

RESEARCH

Open Access



# Molecular characterization and genotype distribution of thioester-containing protein 1 gene in *Anopheles gambiae* mosquitoes in western Kenya

Shirley A. Onyango<sup>1,2</sup>, Kevin O. Ochwedo<sup>2,5</sup>, Maxwell G. Machani<sup>3</sup>, Julius O. Olumeh<sup>2</sup>, Isaiah Debrah<sup>2,7</sup>, Collince J. Omondi<sup>2,5</sup>, Sidney O. Ogolla, Ming-Chieh Lee<sup>4</sup>, Guofa Zhou<sup>4</sup>, Elizabeth Kokwaro<sup>1</sup>, James W. Kazura<sup>8</sup>, Yaw A. Afrane<sup>6</sup>, Andrew K. Githeko<sup>3</sup>, Daibin Zhong<sup>4\*</sup>  and Guiyun Yan<sup>4\*</sup>

## Abstract

**Background:** Evolutionary pressures lead to the selection of efficient malaria vectors either resistant or susceptible to *Plasmodium* parasites. These forces may favour the introduction of species genotypes that adapt to new breeding habitats, potentially having an impact on malaria transmission. Thioester-containing protein 1 (TEP1) of *Anopheles gambiae* complex plays an important role in innate immune defenses against parasites. This study aims to characterize the distribution pattern of TEP1 polymorphisms among populations of *An. gambiae* sensu lato (*s.l.*) 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 in 627 anopheles mosquitoes using restriction fragment length polymorphisms-polymerase chain reaction (RFLP-PCR) and to validate the TEP1 genotyping results, a representative sample of the 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 *Anopheles gambiae* and *Anopheles arabiensis*. There was no significant difference detected among the populations 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.

\*Correspondence: dzhong@uci.edu; guiyun@uci.edu

<sup>4</sup> Program in Public Health, College of Health Sciences, University of California at Irvine, Irvine, CA 92697, USA  
Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

**Conclusion:** Genotyping TEP1 has identified two common TEP1 alleles (TEP1\*S1 and TEP1\*R2) and three corresponding genotypes (\*S1/S1, \*R2/S1, and \*R2/R2) in *An. gambiae s.l.* The TEP1 allele genetic diversity and population structure are low in western Kenya.

**Keywords:** *Anopheles gambiae*, Thioester-containing protein 1, Genetic diversity, Population structure, Signature of selection, Malaria transmission

## 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 may select vector genotypes or species that adapt to new breeding habitats. These factors may cause vectorial rearrangement exerting selection pressure that could change TEP1 allele frequencies and subsequently, efficient vectors could thrive and continue transmitting malaria. Despite the increased vector densities, malaria transmission is dependent on infectious parasites and competent vectors to influence susceptibility to infections in local vector populations. A vector's susceptibility and/or resistance to *Plasmodium* parasites is a determining factor for vector competence and is in part influenced by the thioester containing protein 1 (TEP1).

In *Anopheles gambiae*, TEP1 exhibits allelic variations that alter vector competence and subsequently influence malaria infectivity [10, 11]. These variations may be as a result of selective pressures such as climate change and vector control interventions acting on the TEP1 gene that eventually influence the vector's ability to transmit the *Plasmodium* parasite [11]. The TEP1 gene was reported to target the *Plasmodium* parasite in the early stages of infection in the mosquito host mostly the ookinetes [12, 13] either by melanization or lysis [14, 15] effectively reducing oocysts and sporozoite numbers in the vector. However, there is a lack of knowledge on how these allelic polymorphisms in vector competence affect malaria transmission [16, 17]. Furthermore, the distribution of the TEP1 allele in western Kenya regions with varying malaria transmission intensities is unknown. Therefore, understanding molecular mechanisms underlying mosquito genotypes and *Plasmodium* adaptations to different *Anopheles* species is important and could be used to monitor infection trends in vectors that directly have an impact on malaria transmission.

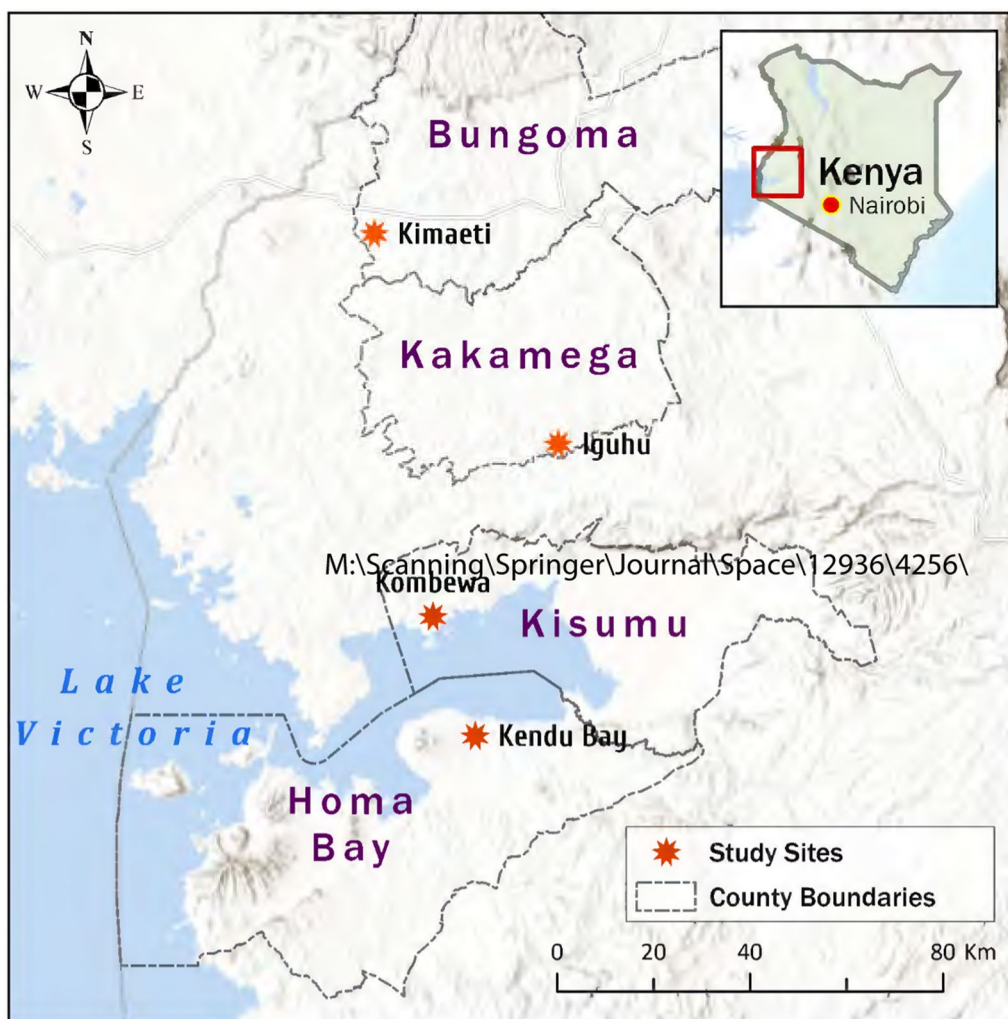
The complement-like thioester-containing protein 1 (TEP1) plays a key role in immunity against pathogens [17–20]. TEP1 is a highly polymorphic protein [21–23] located in the thioester domain (TED) on chromosome 3L coding for 1338 amino acids long protein contributing to phenotypic divergence and demonstrates genetic variations associated with distinct genotypes in its refractoriness to *Plasmodium* parasites. Six allelic classes;

TEP1\*S1, TEP1\*S2, TEP1\*S3, TEP1\*R1, TEP1\*R2, and TEP1\*R3 have recently been characterized in the *An. gambiae* complex in Africa [13, 14, 24]. TEP1\*S1 and TEP1\*R2 are the most common TEP1 alleles identified across Africa. The TEP1\*S1 however lacks a defined geographical structure. The TEP1\*S2 allele identified in the 4Arr strain is specific to *Anopheles coluzzii* [24] and gets rid of the damaged sperm cells in the male mosquitoes [25] bringing forth varying *Anopheles* population abundance. TEP1\*S3 allele closely related to TEP1\*S1 is fixed in the G3 strain associated with susceptibility of infection to *Plasmodium berghei* [13]. TEP1\*R1 identified in the L3-5 strain depicts the highest level of resistance to *Plasmodium* associated with melanization [13, 25, 26] and documented in *An. coluzzii* in West Africa [24]. A newly identified allele TEP1\*R3 is specific to the saline water mosquito, *Anopheles merus* found at the Kenyan coast. Selective pressures influence these variations in the genetic structure of the natural *An. gambiae* populations in different ecological settings and differences in their refractoriness to *Plasmodium* parasites are not clear. Genotyping TEP1 in local vector populations is, therefore, critical for monitoring changes in abundance that could explain sporozoite rates and potential malaria prevalence in varying levels of endemicities and is a potential tool for developing vector control interventions. Furthermore, information regarding the impact of vector control and environment changes on vector competence and underlying molecular mechanisms will significantly improve our understanding of malaria transmission dynamics. This study was designed to determine the distribution of TEP1 alleles circulating in *An. gambiae sensu lato (s.l.)* 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 1430–1545 m above sea level) in Bungoma, and Iguhu (34°45'E, 0°10'N; 1430–1580 m above sea level) in Kakamega, and two malaria-endemic lowland sites located around Lake Victoria; Kombewa (34°30'E,



**Fig. 1** Geographical location of the mosquito collection sites in Western Kenya

0°07'N; 1150–1300 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 1740 and 1940 mm annually. *Plasmodium falciparum* is the most common cause of malaria and is primarily transmitted by *An. gambiae sensu stricto (s.s.)*, *Anopheles funestus* and *Anopheles arabiensis* [27, 28]. The key vector control interventions are long-lasting insecticidal nets (LLINs) and indoor residual spraying (IRS) [29]. 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. Mosquito sampling was done during the middle of the dry season in February–March and 4 weeks after the start of the long rainy season in May–July between 2017 and 2020. Collections were conducted between 0630 and 1000 h 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  $-20^{\circ}\text{C}$  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 [30]. A total number of dips taken from each habitat was between 5 and 20 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 [31]. Only *Anopheles* larvae were sorted and transported to the ICEMR insectary. The larvae were reared to adults using standardized rearing methods [32]. Emerged adults were anesthetized using chloroform and identified using the morphological key in the laboratory as described by Gillies and Coetzee to species [33, 34].

#### Molecular identification of mosquito species

Genomic DNA was extracted from randomly selected single *An. gambiae* female adult using the Chelex resin (chelex<sup>®</sup> -100) method following a protocol by Musapa et al. [35]. 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 min. 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  $85^{\circ}\text{C}$  for 10 min, centrifuged at  $20,000 \times g$  for a minute, and DNA transferred into pre-labelled storage vials. *Anopheles gambiae* was identified to sibling species using polymerase chain reaction (PCR) as described by Scott et al. [36].

#### 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. [24]. Briefly, the initial PCR was conducted using Nest 1 primers - VB3 5' GATGTGGTG AGCAGAATATGG 3' and VB4- 5' ACATCAATTTGC TCCGAGTT 3' targeting 892 base pairs, followed by a second PCR performed on 5  $\mu\text{l}$  of the resulting product from Nest 1 with Nest 2 primers VB1 5' ATCTAATCG ACAAGCTACGAATTT 3' and VB2 5' CTTCAGTTG AACGGTGTAGTCGTT 3' producing a final fragment

length of 758 base pairs. Both PCR reaction conditions were set as denaturation at  $95^{\circ}\text{C}$  for 3 min, 35 cycles of  $94^{\circ}\text{C}$  for 30 s, annealing at  $55^{\circ}\text{C}$  for 30 s, extension at  $72^{\circ}\text{C}$  for 30 s, and a final step at  $72^{\circ}\text{C}$  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) (Additional file 1: Table S1) according to the manufacturer's instructions and analyzed the result with 2.5% agarose gel electrophoresis. The TEP1 allelic classes were then determined by fragment size of restriction enzyme digestion (Additional file 1: Table S1). A subset of samples with identified TEP1 alleles were further used for confirmatory purposes by sequencing 9 respective Nested II amplicons. Sequencing was done using 3700/3730 BigDye<sup>®</sup> Terminator v3.1 Sequencing Standard kit (ABI PRISM<sup>®</sup> 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 ( $H_o$ ), and expected heterozygosity ( $H_e$ ), the inbreeding coefficients ( $F_{IS}$ ), departure from Hardy-Weinberg expectations were analyzed using GenEalex version 6.053 software [37]. 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 [38] 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 [39]. AMOVA was used to determine the level of genetic allele differentiation among populations and within individuals. The  $F_{ST}$  values  $0 \leq 0.05$  were interpreted as low differentiation,  $0.05 \geq 0.15$  moderate differentiation and  $0.15 \geq 0.25$  high levels [40].

## Results

#### Species composition of *An. gambiae s.l.* across study sites

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 s.s.* and *An. arabiensis* constituting 49.28%



**Table 1** Molecular determined species composition in western Kenya

Sampling sites	<i>An. arabiensis</i> , n (%)	<i>An. gambiae</i> , 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)

(309/627) and 50.72% (318/627) of the total samples genotyped respectively (Table 1). There was a significant difference in species abundance (*An. gambiae* s.s. versus *An. arabiensis*) in the total analyzed samples ( $P < 0.0001$ ) however a significant difference in species catches was only observed in Kisumu and Homa bay ( $P < 0.0001$ ) (Table 1).

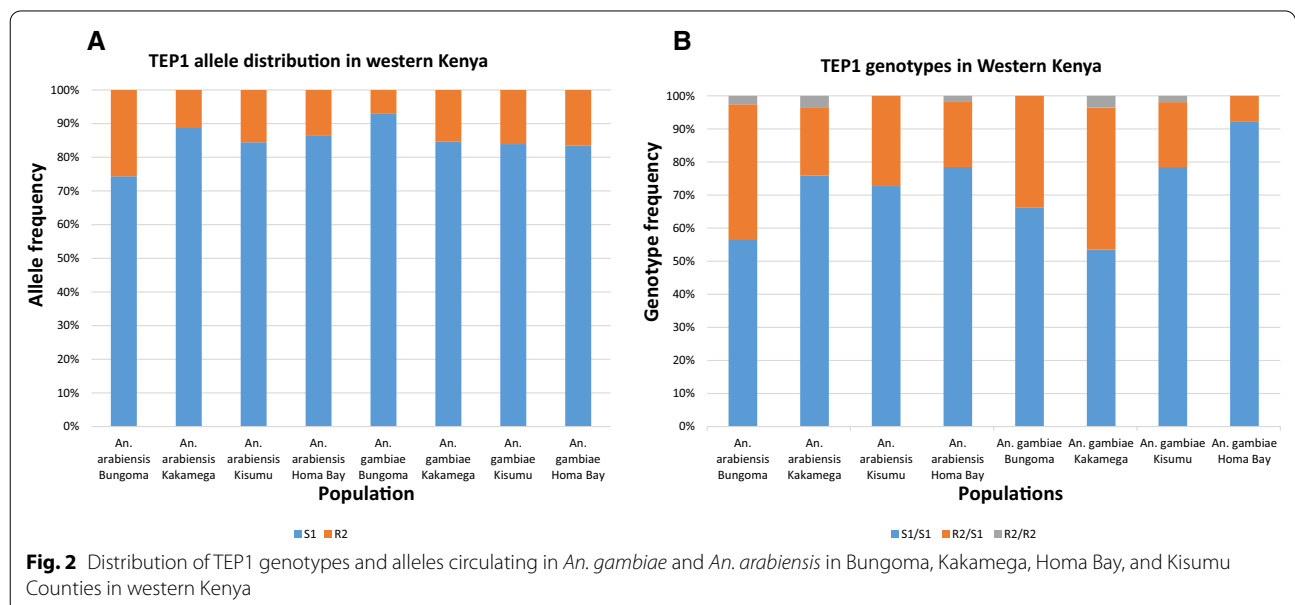
**TEP1 allele distribution in the study sites**

Overall, two TEP1 alleles (TEP1\*S1, and TEP1\*R2) were identified with average frequencies of 84.9% and 15.1%, respectively. *Anopheles arabiensis* populations from Homa Bay had the highest TEP1\*S1 allele frequency (89%, 95% CI 85.8%–92.2%) which significantly differed from observed proportions in Kisumu (86.4%, 95% CI 79.8%–92.9%), Kakamega (84.5%, 95% CI 74.9%–94.1%) and Bungoma (74.4%, 95% CI 64.5%–84.3%) (Two-tailed  $p < 0.0001$ ). Among *An. gambiae* s.s. populations from Bungoma displayed the highest TEP1\*S1 allele frequency (93.1%, 95% CI 88.7%–97.5%) followed by Homa Bay (84.6%, 95% CI 76.4%–92.8%), Kakamega (83.9%, 95%

CI 77%–90.8%), and Kisumu (83.5%, 95% CI 79.4%–87.7%) respectively (Fig. 2A). The observed TEP1\*S1 allele frequency in Bungoma significantly differed from Kakamega (two-tailed  $p = 0.0466$ ) and Kisumu (two-tailed  $p < 0.0001$ ). The highest TEP1\*R2 allele frequency among *An. arabiensis* was observed in vector populations from Bungoma (26%, 95% CI 15.7%–35.5%) followed by Kakamega (15.5%, 95% CI 5.91%–25.1%), Kisumu (13.6%, 95% CI 7.12%–20.2%), and Homa Bay (11%, 95% CI 8.06%–14.5%). In *An. gambiae* the TEP1\*R2 allele frequency was highest in populations from Kisumu and Kakamega displaying allele frequencies of 16.5%, 95% CI 12.3%–20.5% and 16.1%, 95% CI 9.16%–23%, respectively, followed by Homa Bay (15.4%, 95% CI 7.20%–23.6%) and Bungoma (7%, 95% CI 7.20%–23.6%), respectively. No significant differences in allele frequency were observed between species ( $P = 0.799$ ) and between site variation ( $P > 0.05$ ).

**TEP1 genotype distribution in the study sites**

A total of 3 genotypes were identified in populations of *An. gambiae* s.l. in western Kenya. 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



**Fig. 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

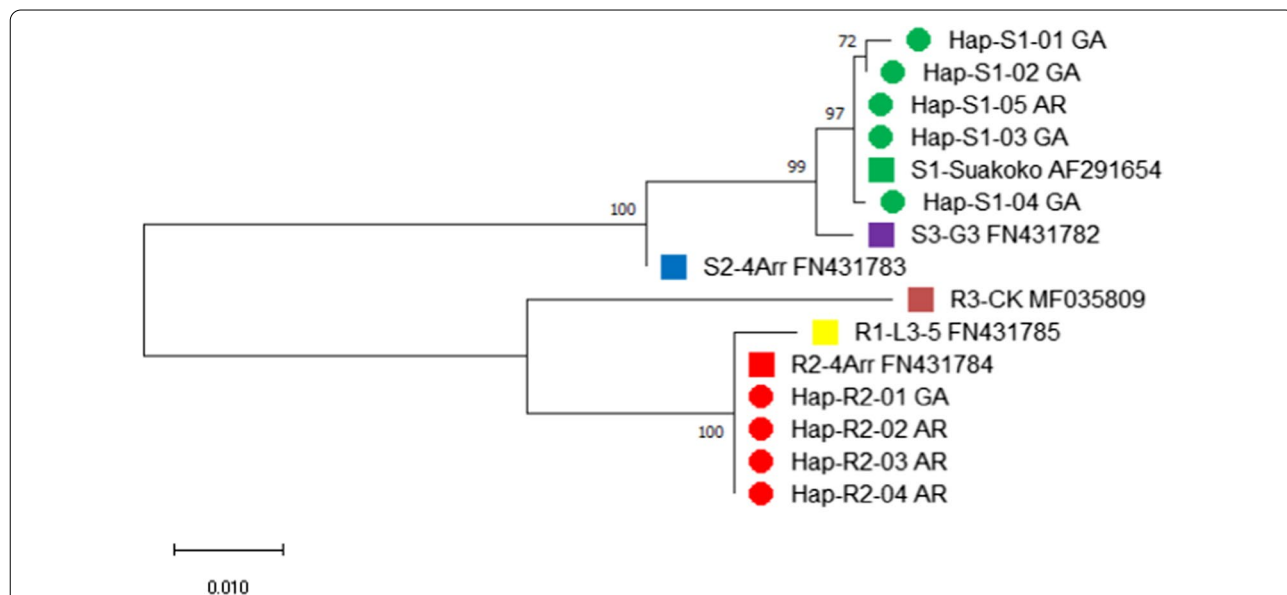
in the lowest average frequency of 1.64% (Fig. 2B). The TEP1\*S1/S1 genotype was predominant followed by TEP1\*R2/S1 but in low varied frequencies among species across all sampling sites. The TEP1\*S1/S1 genotype frequency was highest in *An. gambiae* as compared to *An. arabiensis* from all sites except 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* s.s. 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 = 0.07, 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 to 100%. Out of the sequences retrieved from the gene bank, TEP1\*S1 alleles identified in the study sites 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 the study sites 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 ( $H_o$ ) and expected heterozygosity ( $H_e$ ) of TEP domain in *An. gambiae* s.s. and *An. arabiensis* across all sites were  $0.270 \pm 0.035$  and  $0.251 \pm 0.025$ , respectively. There were slight variations between observed heterozygosity,  $H_o$



**Fig. 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

**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

ranges 0.188–0.462 in *An. arabiensis* and 0.138–0.321 in *An. gambiae* s.s. The *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 ().

The  $F_{IS}$  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  $F_{IS}$  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 analysed population was at HWE as all the computed values were non-significant ( $P > 0.05$ ). 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 ( $F_{ST}$ ) values revealed a low genetic differentiation among *An. arabiensis* and *An. gambiae*. Zero value represented complete Panmixis between species in the subpopulations. The  $F_{ST}$  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 ( $F_{ST} = 0.036$ ). The  $F_{ST}$  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 ( $F_{ST} = 0$ ). All pairwise  $F_{ST}$  values for *An. gambiae* and *An. arabiensis* from all regions across western Kenya demonstrated low population differentiation ( $0 \leq 0.05$ ) except *An. arabiensis* and *An. gambiae* from Bungoma that showed moderate differentiation ( $0.05 \geq 0.15$ ). The overall low levels in population structure ( $F_{ST} = 0.019$ ) 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.

The AMOVA results revealed that 99 percent of the observed variations in allele frequency were within each

**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

**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							
AR-Homa Bay	0.036	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

of the mosquitoes sampled ( $n=627$ ) within respective populations, and a 1% variation was observed among the eight populations, but no variations were observed among individuals (Table 4) indicating that the level of genetic differentiation between populations was very low.

## Discussion

*Plasmodium falciparum* triggers an immune response in *An. gambiae* mosquitoes [41]. Following infections with *P. falciparum* in *An. gambiae*, the midgut mounts specific and nonspecific immune responses to minimize epithelial damage [42]. Interactions between specific TEP1 and leucine-rich protein complex (LRIM1 and APL1C) are important components of the mounted immune response [18, 43]. This study identified two alleles (TEP1\*R2 and TEP1\*S1) in *An. gambiae* s.l from western Kenya regions with different malaria endemicities. 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 [24, 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 [13, 21, 22, 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 [11, 44, 46] as well as encourage evolutionary processes in the TEP1 loci [21]. Additionally, western Kenya still has relatively high malaria cases according to a recent study by Ochwedo et al. [47]. Implemented vector control interventions, such as insecticide-treated nets, indoor residual spraying, and environmental factors may determine the population structure. ITNs and IRS are two commonly used vector control interventions in Africa, and they have a direct impact on vector densities and species composition [48, 49]. For example, 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, consistent with this study, there was a drift in the local species composition with 99% of vectors in the post-spray period being *An. arabiensis* in the same sites [49–51]. Indoor interventions targeting *An. gambiae* complex remain the preferred method of lowering malaria transmission risk in endemic areas. *An. gambiae* s.s. is an anthropophilic indoor malaria vector [52, 53] and is susceptible to *P. falciparum*, which may explain

higher TEP1\*S1/S1 frequencies unlike *An. arabiensis* that is zoophilic and an outdoor dweller [47] which harboured TEP1\*R2/S1 genotypes. The susceptibility rate between the three TEP1 genotypes identified among *An. arabiensis* and *An. gambiae* s.s., would have been confirmed by examining receptivity and sporozoite prevalence, which were, however, outside the scope of this study. Understanding the underlying molecular mechanisms that determine vector competence remains important 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 [54, 55] and gene flow that may directly be shaping TEP1 alleles in *An. gambiae* s.l. in western Kenya. 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 [56].

## Conclusion

This study reveals minimal genetic differentiation and a low population structure of the TEP1 alleles 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, further studies should be carried out to investigate the implication of the current distribution of TEP1 alleles on vector competence and sporozoite rates in mosquito populations and the importance of TEP1 surveillance for malaria control.



## Abbreviations

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

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12936-022-04256-w>.

**Additional file 1: Table S1.** Restriction enzyme digestion of PCR products of TEPI gene in *An. gambiae* and *An. arabiensis* in western Kenya

## Acknowledgements

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.

## Author 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. All authors read and approved the final manuscript.

## Funding

The study was supported by grants from the National Institutes of Health (U19 AI129326, D43 TW001505, R01 AI050243, and R01 AI123074).

## Availability of data and materials

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

## Declarations

### Ethics approval and consent to participate

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.

### Competing interests

All authors declare that they have no conflict of interest.

### Author details

<sup>1</sup>Department of Zoological Sciences, School of Science and Technology, Kenyatta University, Nairobi, Kenya. <sup>2</sup>Sub-Saharan Africa International Centre of Excellence for Malaria Research, Homa bay, Kenya. <sup>3</sup>Centre for Global Health Research, Kenya Medical Research Institute, Kisumu, Kenya. <sup>4</sup>Program in Public Health, College of Health Sciences, University of California at Irvine, Irvine, CA 92697, USA. <sup>5</sup>Department of Biology, Faculty of Science and Technology, University of Nairobi, Nairobi, Kenya. <sup>6</sup>Department of Medical Microbiology, Medical School, University of Ghana, University of Ghana, Accra, Ghana. <sup>7</sup>Department of Biochemistry, Cell and Molecular Biology, West Africa Centre for Cell Biology of Infectious Pathogen, University of Ghana, Accra, Ghana. <sup>8</sup>Center for Global Health and Diseases, Case Western Reserve University, LC 4983, Cleveland, OH 44106, USA.

Received: 14 December 2021 Accepted: 3 August 2022

Published online: 10 August 2022

## References

- Lanzaro CG, Lee Y. Speciation in *Anopheles mosquitoes*—The distribution of genetic polymorphism and patterns of reproductive isolation among natural populations. In: Manguin S, editor. New insights into malaria vectors. Intech Open Publ; 2013. p. 173–96.
- 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. *Parasit Vectors*. 2010;3:117.
- Zhou G, Afrane YA, Dixit A, Ming-Chieh L, Wanjala CL, Beilhe LB, et al. Modest additive effects of integrated vector control measures on malaria prevalence and transmission in western Kenya. *Malar J*. 2013;12:256.
- 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. *Am J Trop Med Hyg*. 2003;68(4):16–22.
- 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;516:387–90.
- Tonnang H, Kangalawe R, Yanda PZ. Predicting and mapping malaria under climate change scenarios: the potential redistribution of malaria vectors in Africa. *Malar J*. 2010;9:111.
- Tanser FC, Sharp B, le Sueur D. Potential effect of climate change on malaria transmission in Africa. *Lancet*. 2003;362:1792–8.
- Lindsay SW, Birley MH. Climate change and malaria transmission. *Ann Trop Med Parasitol*. 1996;90:573–88.
- 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. *J Med Entomol*. 2002;39:833–41.
- Sinden RE, Alavi Y, Raine JD. Mosquito-malaria interactions: a reappraisal of the concepts of susceptibility and refractoriness. *Insect Biochem Mol Biol*. 2004. <https://doi.org/10.1016/j.ibmb.2004.03.015>.
- Le BV, Williams M, Logarajah S, Baxter RH. Molecular basis for genetic resistance of *Anopheles gambiae* to *Plasmodium*: structural analysis of TEPI 1 susceptible and resistant alleles. *PLoS Pathog*. 2012;8: e1002958.
- 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;3:364–74.
- 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;326:147–50.
- Blandin S, Shiao SH, Moita LF, Janse CJ, Waters AP, Kafatos FC, Levashina EA. Complement-like protein TEPI 1 is a determinant of vectorial capacity in the malaria vector *Anopheles gambiae*. *Cell*. 2004;116:661–70.
- Volohonsky G, Hopp AK, Saenger M, Soichot J, Scholze H, Boch J, et al. Transgenic expression of the anti-parasitic factor TEPI 1 in the malaria mosquito *Anopheles gambiae*. *PLoS Pathog*. 2017;13: e1006113.
- 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;11:456–69.
- Levashina EA, Baxter RHG. Complement-like system in the mosquito responses against malaria parasites. In: Stoute J, editor. Complement activation in malaria immunity and pathogenesis. Cham: Springer; 2018. p. 139–46.
- 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 Pathog*. 2013;9: e1003623.
- 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 Mol Biol*. 2000;9:75–84.
- Richman AM, Dimopoulos G, Seeley D, Kafatos FC. *Plasmodium* activates the innate immune response of *Anopheles gambiae* mosquitoes. *Embo J*. 1997;16:6114–9.
- Obbard DJ, Callister DM, Jiggins FM, Soares DC, Yan G, Little TJ. The evolution of TEPI 1, an exceptionally polymorphic immunity gene in *Anopheles gambiae*. *BMC Evol Biol*. 2008;8:274.

22. White BJ, Lawniczakb MKN, Chenga C, Coulibaly BM, Wilson MD, Sagnone N, et al. Adaptive divergence between incipient species of *Anopheles gambiae* increases resistance to *Plasmodium*. *Proc Natl Acad Sci USA*. 2011;108:244–9.
23. 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;16:480–6.
24. Gildenhard M, Rono EK, Diarra A, Boissiere A, Bascunan P, et al. Mosquito microevolution drives *Plasmodium falciparum* dynamics. *Nat Microbiol*. 2019;4:941–7.
25. Pompon J, Levashina EA. A new role of the mosquito complement-like cascade in male fertility in *Anopheles gambiae*. *PLoS Biol*. 2015;13:e1002255.
26. 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;234:607–10.
27. 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:e20318.
28. Githeko AK, Ayisi JM, Odada PK, Atieli FK, Ndenga BA, Githure JJ, et al. Topography and malaria transmission heterogeneity in western Kenya highlands: prospects for focal vector control. *Malar J*. 2006;5:107.
29. 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 parasite infection, clinical malaria and anemia in an area of perennial transmission and moderate coverage of insecticide treated nets in western Kenya. *PLoS ONE*. 2016;11:e0145282.
30. WHO. Entomological field techniques for malaria control. Part II. Tutor's guide. Geneva: World Health Organization; 1992.
31. Coetzee M. Key to the females of Afrotropical *Anopheles* mosquitoes (Diptera: Culicidae). *Malar J*. 2020;19:70.
32. Das S, Garver L, Dimopoulos G. Protocol for Mosquito Rearing (*A. gambiae*). *J Vis Exp*. 2007. <https://doi.org/10.3791/221>.
33. Gillies MT, Coetzee M. A supplement to the Anophelinae of Africa South of the Sahara (Afrotropical Region). *Publ S Afr Inst Med Res*. 1987;55:1–43.
34. Coetzee M. Distribution of the African malaria vectors of the *Anopheles gambiae* complex. *Am J Trop Med Hyg*. 2004;70:103–4.
35. 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;71:3281.
36. 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.
37. Peakall R, Smouse PE. GENALEX 6: genetic analysis in excel. Population genetic software for teaching and research. *Mol Ecol Notes*. 2006;6:288–95.
38. Brown JL, Mucci D, Whiteley M, Dirksen M, Kassis J. The drosophila polycomb group gene pleiohomeotic encodes a DNA binding protein with homology to the transcription factor YY1. *Mol Cell*. 1998;1:1057–64.
39. Kumar S, Stecher G, Tamura. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol*. 2016;33:1870–4.
40. Hartl DL, Clark AG. Principles of population genetics. 1st ed. Mass city: Sinauer Assoc; 1997.
41. Garver LS, de Almeida OG, Barillas-Mury C. The JNK pathway is a key mediator of *Anopheles gambiae* antiplasmodial immunity. *PLoS Pathog*. 2013;9:e1003622.
42. 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:e52.
43. Fraiture M, Baxter RHG, Steinert S, Chelliah Y, Frolet C, Quispe-Tintaya W, et al. Two mosquito LRR proteins function as complement control factors in the tep1-mediated killing of *Plasmodium*. *Cell Host Microbe*. 2009;5:273–84.
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: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 TEPI mediated killing in *Anopheles gambiae* mosquitoes. *Sci Rep*. 2016;6:20440.
47. Ochwedo KO, Omondi CJ, Magomere EO, Olumeh JO, Debrah I, Onyango SA, et al. Hyper-prevalence of submicroscopic *Plasmodium falciparum* infections in a rural area of western Kenya with declining malaria cases. *Malar J*. 2021;20:472.
48. WHO. Guidelines for malaria vector control [Internet]. 2019 [cited 2022 Aug 6]. Available from: <https://www.who.int/news/item/20-02-2019-who-publishes-new-guidelines-for-malaria-vector-control>
49. Russell TL, Govella NJ, Azizi S, Drakeley CJ, Kachur SP, Killeen GF. Increased proportions of outdoor feeding among residual malaria vector populations following increased use of insecticide-treated nets in rural Tanzania. *Malar J*. 2011;10:80.
50. Orondo PW, Nyanjom SG, Atieli H, Githure J, Ondeto BM, Ochwedo KO, et al. Insecticide resistance status of *Anopheles arabiensis* in irrigated and non-irrigated areas in western Kenya. *Parasit Vectors*. 2021;14:335.
51. Initiative USPM: Kenya Annual Entomological Monitoring Report. October 2017–September 2018.
52. Gillies MT, De Meillon B. The Anophelinae of Africa south of the Sahara (Ethiopian zoogeographical region). The Anophelinae of Africa south of the Sahara (Ethiopian Zoogeographical Region) 1968. 54.
53. White G, Magayuka SA, Boreham P. Comparative studies on sibling species of the *Anopheles gambiae* Giles complex (Diptera, Culicidae): bionomics and vectorial activity of species A and species B at Segera. *Tanzania Bull Entomol Res*. 1972;62:295–317.
54. 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:e30849.
55. Horton AA, Lee Y, Coulibaly CA, Rashbrook VK, Cornel AJ, Lanzaro GC. Identification of three single nucleotide polymorphisms in *Anopheles gambiae* immune signaling genes that are associated with natural *Plasmodium falciparum* infection. *Malar J*. 2010;9:160.
56. 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.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

