Molecular Characterization and Genotype Distribution of Thioester-containing Protein 1 Gene, A Key Regulator of Malaria Transmission in An. Gambiae Mosquitoes in Western Kenya

Shirley A. Onyango  
Kenyatta University School of Pure and Applied Sciences  
https://orcid.org/0000-0001-8411-8310

Kevin O. Ochwedo  
University of Nairobi School of Biological Sciences

Maxwell G. Machani  
KEMRI: Kenya Medical Research Institute

Collince J. Omondi  
University of Nairobi School of Biological Sciences

Isaiah Debrah  
University of Ghana Medical School: University of Ghana School of Medicine and Dentistry

Julius O. Olumeh  
University of Nairobi School of Biological Sciences

Sidney S. Ogolla  
KEMRI: Kenya Medical Research Institute

Ming-Chieh Lee  
University of California Irvine College of Medicine: University of California Irvine School of Medicine

Guofa Zhou  
University of California Irvine College of Medicine: University of California Irvine School of Medicine

Elizabeth Kokwaro  
Kenyatta University School of Pure and Applied Sciences

Yaw A. Afrane  
University of Ghana Medical School: University of Ghana School of Medicine and Dentistry

James W. Kazura  
Case Western Reserve University School of Medicine

Andrew K. Githeko  
KEMRI: Kenya Medical Research Institute

Daibin Zhong (mailto:dzhong@uci.edu)  
University of California at Irvine  
https://orcid.org/0000-0002-2771-9598

Guiyun Yan  
University of California Irvine School of Medicine
Abstract

Background

Evolutionary pressures lead to the selection of efficient malaria vectors either resistant or susceptible to *Plasmodium* parasites. These forces may elevate the introduction of new species genotypes that adapt to new breeding habitats which could have serious implications on malaria transmission. Thioester-containing protein 1 (TEP1) of *Anopheles gambiae* plays an important role in innate immune defenses against parasites. This study aims to characterize the distribution pattern of TEP1 polymorphisms determining vector competence and subsequently malaria transmission in western Kenya.

Methods

*Anopheles gambiae* adult and larvae were collected using pyrethrum spray catches (PSC) and plastic dippers respectively from Homa Bay, Kakamega, Bungoma, and Kisumu counties between 2017 and 2020. Collected adults and larvae reared to the adult stage were morphologically identified and then identified to sibling species by PCR. TEP1 alleles were determined using restriction fragment length polymorphisms-polymerase chain reaction (RFLP-PCR) and to validate the TEP1 genotyping results, a representative sample of alleles was sequenced.

Results

Two TEP1 alleles (TEP1*S1 and TEP1*R2) and three corresponding genotypes (*S1/S1, *R2/S1, and *R2/R2) were identified. TEP1*S1 and TEP1*R2 with their corresponding genotypes, homozygous *S1/S1 and heterozygous *R2/S1 were widely distributed across all sites with allele frequencies of approximately 80% and 20%, respectively both in *An. gambiae* and *An. arabiensis*. There was no significant difference detected among the population and between the two mosquito species in TEP1 allele frequency and genotype frequency. The overall low levels in population structure ($F_{ST} = 0.019$) across all sites corresponded to an effective migration index ($Nm = 12.571$) and low Nei’s genetic distance values (<0.500) among the subpopulation. The comparative fixation index values revealed minimal genetic differentiation between species and high levels of gene flow among populations.

Conclusion

There is a low genetic diversity and population structure in western Kenya. TEP1* R2 and TEP1*S1 were the most common alleles in both species which may have been maintained through generations in time. However, the TEP1*R2 allele was in low frequencies and may be used to estimate malaria prevalence. Continued surveillance of the distribution of TEP1 is essential for monitoring the population dynamics of local vectors and their implications on malaria transmission and hence designing targeted vector interventions.

Background

*Anopheles gambiae* mosquitoes are competent vectors for malaria in sub-Saharan Africa (1, 2). Ongoing vector control interventions (3, 4), climate change (5-9), and environmental modifications modulate the mosquito abundance promoting the abundance of efficient malaria vectors either refractory or susceptible to...
The characterization and distribution of TEP1 alleles in Anopheles gambiae vectors in malaria-endemic regions in western Kenya.

**Methods**

**Study sites and design**
This study was conducted in four counties in western Kenya namely, Bungoma, Kakamega, Kisumu, and Homa Bay (Fig. 1). Two malaria epidemic-prone highland sites including Kimaeti (00.6029° N, 034.4073° E; altitude 1,430–1545 m above sea level) in Bungoma, and Iguhu (34°45′E, 0°10′N; 1,430–1,580 m above sea level) in Kakamega, and two lowland sites located around Lake Victoria; Kombewa (34°30′E, 0°07′N; 1,150–1,300 m above sea level) in Kisumu and Kendu Bay (34.64190°E-0.38000°S; 1134-1330 m above sea level) in Homa Bay. The climate in western Kenya consists of long and short rainy seasons that malaria transmission peaks between March to May and October to November respectively. Temperature ranges from a minimum of 14-18°C to a maximum of 30-36°C and average rainfall ranges between 1740mm and 1940mm annually. *Plasmodium falciparum* is the most common cause of malaria and is transmitted by *An. arabiensis*, *An. gambiae* and *An. funestus* (28, 29). The key vector control interventions are long-lasting insecticide treated nets (LLINs) and indoor residual spraying (IRS) (30). Indoor residual spray was conducted in Homa Bay County once a year in 2017 and 2018, unlike the other sampling sites.

**Adult Sampling**

*Anopheles* mosquitoes were collected in a cross-sectional study design using Pyrethrum Spray catch (PSC) from 30 randomly selected houses per site between 2017 and 2020 during the dry and rainy seasons. Collections were conducted between 0630 and 1000hrs in the morning and transported to the Sub-Saharan Africa International Center of Excellence for Malaria Research (ICEMR), Homa Bay, Kenya. Samples were stored at -20°C in 1.5 ml Eppendorf tubes containing silica gel and assigned a unique code for further molecular processing.

**Larval Sampling**

Larval sampling was conducted using 350 ml standard dippers and hand pipettes (31). A maximum of 10 dips was taken at each habitat and the presence or absence of larvae was recorded. To avoid collecting siblings from the same pool, larvae were randomly sampled from different breeding habitats. Collected larvae were labeled by habitat type and identified morphologically using the referenced keys (32). Only *Anopheles* larvae were sorted and transported to the ICEMR insectary. The larvae were reared to adults using standardized rearing methods (33). Emerged adults were anesthetized using chloroform and identified using the morphological key in the laboratory as described by Gillies and Coetzee to species (34, 35).

**Molecular Identification Of Mosquito Species**

Genomic DNA was extracted from randomly selected single *An. gambiae* female adult using the Chelex resin (chelex® -100) method following a protocol by Musapa et al (36). Briefly, deionized water was added into single mosquito sample tubes and ground into a uniform suspension. Phosphate buffer saline 1X and 10% Saponin was then added to sample homogenates, mixed gently, and incubated at room temperature for 20 minutes. The suspension was then centrifuged and the supernatant discarded. The pellets were then resuspended in PBS 1X and centrifuged, supernatant discarded, and gently vortexed. The pellets were then suspended in sterile deionized water and 20% Chelex-resin suspension in deionized water. The samples were
incubated at 85°C for 10 minutes, centrifuged at 20,000 x g for a minute, and DNA transferred into prelabelled storage vials. *Anopheles gambiae* was identified to sibling species using polymerase chain reaction (PCR) as described by Scott *et al.* (37).

**Genotyping and DNA sequencing of TEP1 alleles in *Anopheles gambiae* mosquitoes**

Genotyping of TEP1 alleles was performed using polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) method as described by Gildenhard *et al.* (25). Briefly, the initial PCR was conducted using Nest 1 primers - VB3 and VB4- targeting 892 base pairs, followed by a second PCR performed on 5µl of the resulting product from Nest 1 with Nest 2 primers VB1 and VB2 producing a final fragment length of 758 base pairs. Both PCR reaction conditions were set as denaturation at 95°C for 3 minutes, 35 cycles of 94°C for 30 seconds, annealing at 55°C for 30 seconds, extension at 72°C for 30 seconds, and a final step at 72°C for 6 min using DreamTaq Green Master Mix (Thermo Fisher Scientific). PCR products were digested by restriction enzymes *Bam* HI, *Hind* III, or *Bse* NI (New England Biolabs Inc) ([S1 Table](#)) according to the manufacturer’s instructions and analyzed the result with 2.5% agarose gel electrophoresis. The TEP 1 allelic classes were then determined by fragment size of restriction enzyme digestion ([S1 Table](#)). A subset of samples with identified TEP1 alleles was further confirmed through sequencing of respective Nested II amplicons. Sequencing was done using 3700/3730 BigDye® Terminator v3.1 Sequencing Standard kit (ABI PRISM® 3700 DNA Analyzer).

**Statistical analysis**

Descriptive statistical analyses were performed using GraphPad Prism v.8.0.1 Software and SPSS version 25 for Windows. Statistical significance was set at $P \leq 0.05$. TEP1 allele frequencies observed heterozygosity ($H_0$), and expected heterozygosity ($H_e$), the inbreeding coefficients ($F_{IS}$), departure from Hardy-Weinberg expectations were analyzed using GenEAlx version 6.053 software (38). DNA sequences of TEP1 haplotypes were compared with published sequences. Basic Local Alignment Search Tool (BLASTN) was used to retrieve sequences from the National Center for Biotechnology Information (NCBI) database with a high similarity index to each of the haplotype sequences. The retrieved sequences with accession numbers AF291654.1, FN431783.1, FN431782.1, FN431785.1, FN431784.1, and MF098591.1 together with the identified haplotype sequences in this study were aligned.MView web-based tools (39) were used to conduct the alignment of the sequences and to calculate pairwise sequence identity and similarity. Phylogenetic analysis of the representative sequenced and GenBank retrieved TEP1 sequences was performed using MEGA 7.0 software (40).

**Results**

**Species composition of *An. gambiae* s. l across western Kenya**

A total of 627 *An. gambiae* s.l. adults were collected and molecularly identified to sibling species based on species-specific PCR. Overall, the species identified were *An. gambiae* and *An. arabiensis* constituting 49.28% (309/627) and 50.72% (318/627) of the total samples genotyped respectively (Table 1). A significant difference in species abundance (*An. gambiae* versus *An. arabiensis*) was observed in all study sites ($\chi^2 = 592.012$ df=1, $P<0.0001$).
Table 1
Molecular determined species composition in western Kenya.

| Sampling sites | An. arabiensis, n (%) | An. gambiae, n (%) |
|----------------|-----------------------|--------------------|
| Bungoma        | 39 (37.5)             | 65 (62.5)          |
| Kakamega       | 29 (34.1)             | 56 (65.9)          |
| Homa Bay       | 186 (82.7)            | 39 (17.3)          |
| Kisumu         | 55 (25.8)             | 158 (74.2)         |
| Grand Total    | 309 (49.3)            | 318 (50.7)         |

**Tep1 Allele Distribution In Western Kenya**

Overall, two TEP1 alleles (TEP1*S1, and TEP1*R2) were identified across western Kenya with average frequencies of 84.9% and 15.1%, respectively. *Anopheles arabiensis* populations from Homa Bay had the highest TEP1*S1 allele frequency (89%) followed by *An. arabiensis* from Kisumu (86.4%), Kakamega (84.5%), and Bungoma (74.4%) respectively. Among *An. gambiae s.s.* populations from Bungoma displayed the highest TEP1*S1 allele frequency (93.1%) followed by Homa Bay (84.6%), Kakamega (83.9%), and Kisumu (83.5%) respectively (Fig. 2A). The highest TEP1*R2 allele frequency among *An. arabiensis* was observed on vector populations from Bungoma (26%) followed by Kakamega (15.5%), Kisumu (13.6%), and Homa Bay (11%). In *An. gambiae* the TEP1*R2 allele frequency was highest in populations from Kisumu and Kakamega displaying allele frequencies of 16.5% and 16.1% respectively, followed by Homa Bay (15.4%) and Bungoma (7%) respectively. No signicant differences in allele frequency were observed between species (P=0.799).

**Tep1 Genotype Distribution In Western Kenya**

A total of 3 genotypes were identified in western Kenya *An. gambiae* populations. Out of the 3 genotypes, 2 were homozygous (TEP1*S1/S1 and TEP1*R2/R2) and 1 heterozygous (TEP1*R2/S1). Homozygote TEP1*S1/S1 and heterozygote TEP1*R2/S1 genotypes had distinct frequencies (Fig. 2B). TEP1*S1/S1 and TEP1*R2/S1 genotypes were commonly present among species in all sites at an average frequency of 71.75% and 26.61% respectively. TEP1*R2/R2 although rare, was only present in *An. arabiensis* from Bungoma (2.6%), Kakamega (3.4%) and Homa Bay (1.6%) and *An. gambiae s.s.* from Kakamega (3.6%) and Kisumu (1.9%) but in the lowest average frequency of 1.64% (Fig. 2B). The TEP1*S1/S1 genotype was predominant followed by TEP1*R2/S1 but in low varied frequencies among species across all sampling sites. The TEP1*S1/S1 genotype frequency was highest in *An. gambiae* as compared to *An. arabiensis* from all sites except Kakamega populations that displayed higher TEP1*S1/S1 frequencies in *An. arabiensis* (75.9%) than in *An. gambiae* (53.6%) (Fig. 2B). On the contrary, the distribution of TEP1*R2/S1 genotypes was highest in *An. arabiensis* than *An. gambiae* in all sites except populations from Kakamega where higher genotype frequencies (42.9%) were observed in *An. gambiae* than in *An. arabiensis* (20.7%). The observed RFLP results for each TEP1 allele were confirmed by respective sequences upon alignment with reference sequences from the NCBI database. The TEP1*S1 and TEP1*R2 sequences had 100% identity matrix to AF291654.1 and
FN431784.1 respectively. A significant difference in genotype frequency was observed among sites in *An. gambiae* populations (Fisher’s exact test two-sided p-value < 0.001, n=309) whereas no significant difference was observed among sites in *An. arabiensis* population (Fisher’s exact test two-sided p-value = 0.0727, n=318).

**Evolutionary Relationship Based On Tep1 Gene**

The phylogenetic analysis of TEP1 sequences showed that alleles were clustered into susceptible and resistant groups with high bootstrap values, ranging from 72–100%. Out of the sequences retrieved from the gene bank, TEP1*S1 alleles identified in western Kenya have a common lineage with TEP1*S1 (AF291654) from Suakoko, Liberia. TEP1*S1 evolved as a result of a mutation on the mosquito strain G3 with TEP1*S3 (FN431782) which had a close ancestral lineage with strain 4Arr that had the TEP1*S2 (FN431783) allele. TEP1*R2 from western Kenya and TEP1*R1 independently evolved from TEP1*R3 (MF035809) which shared common ancestral lineage with the Susceptible (S) alleles (Fig. 3).

**Heterozygosity And Departure From Hardy Weinberg Equilibrium (Hwe)**

The overall mean observed heterozygosity of TEP domain in *An. gambiae* and *An. arabiensis* across all sites was 0.270±0.035 and a mean ($H_o$) expected heterozygosity of 0.251±0.025. There were slight variations between $H_o$ ranges 0.188-0.462 in *An. arabiensis* and in *An. gambiae* ranges 0.138-0.321. *An. gambiae* populations from Bungoma, Kakamega, and Homa bay, and *An. arabiensis* from Bungoma and Kisumu also showed similar trends of higher observed heterozygosity than the expected with negative $F_{IS}$ values (Table 2). A deviation was observed among *An. gambiae* from Kisumu and *An. arabiensis* from Kakamega and Homa Bay which displayed slightly higher expected heterozygosity than observed signifying the presence of inbreeding among these populations (Table 2).
Table 2
Genetic diversity of *An. gambiae* (GA) and *An. arabiensis* (AR) in western Kenya.

| Population     | N  | Na  | Ho  | He  | F    |
|----------------|----|-----|-----|-----|------|
| AR-Bungoma     | 39 | 2.000 | 0.462 | 0.381 | -0.210 |
| AR-Homa Bay    | 186 | 2.000 | 0.188 | 0.196 | 0.041 |
| AR-Kakamega    | 29 | 2.000 | 0.241 | 0.262 | 0.079 |
| AR-Kisumu      | 55 | 2.000 | 0.236 | 0.236 | -0.004 |
| GA-Bungoma     | 65 | 2.000 | 0.138 | 0.129 | -0.074 |
| GA-Homa Bay    | 39 | 2.000 | 0.308 | 0.260 | -0.182 |
| GA-Kakamega    | 56 | 2.000 | 0.321 | 0.270 | -0.191 |
| GA-Kisumu      | 158 | 2.000 | 0.266 | 0.275 | 0.033 |

N represents the total number of mosquitoes sampled per study site, Na- Number of alleles per site, Ho- Observed heterozygosity, He- Expected heterozygosity, F- Fixation index

The *F*<sub>S</sub><sub>IS</sub> showed a negative and non-significant value in *An. arabiensis* population from Bungoma (-0.210) and Kisumu (-0.004) and *An. gambiae* from Bungoma (-0.074), Kakamega (-0.191), and Homa Bay (-0.182). These results indicate a slight departure from HWE and excess of heterozygotes in these populations. The *F*<sub>S</sub><sub>IS</sub> results for *An. arabiensis* from Homa Bay (0.041) and Kakamega (0.079) and *An. gambiae* from Kisumu (0.033) infer possible inbreeding. None of the analyzed population was at HWE as all the computed values were nonsignificant (P>0.05). The computed HWE values for *An. arabiensis* across the four localities ranged from 0.001 to 0.307 whereas for *An. gambiae* ranged 0.174 to 2.053 which was >1.

Population Structure

The pairwise Wright's fixation index (*F*<sub>ST</sub>) values revealed a low genetic differentiation among *An. arabiensis* and *An. gambiae*. Sub-populations in western Kenya. The *F*<sub>ST</sub> values 0 ≤ 0.05 were interpreted as low differentiation, 0.05 ≥ 0.15 moderate differentiation and 0.15 ≥ 0.25 high levels (41). Zero value represented complete Panmixis between species in the subpopulations. The *F*<sub>ST</sub> values ranged from no subdivision to moderate differentiation (0.000 - 0.036) among *An. arabiensis* from the four study sites (Table 3). A moderate differentiation in *An. arabiensis* was observed between Bungoma and Homa bay subpopulations (*F*<sub>ST</sub> = 0.036). The *F*<sub>ST</sub> values ranged from 0.000 to 0.022 among the *An. gambiae* subpopulations across the four regions. No population differentiation was observed between Kakamega and Homa Bay, Kisumu and Homa Bay, and Kakamega and Kisumu subpopulations (*F*<sub>ST</sub> =0). All pairwise *F*<sub>ST</sub> values for *An. gambiae and An. arabiensis* from all regions across western Kenya demonstrated low population differentiation (0 ≤ 0.05) except *An. arabiensis and An. gambiae* from Bungoma that showed moderate differentiation (0.05 ≥ 0.15). The overall low levels in population structure (*F*<sub>ST</sub> = 0.019) across all sites were supported by the high level of gene flow (Nm= 12.571) and low Nei's genetic distance values (<0.5) among the subpopulation.
Table 3
Pairwise comparison of $F_{ST}$ among *An.gambiae* and *An. arabiensis* populations in western Kenya

| Population | AR-Bungoma | AR-Homa Bay | AR-Kakamega | AR-Kisumu | GA-Bungoma | GA-Homa Bay | GA-Kakamega | GA-Kisumu |
|------------|------------|-------------|-------------|-----------|------------|-------------|-------------|-----------|
| AR-Bungoma | 0          | 0           |             |           |            |             |             |           |
| AR-Homa Bay| 0.036      | 0           | 0           |           |            |             |             |           |
| AR-Kakamega| 0.016      | 0.004       | 0           |           |            |             |             |           |
| AR-Kisumu | 0.023      | 0.002       | 0.001       | 0         |            |             |             |           |
| GA-Bungoma | 0.064      | 0.005       | 0.019       | 0.012     | 0          |             |             |           |
| GA-Homa Bay| 0.016      | 0.004       | 0           | 0.001     | 0.018      | 0           |             |           |
| GA-Kakamega| 0.014      | 0.005       | 0           | 0.001     | 0.021      | 0           | 0           |           |
| GA-Kisumu | 0.013      | 0.006       | 0           | 0.002     | 0.022      | 0           | 0           | 0         |

The AMOVA results revealed that 99% of the observed variations in allele frequency were among individuals within respective populations, and a 1% variation was observed among populations and within individuals (Table 4). These results show that the level of genetic differentiation among populations was very low.

Table 4
Analysis of molecular variance of the TEP1 gene in An. gambiae populations circulating in western Kenya

| Source          | df | SS     | MS    | Est. Var. | %  |
|-----------------|----|--------|-------|-----------|----|
| Among Pops      | 7  | 2.306  | 0.329 | 0.001     | 1% |
| Among Individuals| 619| 71.992 | 0.116 | 0.000     | 0% |
| Within Individuals| 627| 77.000 | 0.123 | 0.123     | 99%|
| Total           | 1253| 151.298| 0.124 |           | 100%|

*df*, degrees of freedom; *SS*, sum of squares; *MS*, mean squares.

**Discussion**

*Plasmodium falciparum* triggers an immune response in *An. gambiae* mosquitoes (42). Following infections with *P. falciparum* in *An. gambiae*, the midgut mounts specific and nonspecific immune responses to minimize epithelial damage (43). A key component of the immune system is the thioester containing protein 1 (TEP1).
that displays allelic variations associated with distinct genotypes in their ability to eliminate *Plasmodium* parasites. Originally, TEP1 variations have been characterized into five allelic classes including *R1, *R2, *R3 *S1, and *S2 (14, 23). This study identified two alleles (TEP1*R2 and, TEP1*S1) in *An. gambiae* s.l from western Kenya. These alleles were characterized from western Kenya regions with different malaria endemcities. A high similarity index was observed among sequenced alleles that were initially identified by RFLP-PCR and sequences retrieved from NCBI. Consistent with previous reports, TEP1*R2 and TEP1*S1 were the most common identified alleles (25, 44) circulating in western Kenya and did not display a defined distribution in sampled regions implying that they are conserved and may represent ancestral alleles maintained over generations in time. Furthermore, why these alleles have been maintained in the local populations and their roles and significance in vector competence is still not clear (14, 22, 23, 45).

Low TEP1*R allele frequencies observed in these malaria-endemic areas in our study sites may be a product of selective pressures in the TEP1 gene resulting in functional variations that select for susceptible mosquitoes to *Plasmodium* infection (12, 44, 46) as well as encourage evolutionary processes in the TEP1 loci (22). Consequently, these regions still record relatively high malaria prevalence. Implemented vector control interventions such as insecticide treated nets, indoor residual spraying, and environmental factors determine the population structure. In Africa, commonly used vector control interventions are ITNs and IRS (47). In 2017, IRS was deployed in Homa Bay to supplement the existing malaria interventions. The pre-spray period constituted 83% *An. funestus* and 16% *An. gambiae* s.l. However, there was a drift in the local species composition with 99% of mosquitoes in the post-spray period being *An. arabiensis* in 2018 in the same site (Kenya Annual Entomological Monitoring Report, October 2017-September 2018). Furthermore, ecological niches contributing towards selection forces acting on genetic variations shape the population structure of the local species populations in time and hence the adaptations of these malaria vectors to available breeding habitats (48). Understanding the underlying molecular mechanisms that determine vector competence is crucial and will thereafter contribute towards developing new vector control interventions and also complement existing control methods.

The overall $F_{ST}$ values for the pairwise comparison for all populations demonstrate very minimal genetic differentiation between species and sites representing the western Kenya highlands (Bungoma and Kakamega) and lowlands (Homa Bay and Kisumu) suggesting the absence of barriers across regions. This observation does not support that ongoing intervention and ecological changes impacted on allele frequency of TEP1 in the region. The low levels of genetic differentiation correspond to an effective migration index ($Nm=12.571$) indicating high levels of gene flow across the sampling sites. Expected heterozygosity values were higher than the observed heterozygosity among *An. arabiensis* from Homa Bay and Kakamega and *An. gambiae* from Kisumu implies the presence of null alleles and maybe as a result of inbreeding and non-random mating of individuals within those populations ($F_{IS}$ 0.041, 0.079, and 0.033 respectively). The insignificant deviations from HWE imply that the TEP1 loci are under strong selection and confirm other forces such as natural mutations (49, 50) and gene flow that may directly be shaping TEP1 alleles in *An. gambiae* s.l mosquitoes in western Kenya.

**Conclusion**
This study reveals minimal genetic differentiation and a low population structure in the highland and lowland regions of western Kenya with different malaria transmission patterns. TEP1*R2 and TEP1*S1 were the most common alleles across all regions indicating that *An. gambiae* and *An. arabiensis* may have had these specific alleles before inhabiting new ecological niches. However, TEP1*R allele frequencies observed especially in the highlands may be used to estimate malaria prevalence as compared to the lowlands. Therefore, continued TEP1 surveillance is essential for monitoring the population dynamics of local malaria vectors especially those that have implications on malaria transmission then contribute towards developing targeted vector interventions.

**Abbreviations**

- **RT-PCR**: Real-time polymerase chain reaction
- **LLIN**: Long-lasting insecticide-treated net
- **IRS**: Indoor residual spraying
- **DBS**: Dried blood spots
- **DNA**: Deoxyribonucleic acid
- **PCR**: Polymerase chain reaction

**Declarations**

**Ethics approval**

This study was approved by the University of California, Irvine Institutional Review Board (UCI IRB) and Maseno University Ethics Review Committee (MUERC protocol No. 00456) and authorized by the Ministry of Health.

**Consent for publication**

Not applicable

**Availability of data and materials**

The data sets generated and analyzed during this study are available from the corresponding authors on reasonable requests.

**Conflict of interest**

All authors declare that they have no conflict of interest.

**Funding**
The study was supported by grants from the National Institutes of Health (U19 AI129326, D43 TW001505, R01 AI050243, and R01 A1123074).

**Authors’ contributions**

Project conceptualization: SAO, KOO and DZ, Project implementation: SAO, SOO, and AKG, Data collection and sample analysis: SAO, MGM, CJO, ID, and JOO, Formal analysis: SAO, KOO, and DZ. Drafting manuscript: SAO. Editing and revising manuscript: KOO, MGM, AKG, SOO, ML, GZ, EK, JWK, YAA, DZ, and GY. Funded project: JWK and GY

**Acknowledgments**

We thank all the community health workers and field assistants and the entire ICEMR staff who participated in the entomological surveys and the village residents for their continued support.

**References**

1. Lanzaro CG, Lee Y. Speciation in *Anopheles gambiae* — The Distribution of Genetic Polymorphism and Patterns of Reproductive Isolation Among Natural Populations. New Insights into Malaria Vectors,. 2013:173-96.

2. Sinka ME, Bangs MJ, Manguin S, Coetzee M, Mbogo CM, Hemingway J, et al. The dominant *Anopheles* vectors of human malaria in Africa, Europe and the Middle East: occurrence data, distribution maps and bionomic précis. Parasites & Vectors. 2010;3:117.

3. Zhou G, Afrane YA, Dixit A, Ming-Chieh L, Wanjala CL, Beilhe LB, Andrew K, et al. Modest additive effects of integrated vector control measures on malaria prevalence and transmission in western Kenya. Malaria journal. 2013;12:256.

4. Gimnig JE, Vulule JM, Lo TQ, Kamau L, Kolczak MS, Phillips-Howard PA, et al. Impact of permethrin-treated bed nets on entomologic indices in an area of intense year-round malaria transmission. The American journal of tropical medicine and hygiene. 2003;68(4 Suppl):16-22.

5. Dao A, Yaro A, Diallo M, Timbiné S, Huestis D, Kassogué Y, et al. Signatures of aestivation and migration in Sahelian malaria mosquito populations. Nature. 2014;516(7531):387–90.

6. Tonning H, Kangalawe R, Yanda PZ. Predicting and mapping malaria under climate change scenarios: the potential redistribution of malaria vectors in Africa. Malaria journal. 2010;9:111.

7. Tanser FC, Sharp B, le Sueur D. Potential effect of climate change on malaria transmission in Africa. Lancet (London, England). 2003;362(9398):1792-8.

8. Lindsay SW, Birley MH. Climate change and malaria transmission. Annals of Tropical Medicine & Parasitology. 1996;90:573-88.

9. Minakawa N, Sonye G, Mogi M, Githeko A, Yan G. The effects of climatic factors on the distribution and abundance of malaria vectors in Kenya. Journal of medical entomology. 2002;39(6):833-41.

10. Eldering M, Morlais I, Gemert G, Vegte-Bolmer M, Graumans W, Siebelink-Stoter R, et al. Variation in susceptibility of African Plasmodium falciparum malaria parasites to TEP1 mediated killing in *Anopheles gambiae* mosquitoes. Scientific reports. 2016;6:20440.
11. Sinden RE, Alavi Y, Raine JD. Mosquito-malaria interactions: A reappraisal of the concepts of susceptibility and refractoriness. Insect Biochem Mol Biol 2004(34):625-9.

12. Le BV, Williams M, Logarajah S, Baxter RH. Molecular basis for genetic resistance of Anopheles gambiae to Plasmodium: structural analysis of TEP1 susceptible and resistant alleles. PLoS Pathog. 2012;8(10):e1002958.

13. Blandin SA, Marois E, Levashina EA. Antimalarial responses in Anopheles gambiae: from a complement-like protein to a complement-like pathway. Cell Host Microbe. 2008;3(6):364-74.

14. Blandin SA, Wang-Sattler R, Lamacchia M, Gagneur J, Lycett G, Ning Y, et al. Dissecting the genetic basis of resistance to malaria parasites in Anopheles gambiae. Science. 2009;326(5949):147-50.

15. Blandin S, Shiao SH, Moita LF, Janse CJ, Waters AP, Kafatos FC, et al. Complement-like protein TEP1 is a determinant of vectorial capacity in the malaria vector Anopheles gambiae. Cell. 2004;116(5):661-70.

16. Volohonsky G, Hopp AK, Saenger M, Soichot J, Scholze H, Boch J, et al. Transgenic Expression of the Anti-parasitic Factor TEP1 in the Malaria Mosquito Anopheles gambiae. PLoS Pathog. 2017;13(1):e1006113.

17. Lefevre T, Ohm J, Dabire KR, Cohuet A, Choisy M, Thomas MB, et al. Transmission traits of malaria parasites within the mosquito: Genetic variation, phenotypic plasticity, and consequences for control. Evol Appl. 2018;11(4):456-69.

18. Levashina EA, Baxter RHG. Complement-Like System in the Mosquito Responses Against Malaria Parasites. Complement Activation in Malaria Immunity and Pathogenesis2018. p. 139-46.

19. Povelones M, Bhagavatula L, Yassine H, Tan LA, Upton LM, Osta MA, et al. The CLIP-Domain Serine Protease Homolog SPCLIP1 Regulates Complement Recruitment to Microbial Surfaces in the Malaria Mosquito Anopheles gambiae. PLOS/ Pathogens. 2013;9(9): e1003623. doi:10.1371/journal.pp.1003623.

20. Vizioli J, Bulet P, Charlet M, Lowenberger C, Blass C, Muller HM, et al. Cloning and analysis of a cecropin gene from the malaria vector mosquito, Anopheles gambiae. Insect molecular biology. 2000;9(1):75-84.

21. Richman AM, Dimopoulos G, Seeley D, Kafatos FC. Plasmodium activates the innate immune response of Anopheles gambiae mosquitoes. The EMBO journal. 1997;16(20):6114-9.

22. Obbard DJ, Callister DM, Jiggins FM, Soares DC, Yan G, Little TJ. The evolution of TEP1, an exceptionally polymorphic immunity gene in Anopheles gambiae. BMC Evol Biol. 2008;8:274.

23. White BJ, Lawniczakb MKN, Chenga C, Coulibaly BM, Wilsond MD, Sagnone N, et al. Adaptive divergence between incipient species of Anopheles gambiae increases resistance to Plasmodium. |PNAS 2011;vol. 108(1):244–9

24. Fabrigar DJ, Hubbart C, Miles A, Rockett K. High-throughput genotyping of Anopheles mosquitoes using intact legs by Agena biosciences iPLEX. Mol Ecol Resour. 2016;16(2):480-6.

25. Gildenhard M, Rono EK, Diarra A, Boissiere A, Bascunan P, Carrillo-Bustamante P, et al. Mosquito microevolution drives Plasmodium falciparum dynamics. Nat Microbiol. 2019;4(6):941-7.

26. Pompon J, Levashina EA. A New Role of the Mosquito Complement-like Cascade in Male Fertility in Anopheles gambiae. PLoS Biol. 2015;13(9):e1002255.

27. Collins FH, Sakai RK, Vernick KD, Paskewitz S, Seeley DC, Miller LH, et al. Genetic selection of a Plasmodium-refractory strain of the malaria vector Anopheles gambiae. Science. 1986;234(4776):607-10.
28. Zhou G, Afrane AY, Vardo-Zalik AM, Atieli HE, Zhong D, Wamae P, et al. Changing Patterns of Malaria Epidemiology between 2002 and 2010 in Western Kenya: The Fall and Rise of Malaria. PLOS ONE. 2011;6(5).

29. Githeko AK, Ayisi JM, Odada PK, Atieli FK, Ndenga BA, Githure JI, et al. Topography and malaria transmission heterogeneity in western Kenya highlands: prospects for focal vector control. Malaria Journal. 2006;5:107.

30. Gimnig JE, Otieno P, Were V, Marwanga D, Abong’o D, Wiegand R, et al. The Effect of Indoor Residual Spraying on the Prevalence of Malaria Infection, Clinical Malaria and Anemia in an Area of Perennial Transmission and Moderate Coverage of Insecticide Treated Nets in Western Kenya. PLOS ONE. 2016( DOI:10.1371/journal.pone.0145282):1-29.

31. WHO. Entomological field techniques for malaria control. Part II. Tutor’s guide. World Health Organization. 1992.

32. Coetzee M. Key to the females of Afrotropical Anopheles mosquitoes (Diptera: Culicidae). Malaria Journal. 2020; 19; :1-20.

33. Das S, Garver L, Dimopoulos G. Protocol for Mosquito Rearing (A. gambiae). JoVE. (2007);5.

34. Gillies M, and Coetzee, M. A Supplement to the Anophlineae of Africa South of the Shaara (Afrotropical Region). Publications of the South African Institute for Medical Research. 1987;55.

35. Coetzee M. Editorial Distribution of the African Malaria Vectors of the Anopheles gambiae Complex. The American Society of Tropical Medicine and Hygiene. 2004;70(2):103-4.

36. Musapa M, Kumwenda T, Mkulama M, Chishimba S, Norris DE, Thuma PE, et al. A simple Chelex protocol for DNA extraction from Anopheles spp. J Vis Exp. 2013(71).

37. Scott JA, Brogdon WG, Collins FH. Identification of single specimens of the Anopheles gambiae complex by the polymerase chain reaction. Am J Trop Med Hyg. 1993;49:520-9.

38. Peakall R, and Smouse, P.E. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 2006;6 288-95.

39. Brown JL, Mucci D, Whiteley M, Dirksen M, and , Kassis JA. The Drosophila Polycomb Group Gene pleiohomeotic Encodes a DNA Binding Protein with Homology to the Transcription Factor YY1. Molecular Cell. 1998; 1:1057-64.

40. Kumar S, Stecher G, and, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Molecular Biology and Evolution. 2016;33(7):1870-4.

41. Hartl DLaC, G.C. Principles of Population Genetics. Sinauer Associates, Sunderland. 1997.

42. Garver LS, de Almeida Oliveira G, Barillas-Mury C. The JNK pathway is a key mediator of Anopheles gambiae antiplasmodial immunity. PLoS Pathog. 2013;9(9):e1003622.

43. Dong Y, Aguilar R, Xi Z, Warr E, Mongin E, Dimopoulos G. Anopheles gambiae immune responses to human and rodent Plasmodium parasite species. PLoS Pathog. 2006;2(6):e52.

44. Rono KE. Variation in the Anopheles gambiae TEP1 Gene Shapes Local Population Structures of Malaria Mosquitoes. Germany: University of Berlin; 2017.
45. Mancini E, Spinaci MI, Gordicho V, Caputo B, Pombi M, Vicente JL, et al. Adaptive Potential of Hybridization among Malaria Vectors: Introgression at the Immune Locus TEP1 between Anopheles coluzzii and A. gambiae in ‘Far-West’ Africa. PLoS One. 2015;10(6):e0127804.

46. Eldering M, Morlais I, van Gemert GJ, van de Vegte-Bolmer M, Graumans W, Siebelink-Stoter R, et al. Variation in susceptibility of African Plasmodium falciparum malaria parasites to TEP1 mediated killing in Anopheles gambiae mosquitoes. Sci Rep. 2016;6:20440.

47. WHO. https://www.who.int/malaria/areas/vector_control/core_methods/en/ 2020.

48. Smith HA, White BJ, Kundert P, Cheng C, Romero-Severson J, Andolfatto P, et al. Genome-wide QTL mapping of saltwater tolerance in sibling species of Anopheles (malaria vector) mosquitoes. Heredity 2015;115: 471-9.

49. Harris C, Morlais I, Churcher TS, Awono-Ambene P, Gouagna LC, Dabire RK, et al. Plasmodium falciparum produce lower infection intensities in local versus foreign Anopheles gambiae populations. PLoS One. 2012;7(1):e30849.

50. Horton AA, Lee Y, Coulibaly CA, Rashbrook VK, Cornel AJ, Lanzaro GC, et al. Identification of three single nucleotide polymorphisms in Anopheles gambiae immune signaling genes that are associated with natural Plasmodium falciparum infection. Malaria journal. 2010;9:160.

**Figures**

**Figure 1**

Geographical location of the mosquito collection sites in Western Kenya
Figure 2

Distribution of TEP1 genotypes and alleles circulating in *An. gambiae* and *An. arabiensis* in Bungoma, Kakamega, Homa Bay, and Kisumu Counties in western Kenya.

Figure 3
The evolutionary history was inferred by using the Maximum Likelihood method and Kimura 2-parameter model [1]. The tree with the highest log likelihood (-1872.22) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach and then selecting the topology with a superior log-likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.7700). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 43.64% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 15 nucleotide sequences. There were a total of 873 positions in the final dataset. Evolutionary analyses were conducted in MEGA X [2]. Red and green dots indicate haplotypes identified in this study; squares with different colors represent reference haplotypes extracted from GenBank.

**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable.docx