bioRxiv preprint doi: https://doi.org/10.1101/044701; this version posted April 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The genetic basis of host preference and indoor resting behavior in the major African malaria vector, *Anopheles arabiensis*

Bradley J Main[1*], Yoosook Lee[1], Heather M Ferguson[2], Anicet Kihonda[3], Nicodem J Govella[3], Katharina S Kreppel[2 3], Travis C Collier[1], Anthony J Cornel[4], Eleazar Eskin[5], Eun Yong Kang[5], Catelyn C Nieman[1], Allison M Weakley[1] and Gregory C Lanzaro[1]

[*]Correspondence: bradmain@gmail.com

[1]Vector Genetics Laboratory, Department of Pathology, Microbiology and Immunology/University of California, Davis, 95616, USA

[2]Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Glasgow, UK

[3]Environmental Health and Ecological Sciences group, Ifakara Health Institute, Ifakara, United Republic of Tanzania.

[4]Department of Entomology and Nematology/University of California, Davis, 95616, USA.

[5]Department of Computer Science, University of California Los Angeles, 90095 Los Angeles, USA.

Keywords: *Anopheles arabiensis*, host-preference, resting behavior, association study, population genetics, chromosome inversion.

Abstract

Malaria transmission is dependent on the propensity of *Anopheles* mosquitoes to bite humans (anthropophily) instead of other dead end hosts. Recent increases in the usage of Long Lasting Insecticide Treated Nets (LLINs) in Africa have been associated with reductions in highly anthropophilic vectors such as *Anopheles gambiae s.s.*, leaving more zoophilic species such as *Anopheles arabiensis* as the most prominent remaining source of transmission in many settings. *An. arabiensis* appears to be more of a generalist in terms of host preference and resting behavior, which may be due to phenotypic plasticity or segregating allelic variation. To investigate the genetic basis of host preference and resting behavior in *An. arabiensis* we sequenced and analyzed

bioRxiv preprint doi: https://doi.org/10.1101/044701; this version posted April 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

genomes of 48 human- or cattle-fed *An. arabiensis* that were captured resting indoors or outdoors in a village in the Kilombero Valley, Tanzania. A total of 4,820,851 SNPs were identified and used to conduct the first genome-wide estimates of "SNP heritability" for host-choice and resting behavior in this species. A genetic component was detected for host choice (human vs cow fed; permuted $P = 0.002$), but the genetic component for resting behavior was negligible (indoors versus outside: permuted $P = 0.465$). A principal component analysis (PCA) segregated individuals into three groups which are characterized by the 2Rb and/or 3Ra paracentromeric chromosome inversions. There was a non-random distribution of cattle-fed mosquitoes between the PCA clusters, suggesting that alleles linked to the 2Rb and/or 3Ra inversions may influence host preference. Using a novel inversion genotyping assay developed to test for an association between inversion state and host choice, we detected a significant enrichment of the standard (non-inverted) 3Ra arrangement among cattle-fed mosquitoes (N=129) compared to human-fed (N=134; χ^2, p=0.046) and versus all non-cattle-fed individuals (N=234; χ^2, p=0.007). Thus, tracking the frequency of the 3Ra in *An. arabiensis* populations is important, especially in relation to the emergence of behavioral avoidance (e.g. shifting toward zoophily) in some populations so countermeasures can be implemented. A better understanding of the genetic basis for host preference in *An. arabiensis* may also improve vector control if cattle-biting mosquitoes can be genetically engineered and driven in a population, having an effect similar in concept to zooprophylaxis.

Author summary

Increased insecticide treated bed net usage is associated with a shift in relative abundance from the highly anthropophilic and indoor-seeking vector species *Anopheles gambiae s.s.* to the more generalist species, Anopheles arabiensis. A genetic basis for these important phenotypes has not been determined, but recent work has linked variation in an odorant receptor to host-preference in another mosquito genus, *Aedes aegypti*. To begin addressing the genetic basis of these phenotypes, we performed a large-scale bloodmeal analysis at multiple villages in the Kilombero Valley, Tanzania. To limit the identification of genetic variation associated with different geographic

bioRxiv preprint doi: https://doi.org/10.1101/044701; this version posted April 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

locations, we focused our genetic analysis on the village of Lupiro. We sequenced a total of 48 genomes, including females that had fed on either human or cattle and that were resting indoors or outdoors. Our genomic analysis and subsequent follow-up with a novel molecular karyotyping assay revealed a relationship between individuals with the standard arrangement of the 3Ra inversion and preference for cattle. This is evidence supporting a substantial genetic basis for host preference in *An. arabiensis*. Further study is needed to examine allelic variation at candidate genes between the standard and inverted 3Ra.

Introduction

Blood-feeding insects impose a substantial burden on human and animal health through their role as disease vectors. In particular, mosquito species that feed on human blood pose an enormous public health threat by transmitting numerous pathogens such as dengue virus and malaria, which together kill more than one million people per year (WHO 2012; Murray and Lopez 1997). Human exposure to pathogens transmitted by mosquito vectors is determined by vector behaviors such as: (1) preferring to feeding on humans (anthropophily) and (2) residing in close proximity to humans, as reflected by biting and resting inside houses (endophily) (Takken and Verhulst 2013). These traits are known to vary within and between *Anopheles* mosquito species that transmit malaria (Takken and Verhulst 2013). It has been known since the earliest days of malaria transmission model development (Macdonald 1957), that the degree of anthropophily in vector populations is strongly associated with the “basic reproduction rate” of human malaria. At the same time, the extent to which vectors feed and rest inside houses is a critical determinant of the effectiveness of current frontline control strategies including Long-Lasting Insecticide Treated Nets (LLINs) and Indoor Residual Spraying (IRS) which selectively kill mosquitoes that bite and rest indoors (WHO 2012).

Vector species that are more generalist with respect to host feeding behavior, like *An. arabiensis,* are thought to be better able to persist in areas of high indoor insecticide use. For example, several studies in East Africa have shown dramatic declines in the abundance of *An. gambiae* s.s. relative to *An. arabiensis* in parallel with the use of

bioRxiv preprint doi: https://doi.org/10.1101/044701; this version posted April 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

LLINs (Derua et al. 2012; Gatton et al. 2013; Lyimo and Ferguson 2009; Bugoro et al. 2011; Mwangangi et al. 2013; Bayoh et al. 2010a; Russell et al. 2011a; Lindblade et al. 2006; Zhou et al. 2011; Mutuku et al. 2011). Similar changes in vector species composition in response to LLINs have been reported outside of Africa, including in the Solomon Islands where the highly endophagic, anthropophilic *An. punctulatus* has been nearly eliminated by vector control whereas the more exophilic *An. farauti* remains (Bugoro et al. 2011). Given the importance of mosquito feeding and resting behavior to the effectiveness of disease transmission, there is an urgent need to understand the underlying biological determinants of these behaviors and their impact (short and long term) on the effectiveness of the existing frontline interventions.

While the genetic basis for host preference remains unknown in *Anopheles* mosquitoes, environmental heterogeneity has been shown to have a substantial influence on several important vector behaviors (Ferguson et al. 2010), including host preference (Takken and Verhulst 2013). For example, a recent study in southern Tanzania reported that the proportion of blood meals taken from humans by *An. arabiensis* fell by over 50% when at least one cow is kept at a household (Mayagaya et al. 2015). The resting behavior of mosquito vectors in this study was also highly associated with proximity to livestock; the proportion of *An. arabiensis* resting indoors falling by 50% when cattle were present at the household (Mayagaya et al. 2015). Whilst these studies confirm that the environment influences mosquito vector behavior, far less is known about the influence of mosquito genetics on these behavioral phenotypes. An early study by Gillies (Gillies 1964) was one of the few to experimentally investigate this phenomenon in which it was shown that *An. gambiae* s.l. could be selected to switch host preference to cattle within a few generations. However, this study was conducted before the development of molecular methods to distinguish between sibling species in this complex, thus it remains uncertain whether the shift in host preference was due to selection on allelic variation within *An. arabiensis*, or simply due a reduction in the proportion of *An. gambiae* s.s. relative to *An. arabiensis* throughout successive generations. Other work has associated the *An. arabiensis* 3Ra chromosome inversion with cattle-sheds (Lulu, Hadis, and Makonnen 1998). Understanding the genetic basis for host preference is

bioRxiv preprint doi: https://doi.org/10.1101/044701; this version posted April 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

essential for elucidation of the co-evolutionary forces that stabilize the transmission of vector-borne diseases, and may enable the development of genetic markers that could be used for rapid quantification of the degree of anthropophily in vector populations as required to estimate transmission risk and plan vector control programs (Garrett-Jones 1964).

There is evidence from other mosquito taxa that host feeding behavior has a significant genetic component. For example, a recent study linked allelic variation in the odorant receptor gene *Or4* to human-biting preference in the dengue mosquito vector *Aedes aegypti* (McBride et al. 2014). However, to date, no ortholog for *AaegOr4* has been identified in Anopheline mosquitoes (Bohbot et al. 2007), and no direct functional links between genetic mutations in African malaria vectors and behaviors that influence transmission potential have been identified (Fox et al. 2001; Rinker, Zhou, et al. 2013; Rinker, Pitts, et al. 2013; Takken and Verhulst 2013). As the genera *Aedes* and *Anopheles* diverged before the emergence of the *homo sapiens* (~150 MYA) (Reidenbach et al. 2009), anthropophily likely evolved independently in these species and may involve distinct mechanisms. As mosquito populations evolve and adapt to vector control measures, a better understanding of these important disease transmission-related behaviors is becoming increasingly important. Developing the ability to track and anticipate shifts in biting time (Maxwell et al. 1998), host preference (Takken and Verhulst 2013), and resting behavior (Pates and Curtis 2005) in mosquito populations will be necessary to make long-term progress in mosquito control (Govella, Chaki, and Killeen 2013). Indeed behavioral avoidance may be a significant threat to the long-term goal of malaria elimination (Killeen 2013). Understanding the genetic contribution to these phenotypes is a critical first step toward effective mosquito control in the future.

Due to the role of *An. arabiensis* in maintaining residual malaria transmission across sub-Saharan Africa (Bayoh et al. 2010a; Mwangangi et al. 2013; Russell et al. 2010), we conducted a comprehensive investigation into the genetic basis of host preference and habitat use in this phenotypically variable species. This included the first application

bioRxiv preprint doi: https://doi.org/10.1101/044701; this version posted April 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

of both whole genome sequencing and a population-scale assessment of chromosome inversion frequencies to test for associations between mosquito behavioral phenotypes and genotype. Our aim was to elucidate genetic factors that are associated with these epidemiologically relevant mosquito behaviors, and compare potential candidate genes with other important disease vector species such as *Ae. aegypti,* whose preference for humans has been recently described (McBride et al. 2014). Additionally we hope information gathered here can be of use to future malaria control scenarios by highlighting the potential of *An. arabiensis* to evolve behavioral avoidance strategies that could either decrease transmission (e.g. zoophily) or diminish control effectiveness (e.g. outdoor resting).

Results

Analysis of host preference

We analyzed 1,731 bloodfed *An. arabiensis* females that were captured resting indoors or outdoors from 3 villages in Tanzania: 746 from Lupiro, 299 from Minepa, and 686 from Sagamaganga (see methods; Table S1). Mosquitoes that tested positive for more than one host were rare (~3%). The relative frequencies of blood meals from each host varied by site, but cattle was the most abundant blood source detected in all three sites (Figure 1). We collected a significantly higher proportion of human-fed *An. arabiensis* with outdoor resting traps and indoor aspiration in Lupiro (out=0.20, in=0.16) versus Minepa (out=0.09, in=0.02) and versus Sagamaganga (out=0.01, in=0.02; $P<0.0001$; fisher exact). This trend varied by household as the proportion of human-fed mosquitoes at a given household was inversely correlated with the presence of livestock ($P<0.0001$, coeff= -2.3384; GLMM).

bioRxiv preprint doi: https://doi.org/10.1101/044701; this version posted April 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 1. Relative host choice by site.

This figure describes the bloodmeal analysis from collections made at three field sites: Sagamaganga, Minepa, and Lupiro.

Testing for a Genetic Component Underlying Host Preference and Indoor Resting Behavior

To test for a genetic component to host preference and resting behavior, we sequenced a total of 48 individual *An. arabiensis* genomes (median coverage =18x; Table S2). In terms of host choice, this collection included 25 cattle-fed and 23 human-fed individuals. The resting behavior phenotype was represented by 24 indoor and 24 outdoor individuals. There is no relationship between resting behavior and host choice among these chosen samples (i.e. there was no enrichment of cattle-fed mosquitoes in the outdoor samples; Fisher Exact P=1, N=48). From these genomes we identified a set of 4,820,851 segregating SNPs after a minor allele frequency threshold of 10% was imposed (see methods). Using these data, we estimated the genetic component (or "SNP heritability" (Wray et al. 2013)) for each phenotype (see methods). The sample size of 48 genomes was not sufficient to estimate SNP heritability with confidence. For example, the heritability estimate for human-fed vs. cattle-fed mosquitoes was H^2 =0.94, SE=3.47 and indoor vs. outdoor was H^2 = 0.05, SE = 2.34. Thus, we permuted the phenotypes to simulate the null hypothesis of no connection between the SNP data and the behavior. Then, we compared the estimate of the SNP heritability from the real data with the estimates from each of 10,000 permutations. This test supports the initial heritability results indicating a genetic component for host choice (human vs. cow fed;

bioRxiv preprint doi: https://doi.org/10.1101/044701; this version posted April 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

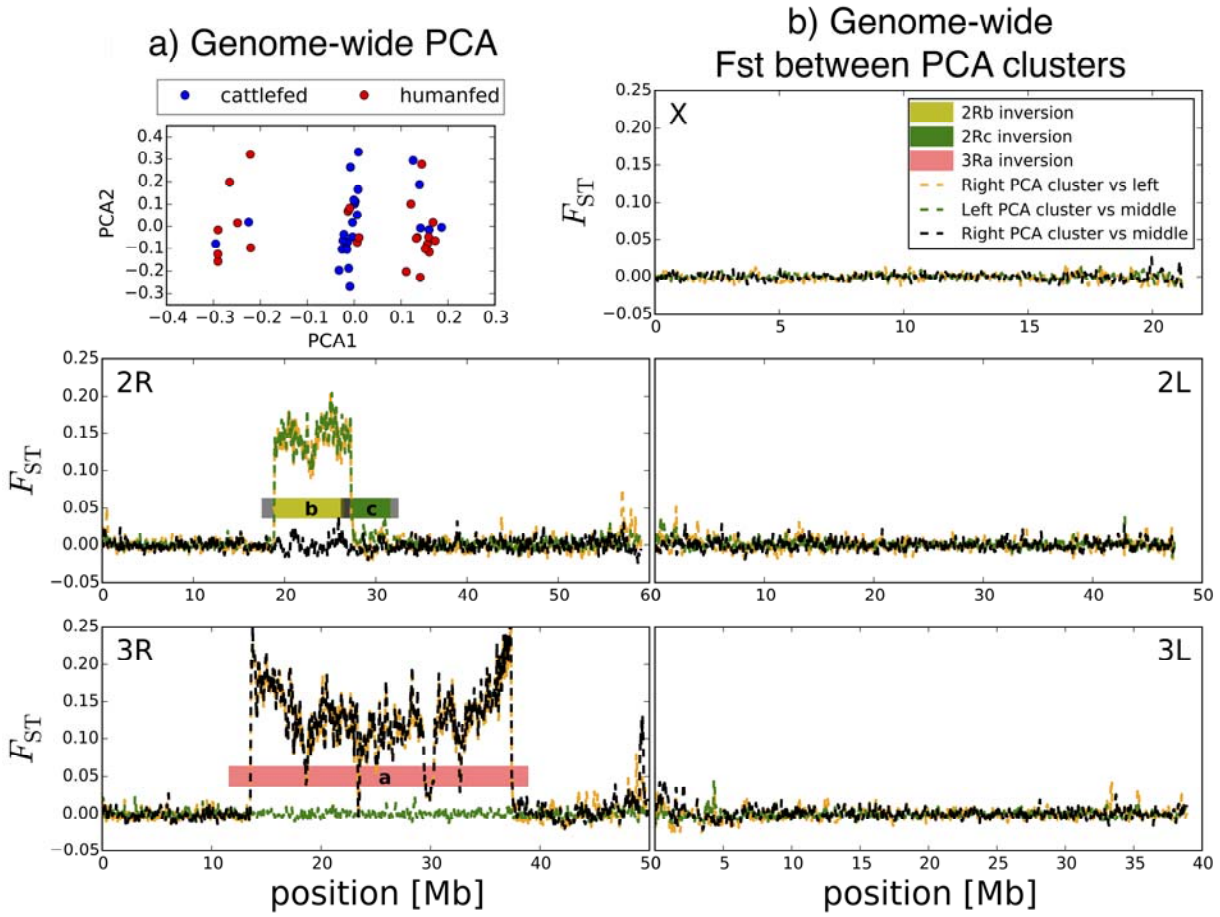
permuted $P = 0.001$), but not for resting behavior (indoor vs. outdoor, permuted $P = 0.470$; see Supporting Information). Due to the lack of evidence for a genetic component for resting behavior, we concluded that this phenotype is unlikely to have a detectable genetic determinant and restricted further analysis to elucidating the observed association between host choice and genotype.

Genetic structure

To test for the existence of genetic structure within our set of 48 sequenced genomes, individuals were partitioned by genetic relatedness using a principle component analysis on all SNPs (PCA; see methods). Using this approach, we observed 3 discrete PCA clusters (Figure 2a). Genome-wide *Fst* in sliding windows between individuals in each PCA cluster (see methods) revealed that the clusters can be explained by distinct combinations of 3Ra and 2Rb chromosome inversion states (Figure 2b). Using a novel inversion genotyping assay (see methods) that was validated on karyotyped samples (Table S4), we determined the genotypes for each of the PCA clusters (2Rb_3Ra): left = bb_a+, middle = bb_++, and right = b+_++. There is an enrichment of cattle-fed mosquitoes among bb_++ individuals ($P < 0.001$; Fisher Exact with Freeman-Halton extension).

bioRxiv preprint doi: https://doi.org/10.1101/044701; this version posted April 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 2. Genetic variation explained by the 2Rb and 3Ra inversions.



a) Genetic structure was assessed using genome-wide SNP data for individual *An. arabiensis* females using a PCA analysis. Three discrete PCA clusters were observed. Red = human-fed and blue = cattle-fed. There is an enrichment of cattle-fed individuals in the middle PCA cluster ($P < 0.001$; Fisher Exact). (b) To reveal differentiated genomic regions underlying the distinct PCA clusters (left, middle, and right) we plotted Fst for each chromosome in 100kb windows with 20kb steps between the PCA clusters. The outside PCA clusters differed at the 2Rb and 3Ra inversions (orange), left versus middle PCA clusters differed at 2Rb only (green), and right versus middle differed at 3Ra only (black).

bioRxiv preprint doi: https://doi.org/10.1101/044701; this version posted April 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Testing for associations between inversion state and host preference

To explore the relationship between the 3Ra and 2Rb inversion state and host preference, we employed a novel inversion genotyping assay (see methods). In total, we genotyped 363 bloodfed females from the village of Lupiro for inversion state, most of which were human-fed (37%) or cattle-fed (36%; Table S5). Multiple blood sources were rare (1%). The 2Rb and 3Ra inversion frequencies were within Hardy-Weinberg (HW) expectations for all samples (P = 0.55 and 0.90, respectively). However, the 3Ra inversion was outside of HW among dog-fed individuals (N=40, P = 0.02; 2). Only four 3Ra homozygotes were observed (N=363); three fed on dog and one fed on human. The frequency of the 3Ra inversion in Lupiro ranged from 7.94% in cattle to 16.67% in pig-fed mosquitoes. The 2Rb inversion ranged from 81.06% in human to 95% in dog-fed specimens (Table S5). We focused on three major comparisons to test for a relationship between inversion state and host preference: 1) cattle-fed versus human-fed, 2) human-fed versus non-human-fed, and 3) cattle-fed versus non-cattle-fed. From these comparisons, we observed a significant deficiency of 3Ra in cattle-fed mosquitoes compared to human (P = 0.047, χ^2; N=263; Table 1b) and a significant deficiency of 3Ra in cattle-fed versus non-cattle-fed (P = 0.007, χ^2, N=363; Table 1b).

Discussion

Anopheles arabiensis is much more of a generalist with respect to resting behavior and host preference, compared to its sibling species *An. gambiae s.s.,* which is highly anthropophilic and endophilic (Takken and Verhulst 2013). Generalism in host preference should evolve when the relative benefit (energetic gain from blood) between hosts is small to moderate. How and why *An. gambiae* ("specialist") and *An. arabiensis* ("generalist") can coexist in sympatry (as seen in Tanzania prior to 2004) has not been fully explored, but may be due to a relatively recent immigration of one species type (Egas, Dieckmann, and Sabelis 2004). Here, we elucidate the genetic basis of host preference and resting behavior in *An. arabiensis* using whole genome sequencing and a custom chromosome inversion genotyping assay. We did not detect a genetic component ("SNP heritability") to resting behavior (endo- versus exo-phily). This may be explained by "behavioral plasticity" in this phenotype (Githeko et al. 1996; Lines, Lyimo,

bioRxiv preprint doi: https://doi.org/10.1101/044701; this version posted April 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

and Curtis 1986). However, a genetic component was detected for host preference based on genome-wide SNP data. Using a novel inversion genotyping assay, we show that the 3Ra inversion (or linked alleles) is involved. The prospect of identifying functional alleles underlying host preference in *An. arabiensis* is particularly exciting because this species has become the dominant malaria vector in many parts of East Africa, where insecticide use is common (Braimah et al. 2005; I. Tirados et al. 2006; Bayoh et al. 2010b; Russell et al. 2011b). As host preference is directly linked to malaria transmission, elucidating the genetic basis of this behavioral phenotype may lead innovative tools for vector control. The inversion genotyping assay described herein may be a valuable monitoring tool (e.g. after GMM release or zooprophylaxis); potentially indicating the relative feeding plasticity of a population based on the frequency of 3Ra.

Associating SNPs with human- and cattle-fed *An. arabiensis*

"SNP heritability" provides an estimate of the correlation between phenotype and genome-wide SNP genotypes from pairs of individuals sampled from a population (Wray et al. 2013). A strength of this metric is its robustness to complex phenotypes that are influenced by many small-effect mutations, which may be the case for host preference in *An. arabiensis*. In this study, we collected mosquitoes that were blood-fed and resting indoors or outdoors to assess the genetic basis of host preference and indoor resting behavior. Thus, we infer preference from choice, which is informative, but may not always be accurate due to potential environmental influences (e.g. lack of preferred host nearby or lack of suitable indoor resting area). Despite this potential limitation, the SNP heritability analysis detected a genetic component for host preference. Increased samples sizes (e.g. 100-1000) are needed to get a quantitative estimate of the SNP heritability of host preference and potentially uncover important candidate genes. However, low LD across the genome of this species may limit the outcome of this approach to large-effect mutations (Marsden et al. 2014). Larger sample sizes may also uncover a genetic component to resting behavior, which we did not detect here but cannot rule out. Previously, high inversion polymorphism has been detected in *An. arabiensis* in malarious areas in Nigeria with some inversions showing

bioRxiv preprint doi: https://doi.org/10.1101/044701; this version posted April 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

changes in frequencies linked to different geographical areas (Coluzzi et al. 1979). This could be linked to selection pressures driven by vector control and/or host availability on resting and feeding behavior.

Cattle-feeding linked to the 3Ra inversion

A principal component analysis on genome-wide SNPs resulted in 3 discrete clusters distinguishable by the 3Ra and 2Rb inversion (Figure 1). There was no significant enrichment among the 48 sequenced individuals in any given cluster (chi sq; P=0.23), but the distribution of human- and cattle-fed mosquitoes among the clusters was significantly different (P < 0.01; 2x3 Fisher Exact). This is strong evidence that the inversion/s may contain alleles related to host preference. In *An. arabiensis*, indirect associations have also been made between host preference and inversions, like 3Ra in Ethiopia (Lulu, Hadis, and Makonnen 1998) and Kenya (Mnzava, Mutinga, and Staak 1994). A non-random distribution of the 2Rb inversion has also been reported between human- and cattle-fed mosquitoes (Petrarca and Beier 1992), but we are unaware of *An. arabiensis* studies with paired karyotype and host choice information from each individual mosquito.

To test for an association between host preference and these inversions with a much larger sample size, we developed a novel inversion genotyping assay (see methods). It should be noted that the inversions represent one or more linked alleles among many possible other contributing alleles throughout the genome. This method was validated using 15 karyotyped samples, which matched all of our genotype-based predictions. This is likely because there is built-in redundancy in the design (7 SNPs for 3Ra and 5 SNPs for 2Rb). This allowed us to determine the bloodmeal source (host) and inversion state from each individual in a high-throughput and economical fashion. More testing is needed to assess how well this assay would perform with samples from outside our study sites in Tanzania and elsewhere throughout Africa.

Using this molecular karyotyping method, we observed an enrichment of the standard arrangement of 3Ra among cattle-fed mosquitoes (p=0.007, 2, N=363; Table 1b). The

bioRxiv preprint doi: https://doi.org/10.1101/044701; this version posted April 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

frequency of the 3Ra inversion in dog-fed, goat-fed, and human-fed mosquitoes was substantially higher than cattle-fed mosquitoes (Table 1a). One possible explanation for this pattern is that the standard (non-inverted) 3Ra is the ancestral state and alleles therein facilitate specialization on cattle and these mechanisms are disrupted in the derived 3Ra allele, resulting in more opportunistic feeding behaviors. This hypothesis is also consistent with studies showing that zoophily (cattle-feeding and standard 3Ra) can be selected for (Gillies 1964).

While we provide strong evidence for a role of 3Ra in host preference in *An. arabiensis*, the effect size (i.e. relative contribution to the phenotype) is unclear. Correcting for environmental variation is likely very important when choosing representative samples for each genotype. For example, a human-fed mosquito may be more meaningful if there is an abundance of alternative hosts nearby (e.g. cattle). This was shown by Tirados et al. (Iňaki Tirados et al. 2011), where *An. arabiensis* was found to persistently bite humans despite being surrounded by cattle, negating a zooprophylactic effect of cattle. This highlights the importance of integrating genetic analyses into a wider context. Colony-based host preference assays involving representatives from each 3Ra inversion state in a controlled environment may be the most effective way forward. Previous tests for population structure only revealed differentiation between distant villages (Marsden et al. 2014). Thus, by comparing individual genomes representing host preference phenotypes (and resting behavior) from within the same village (Lupiro), we limited the identification of demographic SNPs in our data set. However, to assess the role of 3Ra more broadly, additional studies involving study sites across the range of *An. arabiensis* are needed.

Variation in host choice between villages

Cattle was the preferred host at each collection site, but we found differences in relative host-choice patterns between villages. For example, the frequency of human-fed mosquitoes was dramatically higher in Lupiro (18.2%, N=746) versus in Minepa (5.3%, N=399) and Sagamaganga (1.2%, N=686; Table S1). This trend may be due to the lower livestock numbers in Lupiro as the frequency of human-fed mosquitoes appears

bioRxiv preprint doi: https://doi.org/10.1101/044701; this version posted April 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

inversely influenced by livestock presence at the household level (Mayagaya et al. 2015). As individuals with the standard 3Ra inversion prefer cattle (Table 1a), the effect of host availability on host choice will likely be stronger in populations where the 3Ra inversion is relatively rare.

Future directions

This study presents important data suggesting a genetic component to host preference in the malaria vector *An. arabiensis*. We show that the 3Ra inversion is involved, at least in part. This association and the introduction of a novel inversion genotyping assay may be a valuable tool for future malaria control strategies involving *An. arabiensis*. For example, tracking the frequency of the 3Ra in *An. arabiensis* may elucidate the emergence of behavioral avoidance (e.g. shifting toward zoophily) so countermeasures can be implemented. A better understanding of the genetic basis for host preference in *An. arabiensis* may also improve vector control if cattle-biting mosquitoes can be genetically engineered and released in the population, having an effect similar in concept to zooprophylaxis (Burkot 1988). To identify functional alleles within the 3Ra inversion, a beneficial next step would be to 1) establish colonies representative of each inversion state from Lupiro, 2) make controlled genetic crosses, and 3) perform choice assays in controlled environmental conditions to select for recombinants for each phenotype.

Materials and Methods

Mosquito collection area

The mosquitoes were collected within 3 villages in the Kilombero River Valley in south-eastern Tanzania: Lupiro (S08°23.2956'; E036°40.6122'), Minepa (S08°16.4974'; E036°40.7640') and Sagamaganga (S08°03.8392'; E036°47.7709'). The Kilombero Valley is dominated by irrigated and rain-fed rice paddies and maize fields bordered by woodland. The annual rainfall is 1200-1800 mm with two rainy seasons. The average daily temperatures range between 20°C and 33°C. Most people in this area are subsistence farmers and/or livestock keepers. Mud or brick houses stand in clusters

bioRxiv preprint doi: https://doi.org/10.1101/044701; this version posted April 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

among a few trees or banana trees. If a household owns livestock, the animals are kept outside a few meters away from the house in sheds (pigs and goats) or within simple cattle fences. Animal sheds with walls and a roof were considered indoor resting areas. Inside houses you will regularly find chickens, cats and sometimes dogs. The mosquitoes will encounter bed nets inside almost all houses in the valley, but no repellents are currently used by people outdoors (Sangoro et al. 2014) and livestock are not treated with insecticide (Rowland et al. 2001). Malaria is endemic in these communities and although prevalence is declining, almost all inhabitants have antibodies for the disease (Kamuyu et al. 2014). The dominant malaria vector species are *An. arabiensis* and *An. funestus* s.l. (Lwetoijera et al. 2014).

Collection methods

In each village, households chosen for collection were within 100-200m of one another. Indoor mosquito collection method was aspiration using a standard battery-powered CDC Back Pack aspirator (BP, Model 1412, John Hock, Florida USA) (Clark, Seda, and Gubler 1994). In these collections, the aspirator was used to collect mosquitoes from the main bedroom by sweeping the nozzle over the interior walls, roof and furniture for a fixed period of ten minutes. BP collections were timed to standardize sampling effort across houses. A resting bucket trap (RBu) was used to trap mosquitoes outdoors. The RBu is made from a standard 20 liter plastic bucket lined with black cotton cloth, and set by placing it on its side with the open end facing a house at a distance of approximately 5m. A small wet cloth is placed inside the bucket to increase humidity. Mosquitoes resting inside RBus were collected at dawn by placing the nozzle of a battery-powered modified CDC backpack aspirator at the open end of the bucket and aspirating for 10-20 seconds.

Ethics

Before collection, meetings were held with community leaders in all villages during which they were informed about the purpose of the study and their participation requested. After their permission had been granted, the study team visited each village

bioRxiv preprint doi: https://doi.org/10.1101/044701; this version posted April 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

and informed consent was obtained from each head of household where trapping was conducted. Research clearance was obtained from the institutional review board of Ifakara Health Institute in Tanzania (IHI/IRB/No: 16-2013) and by the National Institute for Medical Research in Tanzania (NIMR/HQ/R.8c/Vol. II/304).

DNA extraction

For each specimen, the abdomen was separated from the head and thorax; DNA was extracted separately from each using the QIAGEN Biosprint 96 system and QIAGEN blood and tissue kits (QIAGEN, Valencia, CA). *Anopheles arabiensis* samples were distinguished from other *An. gambiae* s.l. species complex members with the Scott polymerase chain reaction assay (Scott, Brogdon, and Collins 1993) and their DNA content was quantified using the Qubit 2.0 Fluorometer (Life technologies, Grand Island, NY).

Bloodmeal analysis

The specific host species that each mosquito had fed upon was determined by a multiplex genotyping assay on DNA extracted from abdomens (Lee et al. 2015). This multiplex genotyping assay can distinguish between blood from cattle, goat, pig, dog, chicken and human.

Analysis of host preference

Statistical analysis was conducted to compare the proportion of human-fed mosquitoes in total between villages and of these the proportion caught resting indoors using the statistical software R (Core-Team RD, 2013). Variation in the proportion of human-fed *An. arabiensis* within the total catch was investigated. Samples found to contain any human blood represented one category and those containing animal blood another. Generalized linear mixed effects models (GLMM, package lme4 in R (Bates et al. 2014)) were used, with human-fed mosquitoes versus animal-fed mosquitoes as a response variable with a binomial distribution and fitting village and livestock presence as fixed effects, and date and house of collection as random effects. To be able to

bioRxiv preprint doi: https://doi.org/10.1101/044701; this version posted April 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

explore the resting preference of *An. arabiensis*, only mosquitoes resting in houses or outdoors but not those caught resting in animal sheds were used for analysis. Here the GLMM were fitted for each village separately with human-fed mosquitoes caught indoors versus outdoors as a response variable with a binomial distribution and livestock as fixed effect and date and house of collection as random effects.

Cytogenetic analysis

To identify 3Ra, 2Rb, and 2Rc chromosomal inversions, polytene chromosomes were extracted from ovarian nurse cells from half gravid indoor resting mosquitoes using the protocol described by Hunt (Hunt 1973). Chromosome banding patterns were examined using a Nikon Eclipse e600 phase contrast microscope. The genotypes of the chromosome inversions were scored for each individual mosquito. Photographic images of chromosomes for the majority of individual mosquitoes used in this study are available on PopI OpenProject page - AaGenome (https://popi.ucdavis.edu/PopulationData/OpenProjects/AaGenome).

Genomic library preparation and sequencing

To avoid identifying SNPs associated with demography or other environmental factors, we chose to sequence mosquitoes collected from only one village, Lupiro. We focused on this village because it had the highest sample sizes for cattle- and human-fed mosquitoes (Figure 1). Genomic DNA was quantified using a Qubit 2.0 fluorometer (Life Technologies). We used 25-50ng of input DNA for library construction. DNA was then cleaned and concentrated with the DNA Clean and Concentrator kit (Zymo Research Corporation). Library preparations were made with the Nextera DNA Sample Preparation Kit (Illumina), using TruSeq dual indexing barcodes (Illumina). Libraries were size-selected with Agencourt AMPure XP beads (Beckman Coulter). We assessed the insert size distribution of the final libraries using a QIAxcel instrument (Qiagen, Valencia, CA) or Bioanalyzer 2100 (Agilent), and the final library concentration was measured with a Qubit 2.0 fluorometer (Life Technologies). Individually barcoded libraries were sequenced with the Illumina HiSeq2500 platform with paired-end 100

bioRxiv preprint doi: https://doi.org/10.1101/044701; this version posted April 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

base pair reads, at the QB3 Vincent J Coates Genomics Sequencing Laboratory at UC Berkeley. See Table S1 for raw sequence output per sample.

Genome sequence mapping and SNP identification

We assessed the quality of our genome sequencing reads using the FastQC software (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Adaptor sequences and poor quality sequence were trimmed from the raw Illumina Fastq reads using the Trimmomatic software, version 0.30 (Bolger, Lohse, and Usadel 2014), with default options. Reads were aligned to *An. arabiensis* reference genome version AaraCHR (generously provided by the Sharakhov laboratory) using BWA-mem (Li 2013). We used the MarkDuplicates module from Picard tools to remove PCR duplicates and the Genome Analysis Tool Kit (GATK) v1.7 to realign reads around indels (McKenna et al. 2010). The resulting sorted BAM (Binary sequence Alignment/Map) files containing sequences for each read and its mapping position were then used to make a VCF (Variant Call Format) file using samtools (v1.1-12) 'mpileup' and bcftools (v1.1-36) multiallelic-caller. We removed indels using VCFtools (v0.1.13; "--remove-indels") and filtered for variable sites using a minor allele frequency threshold of 0.10 ("--maf 0.1") and a major allele threshold of 0.9 ("--max-maf 0.9").

Estimating SNP heritability of each phenotype

Host preference and resting behavior phenotypes may be influenced by many small-effect mutations across the genome. SNP heritability is the correlation between the genome-wide genotypic variation and phenotypic variance (V(G) / V(p)). To estimate SNP heritability, the VCF file containing genome-wide SNP data for all samples was converted to PLINK with VCFtools (command "vcftools --plink") and then binary ped files (GCTA option: "--make-bed") for analysis with the Genome-Wide Complex Trait Analysis software (GCTA; (Yang et al. 2011). To calculate "SNP heritability" with GCTA, we first generated a genetic relationship matrix. Then we calculated SNP heritability for host preference (estimated human-fed prevalence = 20%) and resting behavior (estimated indoor prevalence = 43%). To estimate the permuted p-value, we used a

bioRxiv preprint doi: https://doi.org/10.1101/044701; this version posted April 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

custom python script to randomly permute the phenotype key for 10000 iterations (see supporting information). The permuted p-value was estimated from the proportion of heritability estimates from the randomly permuted phenotype key that were greater than the heritability estimate from the real data.

SNP genotyping of inversion state

We used GCTA (Yang et al. 2011) to perform a principle component analysis (PCA) on all whole genome sequenced individuals from Lupiro. This partitioned the individuals into at least three clusters. Genomic differentiation among the three clusters was concentrated in regions corresponding to 2Rb and 3Ra inversions (Figure 2). We identified candidate diagnostic SNPs between the three clusters using F_{ST} values. We selected 7 diagnostic SNPs for 3Ra that span over 20Mbp, and 5 diagnostic SNPs for 2Rb spanning 6Mbp (Table S3-4). A multiplex SNP genotyping assay was designed for an iPLEX assay platform using Sequenom Typer AssayDesigner program (Sequenom). See supplemental materials for detailed primer information. The Veterinary Genetics Laboratory at UC Davis performed Genotyping using the Sequenom iPLEX.

Data accessibility

The genetic information and meta data associated with this study are available on dryad and on the open source online vector database PopI: AaGenome (https://popi.ucdavis.edu/PopulationData/OpenProjects/AaGenome/).

Competing interests

The authors declare that they have no competing interests.

Author's contributions

BJM conducted the experiment, data analysis and wrote manuscript. YL and GCL conceived the experiment, conducted field collections, and wrote the manuscript. HF

bioRxiv preprint doi: https://doi.org/10.1101/044701; this version posted April 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

conceived the overall study and helped with the manuscript. KSK coordinated and conducted field collections, analyzed data and contributed to the manuscript. TCC conducted data analysis. AJC conducted field collections and cytogenetic analysis. NJG contributed to the manuscript.

Acknowledgements

We thank Catelyn Neiman and Allison M. Weakley of the Vector Genetics Laboratory for performing DNA extraction and genomic library preparation, and Julia Malvick at the Veterinary Genetics Laboratory for her assistance in iPLEX SNP genotyping. We also thank the inhabitants of Lupiro, Minepa and Sagamaganga for their collaboration during field sampling. Financial support was provided by the National Institutes of Health grants R01AI085175-03 and T32AI074550. This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 Instrumentation Grants S10RR029668 and S10RR027303.

bioRxiv preprint doi: https://doi.org/10.1101/044701; this version posted April 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Tables

Table 1a: Host-specific 3Ra Inversion frequencies.

Host	++	a+	aa	N	a	+	freq a
human	99	32	1	132	34	230	12.88%
cattle	106	20	0	126	20	232	7.9%
pig	38	19	0	57	19	95	16.67%
dog	30	7	3	40	13	67	16.25%
goat	2	1	0	3	1	5	
cattle+goat	2	0	0	2	0	4	
human+cattle	1	0	0	1	0	2	
dog+human	0	1	0	1	1	1	
dog+pig	1	0	0	1	0	2	
				human	35	233	13.06%
				non-human	53	405	11.57%
				cattle	20	238	7.75%
				non-cattle	68	400	14.53%

Mosquitoes were collected from the village of Lupiro. Inversion frequencies (freq a) were not calculated for host categories with low sample sizes. The sum of human- and cattle-fed mosquitoes (bottom four categories) included pure (e.g. human) and mixed host (e.g. dog+human) samples.

Table 1b: 3Ra Inversion frequency differences by host choice.

Host 1	vs	Host 2	p
Human	vs	Cattle	**0.047**
Human	vs	Non-Human	0.553
Cattle	vs	Non-Cattle	**0.007**

P-values were calculated using a chi-square test on the 3Ra count data.

bioRxiv preprint doi: https://doi.org/10.1101/044701; this version posted April 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table 2a: Host-specific 2Rb Inversion frequencies.

Host	++	b+	bb	N	b	+	freq b
human	4	42	86	132	214	50	81.06%
cattle	4	33	89	126	211	41	83.73%
pig	1	18	38	57	94	20	82.46%
dog	0	4	36	40	76	4	95.00%
goat	0	1	2	3	5	1	
cattle+goat	0	1	1	2	3	1	
human+cattle	0	0	1	1	2	0	
human+dog	0	1	0	1	1	1	
dog+pig	0	0	1	1	2	0	
				human	217	51	85.37%
				non-human	391	67	85.37%
				cattle	216	42	83.72%
				non-cattle	392	76	83.76%

Mosquitoes were collected from the village of Lupiro. Inversion frequencies (freq b) were not calculated for host categories with low sample sizes. The sum of human- and cattle-fed mosquitoes (bottom four categories) included pure (e.g. human) and mixed host (e.g. dog+human) samples.

Table 2b: 2Rb Inversion frequency differences by host choice.

Host 1	vs	Host 2	p
Human	vs	Cattle	0.4255
Human	vs	Non-Human	0.1442
Cattle	vs	Non-Cattle	1.0

P-values were calculated using a chi-square test on the 2Rb count data.

References

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2014. “Fitting Linear Mixed-Effects Models Using lme4.” *arXiv [stat.CO]*. arXiv. http://arxiv.org/abs/1406.5823.

bioRxiv preprint doi: https://doi.org/10.1101/044701; this version posted April 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Bayoh, M. Nabie, Derrick K. Mathias, Maurice R. Odiere, Francis M. Mutuku, Luna Kamau, John E. Gimnig, John M. Vulule, William A. Hawley, Mary J. Hamel, and Edward D. Walker. 2010a. "Anopheles Gambiae: Historical Population Decline Associated with Regional Distribution of Insecticide-Treated Bed Nets in Western Nyanza Province, Kenya." *Malaria Journal* 9 (1). : 62.

Bayoh MN, Mathias DK, Odiere MR, Mutuku FM, Kamau L, Gimnig JE, Vulule JM, Hawley WA, Hamel MJ, Walker ED. 2010b. "Anopheles Gambiae: Historical Population Decline Associated with Regional Distribution of Insecticide-Treated Bed Nets in Western Nyanza Province, Kenya." *Malaria Journal* 9 (February): 62.

Bohbot, J., R. J. Pitts, H. W. Kwon, and M. Rützler. 2007. "Molecular Characterization of the Aedes Aegypti Odorant Receptor Gene Family - Bohbot - 2007 - Insect Molecular Biology - Wiley Online Library." *Insect Molecular …*, January. http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2583.2007.00748.x/full.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." *Bioinformatics* 30 (15): 2114–20.

Braimah, Nafisah, Chris Drakeley, Eliningaya Kweka, Frank Mosha, Michelle Helinski, Helen Pates, Caroline Maxwell, Thedy Massawe, Michael G. Kenward, and Chris Curtis. 2005. "Tests of Bednet Traps (Mbita Traps) for Monitoring Mosquito Populations and Time of Biting in Tanzania and Possible Impact of Prolonged Insecticide Treated Net Use." *International Journal of Tropical Insect Science* 25 (03). Cambridge University Press: 208–13.

Bugoro, Hugo, Jeffery Hii, Tanya Russell, Robert Cooper, Benny Chan, Charles Iro'ofa, Charles Butafa, Allen Apairamo, Albino Bobogare, and Cheng-Chen Chen. 2011. "Influence of Environmental Factors on the Abundance of Anopheles Farauti Larvae in Large Brackish Water Streams in Northern Guadalcanal, Solomon Islands." *Malaria Journal* 10 (1): 262.

Burkot, T. R. 1988. "Non-Random Host Selection by Anopheline Mosquitos." *Parasitology Today* 4 (6): 156–62.

Clark, G. G., H. Seda, and D. J. Gubler. 1994. "Use of the 'CDC Backpack Aspirator' for Surveillance of Aedes Aegypti in San Juan, Puerto Rico." *Journal of the American Mosquito Control Association* 10 (1): 119–24.

Coluzzi, M., A. Sabatini, V. Petrarca, and M. A. Di Deco. 1979. "Chromosomal Differentiation and Adaptation to Human Environments in the Anopheles Gambiae Complex." *Transactions of the Royal Society of Tropical Medicine and Hygiene* 73 (5): 483–97.

Derua, Yahya A., Michael Alifrangis, Kenneth M. Hosea, Dan W. Meyrowitsch, Stephen M. Magesa, Erling M. Pedersen, and Paul E. Simonsen. 2012. "Change in Composition of the Anopheles Gambiae Complex and Its Possible Implications for the Transmission of Malaria and Lymphatic Filariasis in North-Eastern Tanzania." *Malaria Journal* 11 (1). : 188.

Egas, Martijn, Ulf Dieckmann, and Maurice W. Sabelis. 2004. "Evolution Restricts the Coexistence of Specialists and Generalists: The Role of Trade-off Structure." *The American Naturalist* 163 (4): 518–31.

Ferguson, Heather M., Anna Dornhaus, Arlyne Beeche, Christian Borgemeister, Michael Gottlieb, Mir S. Mulla, John E. Gimnig, Durland Fish, and Gerry F. Killeen. 2010. "Ecology: A Prerequisite for Malaria Elimination and Eradication." *PLoS Medicine* 7 (8): e1000303.

Fox, A. N., R. J. Pitts, H. M. Robertson, J. R. Carlson, and L. J. Zwiebel. 2001. "Candidate Odorant Receptors from the Malaria Vector Mosquito Anopheles Gambiae and Evidence of down-Regulation in Response to Blood Feeding." *Proceedings of the National Academy of Sciences of the United States of America* 98 (25): 14693–97.

Garrett-Jones, C. 1964. "The human blood index of malaria vectors in relation to epidemiological assessment." *Bulletin of the World Health Organization* 30: 241–61.

Gatton, Michelle L., Nakul Chitnis, Thomas Churcher, Martin J. Donnelly, Azra C. Ghani, H. Charles J. Godfray, Fred Gould, et al. 2013. "The Importance of Mosquito Behavioural

bioRxiv preprint doi: https://doi.org/10.1101/044701; this version posted April 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Adaptations to Malaria Control in Africa." *Evolution; International Journal of Organic Evolution* 67 (4): 1218–30.

Gillies, M. T. 1964. "Selection for host preference in Anopheles gambiae." *Nature* 203 (August). Nature Publishing Group: 852–54.

Githeko, A. K., M. W. Service, C. M. Mbogo, and F. K. Atieli. 1996. "Resting Behaviour, Ecology and Genetics of Malaria Vectors in Large Scale Agricultural Areas of Western Kenya." *Parassitologia* 38 (3): 481–89.

Govella, Nicodem J., Prosper P. Chaki, and Gerry F. Killeen. 2013. "Entomological Surveillance of Behavioural Resilience and Resistance in Residual Malaria Vector Populations." *Malaria Journal* 12 (April): 124.

Hunt, R. H. 1973. "A Cytological Technique for the Study of Anopheles Gambiae Complex." *Parassitologia* 15 (1): 137–39.

Kamuyu, Gathoni, Christian Bottomley, James Mageto, Brett Lowe, Patricia P. Wilkins, John C. Noh, Thomas B. Nutman, et al. 2014. "Exposure to Multiple Parasites Is Associated with the Prevalence of Active Convulsive Epilepsy in Sub-Saharan Africa." *PLoS Neglected Tropical Diseases* 8 (5): e2908.

Killeen, Gerry F. 2013. "A Second Chance to Tackle African Malaria Vector Mosquitoes That Avoid Houses and Don't Take Drugs." *The American Journal of Tropical Medicine and Hygiene* 88 (5): 809–16.

Lee, Yoosook, Allison M. Weakley, Catelyn C. Nieman, Julia Malvick, and Gregory C. Lanzaro. 2015. "A Multi-Detection Assay for Malaria Transmitting Mosquitoes." *Journal of Visualized Experiments: JoVE*, no. 96 (February): e52385.

Li, Heng. 2013. "Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM." *arXiv.org*, March. Cornell University Library. http://arxiv.org/abs/1303.3997.

Lindblade, K. A., J. E. Gimnig, L. Kamau, W. A. Hawley, F. Odhiambo, G. Olang, F. O. Ter Kuile, J. M. Vulule, and L. Slutsker. 2006. "Impact of Sustained Use of Insecticide-Treated Bednets on Malaria Vector Species Distribution and Culicine Mosquitoes." *Journal of Medical Entomology* 43 (2): 428–32.

Lines, J. D., E. O. Lyimo, and C. F. Curtis. 1986. "Mixing of Indoor- and Outdoor-Resting Adults of *Anopheles Gambiae* Giles S.l. and *A. Funestus* Giles (Diptera: Culicidae) in Coastal Tanzania." *Bulletin of Entomological Research* 76 (01). Cambridge University Press: 171–78.

Lulu, M., M. Hadis, and Y. Makonnen. 1998. "Chromosomal Inversion Polymorphisms of Anopheles Arabiensis from Some Localities in Ethiopia in Relation to Host Feeding Choice." *… JOURNAL OF HEALTH …*, January. http://www.ethiopianreview.com/1995/Mesfin%20Lulu_Mamuye_Hadis_Yared_Makonnen_Tefera_Asfaw_1995.pdf.

Lwetoijera, Dickson W., Caroline Harris, Samson S. Kiware, Stefan Dongus, Gregor J. Devine, Philip J. McCall, and Silas Majambere. 2014. "Increasing Role of Anopheles Funestus and Anopheles Arabiensis in Malaria Transmission in the Kilombero Valley, Tanzania." *Malaria Journal* 13 (August). biomedcentral.com: 331.

Lyimo, Issa N., and Heather M. Ferguson. 2009. "Ecological and Evolutionary Determinants of Host Species Choice in Mosquito Vectors." *Trends in Parasitology* 25 (4): 189–96.

Macdonald, George. 1957. "The Epidemiology and Control of Malaria." *The Epidemiology and Control of Malaria.* London, Oxford Univ. Pr. http://www.cabdirect.org/abstracts/19581000237.html.

Marsden, Clare Diana, Yoosook Lee, Katharina Kreppel, Allison Weakley, Anthony Cornel, Heather M. Ferguson, Eleazar Eskin, and Gregory C. Lanzaro. 2014. "Diversity, Differentiation, and Linkage Disequilibrium: Prospects for Association Mapping in the Malaria Vector Anopheles Arabiensis." *G3* 4 (1): 121–31.

Maxwell, C. A., J. Wakibara, S. Tho, and C. F. Curtis. 1998. "Malaria-Infective Biting at Different

bioRxiv preprint doi: https://doi.org/10.1101/044701; this version posted April 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Hours of the Night." *Medical and Veterinary Entomology* 12 (3): 325–27.

Mayagaya, Valeriana, Gamba Nkwengulila, Issa Lyimo, Japheti Kihonda, Hassan Mtambala, Hassan Ngonyani, Tanya Russell, and Heather Ferguson. 2015. "The Impact of Livestock on the Abundance, Resting Behaviour and Sporozoite Rate of Malaria Vectors in Southern Tanzania." *Malar. J.* doi:10.1186/s12936-014-0536-8.

McBride, Carolyn S., Felix Baier, Aman B. Omondi, Sarabeth A. Spitzer, Joel Lutomiah, Rosemary Sang, Rickard Ignell, and Leslie B. Vosshall. 2014. "Evolution of Mosquito Preference for Humans Linked to an Odorant Receptor." *Nature* 515 (7526). Nature Publishing Group: 222–27.

McKenna, A., M. Hanna, E. Banks, and A. Sivachenko. 2010. "The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data." *Genome / National Research Council Canada = Genome / Conseil National de Recherches Canada*, January. http://genome.cshlp.org/content/20/9/1297.short.

Mnzava, A. E., M. J. Mutinga, and C. Staak. 1994. "Host Blood Meals and Chromosomal Inversion Polymorphism in Anopheles Arabiensis in the Baringo District of Kenya." *Journal of the American Mosquito Control Association* 10 (4): 507–10.

Murray, C. J., and A. D. Lopez. 1997. "Global Mortality, Disability, and the Contribution of Risk Factors: Global Burden of Disease Study." *The Lancet* 349 (9063): 1436–42.

Mutuku, Francis M., Charles H. King, Peter Mungai, Charles Mbogo, Joseph Mwangangi, Eric M. Muchiri, Edward D. Walker, and Uriel Kitron. 2011. "Impact of Insecticide-Treated Bed Nets on Malaria Transmission Indices on the South Coast of Kenya." *Malaria Journal* 10 (December): 356.

Mwangangi, Joseph, Ephantus Muturi, Simon Muriu, Joseph Nzovu, Janet Midega, and Charles Mbogo. 2013. "The Role of Anopheles Arabiensis and Anopheles Coustani in Indoor and Outdoor Malaria Transmission in Taveta District, Kenya." *Parasites & Vectors* 6 (1): 114.

Pates, Helen, and Christopher Curtis. 2005. "Mosquito Behavior and Vector Control." *Annual Review of Entomology* 50: 53–70.

Petrarca, V., and J. C. Beier. 1992. "Intraspecific Chromosomal Polymorphism in the Anopheles Gambiae Complex as a Factor Affecting Malaria Transmission in the Kisumu Area of Kenya." *The American Journal of Tropical Medicine and Hygiene* 46 (2): 229–37.

Reidenbach, Kyanne, Shelley Cook, Matthew Bertone, Ralph Harbach, Brian Wiegmann, and Nora Besansky. 2009. "Phylogenetic Analysis and Temporal Diversification of Mosquitoes (Diptera: Culicidae) Based on Nuclear Genes and Morphology." *BMC Evolutionary Biology* 9 (1): 298.

Rinker, David C., R. Jason Pitts, Xiaofan Zhou, Eunho Suh, Antonis Rokas, and Laurence J. Zwiebel. 2013. "Blood Meal-Induced Changes to Antennal Transcriptome Profiles Reveal Shifts in Odor Sensitivities in Anopheles Gambiae." *Proceedings of the National Academy of Sciences* 110 (20): 8260–65.

Rinker, David C., Xiaofan Zhou, Ronald Jason Pitts, AGC Consortium, Antonis Rokas, and Laurence J. Zwiebel. 2013. "Antennal Transcriptome Profiles of Anopheline Mosquitoes Reveal Human Host Olfactory Specialization in Anopheles Gambiae." *BMC Genomics* 14 (1): 749.

Rowland, M., N. Durrani, M. Kenward, N. Mohammed, H. Urahman, and S. Hewitt. 2001. "Control of Malaria in Pakistan by Applying Deltamethrin Insecticide to Cattle: A Community-Randomised Trial." *The Lancet* 357 (9271): 1837–41.

Russell, Tanya L., Nicodem J. Govella, Salum Azizi, Christopher J. Drakeley, S. Patrick Kachur, and Gerry F. Killeen. 2011a. "Increased Proportions of Outdoor Feeding among Residual Malaria Vector Populations Following Increased Use of Insecticide-Treated Nets in Rural Tanzania." *Malaria Journal* 10 (1). : 80.

———. 2011b. "Increased Proportions of Outdoor Feeding among Residual Malaria Vector Populations Following Increased Use of Insecticide-Treated Nets in Rural Tanzania."

bioRxiv preprint doi: https://doi.org/10.1101/044701; this version posted April 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Malaria Journal 10 (April): 80.

Russell, Tanya L., Dickson W. Lwetoijera, Deodatus Maliti, Beatrice Chipwaza, Japhet Kihonda, J. Derek Charlwood, Thomas A. Smith, et al. 2010. "Impact of Promoting Longer-Lasting Insecticide Treatment of Bed Nets upon Malaria Transmission in a Rural Tanzanian Setting with Pre-Existing High Coverage of Untreated Nets." *Malaria Journal* 9 (June): 187.

Sangoro, Onyango, Ann H. Kelly, Sarah Mtali, and Sarah J. Moore. 2014. "Feasibility of Repellent Use in a Context of Increasing Outdoor Transmission: A Qualitative Study in Rural Tanzania." *Malaria Journal* 13 (September): 347.

Scott, J. A., W. G. Brogdon, and F. H. Collins. 1993. "Identification of Single Specimens of the Anopheles Gambiae Complex by the Polymerase Chain Reaction." *The American Journal of Tropical Medicine and Hygiene* 49 (4): 520–29.

Takken, Willem, and Niels O. Verhulst. 2013. "Host Preferences of Blood-Feeding Mosquitoes." *Annual Review of Entomology* 58 (1). Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA: 433–53.

Tirados, I., C. Costantini, G. Gibson, and S. J. Torr. 2006. "Blood-Feeding Behaviour of the Malarial Mosquito Anopheles Arabiensis: Implications for Vector Control." *Medical and Veterinary Entomology* 20 (4): 425–37.

Tirados, Iňaki, Gabriella Gibson, Stephen Young, and Stephen J. Torr. 2011. "Are Herders Protected by Their Herds? An Experimental Analysis of Zooprophylaxis against the Malaria Vector Anopheles Arabiensis." *Malaria Journal* 10 (March): 68.

WHO. 2012. "Global Plan for Insecticide Resistance Management in Malaria Vectors." *World Health Organization Press, Geneva, Switzerland*. http://whqlibdoc.who.int/publications/2012/9789241564472_eng.pdf.

Wray, Naomi R., Jian Yang, Ben J. Hayes, Alkes L. Price, Michael E. Goddard, and Peter M. Visscher. 2013. "Pitfalls of Predicting Complex Traits from SNPs." *Nature Reviews. Genetics* 14 (7). Nature Publishing Group: 507–15.

Yang, Jian, S. Hong Lee, Michael E. Goddard, and Peter M. Visscher. 2011. "GCTA: A Tool for Genome-Wide Complex Trait Analysis." *American Journal of Human Genetics* 88 (1): 76–82.

Zhou, Guofa, Yaw A. Afrane, Anne M. Vardo-Zalik, Harrysone Atieli, Daibin Zhong, Peter Wamae, Yousif E. Himeidan, Noboru Minakawa, Andrew K. Githeko, and Guiyun Yan. 2011. "Changing Patterns of Malaria Epidemiology between 2002 and 2010 in Western Kenya: The Fall and Rise of Malaria." *PloS One* 6 (5). Public Library of Science: e20318.