ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Population genomic analyses reveal population structure and major hubs of invasive *Anopheles stephensi* in the Horn of Africa

Jeanne N. Samake[1] ![ORCID] | Philip Lavretsky[2] ![ORCID] | Isuru Gunarathna[1] ![ORCID] | Madison Follis[1] ![ORCID] | Joshua I. Brown[3] ![ORCID] | Said Ali[4] | Solomon Yared[5] | Tamar E. Carter[1] ![ORCID]

[1]Department of Biology, Baylor University, Waco, Texas, USA

[2]Department of Biological Sciences, University of Texas at El Paso, El Paso, Texas, USA

[3]Department of Life, Earth, and Environmental Sciences, West Texas A&M University, Canyon, Texas, USA

[4]Ministry of Health Somaliland, Hargeisa, Somalia

[5]Department of Biology, Jigjiga University, Jigjiga, Ethiopia

Correspondence
Tamar E. Carter, Department of Biology, Baylor University, Waco, TX, USA.
Email: tamar_carter@baylor.edu

Abstract

Anopheles stephensi invasion in the Horn of Africa (HoA) poses a substantial risk of increased malaria disease burden in the region. An understanding of the history of introduction(s), establishment(s) and potential *A. stephensi* sources in the HoA is needed to predict future expansions and establish where they may be effectively controlled. To this end, we take a landscape genomic approach to assess *A. stephensi* origins and spread throughout the HoA, information essential for vector control. Specifically, we assayed 2070 genome-wide single nucleotide polymorphisms across 214 samples spanning 13 populations of *A. stephensi* from Ethiopia and Somaliland collected in 2018 and 2020, respectively. Principal component and genetic ancestry analyses revealed clustering that followed an isolation-by-distance pattern, with genetic divergence among the Ethiopian samples significantly correlating with geographical distance. Additionally, genetic relatedness was observed between the northeastern and east central Ethiopian *A. stephensi* populations and the Somaliland *A. stephensi* populations. These results reveal population differentiation and genetic connectivity within HoA *A. stephensi* populations. Furthermore, based on genetic network analysis, we uncovered that Dire Dawa, the site of a spring 2022 malaria outbreak, was one of the major hubs from which sequential founder events occurred in the rest of the eastern Ethiopian region. These findings can be useful for the selection of sites for heightened control to prevent future malaria outbreaks. Finally, we did not detect significant genotype–environmental associations, potentially due to the recency of their colonization and/or other anthropogenic factors leading to the initial spread and establishment of *A. stephensi*. Our study highlights how coupling genomic data at landscape levels can shed light into even ongoing invasions.

KEYWORDS
ddRAD-seq, invasive mosquito, landscape genomics, malaria

1 | INTRODUCTION

Malaria remains one of the leading global health concerns, with an estimated 241 million cases and over 600,000 deaths worldwide in 2020 (WHO, 2021). In Africa, members of the *Anopheles gambiae* complex have been the predominant vectors, with *Anopheles arabiensis* acting as the primary vector in parts of the Horn of Africa (HoA) which includes Djibouti, Eritrea, Ethiopia and Somalia. However, recently *Anopheles stephensi*, a natural vector in South Asian and Middle Eastern countries, is now being readily found in some East African countries and across Indian Ocean Islands (Ishtiaq et al., 2021). In Africa, *A. stephensi* was first detected in Djibouti in 2012 (Faulde et al., 2014), Ethiopia in 2016 (Carter et al., 2018), Sudan and Somaliland in 2019 (Ahmed et al., 2021; Ali et al., 2022) and Nigeria in 2020 (WHO, 2022a). Among these, the detection of *A. stephensi* in Djibouti in 2012 correlated with an exponential growth of malaria cases from 1684 in 2013 to over 60,000 in 2020 (de Santi et al., 2021; Faulde et al., 2014; Seyfarth et al., 2019; WHO, 2021). This rapid spread of *A. stephensi* led the World Health Organization (WHO) to issue a 'vector alert' in 2022, aiming to halt the further spread of this invasive malaria vector in Africa (WHO, 2022b).

Although often associated with urban settings, *A. stephensi* can also proliferate in rural environments due to its broad and general habitat needs during its various lifecycle stages. In particular, the ability to use any form of water storage containers, including artificial containers at the larval stage allows *A. stephensi* the ability to readily move between and within rural and urban settings (Imwong et al., 2011; Thomas et al., 2017). Along with their generalist ecology, *A. stephensi* is a highly potent malaria vector that can transmit both *Plasmodium falciparum* and *P. vivax* (Bala-baskaran Nina et al., 2017). Thus, *A. stephensi* threatens malaria disease reduction and elimination efforts in Africa that have successfully lowered malaria transmission rates due in part to urbanization and improved housing in the past decade (Hay et al., 2005; Mathanga et al., 2016). This concern of an invasive urban malaria vector in the HoA led the WHO to recommend increased vector surveillance, the use of larvicides and the removal or modification of breeding sites in urban and peri-urban environments to mitigate and prevent the further spread of *A. stephensi* in Africa (WHO, 2019). Population models predict that uncontrolled range expansion of *A. stephensi* could lead to a substantial increase in at-risk populations (Sinka et al., 2020), including a potential 50% increase of malaria cases in Ethiopia (Hamlet et al., 2022). In fact, a recent malaria outbreak in Dire Dawa City in eastern Ethiopia in the spring of 2022 (Tadesse et al., 2023; preprint) exemplifies the threat of more malaria outbreaks associated with *A. stephensi* range expansions. Thus, the invasion of *A. stephensi* leading to a resurgence of malaria in the HoA could lead to the same trend further into Africa's interior countries.

An understanding of the history of introduction(s), establishment(s) and potential of these incoming *A. stephensi* sources in the HoA is needed not only to predict future expansions but also to establish where they may be effectively controlled. Specifically, establishing population structure of the invading *A. stephensi* populations is essential for uncovering vector invasion routes, distribution dynamics and evolution in its new environments. However, to demarcate source(s) and the demographic history of founding population(s), it is essential to not only sample at the leading edge of the invasion but also include long-established *A. stephensi* populations (Carter et al., 2018). Specifically, we expect high genetic similarity and a single population model if there is ongoing genetic connectivity (i.e. gene flow) among populations (Hague & Routman, 2016; Lerner, 2001). Alternatively, we expect genetic distinctiveness and an isolation-by-distance model under an invasion scenario followed by sequential founder events (Hague & Routman, 2016; Lerner, 2001). Moreover, if meta-population dynamics exist, then we expect the source of each invasion to harbour the highest levels of genetic diversity as they carry all the diversity present in the subsequent founder populations (Schmidt et al., 2021; Sherpa et al., 2019). Furthermore, understanding how invading populations respond to their new environments is critical to determine whether there may be any geographical limitations to their invasion (Lambrechts et al., 2006), which is currently lacking for *A. stephensi*. For example, genetic diversity found in *Aedes aegypti*, a vector found to share larval habitat with *A. stephensi* (Balkew et al., 2020), showed local adaptation and structured to local environmental conditions (Bennett et al., 2021).

Towards reconstructing *A. stephensi* invasion, establishment and expansion, we take a landscape genomics approach by assessing thousands of nuclear loci and mitochondrial DNA for *A. stephensi* ranging across eastern Ethiopia and Somaliland. In addition to characterizing population structure and genetic connectivity, we associate standing genetic diversity with various environmental factors to understand how these may be contributing to their expansion. Together, these data will not only help us understand how *A. stephensi* invaded the region but also inform targeted vector control approaches and surveillance of this invasive malaria species in the HoA.

2 | METHODS

2.1 | Sample descriptions

Samples originated from 13 locations where *A. stephensi* was found across eastern Ethiopia and Somaliland and collected in 2018 and 2020, respectively (Table 1; also see Ali et al., 2022 and Balkew et al., 2020). Briefly, adult mosquitoes were collected using pyrethrum spray sheet collections (PSC) and Centers for Disease Control (CDC) light traps. Larvae and pupae of *Anopheles* were collected from different larval breeding habitats per site, including artificial water containers, to reduce sampling siblings and reared to adults in field insectaries, and then morphologically and molecularly identified as *A. stephensi*. We further categorized collection sites in eastern Ethiopia (northeastern, east-central and southeastern) based on proximity to specific major roads and relative location across sites.

TABLE 1 *Anopheles stephensi* sample description from 13 populations. Altitude in metres above sea level (masl).

Population ID	Site	Samples passing QC	Altitude (masl)	Coordinates
Northeastern Ethiopia	Northeastern Ethiopia			
BA	Bati	21	2055	11°1′92″ N, 40°0′17″ E
GW	Gewane	16	617	10°1′66″ N, 40°6′46″ E
SM	Semera	21	431	11°7′94″ N, 41°0′08″ E
East-central Ethiopia	East-central Ethiopia			
AW	Awash Sebat Kilo	11	916	8.9′89″ N, 40°1′64″ E
DD	Dire Dawa	20	1178	9°5′96″ N, 41°8′54″ E
ER	Erer Gota	17	1090	9°5′56″ N, 41°3′84″ E
JJ	Jigjiga	14	1657	9°3′51″ N, 42°7′93″ E
Southeastern Ethiopia	Southeastern Ethiopia			
DE	Degehabur	21	1065	8°2′23″ N, 43°5′58″ E
GD	Godey	20	294	5°9′49″ N, 43°5′53″ E
KB	Kebridehar	22	532	6°7′38″ N, 44°2′77″ E
Somaliland	Somaliland			
SL	Lawyacado	11	7	11°27′30″ N, 43°15′47″ E
SB	Berbera	11	10	10°26′17″ N, 45°1′12″ E
MH	Hargeisa	9	1334	9°34′6″ N, 44°4′55″ E

'Northeastern' sites included the northern-most sites (Bati, Gewane, Semera and Awash Sebat Kilo) off the B11 and A1 roads (i.e. the road from Mille town to Kombolcha town and the main road between Addis Ababa and Djibouti, respectively). 'East-central' sites (Dire Dawa, Erer Gota and Jigjiga) were located in the centre relative to the rest of the collection sites along the A10 road (i.e. the road between Addis Ababa and Degehabur) running west/east. 'South-eastern' sites were the southern-most sites (Degehabur, Godey and Kebridehar) off the A10 road running north/south (Table 1, Figure 1a; also see Carter et al., 2021). Being 21.6 km away, we used Lawyacado in this nuclear dataset as a proxy for Djibouti City, where *A. stephensi* was first detected in the HoA.

2.2 | DNA extraction

Genomic DNA was extracted from a total of 214 (173 lab-reared and 41 wild-caught) adult *A. stephensi* mosquitoes using Qiagen DNeasy Blood and Tissue kit (Qiagen). DNA quality was assessed on a 1% agarose gel to ensure high-molecular-weight bands and quantified using Nanodrop One Spectrophotometer (Thermo Fisher Scientific Inc.) to ensure a minimum concentration of 20 ng/μL.

2.3 | ddRAD-seq library preparation, sequencing and bioinformatics

Genome-wide single nucleotide polymorphism (SNP) data were collected using the double-digest restriction-site-associated DNA

sequencing (ddRAD-seq) protocol outlined in Lavretsky et al. (2015), but with fragment size selection following Hernández et al. (2021). Briefly, genomic DNA was enzymatically fragmented using SbfI and EcoRI restriction enzymes, and Illumina TruSeq compatible barcodes were ligated to the sticky ends generated for demultiplexing purposes. The libraries were quantified and pooled in equimolar concentrations, and the multiplexed library was sequenced on Illumina HiSeq X using single-end 150 bp chemistry at Novogene (Novogene CO., Ltd.; see detailed methods, Document S1). Raw Illumina reads have been deposited in NCBI's Sequence Read Archive (Accession # SAMN31227709–SAMN31227956).

Raw Illumina reads were first demultiplexed based on perfect barcode/index sequences using the script *ddRADparser.py* (DaCosta & Sorenson, 2014). We then used Trimmomatic (Bolger et al., 2014) to trim or discard poor-quality sequences using a PHRED score of ≥30 to ensure only high-quality sequences were retained. We then used the Burrows–Wheeler Aligner (bwa; Li & Durbin, 2011) to align the remaining quality reads to the *A. stephensi* reference genome (Accession PRJNA661063; Chakraborty et al. (2022)). Next, samples were sorted and indexed in Samtools (Li et al., 2009). Samples were then genotyped using the 'mpileup' function in BCFtools (Li, 2011). At the BCFtools mpileup genotyping step, we filtered out any base pair with PHRED score < 25 and sequences with average PHRED score < 30. Lavretsky et al. (2020) described these methods in detail. In short, bioinformatics followed Lavretsky et al. (2020) with all steps through genotyping automated using custom in-house Python scripts (Python scripts available at https://github.com/jonmohl/PopGen; Lavretsky et al., 2020).
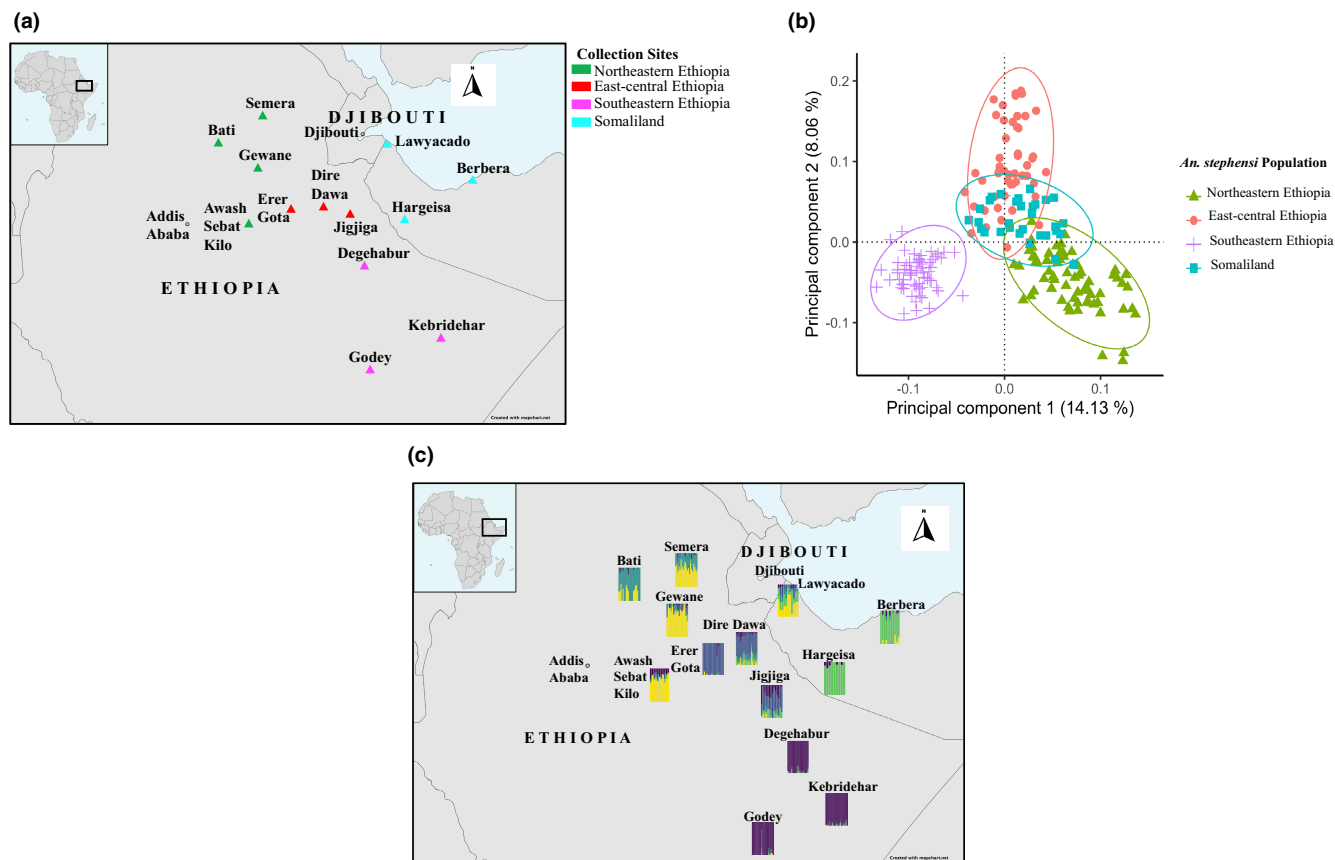
FIGURE 1 (a) *Anopheles stephensi* sampling locations in eastern Ethiopia and Somaliland. Green triangles represent populations from northeastern Ethiopia (Bati, Gewane, Semera and Awash Sebat Kilo), red triangles represent populations from east-central Ethiopia (Dire Dawa, Erer Gota and Jigjiga), purple triangles represent populations from southeastern Ethiopia (Degehabur, Godey and Kebridehar), and populations from Somaliland (Hargeisa, Berbera and Lawyacado) are colour-coded in blue. (b) Scatterplot of subgroup variation based on principal component analysis (PCA). The amounts of variation explained by each principal component (PC 1 on the *x*-axis, PC 2 on the *y*-axis) are given in percentages. *A. stephensi* population subgroups are colour-coded as described in panel (a). (c) Map with collection site-specific admixture plots.

2.4 | Nuclear population structure

Nuclear population structure was assessed across samples using an independent set of biallelic SNPs filtered in PLINK v.1.9 (Purcell et al., 2007) for singletons (i.e. minimum allele frequency (--maf 0.004) and any SNP missing ≥20% of data across samples (--geno 0.2)). Independence between SNPs was based on pairwise analysis of linkage disequilibrium (LD) (--indep-pairwise 2 1 0.5) in which one of two linked SNPs are randomly excluded if an LD correlation factor (r^2) > .5 is obtained.

First, we identified variation among samples with a principal component analysis (PCA) using the --pca function in PLINK v.1.9 (Purcell et al., 2007) and visualized with the R package ggplot2 (Wickham, 2016). Next, individual maximum likelihood population assignment probabilities were attained across samples using ADMIXTURE v.1.3 (Alexander et al., 2009), and datasets were formatted using PLINK v.1.9 (Purcell et al., 2007). Each ADMIXTURE analysis was run with a 10-fold cross-validation (CV) and with a quasi-Newton algorithm to accelerate convergence (Zhou et al., 2011). To limit possible stochastic effects, each analysis was

based on 1000 bootstraps for each population *K* value of 1–10. The block relaxation algorithm for point estimation was used for each analysis and terminated once the log-likelihood of the point estimation increased by <0.0001. The optimum population value was based on the average of CV errors across the analyses per *K* value. ADMIXTURE assignment probability outputs were visualized using the R package ggplot2 (Wickham, 2016). Additionally, we evaluated patterns of co-ancestry using fineRADstructure (Malinsky et al., 2018), which infers a matrix of co-ancestry coefficient based on the distribution of identical or nearest neighbour haplotypes among samples. Co-ancestry at each locus is equally divided among all individuals with identical haplotypes, or in the case of a unique allele with the nearest neighbour haplotype (Malinsky et al., 2018). Hence, rare haplotypes characterized by rare SNPs, which are on average of more recent origin (Kimura & Ohta, 1973), contribute the most to the co-ancestry index, providing a measure that highlights recent co-ancestry. A burn-in of 100,000 iterations, followed by 100,000 Markov chain Monte Carlo iterations, was completed, followed by tree building using default parameters. To visualize the results, we used the R scripts

fineradstructureplot.r and finestructurelibrary.r (available at http://cichlid.gurdon.cam.ac.uk/fineRADstructure.html).

Additionally, nucleotide diversity (π), counts of segregating SNPs (S), Tajima's D and pairwise estimates of fixation index (F_{st}) by site were calculated in the R package PopGenome (Pfeifer et al., 2014) and pairwise F_{st} by SNP locus was calculated with VCFtools (Danecek et al., 2011). Finally, we used the Mantel test statistic (r) implemented in the R package Adegenet to test for any isolation-by-distance effects by assessing the correlation between genetic and geographic distances (Jombart, 2008).

2.5 | Genetic network

To uncover genetic connectivity among the sampled populations, we performed a network analysis using EDENetworks (Kivelä et al., 2015), which allows network analyses based on genetic distance matrices without a prior assumption. The network consists of nodes representing populations connected by edges/links weighted by their F_{st}-based Reynolds' genetic distances (D; Reynolds et al., 1983), which provide the strength of connectivity between pairs of populations (Kivelä et al., 2015). The thicker the edge/link, the stronger the genetic connectivity between the two populations. Moreover, node size is proportional to the cumulative weighted edge linkages for each population. Thus, the larger the node the higher the connectivity hub or sink. Statistical confidence of the nodes was evaluated using 1000 bootstrap replicates. Nodes that appear in the top 5 and top 1 lists of betweenness centrality (BC) values (number of shortest genetic paths passing through a node) can be considered as statistically significant (Kivelä et al., 2014).

2.6 | Updated mitochondrial DNA genetic diversity and structure

Anopheles stephensi cytochrome oxidase subunit 1 (COI) sequences were retrieved from previously published data from eastern Ethiopia ($n = 191$, Carter et al. (2021)), Somaliland ($n = 33$, Ali et al. (2022)) and Djibouti ($n = 20$, de Santi et al. (2021)). Population genetic statistics were generated in DNAsp version 5 (Rozas et al., 2003) for each site from the studies listed above. The statistics generated included the number of polymorphic (segregating) sites (s), number of haplotypes (h), haplotype diversity (H_d), nucleotide diversity (π) and average number of nucleotide differences (k). Population differentiation between sites was determined through pairwise F_{st} (differentiation based on haplotype frequencies only) in Arlequin version 3.5.2.2 software (Excoffier & Lischer, 2010). To test the significance of the derived pairwise F_{st}, 100 permutations were performed, and p-values generated. Significance was set at alpha < 0.001. Next, we mapped mitochondrial COI haplotype proportions across sampled regions. Haplotype numbering is the same as in Carter et al. (2021). *Anopheles stephensi* COI sequence alignment was created in MAFFT version 7

(Katoh & Standley, 2013). Finally, a COI phylogeographic tree was reconstructed using a maximum likelihood approach through RAxML (Randomized Accelerated Maximum Likelihood, Stamatakis (2014)) following methods described in Carter et al. (2021). Trees were annotated in FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

2.7 | Genotype–environment association

To better understand factors contributing to *A. stephensi* genetic variation and expansion across eastern Ethiopia and Somaliland, we performed a genotype–environment association test following the gradient forest (GF) modelling approach detailed in Brown et al. (2022). Briefly, we used a total of 29 environmental predictor variables, which are reported to influence the ecology and expansion of *A. stephensi* across eastern Ethiopia (Balkew et al., 2020; Sinka et al., 2020) and Somaliland (Ali et al., 2022; Sinka et al., 2020). Specifically, our environmental climate variables include 19 Bioclimatic variables from the CHELSA v2.1 database (Karger et al., 2017, 2021), seasonal and annual vegetation indices (NDVI, EVI), elevation (STRM), land cover (Sentinel 2), net primary productivity (NPP) from the USGS AppEEARS database (https://lpdaacsvc.cr.usgs.gov/appeears) and population density from NASA Socioeconomic Data and Applications Center (SEDAC, https://sedac.ciesin.columbia.edu; Table S4). In order to differentiate the effects of annual versus seasonal vegetation processes, we calculated an average annual, first rainy season (April) and second rainy season (October) value in the HoA for NDVI and EVI based on data collected from 2000 to 2019.

Following the approach of Brown et al. (2022); also see Bay et al. (2018) and Fitzpatrick and Keller (2015), we used a machine learning regression tree-based analysis (i.e. random forests from the R package GradientForest (Ellis et al., 2012)). Briefly, GF detects the effects of environmental predictor variables across the landscape on turnover in allele frequencies between sample sites. For use as response variables, we used the R package PopGenome (Pfeifer et al., 2014) to calculate the minor allele frequencies from the previously filtered independent biallelic autosomal SNPs. Additionally, we filtered any SNP that was polymorphic in fewer than five sample sites in order to reduce the potential for false positives. Using a large number of trees ($N = 5000$), GF produced an R^2 ranked list of weighted importance for all environmental and human disturbance variables. To assess the strength of actual GF models, we randomized the environmental predictor data in relation to sampling sites; we then compared the performance of these 100 randomized models to our observed data. To visualize the GF model across eastern Ethiopia and Somaliland, we extracted values for the top five predictor variables from random points generated across this range. We then used a principal component analysis (PCA) to summarize the transformed values from the top five predictor variables (based on GF goodness of fit R^2 weighted importance) for each point. Based on the GF goodness of fit R^2, only SNPs with positive R^2 values are significantly associated with genetic turnover associated with environmental predictor variables (Ellis et al., 2012; Fitzpatrick &

Keller, 2015). Finally, we transformed the top three principal components to create a RGB colour scale that was used to visualize patterns of adaptive genetic diversity across the landscape. In the end, colours reflect associations between allele frequencies and the environmental predictor variables that allow us to draw conclusions about how the environment has affected genetic diversity and putatively driven adaptation.

3 | RESULTS

3.1 | Nuclear population structure

Across samples, we recovered 2070 ddRAD-seq autosomal loci (5675 base pairs) that met our criteria for sequencing coverage and missing data. Mean coverage per sample was 135× (range 10–179×). Moreover, genetic diversity and population structure analyses were based on 1680 (of 2070) independent biallelic SNP dataset collected across samples.

In the *A. stephensi* populations that were sampled, the lowest genetic diversity was observed in the southeastern Ethiopia region and the highest in the northeastern Ethiopia region, as assessed by basic summary statistics (Table 2). The lowest nucleotide diversity was observed in Godey ($\pi=0.1942$), and the highest nucleotide diversity was observed in Awash Sebat Kilo ($\pi=0.2233$) (Table 2). *Anopheles stephensi* populations from the southeastern Ethiopian region were also found to have the highest Tajima's *D* values, 1.37, 1.47 and 1.13 for Degehabur, Kebridehar and Godey, respectively, indicating a lack of rare variants relative to neutral expectation (Table 2; Tajima, 1989). Moreover, Somaliland *A. stephensi* populations had high

nucleotide diversity even with limited sample sizes compared with the Ethiopian *A. stephensi* populations (Table 2). Specifically, in the Somaliland region, the lowest nucleotide diversity was observed in Hargeisa ($\pi=0.2014$, $N=9$) and the highest nucleotide diversity was observed in Lawyacado ($\pi=0.2191$, $N=11$; Table 2). Assessing population structure using principal component analysis (PCA; Figure 1b; Figure S1), ADMIXTURE (Figure 1c; Figure S2) and fineRADstructure (Figure 2) analyses consistently recovered four semidiscrete genetic clusters and with general genetic structuring and pairwise F_{st} estimates (Figure 3) consistent with an isolation-by-distance (IBD) pattern (Mantel test observation: $r=.54$, p-value$<.001$; Figure S3). On average, 91 out of 1680 SNPs were recovered as outliers driving this observed IBD pattern based on the F_{st} values per SNP locus (Figure S4, Table S1). In general, plotting the first two principal components clustered samples regionally (Figure 1b; Figure S1), while ADMIXTURE analysis based on an optimum population *K* model of 5 more specifically identified four main genetic similarities across the regions that include (1) Lawyacado, Semera, Bati, Gewane and Awash Sebat Kilo; (2) Erer Gota, Dire Dawa and Jigjiga; (3) Degehabur, Kebridehar and Godey; and (4) Hargeisa and Berbera (Figure 1c). From these genetic clusters, Lawyacado, Dire Dawa and Jigjiga showed substantially higher admixture proportions (Figure 1c; Figure S2). In fact, the derived co-ancestry matrix from the fineRADstructure analysis (Figure 2) provided further resolution and confirmation of the above genetic clusters. Specifically, fineRADstructure results revealed some levels of recent co-ancestry between Awash Sebat Kilo, Gewane, Lawyacado, Semera and Bati *A. stephensi*, followed by Erer Gota, Dire Dawa and Jigjiga *A. stephensi*, then Hargeisa and Berbera *A. stephensi*, and finally Degehabur, Kebridehar and Godey *A. stephensi* (Figure 2). Similar to PCA and ADMIXTURE analyses,

TABLE 2 Basic summary statistics based on autosomal loci.

Collection site	Sample size *N*	Nucleotide diversity π	Counts of segregating SNPs S	Tajima's *D*
Northeastern Ethiopia				
Semera	21	0.2205	1332	0.74
Bati	21	0.2036	1174	0.95
Gewane	16	0.2163	1238	0.71
Awash Sebat Kilo	11	0.2233	1167	0.71
East-central Ethiopia				
Erer Gota	17	0.2137	1171	0.94
Dire Dawa	20	0.2199	1327	0.64
Jigjiga	14	0.2057	1155	0.65
Southeastern Ethiopia				
Degehabur	21	0.1958	1036	1.37
Kebridehar	22	0.2028	1057	1.47
Godey	20	0.1942	1062	1.13
Somaliland				
Hargeisa	9	0.2014	977	0.82
Berbera	11	0.2071	1078	0.73
Lawyacado	11	0.2191	1234	0.74

FineRADstructure analysis also showed Degehabur, Kebridehar and Godey *A. stephensi* populations as an isolated branch in the fineRAD-structure dendrogram suggesting a separate founder event in the southeastern Ethiopian region (Figure 2).

Pairwise F_{st} estimates also revealed genetic differentiation within these studied populations, with F_{st} values ranging from 0.04 to 0.15 (Figure 3). Southeastern Ethiopia *A. stephensi* populations were again highly differentiated (Figure 3), whereas the Dire Dawa *A. stephensi* population in east-central Ethiopia was least differentiated from the other studied *A. stephensi* populations (Figure 3).

3.2 | Genetic network

Network reconstruction was based on the same 1680 independent biallelic SNP dataset. Although all sites were recovered to be interconnected, Dire Dawa was recovered as the statistically significant node in the genetic network, followed by Godey, Semera, Jigjiga, Kebridehar, Gewane and Lawyacado (Figure 4), as they appeared in both the top 5

and top 1 lists of nodes from the bootstrapping procedure (Figure S5). The network revealed high genetic connectivity between Dire Dawa and sites in northeastern Ethiopia (Semera (SM), Gewane (GW)), east-central Ethiopia (Erer Gota (ER), Jigjiga (JJ)), southeastern Ethiopia (Kebridehar (KB), Godey (GD)) and Somaliland (Lawyacado (SL); Figure 4). This finding suggests Dire Dawa is a major *A. stephensi* hub among the studied populations. Other pairs of sites with high genetic connectivity were Lawyacado (SL) and Semera (SM), Bati (BA) and Semera (SM), Gewane (GW) and Semera (SM) and between the trio Degehabur (DE), Kebridehar (KB) and Godey (GD; Figure 4).

3.3 | Updated mitochondrial DNA genetic diversity and structure

Mitochondrial COI analyses were based on 317 base pairs of overlapping sequences for 244 samples (Table S2). Within the HoA, we recovered eight COI haplotypes, with different frequencies across the region (Figure 5; Table S2). The haplotypes are based
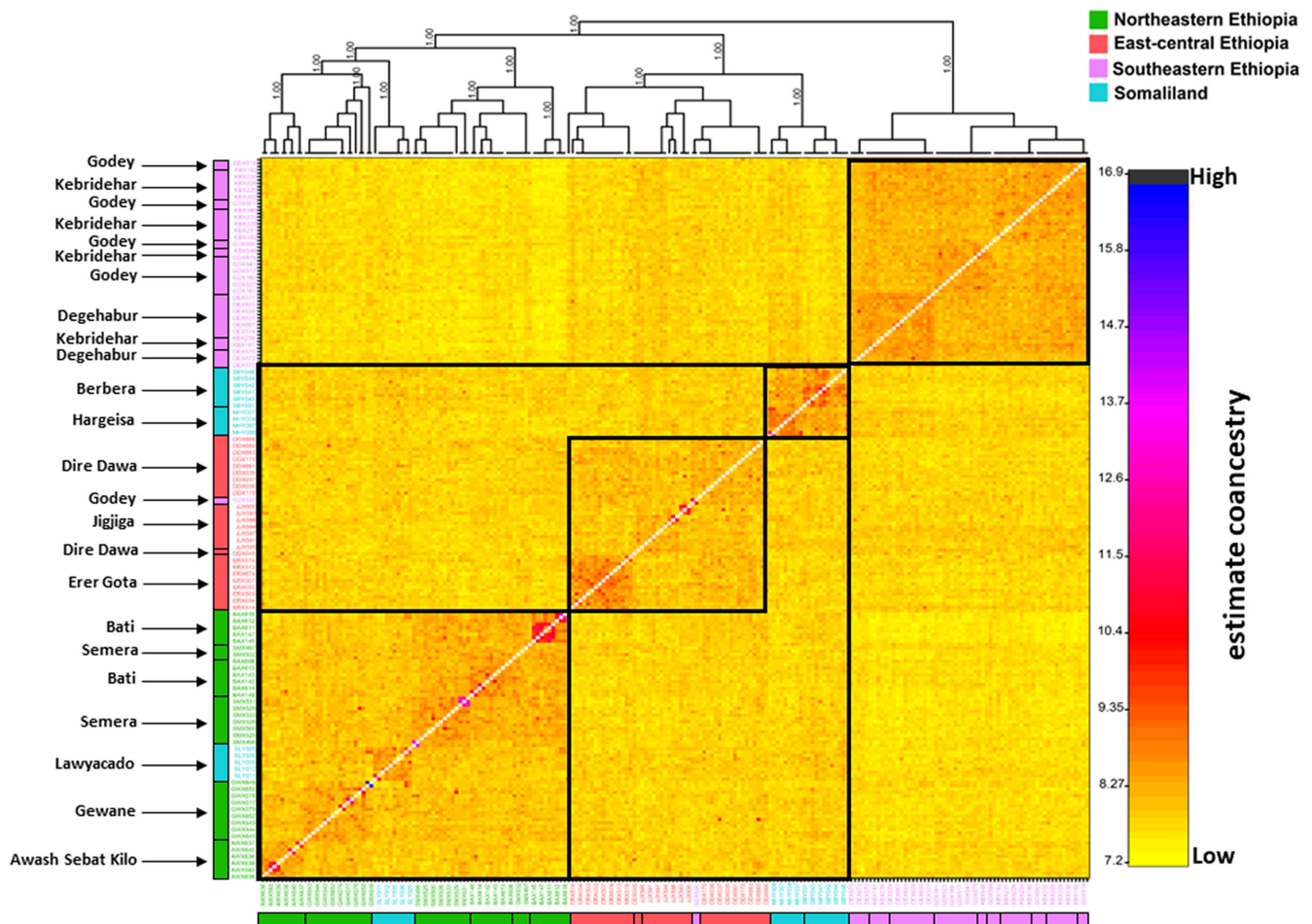

FIGURE 2 Co-ancestry matrix based on fineRADstructure analysis of autosomal ddRAD-seq loci of *Anopheles stephensi*. The dendrogram depicts clustering of individual samples based on the pairwise matrix of co-ancestry coefficients. Pairwise coefficients of co-ancestry are colour-coded from low (yellow) to high (blue) according to the scale shown on the right. Samples are colour-coded based on sample descriptions (see Table 1, Figure 1). Square black box demarcations identify individuals showing higher co-ancestry.

		Northeastern Ethiopia				East-central Ethiopia			Southeastern Ethiopia			Somaliland		
		Semera	Bati	Gewane	Awash Sebat Kilo	Erer Gota	Dire Dawa	Jigjiga	Degehabur	Kebridehar	Godey	Hargeisa	Berbera	Lawyacado
Northeastern Ethiopia	Semera	.	0.05	0.04	0.07	0.10	0.07	0.08	0.12	0.11	0.10	0.12	0.09	0.06
	Bati	0.05	.	0.08	0.10	0.12	0.08	0.10	0.15	0.14	0.13	0.14	0.12	0.08
	Gewane	0.04	0.08	.	0.06	0.09	0.06	0.08	0.11	0.11	0.10	0.12	0.08	0.06
	Awash Sebat Kilo	0.07	0.10	0.06	.	0.10	0.07	0.08	0.11	0.10	0.10	0.12	0.09	0.07
East-central Ethiopia	Erer Gota	0.10	0.12	0.09	0.10	.	0.05	0.08	0.12	0.11	0.12	0.13	0.11	0.09
	Dire Dawa	0.07	0.08	0.06	0.07	0.05	.	0.04	0.09	0.08	0.08	0.09	0.08	0.05
	Jigjiga	0.08	0.10	0.08	0.08	0.08	0.04	.	0.08	0.08	0.08	0.09	0.09	0.07
Southeastern Ethiopia	Degehabur	0.12	0.15	0.11	0.11	0.12	0.09	0.08	.	0.04	0.05	0.11	0.12	0.11
	Kebridehar	0.11	0.14	0.11	0.10	0.11	0.08	0.08	0.04	.	0.04	0.11	0.11	0.10
	Godey	0.10	0.13	0.10	0.10	0.12	0.08	0.08	0.05	0.04	.	0.12	0.12	0.10
Somaliland	Hargeisa	0.12	0.14	0.12	0.12	0.13	0.09	0.09	0.11	0.11	0.12	.	0.07	0.09
	Berbera	0.09	0.12	0.08	0.09	0.11	0.08	0.09	0.12	0.11	0.12	0.07	.	0.08
	Lawyacado	0.06	0.08	0.06	0.07	0.09	0.05	0.07	0.11	0.10	0.10	0.09	0.08	.

FIGURE 3 Population pairwise F_{st} based on autosomal loci. Colour gradient based on F_{st} values (yellow = highest, dark blue = lowest).

on five polymorphic sites (Table S2). The most common haplotype was Haplotype 2 observed in Djibouti, Somaliland and Ethiopia (Figure 5). The highest haplotype diversity was observed in Lawyacado, Semera and Djibouti ($H_d > 0.620$) and the lowest was in Erer Gota, Godey and Kebridehar ($H_d < 0.100$) (Table S2). We also calculated pairwise F_{st} based on genetic distances to determine the relationship between the major subregions in the HoA. The highest average F_{st} values were for Bati (0.582), Kebridehar (0.387) and Lawyacado (0.371), and the lowest average F_{st} was for Berbera (0.110), Dire Dawa (0.174) and Degehabur (0.214) (Table S3). A phylogeographic tree was created using *A. maculatus* as an outgroup. Overall, the tree provides low-to-moderate bootstrap (bs) support for the nodes (max bs = 57, excluding outgroup; Figure S6). In the tree, haplotypes 1, 3 and 4, predominately found in northeastern and east-central Ethiopia, Djibouti and

Somaliland, all cluster together with moderate-to-low bootstrap support (bs = 57) (Figure S6). Within this cluster, Hap 1 is basal to Hap 3 and Hap 4 and another rarer haplotype found in Djibouti. This cluster is sister to Hap 2, the most prevalent haplotype and the one with the broadest geographic distribution (Figure S6).

3.4 | Genotype–environment association

Overall, GF models were not found to be significant compared with randomized models (SNPs with positive R^2 value: $N = 153$; mean R^2 value: $N = 0.160$) (Figure S8), suggesting that genotypic turnover does not seem to be strongly associated with the environment (Figures S7A,B). However, GF still recovered 137 SNPs (out of 1680, 8.2%) with a positive GF goodness of fit R^2 value,
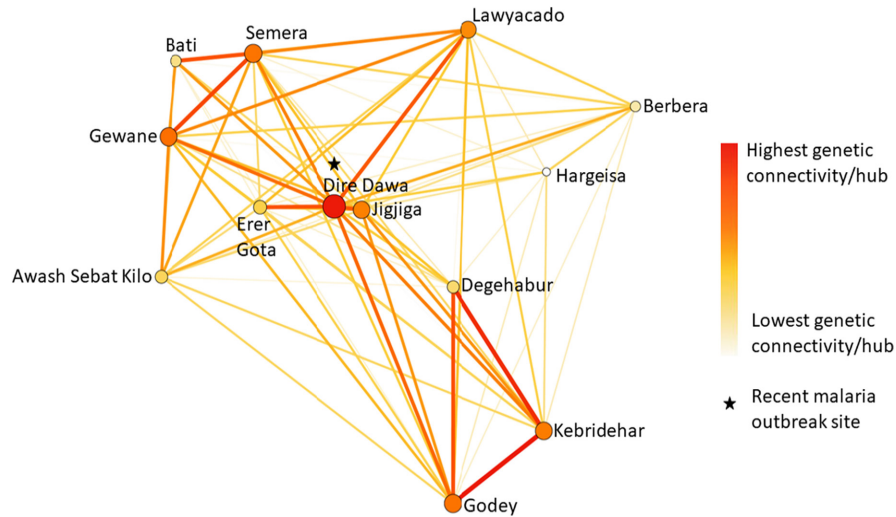
FIGURE 4 Genetic network of *Anopheles stephensi* populations in eastern Ethiopia and Somaliland. Network nodes represent populations/hubs and links represent weighted genetic distances/genetic connectivities. The figure is produced by EDENetworks based on a genotype autosomal ddRAD-seq loci matrix by applying a single realization of bootstrapping with 0.85 percentage of nodes at each location and thresholded at 0.15. The colours and sizes of the links represent the strength of genetic connectivity from lowest (white) to highest (red). The colours and sizes of the nodes represent the cumulative weighted links from lowest (white) to highest (red). The sample sites are Erer Gota, Dire Dawa, Jigjiga, Berbera, Hargeisa, Lawyacado, Awash Sebat Kilo, Gewane, Semera, Bati, Degehabur, Kebriderar and Godey.
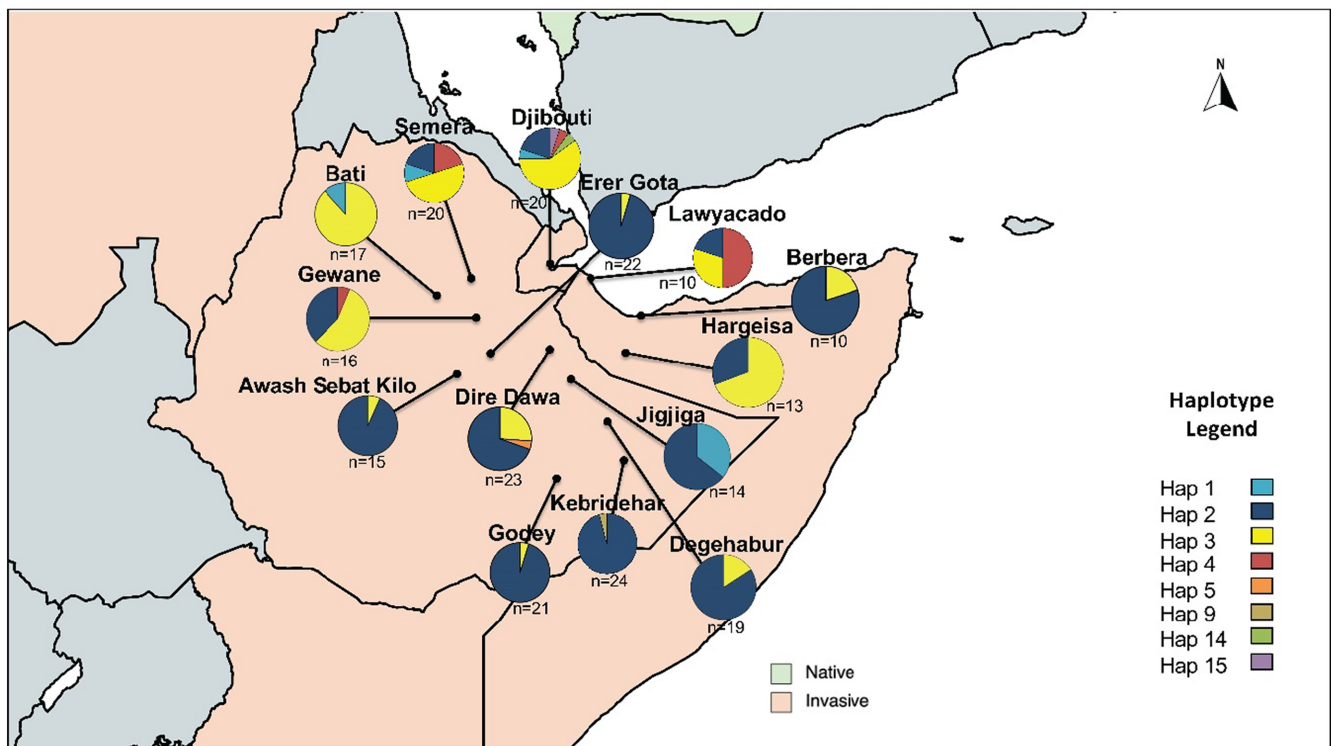
FIGURE 5 Map of HoA *Anopheles stephensi* COI haplotypes frequencies.

representing a number of alleles that are associated with the environment. While three of the top five environmental predictors were related to seasonal changes such as precipitation of coldest quarter (Bio19), temperature seasonality (Bio4) and normalized difference vegetation index (NDVI.April) (Figures S7C and S9), the strongest predictive variable in the region was the mean diurnal range (Bio2; Figures S7C and S9). Finally, while the GF PCA showed a minimal signal of local adaptation (Figure S7B), there is clear genotypic turnover between Somaliland and east-central and south-eastern Ethiopia (Figure S7A).

4 | DISCUSSION

4.1 | *Anopheles stephensi* population structure reveals dispersal patterns within the Horn of Africa

We provide the most complete analysis of the population structure for the recently invaded and currently expanding *A. stephensi* in the HoA. First, our data showed an *A. stephensi* population structure that appears to be consistent with an isolation-by-distance pattern (Figure S3) with various levels of genetic connectivity, including across international borders (Figure 1b; Figure S1). This observed population structure seems to be driven by about 91 outlier SNPs potentially due to Bottlenecks or any kind of selection (Figure S4, Table S1). Additionally, the highest level of nuclear diversity was observed closest to Djibouti, where *A. stephensi* was first detected in the HoA [i.e. Lawyacado (Somaliland) and Semera (Northeastern Ethiopia)] or at transportation hubs [i.e. Awash Sebat Kilo (North-eastern Ethiopia) and Dire Dawa (east-central Ethiopia)] (Figure 1a, Table 2). This pattern of diversity indicates these sites contain older populations. These findings support *A. stephensi* populations dispersal patterns from north central to all the other regions in the HoA, suggesting that north central HoA populations are more likely to be the source populations leading to subsequent founding populations moving outward, with resulting new populations characterized by low genetic variation (Howe et al., 1991). Indeed, southeastern *A. stephensi* populations were differentiated from Somaliland, northeastern and east-central Ethiopia *A. stephensi* populations (Figures 1b,c and 3). This finding coupled with the low genetic diversity observed in these populations (Table 2) suggests a potential recent founder event such as a Bottleneck for the southeastern Ethiopian *A. stephensi* populations compared with other studied populations (Haddrill et al., 2005; Nei et al., 1975). This scenario is further supported by the co-ancestry matrix model, which revealed a separate dendrogram branch for southeastern Ethiopia *A. stephensi* populations, but a shared one for *A. stephensi* populations in Somaliland, northeastern and east-central Ethiopia regions (Figure 2).

Furthermore, the discovery of Dire Dawa as a significant hub (Figure 4; Figure S5) coupled with the isolation-by-distance population structure pattern suggests the potential role of roads and human transportation in the dispersal of *A. stephensi* from one site to another. For instance, Dire Dawa which was reported as the site of a recent malaria outbreak associated with *A. stephensi* (Tadesse et al., 2023) (preprint) had a high genetic connectivity with Lawyacado (SL), which is 21.6km away from Djibouti City (Figure 4). Thus, this high genetic connectivity observed between Dire Dawa and Lawyacado despite an isolation-by-distance population structure pattern could potentially be explained by the fact that there is a direct train route between Djibouti City and Dire Dawa. Overall, the unbalanced sample size between the Ethiopian and Somaliland studied sites posed a limitation. However, the number of independent biallelic SNPs recovered helped our analyses remain informative. Thus, future studies will need to test the potential impact of roads and human transportation on the dispersal of *A. stephensi* within the HoA.

4.2 | Updated *Anopheles stephensi* mitochondrial DNA structure confirms patterns of dispersal into the Horn of Africa

Analysing HoA *A. stephensi* mtDNA COI gene provided further support for the observed nuclear population structure and dispersal patterns of *A. stephensi* into the HoA region. Our findings showed distinct genetic clusters and levels of mtDNA COI genetic variations among the studied *A. stephensi* populations in the HoA region that suggest a radiative dispersal pattern from north central to the other regions (Figure 5; Figure S6). Specifically, Djibouti City *A. stephensi* had the most genetic variation, followed by *A. stephensi* from the closest sites such as Lawyacado and Semera, in Somaliland and Ethiopia, respectively (Figure 5; Table S2). Additionally, the low mitochondrial COI diversity observed in *A. stephensi* populations from the east-central and southeastern Ethiopian regions, coupled with the fact that one or two of the prevalent Djiboutian COI haplotypes were found in every site, again supports potential sequential founder events of *A. stephensi* from Djibouti City outward in the HoA (Figure 5; Figure S6 and Table S2). Again, this mtDNA COI population structure leading to the observed dispersal patterns could potentially be explained by the fact that Ethiopia, a landlocked country, relies on seaports in Djibouti City and Berbera for goods imports and exports. The role that maritime sea transports and port cities play in invasive mosquito species incursion in new geographic areas has been well documented (Ahn et al., 2023; Ammar et al., 2019; Hulme, 2009; Swan et al., 2022). Thus, future work will require the testing of this hypothesis for the movement of the invasive *A. stephensi* populations by a simultaneous and continued sampling of port cities and inland cities across the region to decipher the potential role of seaports on the introduction, dispersal and genetic variation of *A. stephensi* in the HoA.

4.3 | Minimal environmental association with *Anopheles stephensi* genetic variation

Our genotype–environment association analysis revealed a minimal environmental association with genetic turnover in eastern Ethiopia and Somaliland (Figures S7 and S8). These results emphasize the more impactful role that genetic drift is currently having on genetic diversity in the HoA, as *A. stephensi* continues to expand. In general, the weak predictive power of GF is likely due to the recent establishment of *A. stephensi*, meaning that selective pressures have not had sufficient time to act throughout the genome in a way that would be detected from our relatively low genome coverage (i.e. ~0.13%). However, other studies have identified rapid changes in *A. stephensi* going from primarily seasonal to year-round activity that could be the result of strong selective forces (Seyfarth et al., 2019; Sinka et al., 2020; Whittaker et al., 2023), such as the annual temperature variables GF identified as high ranking in predictive power (Figures S7C and S9). These kinds of temperature variations were also reported to impact *A. stephensi* activity and survival in lab settings (Paaijmans

et al., 2010, 2013). Therefore, while we do not see a strong signal in our study, tracking the signal over time could be beneficial in understanding if local adaptation is occurring. Future work will require continued sampling and full genome analyses to better understand how genetic diversity continues to respond to environmental pressures in the HoA.

4.4 | Potential *Anopheles stephensi* expansion controls and future malaria outbreaks

Our findings have significant implications for *A. stephensi* vector control in the HoA. First, the population structure and dispersal patterns could inform strategies to manage the current invasion and prevent the further spread of *A. stephensi* within the HoA.

Also, the finding of Dire Dawa, a recent *A. stephensi*-associated malaria outbreak site, as a significant hub may aid in predicting/preventing other malaria outbreaks within the region. Therefore, we highly recommend heightening vector control surveillance in the observed significant hub sites (i.e. Dire Dawa, Godey, Semera, Jigjiga, Kebridehar, Gewane and Lawyacado) to prevent potential future malaria outbreaks. Also, a genotype–environment association model with a longitudinal nuclear SNP dataset could help monitor and determine critical environmental and climate variables contributing to establishing this invasive *A. stephensi* population, as ecological variables are known factors influencing the development and vectorial capacity of established insect vectors (Alto et al., 2017; Holmes & Benoit, 2019; Paaijmans et al., 2009; Zapletal et al., 2018). Thus, incorporating both genomic and environmental analyses into future *A. stephensi* vector surveillance could significantly aid in designing timely and targeted vector control strategies against this malaria vector in the HoA and beyond.

5 | CONCLUSION AND FUTURE DIRECTIONS

Overall, our present study answered many questions regarding the genetic population structure and dispersal patterns of *A. stephensi* in the HoA. Genetic drift via sequential founder events, rather than adaptation to local environmental conditions, seems to be the main evolutionary force currently impacting *A. stephensi* genetic variation in the HoA though we posit that natural selectivity processes will begin to take shape as this mosquito further establishes in the region. Our study also provides a tool to predict and prevent potential malaria outbreaks in a timely manner by identifying important *A. stephensi* hubs in the region. However, future studies could benefit from longitudinal and more extensive *A. stephensi* sampling across the HoA. Finally, with our genomic data being the only African *A. stephensi* SNPs dataset, we recommend the incorporation of genome-wide multilocus data into future studies of African *A. stephensi* for higher resolution comparative analyses between populations and

better prediction of *A. stephensi* dispersal within the region and beyond.

CONFLICT OF INTEREST STATEMENT

All authors declare that they have no conflicts of interest.

DATA AVAILABILITY STATEMENT

DNA sequences: Bioproject PRJNA888109 and Genbank accessions SAMN31227709–SAMN31227956.

DISCLAIMER

The authors alone are responsible for the views expressed in this article, and they do not necessarily represent the views, decisions or policies of the institutions with which they are affiliated.

ORCID

Jeanne N. Samake https://orcid.org/0000-0003-2391-6583
Philip Lavretsky https://orcid.org/0000-0002-5904-8821
Isuru Gunarathna https://orcid.org/0000-0001-9729-2879
Madison Follis https://orcid.org/0009-0000-6676-9830
Joshua I. Brown https://orcid.org/0000-0001-6299-9815
Tamar E. Carter https://orcid.org/0000-0002-9438-9541

REFERENCES

Ahmed, A., Khogali, R., Elnour, M.-A. B., Nakao, R., & Salim, B. (2021). Emergence of the invasive malaria vector *Anopheles stephensi* in Khartoum State, Central Sudan. *Parasites & Vectors*, 14(1), 511. https://doi.org/10.1186/s13071-021-05026-4

Ahn, J., Sinka, M., Irish, S., & Zohdy, S. (2023). Modeling marine cargo traffic to identify countries in Africa with greatest risk of invasion by *Anopheles stephensi*. *Scientific Reports*, 13(1), 876. https://doi.org/10.1038/s41598-023-27439-0

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9), 1655–1664. https://doi.org/10.1101/gr.094052.109

Ali, S., Samake, J. N., Spear, J., & Carter, T. E. (2022). Morphological identification and genetic characterization of *Anopheles stephensi* in Somaliland. *Parasites & Vectors*, 15(1), 247. https://doi.org/10.1186/s13071-022-05339-y

Alto, B. W., Wiggins, K., Eastmond, B., Ortiz, S., Zirbel, K., & Lounibos, L. P. (2017). Diurnal temperature range and chikungunya virus infection in invasive mosquito vectors. *Journal of Medical Entomology*, 55(1), 217–224. https://doi.org/10.1093/jme/tjx182

Ammar, S. E., McLntyre, M., Swan, T., Kasper, J., Derraik, J. G. B., Baker, M. G., & Hales, S. (2019). Intercepted mosquitoes at New Zealand's ports of entry, 2001 to 2018: Current status and future concerns. *Tropical Medicine and Infectious Disease*, 4(3), 101. https://doi.org/10.3390/tropicalmed4030101

Balabaskaran Nina, P., Mohanty, A. K., Ballav, S., Vernekar, S., Bhinge, S., D'souza, M., Walke, J., Manoharan, S. K., Mascarenhas, A., Gomes, E., Chery, L., Valecha, N., Kumar, A., & Rathod, P. K. (2017). Dynamics of plasmodium vivax sporogony in wild *Anopheles stephensi* in a malaria-endemic region of Western India. *Malaria Journal*, 16(1), 284. https://doi.org/10.1186/s12936-017-1931-8

Balkew, M., Mumba, P., Dengela, D., Yohannes, G., Getachew, D., Yared, S., Chibsa, S., Murphy, M., George, K., Lopez, K., Janies, D., Choi, S. H., Spear, J., Irish, S. R., & Carter, T. E. (2020). Geographical distribution of *Anopheles stephensi* in eastern Ethiopia. *Parasites & Vectors*, 13(1), 35. https://doi.org/10.1186/s13071-020-3904-y

Bay, R. A., Harrigan, R. J., Underwood, V. L., Gibbs, H. L., Smith, T. B., & Ruegg, K. (2018). Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science*, 359(6371), 83–86. https://doi.org/10.1126/science.aan4380

Bennett, K. L., McMillan, W. O., & Loaiza, J. R. (2021). The genomic signal of local environmental adaptation in *Aedes aegypti* mosquitoes. *Evolutionary Applications*, 14(5), 1301–1313. https://doi.org/10.1111/eva.13199

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Brown, J. I., Harrigan, R. J., & Lavretsky, P. (2022). Evolutionary and ecological drivers of local adaptation and speciation in a North American avian species complex. *Molecular Ecology*, 31(9), 2578–2593. https://doi.org/10.1111/mec.16423

Carter, T. E., Yared, S., Gebresilassie, A., Bonnell, V., Damodaran, L., Lopez, K., Ibrahim, M., Mohammed, S., & Janies, D. (2018). First detection of *Anopheles stephensi* Liston, 1901 (Diptera: Culicidae) in Ethiopia using molecular and morphological approaches. *Acta Tropica*, 188, 180–186. https://doi.org/10.1016/j.actatropica.2018.09.001

Carter, T. E., Yared, S., Getachew, D., Spear, J., Choi, S. H., Samake, J. N., Mumba, P., Dengela, D., Yohannes, G., Chibsa, S., Murphy, M., Dissanayake, G., Flately, C., Lopez, K., Janies, D., Zohdy, S., Irish, S. R., & Balkew, M. (2021). Genetic diversity of *Anopheles stephensi* in Ethiopia provides insight into patterns of spread. *Parasites & Vectors*, 14(1), 602. https://doi.org/10.1186/s13071-021-05097-3

Chakraborty, M., Ramaiah, A., Adolfi, A., Halas, P., Kaduskar, B., Ngo, L. T., Jayaprasad, S., Paul, K., Whadgar, S., Srinivasan, S., Subramani, S., Bier, E., James, A. A., & Emerson, J. J. (2022). Author correction: Hidden genomic features of an invasive malaria vector, *Anopheles stephensi*, revealed by a chromosome-level genome assembly. *BMC Biology*, 20(1), 96. https://doi.org/10.1186/s12915-022-01314-2

DaCosta, J. M., & Sorenson, M. D. (2014). Amplification biases and consistent recovery of loci in a double-digest RAD-seq protocol. *PLoS One*, 9(9), e106713. https://doi.org/10.1371/journal.pone.0106713

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & Group, G. P. A. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

de Santi, V. P., Khaireh, B. A., Chiniard, T., Pradines, B., Taudon, N., Larréché, S., Mohamed, A. B., de Laval, F., Berger, F., Gala, F., Mokrane, M., Benoit, N., Malan, L., Abdi, A. A., & Briolant, S. (2021). Role of *Anopheles stephensi* mosquitoes in malaria outbreak, Djibouti, 2019. *Emerging Infectious Diseases*, 27(6), 1697–1700. https://doi.org/10.3201/eid2706.204557

Ellis, N., Smith, S. J., & Pitcher, C. R. (2012). Gradient forests: Calculating importance gradients on physical predictors. *Ecology*, 93(1), 156–168. https://doi.org/10.1890/11-0252.1

Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10(3), 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Faulde, M. K., Rueda, L. M., & Khaireh, B. A. (2014). First record of the Asian malaria vector *Anopheles stephensi* and its possible role in the resurgence of malaria in Djibouti, horn of Africa. *Acta Tropica*, 139, 39–43. https://doi.org/10.1016/j.actatropica.2014.06.016

Fitzpatrick, M. C., & Keller, S. R. (2015). Ecological genomics meets community-level modelling of biodiversity: Mapping the genomic landscape of current and future environmental adaptation. *Ecology Letters*, 18(1), 1–16. https://doi.org/10.1111/ele.12376

Haddrill, P. R., Thornton, K. R., Charlesworth, B., & Andolfatto, P. (2005). Multilocus patterns of nucleotide variability and the demographic and selection history of *Drosophila melanogaster* populations. *Genome Research*, 15(6), 790–799. https://doi.org/10.1101/gr.3541005

Hague, M. T. J., & Routman, E. J. (2016). Does population size affect genetic diversity? A test with sympatric lizard species. *Heredity*, 116(1), 92–98. https://doi.org/10.1038/hdy.2015.76

Hamlet, A., Dengela, D., Tongren, J. E., Tadesse, F. G., Bousema, T., Sinka, M., Seyoum, A., Irish, S. R., Armistead, J. S., & Churcher, T. (2022). The potential impact of *Anopheles stephensi* establishment on the transmission of *Plasmodium falciparum* in Ethiopia and prospective control measures. *BMC Medicine*, 20(1), 135. https://doi.org/10.1186/s12916-022-02324-1

Hay, S. I., Guerra, C. A., Tatem, A. J., Atkinson, P. M., & Snow, R. W. (2005). Urbanization, malaria transmission and disease burden in Africa. *Nature Reviews. Microbiology*, 3(1), 81–90. https://doi.org/10.1038/nrmicro1069

Hernández, F., Brown, J. I., Kaminski, M., Harvey, M. G., & Lavretsky, P. (2021). Genomic evidence for rare hybridization and large demographic changes in the evolutionary histories of four North American dove species. *Animals*, 11(9), 2677.

Holmes, C. J., & Benoit, J. B. (2019). Biological adaptations associated with dehydration in mosquitoes. *Insects*, 10(11), 375. https://doi.org/10.3390/insects10110375

Howe, R. W., Davis, G. J., & Mosca, V. (1991). The demographic significance of 'sink' populations. *Biological Conservation*, 57(3), 239–255. https://doi.org/10.1016/0006-3207(91)90071-G

Hulme, P. E. (2009). Trade, transport and trouble: Managing invasive species pathways in an era of globalization. *Journal of Applied Ecology*, 46(1), 10–18. https://doi.org/10.1111/j.1365-2664.2008.01600.x

Imwong, M., Nakeesathit, S., Day, N. P. J., & White, N. J. (2011). A review of mixed malaria species infections in

anopheline mosquitoes. *Malaria Journal*, 10(1), 253. https://doi.org/10.1186/1475-2875-10-253

Ishtiaq, F., Swain, S., & Kumar, S. S. (2021). *Anopheles stephensi* (Asian malaria mosquito). *Trends in Parasitology*, 37(6), 571–572. https://doi.org/10.1016/j.pt.2021.03.009

Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., & Kessler, M. (2017). Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, 4(1), 170122. https://doi.org/10.1038/sdata.2017.122

Karger, D. N., Nobis, M. P., Normand, S., Graham, C. H., & Zimmermann, N. E. (2021). CHELSA-TraCE21k v1.0. Downscaled transient temperature and precipitation data since the last glacial maximum. *Climate of the Past Discussions*, 2021, 1–27. https://doi.org/10.5194/cp-2021-30

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kimura, M., & Ohta, T. (1973). The age of a neutral mutant persisting in a finite population. *Genetics*, 75(1), 199–212. https://doi.org/10.1093/genetics/75.1.199

Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y., & Porter, M. A. (2014). Multilayer networks. *Journal of Complex Networks*, 2(3), 203–271. https://doi.org/10.1093/comnet/cnu016

Kivelä, M., Arnaud-Haond, S., & Saramäki, J. (2015). EDENetworks: A user-friendly software to build and analyse networks in biogeography, ecology and population genetics. *Molecular Ecology Resources*, 15(1), 117–122. https://doi.org/10.1111/1755-0998.12290

Lambrechts, L., Chavatte, J. M., Snounou, G., & Koella, J. C. (2006). Environmental influence on the genetic basis of mosquito resistance to malaria parasites. *Proceedings of the Biological Sciences*, 273(1593), 1501–1506. https://doi.org/10.1098/rspb.2006.3483

Lavretsky, P., Dacosta, J. M., Hernández-Baños, B. E., Engilis, A., Jr., Sorenson, M. D., & Peters, J. L. (2015). Speciation genomics and a role for the Z chromosome in the early stages of divergence between Mexican ducks and mallards. *Molecular Ecology*, 24(21), 5364–5378. https://doi.org/10.1111/mec.13402

Lavretsky, P., McInerney, N. R., Mohl, J. E., Brown, J. I., James, H. F., McCracken, K. G., & Fleischer, R. C. (2020). Assessing changes in genomic divergence following a century of human-mediated secondary contact among wild and captive-bred ducks. *Molecular Ecology*, 29(3), 578–595. https://doi.org/10.1111/mec.15343

Lerner, K. L. (2001). *Evolution and evolutionary mechanisms controlling genetic variation. Draft copy. Part of a series of essays identifying and explaining theories essential to understanding modern scientific thought. Updated 2010*. Originally Published.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, 27(21), 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, 475(7357), 493–496. https://doi.org/10.1038/nature10231

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Malinsky, M., Trucchi, E., Lawson, D. J., & Falush, D. (2018). RADpainter and fineRADstructure: Population inference from RADseq data. *Molecular Biology and Evolution*, 35(5), 1284–1290. https://doi.org/10.1093/molbev/msy023

Mathanga, D. P., Tembo, A. K., Mzilahowa, T., Bauleni, A., Mtimaukenena, K., Taylor, T. E., Valim, C., Walker, E. D., & Wilson, M. L. (2016). Patterns and determinants of malaria risk in urban and peri-urban areas of Blantyre, Malawi. *Malaria Journal*, 15(1), 590. https://doi.org/10.1186/s12936-016-1623-9

Nei, M., Maruyama, T., & Chakraborty, R. (1975). The bottleneck effect and genetic variability in populations. *Evolution*, 29(1), 1–10. https://doi.org/10.1111/j.1558-5646.1975.tb00807.x

Paaijmans, K. P., Blanford, S., Bell, A. S., Blanford, J. I., Read, A. F., & Thomas, M. B. (2010). Influence of climate on malaria transmission depends on daily temperature variation. *Proceedings of the National Academy of Sciences of the United States of America*, 107(34), 15135–15139. https://doi.org/10.1073/pnas.1006422107

Paaijmans, K. P., Heinig, R. L., Seliga, R. A., Blanford, J. I., Blanford, S., Murdock, C. C., & Thomas, M. B. (2013). Temperature variation makes ectotherms more sensitive to climate change. *Global Change Biology*, 19(8), 2373–2380. https://doi.org/10.1111/gcb.12240

Paaijmans, K. P., Read, A. F., & Thomas, M. B. (2009). Understanding the link between malaria risk and climate. *Proceedings of the National Academy of Sciences of the United States of America*, 106(33), 13844–13849. https://doi.org/10.1073/pnas.0903423106

Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E., & Lercher, M. J. (2014). PopGenome: An efficient Swiss army knife for population genomic analyses in R. *Molecular Biology and Evolution*, 31(7), 1929–1936. https://doi.org/10.1093/molbev/msu136

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81(3), 559–575. https://doi.org/10.1086/519795

Reynolds, J., Weir, B. S., & Cockerham, C. C. (1983). Estimation of the coancestry coefficient: Basis for a short-term genetic distance. *Genetics*, 105(3), 767–779. https://doi.org/10.1093/genetics/105.3.767

Rozas, J., Sánchez-DelBarrio, J. C., Messeguer, X., & Rozas, R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, 19(18), 2496–2497. https://doi.org/10.1093/bioinformatics/btg359

Schmidt, T. L., Swan, T., Chung, J., Karl, S., Demok, S., Yang, Q., Field, M. A., Muzari, M. O., Ehlers, G., Brugh, M., Bellwood, R., Horne, P., Burkot, T. R., Ritchie, S., & Hoffmann, A. A. (2021). Spatial population genomics of a recent mosquito invasion. *Molecular Ecology*, 30(5), 1174–1189. https://doi.org/10.1111/mec.15792

Seyfarth, M., Khaireh, B. A., Abdi, A. A., Bouh, S. M., & Faulde, M. K. (2019). Five years following first detection of *Anopheles stephensi* (Diptera: Culicidae) in Djibouti, horn of Africa: Populations established-malaria emerging. *Parasitology Research*, 118(3), 725–732. https://doi.org/10.1007/s00436-019-06213-0

Sherpa, S., Blum, M. G. B., Capblancq, T., Cumer, T., Rioux, D., & Després, L. (2019). Unravelling the invasion history of the Asian tiger mosquito in Europe. *Molecular Ecology*, 28(9), 2360–2377. https://doi.org/10.1111/mec.15071

Sinka, M. E., Pironon, S., Massey, N. C., Longbottom, J., Hemingway, J., Moyes, C. L., & Willis, K. J. (2020). A new malaria vector in Africa: Predicting the expansion range of *Anopheles stephensi* and identifying the urban populations at risk. *Proceedings of the National Academy of Sciences of the United States of America*, 117(40), 24900–24908. https://doi.org/10.1073/pnas.2003976117

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Swan, T., Russell, T. L., Staunton, K. M., Field, M. A., Ritchie, S. A., & Burkot, T. R. (2022). A literature review of dispersal pathways of *Aedes albopictus* across different spatial scales: Implications for vector surveillance. *Parasites & Vectors*, 15(1), 303. https://doi.org/10.1186/s13071-022-05413-5

Tadesse, F., Emiru, T., Getachew, D., Murphy, M., Sedda, L., Ejigu, L., Bulto, M., Byrne, I., Demisse, M., Abdo, M., Chali, W., Elliot, A., Vickers, E., Diaz, A., Alemayehu, L., Behaksera, S., Jebessa, G., Dinka, H., Tsegaye, T., ... Bousema, T. (2023). *Anopheles stephensi* is implicated in an outbreak of *Plasmodium falciparum* parasites that carry markers of drug and diagnostic resistance in Dire Dawa City, Ethiopia, January–July 2022. *Research Square Platform LLC*. https://doi.org/10.21203/rs.3.rs-2847814/v1

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*(3), 585–595. https://doi.org/10.1093/genetics/123.3.585

Thomas, S., Ravishankaran, S., Justin, N. A. J. A., Asokan, A., Mathai, M. T., Valecha, N., Montgomery, J., Thomas, M. B., & Eapen, A. (2017). Resting and feeding preferences of *Anopheles stephensi* in an urban setting, perennial for malaria. *Malaria Journal*, *16*(1), 111. https://doi.org/10.1186/s12936-017-1764-5

Whittaker, C., Hamlet, A., Sherrard-Smith, E., Winskill, P., Cuomo-Dannenburg, G., Walker, P. G. T., Sinka, M., Pironon, S., Kumar, A., Ghani, A., Bhatt, S., & Churcher, T. S. (2023). Seasonal dynamics of *Anopheles stephensi* and its implications for mosquito detection and emergent malaria control in the Horn of Africa. *Proceedings of the National Academy of Sciences of the United States of America*, *120*(8), e2216142120. https://doi.org/10.1073/pnas.2216142120

WHO. (2019). *Vector alert: Anopheles stephensi invasion and spread: Horn of Africa, the Republic of the Sudan and surrounding geographical areas, and Sri Lanka: information note.* https://www.who.int/publications/i/item/WHO-HTM-GMP-2019.09

WHO. (2021). *World malaria report 2021*. World Health Organization. https://www.who.int/publications/i/item/9789240040496

WHO. (2022a). *2022*. World Health Organization. https://apps.who.int/malaria/maps/threats/?theme=invasive&mapType=invasive

WHO. (2022b). *WHO launches new initiative to stop the spread of invasive malaria vector in Africa.* https://www.who.int/news-room/feature-stories/detail/mosquito-on-the-move

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis.* Springer-Verlag. https://doi.org/10.1007/978-3-319-24277-4

Zapletal, J., Erraguntla, M., Adelman, Z. N., Myles, K. M., & Lawley, M. A. (2018). Impacts of diurnal temperature and larval density on aquatic development of Aedes aegypti. *PLoS One*, *13*(3), e0194025. https://doi.org/10.1371/journal.pone.0194025

Zhou, H., Alexander, D., & Lange, K. (2011). A quasi-Newton acceleration for high-dimensional optimization algorithms. *Statistics and Computing*, *21*(2), 261–273. https://doi.org/10.1007/s11222-009-9166-3

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.