides represented from a given protein on the array.

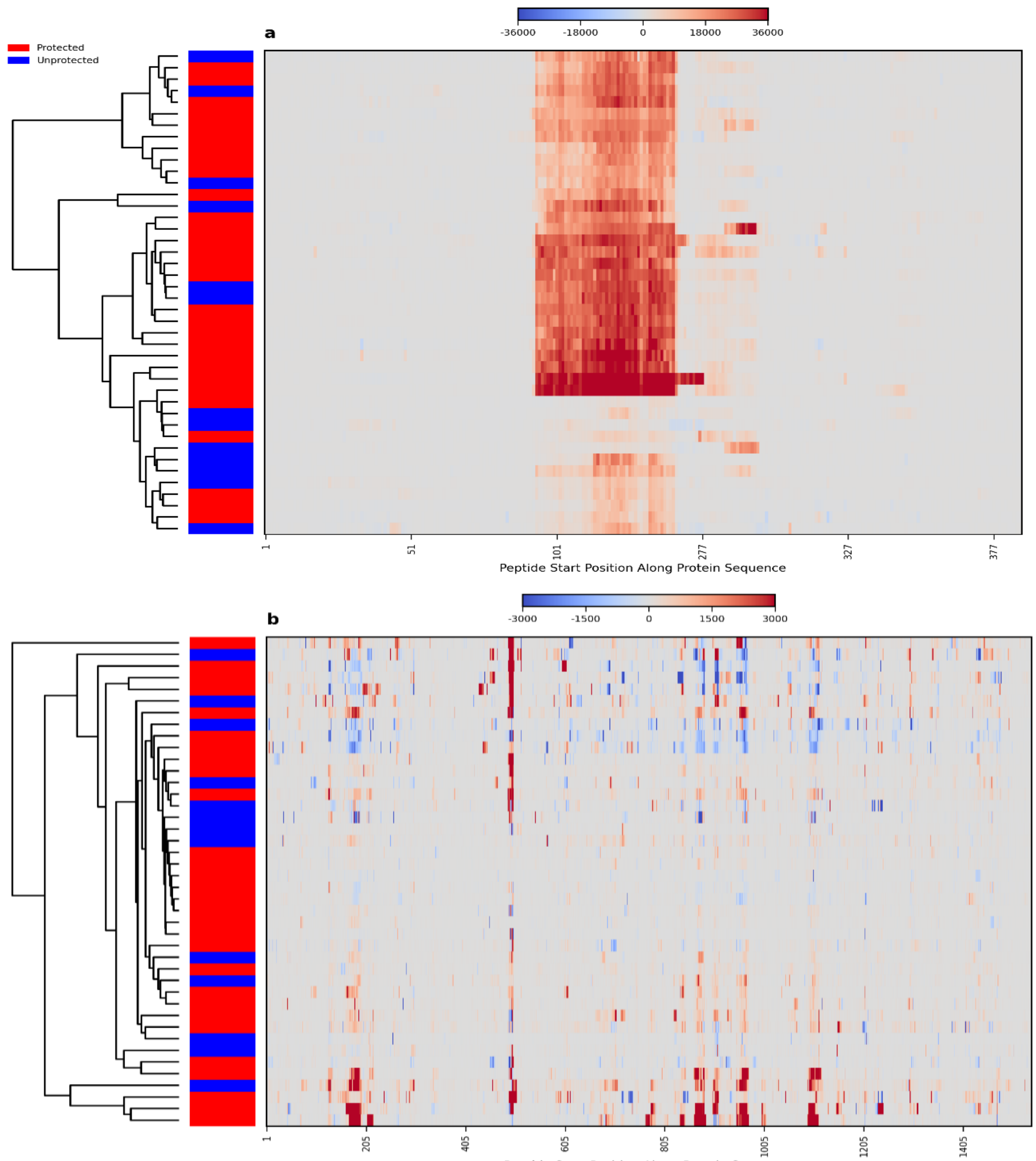


Figure 3.2. Agglomerative hierarchical clustering heatmap of the two proteins with the highest percentage of reactive peptides after vaccination. (a) Circumsporozoite protein (PfCSP, PF3D7_0304600) and (b) putative protein splicing factor 3B subunit 4 (PF3D7_1420000) each with protection status of the absolute change in reactivity due to vaccination. The y-axis is the position along the full protein sequence that aligns with the first amino acid of the peptide. The x-axis corresponds to the peptide responses for each vaccinee with protected participants colored red and unprotected colored blue. A dendrogram indicates the clusters of responses.

3.3.1 **Serological Targets Associated with Protection**

Peptide-level antibody reactivity was evaluated in relation to protection status across pre-vaccination, post-vaccination, and difference (post minus pre) data sets. Few peptides showed statistically significant differences between protected and unprotected individuals, and most exhibited higher reactivity in the unprotected group. This pattern was consistent across all time points. To explore the biological relevance of these findings, gene ontology (GO) enrichment analysis was performed on proteins corresponding to significant peptides associated with protection.

Sera collected prior to vaccination was examined for baseline associations with protection. The baseline seroreactivities of 3 peptides, each from a distinct protein, showed higher baseline reactivity in individuals who were later protected after vaccination and challenge (Figure 3.3a). Of these three source proteins, Rhoptry Neck Protein 3 (PfRON3, PF3D7_1252100), was the only significantly enriched protein in the Gene Ontology (GO) analysis. PfRON3 was functionally annotated to be involved in protein transport and is localized to the cell surface, extracellular space, and rhoptry, with a molecular function classification of host cell surface binding. In contrast, 71 peptides corresponding to 61 proteins showed higher baseline seroreactivity in individuals who were not protected following vaccination and challenge. GO enrichment analysis revealed that 6 of 61 proteins were significantly associated with five biological processes, including DNA repair, intracellular receptor signaling, and various histone modifications. Furthermore, three of these proteins were functionally linked to three molecular functions: methylated histone binding, oxidoreductase activity, and ubiquitin binding. However, none of these proteins showed significant enrichment in the GO category for cellular components.

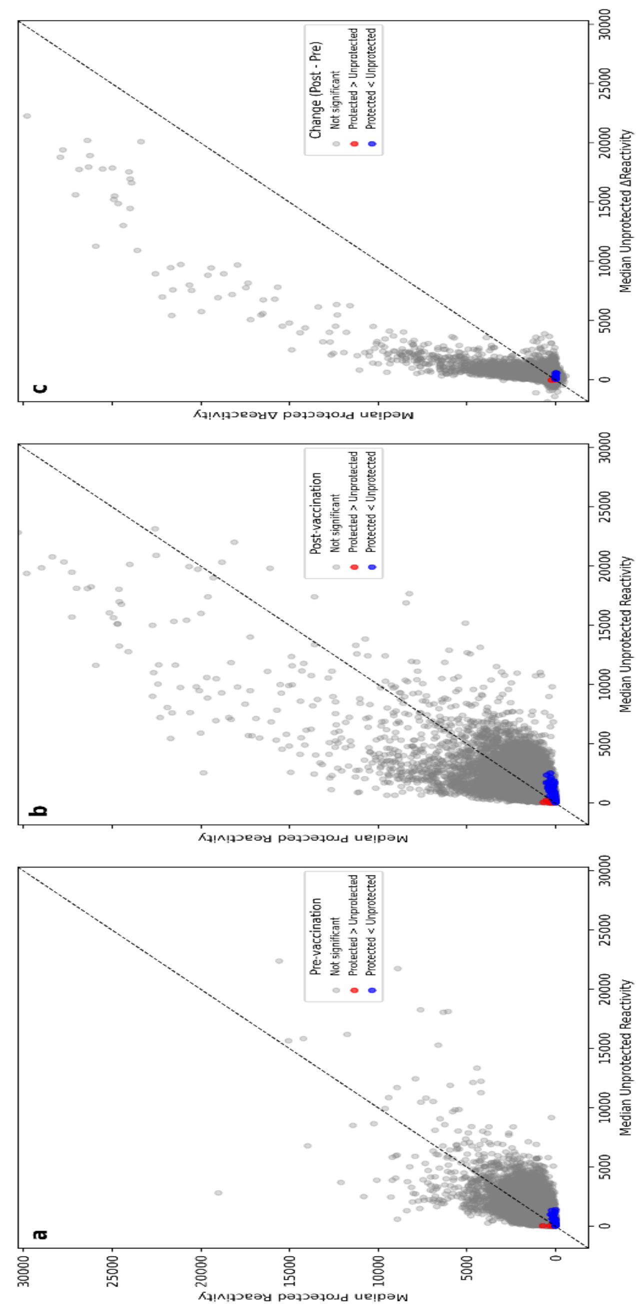


Figure 3.3: Scatter plots of median peptide fluorescence intensities (FI) among protected versus unprotected volunteers. (a) Pre-vaccination, (b) Post-vaccination, and (c) Difference (post minus pre). Peptides with higher median responses in protected individuals are shown in red, the converse shown in blue, and with insignificant responses (Generalized Wilcoxon, Benjamini-Hochberg False Discovery Rate adjusted P < 0.05) shown in grey.

Sera collected prior to vaccination was examined for baseline associations with protection. The baseline seroreactivities of 3 peptides, each from a distinct protein, showed higher baseline reactivity in individuals who were later protected after vaccination and challenge (Figure 3.3a). Of these three source proteins, Rhoptry Neck Protein 3 (PfRON3, PF3D7_1252100), was the only significantly enriched protein in the Gene Ontology (GO) analysis. PfRON3 was functionally annotated to be involved in protein transport and is localized to the cell surface, extracellular space, and rhoptry, with a molecular function classification of host cell surface binding. In contrast, 71 peptides corresponding to 61 proteins showed higher baseline seroreactivity in individuals who were not protected following vaccination and challenge. GO enrichment analysis revealed that 6 of 61 proteins were significantly associated with five biological processes, including DNA repair, intracellular receptor signaling, and various histone modifications. Furthermore, three of these proteins were functionally linked to three molecular functions: methylated histone binding, oxidoreductase activity, and ubiquitin binding. However, none of these proteins showed significant enrichment in the GO category for cellular components.

Post-vaccination antibody responses reflected a similar trend to the baseline (pre-vaccination) data of which relatively few peptides exhibited significantly higher reactivity in protected individuals, whereas a larger set of peptides was elevated in unprotected individuals (Figure 3.3b). Specifically, six peptides, corresponding to six proteins, were significantly more reactive in the protected group, while 222 peptides, mapping to 167 distinct proteins, were elevated in the unprotected group. Among the six proteins enriched in protected individuals, three were associated with five biological processes, including microtubule cytoskeleton organization, anatomical structure development, protein depolymerization, coenzyme A biosynthesis, and supramolecular fiber organization. Four of these proteins were also localized to distinct cellular components such as the cell surface, microtubules, subpellicular microtubules, and the symbiont-containing vacuole. Additionally, two proteins were linked to molecular functions including RNA endonuclease activity and ligase activity, suggesting involvement in nucleic acid metabolism. In contrast, proteins associated with higher reactivity in the unprotected group (167 total) showed extensive enrichment across several GO categories. These included seven biological processes such as DNA replication, DNA biosynthesis, RNA biosynthesis, endocytosis, and histone methylation. Proteins in this group were also significantly enriched in seven cellular component categories, with strong representation in nuclear and chromatin-associated structures such as the DNA polymerase complex, heterochromatin, and nuclear replisome. Molecular function enrichment among these proteins included DNA polymerase activity, topoisomerase activity, histone binding, and sequence-specific DNA binding.

To assess how antibody responses changed after vaccination, I examined the absolute change in peptide reactivity between pre- and post-vaccination timepoints. Only two peptides showed greater increases in protected individuals (Fig. 3.3c). One peptide was derived from a putative protein, Nucleoporin NUP390 (PfNUP390, PF3D7_0212400) and the other from Reticulocyte-binding Protein Homologue 3 (PfRH3, PF3D7_1252400). Only NUP390 was enriched for the nuclear periphery among cellular components. In contrast, 23 peptides, each from a different protein, exhibited larger increases in the unprotected group. Of the 23 proteins with greater post-vaccination signal increases in unprotected individuals were significantly enriched among cellular components such as the 90S pre-ribosome, heterochromatin, COPII vesicle coat, subpellicular microtubule, and cell surface. Enriched molecular functions included cyclase activity, hydro-lyase activity, and methylated histone binding. Biological process enrichment was limited but included anatomical structure development and protein transport by the Sec complex.

3.3.2 Predictive Modeling of Protection

Random forests and elastic net logistic regression, two machine learning approaches adept at handling high-dimensional data, were employed to identify serological responses predictive of protection. The random forest models included 100,000 classification trees with under-sampling of the majority class to address class imbalance. Separate analyses were conducted for pre-vaccination, post-vaccination, and change (post minus pre) in antibody responses for each peptide. Each model incorporated covariates including PfSPZ dosage and administration details, originating clinical trial, and array processing batch. The random forest model yielded overall accuracies of 54.8% for pre-vaccination, 69.0% for post-vaccination, and 61.9% for the change data set, all of which are not substantially different from the majority-class baseline of 69% (29/42).

For the elastic net logistic regression, a nested cross-validation framework was implemented to optimize hyperparameters and robustly evaluate model performance. In this framework, the inner folds tuned the model parameters while the outer folds assessed performance using the area under the receiver operating characteristic curve (AUC-ROC) as the accuracy metric. This approach produced mean AUC-ROC values of 44.3% (S.D. = 7.03%) for pre-vaccination, 47.3% (S.D. = 22.7%) for post-vaccination, and 58.7% (S.D. = 22.8%) for the pre-to-post change. The final predictive model combined random forest for feature selection with elastic net logistic regression within a nested cross-validation framework to prevent data leakage. However, this combined approach also resulted in poor predictive performance, with AUC-ROC accuracies of 48.2% (SD = 13.8%) for pre-vaccination, 50.4% (SD = 31.2%) for post-vaccination, and 37.9% (SD = 20.6%) for the change dataset. Overall, none of the models outperformed a naïve baseline prediction.

3.4 **Discussion**

In this study, I analyzed sera from malaria-naïve recipients of the PfSPZ Vaccine who were either protected or unprotected following controlled human malaria infection. Using a high-density peptide microarray tiling the *Plasmodium falciparum* NF54 proteome, I observed that PfSPZ vaccination elicited broad IgG responses, with 76,695 peptides from 4,292 proteins showing significantly increased reactivity relative to baseline. Despite this extensive reactivity, few responses were associated with protection against homologous challenge, and the magnitude of these differences was generally modest.

Global patterns of antibody binding visualized by UMAP revealed individual-specific profiles, with pre- and post-vaccination samples from the same participant showing greater similarity than samples from different individuals, despite direct venous administration of over 500,000 irradiated sporozoites between these time points. This observation aligns with previous reports describing distinctive antibody profiles called immunosignatures, that are unique to each individual, possibly arising from prior immune exposures or antibody cross-reactivity [247-249]. Such inter-individual variation likely complicates the identification of group-level associations and may inflate false positive or obscure meaningful signals, underscoring the value of developing robust peptide-level internal controls.

Despite broad serological reactivity, machine learning models built from pre-vaccination, post-vaccination, and change-in-response datasets failed to predict protection above chance levels, irrespective of algorithm, feature selection strategy, or data subset. These findings suggest that peptide-specific IgG responses alone are insufficient to reliably predict protection following PfSPZ vaccination. Comparisons between protected and unprotected individuals revealed an unexpected pattern: most differential peptide responses were more reactive in unprotected participants relative to protected participants, and this difference was detectable prior to vaccination. Since all participants were malaria-naïve, such pre-vaccination differences are unlikely to reflect prior malaria exposure. Instead, these pre-vaccination differences may reflect pre-existing antibodies elicited by prior exposure to unrelated pathogens, resulting in cross-reactive binding to malaria peptides. Supporting this interpretation, GO enrichment of proteins corresponding to peptides inversely associated with protection revealed overrepresentation of nuclear pathways, including histone modification and DNA repair, functions localized to compartments that are typically inaccessible to antibody-mediated recognition. Although the median magnitude of these pre-existing responses was low, they may reflect prior exposures to pathogens similar to *P. falciparum*. This is consistent with original antigenic sin, which is the immunological phenomenon where the immune system preferentially recalls and amplifies responses to an initial antigen exposure, potentially at the expense of mounting optimal responses to new but related antigens [250]. A similar pattern of pre-existing immune responses being inversely associated with protection was found during a serological analysis of malaria-naïve North American vaccinees who developed higher anti-PfCSP antibody responses than Tanzanian adults receiving the same vaccine, despite the latter's prior exposure to malaria [251, 252]. The lower efficacy observed in Tanzanian participants may indicate that immune history or antigen familiarity modulates both vaccine responsiveness and protection. In this context, the inversely associated antibody responses identified here may serve as markers of historical immune activation that influence the likelihood of protection [75, 76, 253, 254]. However, this interpretation remains an area of active research and warrants further investigation in larger, immunologically diverse cohorts.

Among proteins showing the strongest increases in response after vaccination, circumsporozoite protein (PfCSP) was the most immunogenic antigen, with approximately 20% of its peptides exhibiting increased reactivity after vaccination. Additional proteins with high proportions of reactive peptides included the putative Splicing Factor 3B Subunit 4 protein (PF3D7_1420000) and several conserved proteins of unknown function. Though these proteins showed high-intensity changes, they did not contain peptides significantly associated with protection. In the case of PfCSP, while several central-repeat region peptides were nominally associated with protection prior to multiple comparison correction, none remained significant after FDR adjustment. Clustering analysis of PfCSP responses revealed that among the 5 largest clusters, four showed strong reactivity targeting the central-repeat region, and individuals in these clusters were predominantly protected (clusters 1-4 combined: 21/27(77.8%) protected compared to 6/27(22.2%) unprotected). In contrast, the fifth cluster, which showed the weakest reactivity to the central-repeat region, included a similar number of protected (8) and unprotected (7) individuals. This suggests that some protected individuals generate low-level PfCSP antibody responses comparable to those of unprotected individuals. While these findings establish PfCSP antibodies as a definitive correlate of protection, they suggest that heterogeneity in responses targeting PfCSP may contribute to variability in vaccine-induced protection. These results are consistent with previous results from Camponovo, et al. [253], which analyzed serological responses of malaria-exposed, Tanzanian PfSPZ Vaccine recipients against 7,455 full length or fragmented *P. falciparum* proteins. In that study, reactivity against PfCSP was initially associated with protection, but the association lost statistical significance after adjustment for multiple comparisons [253]. Larger sample sizes may improve the power to detect PfCSP as a correlate of protection; however, if substantial individual variation exists even among protected individuals, PfCSP may have limited utility as a universal correlate of protection.

This study had several notable strengths. The use of homologous challenge ensured a clearly defined protective phenotype, eliminating strain-specific variation as a confounder. The proteome-wide peptide array enabled a systematic and unbiased high-resolution evaluation of humoral responses to vaccination, including potential targets not captured in more targeted assays. Limitations of this study include detection of total IgG without subclass or avidity resolution, and the inherent restriction of peptide-based arrays to linear epitopes, excluding large conformational and discontinuous epitopes [255]. Importantly, other components of the immune response, such as T cell activity, antibody effector functions, and innate mechanisms, were not assessed, though they are likely critical to protection [72, 73, 251].

Together, these findings indicate that while PfSPZ vaccination elicits widespread IgG reactivity, protection from malaria in this setting is not readily explained by antibody responses to against linear peptides alone. The predominance of non-protective or inversely associated responses, particularly to intracellular and conserved parasite antigens, underscores the complexity of vaccine-induced immunity. Protective responses are likely multifactorial and likely require the coordinated engagement of antibody, cell-mediated, and innate immunity mechanisms. Future studies incorporating multi-omic immunoprofiling and larger, diverse cohorts will be essential to identify reliable correlates of protection and to guide the rational design of more effective malaria vaccines.

3.5 **Methods**

3.5.1 **PfSPZ Vaccination and Control Human Malaria Infection (CHMI)**

This study analyzed serum from three clinical trials across four study sites that evaluated the efficacy of the PfSPZ Vaccine against controlled human malaria infection (CHMI) in malaria-naïve adults [72-74]. In each trial, participants received the PfSPZ Vaccine via direct-venous inoculation (DVI) and underwent homologous CHMI three weeks after final vaccination, during which protective efficacy was assessed by active daily monitoring for malaria infection. Blood was collected prior to and after vaccination from which serum was immediately separated by centrifugation and stored at -80°C for immunological assays [71-74]. As each trial assessed similar, but not identical, dosing and vaccination regimens, groups were selected from each trial where incomplete protection or susceptibility to CHMI was observed. Sera collected from these 42 individuals prior to first vaccination and three weeks after the final vaccination with the PfSPZ Vaccine were used to assess seroreactivity by peptide microarray, with analyses performed on pre-vaccination and post-vaccination measurements separately as well as on the change in signal.

3.5.2 **Whole Proteome Peptide Microarray**

Antibody responses before and after vaccination were measured using a high-density peptide microarray covering the proteome of the NF54 strain of *P. falciparum*. Peptides consisting of 16-amino acids with 15 amino acid overlaps were tiled across protein coding sequence consisting of 5,542 open reading frames (ORF). Identical peptides were represented only once per array. High-density peptide synthesis was performed using a Maskless Array Synthesizer by light-directed solid-phase peptide synthesis following established protocols [255]. Arrays were then scanned with the Roche NimbleGen MS200 microarray scanner (Madison, WI) at 635 nm at 2 μm resolution with 25% gain. The resulting images were aligned using the NimbleScan 2 software (Roche NimbleGen, Madison WI). The resulting fluorescence intensities were smoothed for potential spatial bias by 2D loess normalization [256]. Robust multi-array analysis background correction was performed using the R package preprocessCore [218]. Statistical and predictive analyses was performed on spatial- and background-normalized serological intensities from 42 malaria-naïve participants, across three studies, at two time points (pre- and post-vaccination), covering 3,918,353 unique peptides, resulting in over 329 million data points. Unless otherwise specified, all seroreactivity analyses were performed on the spatial- and background-normalized raw fluorescence intensities (not $\log_2$-transformed). Log_2 fold-change values were calculated only for visualization purposes, and $\log_2$-transformed data were used solely for elastic-net modeling.

3.5.3 **Exploratory Data Analysis and Peptide Filtering**

UMAP was performed to explore broad patterns in peptide reactivity and their relationship to protection and the following covariates: study site, single dose sporozoite count, number of vaccine doses, total sporozoites administered, weeks in which vaccine dosages were received, and probing batch. A 2D UMAP embedding was generated using a grid search across combinations of the number of nearest neighbors, minimum distance between clusters, and distance metrics (Euclidean and Manhattan). For each parameter combination, embeddings were visualized with lines connecting paired pre- and post-vaccination samples from the same individual. Points and connecting lines were color-coded based on protection status and other covariates. To identify reactive peptides and filter away unreactive peptides, a global threshold was established by graphing the empirical cumulative density function plot of all peptide responses from all volunteers from both pre- and post-vaccination timepoints combined. The maximum distance "elbow" analysis threshold [257] distinguishing elevated reactivities from the lower baseline plateau. The threshold was determined by calculating the maximal perpendicular distance from each point to the line connecting the extreme ends of the sorted distribution to identify a global threshold separating reactive from unreactive responses. Peptides were considered unreactive and removed from further analyses if less than 95% of the reactivities of that peptide were below the global threshold, thereby reducing noise and dimensionality while preserving the most informative components of the dataset.

3.5.4 **Differential Seroreactivity**

Bivariate analyses identified peptide responses either associated with changes after PfSPZ vaccination or associated with protection status. Changes after vaccination were determined using paired Wilcoxon signed-rank test comparing pre- and post-vaccination peptide reactivities. Associations with protection status compared peptide responses (pre-vaccination, post-vaccination, and their differences) between protected and unprotected groups using the Generalized Wilcoxon test. The Wilcoxon signed-rank test and the Generalized Wilcoxon test were selected for their robustness to small sample sizes, minimal distributional assumptions, and ability to manage skewed distributions and class imbalances. For both analyses, statistical significance was determined at an $\alpha = 0.05$

threshold, with Benjamini-Hochberg False Discovery Rate adjustments applied to control for multiple comparisons. Median fold-changes and group differences were visualized using volcano plots. To indicate the which proteins contained the most reactive peptides while adjusting for protein size, I calculated the proportion of peptides that significantly increased after vaccination relative to the total peptides representing each protein on the array. This metric allowed us to identify proteins in which multiple peptides exhibited increased reactivity, a pattern expected for proteins represented by overlapping peptides and one that may indicate exposure to the vaccine.

To predict protection from serological responses, two machine learning analyses were performed. Elastic net logistic regression and random forest classification methods were used because they are well-suited to datasets with many more predictors than samples. These methods efficiently manage feature selection, handle class imbalance, and capture complex interactions among predictors effectively. For elastic net logistic regression, I utilized the SGDClassifier package from scikit-learn with a combination of L1 (lasso) and L2 (ridge) regularization penalties to predict protection based on peptide reactivity profiles [221]. Elastic net regularization is particularly useful for high-dimensional datasets with highly correlated features such as responses to significantly overlapping tiled peptides, as it effectively manages collinearity among predictors [221]. Nested cross-validation was performed to determine the best-performing hyperparameter configuration and provide a robust estimation of model generalization [258]. In the inner loop, I conducted a grid search over regularization parameters (alpha values of 0.001, 0.01, 0.1, 1, and 10, and L1 ratios of 0.1, 0.5, and 0.9) with parallelized model fitting, using the area under the receiver operating characteristic curve (AUC-ROC) as the evaluation metric [259]. Numeric predictors were log_2-transformed then standardized via z-score transformation prior to model fitting, and under sampling of the majority class ensured balanced class representation. Hyperparameters identified from the inner fold with the highest mean AUC-ROC accuracy and lowest standard deviation across folds were then applied in the outer cross-validation loop, where independent folds were used to assess generalization. In parallel, random forest classification was performed in R using the randomForest package. The random forest model was constructed with 100,000 decision trees, and an under-sampling strategy was implemented during the bagging process to address the noted class imbalance [219, 260]. Random forests are capable of analyzing high-dimensional datasets and non-linear relationships and interactions between features without extensive preprocessing [219]. Model performance was estimated via the out-of-bag (OOB) error.

3.5.5 Gene Ontology(GO) Enrichment

GO enrichment analysis was performed using the pfGO R package [261] to identify molecular functions, biological processes, and cellular components that were enriched in identified genes relative to background genes. Separate GO enrichment analyses were conducted for proteins containing peptides that were significantly associated with increased reactivity after vaccination. GO analysis was also peptides significantly associated with protection among pre-vaccination, post-vaccination, and change in responses. Background genes consisted of all genes that contain sequences of reactive peptides on the array after filtering. The pfGO R package uses a weighting method that applies a hierarchical weighting algorithm that leverages the GO structure to reduce redundancy and down-weight the significance of terms with overlapping gene annotations without adjustment for multiple comparisons [262, 263].

3.5.6 **Protein Seroreactivity Clustering and Visualization**

Agglomerative hierarchical clustering of peptides responses from all identified proteins was performed using complete linkage with Euclidean distance on peptide intensity values.

Chapter 4

Intertemporal Identification of Genetic Determinants of Antigenic Escape

4.1 Abstract

Antigenic diversity in *Plasmodium falciparum* poses a major challenge for effective and durable malaria vaccine design. This chapter presents an intertemporal genomic framework to identify genetic loci under immune-driven balancing selection using whole-genome sequence data from 114 clinical malaria infections across three time-separated studies in Malawi. By applying sliding window analyses of Tajima's D and Hudson's F_{ST} estimator, genes with persistent allelic diversity and temporal shifts in allele frequencies were identified. A combined threshold-based and empirical strategy revealed both stable and time-variable signals of selection, enriching for loci potentially shaped by host immunity. Further analysis prioritized genes with high Tajima's D and significant F_{ST}, including PF3D7_0710200, which was selected for proof-of-concept analysis. Using haplotype clustering and random forest classification of nonsynonymous SNPs, this gene revealed four distinct haplotype clusters, each defined by unique combinations of amino acid variants with moderate to high population frequencies. These clusters exhibited temporal persistence and divergence consistent with immune selection and putative serotype differentiation. This approach demonstrates how integrating population genetics with unsupervised and supervised machine learning can uncover dynamic antigenic variation and inform prioritization of vaccine targets.

4.2 Introduction

Globally, malaria remains one of the deadliest infectious diseases, with the World Health Organization (WHO) estimating approximately 263 million cases and 597,000 deaths in 2023 [1]. Over 90% of these cases occur in sub-Saharan Africa, where transmission intensity is highest and children under five years of age account for an estimated 74% of malaria deaths [1]. Despite extensive control efforts, progress has plateaued, and new strategies are urgently needed [1]. Vaccines offer a promising approach to reduce the malaria burden. However, the first-generation malaria vaccines, RTS,S/AS01 and R21/Matrix-M, confer only moderate and short-lived protection, highlighting the need for more durable and broadly effective immunization strategies.

A major challenge in malaria vaccine development is the extensive genetic diversity of *Plasmodium falciparum*, particularly within antigen-encoding genes [16, 180, 264]. In addition to expressing different antigens at each life stage, *P. falciparum* exhibits high levels of polymorphism within individual antigens, much of which has evolved over millennia as a mechanism of immune evasion [265, 266]. Host immune responses are often allele-specific, targeting certain antigenic variants while failing to recognize others [267]. This allows parasites carrying rare or divergent variants to escape detection and persist. Such antigenic escape drives frequency-dependent balancing selection, a form of positive selection in which parasites harboring rare alleles gain a fitness advantage by avoiding immune responses elicited by more common variants [268]. As these rare alleles rise in frequency in the population, their advantage diminishes due to increased immune recognition, leading to cyclical shifts in allele prevalence [216]. Over time, this dynamic can stabilize, maintaining multiple alleles at intermediate frequencies and contributing to long-term antigenic diversity.

To address the challenge of antigenic diversity and immune escape, this study analyzes genomic data from three time-separated longitudinal studies conducted in

Malawi. By integrating two complementary population genetics metrics, Tajima's D and F_{ST}, the analysis identifies genes exhibiting persistent allelic diversity and temporal shifts in allele frequencies consistent with immune-driven selection. Candidate loci are further examined using unsupervised clustering followed by supervised classification to explore haplotype structure and highlight coding variation that may influence immune recognition. One gene is presented in detail as a proof of concept, illustrating the potential of this approach to reveal dynamic antigenic variation shaped by host immunity. This framework offers a broadly applicable strategy for detecting immune-relevant genetic diversity and prioritizing targets for vaccine design.

4.3 **Methods**

4.3.1 **Studies Contributing Samples**

Each of the three Malawian longitudinal studies from which samples were used in this analysis were performed in catchment areas around local health centres. For all studies, parasite genome sequences were obtained from clinical isolates collected during symptomatic malaria episodes, prior to the initiation of treatment. To preserve a cross-sectional design, one parasite genome per study participant, randomly selected, was included in the genomic analyses, even if the individual experienced multiple symptomatic episodes. The three studies will be referred to as time points 1-3 corresponding to the oldest, middle, and most recent time frames.

A randomized, controlled clinical trial conducted in Blantyre, Malawi from 2007 to 2009 (time point 1) investigated the efficacy of chloroquine monotherapy compared to chloroquine combined with artesunate, azithromycin or atovaquone-proguanil for the treatment of uncomplicated malaria in children [227]. Children from the catchment area of the Ndirande Health Centre who presented with signs and symptoms consistent with uncomplicated malaria and confirmed via thick smear microscopy were enrolled. Enrollment started in February 2007, resulting in 640 children aged six months to five years enrolled and followed over 12 months concluding in August 2009. During both clinical malaria episodes and routine follow-up visits, blood samples were obtained for parasitological examination, clinical evaluation, and molecular diagnostics. Venous whole blood samples were frozen shortly after collection for storage until later transported to the University of Maryland, where DNA was extracted and subjected to whole genome sequencing.

A prospective cohort study was conducted in the Chikwawa district of rural southern Malawi from 2014 to 2017 (time point 2) to estimate the incidence and duration of infections, including both symptomatic and asymptomatic cases [229, 230]. Detailed study procedures have been reported previously [177, 229, 230]. In brief, 120 individuals presenting with uncomplicated malaria at the Mfera Health Centre between June 2014 and March 2015 were enrolled and subsequently followed over a period of up to two years. Participants underwent scheduled monthly visits as well as additional clinical assessments when presenting with illness. At each visit, venous or capillary blood samples were collected and analyzed for *P. falciparum* parasitemia using both light microscopy and PCR-based methods to maximize diagnostic sensitivity. For the purposes of genome sequencing, only specimens collected during passive surveillance from individuals with confirmed symptomatic malaria episodes were selected. Whole blood samples were centrifuged to remove plasma, and the resulting red blood cell pellets were frozen, stored, and later processed for DNA extraction and whole genome sequencing.

The Malawi International Centers of Excellence for Malaria Research conducted a household-based cohort study that enrolled and followed participants in the catchment area of the Namanolo Health Centre in Balaka District and Ntaja Health Centre in Machinga District from 2019 to 2022 (time point 3) [232]. Beginning in April 2019, households were randomly selected with the aim of enrolling all household members of any age. The Namanolo site enrolled 449 participants, while the Ntaja site enrolled 537 participants. Symptoms consistent with malaria and a positive rapid diagnostic test were used to diagnose uncomplicated malaria. The follow-up period was approximately 12 months, with the Balaka site concluding in June 2021 and the Machinga site ending in March 2022. Whole blood samples were collected from participants diagnosed with uncomplicated malaria then mixed with RNAprotect, a chemical RNA preservative, prior to freezing. Frozen samples were shipped to the University of Maryland Baltimore for DNA extraction, processing, and parasite genome sequencing.

4.3.2 **DNA Extraction and Genomic Sequencing**

Frozen isolates from each study were shipped to the University of Maryland Baltimore for DNA extraction, processing, and parasite genome sequencing. DNA was extracted from isolates from time points 1 and 3 using the QIAamp DNA Blood Mini Kit (Valencia, CA, USA) according to the manufacturer's protocol. DNA from frozen red blood cell pellets collected from time point 2 were extracted using a previously published, in-house extraction method [231]. Selective whole genome amplification was performed to enrich for parasite DNA. Extracted DNA from time point 2 was enriched after purification of DNA by size-exclusion filtering and subsequent selective whole genome amplification (sWGA) as described previously [citation]. DNA from time point 3 was purified using via

AMPure XP beads followed by human-DNA digestion with the McrBC enzyme [269] followed by sWGA [233].

4.3.3 Whole Genome Sequencing

Sequencing libraries were generated from extracted DNA as described previously [177, 234]. DNA was fragmented via acoustic shearing and libraries were then prepared using a modified workflow based on the KAPA Library Preparation Kit, incorporating bead-based purification and quantification via AMPure XP beads and qPCR, respectively. Equimolar pooling was followed by high-throughput sequencing using the Illumina HiSeq 2500 or 4000 or NovaSeq 6000 platforms (Illumina, San Diego, CA) to produce 150 bp paired-end reads suitable for whole-genome analysis.

4.3.4 Sequence Data Processing

To ensure the quality and reliability of genomic analyses, whole-genome sequence data from *P. falciparum* infections were processed using a structured, multi-step pipeline as described previously [177, 234, 270]. Initially, raw sequencing reads were aligned to the human GRCh38 reference genome [271] to remove host DNA reads. The unmapped reads were subsequently aligned to the *P. falciparum* 3D7 reference genome (PlasmoDB version 44) using the Bowtie2 aligner [272]. The resulting alignment files were further refined according to the Genome Analysis Toolkit (GATK) best practices, which included base quality score recalibration and duplicate marking, to produce high confidence read alignments [273, 274]. Depth of coverage across the genome was assessed using Bedtools [275], and custom Python scripts were used to quantify the proportion of each isolate achieving various genome coverage thresholds (e.g., greater than 5×, 10×, 25×, and 100×). Genomic variant call format (GVCF) files for each sample were created by Haplotype

Caller followed by joint SNP calling between all samples using GATK [273, 274]. To retain only high-confidence variants, a minimum read depth of five was required and a series of hard filtering steps were applied. Variants were excluded if they did not meet the following conditions: read depth > five, quality by depth of at least two, Fisher Strand score of 60.0 or lower, root mean square of mapping quality of at least 30.0, mapping quality rank sum score of −12.5 or higher, read position rank sum score of −8.0 or higher, variant quality score of at least 50.0, and a symmetric odds ratio of strand bias of nine or less. Only biallelic single-nucleotide polymorphisms with a minimum allele frequency of 1% were retained for analysis. To reduce biases associated with missing genotype data, a stepwise filtering procedure was used. This involved alternating rounds of filtering samples and then variants based on missingness, beginning with permissive thresholds (e.g., excluding samples and variants with more than 85% missing data) and incrementally tightening the criteria. This process was repeated iteratively, first filtering samples then sites, until the final dataset included only those samples with no more than 30% missing genotypes and variants with no more than 20% missing samples. Sites with multi-allelic variants or within short tandem repeats [276] were also removed.

After variant filtration, F_{WS}, a measure of within host diversity, was calculated for each sample to distinguish monoclonal from polyclonal infections [277]. For each genome, F_{WS} was calculated using the R package, moimix [278] and the empirical distribution of the resulting F_{WS} values was examined using kernel density estimation. Based on the empirical distribution of F_{WS}, a local minimum inflection point was identified as threshold that delineates monoclonal from polyclonal infections. These monoclonal samples were then phased using DEploid [279] to resolve haplotypes and missing genotype calls were imputed using Beagle [280, 281] with default window size (40 cM), three cM overlap, and a 3D7 recombination rate map. The phased and imputed monoclonal samples were then used to construct a high-confidence haplotype reference panel, which served as input for the deconvolution of polyclonal infections using DEploid-IBD [282]. Haplotypes were retained from infections where the predominant clone accounted for more than 70% of the within-host parasite population and was at least three times more abundant than any minor clone. These strain-resolved genotypes were then re-imputed with Beagle to complete any remaining missing SNP calls, resulting in fully phased haplotypes [270, 283]. Finally, functional annotation of genetic variants was conducted using SnpEff [284]. Due to extreme variation in the subtelomeric regions resulting in poor sequencing quality, only the core genome [158] was retained followed by removal of any remaining genes encoding variant surface antigens in the of *pfemp1*, *rifin*, and *stevor* gene families.

4.3.5 Sliding Window Tajima's D and F_{ST}

To investigate temporal dynamics of balancing selection and genetic differentiation across the *P. falciparum* genome, SNP data from three longitudinal studies were analyzed using window-based approaches. Sliding windows were defined by fixed minimum SNP counts for time point 1, the time point with the smallest sample size, to maintain consistent statistical resolution across genomic regions with variable recombination rates and gene densities. Three window configurations were employed: 35 SNPs with a 7-SNP step, 50 SNPs with a 10-SNP step, and 80 SNPs with a 20-SNP step. Smaller windows increased the ability to detect localized signals of selection, whereas larger windows offered greater stability and sensitivity to broader genomic patterns. Using multiple overlapping windows also enhanced the ability to localize genes of interest within larger regions and helped confirm the robustness of signals.

To identify loci where host immune pressure may have resulted in genetic patterns consistent with balancing selection, I applied Tajima's *D*, a widely used neutrality test that compares two estimators of the population mutation rate: one is based on the number of segregating sites (Watterson's θ [285]) and the other is the average number of pairwise nucleotide differences (π [286]) [235]. Under the standard neutral model of evolution, these two metrics are expected to be equal, resulting in Tajima's *D* values near zero [235, 287]. However, deviations from this expectation can signal evolutionary forces acting on the population. A significantly positive Tajima's *D* indicates an excess of intermediate-frequency alleles relative to rare variants, consistent with balancing selection, where multiple alleles are maintained in the population due to frequency-dependent selection. In the context of malaria, such patterns may reflect immune-mediated selection at antigenic sites targeted by host immune responses [237]. By contrast, a negative Tajima's *D* suggests an excess of low-frequency alleles relative to intermediate-frequency alleles, which is consistent with positive directional selection [235]. However, it is important to note that demographic processes confound the interpretation of Tajima's *D*: population contraction or bottlenecks can lead to elevated Tajima's *D* values, mimicking signals of balancing selection, while population expansion can depress Tajima's *D*, leading to values more negative than expected under neutrality [287]. Tajima's *D* was calculated using the sci-kit allel Python package [288] on all variants in sliding windows of minimum fixed variant counts to capture localized signals of selection. Windows with Tajima's *D* values exceeding an absolute threshold of Tajima's $D > 2$ or a chromosome-specific empirical threshold (e.g., the top 1% Tajima's D scores among windows per chromosome), were flagged as candidates for balancing selection.

Although balancing selection is often detected in alleles with intermediate frequencies, changes in selective pressures, like variation in immune recognition among diverse antigen variants, could plausibly shift the relative frequencies of alleles across time to either high or low allele frequencies. If these frequency shifts are substantial, measures of genetic differentiation, e.g., F_{ST} [159, 236], combined with Tajima's D may improve detection of antigenic escape among loci that exhibit genetics signatures of balancing selection and temporal allele frequency shifts. F_{ST} quantifies the proportion of total genetic variance that can be attributed to between-subpopulation differences. With values ranging from zero to one, an F_{ST} of zero indicates no genetic differentiation between subpopulations, while an F_{ST} of one reflects complete differentiation, where subpopulations are fixed for distinct alleles. This analysis aimed to identify genes falling within windows with high Tajima's D in at least one time points and which also showed significant F_{ST} between time points, as a potential signal of allele frequency turnover related to frequency-dependent selection.

Multiple approaches exist for estimating F_{ST}, with the Weir-Cockerham estimator being a popular approach as it compares variance of allele frequencies between subpopulations to those of the combined subpopulations to estimate F_{ST} [289]. However, the Weir-Cockerham estimator is sensitive to imbalanced sample sizes between comparator populations, potentially leading to inflated estimates of differentiation [241]. In contrast, Hudson's F_{ST} compares the average allele frequencies of the combined subpopulations relative to the difference between subpopulations, making it a robust alternative for comparator populations with imbalanced sample sizes [241]. In this study, Hudson's method to estimate F_{ST} was performed using scikit-allel [288] for each genomic window for all pairwise comparisons between the three time points (1 vs. 2, 2 vs. 3, 1 vs. 3). Permutation significance was estimated using 10,000 permutations. Genes within windows with consistently high Tajima's D values across all time points and genes with high Tajima's D values and significant F_{ST} were selected for further analysis.

4.3.6 **Identification of Haplotype Clusters and Informative SNPs**

To assess whether variation at specific SNPs may drive changes in the structure or distribution of haplotypes across different time points, genes were prioritized if they fell within genomic windows exhibiting consistently high Tajima's D across all three time points or if they showed elevated Tajima's D in one or two time points along with significant F_{ST} values, indicating temporal differentiation. The goal of this analysis was to identify the nonsynonymous SNPs (nsSNPs) most likely driving these signals of immune escape. For each gene meeting the above criteria, agglomerative hierarchical clustering was applied to each matrix of haplotypes excluding all but nsSNPs. Haplotypes were clustered using complete linkage and Hamming distance to group samples based on sequence similarity [290] and clustering was performed across a range of two to 6 clusters (k). Silhouette coefficients were calculated to assess clustering quality. The silhouette score quantifies the degree of separation between clusters by comparing intra-cluster similarity to inter-cluster dissimilarity with higher scores indicating more distinct and internally coherent groupings. The k value that maximized the silhouette score was retained as the optimal number of clusters for each gene and time point combination. Genes with fewer than three nsSNPs were excluded from this analysis to ensure sufficient data for both clustering and classification.

To identify the specific nsSNPs predictive of haplotype group membership, random forest classifiers were trained for each gene-timepoint combination using the nsSNP haplotype matrix as input and the assigned cluster labels as the response variable. Classifiers were trained with 3,000 decision trees, a maximum depth of three, and class weights adjusted to balance uneven cluster sizes. Model performance was assessed using out-of-bag (OOB) error/accuracy scoring. Gini importance scores were computed for each nsSNP to quantify its contribution to correct classification of haplotype clusters. Among all nsSNPs, the empirical cumulative distribution function (CDF) of Gini scores was used to identify a threshold delineating important from unimportant nsSNPs using the "elbow" method [257] which identifies the largest perpendicular distance between a line drawn from the lowest and highest importance data points in the empirical CDF. This method identifies a threshold based on the maximum curvature or "elbow" of the ranked data, above which nsSNPs are considered important for predicting haplotype clusters. SNPs with Gini importance above this threshold were retained for building discriminative haplotypes for each cluster.

4.3.7 Identification of Mode and Consensus Variants, and Putative Serotypes

Mode and consensus variants were identified for each cluster using only the discriminative nsSNPs selected by the random forest classifier. Reference and alternate alleles at these positions were translated into amino acid residues based on available SNP annotations. For each cluster, the mode variant was defined as the single most frequently observed complete amino acid haplotype across all individuals. In contrast, the consensus variant was generated by determining the most common amino acid residue at each discriminative position and concatenating these residues into a single sequence. This consensus variant represents a theoretical construct that reflects the most frequent residue at each site but does not necessarily correspond to any observed haplotype in the dataset. Putative serotypes were subsequently inferred based on the structure and composition of both the mode and consensus variants within each cluster. Finally, haplotype frequencies and their temporal distributions were calculated and visualized for each cluster to assess representation over time.

4.4 **Results**

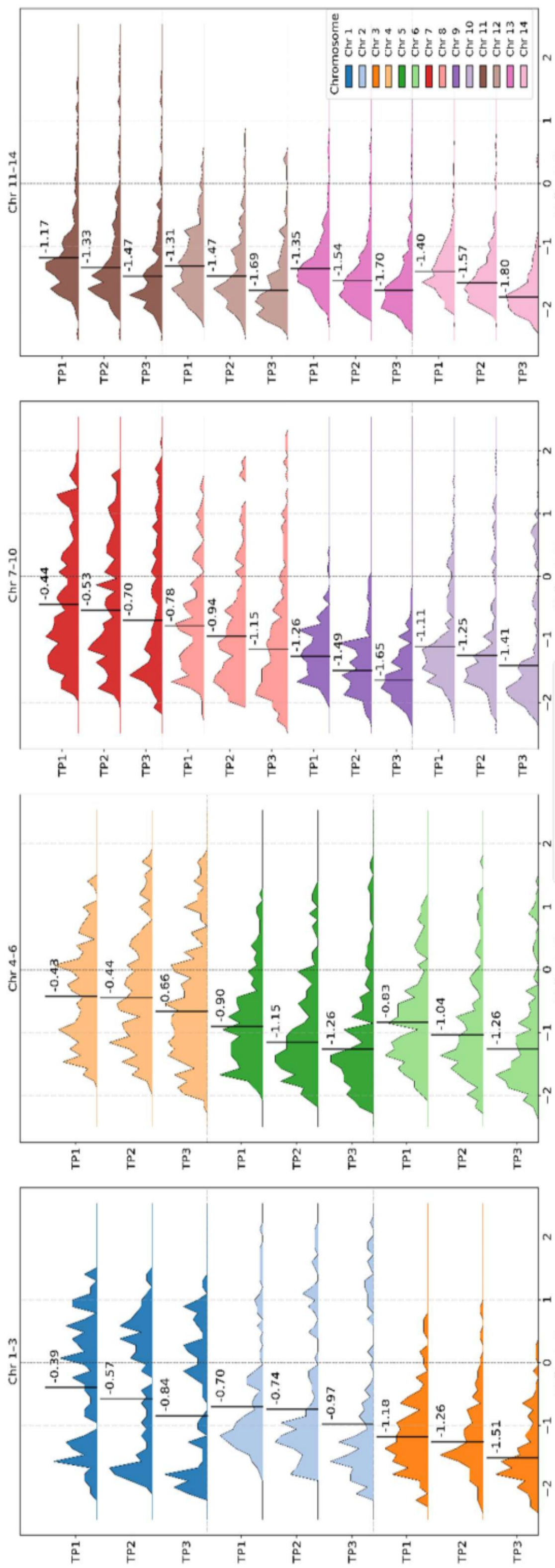
4.4.1 **Sequence Data Processing**

To examine the temporal dynamics of balancing selection and genetic differentiation in *P. falciparum*, a total of 288 *P. falciparum* isolates collected from study participants with clinical malaria across three time-points were subjected to whole genome sequencing, resulting in 900,357 raw SNPs. After applying quality control filters to remove low-confidence sites and variants with a minor allele frequency less than 0.01, 96,956 high-quality SNPs remained. Sample and site missingness thresholds were then applied, reducing the dataset to 66,592 SNPs across 244 samples. Within-host fixation index (F_{WS}) was calculated to assess clonality. The empirical distribution of F_{WS} among all isolates revealed a local minimum close to 1 at 0.899; accordingly, $F_{WS} > 0.9$ was selected to define monoclonality aligning with the observed separation in density. Among the 244 samples, 122 were classified as monoclonal and 122 as polyclonal. Monoclonal samples were phased, and missing calls were imputed, yielding 65,595 biallelic haplotypes. Resolved haplotypes were also generated for 16 polyclonal samples followed by imputation, resulting in a total of 138 complete haplotypes. To ensure genomic reliability, only variants from the core genome were retained, removing genes located in subtelomeric regions and *pfemp1*, *rifin*, and *stevor* gene families and principal component analysis was performed to identify and exclude outliers. The final dataset included 114 samples, with 22 from time point 1, 34 from time point 2, and 58 from time point 3, and a total of 63,665 high-confidence SNPs used for downstream analyses.

4.4.2 **Distribution of Tajima's *D* Across Chromosomes and Time Points**

The distribution of Tajima's *D* values across sliding genomic window was consistently centered below zero for all chromosomes and time points, with full-chromosome values ranging from −0.39 to −1.80 (Figure 4.1). This consistent genome-wide pattern wherein the Tajima's *D* values of all 14 full chromosome windows are consistently negative (though not to the same extent) across each time point, and the distribution of fixed-SNP windows from each chromosome and time point shifting toward negative Tajima's *D*, reflects an excess of rare alleles across each chromosome. To account for global shifts in the distribution of Tajima's *D* that may result from population demographic processes, such as recent expansion, or from systematic technical factors, such as sequencing artifacts, an empirical thresholding approach was applied. Specifically, high-Tajima's *D* windows were defined as those falling within the top 1% of values within each chromosome. This method adjusts for chromosome-specific baselines and facilitates the identification of genomic regions with relatively elevated Tajima's *D* values, even when the absolute magnitude of those values is modest due to global shifts in the background site frequency spectrum.

Figure 4.1: Ridgeline plot of density distributions of Tajima's *D* per chromosome per time point (TP1, TP2, TP3) among sliding genomic windows consisting of 50 SNPs with 10-SNP steps. The vertical black line and corresponding annotated number represents the Tajima's *D* value when all SNPs from the entire chromosome are included in the calculation of Tajima's *D*.


Chr 1-3
Chr 4-6
Chr 7-10
Chr 11-14
Tajima's D
Full Chromosome as Window
Chromosome

4.4.3 **Comparing Absolute and Empirical Thresholds for High Tajima's *D***

Using an absolute threshold of Tajima's $D > 2$ identified high-Tajima's *D* windows on a limited subset of chromosomes, specifically chromosomes 2, 4, 5, 7, 8, 10, and 11, while no windows with high Tajima's *D* were observed on the remaining chromosomes. In contrast, applying an empirical threshold based on the top 1% of Tajima's *D* values per chromosome is by definition able to identify high Tajima's *D* windows on all chromosomes.

When comparing overlap across time points, the absolute threshold approach identified five stretches of contiguous high-Tajima's *D* windows shared among all time points, six shared by two of three time points, and five found only in a single time point, while the top 1% method identified 21 stretches of contiguous high-Tajima's *D* windows shared among all time points, eight shared by two of three time points, and three found only in a single time point. All high-Tajima's *D* regions identified by the absolute threshold were also captured by the empirical threshold. Among the five stretches detected only in a single time point by the absolute threshold method, three were also identified by the empirical threshold method and were also detected at all time points. Figure 4.2 illustrates this comparison on chromosome five, where the windows with high Tajima's *D* identified by the absolute threshold (panel a) are also detected by the top 1% empirical method (panel b), which additionally reveals regions of elevated Tajima's *D* shared across all three time points. These two approaches yielded overlapping but distinct sets of regions, reflecting differences in the number and distribution of high-Tajima's *D* windows across the genome.

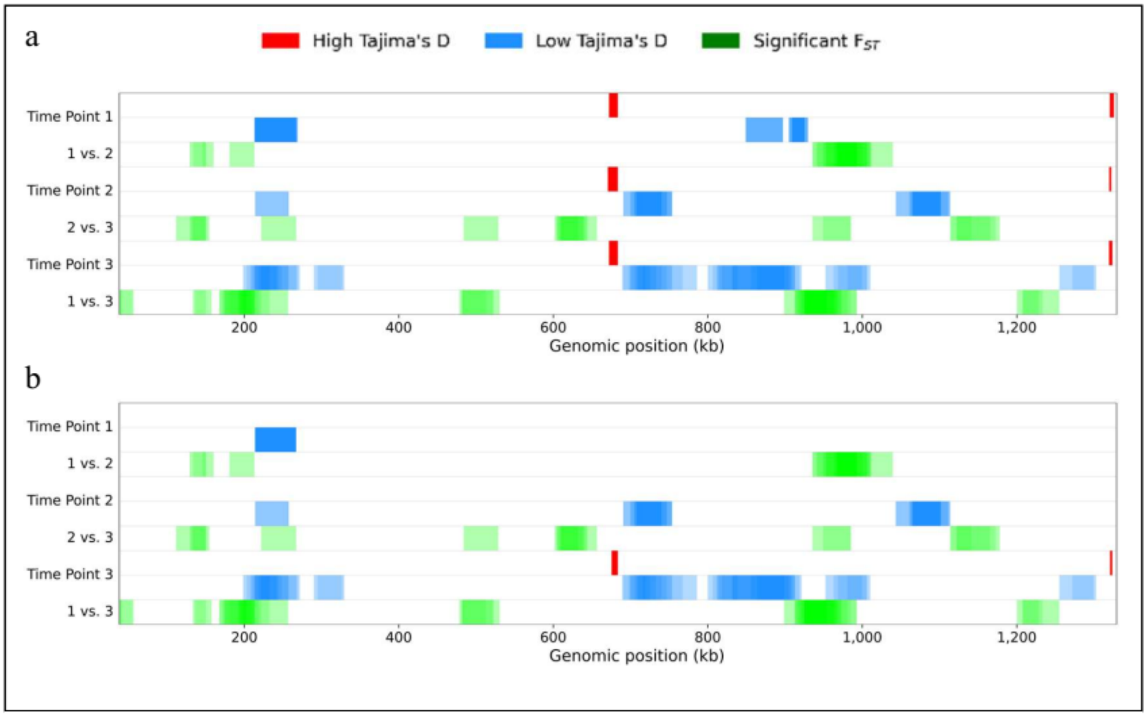


Figure 4.2: Comparison of selection signals on chromosome 5 using absolute and empirical Tajima's *D* thresholds in SNP-based windows. Panel (a) shows results using an absolute threshold of Tajima's $D > 2$ to define high-Tajima's *D* windows, while panel (b) uses an empirical threshold based on the top 1% of Tajima's *D* values per chromosome and time point. Each row represents a time point or pairwise time point comparison. Heatmaps displaying with light to dark colors, the proportion of windows at each genomic position that meet one of three criteria: high Tajima's *D* (red), low Tajima's $D < -2$ (blue), or significant F_{ST} based on permutation testing (green). Color intensity reflects the proportion of overlapping windows meeting each criterion (i.e., darker color = more overlapping windows), with white indicating no signal at that position. The empirical threshold captures all high-Tajima's *D* regions identified by the absolute method and additionally reveals regions shared across multiple time points.

4.4.4 Tajima's *D* and F_{ST}: Joint Signals of Balancing Selection

Consistency across time points was used as a filtering criterion to enrich for loci showing stable signals of balancing selection, reducing the likelihood that elevated values reflected random processes that can influence allele frequencies. Genes in these regions with consistent high Tajima's *D* across all time points, included 21 and 167 genes when employing the Tajima's $D > 2$ (Table 4.1) and top 1% of Tajima's *D* thresholds, respectively. The union across time points included 69 and 317 genes, respectively. Of the 21 genes with consistent signals of balancing selection, 14 (67%) are of unknown function, and five are annotated as "*Plasmodium* exported protein."

To prioritize genes from this list of 21, the intersection of genes identified through two additional approaches: a genomic association study of protected phenotypes [177], and an integrative omics analysis that identified genes with high antigenic potential [291]. Each of these methods and this study converged on PF3D7_0710200, annotated as "Conserved *Plasmodium* Protein, unknown function." Given this overlap, Pf3D7_0710200 was selected as a proof-of-concept candidate for downstream analysis, including investigation of haplotype structure and identification of high-importance nonsynonymous SNPs. Figure 4.3 displays the selection windows across chromosome 7 corresponding to this analysis.

4.4.5 Haplotype Clusters and Informative SNPs

Across all time points, PF3D7_0710200 contained 69 variable nonsynonymous SNPs (nsSNPs) based on variant annotation. Agglomerative clustering identified two clusters and the random forest classifier predicting cluster membership achieved a high out-of-bag (OOB) accuracy of 91.2% (Figure 4.4). Gini importance scores were calculated for each nsSNP. For all time points combined, importance above the determined significance threshold (0.02 Gini importance), yielded a set of 9 high-importance nsSNPs. Figure 4.4 shows the Gini importance and major allele frequency (MAF) for each nsSNP. Five of the nine most influential nsSNPs are in close proximity in a 125-bp stretch (genomic positions 470,356–470,480), corresponding to amino acid residues 2418, 2426, 2427, 2441, and 2459. A second group in close proximity spans from genomic positions 469,224 to 469,357 corresponding to amino acid residues 2040, 2082, and 2085. SNPs in these two groups ranked the highest in importance but also exhibited intermediate MAFs, a pattern consistent with balancing selection.

Table 4.1: Genes in regions with consistently high Tajima's D across all time points

Gene ID	Gene Name	High Tajima's D Total TP	Significant F_{ST} Total TP
PF3D7_0220900	*Plasmodium* Exported Protein, unknown function, pseudogene	3	2
PF3D7_0221000	*Plasmodium* Exported Protein, unknown function	3	2
PF3D7_0221100	*Plasmodium* Exported Protein, unknown function, pseudogene	3	2
PF3D7_0701900	*Plasmodium* Exported Protein, unknown function	3	0
PF3D7_0702000	*Plasmodium* Exported Protein (Hyp12), unknown function	3	0
PF3D7_0710200	Conserved *Plasmodium* Protein, unknown function	3	2
PF3D7_1004600	Conserved *Plasmodium* Membrane Protein, unknown function	3	2
PF3D7_1004700	Conserved Protein, unknown function	3	2
PF3D7_1004800	Adp, Atp Carrier Protein 2	3	2
PF3D7_1004900	Conserved Protein, unknown function	3	2
PF3D7_1005000	Methionine--tRNA Ligase, putative	3	2
PF3D7_1005100	U3 Small Nucleolar RNA-associated Protein 25, putative	3	1
PF3D7_1005200	Conserved *Plasmodium* Protein, unknown function	3	1
PF3D7_1005300	Conserved *Plasmodium* Protein, unknown function	3	1
PF3D7_1005400	Conserved *Plasmodium* Protein, unknown function	3	1
PF3D7_1005500	Regulator Of Nonsense Transcripts 1, putative	3	1
PF3D7_1005600	DnaJ Protein, putative	3	1
PF3D7_1005700	Peptidase, putative	3	1
PF3D7_1005800	Conserved *Plasmodium* Membrane Protein, unknown function	3	1
PF3D7_1149400	*Plasmodium* Exported Protein, unknown function	3	1
PF3D7_1149500	Ring-infected Erythrocyte Surface Antigen 2, pseudogene	3	1

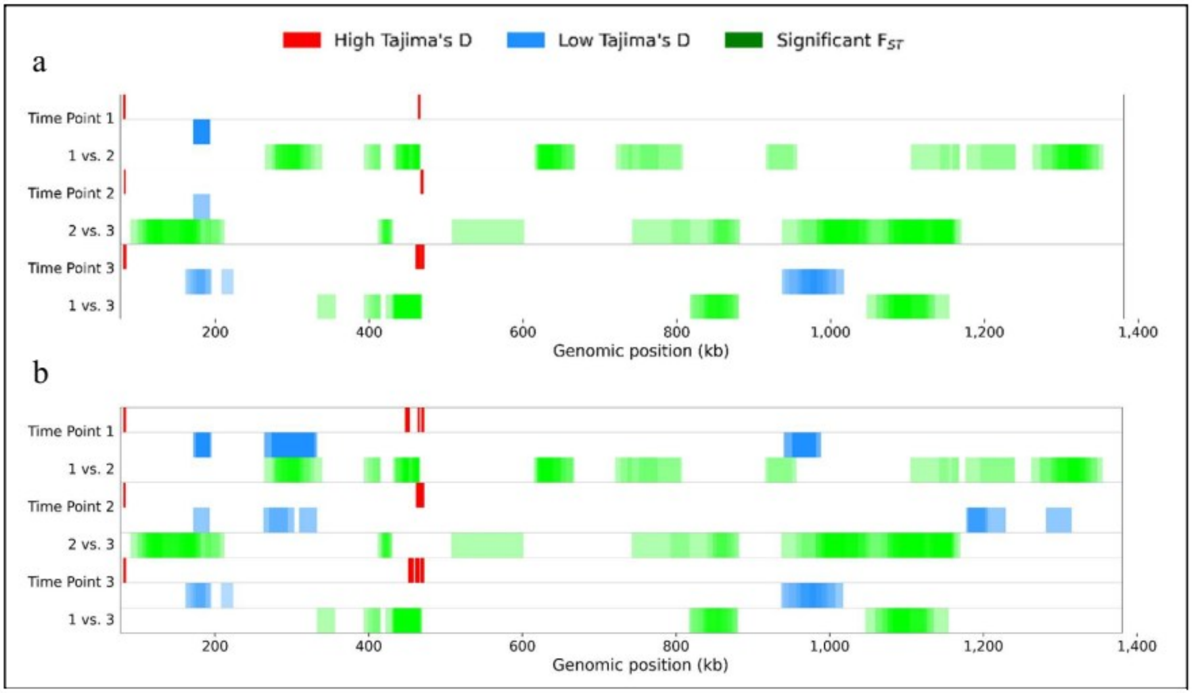
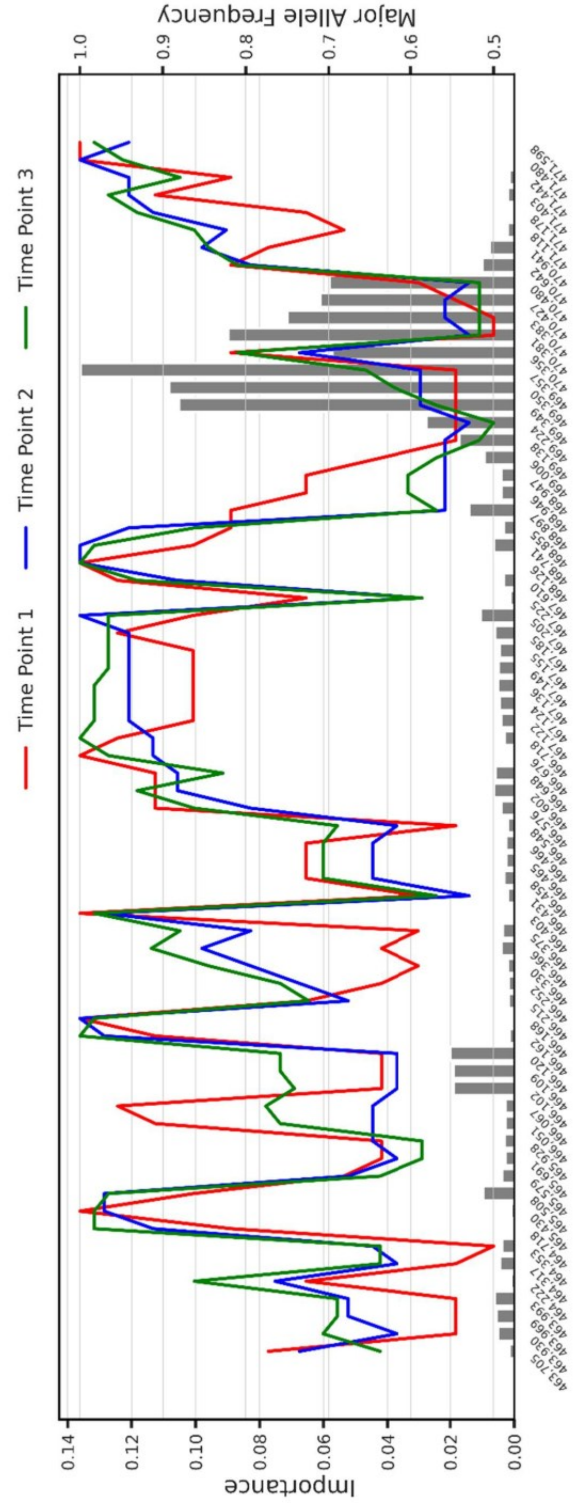


Figure 4.3: Comparison of selection signals on chromosome 7 using absolute and empirical Tajima's *D* thresholds in SNP-based windows. Panel (a) shows results using an absolute threshold of Tajima's $D > 2$ to define high-Tajima's *D* windows, while panel (b) uses an empirical threshold based on the top 1% of Tajima's *D* values per chromosome and time point. Each row represents a time point or pairwise time point comparison. Heatmaps displaying with light to dark colors, the proportion of windows at each genomic position that meet one of three criteria: high Tajima's *D* (red), low Tajima's $D < -2$ (blue), or significant F_{ST} based on permutation testing (green). Color intensity reflects the proportion of overlapping windows meeting each criterion (i.e., darker color = more overlapping windows), with white indicating no signal at that position. The empirical threshold captures all high-Tajima's *D* regions identified by the absolute method and additionally reveals regions shared across multiple time points.

Figure 4.4: Nonsynonymous SNP importance and major allele frequency across PF3D7_0710200. Each bar represents the Gini importance of a nsSNP for classifying haplotype cluster membership based on a random forest model trained across all three time points together. SNPs are ordered by genomic position along the x-axis. Gray bars (left y-axis) indicate the relative contribution of each nsSNP to the classifier's accuracy, as measured by Gini importance. Colored lines (right y-axis) denote the major allele frequency (MAF) of each nsSNP within time point 1 (red), time point 2 (blue), and time point 3 (green). nsSNPs with both high classification importance and intermediate MAFs highlight are consistent with signals of balancing selection.


Time Point 1
Time Point 2
Time Point 3
Importance
Major Allele Frequency

4.4.6 **Haplotype Clusters and Informative SNPs**

Across all time points, PF3D7_0710200 contained 69 variable nonsynonymous SNPs (nsSNPs) based on variant annotation. Agglomerative clustering identified two clusters and the random forest classifier predicting cluster membership achieved a high out-of-bag (OOB) accuracy of 91.2% (Figure 4.4). Gini importance scores were calculated for each nsSNP. For all time points combined, importance above the determined significance threshold (0.02 Gini importance), yielded a set of 9 high-importance nsSNPs. Figure 4.4 shows the Gini importance and major allele frequency (MAF) for each nsSNP. Five of the nine most influential nsSNPs are in close proximity in a 125-bp stretch (genomic positions 470,356–470,480), corresponding to amino acid residues 2418, 2426, 2427, 2441, and 2459. A second group in close proximity spans from genomic positions 469,224 to 469,357 corresponding to amino acid residues 2040, 2082, and 2085. SNPs in these two groups ranked the highest in importance but also exhibited intermediate MAFs, a pattern consistent with balancing selection.

These 9 high-importance SNPs were then used to reconstruct a reduced haplotype matrix across all samples from the combined data set. Agglomerative hierarchical clustering was applied to this reduced matrix. Silhouette analysis supported the selection of four haplotype clusters and for each cluster the resulting consensus sequences were 9-amino-acid in length and were unique to each cluster (Table 4.2). The modal sequences of each cluster were variable in only 8 of the 9 amino acids, though each cluster varied substantially in their temporal distribution. The frequency of each cluster shows that clusters 1 and 2 were the most common overall, observed among 31.6% and 43% of all samples across all three time points (Figure 4.5). In contrast, clusters 3 and 4 were less common with a combined observed percentage of 25.4% (9.6% and 15.4%), though all clusters maintained relatively stable frequencies over time.

Table 4.2: Variant cluster frequencies. Clusters were defined based on haplotype similarity across nine amino acid positions (1006, 2040, 2082, 2085, 2418, 2426, 2427, 2441, and 2459). For each cluster, the modal sequence (most frequent haplotype) and consensus sequence (most frequent amino acid at each position) are shown. The number and percentage of samples belonging to each cluster are provided for the full dataset and for each time point (T1, T2, T3).

Cluster	Modal Sequence	Consensus Sequence	Sample Size (n, %)	T1 (n, %)	T2 (n, %)	T3 (n, %)
1	ENSEKYTNT	ENSEKYTNT	36 (31.6%)	7 (31.8%)	12 (35.3%)	17 (29.3%)
2	EKNQEFSKI	EKNQEFSKI	49 (43.0%)	8 (36.4%)	17 (50.0%)	24 (41.4%)
3	ENSEEFSKI	KNSEEFSKI	11 (9.6%)	3 (13.6%)	2 (5.9%)	6 (10.3%)
4	EKNQEYTNT	EKNQEYTNT	18 (15.8%)	4 (18.2%)	3 (8.8%)	11 (19.0%)

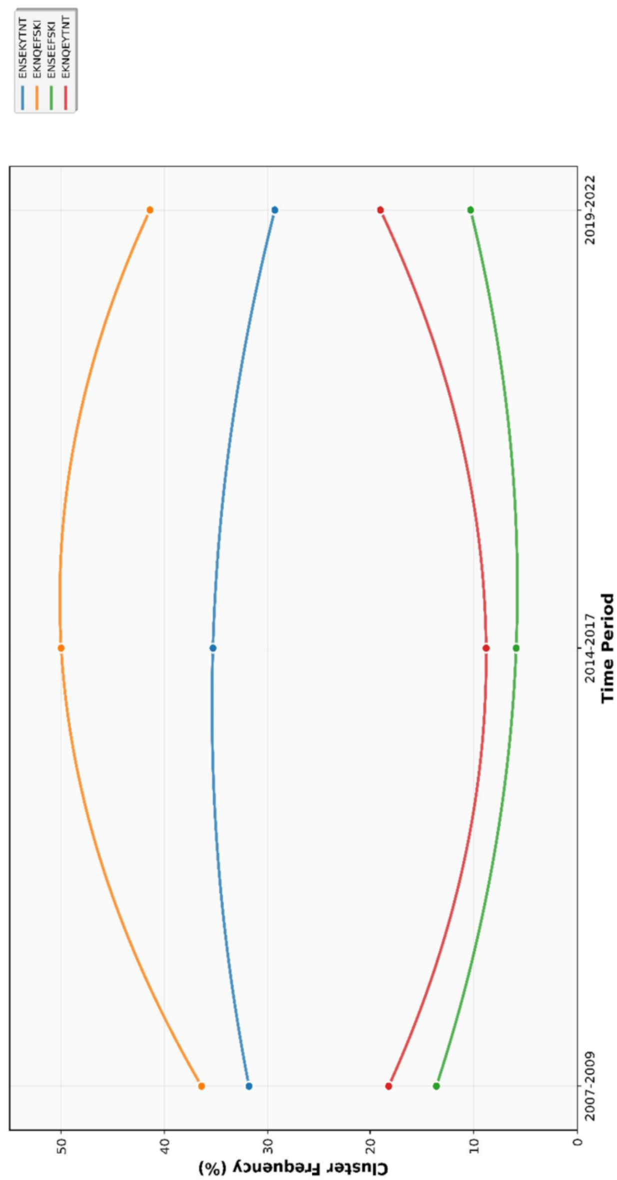


Figure 4.5. Haplotype cluster frequency (%) dynamics over time. Cluster frequencies (%) of across three time periods (2007-2009, 2014-2017, 2019-2022). Each color represents a unique haplotype cluster. Data points indicate actual observed frequency percentages at each cluster, while the legend shows the modal cluster sequences for identification of specific haplotypes and their trajectories over the 15-year period (ENSEKYTNT = cluster 1, EKNQEFSKI = cluster 2, ENSEEFSKI = cluster 3, EKNQEYTNT = cluster 4).

4.5 **Discussion**

A temporally informed genome-wide scan was used to identify loci under balancing selection and genetic differentiation in *P. falciparum* across three time points in Southern Malawi. Chromosome-wide estimates of Tajima's *D* were consistently negative across all chromosomes and time points, a pattern suggesting historic population expansion or bottleneck. However, localized regions showed elevated Tajima's *D*, identifying 21 and 167 genes across all time points using absolute and empirical thresholds, respectively. Among those 21, 19 were in regions exhibiting significant F_{ST}, suggesting a pattern of signals of population differentiation occurring concomitantly with signals of balancing selection. Among the 21 genes, the PF3D7_0710200 gene, which encodes a conserved protein with unknown function, was selected for further characterization. Hierarchical clustering and random forest modelling of nsSNP-based haplotypes revealed four distinct haplotype groups, each characterized by unique 9-amino-acid-long consensus sequences which may represent targets of allele specific immune response analogous to serotypes. Of these four haplotype clusters, the three most common (cluster 1: 31.6%; cluster 2: 43%; cluster 4: 15.8%) together account for approximately over 90% of all sequences in this study over the 15-year period.

In comparison to Pf3D7_0710200, numerous well-characterized vaccine candidates, PfAMA1, PfMSP1, PfMSP3, PfGLURP and PfTRAP, have exhibited elevated Tajima's *D* previously [16, 19, 180, 203, 237, 292-294]. In this study, each was identified by the empirical chromosome-specific threshold, though none of these candidate antigens were identified by the absolute threshold of $D > 2$, suggesting demographic noise may mask their balancing-selection signals. This mirrors previous work, where researchers often report the most extreme genome-wide values instead of relying on a fixed absolute threshold (e.g., $D > 2$) [203, 237, 292]. Other major antigens (PfCSP, PfEBA-175, PfSERA5) lacked detectable signals using both thresholds, which may be a result of filtering short tandem repeats. In contrast, PF3D7_0710200 consistently surpassed chromosome-specific and absolute D thresholds at all time points and exhibited significant temporal F_{ST}, suggesting possible strong, immune-driven selection signature, even on a background of demographic noise, making it a candidate for further characterization for its suitability as a vaccine target.

Located on chromosome 7, PF3D7_0710200 encodes a 2,910-amino acid protein and exhibits strong balancing-selection signals in both Mali [295] and Malawi (this study). Insertional mutagenesis studies suggest that this protein is essential for blood stage growth and survival [296]. Proteomic surveys detect PF3D7_0710200 expression across multiple life-cycle stages, including sporozoites, oocysts, rings, trophozoites, schizonts, and gametocytes [91, 297-304], and *in silico* analyses predict it to be surface-exposed [304]. A previous genome-wide screen flagged PF3D7_0710200 among 25 genes whose allelic turnover was associated with allele-specific natural immunity [177]. An integrative omics approach using positive unlabeled machine learning analyzed proteomic, structural, functional, immunological, genomic, and transcriptomic data to rank order *P. falciparum* [291] and *P. vivax* [305] antigens by antigenicity with the former predicting a set of 200 antigens, of which PF3D7_0710200 was ranked highly for antigenicity. Though PF3D7_0710200 is diverse, 4 major serotype-like haplotypes in Malawi were identified potentially reducing the antigen variants necessary to cover the diversity observed over 15 years. In fact, only three would cover over 90% of all samples in this study. Further validation of this approach might include assessing both immunogenicity and haplotype-specific immune responses, by exposing Malawian sera to haplotype-variants expressed on protein [306-310] or peptide arrays [255, 311] and performing functional assays, including growth inhibition assays [312, 313] and inhibition of liver-stage development assays [314, 315].

A limitation of this study is the temporal resolution of sampling, which may be insufficient to detect allele frequency shifts driven by cyclical or time-variable immune selection. If the periodicity of allele frequency turnover does not align with the multi-year intervals between sampling points, such patterns could be missed, regardless of whether their dynamics unfold over shorter or longer timescales. For example, prior studies of the gene encoding PfAMA1 have shown that allele-specific immunity may be transient, with effects detectable within only weeks following infection and no clear association observed at longer intervals [16]. In the context of our study, where sampling points are separated by several years, such temporally restricted signals of immune selection may be missed, particularly when allele frequency turnover occurs on faster or asynchronous timescales. Additionally, demographic processes such as population expansion and contraction can influence allele frequency distributions in ways that mimic or obscure signals of balancing selection. Tajima's *D*, which assumes a constant population size under the standard neutral model, is known to be sensitive to these deviations [235, 287]. To mitigate this, I used both absolute and chromosome-specific empirical thresholds and focused on signals that were consistent across time points supplemented with allele frequency shifts detected by F_{ST}. Finally, multiallelic SNPs were excluded because current genotype imputation methods and reference panels are optimized for biallelic sites [280, 281], making accurate imputation of multiallelic loci infeasible. This exclusion limits our ability to assess some of the most polymorphic and potentially immunologically relevant SNPs in the genome. However, SNPs in linkage disequilibrium with multiallelic sites may still carry indirect signals of selection, allowing genes harboring such variants to be identified even when the multiallelic sites themselves are not analyzed.

This study presents a novel framework for identifying genetic determinants of antigenic escape using clinical isolates collected from temporally spaced studies. Although the isolates were derived from longitudinal studies, only one cross-sectional sample per individual was analyzed. By examining temporal patterns in Tajima's *D* using both conventional and chromosome-specific thresholds, the approach strengthens the ability to distinguish genuine signals of balancing selection from global (or chromosomal) population demographic noise. Consistent elevation of Tajima's *D* across multiple time points increases the reliability of selection inference, while empirical thresholding within chromosome-specific distributions improves detection of loci that might otherwise be overlooked. Measures of genetic differentiation, such as F_{ST}, provide an additional dimension by highlighting temporal shifts in allele frequencies that may signal changing immune pressures. This combination allows the identification of genes that show both long-term maintenance of allelic diversity and evidence of changing allele frequencies over time, providing a fuller picture of how immune selection acts on parasite populations. Further resolution is achieved through haplotype clustering and classification of key nonsynonymous variants, exemplified by the analysis of PF3D7_0710200, a large, conserved, broadly expressed, and functionally uncharacterized gene that consistently ranked highly across all analytic layers and has been independently predicted to be antigenic.

This analytical framework can be adapted to other endemic settings and vaccine trial cohorts to detect immune-driven allele frequency changes over time. By leveraging complementary signals from Tajima's D, F_{ST}, and haplotype structure, the approach enables the identification of both known and novel antigenic targets that may be overlooked by static, cross-sectional analyses. In conclusion, this study demonstrates the utility of temporally informed population genetic analyses for uncovering signatures of antigenic escape. The combined use of diversity metrics, differentiation measures, and machine learning-based haplotype classification provides a powerful toolset to prioritize vaccine candidates and deepen understanding of host-pathogen evolutionary dynamics.

Chapter 5

Conclusion and Discussion

5.1 **Aim 1**

Identify humoral immune responses directed against *P. falciparum* antigens that are associated with protection following malaria challenge in individuals immunized with a whole organism, irradiated sporozoite vaccine.

5.1.1 **Main Findings in this Study**

This investigation analyzed sera from malaria-naïve adults participating in three independent Phase I clinical trials of the PfSPZ Vaccine, comprising 29 individuals who were protected and 13 who were unprotected following CHMI. Using a high-density peptide microarray covering 3,918,353 unique overlapping 16-mer peptides across the entire *P. falciparum* NF54 proteome represents an agnostic and systematic assessment of vaccine-induced humoral responses against malaria parasites.

Broad but Non-Protective Serological Responses. The PfSPZ Vaccine elicited extensive IgG reactivity across the proteome, with 76,695 peptides from 4,292 proteins (18.2% of reactive peptides) showing significantly increased antibody responses after vaccination. This pattern of broad reactivity was identified across both protected and unprotected vaccinees. However, despite this extensive immunological engagement across a large portion of the parasite proteome, majority of these antibody responses showed no association with protection against subsequent malaria challenge. The results suggest that protection may depend on additional characteristics of the immune response including the quality, specificity, or functional characteristics of responses rather than their quantity.

Pre-existing and Post-vaccination Inverse Association Patterns with Protection. An unexpected finding was a consistent inverse relationship between antibody responses and protection, detectable even before vaccination and persisting throughout the study. At baseline, only 3 peptides showed higher reactivity in individuals who would later be protected compared to 71 peptides in those who would remain unprotected, representing a 24-fold difference. Since all participants were malaria-naïve, these pre-vaccination differences are unlikely to be attributed to prior *P. falciparum* exposure. The 71 peptides, derived from 61 proteins, were significantly enriched for nuclear-localized biological processes such as DNA repair and various histone modifications.

This inverse association pattern became even more pronounced following vaccination. Post-vaccination analysis revealed that merely 6 peptides were elevated in protected individuals versus 222 in unprotected participants, representing a 37-fold difference. The change in response (post- minus pre-vaccination) showed the most extreme imbalance, with only 2 peptides demonstrating greater increases in protected individuals compared to 23 in unprotected participants. The consistency of this inverse pattern across all timepoints was particularly surprising and suggests that higher antibody responses to many parasite antigens may represent individual-specific immunological signatures that include cross-reactive responses not derived from *P. falciparum* exposure but that recognize *P. falciparum* epitopes.

Failure of Predictive Modeling. Machine learning approaches failed to identify serological signatures predictive of protection. Random forest models were unable to achieve accuracies substantially better than the majority-class baseline for pre-vaccination, post-vaccination, and change. Elastic net logistic regression performed even worse, with

AUC-ROC values of 44.3%, 47.3%, and 58.7% for the respective data sets. Even a combined approach using random forest for feature selection followed by elastic net classification failed, yielding AUC-ROC accuracies consistent with chance. Agreement across multiple algorithms may suggest that, in this data set, linear peptide-specific IgG responses alone are insufficient to predict protection, and that protective immunity requires immunological components not captured by this analysis.

Circumsporozoite Protein as Primary Vaccine Target. PfCSP emerged as the most immunogenic protein, with 77 of its 386 tiled peptides (19.9%) showing significant increases after vaccination, the highest proportion among all proteins. Hierarchical clustering revealed immunological heterogeneity: 12 individuals formed a low-reactivity cluster for the central-repeat region (5 protected, 7 unprotected) while 30 showed high reactivity (24 protected, 6 unprotected). However, despite this incongruence and PfCSP being the target of current WHO-approved vaccines, no PfCSP peptides remained significantly associated with protection after multiple comparison correction.

Gene Ontology Insights into Protection Mechanisms. Functional enrichment analysis revealed biologically meaningful distinctions between protective and non-protective responses. Protection-associated proteins were significantly enriched in accessible cellular locations, including the cell surface, extracellular space, and membrane components, with molecular functions such as host cell surface binding (Rhoptry Neck Protein 3) and biological processes like microtubule cytoskeleton organization. In stark contrast, proteins inversely associated with protection showed extensive enrichment for typically inaccessible nuclear and intracellular functions including DNA replication, histone methylation, chromatin-associated structures, and nuclear complexes. This pattern suggests that protective antibodies may target parasite antigens that are accessible for neutralization, while non-protective responses recognize conserved intracellular epitopes. In the context of pre-vaccination responses to intracellular epitopes among the unprotected, this might suggest cross reactivity with preexposure to other similar pathogens. In contrast, in the context of post-vaccination responses the enrichment for intracellular peptide recognition may suggests preexposure cross reactivity or possibly exposure to lysed parasites from the vaccine, or both.

5.1.2 Strengths and Implications

Methodological Strengths. The use of a comprehensive peptide microarray covering 3,918,353 unique overlapping 16-mer peptides across the entire *P. falciparum* proteome enables an extensive, unbiased assessment of vaccine-induced humoral responses. This proteome-wide approach eliminated the selection bias inherent in targeted immunological assays, enabling the systematic evaluation of both known and previously uncharacterized antigens. The overlapping peptide design, with a 15-amino acid overlap, provided high-resolution mapping of antigenic regions while maintaining comprehensive coverage.

The use samples from CHMI trials provided a clearly defined protective phenotype that eliminated confounding variables present in field-based efficacy studies. The homologous challenge design ensured that protection assessment was standardized across all participants and, importantly, was not confounded by antigenic diversity. The inclusion of malaria-naïve volunteers was intended to mitigate the complexity of pre-existing immunity, which complicates interpretation in endemic populations. The integration of multiple analytical approaches, including multiple machine learning algorithms and bivariate statistical testing, provided findings across different methodological frameworks.

Implications for Understanding Protective Immunity. The finding that most differential antibody responses were inversely associated with protection, i.e., unprotected individuals had statistically significantly higher responses among these peptides than protected individuals, suggests that protective immunity may depend on other qualities of immune responses rather than solely on magnitude [129, 316]. The peptide microarray approach, while comprehensive in representing the entire parasite proteome, possesses inherent limitations that constrain the interpretation of results. The array design exclusively captures antibody responses to linear epitopes, completely excluding conformational and discontinuous epitopes that may be important for protective immunity. The failure of machine learning models to predict protection from linear peptide responses alone suggests that protective immunity likely involves complex interactions between multiple immune components, including conformational epitopes, antibody functionality, cellular immune responses, and innate mechanisms that are not captured by linear peptide arrays. The implications for vaccine development suggest that functional evaluation of immune responses should integrate multiple qualitative and quantitative measurements to capture the full spectrum of immunological mechanisms.

Implications for Vaccine Development Strategy. The predominant immunogenicity of circumsporozoite protein, coupled with its lack of association with protection after multiple comparison correction, suggests that the protection conveyed by the PfSPZ vaccine involves additional immune effectors beyond antibodies against linear epitopes or that multiple antigens might combine to result in a protective response or both. While PfCSP is an immunodominant antigen capable of eliciting strong antibody responses, these findings suggest that other antigens or immunological mechanisms may be involved in protective immunity. Implications for next-generation vaccine design might be that that strategies should diversify beyond PfCSP to include additional antigens or alternative approaches that engage alternate protective immune pathways.

Broader Implications for Infectious Disease Vaccines. The methodological approaches and findings from this research extend beyond malaria to other infectious diseases where correlates of protection remain elusive. The demonstration that comprehensive, unbiased immunological profiling can reveal unexpected patterns of immune association provides a framework for investigating vaccine responses against other complex pathogens. The finding that cross-reactive antibodies from prior exposures can influence vaccine outcomes suggests that immune history should be considered more systematically in vaccine development for all infectious diseases. This principle may be particularly relevant for pathogens that share structural or sequence homology with common environmental antigens or other infectious agents.

5.1.3 Limitations

Methodological Limitations of Peptide Array Technology. The peptide microarray approach, while comprehensive in representing the entire parasite proteome, possesses inherent technical limitations that constrain the interpretation of results. The array design exclusively captures antibody responses to linear epitopes, completely excluding conformational and discontinuous epitopes that may be important for protective immunity. Many of the most potent neutralizing antibodies against pathogens recognize complex three-dimensional structures that linear peptides cannot represent. [173, 175, 317]. This limitation is particularly relevant for malaria, where several protective monoclonal antibodies, including the recently identified CIS43 and L9 antibodies derived from protected PfSPZ Vaccine recipients, recognize conformational epitopes at the junction between the N-terminus and central repeat regions of circumsporozoite protein [100, 318]. The analysis was restricted to total IgG responses without differentiation of antibody subclasses, avidity, or functional characteristics. Different IgG subclasses exhibit distinct effector functions [127, 131] and subclasses may have contrasting roles. For example, higher ratios of IgG2 to IgG4 antibodies were associated with reduced risk of infection [135]. The measurement of antibody avidity, which reflects the strength of antigen-antibody binding and often correlates with protective capacity [319], was not assessed. Furthermore, functional antibody activities such as neutralization [320, 321], antibody-dependent cellular cytotoxicity [322], opsonization [323], or complement activation [319, 324] were not evaluated, despite their potential importance in protective immunity.

The focus on humoral immunity excluded components of the immune response that are likely essential for protection against malaria. Cellular immune responses, including CD8+ T cell responses that may be crucial for liver-stage immunity, and CD4+ T cell responses that provide help for antibody production and direct effector functions, were not assessed. Innate immune responses, including natural killer cell activity [118] and cytokine production [117], which may contribute to early parasite control and shape adaptive immunity , were similarly omitted from analysis.

Study Design and Sample Limitations. The study was limited by a relatively small sample size of 42 participants, which limited statistical power for detecting associations and training robust machine learning models. The class imbalance between protected and unprotected individuals, with only 13 unprotected participants compared to 29 protected, created additional analytical challenges and may have contributed to the poor performance of predictive models.

Analytical and Statistical Limitations. The analysis faced substantial multiple comparison challenges inherent in high-dimensional omics data, with nearly 4 million peptides tested simultaneously. Despite the application of false discovery rate correction, the massive number of comparisons may have obscured genuine but modest associations with protection. Conversely, the stringent correction may have eliminated potentially meaningful signals that could have provided insights into protective mechanisms.

The application of machine learning techniques to high-dimensional data with a small sample size presented inherent challenges that may have contributed to poor predictive performance. The curse of dimensionality, where the number of features vastly exceeds the number of observations, can lead to overfitting and poor generalization even with regularization techniques. The complex correlation structure among overlapping peptides from the same proteins may have violated assumptions of independence that underlie many statistical approaches and could have confounded the identification of truly informative features.

The analytical approach was primarily focused on bivariate associations and rank-based relationships, potentially missing complex interactions between immune responses that might be critical for protection. Covariate relationships between responses to different antigens, which could be important for protective immunity, were not systematically evaluated by the bivariate analysis, but were incorporated into the machine learning analyses.

5.1.4 Future Directions

Generalizability and External Validity Limitations. The findings are derived exclusively from the PfSPZ vaccine platform, which may elicit immune responses that differ substantially from those elicited by other vaccine approaches or natural infection. The whole sporozoite vaccine presents proteins expressed during the sporozoite, liver, and possibly blood stages of the parasite's life cycle, whereas subunit vaccines focus on selected antigens, and the immunological correlates identified may not apply to different vaccine strategies. The results may therefore have limited relevance for evaluating other malaria vaccine candidates or platforms.

Additional Functional Immune Profiling Technologies. Functional antibody assessments represent a critical advancement opportunity. Assays for antibody-dependent cellular cytotoxicity [325], complement activation [127], and opsonization [125] could provide mechanistic insights into protective responses. Comprehensive antibody characterization including subclass profiling [129, 132, 326] and avidity measurements [326, 327] could reveal qualities, in addition to magnitude, that delineate protective from unprotective responses.

Expanded Study Populations and Designs. Future studies would benefit from larger sample sizes providing adequate statistical power and more balanced groups of protected and unprotected individuals. Extension to malaria-endemic populations represents a step towards translating findings into deployment settings. Studies in previously exposed individuals could reveal how existing immunity influences vaccine responsiveness and whether correlates differ between naive and pre-exposed populations.

5.2 Aim 2

Identify signatures of antigenic escape in *P. falciparum* genes over time in southern Malawi.

5.2.1 Main Findings in this Study

Genomic Data Set and Identification of Selection Signatures. I successfully analyzed whole-genome sequence data from 114 *P. falciparum* clinical isolates collected across three time points spanning 15 years (2007-2022) in southern Malawi, yielding 63,665 high-confidence single nucleotide polymorphisms (SNPs) for population genetic analysis. The genome-wide Tajima's D distribution was consistently negative across all 14 chromosomes and time points (ranging from -0.39 to -1.80), indicating an overall excess of rare alleles consistent with historical population expansion. Using chromosome-specific empirical thresholds (top 1% of Tajima's D values) was more sensitive than absolute thresholds (Tajima's $D > 2$) for detecting signatures of balancing selection. The empirical approach identified candidate regions on all chromosomes, while the absolute threshold only detected signals on 7 of 14 chromosomes.

Using the absolute threshold approach, 21 genes were identified as exhibiting consistently high Tajima's D across all three time points, suggesting persistent balancing selection. Of these, 19 genes (90%) also showed significant temporal genetic differentiation (F_{ST}), indicating allele frequency shifts over time consistent with frequency-dependent immune selection. Notably, 67% of these genes have no known function and another 24% are annotated as "*Plasmodium* exported proteins," underscoring that the systematic, genome-wide approach revealed previously uncharacterized antigenic targets.

An in-depth analysis of PF3D7_0710200, a conserved protein of unknown function, served as proof of concept for this integrated approach. This gene contained 69 variable nonsynonymous SNPs (nsSNPs) and consistently exceeded both statistical thresholds across all time points. Machine learning approaches identified 9 high-importance nsSNPs that clustered in two genomic regions and exhibited intermediate allele frequencies characteristic of balancing selection. These nsSNPs defined four major haplotype clusters that remained stable over the 15-year study period, with the three most common clusters accounting for over 90% of all parasite isolates. This temporally informed, genome-wide framework successfully identified multiple genes that exhibition signals of immune-mediated balancing selection and highlights both known and novel antigens for future serotype-aware vaccine design.

5.2.2 Strengths and Implications

Methodological Innovation and Robustness. This study represents a methodological advance by integrating temporal genomic surveillance with population genetic- and machine learning-based analysis to detect immune-driven selection. To capture selection signals at varying scales, multiple overlapping window/step combinations were employed both to refine the localization of candidate regions and confirm the robustness of signals across scales. Rather than relying solely on absolute cutoffs (e.g., $D > 2$), this study applied a dual-threshold approach using chromosome-specific empirical cutoffs (top 1% of D-values per chromosome), ensuring that regions with elevated intermediate-frequency alleles were detected across all chromosomes regardless of baseline diversity levels. The 15-year temporal framework provided unique resolution that cross-sectional studies lack, enabling detection of allele frequency shifts indicative of frequency-dependent selection that would be missed in static analyses. Another advantage of applying this analysis over a long timeframe is the improved ability to distinguish alleles with persistently intermediate frequencies from those that only temporarily exhibit intermediate frequencies due to random fluctuations. While population genetic metrics and machine learning approaches have been applied in malaria genomics, this particular integration of temporal F_{ST} analysis with Tajima's D prioritization and both unsupervised and supervised learning for systematic antigen characterization represents a relatively unexplored analytical approach that addresses a major challenge in malaria population genetics where historical population demographic effects can confound neutrality tests.

Implications for Vaccine Development. The identification of PF3D7_0710200 and other novel candidates has immediate implications for rational vaccine design. The finding that only four major haplotype clusters accounted for over 90% of parasites across 15 years suggests that a limited number of antigenic variants may be sufficient to achieve broad coverage against natural parasite diversity. This contrasts with highly polymorphic antigens where dozens of variants may be required, making PF3D7_0710200 and similar targets more tractable for vaccine development.

Broader Scientific and Public Health Impact. This framework offers a generalizable approach for antigen discovery that can be applied to other pathogens exhibiting patterns of immune-driven selection. The discovery that many uncharacterized genes exhibit signatures of immune selection also opens new avenues for functional studies to understand parasite biology and host-pathogen interactions.

5.2.3 Limitations

Temporal Resolution and Sampling Constraints. While the 15-year timeframe provided valuable temporal depth, the multi-year intervals between sampling points may have been insufficient to detect allele frequency oscillations that occur on shorter timescales. Previous studies of AMA1 have demonstrated that allele-specific immunity can be transient, with effects detectable only within weeks following infection, suggesting that immune selection dynamics operating on monthly or seasonal cycles could have been missed entirely. The unequal sample sizes across time points (22, 34, and 58 samples for time points 1, 2, and 3, respectively) may have introduced bias in population genetic parameter estimates, particularly for the earliest time point where statistical power was most limited.

Confounding by Population Demographic Effects. The consistently negative Tajima's *D* values across all chromosomes indicate strong demographic effects that complicate the interpretation of selection signatures. To address this challenge, I employed several mitigation strategies: the empirical threshold approach was designed to increase sensitivity for detecting elevated Tajima's *D* signals against a background of genome-wide population expansion, while the requirement for consistent signals across multiple time points helped distinguish persistent balancing selection from transient random fluctuations. Additionally, the integration of temporal F_{ST} analysis provided an independent line of evidence for allele frequency shifts that could support frequency-dependent immune selection. However, despite these complementary approaches, the possibility remains that some candidate genes reflect demographic rather than selective processes. The study was also geographically constrained to southern Malawi, potentially limiting the generalizability of findings to other transmission settings with different population structures, transmission intensities, or demographic histories. While the loci detected by this approach may not differ greatly, the specific alleles and clusters of SNPs may differ between sites with various epidemiological differences.

Methodological and Technical Limitations. Several sampling and study design factors could have introduced systematic biases from the outset. Although all studies were conducted in southern Malawi, they were carried out in different districts (Blantyre, Chikwawa, and Balaka/Machinga), which have distinct demographic characteristics, including differences in elevation and urbanization, as well as potentially divergent parasite population structures that could confound temporal comparisons. However, the geographical distances spanning these districts are relatively small with the most distant health centres, Mfera and Ntaja, separated by approximately 95 miles. The exclusive focus on clinical malaria cases may have introduced selection bias toward more virulent parasite strains or individuals with less protective immunity, potentially skewing the representation of circulating genetic diversity compared to the broader parasite population that includes asymptomatic infections.

Technical limitations in sequencing and genome processing may have further affected data quality. Coverage biases favoring certain genomic regions and the use of the 3D7 reference genome may have systematically underrepresented variants that are highly divergent from the laboratory strain, potentially biasing against detection of the most evolutionarily distant and potentially immunologically relevant alleles. Genotyping errors arising from sequencing artifacts, misalignment in repetitive regions, and variant calling inaccuracies may have introduced noise that would generally bias results toward the null hypothesis, making genuine selection signals harder to detect. To mitigate these limitations, minimal read depths were employed, reducing the likelihood of sequencing artifacts. Additionally, short tandem repeats and highly variable genes and loci were excluded from the analysis. Low sequencing coverage is likely to disproportionately remove rare alleles compared to intermediate frequency alleles, creating a systematic bias toward detecting more common variants and potentially inflating Tajima's *D* values by reducing the number of segregating sites while having less impact on nucleotide diversity. The exclusion of multiallelic SNPs, although necessary due to current imputation limitations, removed some of the most polymorphic and potentially immunologically relevant sites from analysis. However, through linkage, adjacent SNPs may be sufficient to detect immunological relevant loci, and further fine mapping with targeted sequencing could elucidate relevant multiallelic SNPs. Filtering of short tandem repeats, subtelomeric regions, and highly variable antigenic genes families, which were excluded due to sequencing and alignment challenges, potentially missing important targets of immune selection. Employing imputation to fill missing genotype calls, which, while standard practice, could have introduced systematic errors particularly in highly diverse genomic regions or where population structure differs from the reference panel, potentially creating spurious associations or masking genuine selection signatures.

Downstream analytical limitations included the use of machine learning-based haplotype classification which, while innovative, relied on unsupervised clustering that may not reflect biologically meaningful antigenic groupings, and the functional significance of the identified discriminative SNPs remains unvalidated through experimental approaches.

Interpretive and Validation Constraints. The identification of genes under putative immune selection does not directly demonstrate their antigenicity or protective relevance, as elevated Tajima's D and temporal F_{ST} could result from other evolutionary pressures or technical artifacts. The study lacked experimental validation through immunological assays, functional studies, or structural analysis to confirm that the identified candidates are targets of immune recognition. The immunological characterization of these results is an important next step to further validate this approach. The heavy representation of genes with unknown function among the candidates, while highlighting potential novel targets, also reflects the current limitations in *P. falciparum* functional genomics and complicates the biological interpretation of findings. Additionally, the framework's ability to distinguish true balancing selection from complex demographic scenarios requires further validation in diverse epidemiological settings with known population histories and transmission dynamics.

5.2.4 **Future Directions**

The temporal population genomic framework developed in this study opens multiple avenues for advancing malaria antigen discovery and our understanding of parasite evolution under immune pressure. Building upon the methodological foundation established here, future research could benefit from focusing on four interconnected areas: expanding the temporal and spatial scope of genomic surveillance to capture immune selection dynamics across diverse epidemiological contexts; refining population genetics approaches using machine learning; validating evolutionary predictions through comprehensive experimental characterization of identified candidates; and translating findings into practical applications for vaccine development and malaria control strategies.

Enhanced Temporal and Geographic Sampling Strategies. Future studies could implement higher-resolution temporal sampling with monthly or quarterly collection intervals to capture short-term allele frequency oscillations that may reflect seasonal immune selection dynamics. Longitudinal cohort designs tracking the same individuals over multiple malaria episodes would provide valuable insights into within-host parasite dynamics and the relationship between immune memory and antigenic diversity. Expanding geographic coverage to include multiple transmission settings across sub-Saharan Africa, Southeast Asia, and South America could enable assessment of the generalizability of identified selection signatures and reveal region-specific patterns of immune selection. Balanced sampling designs with equal sample sizes across time points and the inclusion of both symptomatic and asymptomatic infections might also provide more representative estimates of population genetic parameters and capture the full spectrum of circulating parasite diversity.

Novel Machine Learning Approaches to Detect Balancing Selection. Developing machine learning-based approaches specifically designed to identify signatures of balancing selection could offer advantages over traditional population genetic statistics by distinguishing selection signals from complex demographic scenarios without requiring explicit demographic or other restrictive assumptions [328, 329]. Integration with epidemiological models that link transmission intensity, intervention history, and population structure may further refine the interpretation of temporal genetic patterns.

Methodological Advances and Technical Innovation. Advancing sequencing and bioinformatics approaches to substantially increase read depth and genome coverage, especially for asymptomatic infections, and complementary bioinformatics improvements to enable the resolution of polyclonal haplotypes and the inclusion of multiallelic SNPs, short tandem repeats, and subtelomeric regions would be valuable for capturing the full spectrum of antigenic diversity.

Experimental Validation and Functional Characterization. The most critical next step would involve experimental validation of identified candidates through comprehensive immunological assays. Protein microarrays [307, 308, 310], peptide arrays [255, 311], and ELISA-based [326] approaches using sera from malaria-exposed individuals could be employed to confirm that genes showing population genetic signatures of immune selection are indeed antigenic. Functional assays, including growth inhibition and inhibition of liver-stage development assays, which assess the functional impact of humoral immune responses against these targets, would provide evidence that the identified antigens have potential as vaccine candidates.

Translation to Vaccine Development and Public Health Applications. Long-term goals could include incorporating validated targets into next-generation malaria vaccine designs, potentially as components of multivalent vaccines that address antigenic diversity through inclusion of multiple protein variants. Field trials in diverse transmission settings might assess the real-world performance of vaccines targeting these evolutionarily validated antigens. The analytical framework could potentially be adapted for real-time genomic surveillance systems that monitor parasite evolution in response to vaccine deployment, enabling early detection of potential escape variants. Finally, extending this approach to other pathogens experiencing similar immune selection pressures might demonstrate the broader utility of temporal population genomics for infectious disease control and vaccine development.

5.3 **Summary**

Malaria vaccine development faces significant challenges due to the complex immune responses required for protection and the extensive genetic diversity that enables parasites to evade host immunity. The limited efficacy of current vaccines and the difficulty in identifying reliable correlates of protection highlight the need for new approaches that integrate immunological and evolutionary perspectives. This study was dedicated to addressing fundamental questions about which immune responses correlate with protection and how parasite evolution under immune pressure can inform rational vaccine design. Using proteome-wide peptide arrays, I characterized vaccine-induced immune responses in PfSPZ Vaccine recipients and discovered distinct immunological signatures between protected and unprotected individuals, revealing that protective immunity involves complex mechanisms that likely extend beyond linear peptide-binding antibody measurements. Through temporal population genomics analysis of 15 years of parasite surveillance data, I developed a framework that successfully identified genes under immune selection pressure, identifying 21 candidates exhibiting persistent balancing selection with the majority representing previously uncharacterized antigenic targets. The integration of these approaches demonstrated that effective vaccine development requires understanding both the immunological requirements for protection and the evolutionary constraints that maintain antigenic diversity over time. By establishing methodological frameworks for unbiased immune profiling and evolutionary validation of vaccine targets, I anticipate that these tools will enhance rational vaccine design efforts and accelerate the identification of next-generation malaria vaccine candidates, while also providing broadly applicable strategies for other complex pathogens where traditional vaccine development approaches have proven insufficient.

References

1. World Health Organization: World Malaria Report 2024. Geneva, 2024.

2. Plowe CV, Alonso P, Hoffman SL: The potential role of vaccines in the elimination of *falciparum* malaria and the eventual eradication of malaria. *J Infect Dis* 2009, 200:1646-1649.

3. Clyde DF, Most H, McCarthy VC, Vanderberg JP: Immunization of man against sporozoite-induced *falciparum* malaria. *American Journal of Medical Sciences* 1973, 266:169-177.

4. WHO recommends R21/Matrix-M vaccine for malaria prevention in updated advice on immunization https://www.who.int/news/item/02-10-2023-who-recommends-r21-matrix-m-vaccine-for-malaria-prevention-in-updated-advice-on-immunization

5. Datoo MS, Natama HM, Some A, Bellamy D, Traore O, Rouamba T, Tahita MC, Ido NFA, Yameogo P, Valia D, et al: Efficacy and immunogenicity of R21/Matrix-M vaccine against clinical malaria after 2 years' follow-up in children in Burkina Faso: a phase 1/2b randomised controlled trial. *Lancet Infect Dis* 2022, 22:1728-1736.

6. Datoo MS, Natama MH, Some A, Traore O, Rouamba T, Bellamy D, Yameogo P, Valia D, Tegneri M, Ouedraogo F, et al: Efficacy of a low-dose candidate malaria vaccine, R21 in adjuvant Matrix-M, with seasonal administration to children in Burkina Faso: a randomised controlled trial. *Lancet* 2021, 397:1809-1818.

7. Laurens MB: RTS,S/AS01 vaccine (Mosquirix): an overview. *Hum Vaccin Immunother* 2020, 16:480-489.

8. Davies DH, Duffy P, Bodmer JL, Felgner PL, Doolan DL: Large screen approaches to identify novel malaria vaccine candidates. *Vaccine* 2015, 33:7496-7505.

9. Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, et al: Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 2002, 419:498-511.

10. Barry AE, Schultz L, Buckee CO, Reeder JC: Contrasting population structures of the genes encoding ten leading vaccine-candidate antigens of the human malaria parasite, *Plasmodium falciparum*. *PLoS One* 2009, 4:e8497.

11. Reiling L, Richards JS, Fowkes FJ, Wilson DW, Chokejindachai W, Barry AE, Tham WH, Stubbs J, Langer C, Donelson J, et al: The *Plasmodium falciparum* erythrocyte invasion ligand Pfrh4 as a target of functional and protective human antibodies against malaria. *PLoS One* 2012, 7:e45253.

12. Langhorne J, Ndungu FM, Sponaas AM, Marsh K: Immunity to malaria: more questions than answers. *Nat Immunol* 2008, 9:725-732.

13. Barry AE, Arnott A: Strategies for designing and monitoring malaria vaccines targeting diverse antigens. *Front Immunol* 2014, 5:359.

14. Kusi KA, Manu EA, Manful Gwira T, Kyei-Baafour E, Dickson EK, Amponsah JA, Remarque EJ, Faber BW, Kocken CHM, Dodoo D, et al: Variations in the quality of malaria-specific antibodies with transmission intensity in a seasonal malaria transmission area of Northern Ghana. *PLoS One* 2017, 12:e0185303.

15. Doolan DL, Dobano C, Baird JK: Acquired immunity to malaria. *Clin Microbiol Rev* 2009, 22:13-36, Table of Contents.

16. Takala SL, Coulibaly D, Thera MA, Batchelor AH, Cummings MP, Escalante AA, Ouattara A, Traore K, Niangaly A, Djimde AA, et al: Extreme polymorphism in a vaccine antigen and risk of clinical malaria: implications for vaccine development. *Sci Transl Med* 2009, 1:2ra5.

17. Thera MA, Doumbo OK, Coulibaly D, Laurens MB, Ouattara A, Kone AK, Guindo AB, Traore K, Traore I, Kouriba B, et al: A field trial to assess a blood-stage malaria vaccine. *N Engl J Med* 2011, 365:1004-1013.

18. Neafsey DE, Juraska M, Bedford T, Benkeser D, Valim C, Griggs A, Lievens M, Abdulla S, Adjei S, Agbenyega T, et al: Genetic Diversity and Protective Efficacy of the RTS,S/AS01 Malaria Vaccine. *N Engl J Med* 2015, 373:2025-2037.

19. Takala SL, Coulibaly D, Thera MA, Dicko A, Smith DL, Guindo AB, Kone AK, Traore K, Ouattara A, Djimde AA, et al: Dynamics of polymorphism in a malaria vaccine antigen at a vaccine-testing site in Mali. *PLoS Med* 2007, 4:e93.

20. Manson-Bahr P: The Jubilee of Sir Patrick Manson (1878-1938): A Tribute to his Work on the Malaria Problem. *Postgrad Med J* 1938, 14:345-357.

21. Organization WH: World Malaria Report 2021. Geneva, Switzerland: World Health Organization; 2021.

22. Bartoloni A, Zammarchi L: Clinical aspects of uncomplicated and severe malaria. *Mediterr J Hematol Infect Dis* 2012, 4:e2012026.

23. Barry A, Hansen D: Naturally acquired immunity to malaria. *Parasitology* 2016, 143:125-128.

24. Pinkevych M, Petravic J, Chelimo K, Kazura JW, Moormann AM, Davenport MP: The dynamics of naturally acquired immunity to *Plasmodium falciparum* infection. *PLoS Comput Biol* 2012, 8:e1002729.

25. Buchanan HD, Goodman CD, McFadden GI: Roles of the apicoplast across the life cycles of rodent and human malaria parasites. *J Eukaryot Microbiol* 2022, 69:e12947.

26. Venugopal K, Hentzschel F, Valkiunas G, Marti M: *Plasmodium* asexual growth and sexual development in the haematopoietic niche of the host. *Nat Rev Microbiol* 2020, 18:177-189.

27. Soulard V, Bosson-Vanga H, Lorthiois A, Roucher C, Franetich JF, Zanghi G, Bordessoulles M, Tefit M, Thellier M, Morosan S, et al: *Plasmodium falciparum* full life cycle and *Plasmodium ovale* liver stages in humanized mice. *Nat Commun* 2015, 6:7690.

28. Delves MJ, Ruecker A, Straschil U, Lelievre J, Marques S, Lopez-Barragan MJ, Herreros E, Sinden RE: Male and female *Plasmodium falciparum* mature gametocytes show different responses to antimalarial drugs. *Antimicrob Agents Chemother* 2013, 57:3268-3274.

29. Shiff C: Integrated approach to malaria control. *Clin Microbiol Rev* 2002, 15:278-293.

30. Gendrot M, Fawaz R, Dormoi J, Madamet M, Pradines B: Genetic diversity and deletion of *Plasmodium falciparum* histidine-rich protein 2 and 3: a threat to diagnosis of *P. falciparum* malaria. *Clin Microbiol Infect* 2019, 25:580-585.

31. Pati P, Dhangadamajhi G, Bal M, Ranjit M: High proportions of pfhrp2 gene deletion and performance of HRP2-based rapid diagnostic test in *Plasmodium falciparum* field isolates of Odisha. *Malar J* 2018, 17:394.

32. Hamilton WL, Amato R, van der Pluijm RW, Jacob CG, Quang HH, Thuy-Nhien NT, Hien TT, Hongvanthong B, Chindavongsa K, Mayxay M, et al: Evolution and expansion of multidrug-resistant malaria in southeast Asia: a genomic epidemiology study. *Lancet Infect Dis* 2019, 19:943-951.

33. Implications of Insecticide Resistance C: Implications of insecticide resistance for malaria vector control with long-lasting insecticidal nets: trends in pyrethroid resistance during a WHO-coordinated multi-country prospective study. *Parasit Vectors* 2018, 11:550.

34. Riley EM, Stewart VA: Immune mechanisms in malaria: new insights in vaccine development. *Nat Med* 2013, 19:168-178.

35. Mu J, Awadalla P, Duan J, McGee KM, Keebler J, Seydel K, McVean GA, Su XZ: Genome-wide variation and identification of vaccine targets in the *Plasmodium falciparum* genome. *Nat Genet* 2007, 39:126-130.

36. Beeson JG, Kurtovic L, Dobano C, Opi DH, Chan JA, Feng G, Good MF, Reiling L, Boyle MJ: Challenges and strategies for developing efficacious and long-lasting malaria vaccines. *Sci Transl Med* 2019, 11.

37. Ouattara A, Laurens MB: Vaccines against malaria. *Clin Infect Dis* 2015, 60:930-936.

38. World Health Organization: WHO recommends groundbreaking malaria vaccine for children at risk. Geneva, 2021.

39. Mahmoudi S, Keshavarz H: Efficacy of phase 3 trial of RTS, S/AS01 malaria vaccine: The need for an alternative development plan. *Hum Vaccin Immunother* 2017, 13:2098-2101.

40. Regules JA, Cummings JF, Ockenhouse CF: The RTS,S vaccine candidate for malaria. *Expert Rev Vaccines* 2011, 10:589-599.

41. Collins KA, Snaith R, Cottingham MG, Gilbert SC, Hill AVS: Enhancing protective immunity to malaria with a highly immunogenic virus-like particle vaccine. *Sci Rep* 2017, 7:46621.

42. Conway DJ: Paths to a malaria vaccine illuminated by parasite genomics. *Trends Genet* 2015, 31:97-107.

43. Chen E, Salinas ND, Ntumngia FB, Adams JH, Tolia NH: Structural analysis of the synthetic Duffy Binding Protein (DBP) antigen DEKnull relevant for *Plasmodium vivax* malaria vaccine design. *PLoS Negl Trop Dis* 2015, 9:e0003644.

44. Duffy PE, Patrick Gorres J: Malaria vaccines since 2000: progress, priorities, products. *NPJ Vaccines* 2020, 5:48.

45. Sagara I, Dicko A, Ellis RD, Fay MP, Diawara SI, Assadou MH, Sissoko MS, Kone M, Diallo AI, Saye R, et al: A randomized controlled phase 2 trial of the blood stage AMA1-C1/Alhydrogel malaria vaccine in children in Mali. *Vaccine* 2009, 27:3090-3098.

46. Ouattara A, Takala-Harrison S, Thera MA, Coulibaly D, Niangaly A, Saye R, Tolo Y, Dutta S, Heppner DG, Soisson L, et al: Molecular basis of allele-specific efficacy of a blood-stage malaria vaccine: vaccine development implications. *J Infect Dis* 2013, 207:511-519.

47. Srinivasan P, Ekanem E, Diouf A, Tonkin ML, Miura K, Boulanger MJ, Long CA, Narum DL, Miller LH: Immunization with a functional protein complex required for erythrocyte invasion protects against lethal malaria. *Proc Natl Acad Sci U S A* 2014, 111:10311-10316.

48. Crosnier C, Bustamante LY, Bartholdson SJ, Bei AK, Theron M, Uchikawa M, Mboup S, Ndir O, Kwiatkowski DP, Duraisingh MT, et al: Basigin is a receptor

essential for erythrocyte invasion by *Plasmodium falciparum*. *Nature* 2011, 480:534-537.

49. Minassian AM, Silk SE, Barrett JR, Nielsen CM, Miura K, Diouf A, Loos C, Fallon JK, Michell AR, White MT, et al: Reduced blood-stage malaria growth and immune correlates in humans following RH5 vaccination. *Med (N Y)* 2021, 2:701-719 e719.

50. Raj DK, Nixon CP, Nixon CE, Dvorin JD, DiPetrillo CG, Pond-Tor S, Wu HW, Jolly G, Pischel L, Lu A, et al: Antibodies to PfSEA-1 block parasite egress from RBCs and protect against malaria infection. *Science* 2014, 344:871-877.

51. Nixon CE, Park S, Pond-Tor S, Raj D, Lambert LE, Orr-Gonzalez S, Barnafo EK, Rausch KM, Friedman JF, Fried M, et al: Identification of Protective B-Cell Epitopes within the Novel Malaria Vaccine Candidate *Plasmodium falciparum* Schizont Egress Antigen 1. *Clin Vaccine Immunol* 2017, 24.

52. Mordmuller B, Sulyok M, Egger-Adam D, Resende M, de Jongh WA, Jensen MH, Smedegaard HH, Ditlev SB, Soegaard M, Poulsen L, et al: First-in-human, Randomized, Double-blind Clinical Trial of Differentially Adjuvanted PAMVAC, A Vaccine Candidate to Prevent Pregnancy-associated Malaria. *Clin Infect Dis* 2019, 69:1509-1516.

53. Challenger JD, Olivera Mesa D, Da DF, Yerbanga RS, Lefevre T, Cohuet A, Churcher TS: Predicting the public health impact of a malaria transmission-blocking vaccine. *Nat Commun* 2021, 12:1494.

54. Gwadz RW: Successful immunization against the sexual stages of *Plasmodium gallinaceum*. *Science* 1976, 193:1150-1151.

55. Carter R, Chen DH: Malaria transmission blocked by immunisation with gametes of the malaria parasite. *Nature* 1976, 263:57-60.

56. Kaslow DC, Quakyi IA, Syin C, Raum MG, Keister DB, Coligan JE, McCutchan TF, Miller LH: A vaccine candidate from the sexual stage of human malaria that contains EGF-like domains. *Nature* 1988, 333:74-76.

57. Duffy PE, Kaslow DC: A novel malaria protein, Pfs28, and Pfs25 are genetically linked and synergistic as *falciparum* malaria transmission-blocking vaccines. *Infect Immun* 1997, 65:1109-1113.

58. Williamson KC, Criscio MD, Kaslow DC: Cloning and expression of the gene for *Plasmodium falciparum* transmission-blocking target antigen, Pfs230. *Mol Biochem Parasitol* 1993, 58:355-358.

59. Kocken CH, Jansen J, Kaan AM, Beckers PJ, Ponnudurai T, Kaslow DC, Konings RN, Schoenmakers JG: Cloning and expression of the gene coding for the transmission blocking target antigen Pfs48/45 of *Plasmodium falciparum*. *Mol Biochem Parasitol* 1993, 61:59-68.

60. Armistead JS, Morlais I, Mathias DK, Jardim JG, Joy J, Fridman A, Finnefrock AC, Bagchi A, Plebanski M, Scorpio DG, et al: Antibodies to a single, conserved epitope in *Anopheles* APN1 inhibit universal transmission of *Plasmodium falciparum* and *Plasmodium vivax* malaria. *Infect Immun* 2014, 82:818-829.

61. Takashima E, Tachibana M, Morita M, Nagaoka H, Kanoi BN, Tsuboi T: Identification of Novel Malaria Transmission-Blocking Vaccine Candidates. *Front Cell Infect Microbiol* 2021, 11:805482.

62. Wu Y, Ellis RD, Shaffer D, Fontes E, Malkin EM, Mahanty S, Fay MP, Narum D, Rausch K, Miles AP, et al: Phase 1 trial of malaria transmission blocking vaccine candidates Pfs25 and Pvs25 formulated with montanide ISA 51. *PLoS One* 2008, 3:e2636.

63. Richie TL, Billingsley PF, Sim BK, James ER, Chakravarty S, Epstein JE, Lyke KE, Mordmuller B, Alonso P, Duffy PE, et al: Progress with *Plasmodium falciparum* sporozoite (PfSPZ)-based malaria vaccines. *Vaccine* 2015, 33:7452-7461.

64. Nussenzweig RS, Vanderberg J, Most H, Orton C: Protective immunity produced by the injection of x-irradiated sporozoites of *Plasmodium berghei*. *Nature* 1967, 216:160-162.

65. Clyde DF, McCarthy VC, Miller RM, Hornick RB: Specificity of protection of man immunized against sporozoite-induced *falciparum* malaria. *Am J Med Sci* 1973, 266:398-403.

66. Clyde DF: Immunization of man against *falciparum* and *vivax* malaria by use of attenuated sporozoites. *Am J Trop Med Hyg* 1975, 24:397-401.

67. Clyde DF: Immunity to *falciparum* and *vivax* malaria induced by irradiated sporozoites: a review of the University of Maryland studies, 1971-75. *Bull World Health Organ* 1990, 68 Suppl:9-12.

68. McCarthy VC, Clyde DF: *Plasmodium vivax*: correlation of circumsporozoite precipitation (CSP) reaction with sporozoite-induced protective immunity in man. *Exp Parasitol* 1977, 41:167-171.

69. Rieckmann KH, Carson PE, Beaudoin RL, Cassells JS, Sell KW: Letter: Sporozoite induced immunity in man against an Ethiopian strain of *Plasmodium falciparum*. *Trans R Soc Trop Med Hyg* 1974, 68:258-259.

70. Rieckmann KH, Beaudoin RL, Cassells JS, Sell KW: Use of attenuated sporozoites in the immunization of human volunteers against *falciparum* malaria. *Bull World Health Organ* 1979, 57 Suppl 1:261-265.

71. Hoffman SL, Billingsley PF, James E, Richman A, Loyevsky M, Li T, Chakravarty S, Gunasekera A, Chattopadhyay R, Li M, et al: Development of a metabolically

active, non-replicating sporozoite vaccine to prevent *Plasmodium falciparum* malaria. *Hum Vaccin* 2010, 6:97-106.

72. Seder RA, Chang LJ, Enama ME, Zephir KL, Sarwar UN, Gordon IJ, Holman LA, James ER, Billingsley PF, Gunasekera A, et al: Protection against malaria by intravenous immunization with a nonreplicating sporozoite vaccine. *Science* 2013, 341:1359-1365.

73. Ishizuka AS, Lyke KE, DeZure A, Berry AA, Richie TL, Mendoza FH, Enama ME, Gordon IJ, Chang LJ, Sarwar UN, et al: Protection against malaria at 1 year and immune correlates following PfSPZ vaccination. *Nat Med* 2016, 22:614-623.

74. Epstein JE, Paolino KM, Richie TL, Sedegah M, Singer A, Ruben AJ, Chakravarty S, Stafford A, Ruck RC, Eappen AG, et al: Protection against *Plasmodium falciparum* malaria by PfSPZ Vaccine. *JCI Insight* 2017, 2:e89154.

75. Sissoko MS, Healy SA, Katile A, Omaswa F, Zaidi I, Gabriel EE, Kamate B, Samake Y, Guindo MA, Dolo A, et al: Safety and efficacy of PfSPZ Vaccine against *Plasmodium falciparum* via direct venous inoculation in healthy malaria-exposed adults in Mali: a randomised, double-blind phase 1 trial. *Lancet Infect Dis* 2017, 17:498-509.

76. Jongo SA, Shekalaghe SA, Church LWP, Ruben AJ, Schindler T, Zenklusen I, Rutishauser T, Rothen J, Tumbo A, Mkindi C, et al: Safety, Immunogenicity, and Protective Efficacy against Controlled Human Malaria Infection of *Plasmodium falciparum* Sporozoite Vaccine in Tanzanian Adults. *Am J Trop Med Hyg* 2018, 99:338-349.

77. Bastiaens GJH, van Meer MPA, Scholzen A, Obiero JM, Vatanshenassan M, van Grinsven T, Sim BKL, Billingsley PF, James ER, Gunasekera A, et al: Safety, Immunogenicity, and Protective Efficacy of Intradermal Immunization with Aseptic, Purified, Cryopreserved *Plasmodium falciparum* Sporozoites in Volunteers Under Chloroquine Prophylaxis: A Randomized Controlled Trial. *Am J Trop Med Hyg* 2016, 94:663-673.

78. Roestenberg M, Walk J, van der Boor SC, Langenberg MCC, Hoogerwerf MA, Janse JJ, Manurung M, Yap XZ, Garcia AF, Koopman JPR, et al: A double-blind, placebo-controlled phase 1/2a trial of the genetically attenuated malaria vaccine PfSPZ-GA1. *Sci Transl Med* 2020, 12.

79. VanBuskirk KM, O'Neill MT, De La Vega P, Maier AG, Krzych U, Williams J, Dowler MG, Sacci JB, Jr., Kangwanrangsan N, Tsuboi T, et al: Preerythrocytic, live-attenuated *Plasmodium falciparum* vaccine candidates by design. *Proc Natl Acad Sci U S A* 2009, 106:13004-13009.

80. Goswami D, Betz W, Locham NK, Parthiban C, Brager C, Schafer C, Camargo N, Nguyen T, Kennedy SY, Murphy SC, et al: A replication-competent late liver stage-attenuated human malaria parasite. *JCI Insight* 2020, 5.

81. Butler NS, Schmidt NW, Vaughan AM, Aly AS, Kappe SH, Harty JT: Superior antimalarial immunity after vaccination with late liver stage-arresting genetically attenuated parasites. *Cell Host Microbe* 2011, 9:451-462.

82. Stanisic DI, Fink J, Mayer J, Coghill S, Gore L, Liu XQ, El-Deeb I, Rodriguez IB, Powell J, Willemsen NM, et al: Vaccination with chemically attenuated *Plasmodium falciparum* asexual blood-stage parasites induces parasite-specific cellular immune responses in malaria-naive volunteers: a pilot study. *BMC Med* 2018, 16:184.

83. Ssemaganda A, Giddam AK, Low LM, Liu XQ, Ho MF, Zaman M, Hussein WM, Skwarczynski M, Toth I, Stanisic DI, Good MF: Mannosylated liposomes formulated with whole parasite *P. falciparum* blood-stage antigens are highly immunogenic in mice. *Vaccine* 2020, 38:1494-1504.

84. Amaratunga D, Cabrera J, Lee YS: Enriched random forests. *Bioinformatics* 2008, 24:2010-2014.

85. Jongo SA, Urbano V, Church LWP, Olotu A, Manock SR, Schindler T, Mtoro A, Kc N, Hamad A, Nyakarungu E, et al: Immunogenicity and Protective Efficacy of Radiation-Attenuated and Chemo-Attenuated PfSPZ Vaccines in Equatoguinean Adults. *Am J Trop Med Hyg* 2021, 104:283-293.

86. Jongo SA, Church LWP, Nchama V, Hamad A, Chuquiyauri R, Kassim KR, Athuman T, Deal A, Natasha KC, Mtoro A, et al: Multi-Dose Priming Regimens of PfSPZ Vaccine: Safety and Efficacy against Controlled Human Malaria Infection in Equatoguinean Adults. *Am J Trop Med Hyg* 2022.

87. Diawara H, Healy SA, Mwakingwe-Omari A, Issiaka D, Diallo A, Traore S, Soumbounou IH, Gaoussou S, Zaidi I, Mahamar A, et al: Safety and efficacy of PfSPZ Vaccine against malaria in healthy adults and women anticipating pregnancy in Mali: two randomised, double-blind, placebo-controlled, phase 1 and 2 trials. *Lancet Infect Dis* 2024, 24:1366-1382.

88. Kublin JG, Mikolajczak SA, Sack BK, Fishbaugher ME, Seilie A, Shelton L, VonGoedert T, Firat M, Magee S, Fritzen E, et al: Complete attenuation of genetically engineered *Plasmodium falciparum* sporozoites in human subjects. *Sci Transl Med* 2017, 9.

89. Roozen GVT, van Schuijlenburg R, Hensen ADO, Koopman JPR, Lamers OAC, Geurten FJA, Sijtsma JC, Baalbergen E, Janse JJ, Chevalley-Maurel S, et al: Single immunization with genetically attenuated PfΔmei2 (GA2) parasites by mosquito bite in controlled human malaria infection: a placebo-controlled randomized trial. *Nat Med* 2025, 31:218-222.

90. Hammershaimb EA, Berry AA: Pre-erythrocytic malaria vaccines: RTS,S, R21, and beyond. *Expert Rev Vaccines* 2024, 23:49-52.

91. Lindner SE, Swearingen KE, Harupa A, Vaughan AM, Sinnis P, Moritz RL, Kappe SH: Total and putative surface proteomics of malaria parasite salivary gland sporozoites. *Mol Cell Proteomics* 2013, 12:1127-1143.

92. Coppi A, Natarajan R, Pradel G, Bennett BL, James ER, Roggero MA, Corradin G, Persson C, Tewari R, Sinnis P: The malaria circumsporozoite protein has two functional domains, each with distinct roles as sporozoites journey from mosquito to mammalian host. *J Exp Med* 2011, 208:341-356.

93. Menard R, Sultan AA, Cortes C, Altszuler R, van Dijk MR, Janse CJ, Waters AP, Nussenzweig RS, Nussenzweig V: Circumsporozoite protein is required for development of malaria sporozoites in mosquitoes. *Nature* 1997, 385:336-340.

94. Zhao J, Bhanot P, Hu J, Wang Q: A Comprehensive Analysis of *Plasmodium* Circumsporozoite Protein Binding to Hepatocytes. *PLoS One* 2016, 11:e0161607.

95. Plassmeyer ML, Reiter K, Shimp RL, Jr., Kotova S, Smith PD, Hurt DE, House B, Zou X, Zhang Y, Hickman M, et al: Structure of the *Plasmodium falciparum* circumsporozoite protein, a leading malaria vaccine candidate. *J Biol Chem* 2009, 284:26951-26963.

96. Aly AS, Vaughan AM, Kappe SH: Malaria parasite development in the mosquito and infection of the mammalian host. *Annu Rev Microbiol* 2009, 63:195-221.

97. Loubens M, Vincensini L, Fernandes P, Briquet S, Marinach C, Silvie O: *Plasmodium* sporozoites on the move: Switching from cell traversal to productive invasion of hepatocytes. *Mol Microbiol* 2021, 115:870-881.

98. Coppi A, Pinzon-Ortiz C, Hutter C, Sinnis P: The *Plasmodium* circumsporozoite protein is proteolytically processed during cell invasion. *J Exp Med* 2005, 201:27-33.

99. Le TP, Church LW, Corradin G, Hunter RL, Charoenvit Y, Wang R, de la Vega P, Sacci J, Ballou WR, Kolodny N, et al: Immunogenicity of *Plasmodium falciparum* circumsporozoite protein multiple antigen peptide vaccine formulated with different adjuvants. *Vaccine* 1998, 16:305-312.

100. Kisalu NK, Idris AH, Weidle C, Flores-Garcia Y, Flynn BJ, Sack BK, Murphy S, Schön A, Freire E, Francica JR, et al: A human monoclonal antibody prevents malaria infection by targeting a new site of vulnerability on the parasite. *Nature Medicine* 2018, 24:408-416.

101. Kayentao K, Ongoiba A, Preston AC, Healy SA, Doumbo S, Doumtabe D, Traore A, Traore H, Djiguiba A, Li S, et al: Safety and Efficacy of a Monoclonal Antibody against Malaria in Mali. *N Engl J Med* 2022, 387:1833-1842.

102. Gaudinski MR, Berkowitz NM, Idris AH, Coates EE, Holman LA, Mendoza F, Gordon IJ, Plummer SH, Trofymenko O, Hu Z, et al: A Monoclonal Antibody for Malaria Prevention. *N Engl J Med* 2021, 385:803-814.

103. Lyke KE, Berry AA, Mason K, Idris AH, O'Callahan M, Happe M, Strom L, Berkowitz NM, Guech M, Hu Z, et al: Low-dose intravenous and subcutaneous CIS43LS monoclonal antibody for protection against malaria (VRC 612 Part C): a phase 1, adaptive trial. *Lancet Infect Dis* 2023, 23:578-588.

104. Aleshnick M, Florez-Cuadros M, Martinson T, Wilder BK: Monoclonal antibodies for malaria prevention. *Mol Ther* 2022, 30:1810-1821.

105. Enea V, Ellis J, Zavala F, Arnot DE, Asavanich A, Masuda A, Quakyi I, Nussenzweig RS: DNA cloning of *Plasmodium falciparum* circumsporozoite gene: amino acid sequence of repetitive epitope. *Science* 1984, 225:628-630.

106. Dame JB, Williams JL, McCutchan TF, Weber JL, Wirtz RA, Hockmeyer WT, Maloy WL, Haynes JD, Schneider I, Roberts D, et al.: Structure of the gene encoding the immunodominant surface antigen on the sporozoite of the human malaria parasite *Plasmodium falciparum*. *Science* 1984, 225:593-599.

107. Wang LT, Pereira LS, Flores-Garcia Y, O'Connor J, Flynn BJ, Schon A, Hurlburt NK, Dillon M, Yang ASP, Fabra-Garcia A, et al: A Potent Anti-Malarial Human Monoclonal Antibody Targets Circumsporozoite Protein Minor Repeats and Neutralizes Sporozoites in the Liver. *Immunity* 2020, 53:733-744 e738.

108. Huang HY, Liang XY, Lin LY, Chen JT, Ehapo CS, Eyi UM, Li J, Jiang TT, Zheng YZ, Zha GC, et al: Genetic polymorphism of *Plasmodium falciparum* circumsporozoite protein on Bioko Island, Equatorial Guinea and global comparative analysis. *Malar J* 2020, 19:245.

109. Pringle JC, Carpi G, Almagro-Garcia J, Zhu SJ, Kobayashi T, Mulenga M, Bobanga T, Chaponda M, Moss WJ, Norris DE: RTS,S/AS01 malaria vaccine mismatch observed among *Plasmodium falciparum* isolates from southern and central Africa and globally. *Sci Rep* 2018, 8:6622.

110. Lyke KE, Laurens MB, Strauss K, Adams M, Billingsley PF, James E, Manoj A, Chakravarty S, Plowe CV, Li ML, et al: Optimizing Intradermal Administration of Cryopreserved *Plasmodium falciparum* Sporozoites in Controlled Human Malaria Infection. *Am J Trop Med Hyg* 2015, 93:1274-1284.

111. Stanisic DI, McCarthy JS, Good MF: Controlled Human Malaria Infection: Applications, Advances, and Challenges. *Infect Immun* 2018, 86.

112. Laurens MB, Billingsley P, Richman A, Eappen AG, Adams M, Li T, Chakravarty S, Gunasekera A, Jacob CG, Sim BK, et al: Successful human infection with *P. falciparum* using three aseptic *Anopheles stephensi* mosquitoes: a new model for controlled human malaria infection. *PLoS One* 2013, 8:e68969.

113. Njue M, Njuguna P, Kapulu MC, Sanga G, Bejon P, Marsh V, Molyneux S, Kamuya D: Ethical considerations in Controlled Human Malaria Infection studies in low resource settings: Experiences and perceptions of study participants in a malaria Challenge study in Kenya. *Wellcome Open Res* 2018, 3:39.

114. Friedman-Klabanoff DJ, Laurens MB, Berry AA, Travassos MA, Adams M, Strauss KA, Shrestha B, Levine MM, Edelman R, Lyke KE: The Controlled Human Malaria Infection Experience at the University of Maryland. *Am J Trop Med Hyg* 2019, 100:556-565.

115. Pohl K, Cockburn IA: Innate immunity to malaria: The good, the bad and the unknown. *Front Immunol* 2022, 13:914598.

116. Clark R, Kupper T: Old meets new: the interaction between innate and adaptive immunity. *J Invest Dermatol* 2005, 125:629-637.

117. Liehl P, Zuzarte-Luis V, Chan J, Zillinger T, Baptista F, Carapau D, Konert M, Hanson KK, Carret C, Lassnig C, et al: Host-cell sensors for *Plasmodium* activate innate immunity against liver-stage infection. *Nat Med* 2014, 20:47-53.

118. Gowda DC, Wu X: Parasite Recognition and Signaling Mechanisms in Innate Immune Responses to Malaria. *Front Immunol* 2018, 9:3006.

119. Fernandez-Ruiz D, Ng WY, Holz LE, Ma JZ, Zaid A, Wong YC, Lau LS, Mollard V, Cozijnsen A, Collins N, et al: Liver-Resident Memory CD8(+) T Cells Form a Front-Line Defense against Malaria Liver-Stage Infection. *Immunity* 2016, 45:889-902.

120. Zhu C, Jiao S, Xu W: CD8(+) Trms against malaria liver-stage: prospects and challenges. *Front Immunol* 2024, 15:1344941.

121. Perez-Mazliah D, Langhorne J: CD4 T-cell subsets in malaria: TH1/TH2 revisited. *Front Immunol* 2014, 5:671.

122. Schroeder HW, Jr., Cavacini L: Structure and function of immunoglobulins. *J Allergy Clin Immunol* 2010, 125:S41-52.

123. Wang LT, Cooper AJR, Farrell B, Miura K, Diouf A, Muller-Sienerth N, Crosnier C, Purser L, Kirtley PJ, Maciuszek M, et al: Natural malaria infection elicits rare but potent neutralizing antibodies to the blood-stage antigen RH5. *Cell* 2024, 187:4981-4995 e4914.

124. Thiam LG, McHugh K, Ba A, Li R, Guo Y, Pouye MN, Cisse A, Pipini D, Diallo F, Sene SD, et al: Vaccine-induced human monoclonal antibodies to PfRH5 show broadly neutralizing activity against *P. falciparum* clinical isolates. *NPJ Vaccines* 2024, 9:198.

125. Feng G, Wines BD, Kurtovic L, Chan JA, Boeuf P, Mollard V, Cozijnsen A, Drew DR, Center RJ, Marshall DL, et al: Mechanisms and targets of Fcgamma-receptor mediated immunity to malaria sporozoites. *Nat Commun* 2021, 12:1742.

126. Kyei-Baafour E, Kusi KA, Arthur FKN, Tiendrebeogo RW, Owusu-Yeboa E, Singh SK, Friedrich S, Gerds TA, Dodoo D, Theisen M, Adu B: High opsonic phagocytosis activity and growth inhibition of merozoites are associated with RON4 antibody levels and protect against febrile malaria in Ghanaian children. *Front Immunol* 2023, 14:1161301.

127. Boyle MJ, Reiling L, Feng G, Langer C, Osier FH, Aspeling-Jones H, Cheng YS, Stubbs J, Tetteh KK, Conway DJ, et al: Human antibodies fix complement to inhibit *Plasmodium falciparum* invasion of erythrocytes and are associated with protection against malaria. *Immunity* 2015, 42:580-590.

128. Clements T, Rice TF, Vamvakas G, Barnett S, Barnes M, Donaldson B, Jones CE, Kampmann B, Holder B: Update on Transplacental Transfer of IgG Subclasses: Impact of Maternal and Fetal Factors. *Front Immunol* 2020, 11:1920.

129. Dobano C, Santano R, Vidal M, Jimenez A, Jairoce C, Ubillos I, Dosoo D, Aguilar R, Williams NA, Diez-Padrisa N, et al: Differential Patterns of IgG Subclass Responses to *Plasmodium falciparum* Antigens in Relation to Malaria Protection and RTS,S Vaccination. *Front Immunol* 2019, 10:439.

130. Vidarsson G, Dekkers G, Rispens T: IgG subclasses and allotypes: from structure to effector functions. *Front Immunol* 2014, 5:520.

131. Damelang T, de Taeye SW, Rentenaar R, Roya-Kouchaki K, de Boer E, Derksen NIL, van Kessel K, Lissenberg-Thunnissen S, Rooijakkers SHM, Jongerius I, et al: The Influence of Human IgG Subclass and Allotype on Complement Activation. *J Immunol* 2023, 211:1725-1735.

132. Stanisic DI, Richards JS, McCallum FJ, Michon P, King CL, Schoepflin S, Gilson PR, Murphy VJ, Anders RF, Mueller I, Beeson JG: Immunoglobulin G subclass-specific responses against *Plasmodium falciparum* merozoite antigens are associated with control of parasitemia and protection from symptomatic illness. *Infect Immun* 2009, 77:1165-1174.

133. Bredius RG, Fijen CA, De Haas M, Kuijper EJ, Weening RS, Van de Winkel JG, Out TA: Role of neutrophil Fc gamma RIIa (CD32) and Fc gamma RIIIb (CD16) polymorphic forms in phagocytosis of human IgG1- and IgG3-opsonized bacteria and erythrocytes. *Immunology* 1994, 83:624-630.

134. Chumpitazi BF, Lepers JP, Simon J, Deloron P: IgG1 and IgG2 antibody responses to *Plasmodium falciparum* exoantigens correlate inversely and positively, respectively, to the number of malaria attacks. *FEMS Immunol Med Microbiol* 1996, 14:151-158.

135. Aucan C, Traore Y, Tall F, Nacro B, Traore-Leroux T, Fumoux F, Rihet P: High immunoglobulin G2 (IgG2) and low IgG4 levels are associated with human resistance to *Plasmodium falciparum* malaria. *Infect Immun* 2000, 68:1252-1258.

136. Schreiber N, Brattig N, Evans J, Tsiri A, Horstmann RD, May J, Klinkert MQ: Cerebral malaria is associated with IgG2 and IgG4 antibody responses to recombinant *Plasmodium falciparum* RIFIN antigen. *Microbes Infect* 2006, 8:1269-1276.

137. Aitken EH, Mahanty S, Rogerson SJ: Antibody effector functions in malaria and other parasitic diseases: a few needles and many haystacks. *Immunol Cell Biol* 2020, 98:264-275.

138. Julien JP, Wardemann H: Antibodies against *Plasmodium falciparum* malaria at the molecular level. *Nat Rev Immunol* 2019, 19:761-775.

139. Plotkin SA: Correlates of protection induced by vaccination. *Clin Vaccine Immunol* 2010, 17:1055-1065.

140. Crompton PD, Pierce SK, Miller LH: Advances and challenges in malaria vaccine development. *J Clin Invest* 2010, 120:4168-4178.

141. Stanisic DI, McCall MBB: Correlates of malaria vaccine efficacy. *Expert Rev Vaccines* 2021, 20:143-161.

142. World Health O: WHO Technical Consultation on Immune Correlates of Protection for the Evaluation of Malaria Vaccines: Meeting Report. In *Meeting Report, 24–25 June 2024*. Geneva, Switzerland: World Health Organization; 2024.

143. Kester KE, Cummings JF, Ofori-Anyinam O, Ockenhouse CF, Krzych U, Moris P, Schwenk R, Nielsen RA, Debebe Z, Pinelis E, et al: Randomized, double-blind, phase 2a trial of *falciparum* malaria vaccines RTS,S/AS01B and RTS,S/AS02A in malaria-naive adults: safety, efficacy, and immunologic associates of protection. *J Infect Dis* 2009, 200:337-346.

144. Williams TN, Mwangi TW, Roberts DJ, Alexander ND, Weatherall DJ, Wambua S, Kortok M, Snow RW, Marsh K: An immune basis for malaria protection by the sickle cell trait. *PLoS Med* 2005, 2:e128.

145. Gong L, Maiteki-Sebuguzi C, Rosenthal PJ, Hubbard AE, Drakeley CJ, Dorsey G, Greenhouse B: Evidence for both innate and acquired mechanisms of protection from *Plasmodium falciparum* in children with sickle cell trait. *Blood* 2012, 119:3808-3814.

146. Kurup SP, Butler NS, Harty JT: T cell-mediated immunity to malaria. *Nat Rev Immunol* 2019, 19:457-471.

147. Lefebvre MN, Harty JT: You Shall Not Pass: Memory CD8 T Cells in Liver-Stage Malaria. *Trends Parasitol* 2020, 36:147-157.

148. Weiss WR, Jiang CG: Protective CD8+ T lymphocytes in primates immunized with malaria sporozoites. *PLoS One* 2012, 7:e31247.

149. Kisalu NK, Idris AH, Weidle C, Flores-Garcia Y, Flynn BJ, Sack BK, Murphy S, Schon A, Freire E, Francica JR, et al: Author Correction: A human monoclonal antibody prevents malaria infection by targeting a new site of vulnerability on the parasite. *Nat Med* 2019, 25:188-189.

150. Zalevsky J, Chamberlain AK, Horton HM, Karki S, Leung IW, Sproule TJ, Lazar GA, Roopenian DC, Desjarlais JR: Enhanced antibody half-life improves in vivo activity. *Nat Biotechnol* 2010, 28:157-159.

151. Barry AE, Schultz L, Senn N, Nale J, Kiniboro B, Siba PM, Mueller I, Reeder JC: High levels of genetic diversity of *Plasmodium falciparum* populations in Papua New Guinea despite variable infection prevalence. *Am J Trop Med Hyg* 2013, 88:718-725.

152. Robert F, Ntoumi F, Angel G, Candito D, Rogier C, Fandeur T, Sarthou JL, Mercereau-Puijalon O: Extensive genetic diversity of *Plasmodium falciparum* isolates collected from patients with severe malaria in Dakar, Senegal. *Trans R Soc Trop Med Hyg* 1996, 90:704-711.

153. Belachew EB: Immune Response and Evasion Mechanisms of *Plasmodium falciparum* Parasites. *J Immunol Res* 2018, 2018:6529681.

154. Takala-Harrison S, Laufer MK: Antimalarial drug resistance in Africa: key lessons for the future. *Ann N Y Acad Sci* 2015, 1342:62-67.

155. Long CA, Zavala F: Malaria vaccines and human immune responses. *Curr Opin Microbiol* 2016, 32:96-102.

156. Koita OA, Doumbo OK, Ouattara A, Tall LK, Konare A, Diakite M, Diallo M, Sagara I, Masinde GL, Doumbo SN, et al: False-negative rapid diagnostic tests for malaria and deletion of the histidine-rich repeat region of the hrp2 gene. *Am J Trop Med Hyg* 2012, 86:194-198.

157. Volkman SK, Barry AE, Lyons EJ, Nielsen KM, Thomas SM, Choi M, Thakore SS, Day KP, Wirth DF, Hartl DL: Recent origin of *Plasmodium falciparum* from a single progenitor. *Science* 2001, 293:482-484.

158. Manske M, Miotto O, Campino SG: Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. *Nature*.

159. Wright S: Genetical structure of populations. *Nature* 1950, 166:247-249.

160. Hartl DL, Clark AG: *Principles of population genetics.* 3rd edn. Sunderland, MA: Sinauer Associates; 1997.

161. Song S, Dey DK, Holsinger KE: Differentiation among populations with migration, mutation, and drift: implications for genetic inference. *Evolution* 2006, 60:1-12.

162. Star B, Spencer HG: Effects of genetic drift and gene flow on the selective maintenance of genetic variation. *Genetics* 2013, 194:235-244.

163. Neafsey DE, Taylor AR, MacInnis BL: Advances and opportunities in malaria population genomics. *Nat Rev Genet* 2021, 22:502-517.

164. Pumpaibool T, Arnathau C, Durand P, Kanchanakhan N, Siripoon N, Suegorn A, Sitthi-Amorn C, Renaud F, Harnyuttanakorn P: Genetic diversity and population structure of *Plasmodium falciparum* in Thailand, a low transmission country. *Malar J* 2009, 8:155.

165. Anderson TJ, Haubold B, Williams JT, Estrada-Franco JG, Richardson L, Mollinedo R, Bockarie M, Mokili J, Mharakurwa S, French N, et al: Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. *Mol Biol Evol* 2000, 17:1467-1482.

166. Sakihama N, Kaneko A, Hattori T, Tanabe K: Limited recombination events in merozoite surface protein-1 alleles of *Plasmodium falciparum* on islands. *Gene* 2001, 279:41-48.

167. Mulenge FM, Hunja CW, Magiri E, Culleton R, Kaneko A, Aman RA: Genetic Diversity and Population Structure of *Plasmodium falciparum* in Lake Victoria Islands, A Region of Intense Transmission. *Am J Trop Med Hyg* 2016, 95:1077-1085.

168. Roh ME, Tessema SK, Murphy M, Nhlabathi N, Mkhonta N, Vilakati S, Ntshalintshali N, Saini M, Maphalala G, Chen A, et al: High Genetic Diversity of *Plasmodium falciparum* in the Low-Transmission Setting of the Kingdom of Eswatini. *J Infect Dis* 2019, 220:1346-1354.

169. Hendry JA, Kwiatkowski D, McVean G: Elucidating relationships between *P. falciparum* prevalence and measures of genetic diversity with a combined genetic-epidemiological model of malaria. *PLoS Comput Biol* 2021, 17:e1009287.

170. Das S, Muleba M, Stevenson JC, Pringle JC, Norris DE: Beyond the entomological inoculation rate: characterizing multiple blood feeding behavior and *Plasmodium falciparum* multiplicity of infection in *Anopheles* mosquitoes in northern Zambia. *Parasit Vectors* 2017, 10:45.

171. Abukari Z, Okonu R, Nyarko SB, Lo AC, Dieng CC, Salifu SP, Gyan BA, Lo E, Amoah LE: The Diversity, Multiplicity of Infection and Population Structure of *P.*

falciparum Parasites Circulating in Asymptomatic Carriers Living in High and Low Malaria Transmission Settings of Ghana. *Genes (Basel)* 2019, 10.

172. Amambua-Ngwa A, Amenga-Etego L, Kamau E, Amato R, Ghansah A, Golassa L, Randrianarivelojosia M, Ishengoma D, Apinjoh T, Maiga-Ascofare O, et al: Major subpopulations of *Plasmodium falciparum* in sub-Saharan Africa. *Science* 2019, 365:813-816.

173. Ouattara A, Barry AE, Dutta S, Remarque EJ, Beeson JG, Plowe CV: Designing malaria vaccines to circumvent antigen variability. *Vaccine* 2015, 33:7506-7512.

174. Takala SL, Plowe CV: Genetic diversity and malaria vaccine design, testing and efficacy: preventing and overcoming 'vaccine resistant malaria'. *Parasite Immunol* 2009, 31:560-573.

175. Dutta S, Lee SY, Batchelor AH, Lanar DE: Structural basis of antigenic escape of a malaria vaccine candidate. *Proc Natl Acad Sci U S A* 2007, 104:12488-12493.

176. Ernst JD: Antigenic Variation and Immune Escape in the MTBC. *Adv Exp Med Biol* 2017, 1019:171-190.

177. Shah Z, Naung MT, Moser KA, Adams M, Buchwald AG, Dwivedi A, Ouattara A, Seydel KB, Mathanga DP, Barry AE, et al: Whole-genome analysis of Malawian *Plasmodium falciparum* isolates identifies possible targets of allele-specific immunity to clinical malaria. *PLoS Genet* 2021, 17:e1009576.

178. Ouattara A, Mu J, Takala-Harrison S, Saye R, Sagara I, Dicko A, Niangaly A, Duan J, Ellis RD, Miller LH, et al: Lack of allele-specific efficacy of a bivalent AMA1 malaria vaccine. *Malar J* 2010, 9:175.

179. Gandhi K, Thera MA, Coulibaly D, Traore K, Guindo AB, Ouattara A, Takala-Harrison S, Berry AA, Doumbo OK, Plowe CV: Variation in the circumsporozoite protein of *Plasmodium falciparum*: vaccine development implications. *PLoS One* 2014, 9:e101783.

180. Naung MT, Martin E, Munro J, Mehra S, Guy AJ, Laman M, Harrison GLA, Tavul L, Hetzel M, Kwiatkowski D, et al: Global diversity and balancing selection of 23 leading *Plasmodium falciparum* candidate vaccine antigens. *PLoS Comput Biol* 2022, 18:e1009801.

181. Escalante AA, Grebert HM, Chaiyaroj SC, Magris M, Biswas S, Nahlen BL, Lal AA: Polymorphism in the gene encoding the apical membrane antigen-1 (AMA-1) of *Plasmodium falciparum*. X. Asembo Bay Cohort Project. *Mol Biochem Parasitol* 2001, 113:279-287.

182. Escalante AA, Grebert HM, Isea R, Goldman IF, Basco L, Magris M, Biswas S, Kariuki S, Lal AA: A study of genetic diversity in the gene encoding the circumsporozoite protein (CSP) of *Plasmodium falciparum* from different

transmission areas--XVI. Asembo Bay Cohort Project. *Mol Biochem Parasitol* 2002, 125:83-90.

183. Conway DJ, Roper C, Oduola AM, Arnot DE, Kremsner PG, Grobusch MP, Curtis CF, Greenwood BM: High recombination rate in natural populations of *Plasmodium falciparum*. *Proc Natl Acad Sci U S A* 1999, 96:4506-4511.

184. Su X, Ferdig MT, Huang Y, Huynh CQ, Liu A, You J, Wootton JC, Wellems TE: A genetic map and recombination parameters of the human malaria parasite *Plasmodium falciparum*. *Science* 1999, 286:1351-1353.

185. Talisuna AO, Okello PE, Erhart A, Coosemans M, D'Alessandro U: Intensity of malaria transmission and the spread of *Plasmodium falciparum* resistant malaria: a review of epidemiologic field evidence. *Am J Trop Med Hyg* 2007, 77:170-180.

186. Jiang H, Li N, Gopalan V, Zilversmit MM, Varma S, Nagarajan V, Li J, Mu J, Hayton K, Henschen B, et al: High recombination rates and hotspots in a *Plasmodium falciparum* genetic cross. *Genome Biol* 2011, 12:R33.

187. Morlais I, Nsango SE, Toussile W, Abate L, Annan Z, Tchioffo MT, Cohuet A, Awono-Ambene PH, Fontenille D, Rousset F, Berry A: *Plasmodium falciparum* mating patterns and mosquito infectivity of natural isolates of gametocytes. *PLoS One* 2015, 10:e0123777.

188. O'Brien JD, Amenga-Etego L, Li R: Approaches to estimating inbreeding coefficients in clinical isolates of *Plasmodium falciparum* from genomic sequence data. *Malar J* 2016, 15:473.

189. Sa JM, Kaslow SR, Krause MA, Melendez-Muniz VA, Salzman RE, Kite WA, Zhang M, Moraes Barros RR, Mu J, Han PK, et al: Artemisinin resistance phenotypes and K13 inheritance in a *Plasmodium falciparum* cross and Aotus model. *Proc Natl Acad Sci U S A* 2018, 115:12513-12518.

190. Camponovo F, Buckee CO, Taylor AR: Measurably recombining malaria parasites. *Trends Parasitol* 2023, 39:17-25.

191. Duffy MF, Byrne TJ, Carret C, Ivens A, Brown GV: Ectopic recombination of a malaria var gene during mitosis associated with an altered var switch rate. *J Mol Biol* 2009, 389:453-469.

192. Bopp SE, Manary MJ, Bright AT, Johnston GL, Dharia NV, Luna FL, McCormack S, Plouffe D, McNamara CW, Walker JR, et al: Mitotic evolution of *Plasmodium falciparum* shows a stable core genome but recombination in antigen families. *PLoS Genet* 2013, 9:e1003293.

193. Anthony TG, Conway DJ, Cox-Singh J, Matusop A, Ratnam S, Shamsul S, Singh B: Fragmented population structure of *Plasmodium falciparum* in a region of declining endemicity. *J Infect Dis* 2005, 191:1558-1564.

194. Shetty AC, Jacob CG, Huang F, Li Y, Agrawal S, Saunders DL, Lon C, Fukuda MM, Ringwald P, Ashley EA, et al: Genomic structure and diversity of *Plasmodium falciparum* in Southeast Asia reveal recent parasite migration patterns. *Nat Commun* 2019, 10:2665.

195. Bogreau H, Renaud F, Bouchiba H, Durand P, Assi SB, Henry MC, Garnotel E, Pradines B, Fusai T, Wade B, et al: Genetic diversity and structure of African *Plasmodium falciparum* populations in urban and rural areas. *Am J Trop Med Hyg* 2006, 74:953-959.

196. Villena FE, Sanchez JF, Nolasco O, Braga G, Ricopa L, Barazorda K, Salas CJ, Lucas C, Lizewski SE, Joya CA, et al: Drug resistance and population structure of *Plasmodium falciparum* and *Plasmodium vivax* in the Peruvian Amazon. *Sci Rep* 2022, 12:16474.

197. Tiedje KE, Zhan Q, Ruybal-Pesantez S, Tonkin-Hill G, He Q, Tan MH, Argyropoulos DC, Deed SL, Ghansah A, Bangre O, et al: Measuring changes in *Plasmodium falciparum* census population size in response to sequential malaria control interventions. *medRxiv* 2024.

198. Brokhattingen N, Matambisso G, da Silva C, Neubauer Vickers E, Pujol A, Mbeve H, Cistero P, Maculuve S, Cuna B, Melembe C, et al: Genomic malaria surveillance of antenatal care users detects reduced transmission following elimination interventions in Mozambique. *Nat Commun* 2024, 15:2402.

199. Mita T, Jombart T: Patterns and dynamics of genetic diversity in *Plasmodium falciparum*: what past human migrations tell us about malaria. *Parasitol Int* 2015, 64:238-243.

200. Takala-Harrison S, Clark TG, Jacob CG, Cummings MP, Miotto O, Dondorp AM, Fukuda MM, Nosten F, Noedl H, Imwong M, et al: Genetic loci associated with delayed clearance of *Plasmodium falciparum* following artemisinin treatment in Southeast Asia. *Proc Natl Acad Sci U S A* 2013, 110:240-245.

201. Dobzhansky T: A review of some fundamental concepts and problems of population genetics. *Cold Spring Harb Symp Quant Biol* 1955, 20:1-15.

202. Charlesworth D: Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genet* 2006, 2:e64.

203. Tetteh KK, Stewart LB, Ochola LI, Amambua-Ngwa A, Thomas AW, Marsh K, Weedall GD, Conway DJ: Prospective identification of malaria parasite genes under balancing selection. *PLoS One* 2009, 4:e5568.

204. Wright GJ, Rayner JC: *Plasmodium falciparum* erythrocyte invasion: combining function with immune evasion. *PLoS Pathog* 2014, 10:e1003943.

205. Persson KE, Fowkes FJ, McCallum FJ, Gicheru N, Reiling L, Richards JS, Wilson DW, Lopaticki S, Cowman AF, Marsh K, Beeson JG: Erythrocyte-binding antigens of *Plasmodium falciparum* are targets of human inhibitory antibodies and function to evade naturally acquired immunity. *J Immunol* 2013, 191:785-794.

206. Tham WH, Healer J, Cowman AF: Erythrocyte and reticulocyte binding-like proteins of *Plasmodium falciparum*. *Trends Parasitol* 2012, 28:23-30.

207. Kingsolver JG, Pfennig DW: Patterns and Power of Phenotypic Selection in Nature. *BioScience* 2007, 57:561-572.

208. Sidhu AB, Verdier-Pinard D, Fidock DA: Chloroquine resistance in *Plasmodium falciparum* malaria parasites conferred by pfcrt mutations. *Science* 2002, 298:210-213.

209. Djimde A, Doumbo OK, Cortese JF, Kayentao K, Doumbo S, Diourte Y, Coulibaly D, Dicko A, Su XZ, Nomura T, et al: A molecular marker for chloroquine-resistant *falciparum* malaria. *N Engl J Med* 2001, 344:257-263.

210. Narh CA, Ghansah A, Duffy MF, Ruybal-Pesantez S, Onwona CO, Oduro AR, Koram KA, Day KP, Tiedje KE: Evolution of Antimalarial Drug Resistance Markers in the Reservoir of *Plasmodium falciparum* Infections in the Upper East Region of Ghana. *J Infect Dis* 2020, 222:1692-1701.

211. Dokomajilar C, Nsobya SL, Greenhouse B, Rosenthal PJ, Dorsey G: Selection of *Plasmodium falciparum* pfmdr1 alleles following therapy with artemether-lumefantrine in an area of Uganda where malaria is highly endemic. *Antimicrob Agents Chemother* 2006, 50:1893-1895.

212. Mu J, Myers RA, Jiang H, Liu S, Ricklefs S, Waisberg M, Chotivanich K, Wilairatana P, Krudsood S, White NJ, et al: *Plasmodium falciparum* genome-wide scans for positive selection, recombination hot spots and resistance to antimalarial drugs. *Nat Genet* 2010, 42:268-271.

213. World Malaria Report 2024. Geneva: World Health Organization; 2024.

214. Doolan DL, Apte SH, Proietti C: Genome-based vaccine design: the promise for malaria and other infectious diseases. *Int J Parasitol* 2014, 44:901-913.

215. Barry AE, Leliwa-Sytek A, Tavul L, Imrie H, Migot-Nabias F, Brown SM, McVean GA, Day KP: Population genomics of the immune evasion (var) genes of *Plasmodium falciparum*. *PLoS Pathog* 2007, 3:e34.

216. Weedall GD, Conway DJ: Detecting signatures of balancing selection to identify targets of anti-parasite immunity. *Trends Parasitol* 2010, 26:363-369.

217. Deroost K, Pham TT, Opdenakker G, Van den Steen PE: The immunological balance between host and parasite in malaria. *FEMS Microbiol Rev* 2016, 40:208-257.

218. Bolstad B: preprocessCore: A collection of pre-processing functions. 1.66.0 edition: GitHub; 2024.

219. Breiman L: Random forests. *Machine learning* 2001, 45:5-32.

220. Hastie T, Tibshirani R, Friedman J: *The elements of statistical learning: data mining, inference, and prediction.* Springer Science & Business Media; 2009.

221. Zou H, Hastie T: Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2005, 67:301-320.

222. Varma S, Simon R: Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics* 2006, 7:91.

223. Wilcoxin F: Probability tables for individual comparisons by ranking methods. *Biometrics* 1947, 3:119-122.

224. Wilcoxon F: Individual comparisons of grouped data by ranking methods. *J Econ Entomol* 1946, 39:269.

225. Brunner E, Munzel U: The Nonparametric Behrens-Fisher Problem: Asymptotic Theory and a Small-Sample Approximation. *Biometrical Journal* 2000, 42:17-25.

226. Benjamini Y, Hochberg Y: Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)* 1995, 57:289-300.

227. Laufer MK, Thesing PC, Dzinjalamala FK, Nyirenda OM, Masonga R, Laurens MB, Stokes-Riner A, Taylor TE, Plowe CV: A longitudinal trial comparing chloroquine as monotherapy or in combination with artesunate, azithromycin or atovaquone-proguanil to treat malaria. *PLoS One* 2012, 7:e42284.

228. Shah Z, Adams M, Moser KA, Shrestha B, Stucke EM, Laufer MK, Serre D, Silva JC, Takala-Harrison S: Optimization of parasite DNA enrichment approaches to generate whole genome sequencing data for *Plasmodium falciparum* from low parasitaemia samples. *Malar J* 2020, 19:135.

229. Buchwald AG, Sixpence A, Chimenya M, Damson M, Sorkin JD, Wilson ML, Seydel K, Hochman S, Mathanga DP, Taylor TE, Laufer MK: Clinical Implications of Asymptomatic *Plasmodium falciparum* Infections in Malawi. *Clin Infect Dis* 2019, 68:106-112.

230. Buchwald AG, Sorkin JD, Sixpence A, Chimenya M, Damson M, Wilson ML, Seydel K, Hochman S, Mathanga D, Taylor TE, Laufer MK: Association Between Age and *Plasmodium falciparum* Infection Dynamics. *Am J Epidemiol* 2019, 188:169-176.

231. Zainabadi K, Adams M, Han ZY, Lwin HW, Han KT, Ouattara A, Thura S, Plowe CV, Nyunt MM: A novel method for extracting nucleic acids from dried blood spots for ultrasensitive detection of low-density *Plasmodium falciparum* and *Plasmodium vivax* infections. *Malar J* 2017, 16:377.

232. Vareta J, Horstman NA, Adams M, Seydel KB, McCann RS, Cohee LM, Laufer MK, Takala-Harrison S: Genotyping *Plasmodium falciparum* gametocytes using amplicon deep sequencing. *Malar J* 2024, 23:96.

233. Sundararaman SA, Plenderleith LJ, Liu W, Loy DE, Learn GH, Li Y, Shaw KS, Ayouba A, Peeters M, Speede S, et al: Genomes of cryptic chimpanzee *Plasmodium* species reveal key evolutionary events leading to human malaria. *Nat Commun* 2016, 7:11078.

234. Moser KA, Drabek EF, Dwivedi A, Stucke EM, Crabtree J, Dara A, Shah Z, Adams M, Li T, Rodrigues PT, et al: Strains used in whole organism *Plasmodium falciparum* vaccine trials differ in genome structure, sequence, and immunogenic potential. *Genome Med* 2020, 12:6.

235. Tajima F: Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 1989, 123:585-595.

236. Wright S: The genetical structure of populations. *Ann Eugen* 1951, 15:323-354.

237. Amambua-Ngwa A, Tetteh KK, Manske M, Gomez-Escobar N, Stewart LB, Deerhake ME, Cheeseman IH, Newbold CI, Holder AA, Knuepfer E, et al: Population genomic scan for candidate signatures of balancing selection to guide antigen characterization in malaria parasites. *PLoS Genet* 2012, 8:e1002992.

238. Hudson RR, Kreitman M, Aguade M: A test of neutral molecular evolution based on nucleotide data. *Genetics* 1987, 116:153-159.

239. Hudson RR, Slatkin M, Maddison WP: Estimation of levels of gene flow from DNA sequence data. *Genetics* 1992, 132:583-589.

240. Holsinger KE, Weir BS: Genetics in geographically structured populations: defining, estimating and interpreting F(ST). *Nat Rev Genet* 2009, 10:639-650.

241. Bhatia G, Patterson N, Sankararaman S, Price AL: Estimating and interpreting F_{ST}: the impact of rare variants. *Genome Res* 2013, 23:1514-1521.

242. World Health Organization: WHO recommends R21/Matrix-M vaccine for malaria prevention in updated advice on immunization. Geneva, 2023.

243. World Health Organization: Full Evidence Report on the RTS,S/AS01 Malaria Vaccine Geneva, 2021.

244. Draper SJ, Sack BK, King CR, Nielsen CM, Rayner JC, Higgins MK, Long CA, Seder RA: Malaria Vaccines: Recent Advances and New Horizons. *Cell Host Microbe* 2018, 24:43-56.

245. Lyke KE, Singer A, Berry AA, Reyes S, Chakravarty S, James ER, Billingsley PF, Gunasekera A, Manoj A, Murshedkar T, et al: Multidose Priming and Delayed Boosting Improve *Plasmodium falciparum* Sporozoite Vaccine Efficacy Against Heterologous *P. falciparum* Controlled Human Malaria Infection. *Clin Infect Dis* 2021, 73:e2424-e2435.

246. Mordmuller B, Sulyok Z, Sulyok M, Molnar Z, Lalremruata A, Calle CL, Bayon PG, Esen M, Gmeiner M, Held J, et al: A PfSPZ vaccine immunization regimen equally protective against homologous and heterologous controlled human malaria infection. *NPJ Vaccines* 2022, 7:100.

247. Legutki JB, Zhao ZG, Greving M, Woodbury N, Johnston SA, Stafford P: Scalable high-density peptide arrays for comprehensive health monitoring. *Nat Commun* 2014, 5:4785.

248. O'Donnell B, Maurer A, Papandreou-Suppappola A, Stafford P: Time-Frequency Analysis of Peptide Microarray Data: Application to Brain Cancer Immunosignatures. *Cancer Inform* 2015, 14:219-233.

249. Stafford P, Johnston SA, Kantarci OH, Zare-Shahabadi A, Warrington A, Rodriguez M: Antibody characterization using immunosignatures. *PLoS One* 2020, 15:e0229080.

250. Francis T: On the Doctrine of Original Antigenic Sin. *Proceedings of the American Philosophical Society* 1960, 104:572-578.

251. Jongo SA, Church LWP, Mtoro AT, Chakravarty S, Ruben AJ, Swanson PA, Kassim KR, Mpina M, Tumbo AM, Milando FA, et al: Safety and Differential Antibody and T-Cell Responses to the *Plasmodium falciparum* Sporozoite Malaria Vaccine, PfSPZ Vaccine, by Age in Tanzanian Adults, Adolescents, Children, and Infants. *Am J Trop Med Hyg* 2019, 100:1433-1444.

252. Richie TL, Church LWP, Murshedkar T, Billingsley PF, James ER, Chen MC, Abebe Y, K CN, Chakravarty S, Dolberg D, et al: Sporozoite immunization: innovative translational science to support the fight against malaria. *Expert Rev Vaccines* 2023, 22:964-1007.

253. Camponovo F, Campo JJ, Le TQ, Oberai A, Hung C, Pablo JV, Teng AA, Liang X, Sim BKL, Jongo S, et al: Proteome-wide analysis of a malaria vaccine study reveals personalized humoral immune profiles in Tanzanian adults. *Elife* 2020, 9.

254. Mordmuller B, Surat G, Lagler H, Chakravarty S, Ishizuka AS, Lalremruata A, Gmeiner M, Campo JJ, Esen M, Ruben AJ, et al: Sterile protection against human malaria by chemoattenuated PfSPZ vaccine. *Nature* 2017, 542:445-449.

255. Bailey JA, Berry AA, Travassos MA, Ouattara A, Boudova S, Dotsey EY, Pike A, Jacob CG, Adams M, Tan JC, et al: Microarray analyses reveal strain-specific antibody responses to *Plasmodium falciparum* apical membrane antigen 1 variants following natural infection and vaccination. *Sci Rep* 2020, 10:3952.

256. Valentini D, Rao M, Rane L, Rahman S, Axelsson-Robertson R, Heuchel R, Lohr M, Hoft D, Brighenti S, Zumla A, Maeurer M: Peptide microarray-based characterization of antibody responses to host proteins after bacille Calmette-Guerin vaccination. *Int J Infect Dis* 2017, 56:140-154.

257. Suvorov A: Simple method for cutoff point identification in descriptive high-throughput biological studies. *BMC Genomics* 2022, 23:204.

258. Cawley GC, Talbot NL: On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research* 2010, 11:2079-2107.

259. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Duchesnay E: Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 2011, 12:2825-2830.

260. Liaw A, Wiener M: Classification and regression by randomForest. *R News* 2002, 2:18-22.

261. Zhang M, Wang C, Oberstaller J, Thomas P, Otto TD, Casandra D, Boyapalle S, Adapa SR, Xu S, Button-Simons K, et al: The apicoplast link to fever-survival and artemisinin-resistance in the malaria parasite. *Nat Commun* 2021, 12:4563.

262. Alexa AaR, J.: topGO: Enrichment Analysis for Gene Ontology. R package version 2.59.0 edition: Bioconductor; 2024.

263. Alexa A, Rahnenfuhrer J, Lengauer T: Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* 2006, 22:1600-1607.

264. Ochola LI, Tetteh KK, Stewart LB, Riitho V, Marsh K, Conway DJ: Allele frequency-based and polymorphism-versus-divergence indices of balancing selection in a new filtered set of polymorphic genes in *Plasmodium falciparum*. *Mol Biol Evol* 2010, 27:2344-2351.

265. Cowman AF, Healer J, Marapana D, Marsh K: Malaria: Biology and Disease. *Cell* 2016, 167:610-624.

266. Recker M, Buckee CO, Serazin A, Kyes S, Pinches R, Christodoulou Z, Springer AL, Gupta S, Newbold CI: Antigenic variation in *Plasmodium falciparum* malaria involves a highly structured switching pattern. *PLoS Pathog* 2011, 7:e1001306.

267. Early AM, Lievens M, MacInnis BL, Ockenhouse CF, Volkman SK, Adjei S, Agbenyega T, Ansong D, Gondi S, Greenwood B, et al: Host-mediated selection impacts the diversity of *Plasmodium falciparum* antigens within infections. *Nat Commun* 2018, 9:1381.

268. He Q, Pilosof S, Tiedje KE, Day KP, Pascual M: Frequency-Dependent Competition Between Strains Imparts Persistence to Perturbations in a Model of *Plasmodium falciparum* Malaria Transmission. *Front Ecol Evol* 2021, 9.

269. Teyssier NB, Chen A, Duarte EM, Sit R, Greenhouse B, Tessema SK: Optimization of whole-genome sequencing of *Plasmodium falciparum* from low-density dried blood spot samples. *Malar J* 2021, 20:116.

270. Guo B, Borda V, Laboulaye R, Spring MD, Wojnarski M, Vesely BA, Silva JC, Waters NC, O'Connor TD, Takala-Harrison S: Strong positive selection biases identity-by-descent-based inferences of recent demography and population structure in *Plasmodium falciparum*. *Nat Commun* 2024, 15:2499.

271. Genome Reference C: GRCh38.p14 Genome Assembly. 2022.

272. Langmead B, Salzberg SL: Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012, 9:357-359.

273. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, et al: From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* 2013, 43:11 10 11-11 10 33.

274. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, et al: A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011, 43:491-498.

275. Quinlan AR, Hall IM: BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010, 26:841-842.

276. Benson G: Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 1999, 27:573-580.

277. Auburn S, Campino S, Miotto O, Djimde AA, Zongo I, Manske M, Maslen G, Mangano V, Alcock D, MacInnis B, et al: Characterization of within-host *Plasmodium falciparum* diversity using next-generation sequence data. *PLoS One* 2012, 7:e32891.

278. Lee. S, Bahlo M: Moimix: an R package for assessing clonality in high-throughput sequencing data. moimix: an R package for assessing clonality in high-throughput sequencing data. 2016.

279. Zhu SJ, Almagro-Garcia J, McVean G: Deconvolution of multiple infections in *Plasmodium falciparum* from high throughput sequencing data. *Bioinformatics* 2018, 34:9-15.

280. Browning BL, Tian X, Zhou Y, Browning SR: Fast two-stage phasing of large-scale sequence data. *Am J Hum Genet* 2021, 108:1880-1890.

281. Browning BL, Zhou Y, Browning SR: A One-Penny Imputed Genome from Next-Generation Reference Panels. *Am J Hum Genet* 2018, 103:338-348.

282. Zhu SJ, Hendry JA, Almagro-Garcia J, Pearson RD, Amato R, Miles A, Weiss DJ, Lucas TC, Nguyen M, Gething PW, et al: The origins and relatedness structure of mixed infections vary with local prevalence of *P. falciparum* malaria. *Elife* 2019, 8.

283. Guo B, Takala-Harrison S, O'Connor TD: Benchmarking and Optimization of Methods for the Detection of Identity-By-Descent in High-Recombining *Plasmodium falciparum* Genomes. *bioRxiv* 2024.

284. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM: A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* 2012, 6:80-92.

285. Watterson GA: On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* 1975, 7:256-276.

286. Nei M, Li WH: Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci U S A* 1979, 76:5269-5273.

287. DeGiorgio M, Lohmueller KE, Nielsen R: A model-based approach for identifying signatures of ancient balancing selection in genetic data. *PLoS Genet* 2014, 10:e1004561.

288. Miles A, pyup.io b, Rodrigues MF, Ralph P, Kelleher J, Schelker M, Pisupati R, Rae S, Millar T: cggh/scikit-allel: v1.3.13 (v1.3.13). Zenodo; 2024.

289. Weir BS, Cockerham CC: Estimating F-Statistics for the Analysis of Population Structure. *Evolution* 1984, 38:1358-1370.

290. Hamming RW: Error detecting and error correcting codes. *The Bell System Technical Journal* 1950, 29:147-160.

291. Chou RT, Ouattara A, Adams M, Berry AA, Takala-Harrison S, Cummings MP: Positive-unlabeled learning identifies vaccine candidate antigens in the malaria parasite *Plasmodium falciparum*. *NPJ Syst Biol Appl* 2024, 10:44.

292. Ajibola O, Diop MF, Ghansah A, Amenga-Etego L, Golassa L, Apinjoh T, Randrianarivelojosia M, Maiga-Ascofare O, Yavo W, Bouyou-Akotet M, et al: In silico characterisation of putative *Plasmodium falciparum* vaccine candidates in African malaria populations. *Sci Rep* 2021, 11:16215.

293. Naung MT, Martin E, Wong W, Razook Z, Utama D, Guy AJ, Harrison ST, Cowman AF, Lin E, Kiniboro B, et al: Reticulocyte Binding Protein Homologue 5 is a target of balancing selection in the *Plasmodium falciparum* population of Papua New Guinea. *Front Parasitol* 2023, 2:1288867.

294. Kassegne K, Komi Koukoura K, Shen HM, Chen SB, Fu HT, Chen YQ, Zhou XN, Chen JH, Cheng Y: Genome-Wide Analysis of the Malaria Parasite *Plasmodium falciparum* Isolates From Togo Reveals Selective Signals in Immune Selection-Related Antigen Genes. *Front Immunol* 2020, 11:552698.

295. Coulibaly A, Diop MF, Kone A, Dara A, Ouattara A, Mulder N, Miotto O, Diakite M, Djimde A, Amambua-Ngwa A: Genome-wide SNP analysis of *Plasmodium falciparum* shows differentiation at drug-resistance-associated loci among malaria transmission settings in southern Mali. *Front Genet* 2022, 13:943445.

296. Zhang M, Wang C, Otto TD, Oberstaller J, Liao X, Adapa SR, Udenze K, Bronner IF, Casandra D, Mayho M, et al: Uncovering the essential genes of the human malaria parasite *Plasmodium falciparum* by saturation mutagenesis. *Science* 2018, 360.

297. Florens L, Washburn MP, Raine JD, Anthony RM, Grainger M, Haynes JD, Moch JK, Muster N, Sacci JB, Tabb DL, et al: A proteomic view of the *Plasmodium falciparum* life cycle. *Nature* 2002, 419:520-526.

298. Swearingen KE, Lindner SE: Transcriptomics and proteomics reveal two waves of translational repression during the maturation of malaria parasite sporozoites. *Nature Communications* 2019, 10:4964.

299. Treeck M, Sanders JL, Elias JE, Boothroyd JC: The phosphoproteomes of *Plasmodium falciparum* and *gondii* reveal unusual adaptations within and beyond the parasites' boundaries. *Cell Host Microbe* 2011, 10:410-419.

300. Oehring SC, Woodcroft BJ, Moes S, Wetzel J, Dietz O, Pulfer A, Dekiwadia C, Maeser P, Flueck C, Witmer K, et al: Organellar proteomics reveals hundreds of novel nuclear proteins in the malaria parasite *Plasmodium falciparum*. *Genome Biol* 2012, 13:R108.

301. Lasonder E, Rijpma SR, van Schaijk BC, Hoeijmakers WA, Kensche PR, Gresnigt MS, Italiaander A, Vos MW, Woestenenk R, Bousema T, et al: Integrated

transcriptomic and proteomic analyses of *P. falciparum* gametocytes: molecular insight into sex-specific processes and translational repression. *Nucleic Acids Res* 2016, 44:6087-6101.

302. Lasonder E, Janse CJ, van Gemert GJ, Mair GR, Vermunt AM, Douradinha BG, van Noort V, Huynen MA, Luty AJ, Kroeze H, et al: Proteomic profiling of *Plasmodium* sporozoite maturation identifies new proteins essential for parasite development and infectivity. *PLoS Pathog* 2008, 4:e1000195.

303. Lindner SE, Swearingen KE, Shears MJ, Sebastian A, Walker MP, Vrana EN, Hart KJ, Minns AM, Albert I, Sinnis P, et al: Addendum: Transcriptomics and proteomics reveal two waves of translational repression during the maturation of malaria parasite sporozoites. *Nat Commun* 2022, 13:283.

304. El-Manzalawy Y, Munoz EE, Lindner SE, Honavar V: PlasmoSEP: Predicting surface-exposed proteins on the malaria parasite using semisupervised self-training and expert-annotated data. *Proteomics* 2016, 16:2967-2976.

305. Chou RT, Ouattara A, Takala-Harrison S, Cummings MP: *Plasmodium vivax* antigen candidate prediction improves with the addition of *Plasmodium falciparum* data. *NPJ Syst Biol Appl* 2024, 10:133.

306. Berry AA, Obiero JM, Travassos MA, Ouattara A, Coulibaly D, Adams M, de Assis RR, Jain A, Taghavian O, Sy A, et al: Immunoprofiles associated with controlled human malaria infection and naturally acquired immunity identify a shared IgA pre-erythrocytic immunoproteome. *NPJ Vaccines* 2021, 6:115.

307. Ventimiglia NT, Stucke EM, Coulibaly D, Berry AA, Lyke KE, Laurens MB, Bailey JA, Adams M, Niangaly A, Kone AK, et al: Malian adults maintain serologic responses to virulent PfEMP1s amid seasonal patterns of fluctuation. *Sci Rep* 2021, 11:14401.

308. Travassos MA, Niangaly A, Bailey JA, Ouattara A, Coulibaly D, Laurens MB, Pablo J, Jasinskas A, Nakajima-Sasaki R, Berry AA, et al: Seroreactivity to *Plasmodium falciparum* erythrocyte membrane protein 1 intracellular domain in malaria-exposed children and adults. *J Infect Dis* 2013, 208:1514-1519.

309. Zhou AE, Berry AA, Bailey JA, Pike A, Dara A, Agrawal S, Stucke EM, Ouattara A, Coulibaly D, Lyke KE, et al: Antibodies to Peptides in Semiconserved Domains of RIFINs and STEVORs Correlate with Malaria Exposure. *mSphere* 2019, 4.

310. Tewey MA, Coulibaly D, Lawton JG, Stucke EM, Zhou AE, Berry AA, Bailey JA, Pike A, Dara A, Ouattara A, et al: Natural immunity to malaria preferentially targets the endothelial protein C receptor-binding regions of PfEMP1s. *mSphere* 2023, 8:e0045123.

311. Friedman-Klabanoff DJ, Travassos MA, Ifeonu OO, Agrawal S, Ouattara A, Pike A, Bailey JA, Adams M, Coulibaly D, Lyke KE, et al: Epitope-Specific Antibody

Responses to a *Plasmodium falciparum* Subunit Vaccine Target in a Malaria-Endemic Population. *J Infect Dis* 2021, 223:1943-1947.

312. Miura K, Diouf A, Fay MP, Barrett JR, Payne RO, Olotu AI, Minassian AM, Silk SE, Draper SJ, Long CA: Assessment of precision in growth inhibition assay (GIA) using human anti-PfRH5 antibodies. *Malar J* 2023, 22:159.

313. Azasi Y, Gallagher SK, Diouf A, Dabbs RA, Jin J, Mian SY, Narum DL, Long CA, Gaur D, Draper SJ, et al: Bliss' and Loewe's additive and synergistic effects in *Plasmodium falciparum* growth inhibition by AMA1-RON2L, RH5, RIPR and CyRPA antibody combinations. *Sci Rep* 2020, 10:11802.

314. Sacci JB, Jr.: Inhibition of liver-stage development assay. *Methods Mol Med* 2002, 72:517-520.

315. Nicholas J, Kolli SK, Subramani PA, De SL, Ogbondah MM, Barnes SJ, Ntumngia FB, Adams JH: Comparative analyses of functional antibody-mediated inhibition with anti-circumsporozoite monoclonal antibodies against transgenic *Plasmodium berghei*. *Malar J* 2023, 22:335.

316. Nogaro SI, Hafalla JC, Walther B, Remarque EJ, Tetteh KK, Conway DJ, Riley EM, Walther M: The breadth, but not the magnitude, of circulating memory B cell responses to *P. falciparum* increases with age/exposure in an area of low transmission. *PLoS One* 2011, 6:e25582.

317. Patel PN, Diouf A, Dickey TH, Tang WK, Hopp CS, Traore B, Long CA, Miura K, Crompton PD, Tolia NH: A strain-transcending anti-AMA1 human monoclonal antibody neutralizes malaria parasites independent of direct RON2L receptor blockade. *Cell Rep Med* 2025, 6:101985.

318. Tripathi P, Bender MF, Lei H, Da Silva Pereira L, Shen CH, Bonilla B, Dillon M, Ou L, Pancera M, Wang LT, et al: Cryo-EM structures of anti-malarial antibody L9 with circumsporozoite protein reveal trimeric L9 association and complete 27-residue epitope. *Structure* 2023, 31:480-491 e484.

319. Thompson HA, Hogan AB, Walker PGT, White MT, Cunnington AJ, Ockenhouse CF, Ghani AC: Modelling the roles of antibody titre and avidity in protection from *Plasmodium falciparum* malaria infection following RTS,S/AS01 vaccination. *Vaccine* 2020, 38:7498-7507.

320. Douglas AD, Baldeviano GC, Jin J, Miura K, Diouf A, Zenonos ZA, Ventocilla JA, Silk SE, Marshall JM, Alanine DGW, et al: A defined mechanistic correlate of protection against *Plasmodium falciparum* malaria in non-human primates. *Nat Commun* 2019, 10:1953.

321. Pandey AK, Reddy KS, Sahar T, Gupta S, Singh H, Reddy EJ, Asad M, Siddiqui FA, Gupta P, Singh B, et al: Identification of a potent combination of key

Plasmodium falciparum merozoite antigens that elicit strain-transcending parasite-neutralizing antibodies. *Infect Immun* 2013, 81:441-451.

322. Arora G, Hart GT, Manzella-Lapeira J, Doritchamou JY, Narum DL, Thomas LM, Brzostowski J, Rajagopalan S, Doumbo OK, Traore B, et al: NK cells inhibit *Plasmodium falciparum* growth in red blood cells via antibody-dependent cellular cytotoxicity. *Elife* 2018, 7.

323. Leitner WW, Haraway M, Pierson T, Bergmann-Leitner ES: Role of Opsonophagocytosis in Immune Protection against Malaria. *Vaccines (Basel)* 2020, 8.

324. Kurtovic L, Behet MC, Feng G, Reiling L, Chelimo K, Dent AE, Mueller I, Kazura JW, Sauerwein RW, Fowkes FJI, Beeson JG: Human antibodies activate complement against *Plasmodium falciparum* sporozoites, and are associated with protection against malaria in children. *BMC Med* 2018, 16:61.

325. Jafarshad A, Dziegiel MH, Lundquist R, Nielsen LK, Singh S, Druilhe PL: A novel antibody-dependent cellular cytotoxicity mechanism involved in defense against malaria requires costimulation of monocytes FcgammaRII and FcgammaRIII. *J Immunol* 2007, 178:3099-3106.

326. Berry AA, Gottlieb ER, Kouriba B, Diarra I, Thera MA, Dutta S, Coulibaly D, Ouattara A, Niangaly A, Kone AK, et al: Immunoglobulin G subclass and antibody avidity responses in Malian children immunized with *Plasmodium falciparum* apical membrane antigen 1 vaccine candidate FMP2.1/AS02(A). *Malar J* 2019, 18:13.

327. Taylor DW, Bobbili N, Kayatani A, Tassi Yunga S, Kidima W, Leke RFG: Measuring antibody avidity to *Plasmodium falciparum* merozoite antigens using a multiplex immunoassay approach. *Malar J* 2020, 19:171.

328. Korfmann K, Gaggiotti OE, Fumagalli M: Deep Learning in Population Genetics. *Genome Biol Evol* 2023, 15.

329. Isildak U, Stella A, Fumagalli M: Distinguishing between recent balancing selection and incomplete sweep using deep neural networks. *Mol Ecol Resour* 2021, 21:2706-2718.