





Plasmodium falciparum Genetic Diversity in Coincident Human and Mosquito Hosts

 Zena Lapp,^a Andrew A. Obala,^b Lucy Abel,^c David A. Rasmussen,^{d,e} Kelsey M. Sumner,^{f,g} Elizabeth Freedman,^g
 Steve M. Taylor,^{a,g} Wendy Prudhomme-O'Meara^{a,g}

^aDuke Global Health Institute, Duke University, Durham, North Carolina, USA

^bSchool of Medicine, College of Health Sciences, Moi University, Eldoret, Kenya

^cAcademic Model Providing Access to Healthcare (AMPATH), Eldoret, Kenya

^dDepartment of Entomology and Plant Pathology, North Carolina State University, Raleigh, North Carolina, USA

^eBioinformatics Research Center, North Carolina State University, Raleigh, North Carolina, USA

^fDepartment of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, North Carolina, USA

^gDivision of Infectious Diseases, School of Medicine, Duke University, Durham, North Carolina, USA

Steve M. Taylor and Wendy Prudhomme-O'Meara contributed equally to this work. The order reflects WPO's leadership in conceptualization.

ABSTRACT Population genetic diversity of *Plasmodium falciparum* antigenic loci is high despite large bottlenecks in population size during the parasite life cycle. The prevalence of genetically distinct haplotypes at these loci, while well characterized in humans, has not been thoroughly compared between human and mosquito hosts. We assessed parasite haplotype prevalence, diversity, and evenness using human and mosquito *P. falciparum* infections collected from the same households during a 14-month longitudinal cohort study using amplicon deep sequencing of two antigenic gene fragments (*ama1* and *csp*). To a prior set of infected humans ($n = 1,175/2,813$; 86.2% sequencing success) and mosquito abdomens ($n = 199/1,448$; 95.5% sequencing success), we added sequences from infected mosquito heads ($n = 134/1,448$; 98.5% sequencing success). The overall and sample-level parasite populations were more diverse in mosquitoes than in humans. Additionally, haplotype prevalences were more even in the *P. falciparum* human population than in the mosquito population, consistent with balancing selection occurring at these loci in humans. In contrast, we observed that infections in humans were more likely to harbor a dominant haplotype than infections in mosquitoes, potentially due to removal of unfit strains by the human immune system. Finally, within a given mosquito, there was little overlap in genetic composition of abdomen and head infections, suggesting that infections may be cleared from the abdomen during a mosquito's lifespan. Taken together, our observations provide evidence for the mosquito vector acting as a reservoir of sequence diversity in malaria parasite populations.

IMPORTANCE *Plasmodium falciparum* is the deadliest human malaria parasite, and infections consisting of concurrent, multiple strains are common in regions of high endemicity. During transitions within and between the parasite's mosquito and human hosts, these strains are subject to population bottlenecks, and distinct parasite strains may have differential fitness in the various environments encountered. These bottlenecks and fitness differences may lead to differences in strain prevalence and diversity between hosts. We investigated differences in genetic diversity and evenness between *P. falciparum* parasites in human and mosquito hosts collected from the same households during a 14-month longitudinal study in Kenya. Compared to human parasite populations and infections, *P. falciparum* parasites observed in mosquito populations and infections were more diverse by multiple population genetic metrics. This suggests that the mosquito vector acts as a reservoir of sequence diversity in malaria parasite populations.

Editor Louis H. Miller, NIAID/NIH

Copyright © 2022 Lapp et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Steve M. Taylor, steve.taylor@duke.edu.

The authors declare no conflict of interest.

Received 19 August 2022

Accepted 22 August 2022

Published 8 September 2022

KEYWORDS malaria, *Plasmodium falciparum*, transmission, comparative genomics, genetic diversity

Plasmodium falciparum has a complex life cycle that requires it to navigate multiple cellular and host transitions to sustain transmission. These include transitions both between human and mosquito hosts and between compartments within those hosts. In addition, distinct genotypes may be cotransmitted between hosts in a single bite or may accumulate within a host owing to serial superinfections. Such infections consisting of many different strains are particularly commonplace in highly endemic settings, such as some regions of sub-Saharan Africa (1), promoting both outcrossing in mosquito hosts and competition in human hosts. These factors, coupled with population bottlenecks and selective pressures encountered by *P. falciparum* throughout its life cycle, shape overall patterns of parasite genetic diversity and evenness (2).

Comparative population genetics of *P. falciparum* between the hosts and cellular environments through which the parasite transitions in natural cycles of transmission remains relatively unexplored. Several studies have compared markers of drug resistance loci between hosts, and an early report from Zambia observed very different allele frequencies in humans and mosquitoes (3, 4), suggesting differences in parasite population structure between hosts. However, subsequent reports from other settings using different genetic markers have not consistently observed this phenomenon (5, 6). As these studies used marker genes with few polymorphisms, analyses of individuals with complex coinfections were limited. While microsatellite markers overcome some of these limitations (7, 8), prior studies have not, to our knowledge, contrasted the genetic composition or diversity of highly polymorphic targets in naturally occurring infections of humans and mosquitoes that are participating in coincident transmission networks. By exploring these phenomena more closely, we can better understand what factors contribute to the diversity of malaria parasite populations.

We investigated variability in *P. falciparum* population genetic metrics across human and mosquito hosts in an area of high endemicity in Western Kenya. During a 14-month longitudinal cohort study, we detected *P. falciparum* parasites in human participants and in the heads and abdomens of resting Anopheline mosquitoes collected from their households (1). From each *P. falciparum* infection, we used amplicon deep sequencing of polymorphic segments of the parasite genes encoding apical membrane antigen 1 (*ama1*) and circumsporozoite protein (*csp*) to catalog complex *P. falciparum* infections in human blood, mosquito abdomens, and mosquito heads. We previously reported that parasite multiplicity of infection (MOI) as expressed by either marker was higher in mosquito abdomens harboring recently ingested parasites than in humans harboring blood-stage parasites (1). Due to this observation, and the robust immune defenses against *P. falciparum* in humans (9), we hypothesized that the mosquito host acts as a reservoir of sequence diversity of *P. falciparum* antigenic genes.

RESULTS

Data overview and analytic population. Samples were collected over the course of 14 months (June 2017 to July 2018) from 38 households in three Kenyan villages. Mosquitoes were aspirated weekly from each household, and blood samples from household members were collected monthly. To the previously reported data on humans and mosquito abdomens (1), we added data from mosquito heads/thoraces (here denoted as heads). Over one-third of human samples (41.8%; 1,175/2,813) contained *P. falciparum* compared to 13.7% (199/1,448) of mosquito abdomens and 9.2% (134/1,462) of mosquito heads (Fig. S1 in the supplemental material). Of these, sequencing of at least one marker was successful in 86.2% (1,013/1,175) of human, 95.5% (190/199) of mosquito abdomen, and 98.5% (132/134) of mosquito head infections. Haplotype information from these 1,013 infections in 224 people and 322 infections in 244 mosquitoes constituted the analytic population.

Mosquito head infections are not a subset of their abdomen infections. Parasite development within the mosquito host begins in the abdomen, following which sporozoites must traverse the midgut wall to reach the salivary glands in the head; however, it is not

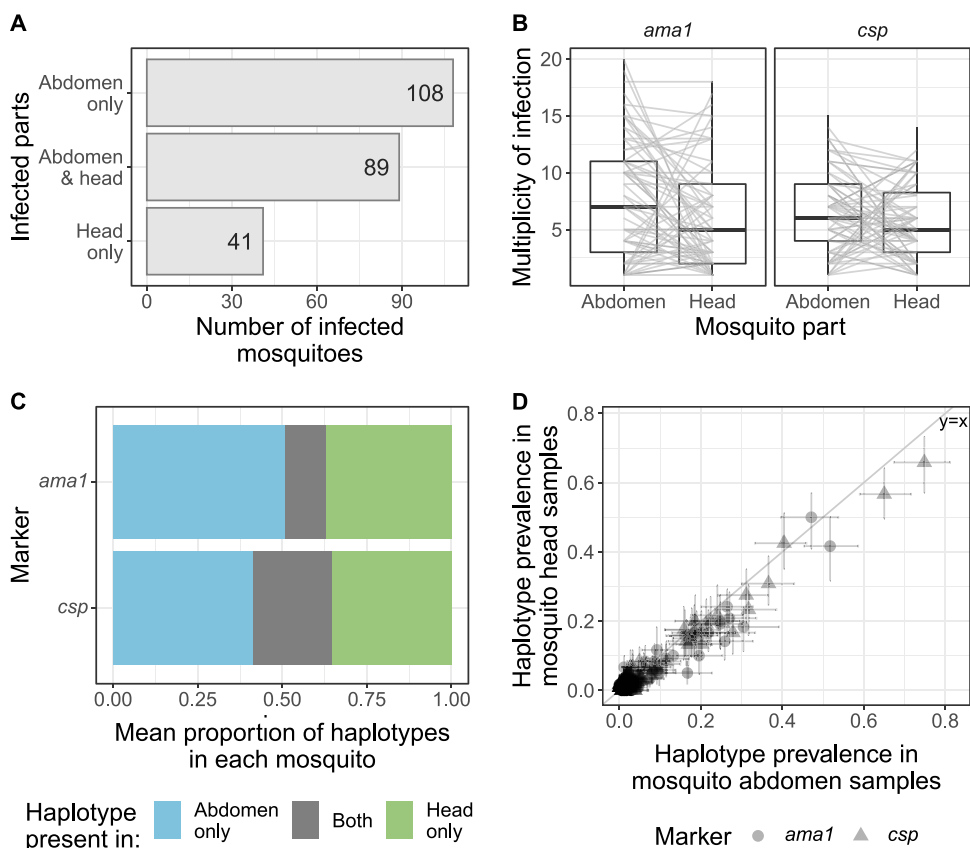


FIG 1 Mosquito abdomens and heads do not contain similar infections. (A) *P. falciparum* infection of mosquito abdomens and heads of mosquitoes for which both compartments were tested using PCR. (B) Multiplicity of infection of infected mosquito abdomen and head samples. Gray lines link MOIs from an individual mosquito. (C) For 89 mosquitoes with infections of both the abdomen and the head, the proportions of the set of haplotypes in the mosquito found in the abdomen only, head only, or both are shown. The mean counts for each of the three groups were used to obtain the proportions. (D) Prevalence of each *ama1* and *csp* haplotype in the mosquito abdomen population compared to the mosquito head population. Each dot represents a unique *ama1* or *csp* haplotype, and bars indicate the 95% bootstrapped confidence intervals.

known how quick and comprehensive this egress is. We hypothesized that if both midgut and salivary gland infections persist throughout the mosquito's lifespan (i.e., incomplete egress from the midgut), haplotypes in a mosquito's head would be a subset of those in the abdomen. Among mosquitoes in which at least one compartment was infected, *P. falciparum* was detected in both the abdomen and the head in 89/238 (37.4%), in only the abdomen in 108/238 (45.4%), and in only the head in 41/238 (17.2%) (Fig. 1A). The latter finding suggests that infections may be completely cleared from the abdomen within the span of a mosquito's lifetime.

We next compared the haplotype compositions of infections in the 89 mosquitoes in which *P. falciparum* was detected in both the head and the abdomen. Median multiplicity of infection (MOI) was generally higher in the abdomen (*csp* = 6; *ama* = 7) than in the head (*csp* = 5; *ama* = 5), but this trend was not consistent at the individual mosquito level (Fig. 1B). We calculated the percentage of *ama1* or *csp* haplotypes found only in the head or the abdomen or observed in both compartments (i.e., the Jaccard distance; intersect/union) within each mosquito. While some haplotypes were observed in both compartments of a given mosquito (mean for *ama1* = 12.0% and *csp* = 23.7%), the majority of haplotypes were either private to the abdomen (mean for *ama1* = 50.7% and *csp* = 41.5%) or head (mean for *ama1* = 37.3% and *csp* = 34.8%) (Fig. 1C and Fig. S2A and B). Despite this limited overlap, sharing between abdomens and heads from the same mosquito was higher than sharing between random pairs of abdomens and heads (Fig. S2C; Kolmogorov-Smirnov test, $P \leq 1 \times 10^{-10}$ for both markers). Additionally, we observed no difference in haplotype overlap between

abdomens and heads from mosquitoes who had recently fed and those who had not (Fig. S2D; both Wilcoxon rank-sum tests, $P > 0.2$).

To determine whether the differences in haplotype composition between abdomen and head infections within a single mosquito corresponded to differences at the host population level, we compared between mosquito compartments haplotype population-level prevalences, defined as the number of samples in which a haplotype was observed. Both *ama1* and *csp* haplotype prevalences were similar between mosquito abdomen and head populations (Fig. 1D), suggesting that the transition from oocyst to sporozoite does not alter the diversity of circulating parasites. Owing to this population-level similarity in prevalences and our observation that abdomen and head parasite populations from the same mosquito appear to frequently represent different infections, we subsequently performed all comparisons between the two *P. falciparum* hosts humans and mosquitoes, where mosquito samples included both abdomen and head samples.

The *P. falciparum* population in mosquitoes is more diverse and more even than the population in humans. To investigate signatures of differential bottlenecks or selection during parasite transition between mosquito and human hosts, we compared population-level differences in parasite haplotype prevalence among mosquitoes and humans, where differences in prevalence may indicate differential bottlenecks or selection. Across all infections, we observed high haplotype richness, with 456 *ama1* and 298 *csp* distinct haplotypes. The vast majority of these were low-frequency haplotypes, many of which were observed in only one host (Fig. S3 and S4). Among 54 distinct haplotypes (both *ama1* and *csp*) with a prevalence above 5% across all samples, we observed 28 haplotypes with differential prevalence across hosts; 19 were more common in mosquito infections, and 9 were more common in human infections (Fig. 2A), consistent with our observation of higher average mosquito MOIs (1) (Fig. S5).

We next used haplotype prevalence to quantify population-level diversity across orders of diversity (q) ranging from equal weight to each haplotype ($q = 0$, equivalent to haplotype richness or the number of distinct haplotypes observed) to downweighting rare haplotypes ($q = 2$, effective number of highly abundant haplotypes) (10). The mosquito parasite population was more diverse than the parasite population in human hosts (Fig. 2B and Fig. S6A). This trend is consistent even when accounting for differences in sample size between hosts, multiple samples per host, different sampling schemes between hosts, potential differences between symptomatic and asymptomatic infections, differences in MOI, haplotypes with rare variants, and limitations of using estimated diversity (Fig. S6B) as well as across villages (Fig. S6C) and transmission seasons (Fig. S6D). To determine whether there may be differences in the changes in diversity during the transition between different parasite host compartments, we also compared separately the diversity of the parasite populations of humans, mosquito abdomens, and mosquito heads (Fig. S6E). The parasite populations of mosquito abdomens and heads were relatively similar, and the human parasite population was more diverse than the parasite population in both mosquito abdomens and heads. Moreover, as evidenced by the steeper decline in diversity with increasing q in mosquitoes than that observed in humans (Fig. 2B), haplotype prevalences were more uneven in the parasite population of the mosquito (Fig. 2C), particularly in the head (Fig. S7A). These trends were generally consistent, although not always significant, when performing the same sensitivity analyses as for the diversity comparison (Fig. S7B to D). This indicates that mosquitoes contained a larger relative number of infrequent haplotypes than humans. Even so, higher diversity in the mosquito host is still apparent when downweighting the contribution of these minor haplotypes. Taken together, these results indicate that there may be a greater relative loss in diversity across the transition from mosquitoes to humans than from humans to mosquitoes.

Dominant haplotypes within infections are more common in humans than in mosquitoes. In addition to lower population-level diversity in humans than in mosquitoes, we also observed lower within-sample diversity (Fig. S5) and proportionately more monoclonal infections in humans (Fig. 3A; both Fisher's exact test, $P < 1 \times 10^{-12}$). To further investigate whether human infections are more often dominated by one or a few haplotypes than mosquito infections, we calculated the haplotype evenness for each infection, which

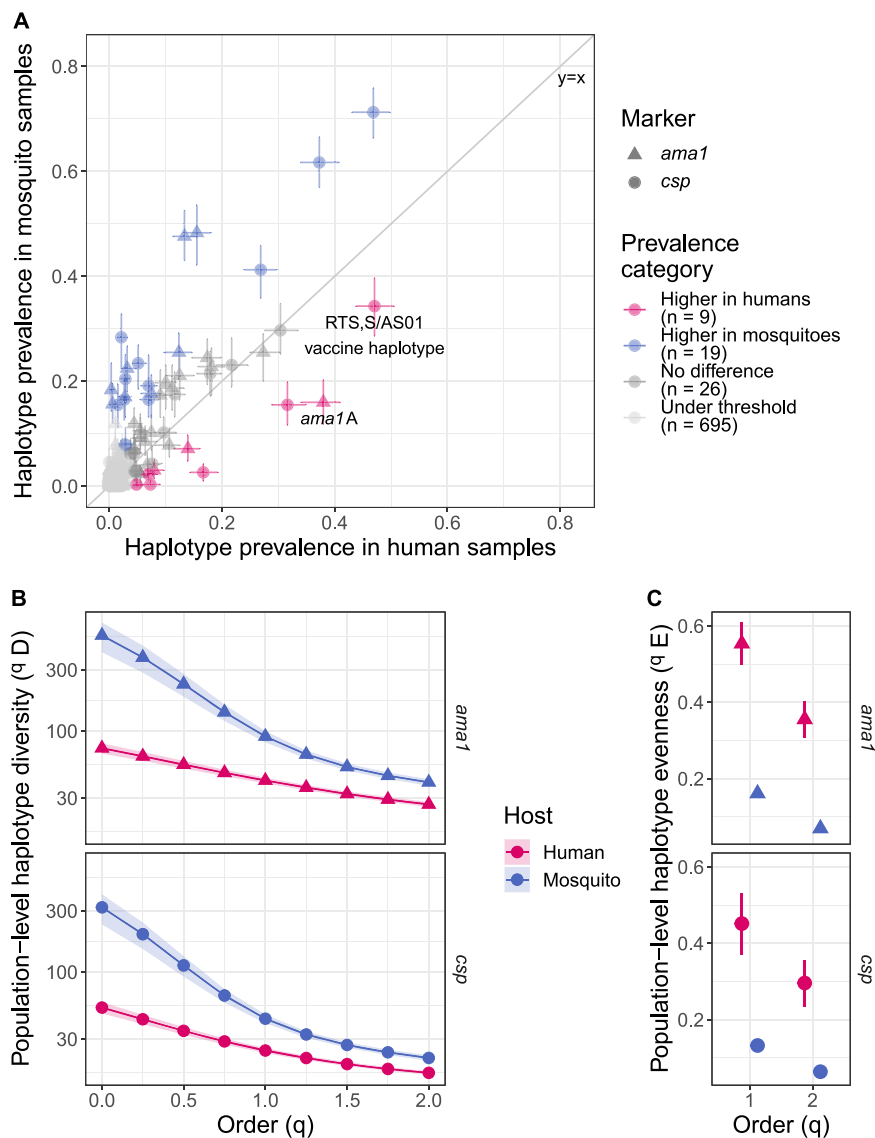


FIG 2 The mosquito *P. falciparum* population is more diverse and less even than the human *P. falciparum* population. (A) Prevalence of each *ama1* and *csp* haplotype in the human population compared to the mosquito population. Each dot represents a unique *ama1* or *csp* haplotype, and bars indicate the 95% bootstrapped confidence intervals. The lower threshold was defined as haplotypes observed in fewer than 5% of combined human and mosquito samples. Haplotype prevalences were considered higher in one compartment if the 95% bootstrapped confidence intervals did not overlap the expected prevalence (i.e., the overall prevalence across all samples). (B) Diversity of *P. falciparum* populations by host and genetic marker across orders of diversity. Ribbons are bootstrapped 95% confidence intervals. Higher values indicate more diversity. The slope of the line across orders *q* is a measure of haplotype evenness in the population. (C) Haplotype evenness of human and mosquito samples. Bars are bootstrapped 95% confidence intervals. Higher values indicate more similar prevalence of haplotypes in the population.

examines the haplotype abundance within an infection based on sequencing reads. A lower value indicates that the infection consists of mostly reads from a single haplotype or, in other words, is dominated by a majority haplotype. Median evenness values for mosquito infections were higher (*ama1* = 0.88; *csp* = 0.84) than those for human infections (*ama1* = 0.51; *csp* = 0.67; all $P < 1 \times 10^{-4}$) (Fig. 3B). This observation was robust to taking the maximum evenness among the two markers and to differences in haplotype filtering (all $P < 1 \times 10^{-10}$) (Fig. S8). This differential composition of polyclonal *P. falciparum* infections between hosts supports a differential in selective landscapes that may further enable the preservation of diverse *P. falciparum* populations in Anopheline mosquitoes.

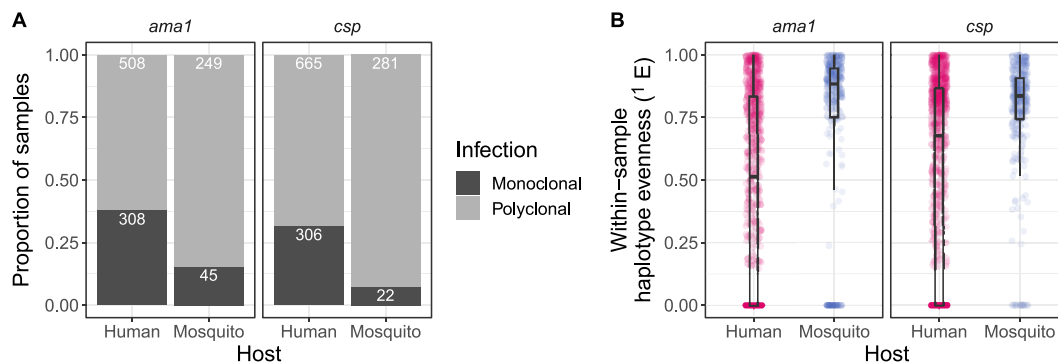


FIG 3 Compared to mosquito samples, human samples are more often dominated by a single haplotype. (A) Proportion of samples with mono-clonal and poly-clonal infections. Numbers are counts for each category. (B) Distributions of within-sample haplotype evenness ($q = 1$) by genetic marker and host. Lower values indicate more dominance by individual haplotypes within the strain mixture.

DISCUSSION

We compared *P. falciparum* genetic diversity across several host compartments that the parasite must successfully navigate to sustain transmission. Parasite genetic diversity was increased relative to humans during the mosquito stages. In addition, individual infections were composed differently in mosquitoes and humans, with human infections more commonly harboring dominant members. Collectively, our observations suggest that mosquito-stage infections not only maintain diversity in *P. falciparum* populations through recombination but also act as a reservoir of parasite sequence diversity.

We observed, using multiple metrics, more parasite genetic diversity in mosquitoes than in humans. This high diversity contrasts with the known marked reduction in parasite biomass during the transition from the human to the mosquito abdomen (11), which might be expected to constrain parasite diversity. One potential explanation for this is the possibility of cryptic genotypes in humans undetected by marker sequencing; this has been reported in experimental studies (7), although the large range of MOIs we observed in humans suggests that these infections were not systematically undersampled. Alternatively, the reduced diversity in humans could result from large reductions in population size and negative selective pressures as the parasite passes from mosquitoes to the human liver and into the blood stage. Mosquitoes are the location of parasite sexual recombination and therefore certainly provide a site for genomic diversification, but this seems unsuited to explain the diversity of these short segments in *ama1* and *csp* that do not harbor known recombination hot spots (12). A probable contributor to this high mosquito diversity is multiple or interrupted feeds on infected hosts, which would allow strains to accumulate in the mosquito abdomen. This feeding behavior has been reported for *Anopheles gambiae* (13), may be enhanced by human *P. falciparum* infection (14), and may increase the likelihood of subsequent infections (15). Additionally, *P. falciparum* adaptation to evade the immune system of local *Anopheles* strains (16), as well as imposition of selective pressures on the *Anopheles* vector to the parasites' benefit (17), may reduce differences in fitness between distinct parasite strains within the mosquito and lead to an accumulation of genetic diversity at the population level. However, some of these novel strains may be unfit to survive the human host. Indeed, prior work on arbovirus infection found an accumulation of mutations in mosquitoes that led to fitness costs during vertebrate infection (18). Despite these plausible explanations for constrained diversity in humans and higher diversity in mosquitoes, the mechanism by which mosquitoes maintain such high parasite diversity when their parasite population is necessarily sampled from the less diverse human population remains to be fully elucidated.

Within individual infections, we observed higher dominance of haplotypes in human infections than in mosquito infections, while on a larger scale, the *P. falciparum* haplotype population was more evenly distributed among humans than among mosquitoes. These differences may result from the differential selection landscapes between hosts, in particular for the proteins encoded by our gene targets AMA1 and CSP, which harbor epitopes that

are known targets of functional human immunity (19). In humans, the concurrent maintenance in the population of multiple viable alleles due to balancing selection, paired with the removal of deleterious alleles due to negative selection, could produce a relatively high evenness of haplotypes in the human parasite population even as individual infections are shaped by directional selection resulting from individual host immune responses. In contrast, the relative lack of differential fitness in the mosquito host described above may lead to even parasite strain abundances within a mosquito.

Comparison of paired abdomens and heads from the same mosquito revealed striking differences between *P. falciparum* presence and haplotype composition. As expected, given the delay between midgut and salivary gland infections, many mosquitoes had haplotypes private to the abdomen that were not present in the head. More surprising was the observation of mosquitoes with haplotypes private to the head that were absent from the abdomen, suggesting that infections do not reliably persist in a mosquito's abdomen throughout its lifespan. While these differences may again be due to cryptic haplotypes, the identification of mosquitoes with infections in the head but not the abdomen using sensitive PCR detection methods (20, 21) indicates that cryptic haplotypes likely cannot explain all of the observed differences. Despite these discrepancies between abdomens and heads from a given mosquito, at the population level, haplotype composition and diversity were similar between mosquito abdomens and heads, suggesting that the selective pressures for or against certain haplotypes (or lack thereof) may be similar in these two compartments.

Our findings highlight the role of the mosquito host in harboring sequence diversity of *P. falciparum* parasites. A unique feature of genetic diversity in *P. falciparum* compared to other organisms is the preponderance of low-frequency alleles (22). A prior modeling study suggested that this phenomenon may be the result of the complex, "unconventional" life cycle of *P. falciparum*, specifically the bottlenecks and host transitions that intensify both random genetic drift as well as natural selection (2). Consistent with this, we observed many haplotypes private to one host, which was more prominent in mosquitoes. As noted above, meiotic recombination is unlikely to be the main contributor to the diversity we cataloged, and the mechanisms by which these low-frequency and private alleles arise remain obscure. However, our observations furnish compelling evidence for a role of the mosquito vector in accumulating genetic diversity in genic regions likely not under positive selection in mosquitoes. While it appears that the majority of these polymorphisms tend to be selected against in humans, the diversity in mosquitoes nevertheless acts as a continual supply of novel alleles and allelic combinations that may by chance be advantageous and exploited by the parasite during human infection.

This study has limitations. First, the inability to sample parasites from mosquitoes without sacrificing them precludes a comprehensive study of paired mosquito abdomen and head infections over time. Even so, we were still able to identify similarities and differences between the haplotype populations in these two compartments. Second, parasites accumulate in mosquitoes over multiple blood meals, which presumably increases diversity in mosquitoes; however, this phenomenon is likely balanced by parasite accumulation in humans owing to superinfection. Third, the mosquito and human sampling schemes were different, potentially biasing sampling comprehensiveness between hosts. To mitigate the risk that this potential imbalance influenced our results, we performed comparative population analyses using empirical methods with a fixed coverage threshold (10) and sensitivity analyses. Fourth, many of the human and mosquito infections had very low parasite densities, which not only increases the possibility of failing to detect infections but also increases the possibility of false haplotype discovery (23). To reduce the inclusion of false haplotypes to the greatest extent possible, we performed strict haplotype censoring to remove potential false positives (1) and performed sensitivity analyses on key findings to determine whether haplotype filtering criteria influenced the results.

In conclusion, our comparison of *P. falciparum* haplotypes observed in natural, coincident infections of humans, mosquito abdomens, and mosquito heads revealed greater genetic diversity in mosquitoes than in human populations and infections. This provides evidence for

the role of the mosquito vector in harboring sequence diversity of the malaria parasite population.

MATERIALS AND METHODS

Ethics statement. All adults and parents or legal guardians for individuals under 18 years old provided written informed consent. Children over 8 years old also provided verbal assent. The study was approved by the ethical review boards of Moi University (2017/36) and Duke University (Pro00082000).

Study design and sampling. The study design and sample processing have been described previously (1). Briefly, a longitudinal cohort of participants (1 year of age or older) residing in 38 households in three villages in Western Kenya were followed from June 2017 to July 2018. For each participant, dried blood spots (DBS) were collected monthly and at any time participants had malaria symptoms. One morning each week, indoor resting mosquitoes were collected from participant households using vacuum aspiration, and, following morphologic identification, the abdomen was separated from the head and thorax of female *Anopheles* mosquitoes. Genomic DNA was isolated from DBS, mosquito abdomens, and mosquito heads/thoraces; *P. falciparum* was detected in these extracts using a real-time PCR assay. Segments of approximately 300 nucleotides of the *P. falciparum* *ama1* and *csp* genes were amplified and sequenced on an Illumina MiSeq platform. Negative controls were included for all PCR amplification assays; all negative controls were negative. Haplotype inference was performed using DADA2 v1.8 (24) with custom read and haplotype filtering as described in Sumner et al. (1). Briefly, haplotypes were removed if they were supported by fewer than 250 reads or fewer than 3% of the total number of reads for a sample, deviated from the expected length, were a minority haplotype differing by only one variant position from a majority haplotype present with greater than 8 times the read depth, or contained a variant position only variable within that haplotype.

The output was a set of quality-filtered *ama1* and *csp* reads and corresponding parasite haplotypes for each *P. falciparum* infection. Sequencing was only possible at two targets because of the combination of a very small amount of material for each sample and very low infection density for many samples.

We performed parallel analyses of amplicon deep-sequenced segments of the *P. falciparum* *ama1* and *csp* marker genes. Since *ama1* and *csp* are unlinked markers found on different chromosomes, to some extent, these parallel analyses can be considered pseudoreplicates, where similar results for both markers increase confidence in our findings.

Within-mosquito comparison. For each mosquito with a *P. falciparum* infection in both the abdomen and the head, the Jaccard distance (25) was calculated for the haplotypes in the abdomen-head pair,

$$J(H_a, H_h) = \left| \frac{H_a \cap H_h}{H_a \cup H_h} \right| \quad (1)$$

where H_a is the set of haplotypes in the abdomen, and H_h is the set of haplotypes in the head.

Haplotype prevalence. For each haplotype, population-level prevalence was determined for 5 distinct populations: the entire sample set, all human samples, all mosquito samples, mosquito abdomens, and mosquito heads. Prevalence was calculated as the proportion of samples harboring that haplotype; 95% confidence intervals were computed from 100 bootstrapped data sets. Haplotypes were considered low frequency if they occurred in fewer than 5% of all samples. Haplotypes that were not low frequency (i.e., above a threshold of 5% prevalence) were considered higher in a given compartment if the range of bootstrapped prevalences did not overlap the expected prevalence (i.e., the overall prevalence across all samples).

Randomized minimum spanning trees. To visualize the relatedness of haplotypes, we calculated pairwise distances using the `dist.dna()` function in the R package `ape` v5.6-2 (26) with the K80 evolutionary model, computed randomized minimum spanning trees (27) using the `rmst()` function in `pegas` v1.1 (28), and visualized the trees in `ggtree` v3.0.4 (29).

Diversity and evenness. For analyses between humans and mosquitoes, all mosquito abdomen and head samples were considered mosquito samples, providing a maximum of 2 samples from each mosquito. All human samples were included unless otherwise noted, as individuals may acquire additional infections over time and the relative abundance of coinfecting strains may change over time.

(i) Population level. For the set of mosquito samples and the set of human samples, we calculated the population-level diversity of haplotypes, rarefaction curves, and population evenness using the R packages `iNEXT.4steps` v1.0.1 (10) and `iNEXT.3D` v1.0.1 (30).

Diversity was calculated using the following equation (10, 31):

$${}^qD = \left(\sum_{i=1}^n p_i^q \right)^{1/(1-q)} \quad (2)$$

where q is the order of diversity, n is the number of distinct haplotypes, and p_i is the prevalence of haplotype i in the sample set. D was computed across a range of values q between 0 and 2, where higher numbers correspond to upweighting haplotypes that are more abundant in the overall population. True diversity was not accurately calculable for low orders of diversity ($q < 1$) due to an abundance of unsampled rare haplotypes. Therefore, to enable comparison of diversity between the human and mosquito haplotype populations, we calculated the empirical diversity at a standardized coverage of the host population's haplotypes (90.1% for *ama1* and 94.7% for *csp*).

Because haplotypes observed within multiple samples from a single individual may either be correlated or more easily detectable owing to repeated testing, we performed a sensitivity analysis in which we included a

single sample from each individual (i.e., one sample from each mosquito and one sample from each person). Additional sensitivity analyses were performed (i) by using true (asymptotic) diversity (for $q \geq 1$), (ii) by using a subsampled data set including the same number of human and mosquito samples, (iii) by using a subsampled data set including the same number of host samples per week (to account for differences in mosquito and human sampling schemes), (iv) by using a data set excluding samples from symptomatic humans to reduce the possibility of undetected cryptic haplotypes that may be present in high-density symptomatic infections, (v) by defining p_i as the frequency of haplotypes in the total set of haplotypes (to account for differences in MOI between infections), and (vi) by using a data set including only haplotypes with variation at amino acid positions that are variable in both hosts (to limit potential false positives by using this stricter set of haplotype-filtering criteria). We also calculated diversity by village and season (we defined high season as April to September, inclusive).

We calculated haplotype evenness using the following equation (32):

$${}^qE = \frac{{}^qD - 1}{H - 1} \quad (3)$$

where H is the haplotype richness or the number of distinct haplotypes in the population. For $q = 0$, evenness is defined as 1, and for $H = 1$, evenness is defined as 0. The same sensitivity analyses described above for diversity were performed for evenness.

(ii) Within sample. For each sample, we computed haplotype diversity and evenness using equations 1 and 2. In this case, p_i in equation 1 is the relative read abundance of each haplotype, qD is the within-host diversity, and H is the MOI of the infection.

To compare evenness between human and mosquito hosts, we computed a zero-one inflated Beta regression model using the R package `gamlss v5.4-1` (33) with host as the main exposure, evenness as the outcome, \log_2 -transformed haplotype reads as a covariate, and individual as a random effect. To determine whether incorporating information from both markers influenced differences in evenness between hosts, for each sample, we selected the highest evenness value (between *ama1* and *csp*) and compared these values between humans and mosquitoes. Finally, to explore if evenness values were biased by the initial enforcement of haplotype quality-filtering criteria that were partially based on within-sample haplotype proportion, we performed a sensitivity analysis using unfiltered haplotypes. These haplotypes were inferred by DADA2 v1.8 (24) from input reads that passed upstream read quality filtering. Using these unfiltered haplotypes, we used the same methods as above to compute and compare evenness.

Data analysis and visualization. Comparison across groups was performed using Wilcoxon rank-sum tests, Fisher's exact tests, or Kolmogorov-Smirnov tests. All data analyses and visualization were performed in R v4.1.1 (34) and RStudio v2021.9.0.351 (<https://www.rstudio.com/>) using the following packages: `tidyverse v1.3.1` (35), `ape v5.6-2` (26), `cowplot v1.1.1` (36), `scales v1.1.1` (37), and `ggtext v0.1.1` (38).

Data availability. Raw sequencing reads are available from NCBI (BioProject [PRJNA646940](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA646940)) (1, 39). All processed data and code to reproduce the analyses and figures can be found on GitHub (40).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, PDF file, 0.006 MB.

FIG S2, PDF file, 0.04 MB.

FIG S3, PDF file, 0.006 MB.

FIG S4, PDF file, 0.1 MB.

FIG S5, PDF file, 0.02 MB.

FIG S6, PDF file, 0.05 MB.

FIG S7, PDF file, 0.01 MB.

FIG S8, PDF file, 0.2 MB.

ACKNOWLEDGMENTS

We thank the field technicians in Webuye for their engagement with the study participants, I. Khaoya, L. Marango, E. Mukeli, E. Nalianya, J. Namae, L. Nukewa, E. Wamalwa, and A. Wekesa; J. Kipkoach Kirui (AMPATH) for operational assistance; V. Liao, A. Nantume, S. Kim, and J. Saelens (each of Duke University) for laboratory sample and data processing; and C. Markwalter (Duke University) for thoughtful discussion. Ultimately, we are indebted to the households for their participation in this study.

Preliminary results were presented at the Annual Meeting of the American Society of Tropical Medicine and Hygiene in November 2020.

We declare no conflict of interest.

D.A.R., K.M.S., S.M.T., W.P.O., and Z.L. conceptualized the study. E.F., S.M.T., W.P.O., K.M.S., and Z.L. developed the methodology. Z.L. and K.M.S. curated the data, performed formal analysis, and performed the computer programming. Z.L. visualized the data. W.P.O., S.M.T., K.M.S., and D.A.R. acquired funding. Z.L., L.A., E.F., and K.M.S. performed the investigation.

S.M.T., W.P.O., A.A.O., and L.A. administered and supervised the project. Z.L., S.M.T., and W.P.O. wrote the original manuscript draft. All authors reviewed and edited the manuscript.

This work was supported by NIAID (R21AI126024 to W.P.O. and R01AI146849 to W.P.O. and S.M.T.) and the Triangle Center for Evolutionary Medicine (Graduate Student Award to K.M.S.).

REFERENCES

- Sumner KM, Freedman E, Abel L, Obala A, Pence BW, Wesolowski A, Meshnick SR, Prudhomme-O'Meara W, Taylor SM. 2021. Genotyping cognate *Plasmodium falciparum* in humans and mosquitoes to estimate onward transmission of asymptomatic infections. *Nat Commun* 12:909. <https://doi.org/10.1038/s41467-021-21269-2>.
- Chang H-H, Moss EL, Park DJ, Ndiaye D, Mboup S, Volkman SK, Sabeti PC, Wirth DF, Neafsey DE, Hartl DL. 2013. Malaria life cycle intensifies both natural selection and random genetic drift. *Proc Natl Acad Sci U S A* 110:20129–20134. <https://doi.org/10.1073/pnas.1319857110>.
- Mharakurwa S, Kumwenda T, Mkulama MAP, Musapa M, Chishimba S, Shiff CJ, Sullivan DJ, Thuma PE, Liu K, Agre P. 2011. Malaria antifolate resistance with contrasting *Plasmodium falciparum* dihydrofolate reductase (DHFR) polymorphisms in humans and *Anopheles* mosquitoes. *Proc Natl Acad Sci U S A* 108:18796–18801. <https://doi.org/10.1073/pnas.1116162108>.
- Mharakurwa S, Sialumano M, Liu K, Scott A, Thuma P. 2013. Selection for chloroquine-sensitive *Plasmodium falciparum* by wild *Anopheles arabiensis* in Southern Zambia. *Malar J* 12:453. <https://doi.org/10.1186/1475-2875-12-453>.
- Conrad MD, Mota D, Musiime A, Kilama M, Rek J, Kamya M, Dorsey G, Rosenthal PJ. 2017. Comparative prevalence of *Plasmodium falciparum* resistance-associated genetic polymorphisms in parasites infecting humans and mosquitoes in Uganda. *Am J Trop Med Hyg* 97:1576–1580. <https://doi.org/10.4269/ajtmh.17-0351>.
- Mendes C, Salgueiro P, Gonzalez V, Berzosa P, Benito A, do Rosário VE, de Sousa B, Cano J, Arez AP. 2013. Genetic diversity and signatures of selection of drug resistance in *Plasmodium* populations from both human and mosquito hosts in continental Equatorial Guinea. *Malar J* 12:114. <https://doi.org/10.1186/1475-2875-12-114>.
- Morlais I, Nsango SE, Toussile W, Abate L, Annan Z, Tchioffo MT, Cohuet A, Awono-Ambene PH, Fontenille D, Rousset F, Berry A. 2015. *Plasmodium falciparum* mating patterns and mosquito infectivity of natural isolates of gametocytes. *PLoS One* 10:e0123777. <https://doi.org/10.1371/journal.pone.0123777>.
- Berry A, Menard S, Nsango SE, Abate L, Concorde D, Tsapi MT, Iriart X, Awono-Ambéné PH, Roche B, Morlais I. 2021. The rare, the best: spread of antimalarial-resistant *Plasmodium falciparum* parasites by *Anopheles* mosquito vectors. *Microbiol Spectr* 9:e0085221. <https://doi.org/10.1128/Spectrum.00852-21>.
- Malaguarnera L, Musumeci S. 2002. The immune response to *Plasmodium falciparum* malaria. *Lancet Infect Dis* 2:472–478. [https://doi.org/10.1016/S1473-3099\(02\)00344-4](https://doi.org/10.1016/S1473-3099(02)00344-4).
- Chao A, Kubota Y, Zelený D, Chiu C-H, Li C-F, Kusumoto B, Yasuhara M, Thorn S, Wei C-L, Costello MJ, Colwell RK. 2020. Quantifying sample completeness and comparing diversities among assemblages. *Ecological Res* 35:292–314. <https://doi.org/10.1111/1440-1703.12102>.
- Graumans W, Jacobs E, Bousema T, Sinnis P. 2020. When is a *Plasmodium*-infected mosquito an infectious mosquito? *Trends Parasitol* 36:705–716. <https://doi.org/10.1016/j.pt.2020.05.011>.
- Mu J, Awadalla P, Duan J, McGee KM, Joy DA, McVean GAT, Su X. 2005. Recombination hotspots and population structure in *Plasmodium falciparum*. *PLoS Biol* 3:e335. <https://doi.org/10.1371/journal.pbio.0030335>.
- Koella JC, Sørensen FL, Anderson RA. 1998. The malaria parasite, *Plasmodium falciparum*, increases the frequency of multiple feeding of its mosquito vector, *Anopheles gambiae*. *Proc Biol Sci* 265:763–768. <https://doi.org/10.1098/rspb.1998.0358>.
- Scott TW, Takken W. 2012. Feeding strategies of anthropophilic mosquitoes result in increased risk of pathogen transmission. *Trends Parasitol* 28:114–121. <https://doi.org/10.1016/j.pt.2012.01.001>.
- Pollitt LC, Bram JT, Blanford S, Jones MJ, Read AF. 2015. Existing infection facilitates establishment and density of malaria parasites in their mosquito vector. *PLoS Pathog* 11:e1005003. <https://doi.org/10.1371/journal.ppat.1005003>.
- Molina-Cruz A, Canepa GE, Kamath N, Pavlovic NV, Mu J, Ramphul UN, Ramirez JL, Barillas-Mury C. 2015. *Plasmodium* evasion of mosquito immunity and global malaria transmission: the lock-and-key theory. *Proc Natl Acad Sci U S A* 112:15178–15183. <https://doi.org/10.1073/pnas.1520426112>.
- Cohuet A, Harris C, Robert V, Fontenille D. 2010. Evolutionary forces on *Anopheles*: what makes a malaria vector? *Trends Parasitol* 26:130–136. <https://doi.org/10.1016/j.pt.2009.12.001>.
- Grubaugh ND, Weger-Lucarelli J, Murrieta RA, Fauver JR, Garcia-Luna SM, Prasad AN, Black WC IV, Ebel GD. 2016. Genetic drift during systemic arbovirus infection of mosquito vectors leads to decreased relative fitness during host switching. *Cell Host Microbe* 19:481–492. <https://doi.org/10.1016/j.chom.2016.03.002>.
- Greenhouse B, Ho B, Hubbard A, Njama-Meya D, Narum DL, Lanar DE, Dutta S, Rosenthal PJ, Dorsey G, John CC. 2011. Antibodies to *Plasmodium falciparum* antigens predict a higher risk of malaria but protection from symptoms once parasitemic. *J Infect Dis* 204:19–26. <https://doi.org/10.1093/infdis/jir223>.
- Demas A, Oberstaller J, DeBarry J, Lucchi NW, Srinivasamoorthy G, Sumari D, Kabanyanyi AM, Villegas L, Escalante AA, Kachur SP, Barnwell JW, Peterson DS, Udhayakumar V, Kissinger JC. 2011. Applied genomics: data mining reveals species-specific malaria diagnostic targets more sensitive than 18S rRNA. *J Clin Microbiol* 49:2411–2418. <https://doi.org/10.1128/JCM.02603-10>.
- Taylor SM, Sumner KM, Freedman B, Mangeni JN, Obala AA, Prudhomme O'Meara W. 2019. Direct estimation of sensitivity of *Plasmodium falciparum* rapid diagnostic test for active case detection in a high-transmission community setting. *Am J Trop Med Hyg* 101:1416–1423. <https://doi.org/10.4269/ajtmh.19-0558>.
- Chang H-H, Park DJ, Galinsky KJ, Schaffner SF, Ndiaye D, Ndir O, Mboup S, Wiegand RC, Volkman SK, Sabeti PC, Wirth DF, Neafsey DE, Hartl DL. 2012. Genomic sequencing of *Plasmodium falciparum* malaria parasites from Senegal reveals the demographic history of the population. *Mol Biol Evol* 29:3427–3439. <https://doi.org/10.1093/molbev/mss161>.
- Early AM, Daniels RF, Farrell TM, Grimmsby J, Volkman SK, Wirth DF, MacInnis BL, Neafsey DE. 2019. Detection of low-density *Plasmodium falciparum* infections using amplicon deep sequencing. *Malar J* 18:219. <https://doi.org/10.1186/s12936-019-2856-1>.
- Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods* 13:581–583. <https://doi.org/10.1038/nmeth.3869>.
- Jaccard P. 1912. The distribution of the flora in the Alpine zone. *New Phytol* 11:37–50. <https://doi.org/10.1111/j.1469-8137.1912.tb05611.x>.
- Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35:526–528. <https://doi.org/10.1093/bioinformatics/bty633>.
- Paradis E. 2018. Analysis of haplotype networks: the randomized minimum spanning tree method. *Methods Ecol Evol* 9:1308–1317. <https://doi.org/10.1111/2041-210X.12969>.
- Paradis E. 2010. pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics* 26:419–420. <https://doi.org/10.1093/bioinformatics/btp696>.
- Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. 2017. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* 8:28–36. <https://doi.org/10.1111/2041-210X.12628>.
- Chao A, Henderson PA, Chiu C-H, Moyes F, Hu K-H, Dornelas M, Magurran AE. 2021. Measuring temporal change in alpha diversity: a framework integrating taxonomic, phylogenetic and functional diversity and the iNEXT.3D standardization. *Methods Ecol Evol* 12:1926–1940. <https://doi.org/10.1111/2041-210X.13682>.
- Hill MO. 1973. Diversity and evenness: a unifying notation and its consequences. *Ecology* 54:427–432. <https://doi.org/10.2307/1934352>.
- Chao A, Ricotta C. 2019. Quantifying evenness and linking it to diversity, beta diversity, and similarity. *Ecology* 100:e02852. <https://doi.org/10.1002/ecy.2852>.
- Rigby RA, Stasinopoulos DM. 2005. Generalized additive models for location, scale and shape. *Appl Statist* 54:507–554. <https://doi.org/10.1111/j.1467-9876.2005.00510.x>.
- R Core Team. 2021. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

35. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Golemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen T, Miller E, Bache S, Müller K, Ooms J, Robinson D, Seidel D, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H. 2019. Welcome to the Tidyverse. *Joss* 4:1686. <https://doi.org/10.21105/joss.01686>.
36. Wilke CO. 2019. cowplot: streamlined plot theme and plot annotations for "ggplot2" (1.0.0). <https://CRAN.R-project.org/package=cowplot>. Retrieved 15 April 2020.
37. Wickham H, Seidel D, RStudio. 2020. scales: scale functions for visualization (1.1.1). <https://CRAN.R-project.org/package=scales>. Retrieved 1 April 2022.
38. Wilke CO. 2020. ggtext: improved text rendering support for "ggplot2" (0.1.1). <https://CRAN.R-project.org/package=ggtext>. Retrieved 1 April 2022.
39. Duke Health. 2020. *Plasmodium falciparum* raw sequence reads from humans and mosquitoes in Western Kenya. <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA646940/>. Accessed 25 August 2022.
40. Lapp Z. 2022. Data and code for "*Plasmodium falciparum* genetic diversity in coincident human and mosquito hosts." <https://github.com/duke-malaria-collaboratory/parasite-host-comparison>. Accessed 25 August 2022.