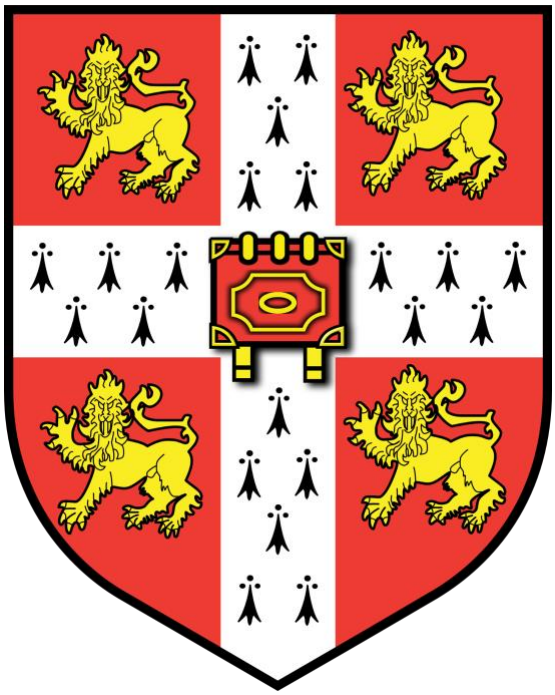
Application of the ANOSPP Amplicon Sequencing Panel to Assess *Anopheles* Mosquito Diversity, Population Structure, and Distribution in Tanzania



University of Cambridge

Wolfson College

Wellcome Sanger Institute

This dissertation is submitted for the degree of Master of Philosophy

Maneno Edson Baravuga

31st August, 2025

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specified in the text. It is not substantially the same as any work that has already been submitted, or is being concurrently submitted, for any degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee.

Author Contribution Statement

This thesis is based on original research and analyses undertaken by the candidate. Most of the data are derived from a major malaria-vector research project in Tanzania conducted prior to my MPhil candidature, supplemented with datasets from collaborators and publicly available repositories, as outlined below.

Mosquito field collections in Tanzania for the main dataset were carried out under the project "Integrating intervention-targetable behaviours of malaria vectors to optimize intervention selection and impact in Tanzania (ITB)", led by Dr Nicodemus Govella (Ifakara Health Institute, Tanzania), in which I contributed to field data collection, leading field teams, engaging communities and local authorities, and overseeing logistics across 25 districts (December 2020 – December 2023). These field activities occurred before the MPhil period and therefore fall outside this thesis.

Additional *Anopheles arabiensis* samples from Tanzania used in the genetic analyses were obtained from a separate study led by Mr Deogratius Kavishe (Ifakara Health Institute, Tanzania). Further *An. arabiensis* datasets from Madagascar, Uganda, Democratic Republic of Congo, Nigeria, Senegal, Ghana, and Burkina Faso were accessed through collaborating partners within the ANOSPP project led by Dr Mara Lawniczak (Wellcome Sanger Institute). I did not participate in the collection of these samples.

For ecological niche partitioning and species distribution modelling, environmental and socio-ecological covariates (climatic, landcover, human-population, and livestock-population data) were obtained from publicly available remote-sensing repositories which are well cited. Primary mosquito-occurrence data were largely derived from the ITB project. Additional records were sourced from the Tanzania Malaria Case Management Programme (Dhibiti Malaria, led by Dr Yeromin Mlacha), where I contributed to data collection, coordination, logistics, local government authorities and community engagement, and field-team training across 22 sentinel districts (insecticide-resistance monitoring) and three regions (*Anopheles stephensi* surveillance) between October 2023 and September 2024; these activities also occurred outside the MPhil period. Additional Secondary presence data were obtained from WHO malaria threat-map metadata, published studies (with consultation from authors), and records shared by Dr Fredros Okumu (Ifakara Health Institute), Mr Deogratius Kavishe

(Ifakara Health Institute), and Mr Bernad Edmund (National Institute for Medical Research, Tanzania).

All laboratory procedures, including DNA extraction, PCR, amplicon sequencing, preprocessing, and species-identification pipelines for mosquitoes and *Plasmodium* (developed by Dr Alex Makunin, Dr Petra Korlevic, and Dr Marilou Bodde), were completed before this thesis and are not part of it. Although I received training to understand these workflows and later performed some hands-on work, the laboratory work associated with the data used here was carried out prior to my candidature.

No analysis or interpretation of these datasets existed prior to my candidature. All bioinformatic processing (including QC and a custom R-based variant-calling pipeline developed by me), dataset integration, population-genetic analyses, ecological modelling, spatial analyses, species distribution modelling, and all other analyses and interpretation, as well as thesis writing, were conducted independently by me.

Abstract

Malaria continues to pose a major public health burden in sub-Saharan Africa, where the ecology, behaviour, and genetic structure of mosquito vectors govern transmission dynamics and shape the effectiveness of control strategies. Traditional surveillance approaches, constrained by morphology-based identification, narrow geographic scope, and limited resolution, are insufficient to capture the complexity of vector populations and their interactions with parasites and environments.

This thesis addresses these gaps by combining ecological surveys, species distribution modelling, and genomic analyses to deliver a multi-dimensional view of malaria vectors in Tanzania and their broader African context. Ecological surveys across 25 districts revealed marked variation in *Anopheles* community composition, ecological niche partitioning, and species co-occurrence, alongside the distribution of *Plasmodium* parasites in nine districts where all four major human malaria species were detected. Use of the ANOSPP panel substantially improved taxonomic resolution over morphology, enabling simultaneous identification of multiple *Anopheles* vectors and their associated *Plasmodium* species, and yielded the first report of *Plasmodium caprae* in *An. arabiensis* in Tanzania.

Species distribution models developed for *An. arabiensis*, *An. gambiae s.s* and *An. funestus s.s* incorporated climate and land cover predictors at 1-km scale, producing high-resolution habitat suitability maps that aligned with known transmission zones while identifying species-specific ecological associations. These models provide a predictive framework to support targeted surveillance and resource allocation in vector control programs.

Genomic analyses of *Anopheles arabiensis* populations from Tanzania revealed broad genetic connectivity overall, but with significant isolation by distance. At a continental scale, three clusters emerged: Eastern-Central Africa, Western Africa, and Madagascar, with patterns reflecting isolation by distance, climatic differences, and resistance surfaces shaped by ecological barriers, as confirmed using Mantel and partial Mantel tests alongside the Maximum-Likelihood Population-Effects Model. Madagascar was clearly separated from the mainland by the ocean, while the Central African rainforest delineated eastern and western populations. Within West Africa, additional substructure was associated with climatic gradients. These findings suggest that the broad genetic connectivity observed across

populations could facilitate the spread of adaptive alleles, such as those conferring insecticide resistance, while localized genomic structuring likely reflects adaptation to ecological conditions, all of which may influence vector control intervention outcomes.

By integrating ecological modelling, spatial, and genomic perspectives, this thesis provides the first comprehensive overview of *Anopheles* mosquitoes distribution and genomic overview of *An. arabiensis* in Tanzania while validating ANOSPP as a scalable tool for both species' identification and genetic surveillance. The results highlight both the opportunities and risks posed by high connectivity and local adaptation, offering actionable insights for malaria control and contributing to a broader shift from reactive, coarse-grained surveillance toward predictive, multi-layered systems capable of strengthening national and regional capacity for malaria elimination.

Acknowledgement

First and foremost, I would like to express my deepest gratitude to my supervisor, Dr. Mara Lawniczak, and to my co-supervisor and advisors, Dr. Petra Korlevic, Dr. Henrik Salje, and Prof. Nicodemus Govella for their unwavering guidance, encouragement, and insightful mentorship throughout this journey. Your combined expertise and support have shaped not only the direction of this thesis but also my growth as a scientist.

I am especially grateful to Dr. Alex Makunin and Dr. Marilou Bodde, whose guidance was instrumental in developing and refining the genomic analyses presented in this work. I also wish to extend my appreciation to Dr. Gerry Ryan and Dr. Nick Golding, who played a pivotal role in equipping me with the knowledge and skills to conduct species distribution modelling. Special thanks go to Dr. Francis Totanes, whose incredible support and expertise in the wet-lab processes of the ANOSPP work were vital to the success of this project.

To my friends and colleagues in Team 222, thank you for making the office feel like a true second family, and for filling each day with camaraderie, motivation, and laughter. I am also grateful to the postgraduate department, particularly Dr. Annabel Smith and Ms. Catherine Ingle, for their consistent support from the earliest stages of my master's scholarship application through to the completion of this thesis. I would also like to acknowledge Dr. Kamil Jaron, for his constructive feedback and guidance during the critical stages of this work.

On a personal note, I owe my deepest gratitude to my wife, Fraiska, for her love, patience, and strength. You carried the responsibility of our home with grace while I was away, and I could not have achieved this without your sacrifices. To my son, Aaron, thank you for your patience in my absence, you have been a constant source of inspiration and joy. I am equally indebted to my parents and family, whose encouragement and belief in me have sustained me throughout this journey.

Finally, I dedicate this thesis to all who have walked this path with me, mentors, colleagues, friends, and family. Your belief in me, your encouragement, and your support have been the foundation of this achievement.

Acronyms and Abbreviations

Climate (Köppen–Geiger Classification)

Af Tropical rainforest climate

Am Tropical monsoon climate

Aw Tropical savanna climate

BSh Hot semi-arid steppe climate

BWh Hot desert climate

Cfa Humid subtropical climate

Cwb Temperate oceanic climate with dry winters (highland)

Cwc Subtropical highland climate with dry winters (cool)

EF Ice Cap climate

ET Tundra climate

Organisations and Institutions

FAO Food and Agriculture Organization

IHI Ifakara Health Institute

IRB Institutional Review Board

NASA National Aeronautics and Space Administration

NIMR National Institute for Medical Research

NMCP National Malaria Control Programme

WHO World Health Organization

Assays and Laboratory Methods

ANOSPP ANOpheles Species and Plasmodium Panel

BLAST Basic Local Alignment Search Tool

DADA2 Divisive Amplicon Denoising Algorithm 2

ELISA Enzyme-Linked Immunosorbent Assay

MAFFT Multiple Alignment using Fast Fourier Transform

PCR Polymerase Chain Reaction

SNP Single Nucleotide Polymorphism

VCF Variant Call Format

ELISA Enzyme Immunosorbent Assay

SNP Single Nucleotide Polymorphism

ITS2 Internal Transcribed Spacer 2

COI/II Cytochrome c Oxidase subunit one/two

Analyses and Modelling

CCA Canonical Correspondence Analysis

CSV Comma-Separated Values (data format)

DAPC Discriminant Analysis of Principal Components

DEM Digital Elevation Model

DHARMa Diagnostics for Hierarchical Regression Models (R package)

ENM Ecological Niche Modelling

FST Fixation Index

GAM Generalized Additive Model

IBD Isolation by Distance

IBE Isolation by Environment

IBR Isolation by Resistance

LD Linkage Disequilibrium

MLPE Maximum Likelihood Population Effects

TSS True Skill Statistic

AUC Area Under the Curve

NDVI Normalized Difference Vegetation Index

PCA Principal Component Analysis

r^2 Squared correlation coefficient

SDM Species Distribution Modelling

STRM Shuttle Radar Topography Mission (digital elevation data)

TWI Topographic Wetness Index

Remote Sensing and Climate Datasets

CHELSA Climatologies at High Resolution for the Earth's Land Surface Areas

MODIS Moderate Resolution Imaging Spectroradiometer

Vector Control Interventions

IRS Indoor Residual Spraying

ITN Insecticide-Treated Net

IVM Integrated Vector Management

LSM Larval Source Management

Table of Content

List of Figures

List of Tables

Chapter One: Malaria Transmission in Context: Global Patterns, Biological Complexities, and Tanzanian Challenges

1.1 Malaria Global Overview

Despite decades of investments in vector control, improved diagnostics, and antimalarial treatments, malaria remains a major global health challenge. Caused by protozoan parasites of the genus *Plasmodium* and transmitted exclusively by infected female *Anopheles* mosquitoes, the disease continues to cause high morbidity and mortality, especially in low- and middle-income countries. Clinically, it presents with fever, chills, headaches, and malaise, and can escalate to severe complications such as cerebral malaria, metabolic acidosis, hypoglycaemia, and multi-organ failure if untreated [1]. Vulnerable groups, particularly young children and pregnant women, suffer disproportionately, facing chronic anaemia, miscarriage, low birth weight, and impaired cognitive development [2–5]. Malaria's global distribution spans tropical and subtropical regions, forming a broad equatorial belt that includes vast portions of sub-Saharan Africa, Southeast Asia, and Latin America (Figure 1). However the burden remains starkly uneven: in 2023, the WHO African Region, primarily sub-Saharan Africa, accounted for 94% of reported cases and 95% of malaria-related deaths, with children under five comprising 76% of those fatalities [1]. While global efforts have made measurable progress, stagnation in high-burden regions reveals the shortcomings of current approaches and emphasizes the urgent need for more innovative, context-specific, and sustainable solutions.

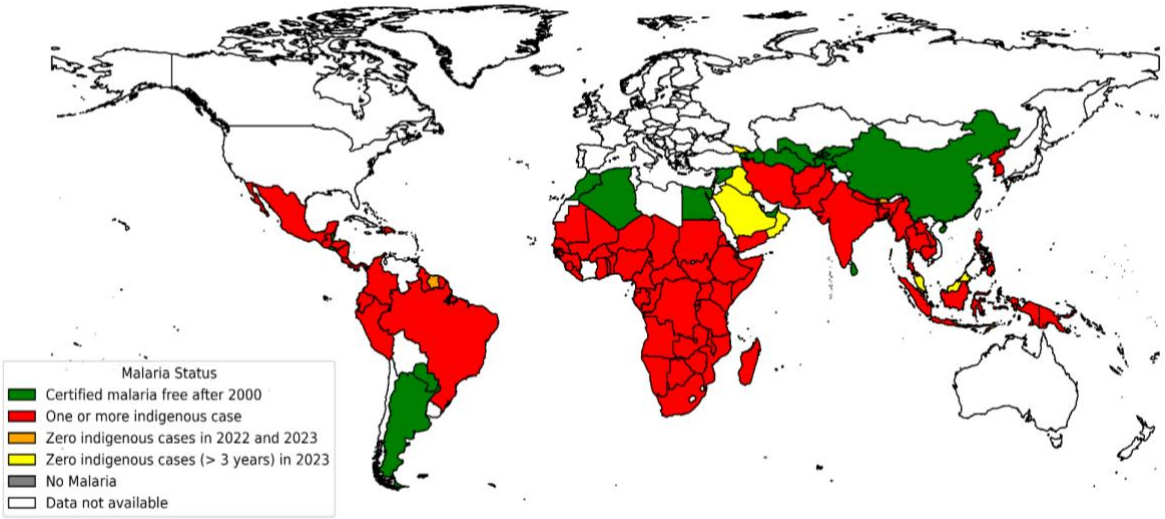


Figure 1. Global Malaria Status as of 2023. Data sourced from the World Health Organization. (Disclaimer: This map was created using publicly available data downloaded from the WHO website)

1.2 Strategies and Challenges in Malaria Control

Malaria control efforts currently adopt a dual strategy, targeting both the *Plasmodium* parasite and the *Anopheles* mosquito vector. On the parasite side, artemisinin-based combination therapies remain the first line of treatment. Preventive strategies such as intermittent preventive treatment in pregnancy and infancy, along with seasonal malaria chemoprevention, are employed in high-transmission settings [1]. However, the growing resistance of *Plasmodium* spp. to many antimalarials, which is linked to mutations in genes such as *kelch13*, *pfcrt*, *pfmdr1*, *dhps*, and *dhfr*, threatens treatment efficacy [6–15]. Compounding this is the diagnostic challenge posed by *pfhrp2* and *pfhrp3* gene deletions, which can cause false-negative Malaria Rapid Diagnostic Test results and lead to misdiagnosis and hence continued transmission [16–20].

Vector control remains a cornerstone of malaria prevention but is increasingly undermined by the rise of insecticide resistance. Core interventions including insecticide-treated nets, indoor residual spraying, and larval source management, are delivered under the Integrated Vector Management framework, which combines chemical, biological, and environmental strategies with community engagement [1]. However, resistance to pyrethroids, the primary insecticide class used in insecticide-treated nets and indoor residual spraying, is now widespread, driven by mechanisms such as metabolic detoxification, target-site mutations, and behavioural avoidance [21–30]. In response, research is advancing innovative approaches, including genetically modified mosquitoes with gene-drive constructs, symbiotic microbes like *Wolbachia* and *Microsporidia MB* to block parasite development, autodissemination strategies that exploit mosquito behaviour to deliver larvicides to cryptic breeding sites, the use of drones for precision larviciding and high-resolution habitat mapping; partially protective vaccines (e.g., RTS,S/AS01, R21/Matrix-M), and ecologically informed strategies integrated within the integrated vector management framework [31–44]. While these emerging tools show promise, many are still in experimental stages or have yet to demonstrate consistent efficacy in real-world settings.

1.3 Taxonomic Complexity and Diagnostics in Malaria Vectors

The genus *Anopheles* consists of nearly 500 described species, which span more than 100 million years of evolution [45,46]. These species are categorized into several subgenera: *Anopheles* (185 cosmopolitan species), *Baimaia* (1 Oriental species), *Cellia* Theobald (224 Old

World species), *Kerteszia* Theobald (12 Neotropical species), *Lophopodomyia* Antunes (6 Neotropical species), *Nyssorhynchus* Blanchard (39 Neotropical species), and *Stethomyia* Theobald (5 Neotropical species) [45]. While only a subset of these species transmit human malaria, this vectorial capacity is distributed throughout the phylogeny [47]. Many species also belong to closely related complexes or groups that are morphologically indistinguishable and exhibit substantial genetic similarity, largely due to their capacity to hybridize in regions of sympatry [48]. These complexities challenge accurate species identification and, by extension, effective vector surveillance.

Traditional morphological classification, while foundational, is often insufficient to differentiate cryptic species, including between malaria vectors and non-vectors [49–51]. To address this, molecular diagnostics have become essential. Polymerase chain reaction assays (PCR) targeting the ITS2 region of ribosomal DNA are widely used, but have limitations, including restricted taxonomic scope, inability to detect hybrids, and sensitivity to primer-binding mutations [52–55]. Moreover, initial morphological misclassification can compromise subsequent molecular results [56]. Current molecular approaches, even with tools like mitochondrial COI/COII and nuclear ITS2, often fail to fully capture the evolutionary and ecological diversity within *Anopheles*, particularly in resolving recently diverged taxa or detecting gene flow [57–59]. This hinders accurate assessment of vectorial capacity and insecticide resistance. While advanced multi-locus genotyping and genome-informed strategies offer higher taxonomic resolution and can identify hybrid zones and track adaptive traits [60,61], their widespread adoption in endemic regions is limited by cost, infrastructure, and the need for specialized bioinformatics expertise. Therefore, scalable and field-adaptable diagnostic platforms that combine molecular precision with logistical feasibility are urgently required.

1. 4. Malaria in Tanzania: Context and Challenges

1.4.1 Epidemiology

Over the past 25 years, Tanzania has made significant progress in reducing malaria-related deaths, primarily through the widespread deployment of vector control measures such as insecticide-treated nets, indoor residual spraying, effective antimalarial therapies, and rapid diagnostic tests [62]. Intermittent preventive treatment in pregnancy has provided additional

protection for vulnerable populations. Supported by international and domestic investments, these efforts have lowered national malaria prevalence from 14.8% in 2015 to 8.1% in 2022 and reduced the death rate from 121.6 per 100,000 people in 2000 to 41.4 in 2021 (Figure. 2A) [62]. Despite these gains, malaria control has recently plateaued, and the disease remains one of the top two causes of mortality nationwide (Figure 2B) [1]. Moreover, national-level data conceal considerable variation in malaria burden across regions, masking the spatial heterogeneity that underlies the current transmission dynamics. In several councils within the Northwestern and Southern zones, malaria prevalence among school-aged children has been reported to exceed 50% [63]. Overall, an estimated 93% of the Tanzanian population remains at risk, with higher transmission intensities reported in areas between 750 and 1,250 meters above sea level, such as the Lake and Southern zones, whereas lower transmission levels are observed in the Central and Southern Highlands (Figure 3) [62]. This marked regional variation underscores the need to understand the ecological drivers of malaria transmission, as such insights are essential for designing targeted, context-specific surveillance systems and interventions.

1.4.2 Vectors, and Parasites of Tanzania

Human malaria transmission in Tanzania is driven by several *Anopheles* species. The principal vectors belong to the Gambiae Complex (*An. gambiae* and *An. arabiensis*) and the *An. funestus* Subgroup (*An. funestus*) [64–66]. Secondary vectors, such as *An. rivulorum, An. parensis*, *An. leesoni, An. coustani*, *An. pharoensis*, and *An. squamosus* have also been reported as *Plasmodium* carriers in the country [64,66–69]. Furthermore, *An. coustani*, which was historically considered a minor vector and is abundant in the country, is now frequently detected harbouring human *Plasmodium* in multiple countries, including Cameroon, Kenya, and Madagascar [70–73], prompting a reassessment of its role in transmission [74,75]. The involvement of multiple vector species corresponds with the circulation of diverse *Plasmodium* species. Recent surveys indicate that all major human *Plasmodium* species, except *P. knowlesi*, are present in the country. A cross-sectional study of 3,456 schoolchildren reported infection rates of 22% for *P. falciparum*, 24% for *P. ovale* spp., 4% for *P. malariae*, and 0.3% for *P. vivax* [76]. Co-infections involving multiple *Plasmodium* species were also frequently observed, underscoring the complexity of malaria transmission in the country.

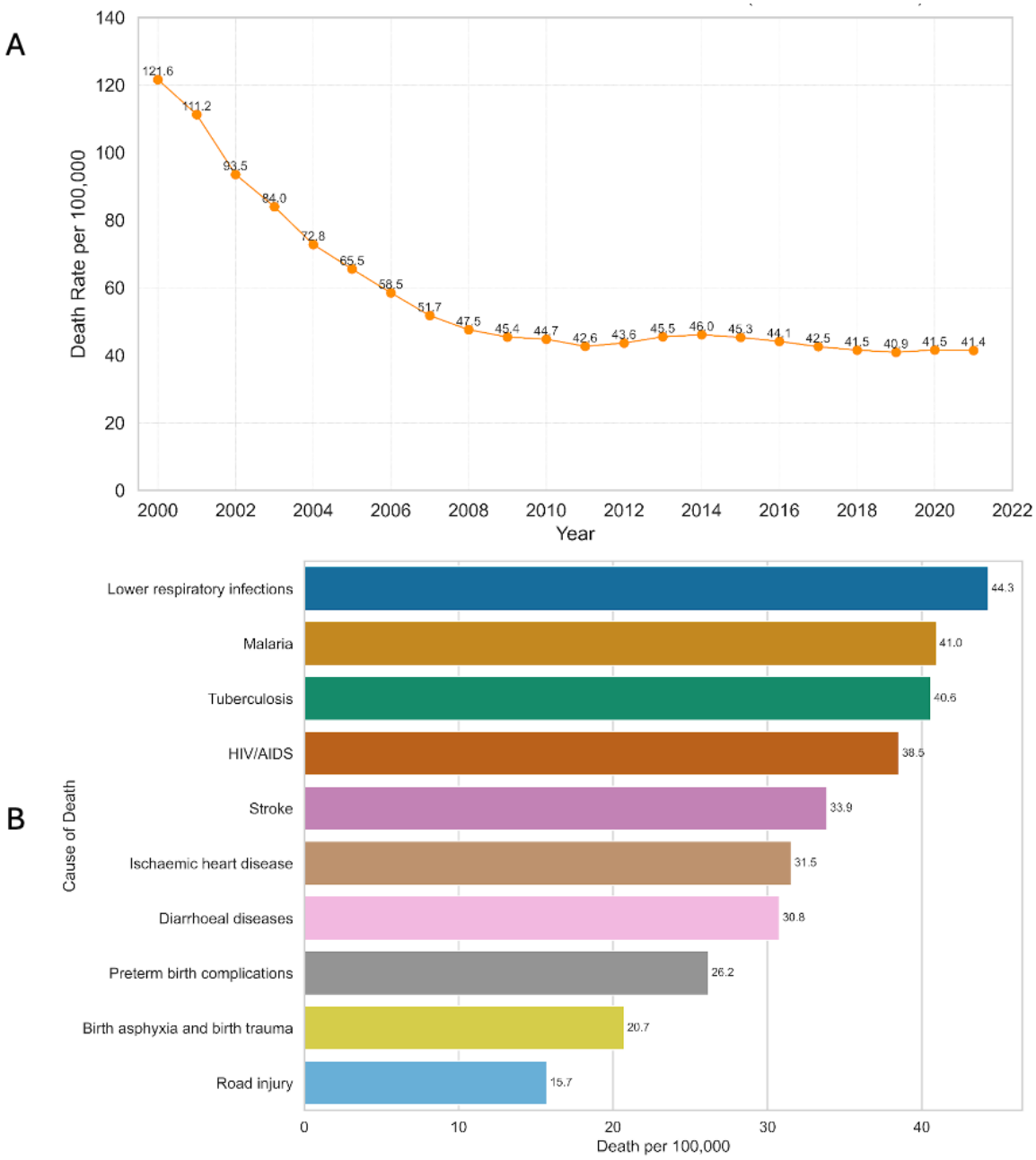


Figure 2. Malaria Trends and Status in Tanzania. (A). Trends in malaria death rate in Tanzania since 2000, according to 2021 WHO data. (B). Top 10 causes of death in Tanzania according to 2021 WHO data. (Disclaimer: This plot was created using publicly available data downloaded from the WHO website)

1.4.3 Challenges of Malaria Control in Tanzania

Gaps in Vector Surveillance: Taxonomic Resolution and Spatial Coverage

Despite notable progress in malaria control over the past two decades, persistent biological, ecological, and technological challenges continue to undermine elimination efforts in the

country. A major limitation lies in the inadequate resolution of malaria surveillance systems, particularly the shortage of detailed, species-level information on *Anopheles* vector populations across diverse ecological zones. Although approximately 50 *Anopheles* species have been reported in the country [77], most identifications relied on morphological keys, an approach with limited reliability for differentiating sibling species without molecular validation. This uncertainty surrounding vector identity, abundance, and distribution hinders a comprehensive understanding of malaria transmission dynamics.

To overcome limitations in traditional vector identification, molecular diagnostics, particularly PCR based assays targeting the ITS2 region have become widely used. However, these tools are not without limitations. Taxonomic resolution remains limited, especially for recently diverged or hybridizing taxa, and primer-binding site mutations can compromise amplification success. Moreover, PCR assays often depend on accurate initial morphological sorting, meaning early-stage misidentification can propagate through molecular workflows, resulting in misclassification or amplification failure [56]. Consequently, both morphological and PCR based methods can introduce uncertainty into species identification, potentially distorting risk assessments and misguiding intervention strategies. These diagnostic limitations are further compounded by surveillance systems that often lack sufficient geographic and temporal coverage, leaving critical data gaps. To address these challenges, Ecological Niche Models and Species Distribution Models have emerged as valuable tools. By predicting habitat suitability from species occurrence and environmental variables [78–81] these models offer an indirect but scalable means to assess vector distributions and ecological preferences, particularly where empirical data are sparse. However, their predictive accuracy is constrained by the absence of high-resolution, species-specific, and up-to-date occurrence records. This limits their effectiveness for fine-scale risk mapping, surveillance prioritization, and the design of geographically targeted vector control strategies.

Non-falciparum Malaria Parasites in Vectors: A Surveillance Gap

Although multiple human malaria *Plasmodium* species have been reported in Tanzania [76,82], surveillance and research have largely focused on *P. falciparum*, leaving the epidemiology and the mosquito vectors of non-*falciparum* malaria poorly characterized. This underrepresentation is concerning, not only because it limits understanding of transmission ecology, but also because co-infections involving multiple *Plasmodium* species have been associated with increased disease severity and mortality compared to single-species infections [83].

Compounding this, the global malaria landscape is becoming increasingly complex due to the emergence of zoonotic *Plasmodium* species, particularly in regions where humans and non-human primates share overlapping habitats [84]. Although such cases have not yet been documented locally, these global trends underscore the potential risk of zoonotic spillover and highlight the need to account for its implications in future malaria control and elimination strategies. Furthermore, in mosquito-based surveillance, circumsporozoite protein enzyme-linked immunosorbent assays, remains the standard for detecting *Plasmodium* infections, yet it has limited sensitivity for low-density infections and is largely restricted to *P. falciparum*, with poor capacity to detect non-*falciparum* species.

Fragmented Genomic Surveillance and Connectivity Data

Knowledge of the population structure, gene flow, and spatial dynamics of malaria vectors remains incomplete. While recent genomic analyses of *An. funestus* [85], have provided valuable national-scale insights, studies on *An. gambiae* and *An. arabiensis* have largely relied on microsatellite data [86,87] or geographically restricted high-resolution whole genome sequencing [88]. Consequently, these datasets offer only a fragmented view of connectivity, dispersal, and genetic structure across the country. A comprehensive genomic surveillance framework covering all major vector species is essential to monitor the spread of adaptive traits such as insecticide resistance, and to inform genetic control strategies.

1.5 Theoretical Synthesis

Malaria vector surveillance remains fragmented. Diagnostic tools often miss cryptic mosquito species or non-*falciparum Plasmodium* because they rely on morphological keys and low-resolution assays, limiting taxonomic precision. Ecological niche models, while statistically robust, rely mainly on correlative associations between occurrence records and environmental variables, offering little insight into the physiological and ecological constraints that shape mosquito distributions. Genetic studies have advanced knowledge of vector diversity but are rarely integrated with the landscapes where populations persist and move, limiting their value for surveillance and control. To address these gaps, this study proposes a techno-ecological systems framework that links three constructs, molecular resolution, ecological suitability, and genetic connectivity. This integration shifts surveillance from static, reactive monitoring to dynamic, predictive systems that can guide proactive, context-specific interventions. The framework rests on three interdependent pillars:

Enhanced Molecular Surveillance: This study applies ANOSPP (the *Anopheles* Species and *Plasmodium* Panel), a multi-locus amplicon sequencing approach, to achieve high-resolution identification of both mosquito vectors and malaria parasites. Unlike morphology-based taxonomy, single-locus PCR assays, or circumsporozoite protein ELISAs, which often fail to distinguish cryptic or emerging taxa, ANOSPP offers a more accurate and comprehensive means of detecting mosquito and parasite diversity. By overcoming these diagnostic limitations, molecular surveillance can more reliably detect species of epidemiological importance and track changes in transmission dynamics with greater precision.

Ecological Niche and Species Distribution Modelling: This study integrates fine-scale spatial and environmental data to model the ecological distribution of primary malaria vectors. Unlike traditional approaches that rely mainly on statistical correlations between occurrence records and environmental layers, the modelling framework here incorporates biological knowledge of mosquito ecology derived from laboratory and field studies. Climatic variables are used to delineate unsuitable areas, from which biologically informed pseudo-absence data are generated. Combining these pseudo-absences with observed presence records allows the models to move beyond purely correlative predictions and instead capture ecological plausibility. In doing so, the models not only estimate habitat suitability and identify the environmental drivers of vector occurrence but also achieve greater biological realism, enabling finer-scale identification of areas at elevated risk of malaria transmission.

Population Genetic Structure and Connectivity: This component uses molecular data to examine genetic diversity and population structure among malaria vector populations. By characterizing patterns of local adaptation, genetic differentiation, and connectivity, it provides critical insights into how mosquito populations are shaped across ecological landscapes. Such information is essential for understanding spatial dynamics of transmission and for anticipating the spread of adaptive traits, including insecticide resistance. It also offers an evidence base for evaluating the potential effectiveness and risks of novel interventions such as gene drive.

Together, these pillars establish a unified paradigm that integrates molecular, ecological, and genetic dimensions of surveillance. The framework generates actionable insights into where vectors occur, how their populations are structured and connected, and which species or sites are most relevant for intervention, supporting more precise and sustainable malaria control in Tanzania and comparable settings.

1.6 Objectives

1. **To assess the composition, diversity, and spatial distribution** of *Anopheles* mosquito vectors, their associated *Plasmodium* parasites across Tanzania using the ***Ano****pheles* **Sp**ecies and ***P****lasmodium* Panel (ANOSPP).

2. **To identify key environmental, climatic, and anthropogenic factors** influencing malaria vector diversity and geographic distribution.

3. **To develop predictive species distribution models** for primary malaria vectors to inform surveillance and targeted vector control strategies in Tanzania.

4. **To examine the population structure** of *Anopheles arabiensis* in Tanzania using ANOSPP generated data.

Chapter Two. Methodological Framework

2.1 Research Design

This study primarily analysed entomological samples and data collected between December 2020 and December 2023 from 25 ecologically diverse districts across mainland Tanzania. As a large country with a significant malaria burden, and substantial ecological variability from coastal lowlands to the Great Rift Valley, Tanzania offers a representative landscape for investigating malaria vector dynamics. The selected districts included Misenyi, Muleba, Ngara, Sengerema, Bariadi, Magu, Moshi Urban, Kigoma Urban, Mpanda, Nkasi, Igunga, Singida Urban, Babati, Muheza, Manyoni, Iringa Urban, Mpwapwa, Kilosa, Bagamoyo, Rufiji, Kilwa, Ruangwa, Tandahimba, Tunduru, and Ludewa (Figure 3A). These sites were part of the national malaria surveillance system coordinated by the National Malaria Control Programme (NMCP) [65], chosen to reflect a range of ecological zones, intervention strategies, malaria endemicity levels, and anthropological contexts.

The study employed a rolling cross-sectional surveillance design. Within each selected district, one of the three NMCP-designated sentinel villages was randomly chosen. In each village, three sub-villages were selected, and four households were enrolled per sub-village for entomological monitoring. Each sub-village was surveyed three times during the study period, with one night of mosquito sampling per visit. To enhance spatial coverage and minimize pseudo-replication, new households were selected for each round. Seasonal variation was addressed by ensuring that each village was sampled at least once during both wet and dry seasons. To strengthen the species distribution modelling component, additional mosquito occurrence records from published sources covering other parts of the country were integrated. In addition, mosquito specimens collected from Ulanga District, including areas within Nyerere National Park under a separate study by Deogratius Kavishe from Ifakara Health Institute, were included in genomic analyses.

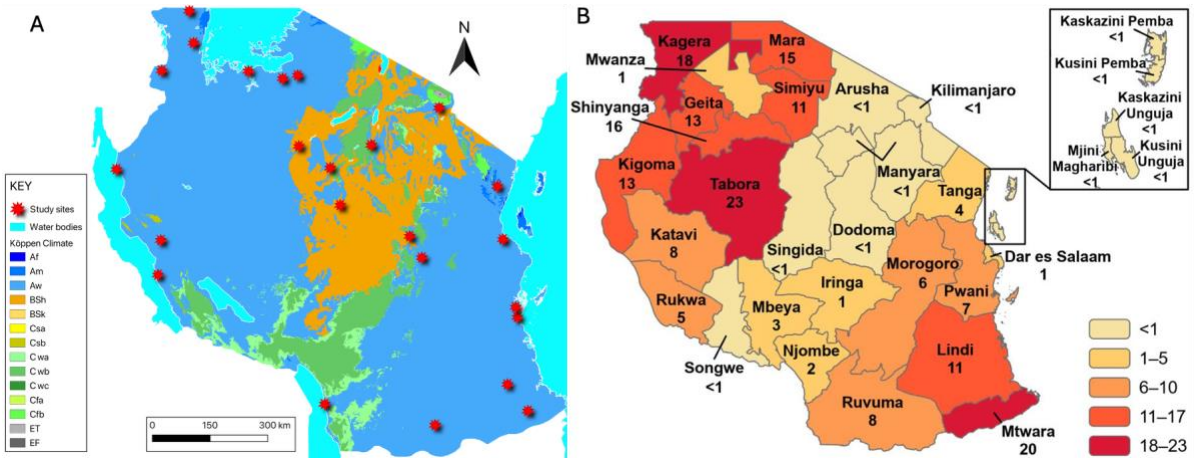


Figure 3. Study Area. Geographic distribution of mosquito sampling sites (red stars) overlaid on the Köppen–Geiger climate classification map of Tanzania, with major water bodies shown in cyan for spatial reference. In Tanzania, the tropical climates include Af (tropical rainforest), Am (tropical monsoon), and Aw (tropical savanna); the arid and semiarid climates include BSh (hot semi-arid steppe) and BWh (hot desert); the temperate climates include Cfa (humid subtropical), Cwb (temperate oceanic with dry winters), and Cwc (subtropical highland with dry winters); and finally, the polar climates are represented by EF (ice cap, top of mountain Kilimanjaro) and ET (tundra-near top of mountain Kilimanjaro). (B). Malaria prevalence across Tanzania based on the 2022 National Malaria Survey, where colour gradients indicate varying prevalence levels (in percentage) as shown in the legend (right side bottom) and highlight regional disparities in malaria burden.

2.2 Data Collection

2.2.1 Mosquito Collection

Mosquito collections were conducted in accordance with institutional and national ethical approvals. At each site (defined as one sub-village per sampling night), we deployed the same three complementary trapping methods with identical relative sampling effort: (i) Mosquito Electrocuting Traps (METs) [89–91]**,** one indoors (living room) and one outdoors (immediately outside the house) at each of four households per site, operated from 18:00–07:00 (≈13 h), yielding 8 MET trap-nights per site per sampling night**;** (ii) Backpack aspirator collections [92] , conducted the following morning for 1 hour per house (4 hours of aspiration effort per site); and (iii) Barrier Screen Interception Traps (BS) [93], two screens per site operated from 18:00–07:00 (≈13 h). This sampling design was applied uniformly at every site and in every sampling round to ensure equal relative sampling effort and allow unbiased comparison of mosquito

abundance and community composition across the study area. To capture fine-scale microclimatic variation, portable weather stations were installed near each sampled household to record nightly temperature, wind speed, and humidity (18:00–07:00).

2.2.2 Morphological and Molecular Species Identification

After collection, mosquitoes were morphologically identified based on the protocols of Coetzee et al. (2020) [49] and Gillies et al. (1987) [50] and stored individually in Eppendorf tubes filled with silica gel. A subset of specimens belonging to the primary malaria vector species was subjected to molecular analysis at the Ifakara Health Institute using species-specific polymerase chain reaction assays [53,55]. The remaining samples, comprising most of the dataset, were preserved in 100% ethanol in 96-well plates and shipped to the Wellcome Sanger Institute in the UK for molecular analysis using the ANOSPP protocol. Transportation of samples complied with the Nagoya Protocol on Access and Benefit-Sharing of Genetic Resources [94], prior to shipment. For the ANOSPP dataset, DNA extraction was carried out using a minimally morphologically destructive protocol by Korlević et al. (2021) [95], and all samples were sequenced using the ANOSPP protocol, as developed by Makunin et al. (2020) [96]. Species identification was done through the NNoVAE species assignment pipeline using anospp_analysis v0.3.5 and reference datasets nnv2, gcrefv1 and plasmv1 [97,98].

2.2.3 ANOSPP Panel Design

The ANOSPP panel is a targeted amplicon-sequencing assay utilizing 64 primer pairs for simultaneous *Anopheles* species identification and associated *Plasmodium* species detection. The panel includes 62 single-copy nuclear targets distributed across all *Anopheles* chromosome arms to maximise species-level resolution and ensure consistent amplification across divergent taxa, and two conserved mitochondrial targets for *Plasmodium* species. *Anopheles* amplicons are short (approximately ~100–250 bp) and cover highly variable regions phylogenetically informative across the genus, conserved loci retained in outgroup alignments, and X-linked loci particularly informative within the *An. gambiae* complex; targets span exonic**,** intron-spanning, and intergenic regions to capture diverse evolutionary signals (see figure 4) [97,98]. Primer binding sites were selected from conserved flanking regions using progressive masking strategies to avoid variable positions, incorporating degenerate bases where needed to maintain binding robustness across taxa. *Plasmodium* targets **(**~170–220 bp) were selected from mitochondrial rRNA-fragment regions aligned across major human malaria species, leveraging the high mitochondrial copy number for sensitive parasite detection in mixed mosquito DNA

extracts [97,98]. DNA extraction uses a non-destructive, high-throughput lysis method (figure 5)[95], enabling subsequent morphological confirmation when required. Library preparation follows a two-step multiplex PCR workflow (pooled target amplification followed by dual-index barcoding), and sequencing is performed on an Illumina MiSeq paired-end platform (see figure 5).

Following sequencing, mosquito species identity is determined using the NNoVAE workflow, beginning with a k-mer nearest-neighbour (NN) classifier against a curated reference database, followed where necessary (e.g. within the *An. gambiae* complex) by refinement using a variational autoencoder (VAE) trained on reference haplotypes to resolve closely related taxa[97,98]. *Plasmodium* detection is based on the two mitochondrial amplicons, which are aligned via BLAST against reference parasite genomes; a mosquito is considered *Plasmodium-positive*, and parasite species assigned, when both loci consistently match the same reference species [97,98]. All wet-lab steps and primer sets followed the original publications and were completed prior to this thesis; my work begins at amplicon sequences and species identified and covers all downstream bioinformatics, variant calling, data integration, and analyses.

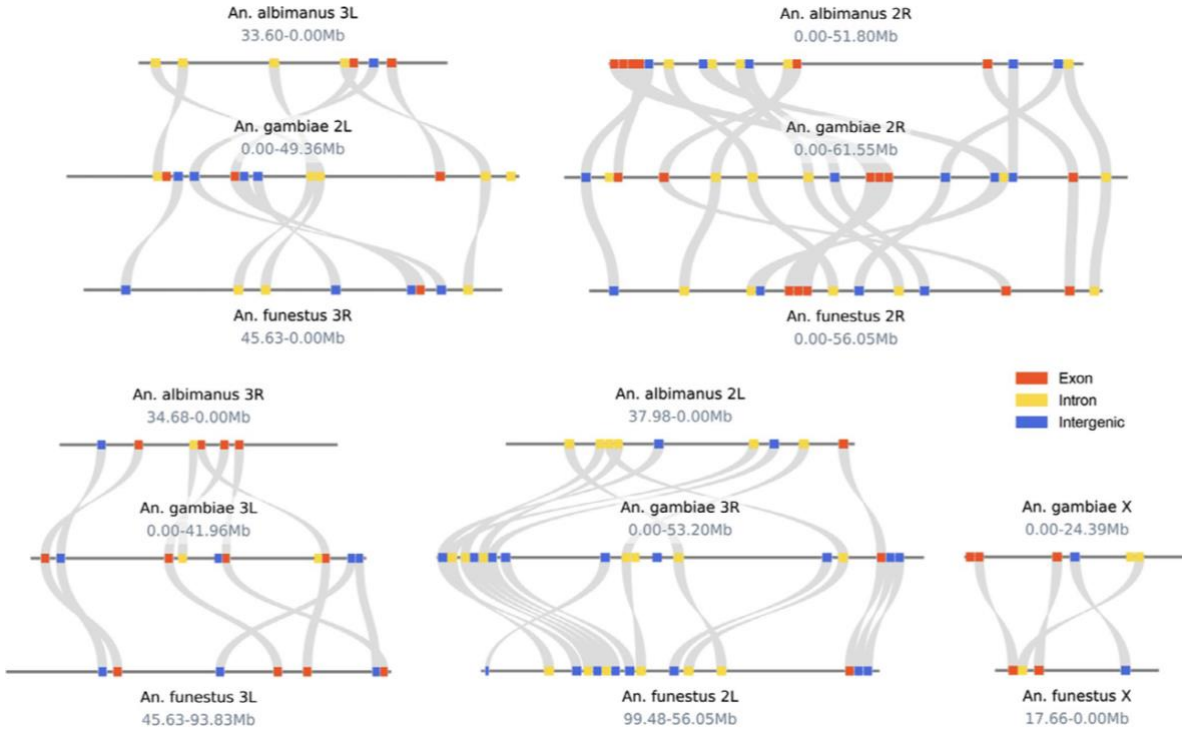


Figure 4. Genomic distribution of the 62 Anopheles nuclear amplicons targeted by the ANOSPP panel across three reference genomes. Amplicon locations shown for *Anopheles albimanus* (top), *An. gambiae* (middle), and *An. funestus* (bottom). Colours represent genomic context based on the AgamP3 annotation: exonic (red), intron-spanning (yellow), and

intergenic (blue) regions. No amplicons map to the *An. albimanus* X chromosome due to the absence of homologous target regions in this species. Figure adopted from Makunin et al. (2022)[97,98]

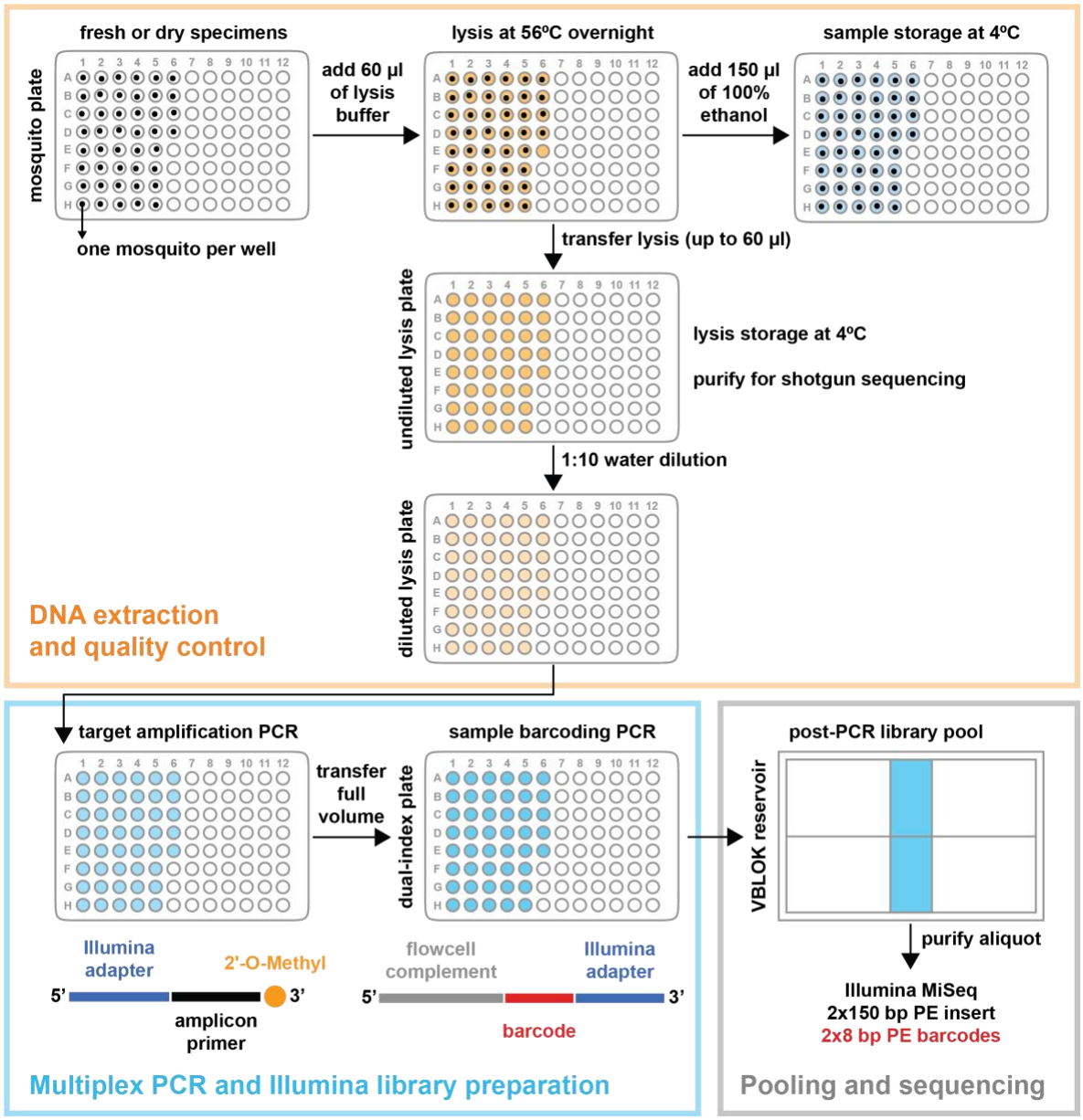


Figure 5. Workflow for the ANOSPP targeted amplicon sequencing panel from DNA extraction to Illumina sequencing. Overview of the laboratory steps used in the ANOSPP assay, including non-destructive mosquito DNA extraction, multiplex PCR amplification, dual-index barcoding, library pooling, and Illumina MiSeq sequencing. Figure courtesy of Dr. Petra Korlević.

2.2.4 *Plasmodium* Species Identification and Tree Generation

Plasmodium detection and species identification were performed using the ANOSPP protocol, which targets two short mitochondrial amplicons (hereafter P1 and P2; ~170–220 bp) for high-throughput sequencing. Parasite presence was inferred from the recovery of *Plasmodium* reads in these loci, and primary species assignment for *Plasmodium*-positive samples was obtained using the ANOSPP species-assignment pipeline (anospp_analysis v0.3.5), which implements a local BLAST-based workflow against the plasmv1 reference database [97,98] as detailed in Section 2.2.2 and 2.2.3. To visualise sequence relationships and provide an additional layer of confirmation for these BLAST-based calls, we constructed a local reference database of complete *Plasmodium* mitochondrial genomes by downloading sequences from NCBI via a custom Bash pipeline using E-utilities (*efetch*). The combined FASTA file was organised by species and accession using a modified header format (>accession|species) and integrated into R workflows for phylogenetic analysis. Per-target alignments (sample amplicons plus matched reference segments) were generated with MAFFT using the L-INS-i algorithm[99], and maximum-likelihood trees were inferred with IQ-TREE 3[100], employing ModelFinder for model selection and UFBoot (1,000 replicates) together with SH-aLRT (1,000 replicates) for branch support. Trees were rooted with *Haemoproteus columbae* as the outgroup and used qualitatively to verify that sample tips clustered with the reference sequences corresponding to their ANOSPP-assigned species (reported as Fig. 7A and Fig. 7B, respectively; species assignments reported in the main analyses are those produced by the ANOSPP pipeline.

2.2.5 Environmental and Ecological Data Collection

To extract the habitat information of *Anopheles* mosquitoes across Tanzania, environmental and ecological data was collected from publicly available sources. The bioclimatic variables (including temperature, rainfall, humidity, and aridity index) were sourced from the CHELSA database [101]. Land cover data essential for habitat delineation were obtained from the Copernicus Global Land Service [102], while human population distribution layers were sourced from WorldPop [103]. Livestock density data for cattle and goats, indicating potential blood meal sources, were retrieved from the FAO [104]. Vegetation indices, notably the Normalized Difference Vegetation Index (NDVI), were accessed from MODIS datasets for raster data, and AppeeARS for point data [105,106]. Climate classification data for isolation-by-environment (IBE) analysis were taken from the global Köppen-Geiger dataset [107]. Topographic metrics, including slope and the Topographic Wetness Index (TWI), were derived from a high-resolution digital elevation model (DEM) of Tanzania, processed using the terra

R package (v1.7-65)[108], with the DEM sourced from NASA Shuttle Radar Topography Mission (SRTM) [109]. All spatial layers were standardized to a 1 km² resolution, optimizing ecological relevance and computational feasibility, as this scale reflects the typical dispersal range of *Anopheles* mosquitoes [110–113]. Data preprocessing, raster operations, and spatial transformations were executed within the R studio environment (R version 4.4.0, RStudio version 2025.05.1-513) using the terra (v1.7-65), raster (v3.6-23), and tidyverse (v2.0.0) [108,114,115] packages, ensuring reproducibility, consistency, and transparency in the analytical workflow.

2.3 Data Analysis Process

2.3.1 Spatiotemporal Analysis and Community Niche Partitioning

To evaluate environmental drivers of mosquito species richness and community structure, a unified set of environmental predictors was applied across all analytical frameworks. These included both short-term variables (sampling-period temperature and humidity) and longer-term climatic factors (mean humidity, annual mean temperature, mean diurnal range of temperature, and annual precipitation), as well as land cover variables (shrub cover, NDVI, cropland, and built environment), and temporal livestock presence during data collection. Determinants of species richness were modelled using a quasi-Poisson generalized linear model. This model estimated the influence of environmental and land-use variables on richness patterns collection sites.

To investigate species co-occurrence patterns, pairwise Fisher's Exact tests were applied to presence–absence data across confirmed *Anopheles* species. Significant associations were classified as positive or negative based on the direction and magnitude of odds ratios, which were then log-transformed and visualized as a symmetrical heatmap. This approach captured species pairs exhibiting consistent co-occurrence or mutual exclusion across sites.

Multivariate species and environment relationships were explored using Canonical Correspondence Analysis (CCA) (using vegan package [116]) on Hellinger-transformed species abundance data. The first two constrained axes (CCA1 and CCA2) were extracted to visualize community-level responses to environmental gradients. Cosine similarity analysis was then applied to the first four significant axes to quantify alignment between individual

species and environmental vectors, and the results were displayed in a clustered similarity heatmap and later used to partition species into their community niches. All statistical analyses and visualizations were conducted in R (v4.3.2) using a fully reproducible workflow. Spatial layers were processed using the sf and terra packages ([108,117], with background maps sourced from rnaturalearth [118], and visual outputs generated with ggplot2 [119] and complexheatmap [120] packages.

2.3.2 Species Distribution Modelling

To estimate the spatial distributions of *Anopheles arabiensis*, *An. gambiae s.s.*, and *An. funestus* across Tanzania, a species distribution modelling (SDM) framework was implemented that integrated ecologically informed pseudo-absence generation with environmental, climatic, demographic, and topographic covariates (Table 4). Occurrence records were spatially aggregated within a 1 km radius to minimize spatial autocorrelation, align with the resolution of predictor rasters, and reflect realistic mosquito dispersal ranges [110–113]. Pseudo-absence points were generated from environmentally unsuitable areas identified using forest-canopy thresholds, aridity indices, and temperature-based suitability limits, defined according to species-specific ecological tolerances (see Chapter 4, Section 4.2). Predictor variables were standardized and screened for multicollinearity using Pearson correlation and Variance Inflation Factor (VIF) analysis.

This study employed an advanced hybrid SDM framework that departs from both traditional correlative models (e.g., MaxEnt[81]) and purely mechanistic approaches (e.g., temperature-driven physiological suitability[121]). The distinguishing feature lies in the construction of the model background: pseudo-absences were generated explicitly from biologically unsuitable zones grounded in species-specific physiological and climatic thresholds. This physiology-guided pseudo-absence selection anchors the model in ecological realism and minimizes bias associated with random or environmentally accessible yet unsuitable background points [122]. Similar principles have been advocated to enhance predictive realism [123,124] and to integrate mechanistic realism within correlative frameworks [125]. The resulting physiology-guided pseudo-absence plus presence model framework bridges the mechanistic precision of studies such as Ryan et al. (2015)[121] with the correlative flexibility of global mosquito-mapping initiatives Sinka et al. (2010, 2012)[81,126]. To my knowledge, this represents the first application of such a hybrid SDM configuration for *Anopheles* distribution mapping in Tanzania, and likely one of the earliest if not the first within sub-Saharan Africa.

Species–environment relationships were modelled using binomial Generalized Additive Models (GAMs)[127] implemented in the mgcv package (R). Restricted maximum likelihood (REML) estimation with penalized smoothing splines was used to capture nonlinear effects while preventing overfitting. Although alternative algorithms such as Random Forests, Boosted Regression Trees, ensemble frameworks, and Bayesian hierarchical methods (e.g., INLA) can incorporate pseudo-absences, GAMs were selected for their optimal balance of flexibility, interpretability, and diagnostic transparency. In contrast, presence-only methods such as MaxEnt rely on background sampling and provide limited insight into residual structure or model calibration.

Each model underwent a rigorous diagnostic and validation workflow. Model residuals were evaluated using the DHARMa package[128] to detect overdispersion, zero inflation, and spatial autocorrelation. Predictor refinement followed an AIC-guided backward elimination procedure, ensuring that each retained term improved both model parsimony and diagnostic behaviour. Model performance and generalizability were assessed via 10-fold spatial block cross-validation [129] implemented in the blockCV package, thereby reducing spatial dependence between training and test data and avoiding overestimation of predictive power. Calibration accuracy was further evaluated using reliability curves and root mean square error metrics[130]. The final modelling framework integrates ecological realism, statistical rigour, and diagnostic transparency to produce spatially explicit habitat-suitability maps for Tanzania's three principal malaria vectors. These models capture complex, nonlinear environmental responses while remaining interpretable and biologically grounded, thereby providing a robust evidence base for geographically targeted, species-specific malaria vector control and surveillance strategies across Tanzania.

2.3.3 Population Structure Analysis

To investigate population structure among *Anopheles* vector populations across Tanzania, sequencing data from the ANOSPP panel, which targets 62 informative short-amplicon regions optimized for species identification, was used [96]. Raw sequence reads underwent sample inference and quality control using DADA2 [131], followed by extraction of corresponding amplicon reference sequences from the *An. arabiensis* Dongola strain reference genome (AaraD3, GCF_016920715.1) [132], which was downloaded from NCBI. For extraction of reference amplicon for each target, the most frequently observed haplotype sequence was

aligned to the *An. arabiensis* reference genome using locally installed BLASTn (BLAST+ v2.13.0). Genomic coordinates from the top hits were parsed and written into BED format via a custom Bash script. Strand-aware reference amplicon sequences were then extracted using bedtools getfasta (v2.30.0) [133]. To verify accuracy, each reference sequence was realigned to the original query, and alignment reports were generated including coordinate ranges, percent identity, and orientation. The genomic annotation of each target, such as chromosomal location and whether the region fell within an exon, intron, or intergenic space, was retrieved from the accompanied GFF file and included in the report for verification. Validated reference sequences were renamed and prepended to the corresponding sample haplotypes, generating one merged FASTA per target. These were subsequently aligned using MAFFT (v7.505) with the *--globalpair* and *--large* settings [99].

Variant calling was performed on the aligned FASTA files using a custom R-based pipeline designed to operate using ANOSPP data or any other multilocus sequence outputs. This pipeline mimicked standard VCF logic [134] to detect SNPs, indels, and complex variants, outputting a long-format CSV with detailed genotype information per specimen. Only biallelic variants were retained for downstream analyses. Missing genotypes were imputed using a nonparametric random forest approach implemented via the missForest package in R [135]. This method was selected for its ability to model complex multilocus genotype structures and its robustness to missing data. Random forest imputation has demonstrated high accuracy in genetic datasets, including under moderate missingness levels up to 20%, and performs comparably to Hidden Markov Model–based approaches [136–138]. The resulting imputed matrix was then used for subsequent downstream analyses.

Population structure was assessed using Principal Component Analysis (PCA) and Discriminant Analysis of Principal Components (DAPC). PCA was performed in R using the *prcomp* function to summarize major axes of genetic variation and to evaluate patterns such as large-scale differentiation, potential chromosomal inversions, and noise or outlier signals in the dataset. DAPC was conducted in the *adegenet* package (v2.1.10) [139], with districts (for Tanzania) and countries (for continental analyses) specified as predefined groups. These groupings were used as priors to guide the analysis but did not impose structure, as DAPC identifies genetic clusters directly from allele frequency variation. The number of retained principal components was determined through cross-validation and score optimization to achieve the best balance between discrimination and avoid overfitting.

To evaluate spatial and environmental drivers of genetic variation, Isolation by Distance (IBD), Isolation by Resistance (IBR), and Isolation by Environment (IBE) using Mantel and partial Mantel correlations, controlling for spatial autocorrelation among genetic, geographic, and ecological distance matrices. These analyses were complemented by maximum-likelihood population-effects (MLPE) models, which explicitly account for spatial autocorrelation and the non-independence of pairwise genetic distances [140,141]. Pairwise FST values were calculated following Weir and Cockerham [142] to quantify genetic differentiation among populations. Together, these approaches provided a multilayered framework for characterizing population structure and ecological differentiation across Tanzania using the amplicon-based ANOSPP dataset.

2.4 Ethical Considerations

Full ethical approval for this study was obtained from the Ifakara Health Institute, Institutional Review Board (IHI/IRB/No:09-2020), the National Institute for Medical Research (NIMR/HQ/R.8c/Vol.1/1984), and the President's Office – Regional Administration and Local Government (AB.307/223/01). Additional permissions were secured from local government authorities and community leaders in each sub-village where the research was conducted. Written informed consent was obtained from all study participants, including heads of households and mosquito collectors. For participants unable to read, the consent form was read aloud and explained in Kiswahili or the local language by trained field staff, in the presence of a community witness. Upon agreeing to participate, the household head or their representative provided a thumbprint on the consent form, which was then signed by the witness. All participants were informed of their right to withdraw from the study at any time without consequence. To ensure household privacy during overnight mosquito collections, only houses with at least two bedrooms and a living room were selected. Priority for mosquito collection roles was given to consenting members of the participating household; if no one consented, volunteers were recruited from nearby households within the same sub-village. As the study began during the COVID-19 pandemic in 2020, field activities were suspended in compliance with national restrictions. Once the ban was lifted, data collection resumed with strict adherence to WHO guidelines, including the use of face masks, regular handwashing, avoiding handshakes, application of hand sanitizers, and maintaining social distance to protect both researchers and participants.

Chapter Three: Diversity, Community Niche Structure, and Geographical Distribution of *Anopheles* Mosquitoes and Their Associated *Plasmodium* Species in Tanzania

Chapter summary

This chapter presents a comprehensive analysis of the spatial distribution, diversity, and ecological patterns of *Anopheles* mosquito species and their associated *Plasmodium* parasites across 25 districts in Tanzania. The aim is to inform vector control strategies by identifying spatial heterogeneity and co-occurrence patterns that can support locally tailored interventions. Understanding species assemblages, co-occurrence frequencies, and the ecological drivers behind them provides insights into interspecific interactions, potential competition, and the use of signature species as indicators in surveillance. Through spatial mapping, statistical plots, and tabular summaries, this chapter describes variation in species richness, abundance, species composition and community structure, and regional distribution, and the ecological drivers influencing these patterns. It also compares conventional morphological identification with ANOSPP based identification, demonstrating substantial gains in taxonomic resolution and identification accuracy. It also identifies *Anopheles* vector species linked to the transmission of specific *Plasmodium* parasites and quantifies co-infections instances where single mosquito specimens carry multiple *Plasmodium* species. Importantly, the chapter reports, for the first time in Tanzania, the detection of *Plasmodium caprae*, a goat-specific *Plasmodium* species, in *Anopheles arabiensis*. While not zoonotic, this finding broadens our understanding of host–parasite–vector interactions and the ecological range of malaria vectors.

3.1 Mosquito Collection and Morphological Identification

Over a three-year surveillance period from December 2020 to December 2023, a total of 71,146 mosquitoes were collected across 25 sentinel districts. Morphological identification categorized these into two primary groups: Culicines and Anophelines (Figure 6). The majority of the collection, n=61,126 (86.0%), were Culicines. Within this group, species from the *Culex* genus predominated, accounting for n=57,620 (80.99%) of the total collection. Other Culicine genera identified included *Coquillettidia* at n=1,942 (2.73%), *Mansonia* at n=1,351 (1.90%), and *Aedes* at n=213 (0.30%). The remaining n=10,020 (14.0%) consisted of *Anopheles* species. Among these, *An. gambiae s.l.* was the most prevalent, representing n=4,168 (5.86%) of the total collection. Other notable *Anopheles* species included *An. pharoensis* at n=2,005 (2.82%),

An. coustani at n=1,574 (2.21%), and *An. funestus s.l.* at n=1,314 (1.85%). Additional *Anopheles* species identified in smaller numbers were *An. squamosus* (n=905, 1.27%), *An. maculipalpis* (n=18, 0.03%), *An. ziemanni* (n=18, 0.03%), *An. rufipes* (n=17, 0.02%), and *An. cinctus* (n=1, 0.001%).

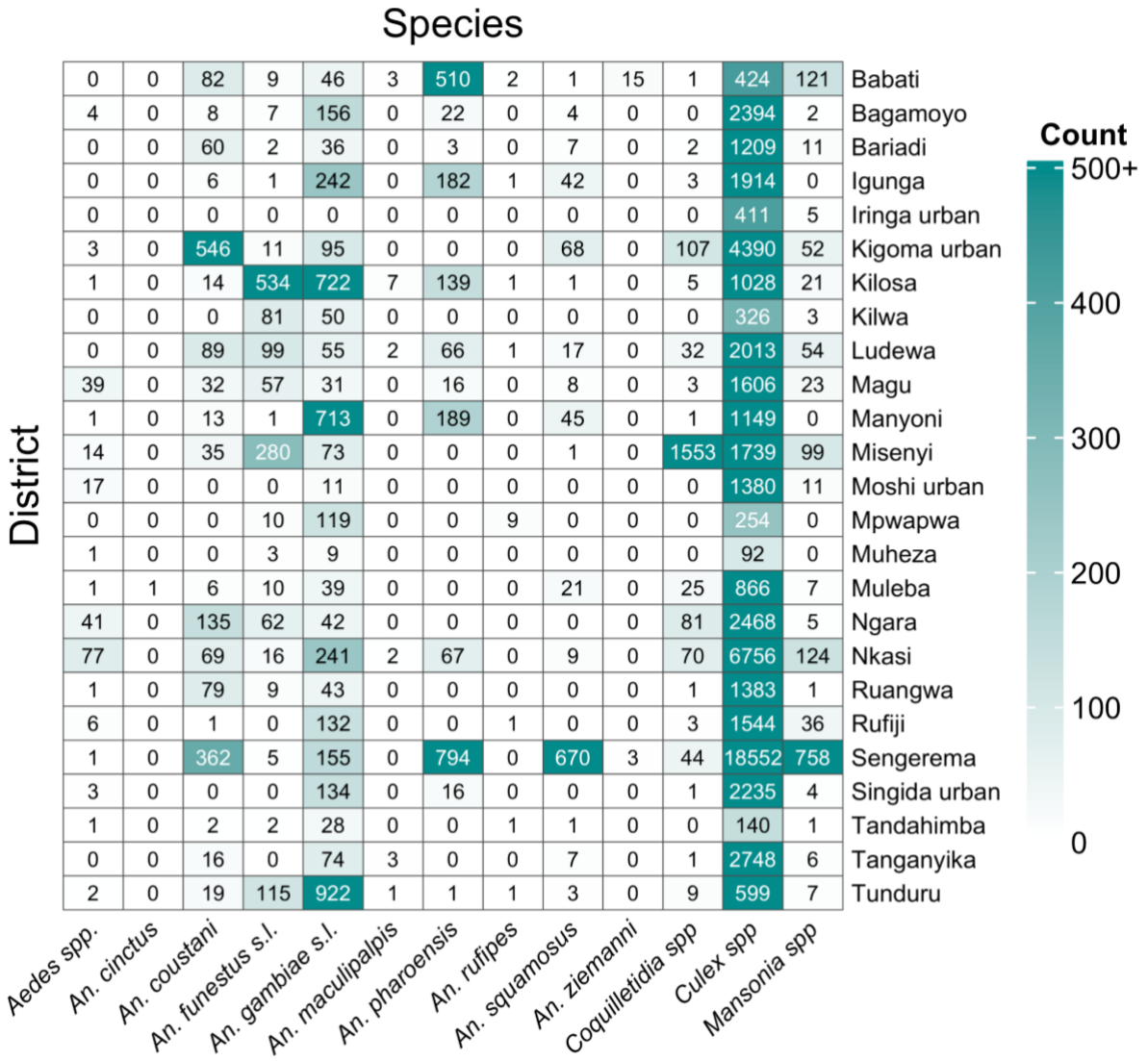


Figure 6. Distribution and relative abundance of mosquito taxa (identified by morphological keys) across districts. The heatmap shows the number of specimens by taxon (columns) and district (rows) collected during the study period. Cell colour intensity represents specimen count, using a white-to-teal gradient capped at 500 (counts above 500 shown at maximum saturation). Numeric values in each cell indicate the exact count per taxon–district pair.

3. 2 Molecular Identification of *Anopheles* Species

Molecular identification was performed using two methods: conventional PCR for a subset of samples and ANOSPP for a larger collection.

3.2.1 PCR-based Identification

A total of 2,642 *Anopheles* mosquitoes, initially identified morphologically as members of the *An. gambiae s.l.* or the *An. funestus s.l.*, were submitted for PCR amplification. Of these, 248 samples failed to amplify, yielding 2,394 successfully amplified and confirmed specimens. PCR-based identification revealed the following sibling species composition: *An. arabiensis* (n = 1,686; 70.43%), *An. funestus* s.s (n = 405; 16.92%), *An. gambiae* s.s (n = 173; 7.23%), *An. quadriannulatus* (n = 62; 2.59%), *An. leesoni* (n = 50; 2.09%), and *An. rivulorum* (n = 18; 0.75%). These specimens were not processed further using the ANOSPP panel but are retained for complementary analyses in this and subsequent chapters, where species-level resolution alone is sufficient. (See Table1)

3.2.2 ANOSPP-based Identification

A total of 6,650 *Anopheles* mosquito specimens were submitted for ANOSPP sequencing. Of these, 1,124 samples failed quality control due to insufficient amplicon recovery, likely attributable to poor sample quality, and 28 mosquitoes was further filtered out dure to potential contamination. The remaining 5498 samples were successfully identified at different levels of taxonomic resolution. The taxonomic resolution achieved across these samples varied. A small fraction (n = 10; 0.20%) were identified only at a coarse taxonomic level, either series or subgenus, indicating incomplete resolution: *Cellia* series (n = 3), Christya series (n = 1), and *Myzomyia* series (n = 6). An additional 1,032 samples (18.77%) were resolved at an intermediate level, typically complex or group level, also reflecting partial resolution: *An. coustani group* (n = 669), *An. marshallii group* (n = 41), *An. funestus group* (n = 28), and *An. gambiae complex* (n = 279). Most samples (n = 4,456; 81.05%) were successfully identified to the fine or sibling species level. Among these, the species distribution was as follows: *An. pharoensis* (n = 1,573; 35.30%), *An. arabiensis* (n = 1,361; 30.54%), *An. squamosus* (n = 671; 15.06%), *An. funestus* (n = 448; 10.05%), *An. rivulorum* (n = 170; 3.82%), *An. ziemanni* (n = 121; 2.72%), *An. gambiae s.s* (n = 70; 1.57%), *An. quadriannulatus* (n = 32; 0.72%), *An. pretoriensis* (n = 18; 0.40%), *An. rufipes* (n = 4; 0.09%), and *An. maculipalpis* (n = 3; 0.07%). The incomplete resolution observed (intermediate level resolution only) was primarily due to

the absence of corresponding reference sequences in existing databases, contamination, and in some cases, due to insufficient amplicon recovery that rendered samples inadequate for full taxonomic resolution. Therefore, subsequent analyses will focus exclusively on fully resolved species, apart from the *An. marshallii* and *An. coustani* groups, for which where necessary the intermediate level will be used given representative species-level data are unavailable. (See Table1)

Table 1. Species composition summary across *Anopheles* species complexes identified in this study

Series	Species Group/Complex	Species	Vectorial Status	PCR	ANOSPP	Total
Pyretophorus	*Gambiae*	*An. gambiae*	Primary vector	173	70	244
		An. arabiensis	Primary vector	1686	1361	3050
		An. quadriannulatus	Non-vector	62	32	94
		Either member of *Gambiae* complex	-	0	279	279
Myzomyia	*Funestus*	*An. funestus*	Primary vector	405	448	853
		An. leesoni	Secondary vector	50	0	50
		An. rivulorum	Secondary vector	18	170	189
		Either member of *Funestus* group	-	0	28	28
	Marshalii	Either member of *Marshalii* group	Some are secondary vectors while most are non-vectors	0	41	41
	-	Either member of *Myzomyia* series	-	0	6	6
Cellia	-	*An. pharoensis*	Secondary vectors	0	1573	1576
	Squamosus	*An. squamosus*	Secondary vectors	0	671	673
	-	Either member of *Cellia* series	-	0	3	4
Christya	-	Either member of *Christya* series	Non vectors	0	1	1
Neocellia	-	*An. maculipalpis*	Non vector	0	3	3
	-	*An. pretoriensis*	Non vector	0	18	18
	-	*An. rufipes*	Secondary vector	0	4	4
Myzorynchus	*Coustani*	*An. ziemanni*	Secondary vector	0	121	121
		Either member of *Coustani* group	-	0	669	682
Grand Total				2394	5496	7890

3.3 *Plasmodium* species detected *in Anopheles* Mosquitoes

Analysis of *Plasmodium* presence in *Anopheles* mosquitoes revealed circulation of multiple parasite species across several mosquito species, including instances where more than one species was detected in a single mosquito. *An. arabiensis* carried *P. falciparum* (n = 5) and *P. caprae* (n = 5), the latter representing the first report of this parasite in Tanzania. In *An. funestus*, both single and multi-species detections were observed: *P. falciparum* (n = 14), *P. malariae* (n = 1), and *P. ovale* (n = 3) occurred alone, while two multi-species detections were recorded, *P. vivax* + *P. malariae* and *P. falciparum* + *P. malariae*. *An. gambiae* carried *P. falciparum* (n = 4), while single detections of *P. falciparum* were also found in *An. pharoensis* (n = 1) and *An. rivulorum* (n = 1) (Table 2, Figure 7).

Table 1. Detection Status of Individual Anopheles Mosquitoes Harbouring Plasmodium Parasites

Mosquitoes Species	*Plasmodium* Status	*Plasmodium* Species	count	Prevalence (Numbers)	Prevalence (Percent)
An. arabiensis	Single species	*P. falciparum*	5	5/1361	0.36
An. arabiensis	Single species	*P. caprae*	5	5/1361	0.44
An. funestus	Two species	*P. vivax; P. malariae*	1	1/448	0.22
An. funestus	Two species	*P. falciparum; P. malariae*	1	1/448	0.22
An. funestus	Single species	*P. falciparum*	14	14/448	3.13
An. funestus	Single species	*P. malariae*	1	1/448	0.22
An. funestus	Single species	*P. ovale*	3	3/448	0.67
An. gambiae	Single species	*P. falciparum*	4	4/70	5.71
An. pharoensis	Single species	*P. falciparum*	1	1/869	0.12
An. rivulorum	Single species	*P. falciparum*	1	1/170	0.59

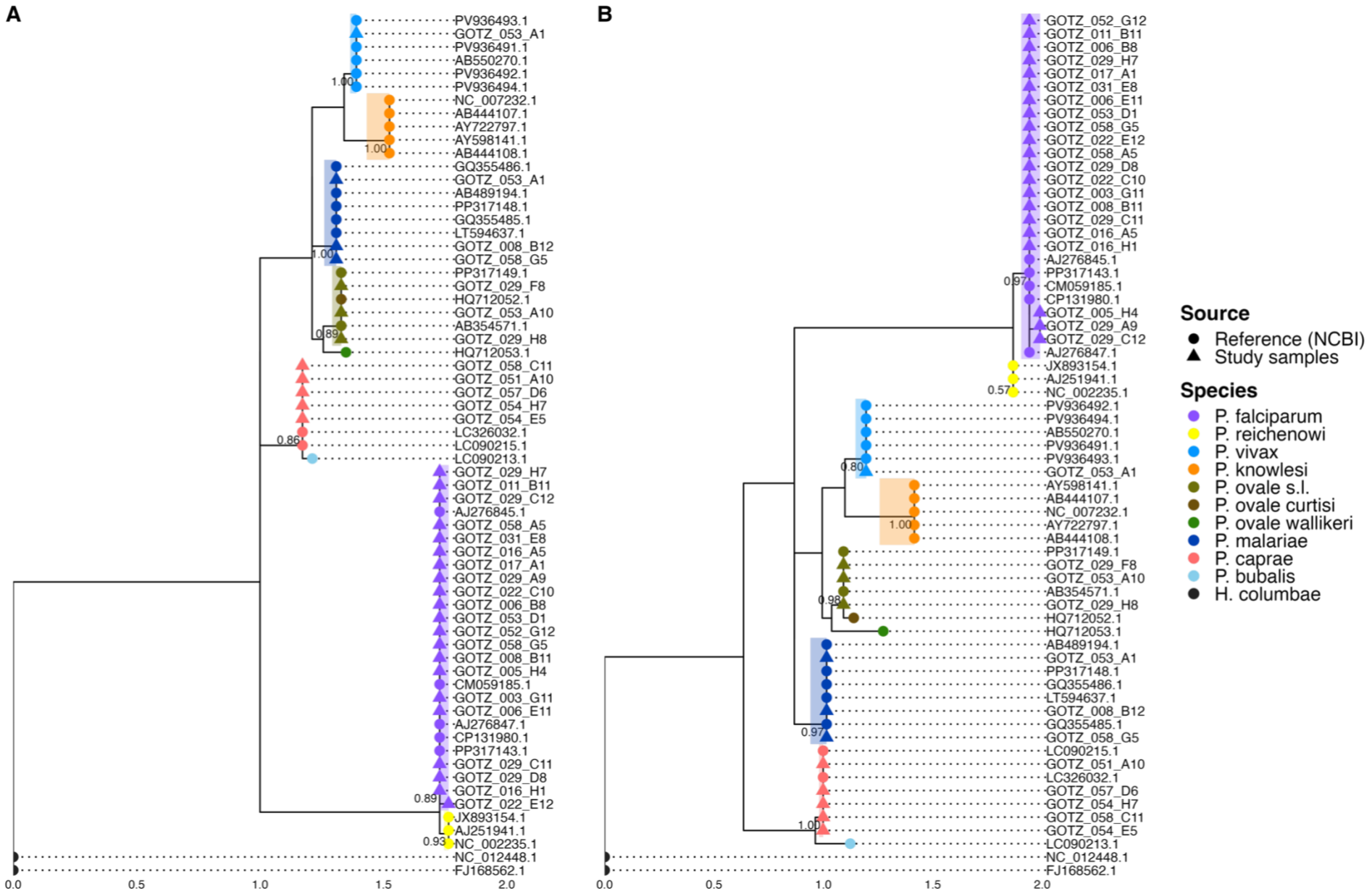

A
B
PV936493.1
GOTZ_053_A1
PV936491.1
AB550270.1
PV936492.1
PV936494.1
NC_007232.1
AB444107.1
AY722797.1
AY598141.1
AB444108.1
GQ355486.1
GOTZ_053_A1
AB489194.1
PP317148.1
GQ355485.1
LT594637.1
GOTZ_008_B12
GOTZ_058_G5
PP317149.1
GOTZ_029_F8
HQ712052.1
GOTZ_053_A10
AB354571.1
GOTZ_029_H8
HQ712053.1
GOTZ_058_C11
GOTZ_051_A10
GOTZ_057_D6
GOTZ_054_H7
GOTZ_054_E5
LC326032.1
LC090215.1
LC090213.1
GOTZ_029_H7
GOTZ_011_B11
GOTZ_029_C12
AJ276845.1
GOTZ_058_A5
GOTZ_031_E8
GOTZ_016_A5
GOTZ_017_A1
GOTZ_029_A9
GOTZ_022_C10
GOTZ_006_B8
GOTZ_053_D1
GOTZ_052_G12
GOTZ_058_G5
GOTZ_008_B11
GOTZ_005_H4
CM059185.1
GOTZ_003_G11
GOTZ_006_E11
AJ276847.1
CP131980.1
PP317143.1
GOTZ_029_C11
GOTZ_029_D8
GOTZ_016_H1
GOTZ_022_E12
JX893154.1
AJ251941.1
NC_002235.1
NC_012448.1
FJ168562.1
1.00
1.00
1.00
0.89
0.86
0.89
0.93
0.97
0.57
0.80
1.00
0.98
0.97
1.00
0.0
0.5
1.0
1.5
2.0
Source
Reference (NCBI)
Study samples
Species
P. falciparum
P. reichenowi
P. vivax
P. knowlesi
P. ovale s.l.
P. ovale curtisi
P. ovale wallikeri
P. malariae
P. caprae
P. bubalis
H. columbae

Figure 7. Maximum-likelihood phylogenetic trees for Plasmodium species identification. Phylogenetic trees inferred from ANOSPP mitochondrial amplicons P1 (A) and P2 (B). Trees were reconstructed in IQ-TREE 3 under the best-fit nucleotide substitution models selected by ModelFinder. Node support values represent ultrafast bootstrap (UFBoot) and SH-aLRT estimates. Each tip corresponds to either a reference mitochondrial genome (circles) or a study-derived haplotype (triangles). Major clades are colour-coded by species according to the key on the right. Trees are rooted with *Haemoproteus columbae* as the outgroup, and branch lengths indicate substitutions per site.

3.4 Comparison of Morphological and Molecular Identification

A total of 5,488 mosquito specimens with paired field-based morphological and ANOSPP molecular identifications were included in the comparison. Morphological identifications were conducted at species, group, or complex levels. Consequently, comparisons between morphological and molecular identifications were limited to equivalent morphological taxonomic resolution.

The ANOSPP assay substantially reduced species misidentification compared with morphology alone. Of the 5,488 specimens examined, 338 were misclassified morphologically when ANOSPP was used as the reference, corresponding to a 6.2% misidentification rate (95% CI: 5.5–6.8%), which was significantly higher than the assumed 5% error threshold (binomial test, $p = 0.00014$). Misidentification rates varied markedly among taxa. *An. pharoensis* showed the lowest rate (9 of 1,556; 0.6%), whereas *An. rufipes* exhibited the highest (9 of 11; 81.8%). High rates were also recorded for *An. maculipalpis* (7 of 9; 77.8%) and *An. ziemanni* (3 of 18; 16.7%) (Figure 4). Intermediate error rates occurred within the *An. gambiae* complex (120 of 1,832; 6.6%), *An. funestus* group (53 of 612; 8.7%), *An. coustani* group (111 of 766; 14.5%), and *An. squamosus* (26 of 684; 3.8%). A chi-squared test on the contingency table of correct versus incorrect identifications by taxon confirmed substantial heterogeneity in misclassification rates across groups ($\chi^2 = 381.8$, $p = 0.001$, Monte Carlo simulation).

Binomial logistic regression using the *An. coustani* group as the reference category further highlighted these contrasts. Taxa with significantly lower odds of misidentification included *An. pharoensis* (OR ≈ 0.03, $p < 2\times10^{-16}$), *An. squamosus* (OR ≈ 0.23, $p = 9.3\times10^{-11}$), *An. gambiae* complex (OR ≈ 0.41, $p = 2.5\times10^{-10}$), and *An. funestus* group (OR ≈ 0.56, $p = 0.001$), reflecting high morphological familiarity with major primary vectors. By contrast, *An. rufipes* (OR ≈ 26.6, $p = 3.2\times10^{-5}$) and *An. maculipalpis* (OR ≈ 20.7, $p = 1.8\times10^{-4}$) had dramatically elevated odds of misclassification, consistent with the very high failure rates observed for these rarely targeted taxa. *An. ziemanni* showed a moderate failure rate (16.7%), but its odds ratio did not differ significantly from the *An. coustani* group, likely due to the small sample size.

In addition to correcting misclassifications within morphologically recognised groups, ANOSPP resolved species that were not detected by morphology at all. Notably, the assay identified 41 specimens belonging to the *An. marshallii* group and 18 specimens of *An. pretoriensis*, none of which had been distinguished as such in the original morphological

sorting. These findings underscore how reliance on morphology alone can under-represent morphologically challenging, or less prioritised taxa in routine surveillance, and illustrate the added value of molecular diagnostics for capturing the full breadth of *Anopheles* diversity in Tanzania.

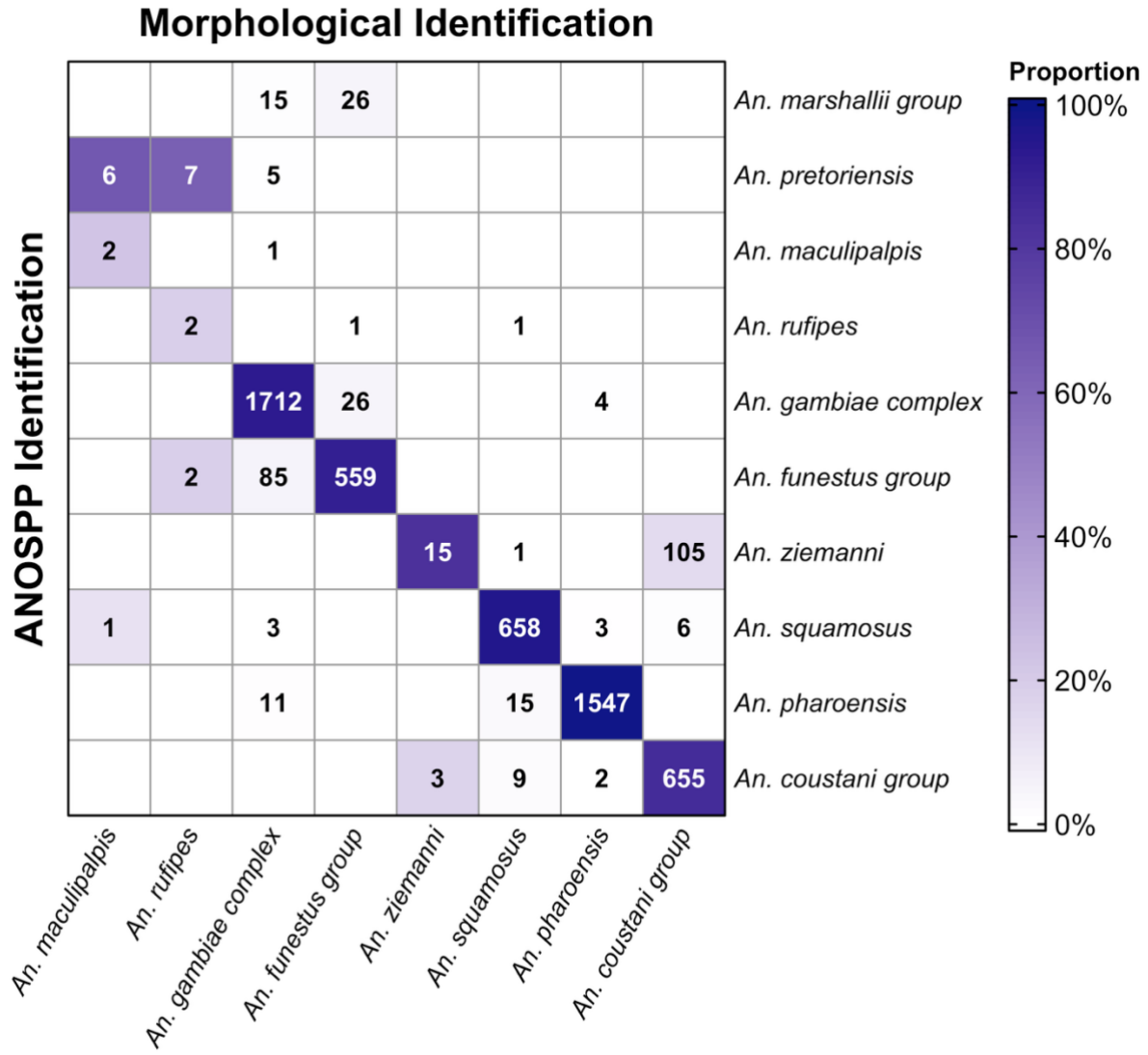


Figure 8. Heatmap of morphological versus ANOSPP identification. Each cell represents the proportion (colour intensity) and absolute number (numeric label) of specimens within each morphological group (columns) assigned to a given molecular confirmed group by ANOSPP (rows). The proportion is calculated relative to the total number identified morphologically as each taxon. Diagonal cells indicate congruent identifications for all species except *An. marshallii* group, and *An. pretoriensis*, which were not detected morphologically; off-diagonal cells reveal misclassification patterns.

3.5 Spatial Distribution and Species Richness

Among the primary malaria vectors, *An. arabiensis* emerged as the most widely distributed species, consistently present in large numbers across Tanzania. However, several high-transmission districts, such as Ngara, Kilwa, Missenyi, and Magu as well as Ludewa (a low-transmission area with a recognized hotspot in Manda ward), showed a different pattern. In these areas, *An. funestus* was found in relatively high densities. Similarly, in Tandahimba District, *An. gambiae* was notably more prevalent than other vector species. Although *An. gambiae s.s* and *An. funestus* were generally less abundant than *An. arabiensis*, their consistent presence in high-transmission zones underscores their strong association with elevated malaria risk, indicating their higher transmission potential relative to other species. Several other *Anopheles* species were also detected in considerable numbers and exhibited broad geographic distributions, including *An. pharoensis*, *An. squamosus*, and members of the *An. coustani* group. In contrast, species from the *An. marshallii* group appeared more localized, with occurrences primarily in the northwestern districts of Missenyi, Ngara, and Muleba, as well as in central Tanzania (Mpwapwa and Kilosa). *An. pretoriensis* was only recorded in Kilosa and Mpwapwa (central Tanzania) and Ludewa (southern Tanzania). Species richness varied markedly across the surveyed districts. Ludewa, Tunduru, and Kilosa exhibited the highest species richness, each with nine or more distinct *Anopheles* species. In contrast, Moshi Urban recorded very low richness, with only one species identified (see figure 9 & 10).

From a parasitological perspective, *P. falciparum* was the most detected species, found in mosquito samples from nine of the twenty-five surveyed districts, confirming its widespread distribution. Conversely, *P. caprae* was only identified in Kilosa District. Notably, districts where multiple human-infecting *Plasmodium* species were detected in mosquitoes often overlapped with areas of high malaria transmission. For example, Missenyi District showed the presence of *P. falciparum*, *P. malariae*, and *P. ovale*; Ngara recorded *P. falciparum*, *P. vivax*, and *P. malariae*; and Kilwa exhibited co-circulation of *P. falciparum* and *P. ovale*. While the absolute counts of these infections were relatively low, the diversity of *Plasmodium* species observed highlights ongoing transmission and the circulation of multiple malaria parasites within vector populations in these high-burden regions (see figure 11).

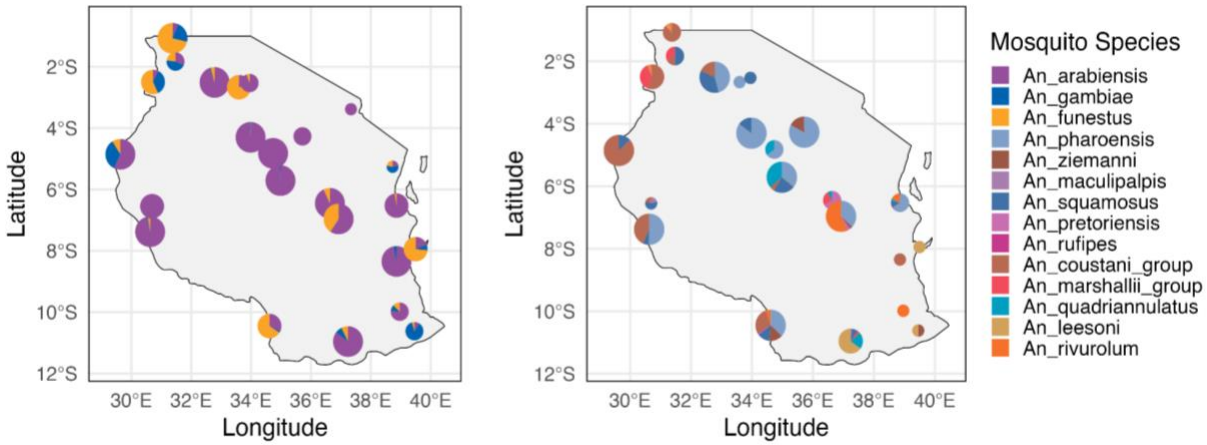


Figure 9. Spatial distribution of Anopheles species composition by district across Tanzania. Pie charts represent the proportional distribution of mosquito species detected at each site, with the size of each pie scaled to the total number of mosquitoes sampled per district (larger pies indicate higher abundance). The left panel displays the composition of primary malaria vectors (*An. arabiensis*, *An. gambiae s.s*, and *An. funestus s.s*), while the right panel shows secondary vectors and other Anopheles species.

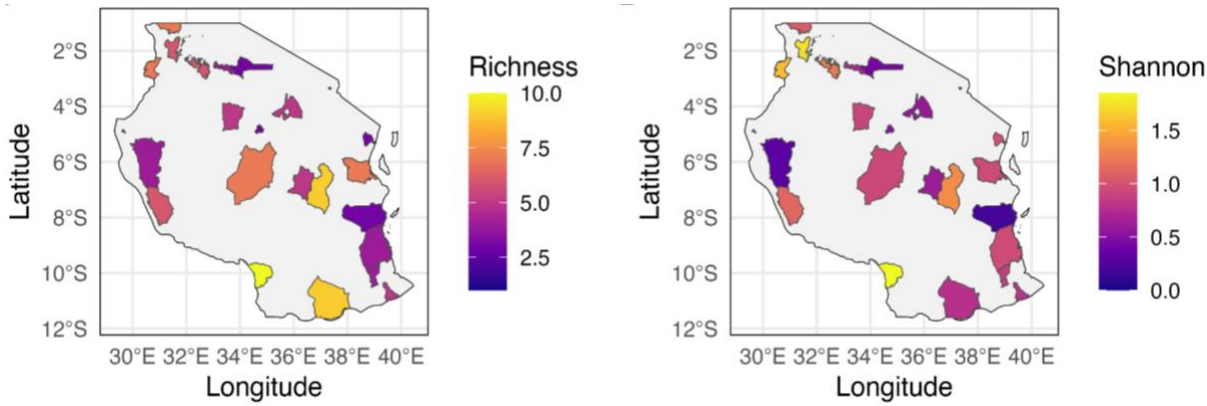


Figure 10. Anopheles species richness and diversity across surveyed districts. The left panel illustrates species richness (total number of distinct Anopheles species), while the right panel shows the Shannon diversity index, reflecting both species abundance and evenness. Higher values in both metrics indicate greater ecological complexity of mosquito populations across regions.

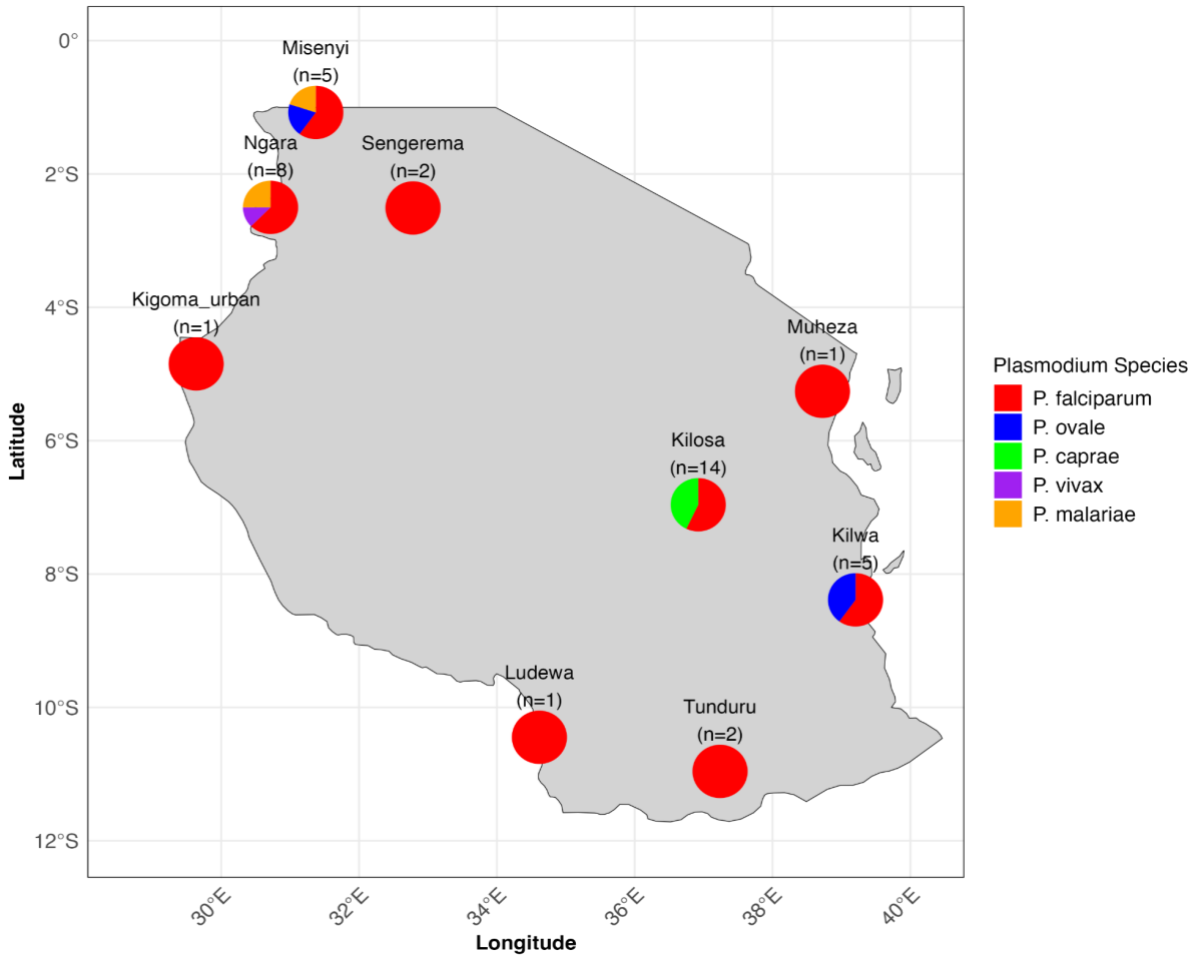


Figure 11. Spatial distribution of Plasmodium detected in mosquito vectors across surveyed districts. Each district is represented by a pie chart showing the proportional composition of Plasmodium species identified, including *P. falciparum, P. malariae, P. ovale, P. vivax,* and *P. caprae*. The figure illustrates both the geographic spread and species diversity of Plasmodium parasites circulating in vector populations.

3.6 Predictors of Species Richness

A quasi-Poisson generalized linear model (GLM) was employed to investigate the factors influencing species richness at the hamlet level, accounting for under dispersion in the count data as indicated by a dispersion parameter of 0.539. The model revealed that Shrubs land cover ($p<0.001$) and presence of livestock ($p<0.001$) were highly significant positive predictors of species richness, while mean diurnal temperature range was a significant negative predictor ($p=0.032$). Statistical significance was assessed using robust (HC3) standard errors to account for potential heteroskedasticity and confirmed by Type II analysis of deviance (ANOVA), as model selection criteria such as AIC are not applicable to quasi-Poisson models due to the absence of a true likelihood function. While other environmental and anthropogenic factors

were included in the model, they did not demonstrate a statistically significant association with species richness (Figure 12, Table 3). Diagnostic plots indicated a generally adequate model fit, with residuals scattered around zero and no critically influential observations, suggesting the model provides a reliable representation of the relationships between the predictors and species richness.

Table 2. Ecologically Relevant Environmental and Climatic Variables Included in Anopheles Species Richness (Quasi-Poisson) and Community Assembly Model (CCA)

Variable	Type	Rationale
Annual Mean Humidity (HumidityMean)	Long-term climate	Captures average background humidity influencing adult mosquito survival.
Temperature	Short-term meteorological	Reflects immediate temperature conditions during sampling; affects mosquito activity.
Humidity	Short-term meteorological	Measures real-time humidity during collection; influences mosquito flight and host-seeking.
Annual Mean Temperature	Long-term climate	Represents baseline thermal regime shaping mosquito distribution and species limits.
Annual Precipitation	Long-term climate	Provides context for rainfall-driven breeding site availability and vector persistence.
Mean Diurnal Range temperature (MDR, Mean Diurnal Range)	Long Term measure of daily temperature fluctuation	Differentiates species adapted to stable vs. fluctuating thermal conditions
Shrubs	Land cover- shrub covered areas	Associated with transitional vegetation zones influencing local microclimates.
Normalized Difference Vegetation Index (NDVI)	Vegetation across the study period	Quantifies greenness and productivity, serving as a general proxy for habitat suitability.
Cropland	Land cover - farm lands	Represents agricultural disturbance, often linked to habitat modification or irrigation.
Livestock	Presence of livestock during collection period	Reflects alternative blood meal sources and possible zoophilic mosquito attraction.
Built	Land cover: Built areas	Indicates guilt areas villages, urban or peri-urban environments; influences host availability and breeding site availability due to land modification and hence species composition.

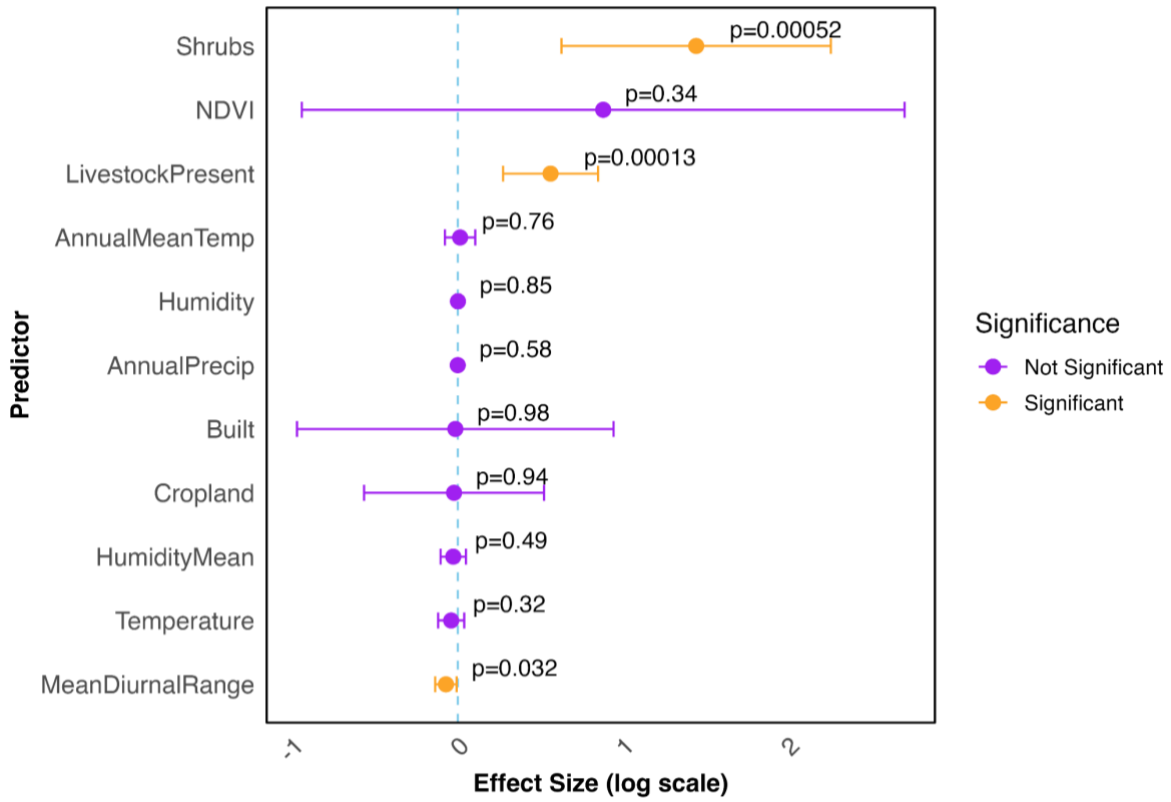


Figure 12. Forest plot of quasi-Poisson GLM predictors for mosquito species richness. The plot show log-scale estimates (dots) with 95% robust CIs (lines). Orange dots indicate $p \leq 0.05$; purple dots $p > 0.05$.

3.7 Species Co-occurrence Patterns

Pairwise co-occurrence patterns among *Anopheles* species, including primary malaria vectors such as *An. arabiensis*, *An. gambiae*, and *An. funestus*, were analysed using Fisher's exact tests (Figure 13). This analysis revealed complex ecological interactions. Among the primary vectors, *An. arabiensis* showed significant negative associations with both *An. gambiae* and *An. funestus*. Conversely, *An. gambiae* and *An. funestus* exhibited a significant positive association with each other. Associations involving *An. arabiensis* and secondary vectors indicated negative associations with species like *An. ziemanni*, but frequent co-occurrence with *An. squamosus* and *An. quadriannulatus*.

An. gambiae demonstrated a significant negative association with *An. pharoensis*. However, *An. gambiae* showed a positive association with *An. leesoni*. For *An. funestus*, negative associations were observed with *An. pharoensis* and *An. quadriannulatus*, while positive associations with *An. rivulorum* and *An. leesoni* were detected. Among secondary vectors, *An.*

pretoriensis displayed multiple strong positive associations with species such as *An. pharoensis*, *An. rivulorum*, *An. ziemanni*, *An. coustani* group, and *An. rufipes*. Similarly, *An. pharoensis* showed positive associations with several other secondary vectors, including *An. ziemanni*, *An. coustani* group, *An. squamosus*, and *An. rufipes*. Overall, the co-occurrence patterns suggest both habitat segregation and shared niche exploitation among diverse *Anopheles* species

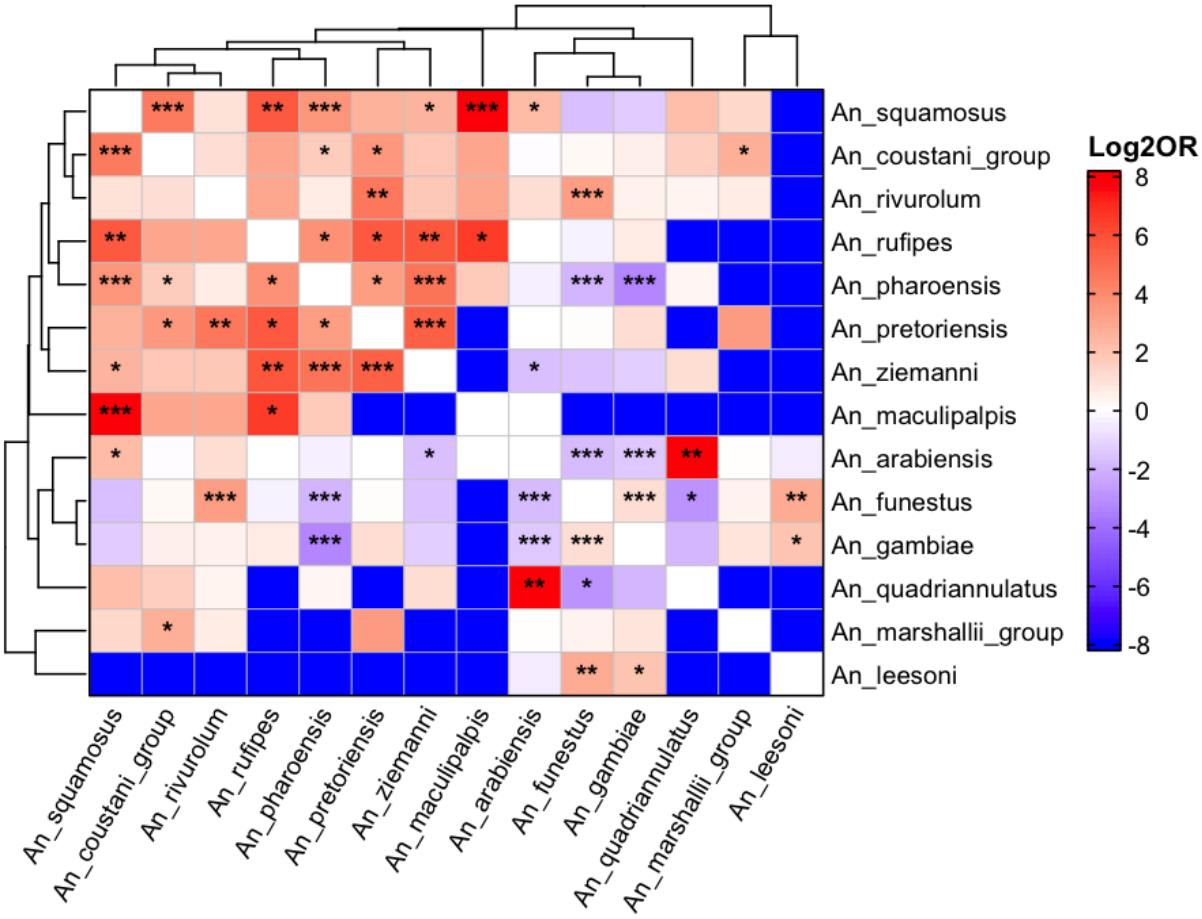


Figure 13. Heatmap of pairwise co-occurrence patterns among Anopheles species. Red and blue indicate positive and negative associations, based on scaled $\log_2$ odds ratios ($\log_2$(OR)) respectively, with colour intensity proportional to association strength. Asterisks denote statistical significance (p-values are Bonferroni corrected) ($p < 0.05$ *, $p < 0.01$ **, $p < 0.001$ ***). Hierarchical clustering groups species by similarity in co-occurrence patterns, suggesting shared habitats or ecological interactions.

3.8 Ecological Niche Partitioning of Mosquito Communities

Canonical Correspondence Analysis (CCA) of Hellinger-transformed *Anopheles* abundance data explained 14.1% of the total variance in species composition ($\chi^2 = 0.81$, $F = 6.43$, $p =$

0.001). The remaining 85.9% of variance was not explained by the model, likely due to unmeasured environmental factors, biotic interactions, collection biases, or stochastic processes. All eleven environmental predictors including short-term weather, long-term climate, land cover, and livestock presence, were individually significant ($p < 0.001$). The first four constrained axes (CCA1–CCA4) together accounted for 87.8% of the constrained variance and were all statistically significant ($p < 0.001$). The CCA biplot (Figure 14) illustrates the distribution of species and environmental variables along the first two constrained axes, highlighting the dominant ecological gradients.

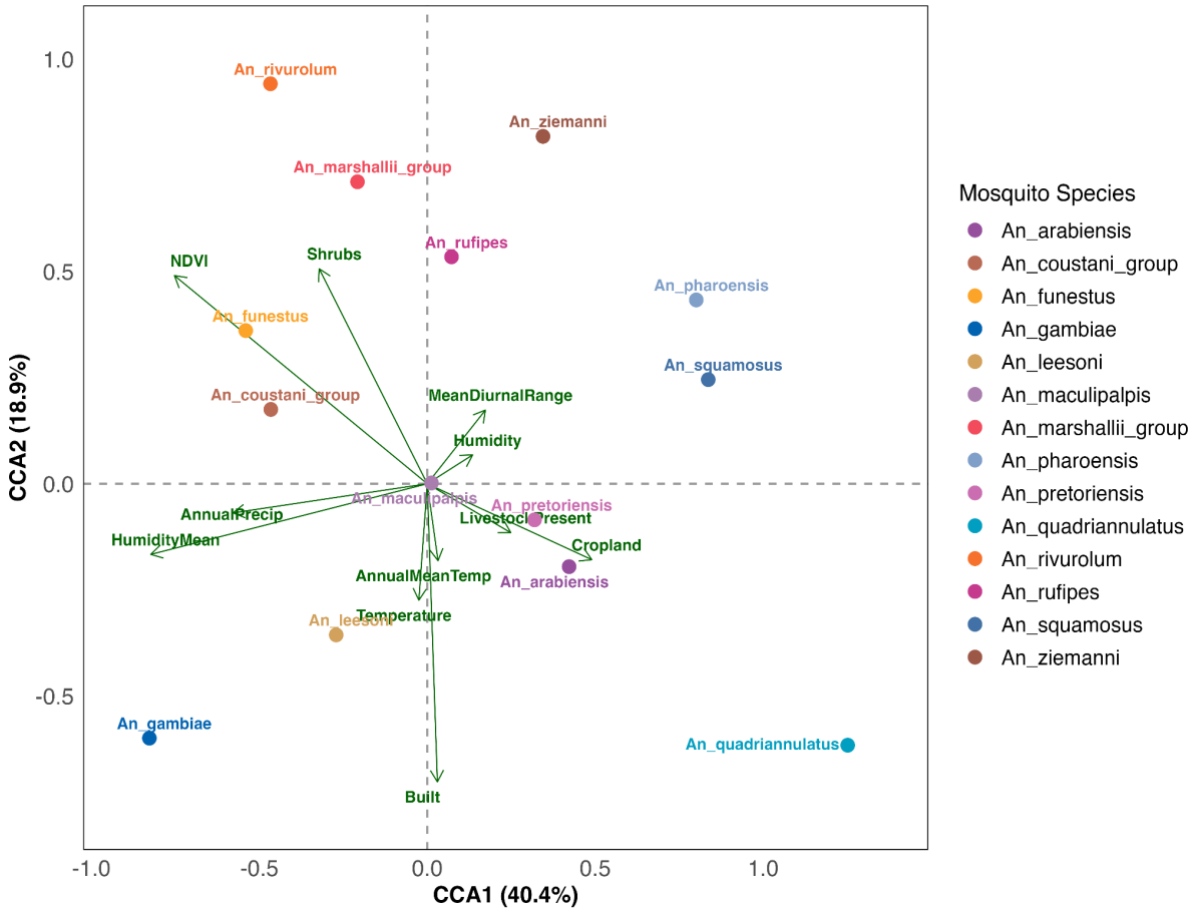


Figure 14. Canonical Correspondence Analysis (CCA) biplot. The plot shows mosquito species (points) and significant environmental variables (arrows) along the first two canonical axes. Arrow length reflects variable importance and direction indicates positive associations. Species proximity and projection onto arrows indicate the strength and direction of ecological associations. As described in the table Humidity is temporal humidity during data collection while HumidityMean is climatic humidity recorded over a long period.

To comprehensively quantify species–environment associations, cosine similarity analyses were performed using the first four constrained axes (Figure 15). Species were grouped by manual inspection of their cosine similarity patterns with key environmental variables, particularly Humidity Mean, Annual Precipitation, and Mean Diurnal Range of Temperature (MDR). 'Wet-Associated' species exhibited positive associations with humidity and precipitation and negative associations with MDR, reflecting adaptation to moist, stable environments. In contrast, 'Arid-Associated' species were characterized by positive associations with MDR and negative associations with humidity and precipitation, indicating adaptation to drier environments with greater daily temperature fluctuations. Species showing neither extreme or mixed associations were classified as 'Moderate/Transitional.' NDVI and Annual Mean Temperature served as additional interpretive variables, with association thresholds based on the sign and magnitude of cosine similarity values.

To further explore ecological differentiation, additional variables, including short-term weather conditions, vegetation cover, cropland, built environments, and livestock presence, were examined to identify potential sub-groupings and assess species' short-term environmental responses. Group cohesion was supported by Fisher's exact tests, which identified significant positive within-group and negative between-group associations (e.g., *An. arabiensis* vs. *An. quadriannulatus*, OR = Inf, p = 0.0026; *An. arabiensis* vs. *An. gambiae*, OR = 0.37, p < 0.00001).

3.8.1 Arid-Associated Group

The Arid-Associated Group, comprising *An. arabiensis*, *An. squamosus*, *An. pharoensis*, *An. pretoriensis*, and *An. quadriannulatus*, consistently showed strong negative associations with Mean Humidity (e.g., -0.81 for *An. arabiensis*, -0.61 for *An. quadriannulatus*), Annual Precipitation (e.g., -0.84 for *An. arabiensis*, -0.58 for *An. pharoensis*), and NDVI (e.g., -0.95 for *An. arabiensis*, -0.82 for *An. quadriannulatus*). They also displayed positive associations with MDR (e.g., 0.76 for *An. pharoensis*, 0.49 for *An. pretoriensis*), suggesting adaptation to arid conditions with high temperature variability.

Association with Habitat related Variables: Positive associations with Cropland (e.g., 0.86 for *An. arabiensis*, 0.95 for *An. quadriannulatus*, 0.55 for *An. pharoensis*) indicate these species may be adapted to breeding in open, agricultural water bodies such as farm-field ponds. *An. pretoriensis* (Cropland: 0.03, NDVI: -0.38) and *An. squamosus* (Cropland: 0.59, NDVI: -0.22)

showed weaker or moderate Cropland associations, suggesting potential breeding in open, sunlit habitats with less dense vegetation.

Association with Host-Related Variables: Associations with host-related variables varied within this group. *An. squamosus* (Livestock Presence: 0.94, Built: -0.03) and *An. pharoensis* (Livestock Presence: 0.11, Built: -0.40) showed positive associations with Livestock Presence and negative or near-zero associations with Built area, suggesting adaptation to feeding on livestock. *An. arabiensis* (Livestock Presence: 0.24, Built: 0.42) and *An. quadriannulatus* (Livestock Presence: 0.60, Built: 0.45) exhibited moderate positive associations with both Livestock Presence and Built area, indicating potential generalist feeding behaviour on both human and livestock hosts. *An. pretoriensis* (Livestock Presence: -0.79, Built: -0.02) showed a strong negative association with Livestock Presence, suggesting adaptation to feeding on wild hosts or in rural areas. Fisher's positive association between *An. pretoriensis* and *An. pharoensis* (OR = 10.3, p = 0.02) supports their shared ecological niche.

***Association with Temporal Variables*:** Temporal variable associations, based on instantaneous Temperature and Humidity, varied across the group. *An. squamosus* (Temperature: -0.83; Humidity: 0.95) exhibited a strong positive association with humidity and a negative association with temperature, suggesting peak activity during cooler, humid conditions, potentially corresponding to late-night or pre-dawn periods. *An. pretoriensis* (Temperature: 0.83; Humidity: -0.85) showed a strong positive association with temperature and a negative association with humidity, suggesting increased activity under warmer, drier conditions. This pattern aligns with its relatively high Mean Diurnal Range (0.49), potentially indicating crepuscular or early evening feeding behaviour when temperatures remain elevated but humidity declines. In contrast, *An. quadriannulatus* (Humidity: 0.37; Temperature: -0.07) exhibited a moderate positive association with humidity and near-neutral association with temperature, suggesting activity peaks during more humid periods, possibly at night or early morning. *An. arabiensis* (Temperature: 0.26; Humidity: 0.06) and *An. pharoensis* (Temperature: -0.16; Humidity: 0.06) exhibited near-neutral associations with both temperature and humidity, indicating temporal flexibility in activity patterns across varying weather conditions. The high Mean Diurnal Range observed for *An. pharoensis* (0.76) further supports its potential adaptation to environments characterized by substantial daily temperature fluctuations.

3.8.2 Wet-Humid Associated Group

The Wet-Associated Group, including *An. funestus*, *An. gambiae*, *An. coustani Group*, and *An. leesoni*, consistently showed positive associations with Mean Humidity (e.g., 0.91 for *An. gambiae*, 0.59 for *An. funestus*) and Annual Precipitation (e.g., 0.85 for *An. gambiae*, 0.7236 for *An. coustani Group*). They also displayed negative associations with Mean Diurnal Range (e.g., -0.91 for *An. gambiae*, -0.48 for *An. coustani Group*), suggesting adaptation to humid conditions with stable temperatures.

Association with Habitat related Variables: Positive associations with NDVI (e.g., 0.88 for *An. funestus*, 0.53 for *An. coustani Group*) and Shrubs (e.g., 0.76 for *An. funestus*, 0.53 for *An. coustani Group*), along with negative associations with Cropland (e.g., -0.96 for *An. funestus*), indicate potential breeding in vegetated wetlands or shrubby areas. *An. gambiae* (NDVI: 0.38; Shrubs: -0.06; Cropland: –0.56) and *An. leesoni* (NDVI: 0.04; Cropland: -0.26; Shrubs: -0.35) exhibited weak associations with vegetation indices, suggesting adaptation to transitional habitats such as sparsely vegetated zones or open sunlit pools. The moderately negative Mean Diurnal Range for *An. leesoni* (-0.46) further supports its affinity for humid, thermally stable environments.

***Association with Host-Related Variables*:** Host-related associations varied within the Wet-Associated Group. *An. gambiae* (Built: 0.52; Livestock Presence: -0.09) showed a strong positive association with built areas, consistent with its well-documented anthropophilic behaviour and preference for human-dominated environments. *An. coustani Group* (Livestock Presence: 0.61; Built: -0.08) exhibited a positive association with livestock presence and a weak negative association with built areas, suggesting a zoophilic inclination, likely favouring peri-domestic livestock. *An. funestus* (Livestock Presence: -0.60; Built: -0.58) showed negative associations with both livestock and built areas, potentially indicating a preference for anthropophilic feeding in less disturbed, rural environments. *An. leesoni* (Built: 0.44; Livestock Presence: -0.46) demonstrated a moderate positive association with built environments, suggesting some degree of human host preference. These patterns are further supported by significant positive co-occurrence from Fisher's exact tests, for example, *An. gambiae* with *An. funestus* (OR = 2.13, $p = 0.0009$) and *An. gambiae* with *An. leesoni* (OR = 3.97, $p = 0.026$), indicating overlapping ecological niches within humid, anthropogenic habitats.

Association with Temporal Variables: Temporal associations showed diversity. *An. coustani Group* (Temperature: -0.755, Humidity: 0.68) had a negative association with Temperature and a positive association with Humidity, suggesting activity in cooler, humid conditions, consistent with its negative Mean Diurnal Range (-0.48). *An. leesoni* (Temperature: 0.67; Humidity: -0.62) exhibited a positive association with temperature and a negative association with humidity, suggesting increased activity during warmer, drier periods within otherwise humid environments, potentially indicating evening or early night-time feeding behaviour. *An. funestus* (Humidity: –0.36; Temperature: -0.002) also showed a negative association with humidity, implying a preference for drier conditions despite its broader humid niche. In contrast, *An. gambiae* (Temperature: 0.25; Humidity: -0.16) showed relatively neutral associations, indicating temporal flexibility. The notably low Mean Diurnal Range for *An. gambiae* (-0.91) further supports its adaptation to thermally stable environments.

3.8.3 Moderate/Transitional Group

The Moderate/Transitional Group, including *An. marshallii Group*, *An. rivulorum*, *An. rufipes*, *An. ziemanni*, and *An. maculipalpis*, exhibited weak or mixed associations with climate variables (e.g., Mean Humidity: 0.14 for *An. marshallii Group*, -0.18 for *An. ziemanni*; Annual Precipitation: 0.42 for *An. marshallii Group*, 0.02 for *An. rufipes*). They also showed a variable Mean Diurnal Range (e.g., 0.54 for *An. ziemanni*, -0.31 for *An. maculipalpis*), suggesting adaptation to moderate conditions with variable temperature fluctuations.

Association with Habitat related Variables: Positive associations with shrub cover and NDVI, such as *An. rufipes* (Shrubs: 0.77; NDVI: 0.44) and *An. rivurolum* (Shrubs: 0.69; NDVI: 0.54), suggest that these species are likely to breed in vegetated or shrubby habitats. Concurrent negative or near-zero associations with cropland (e.g., –0.48 for *An. rivurolum*, -0.054 for *An. marshallii Group*) further supports a preference for more natural vegetation over open agricultural landscapes. *An. ziemanni* (NDVI: 0.15; Shrubs: 0.26) and *An. maculipalpis* (NDVI: 0.063; Shrubs: 0.23) showed weaker associations with vegetation metrics, indicating adaptation to less densely vegetated habitats, potentially including open, sunlit puddles. The relatively high Mean Diurnal Range of *An. ziemanni* (0.54) suggests physiological tolerance to temperature fluctuations, consistent with its broader habitat use. This is further supported by its strong positive co-occurrence with *An. rufipes* (OR = 54.4, p = 0.004), implying shared ecological preferences.

Association with Host-Related Variables: Host-related associations within this group generally leaned toward non-anthropogenic hosts. *An. marshallii Group* (Livestock Presence: 0.72, Built: -0.32) and *An. maculipalpis* (Livestock Presence: 0.77, Built: 0.12) showed strong positive associations with Livestock Presence, suggesting adaptation to livestock feeding. *An. maculipalpis* is classified as zoophilic (Built < 0.2 threshold) to align with Fisher's positive association with *An. rufipes* (OR = 96.4, p = 0.02). *An. rivulorum* (Built: -0.57), *An. rufipes* (Built: -0.78), and *An. ziemanni* (Built: -0.57) showed negative associations with Built area, indicating potential adaptation to feeding on wild hosts.

Association with Temporal Variables: Temporal associations were predominantly cool-active. *An. marshallii Group* (Temperature: -0.97, Humidity: 0.87), *An. rufipes* (Temperature: -0.88, Humidity: 0.59), and *An. maculipalpis* (Temperature: -0.71, Humidity: 0.86) showed negative associations with Temperature and positive associations with Humidity, suggesting activity in cooler, humid conditions, consistent with moderate Mean Diurnal Range values (e.g., 0.42 for *An. rufipes*). *An. rivulorum* (Temperature: -0.44, Humidity: 0.22) had a negative association with Temperature, indicating cool-active behaviour, while *An. ziemanni* (Temperature: -0.13, Humidity: -0.14) showed neutral associations, supported by its high Mean Diurnal Range (0.54) suggesting flexibility across temperature fluctuations and possible endurance in semi-arid regions.

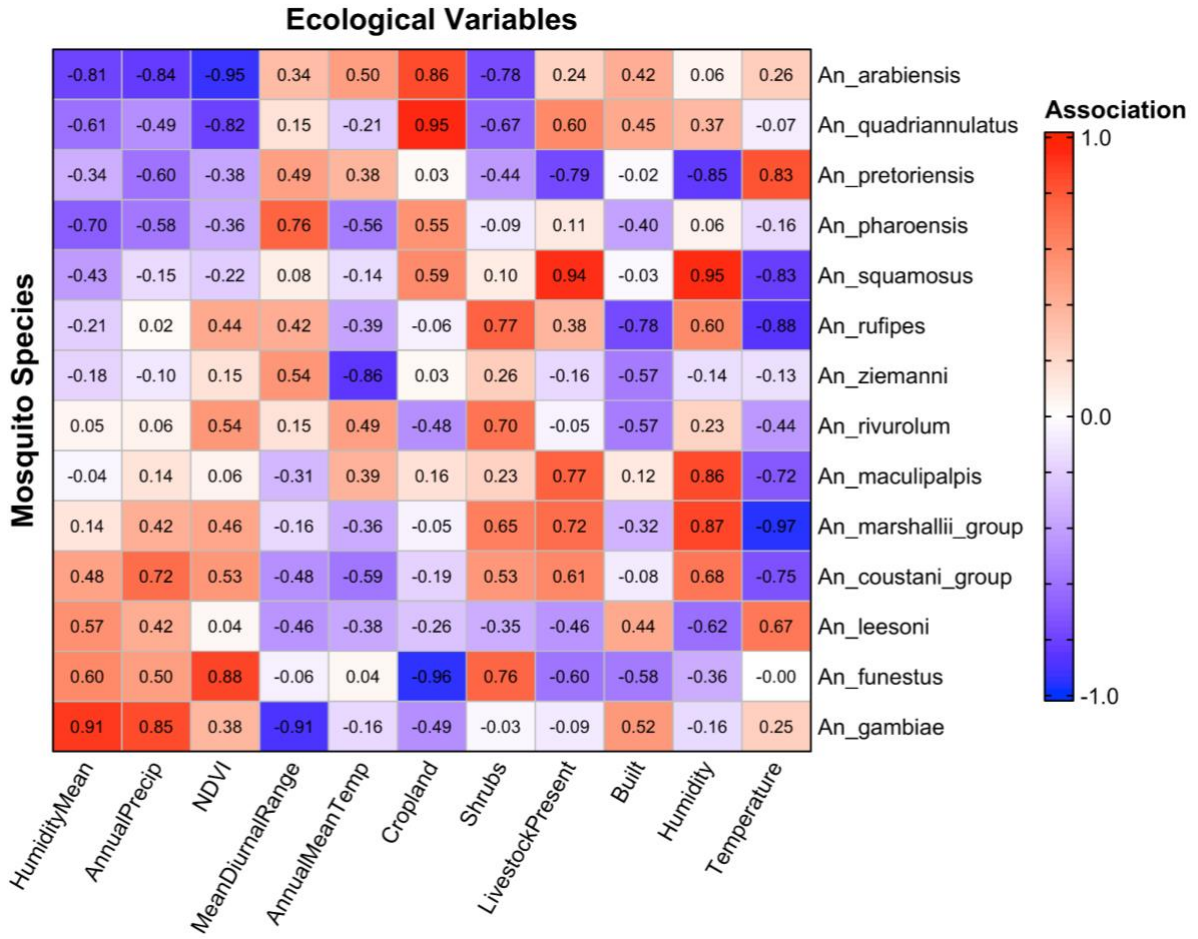


Figure 15. Heatmap of cosine similarities between Anopheles species and environmental variables from the first four CCA Axes. Red indicates strong positive associations, blue indicates strong negative associations, and white/light colours indicate weaker associations. Hierarchical clustering groups species by similarity in association patterns to the environment, suggesting shared habitats or ecological interactions.

3.9 Discussion

For precise epidemiological inference and targeted species-specific control interventions, a higher level of diagnostic accuracy is required. While widely used, conventional morphological identification of *Anopheles* mosquitoes presents significant challenges, as highlighted by our study's substantial 6.0% overall misclassification rate. Though more accurate than rates reported in other regional studies, such as 15% in Zambia [143], and 10.8% in Kenya [144], this level of precision still masks the underlying diversity within groups and remains insufficient for effective surveillance. This is because this seemingly tolerable rate masks the inherent inability of morphological keys to discriminate between species within morphologically similar groups and complexes, such as *An. gambiae* s.l. (comprising nine

species) and *An. funestus* s.l. (comprising 13 species) [56]. This leads to challenges in effective control because, despite their morphological similarity, these species often exhibit distinct behaviours and vectorial capacities, thereby requiring tailored intervention strategies [25,145]. Importantly, such morphological reliance also critically impacts downstream molecular surveillance, because traditional PCR assays, dependent on accurate initial morphological assignments, without which it can lead to non-amplification or costly re-testing if samples are misidentified [56].

Our data specifically showed species like *An. rufipes* and *An. maculipalpis* were particularly prone to misidentification, while *An. pretoriensis*, and species in the *An. marshallii* Group were entirely missed by morphology. These inaccuracies largely stem from human factors such as technician experience and the tediousness of identification in field settings. This is further amplified by limitations in surveillance systems, which often restrict the range of species detected by focusing primarily on the major malaria vectors, thereby masking the crucial and increasingly important role of secondary vectors in transmission.
[64,71,72,74].

The ANOSPP panel [96] offers a transformative solution to these challenges, providing genus-wide species resolution for *Anopheles* mosquitoes with only a simple initial morphological identification to distinguish them from other genera [98]. This reduces the need for detailed and often challenging morphological species or group-level pre-sorting, enabling the identification of virtually any *Anopheles* mosquito to species level where reference sequence data exists. In this study ANOSPP was able to identify 11 distinct *Anopheles* species and place some in their group level, with the potential for full resolution upon obtaining additional reference sequences in the databases. The panel also identified five *Plasmodium* species, including *P. falciparum*, *P. malariae*, *P. ovale*, *P. vivax*, and *P. caprae*. This simultaneous, multi-species *Plasmodium* detection capability is often overlooked in routine surveillance, which predominantly focuses on *P. falciparum*. However, identifying the full spectrum of circulating *Plasmodium* species is crucial, as co-infections can lead to more severe disease outcomes [83], and the growing threat of zoonotic malaria [146–149] demands comprehensive surveillance to truly understand transmission dynamics and safeguard against jeopardizing current malaria control achievements. This capability, along with its potential to reduce dependence on morphological keys, underscores ANOSPP's suitability for integration into routine surveillance activities. However, like all sequence-based methods, the resolution power

of ANOSPP ultimately depends on the breadth and completeness of its reference database. In cases where a species lacks a reference sequence, taxonomic assignment may be limited to the closest available group or, to the nearest series [97,98]. Nonetheless, this limitation is progressively diminishing due to the ongoing expansion of the reference database [98].

It was revealed that shrubland coverage and the presence of livestock were positively associated with higher mosquito species richness, whereas mean diurnal temperature range had a negative correlation. The presence of livestock may increase olfactory cues in the environment, thereby enhancing host detection and supporting a greater diversity of mosquito species [150]. Similarly, shrubland may contribute to the formation of favourable microclimates that enhance mosquito survival and breeding. On the other hand, the negative impact of mean diurnal temperature range on species richness suggests that environments with high temperature fluctuations, often indicative of arid conditions, may be less suitable for the survival of a broad range of mosquito species. This is consistent with findings that show such fluctuations adversely affect mosquito biology and survival [151].

The highest levels of species richness were recorded in Kilosa, an area characterized by year-round agricultural activity and wetland habitats; in Ludewa, particularly Manda ward along the shoreline of Lake Nyasa; and in Tunduru, which is also known for paddy cultivation. These land-use patterns may contribute to increased breeding site availability and resource diversity. Additionally, in both Kilosa, Ludewa and Tunduru, many mosquito species were collected from cattle sheds, further supporting the role of livestock in sustaining higher species richness. The very low species diversity observed in Moshi Urban likely reflects the significant impact of urbanization, which leads to habitat homogenization and imposes strong selective pressures favouring highly adaptable, generalist *Anopheles* species. These species can exploit the limited and altered breeding habitats characteristic of built-up environments. This may explain why only a single *Anopheles* species, *An. arabiensis* was detected in the area. This pattern aligns with previous research indicating that diverse land use in agriculture can enhance mosquito species richness [152].

Importantly, An. *gambiae* and *An. funestus* were more consistently found in high malaria transmission zones, whereas *An. arabiensis* was widespread and detected across much of the country. Notably, no *Anopheles* mosquitoes were recorded in Iringa Municipality. The co-occurrence of *An. gambiae* and *An. funestus* with high transmission zones is consistent with

their high vectorial capacity, while *An. arabiensis* appears to sustain residual transmission in low-transmission settings [65]. Regarding parasite distribution, *P. falciparum* was found in all nine districts where *Plasmodium* was detected. Moreover, districts such as Ngara, Misenyi, and Kilwa identified as high-transmission areas [62] with elevated malaria prevalence and mortality rates were also hotspots for *Plasmodium* species diversity. In these locations, multiple species were detected, including *P. falciparum*, *P. malariae*, *P. ovale*, and *P. vivax*, a pattern that may contribute to challenges in malaria case management due to mixed-species infections [83].

A particularly novel finding was the detection of *P. caprae*, a parasite typically associated with goats, in *An. arabiensis*. Although *P. caprae* has previously been identified in mosquitoes such as *An. subpictus* and *An. aconitus* [153], this is the first report of its presence in *An. arabiensis*. Moreover, while *P. caprae* has been documented in both mosquitoes and goats in parts of Asia and several African countries, this represents its first detection in Tanzania [153–155]. Although *P. caprae* has not been detected in humans and is considered restricted to non-human hosts, its detection in a primary human malaria vector demonstrates the capacity of *An. arabiensis* to harbour a broader range of parasites than previously recognized, with important implications for understanding its transmission ecology. This finding may indicate interactions at the livestock-human interface or demonstrate the vector's capacity to harbour non-human *Plasmodium* species. While there is currently no direct evidence of zoonotic transmission, these observations highlight the need for enhanced entomological and molecular surveillance. Ultimately, the detection of *P. caprae* in this context reinforces the importance of adopting a One Health surveillance framework that integrates data across human, animal, and vector domains to better understand the ecology and transmission dynamics of malaria and related parasites.

Using canonical correspondence analysis and pairwise chi-square co-occurrence analysis, we were able to partition the sampled *Anopheles* species into potential ecological niche groups. This classification was based on climatic and land cover variables derived from satellite remote sensing data, as well as temporal data collected concurrently during field sampling. These groups were defined as the Arid-Adapted Group, Humid-Adapted Group, and Ecologically Flexible Group, further subdivided within each group based on habitat, potential host and temporal weather variables which gave a glimpse of how their general activities are shaped by instantaneous change in temperature and humidity. The Arid-Adapted Group, including *An. arabiensis*, was associated with dry areas and agricultural zones characterized by marked

temperature variability, as well as the presence of livestock and built environments. These patterns confirm its generalist nature in host preference, as reported in previous studies [91]. This finding aligns with earlier reports from East Africa showing the species' association with irrigation systems and farm areas, which facilitate its survival in arid conditions [156,157], and its predominantly zoophilic behaviour [158]. This understanding informs intervention design; for instance, vector control in such contexts should emphasize LSM targeting agricultural water bodies and the use of insecticide-treated livestock to intercept zoophilic vectors [159–161].

The Humid-Adapted Group including species such as *An. gambiae,* showed clear associations with mean humidity and annual rainfall, and a negative association with mean diurnal temperature range, conditions typical of more arid regions. This concurs with earlier studies confirming its affinity for humid areas [157,162]. Its association with built environments also supports its well-established anthropophilic behaviour. Similarly, *An. funestus* was associated with vegetation-dominated land cover and negatively associated with built environments, confirming prior reports of its preference in rural, vegetated habitats [163,164], its sensitivity to temperature fluctuations [151] and its varying zoophilic and anthropophilic tendencies [165].

Meanwhile, species within the Ecologically Flexible Group, such as *An. rufipes* and the *An. marshallii* group, exhibited broad environmental tolerance, occupying both natural and anthropogenic habitats across diverse climatic zones. This adaptability may undermine the effectiveness of static intervention strategies and underscores the need for localized seasonally adaptive tools such as spatial repellents or community-driven larval surveillance. However, it should be acknowledged that species associations with levels of aridity or humidity are context-dependent within the Tanzanian landscape. Given that this study was limited to samples from 25 villages across 25 districts, it may not fully reflect the underlying ecological complexity. Notably, even within a single village, different species may occupy distinct ecological niches [166].

Nonetheless, our findings demonstrate that well-designed ecological modelling, supported by robust data, can generate reliable insights for informing vector control. Additionally, partitioning species into ecologically similar groups offers a practical advantage: a representative "signature species" could serve as a proxy to infer the presence of other group members, particularly in resource-limited surveillance settings. Although members of these ecological groups may share general associations with humid or arid regions, they often differ

significantly in their habitat and host-related variables. Consequently, interventions should be tailored not on the group level but at the subgroup or species level depending on the pattern in sharing characteristics as well articulated in the result section. For example, contrasting associations with vegetation-related variables between *An. gambiae* (negative) and *An. funestus* (positive), along with differences in temporal patterns, suggest distinct breeding site preferences and weather-driven biting behaviours. These ecological nuances offer actionable entry points for species-specific targeting within broader intervention frameworks.

Our classifications align with existing research for major vectors. *An. gambiae* and *An. funestus* are well-documented in humid areas, with human-feeding behaviour, consistent with their **Humid-Adapted Group** placement [167,168], and vegetated habitats subgroup placement for funestus [164,165]. *An. arabiensis* is recognized in drier, agricultural areas with animal-feeding tendencies, supporting its **Arid-Adapted Group** classification [91,157,158]. Notably, the *An. coustani* group's positive association with vegetated areas and its zoophilic nature aligns with findings from northern Tanzania and southern Kenya, where *An. coustani* larvae were predominantly found in short grass habitats (53.6%) and tall grass habitats (45.7%), compared to only 0.7% in open sunlit pools [169]. For less-studied species such as *An. leesoni* and *An. rivulorum*, our study provides novel ecological insights. *An. leesoni* showed a clear preference for built, humid areas, a negative association with livestock presence, and increased activity during warmer periods. These patterns suggest a possible tendency toward early evening biting behaviour, and anthropophilic behaviour which warrants further investigation. Similarly, the associations of *An. rivulorum* and *An. rufipes* with shrub-dominated landscapes and higher NDVI values indicate a likely preference for vegetated microhabitats, particularly in humid zones. These findings highlight potential breeding site characteristics and behavioural adaptations that have not been well described in the literature, underscoring the need for targeted ecological studies to inform surveillance and control strategies for these understudied vector species.

This study is not without limitations. Mosquito sampling was evenly distributed across 25 districts, but logistical challenges, such as limited accessibility, weather conditions, and resource constraints, may have introduced spatial and seasonal sampling biases. Although species identification was validated using both morphological keys and molecular diagnostics (ANOSPP panel), trap-based methods are inherently selective and may under-sample cryptic taxa, microhabitat specialists, or species with divergent activity patterns. Furthermore,

Detection of *Plasmodium* DNA using the ANOSPP assay confirmed parasite presence but does not confirm actual transmission potential. DNA may reflect recent blood meals rather than sporozoite presence in salivary glands, which is essential for transmission. Thus, parasite detection alone does not confirm vector competence. In addition, some species such as *An. rufipes* and *An. maculipalpis* were captured in very low numbers, and further targeted surveillance will be needed to clarify their ecological niches and epidemiological relevance.

Despite these limitations, the study offers practical and policy-relevant insights by delineating spatial patterns of vector distribution and *Plasmodium* presence across ecologically diverse settings, providing a critical foundation for targeted surveillance, adaptive vector control, and future investigations into transmission dynamics and ecological niche shifts.

Chapter Four: *Anopheles* Species Distribution Modelling in Tanzania Under Current Climatic Scenario

Chapter Summary

Traditional entomological surveillance for malaria vectors remains resource-intensive, time-consuming, and spatially constrained, limiting the timely identification of high-risk areas for targeted intervention. This challenge underscores the need for predictive, spatially explicit tools to optimize vector control strategies. Species Distribution Models (SDMs) address this gap by integrating limited occurrence data with environmental predictors to generate high-resolution, spatially continuous estimates of vector habitat suitability. In this chapter, SDMs were developed for the three primary malaria vectors, *An. arabiensis*, *An. gambiae* s.s, and *An. funestus* s.s by combining occurrence records from this study, other research projects, published literature, and WHO threat maps metadata. To improve model robustness, ecologically informed pseudo-absence points were generated using temperature suitability envelopes and forest canopy cover. Key environmental and bioclimatic variables, including precipitation, temperature, and land cover, were incorporated at a 1 km spatial scale. Generalized Additive Models (GAMs) were applied to capture smooth nonlinear responses, and model performance was evaluated using standard diagnostic metrics. The resulting habitat suitability maps align with known malaria transmission zones and reveal both overlapping and species-specific environmental associations, providing actionable spatial insights to guide targeted surveillance and resource allocation in Tanzania's vector control programs.

4.1 Observed Species Presence Data and Spatial Distribution

Occurrence data for *An. gambiae*, *An. funestus*, and *An. arabiensis* were compiled from different sources to maximize spatiotemporal coverage across the study area. This dataset integrates records from different research projects, national surveillance efforts, and publicly available databases. All unique data points were georeferenced and associated with a recorded year of collection. The distribution of *Anopheles* species occurrence records (presence GPS data) by their original source and year of collection is summarized in Supplementary Table A1. In total, 1,040 raw occurrence records were compiled across the three species. *An. arabiensis* data (579 records) spanned the longest period (2011-2024), with early contributions from WHO Threat Maps metadata and more recent data from 25 districts for this project and Dhibiti (control) Malaria project from Ifakara Health Institute (IHI) that was conducting insecticide

resistance monitoring across 22 districts across the country. *An. funestus* (239 records) and *An. gambiae* (222 records) datasets primarily consisted of more recent observations (2017-2024), with significant contributions from this project for both species, and Dhibiti Malaria Project for *An. gambiae*. Data from Deogratius Kavishe, Kavishe et al. 2025 [170], Edmond Bernad (NIMR), Matowo et al. 2021 [162], and Nambunga et al. 2020 [164], along with other data from Dr. Fedros Okumu (IHI) for *An. funestus*, provided additional recent records for various species. National surveillance data contributed consistently across multiple years for all three vectors [65]. Furthermore, data not linked to specific citations represent unpublished records contributed directly by the respective institutions and project leads. The geographical distribution of these diverse, compiled data points across the study area is illustrated in Figure 16. Following compilation, all records (presence data) were subjected to quality control by removing duplicates and spatially thinning presence locations to one record per 1 km cluster using the *geosphere* and *igraph* packages in R [171,172], resulting in 233 unique records for *An. arabiensis*, 102 for *An. gambiae*, and 79 for *An. funestus*, with spatial resolution matching that of environmental predictors

4.2 Pseudoabsence Data Generation and Spatial Distribution

Ecological suitability for *An. gambiae*, *An. funestus*, and *An. arabiensis* was quantified using a Temperature Suitability Index (TSI) parameterized from established physiological and ecological thresholds [81,121,151,173–181]. For each species, TSI rasters were constructed as composite functions of coldest-quarter mean temperature, relative humidity, precipitation of the driest quarter, mean diurnal range of temperature (MDR), and land cover characteristics including cropland, built-up area, closed-canopy forest, and wetlands [164,165,182–184]. The thermal response curves and associated penalties were species-specific, with ecological realism ensured through the integration of modifiers for aridity, humidity, seasonality, anthropogenic habitats, and refugial buffers, reflecting both empirical and recent field-based evidence.

For *An. arabiensis*, a TSI, was constructed based on its thermal, humidity, aridity, and seasonality tolerances. Optimal suitability (TSI = 1) was assigned for mean coldest quarter temperatures between 25–30°C, declining linearly to zero at 15°C and 38°C. Higher minimum humidity extended tolerance to warmer temperatures. A penalty was applied by capping TSI at 0.2 in highly arid regions, defined as areas with a mean diurnal temperature range greater than 12 °C and precipitation in the driest quarter below 20 mm, as well as in locations with extreme

temperature seasonality exceeding 155. The effect of this penalty was partially compensated by the presence of cropland. The final unsuitability mask for *An. arabiensis* combined regions where TSI was less than 0.2 with areas of dense closed forest (tree cover ≥0.9), due to its preference in sunlit open habitats [121,182,185,186].

The pseudo-absence mask for *An. gambiae* utilized a more robust, multi-factorial TSI reflecting its nuanced ecological preferences. Optimal thermal suitability (TSI = 1) occurred between 25–28°C mean coldest quarter temperature, with non-linear transitions to zero at 16°C and 35°C. This core thermal suitability was modulated by a logistic penalty for low minimum humidity in warmer ranges, and a dynamic aridity penalty (for driest quarter precipitation <40 mm) that was partially offset by the presence of cropland. Additional adjustments included a penalty for extreme temperature seasonality, a ±0.5°C evolutionary adaptation buffer at thermal limits, and a Gaussian-shaped suitability boost in peri-urban areas (8–30% built-up land). The final *An. gambiae* unsuitability masks integrated areas where TSI was less than 0.2 and closed forest (tree cover ≥0.9) [121,182,185,186].

For *An. funestus*, the unsuitability mask was derived from a TSI that emphasized its association with stable aquatic habitats and sensitivity to environmental extremes. Its temperature suitability based on the mean coldest quarter temperature was optimized between 25–28°C (TSI = 1), with its upper limit of 35°C strongly dependent on minimum humidity. Additional factors included an aridity penalty the same used to *An. gambiae* (adjusted by cropland presence), a penalty for high mean diurnal temperature range, and a significant logistic decline in suitability with increasing built-up land, reflecting its strong rural preference. Conversely, a suitability boost was applied in wetlands. The final *An. funestus* unsuitability mask combined areas with a TSI less than 0.2 and closed forest (tree cover ≥0.95) [121,164,165].

Pseudo-absence points were generated by overlaying TSI-based ecological suitability thresholds (TSI < 0.2) with exclusion of closed-canopy forest, using species-specific thresholds. Spatial binary masks were validated for congruence with environmental predictor rasters, and random pseudo-absence points were sampled exclusively from unsuitable areas, avoiding ambiguous or missing data regions. Specifically, 50 pseudo-absence points were generated for *An. gambiae*, 50, for *An. arabiensis*, and 40 for *An. funestus* (see figure 16 for spatial distribution). For each pseudo-absence, the full suite of environmental covariates was

extracted to mirror the structure of the presence dataset, and each point was annotated with its exclusion criterion (low TSI or closed forest) to facilitate downstream model validation.

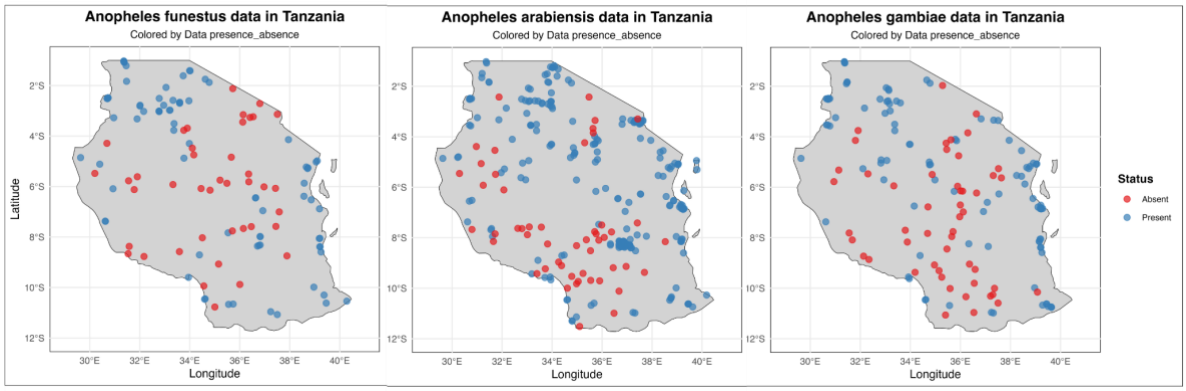


Figure 16. Spatial distribution of presence and pseudo-absence records used for species distribution modelling. Map showing georeferenced input data used to train habitat suitability models for malaria vectors in Tanzania. Blue points represent known presence locations derived from entomological surveys or confirmed species records. Red points represent pseudo-absence locations, selected using ecological exclusion criteria based on established physiological and environmental constraints for each species. These points were used to improve model discrimination and reduce sampling bias during Model fitting.

4.3 Environmental Covariate Patterns and Predictors Selection

Predictor selection began with the acquisition of environmental, climatic, and demographic variables known or hypothesized, based on published literature, mosquito biology, and broader insect ecology, to influence habitat suitability (Chapter 2, Section 2.2.5) [81,121,151,173–181]. The initial inclusion of these variables was therefore guided by established ecological reasoning and data availability rather than subjective preference. However, all subsequent screening and refinement steps followed a statistics-driven workflow, with ecological interpretation reapplied only after objective statistical criteria had been satisfied (Figure 17). During the first screening phase, multicollinearity was assessed using Pearson pairwise correlation and Variance Inflation Factors (VIF). Variables exhibiting strong collinearity ($|r| > 0.7$ or VIF > 5) were excluded to enhance model stability and reduce redundancy. Within each correlated pair, the variable deemed most ecologically relevant to *Anopheles* habitat representation was provisionally retained. In certain cases, variables excluded at this stage were later reintroduced if doing so demonstrably improved model performance, replacing earlier alternatives where statistical support was stronger.

After preliminary filtering, species-specific models were evaluated and, where necessary, refined through a structured and fully objective backward elimination procedure. This step was not applied arbitrarily but was undertaken only when diagnostic outcomes indicated the need for further model optimization. Predictor inclusion or removal was governed entirely by statistical evidence and diagnostic validity. Each iteration followed two sequential criteria. First, variables were assessed for their statistical contribution by examining changes in the Akaike Information Criterion (AIC); a variable was removed only if its exclusion reduced AIC, thereby improving model parsimony without degrading fit. Second, the simplified model was re-evaluated using the DHARMa package[128] to test for residual dispersion, zero-inflation, and spatial autocorrelation (Section 4.4). Refinement proceeded only when diagnostic results justified it, and a predictor was eliminated solely when both statistical and diagnostic conditions were satisfied. This procedure occasionally led to the reinstatement of variables previously discarded during collinearity screening, where their inclusion improved AIC or resolved diagnostic issues introduced by their alternatives. Such reintroductions were justified strictly by empirical model improvement. The elimination process continued until all diagnostic assumptions were met.

Ecological interpretation was reapplied only after the final models had been statistically optimized, to verify that the retained predictors (based on response curves) were biologically & ecologically coherent. This interpretive step did not influence model composition but confirmed that the statistically supported variables aligned with established ecological understanding. All retained predictors contributed to improved model performance and diagnostics. Most were also statistically significant consistent with known ecological patterns. For instance, livestock density was retained and statistically significant only in *An. arabiensis* models, reflecting its zoophilic behaviour and association with livestock-rich habitats. Similarly, isothermality was statistically significant in *An. gambiae s.s.* models, aligning with its sensitivity to stable temperature regimes. In contrast, Mean Diurnal Temperature Range (MDR) in *An. funestus* did not reach statistical significance but was retained because its inclusion improved model performance and diagnostic outcomes, suggesting a plausible, though weaker, influence of thermal variability on its distribution[151]. The final predictor sets therefore represent statistically optimized models that are also ecologically interpretable. A summary of all candidate variables retained predictors, and their ecological rationale is presented in Table 4.

Table 3. Environmental and anthropogenic predictors used in habitat suitability modelling and their ecological rationale

Predictor Variable	*An. arabiensis*	*An. gambiae*	*An. funestus*	Ecological Justification
Annual Mean Temperature	✓	✓	✓	Fundamental driver of mosquito physiology, development rate, and geographic range [177]
Mean Diurnal Temperature Range (MDR)			✓	Captures daily temperature fluctuation; *An. funestus* is sensitive to thermal variability [151].
Isothermality		✓		Reflects temperature stability [187]
Annual Precipitation	✓	✓	✓	Determines the availability and persistence of aquatic breeding habitats [81]
Mean Relative Humidity	✓	✓	✓	Influences adult longevity and host-seeking [179].
Cattle Density	✓			Indicates availability of zoophilic blood sources; highly relevant for *An. arabiensis* [91].
Cropland Cover	✓	✓	✓	Proxy for anthropogenic larval habitats, especially in irrigated/agricultural landscapes [184,188].
Population Density		✓	✓	Proxy for human host availability and intensity of human settlement. Not retained for other species due to weaker predictive performance.[189]
Built Area Cover	✓			Captures the extent of physical habitat modification; relevant for species utilizing peri-domestic structures for resting, breeding, or shelter. Not retained for other species due to weaker predictive performance.[190]
Tree Cover	✓	✓	✓	Influences microclimate, shade, and habitat structure; key for resting and larval sites [81].

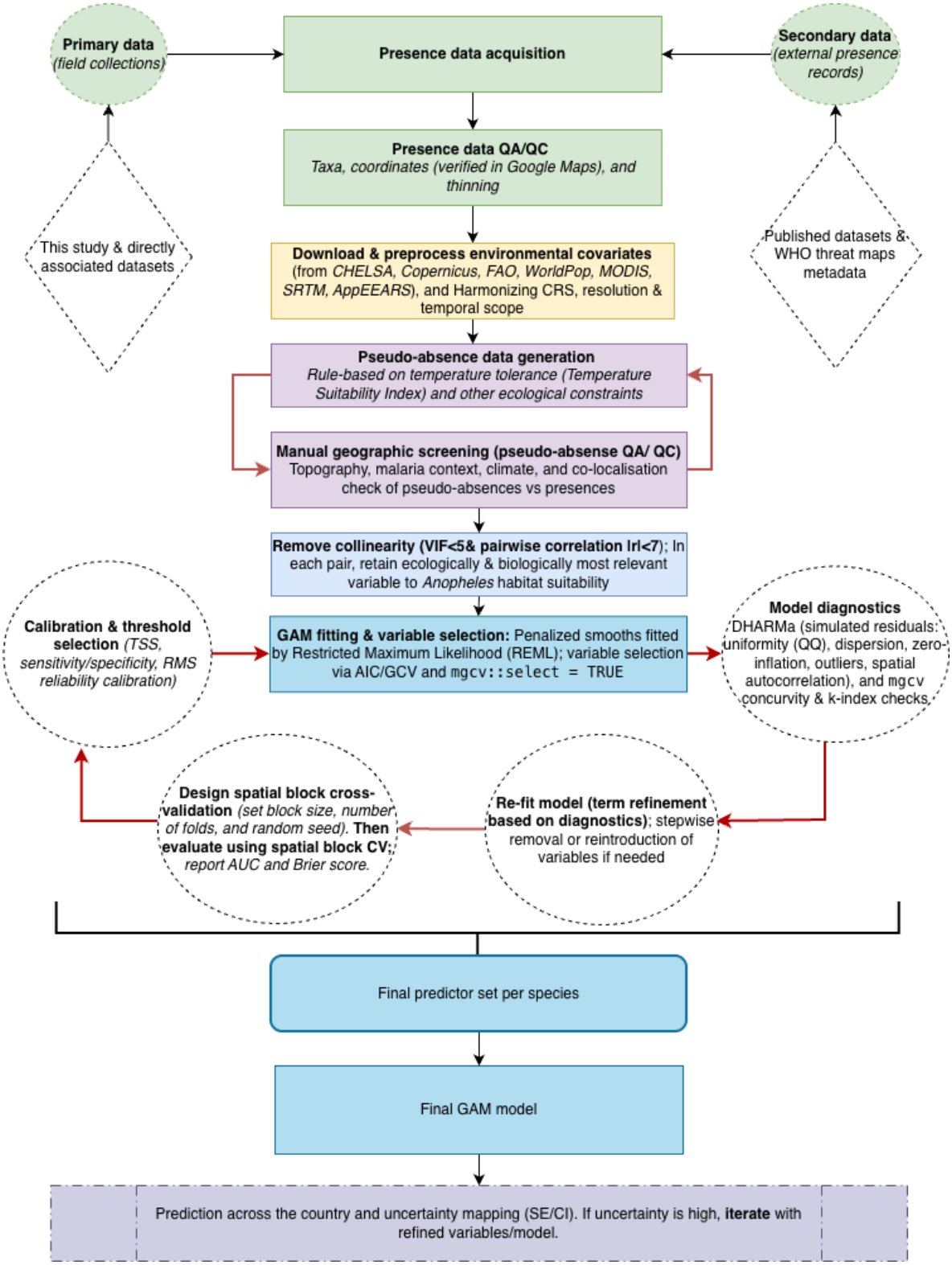


Figure 17. Workflow for species distribution modelling of *Anopheles* mosquitoes. The diagram outlines the sequential steps from field and secondary data acquisition through environmental covariate preprocessing, pseudo-absence generation, model fitting, diagnostics, and predictive mapping.

4.4 Model Fitting, Diagnostics, and Cross-Validation

Binomial Generalized additive models (GAMs) [191] were independently fitted for each species to quantify relationships between species presence and environmental predictors. Model fitting utilized the **mgcv** package in R [127], employing restricted maximum likelihood (REML) estimation and penalized smoothing splines to capture non-linear effects. Residual diagnostics were performed using the DHARMa package [128], with formal tests for dispersion, zero-inflation, and spatial autocorrelation (Moran's I) conducted on simulated residuals. The diagnostics revealed no significant overdispersion (dispersion statistic: *An. arabiensis* = 0.643, p = 0.056; *An. gambiae* = 0.685, p =0.072; *An. funestus* = 0.685, p = 0.064), zero-inflation, or spatial autocorrelation in the residuals for any species. Final model formulas included only predictors that passed multicollinearity thresholds, exhibited strong model fit in diagnostic evaluations, demonstrated high predictive performance in cross-validation, and showed good calibration. Visual assessment of spatial predictions was also used to confirm ecological plausibility in the context of Tanzania's known topography.

Model selection and validation were conducted using 10-fold spatial block cross-validation implemented via the *blockCV* package [129]. For each fold, models were trained on spatially separated data and evaluated on withheld blocks to reduce spatial autocorrelation between training and test sets, thereby preventing overestimation of predictive performance. Predictive performance was quantified using the area under the curve (AUC) and Brier score, with means and standard deviations reported across folds. The models demonstrated strong discriminative power: *An. arabiensis* (AUC = 0.962 ± 0.042), An. *gambiae* (AUC = 0.924 ± 0.065), and *An. funestus* (AUC = 0.830 ± 0.150). Corresponding Brier scores were low, indicating well-calibrated predictions: *An. arabiensis* (0.039), *An. gambiae* (0.097), and *An. funestus* (0.160). Calibration of predicted probabilities was assessed using bootstrap-corrected calibration curves (B = 1,000) implemented via the *rms* package [130], which showed close agreement between predicted and observed probabilities for all species, with mean absolute errors of 0.017 (*An. arabiensis*), 0.043 (*An. gambiae*), and 0.045 (*An. funestus*), confirming minimal deviation from perfect calibration.

Threshold-dependent performance was assessed by calculating sensitivity, specificity, and the True Skill Statistic (TSS) across probability thresholds ranging from 0.1 to 0.9 in 0.1 increments. At each threshold, predicted probabilities from the GAMs were binarized and

evaluated against observed presence–absence labels using confusion matrices. TSS was computed as *sensitivity + specificity − 1*, with the highest TSS value indicating the optimal discrimination threshold for each species. While final predictions were retained as continuous probabilities, these threshold-based metrics served to evaluate each model's classification capacity. TSS values ranged from approximately 0.53 to 0.86 across species, with sensitivities between 0.59 and 0.96 and specificities from 0.61 to 0.95, confirming that all models exhibited strong potential for binary discrimination if required in applied contexts. Collectively, these results indicate that the selected models are both statistically robust and ecologically meaningful, providing well calibrated, interpretable predictions of malaria vector occurrence.

4.5 Environmental Drivers and Variable Importance

The final model for *An. arabiensis* (deviance explained = 78.3%, adjusted $R^2 = 0.849$) revealed distinct, linear and nonlinear responses to environmental gradients (Figure 18). Habitat suitability increased nearly linearly with annual mean temperature across the observed range up to 26°C ($p < 0.001$), indicating a linear increase in suitability across the observed gradient. Tree cover also had a significant effect ($p < 0.001$), with suitability highest below 30–40% canopy density and declining sharply beyond that threshold, suggesting a preference for open landscapes with limited overhead vegetation. Cattle density was positively associated with suitability ($p = 0.024$), particularly across the range of 0 to 500 cattle per km², aligning with known zoophilic feeding behaviour. Annual precipitation ($p = 0.046$) showed a unimodal effect, with predicted suitability peaking around 1,500 to 1,700 mm/year, indicating an optimal moisture window likely related to larval habitat availability. Although built area cover did not reach conventional statistical significance ($p = 0.089$), it displayed a positive association with suitability at low to moderate levels. This trend is ecologically plausible given the peri-domestic habitat preferences of *An. arabiensis*, and it may reflect a true but underpowered relationship due to sampling limitations. In contrast, cropland cover and mean relative humidity were not significant predictors ($p > 0.5$) and exhibited flat response curves across their observed ranges.

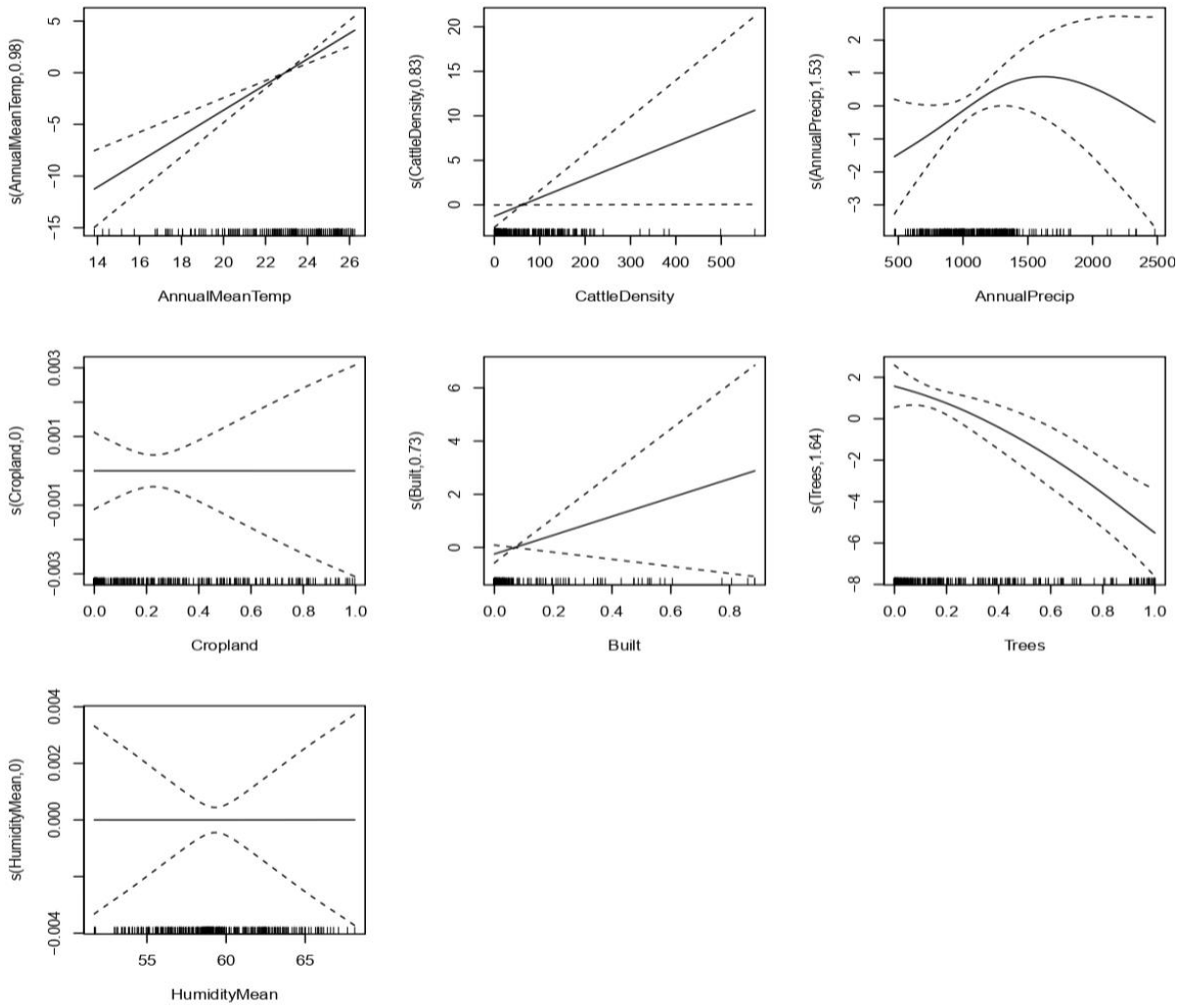


Figure 18. Partial dependence plots for environmental predictors of habitat suitability for *Anopheles arabiensis.* Each plot shows the partial effect of a single predictor on the predicted habitat suitability of An. arabiensis as estimated by the Generalized Additive Model (GAM). The solid line represents the smoothed effect (centred to a mean of zero), while the dashed lines indicate ±2 standard errors (confidence intervals). Rug marks along the x-axis show the distribution of observed values for each predictor, indicating regions with strong model support.

The final model for *An. gambiae* (deviance explained = 68.6%, adjusted $R^2 = 0.762$) revealed distinct, linear and nonlinear responses to environmental gradients (Figure. 19). Habitat suitability increased nearly linearly with annual mean temperature across the observed range up to 26 °C ($p < 0.001$), indicating a linear increase in suitability across the observed gradient, and showed a dome-shaped relationship with annual precipitation, peaking at intermediate levels (~1,500–2,000 mm/year) before declining steeply above 2,500 mm ($p < 0.001$). Suitability was largely flat at low Isothermality values but increased sharply between ~0.65

and 0.80, plateauing at higher values ($p = 0.005$), indicating a preference for environments with moderate-to-high thermal stability. While mean relative humidity (50–70%) was not statistically significant ($p = 0.213$), a slight positive trend was observed. Population density exhibited a weakly unimodal effect, with suitability increasing through moderate densities (~10,000–20,000 persons/km²) and then declining gently at the highest values ($p = 0.014$). Suitability peaked at moderate tree cover (~20–30%), declining sharply above ~40% ($p = 0.005$), and was highest in landscapes with 60–70% cropland, dropping steeply beyond this threshold ($p = 0.029$). Collectively, these patterns indicate that *An. gambiae* is most likely to occur in warm, moderately wet, thermally stable environments with a mosaic of cropland and moderate tree cover landscapes typical of rural to peri-urban settlements and agro-ecological transition zones.

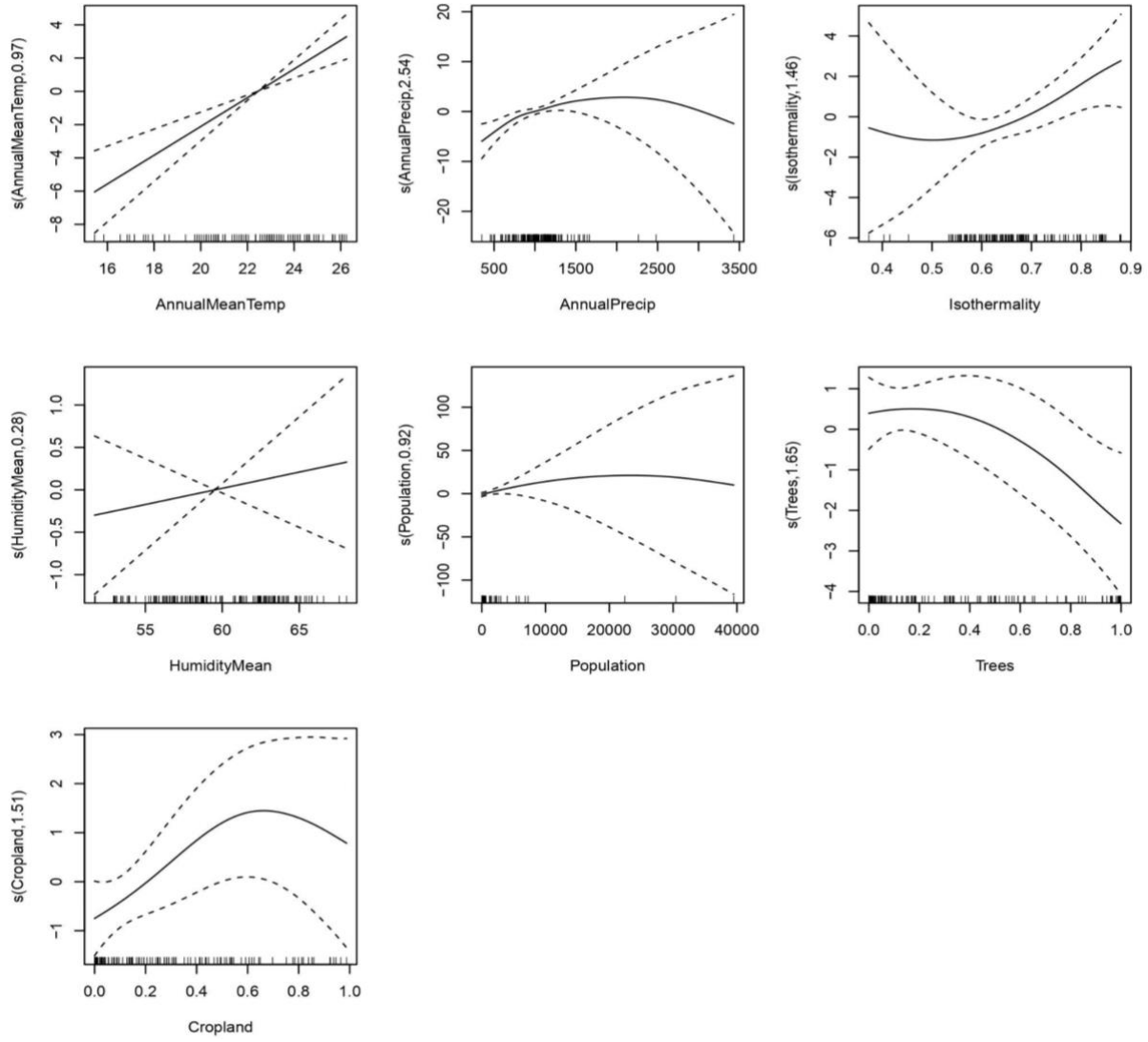


Figure 19. Partial dependence plots for environmental predictors of habitat suitability for *Anopheles gambiae*. Each plot shows the partial effect of a single predictor on the predicted

habitat suitability of *An. gambiae* as estimated by the Generalized Additive Model (GAM). The solid line represents the smoothed effect (centred to a mean of zero), while the dashed lines indicate ±2 standard errors (confidence intervals). Rug marks along the x-axis show the distribution of observed values for each predictor, indicating regions with strong model support.

The final model for *An. funestus* (deviance explained = 64.0%, adjusted $R^2 = 0.690$) identified five significant linear and nonlinear environmental predictors of species presence (Figure. 20). Habitat suitability increased nearly linearly with annual mean temperature up to 26 °C ($p < 0.001$), indicating a strong preference for warmer conditions. Tree cover ($p = 0.0009$) exhibited a unimodal relationship, with suitability peaking at moderate cover (~0.3 fractional cover) and declining sharply at higher levels, suggesting that dense vegetation may limit occurrence. Annual precipitation ($p = 0.0023$) followed a dome-shaped pattern, with suitability highest at intermediate rainfall levels and decreasing at both low and high extremes. Population density ($p = 0.0067$) showed a weak unimodal effect, with suitability increasing at low densities, peaking around 2,000 persons/km², and gradually declining thereafter. Mean relative humidity ($p = 0.0398$) also showed a hump-shaped association, with peak suitability near 60% and lower suitability at both drier and more humid extremes. In contrast, cropland cover and mean diurnal temperature range were not significant predictors ($p > 0.6$), suggesting limited influence on broad-scale occurrence of *An. funestus* within the study area. Overall, these findings indicate that *An. funestus* preferentially occupies warm, moderately wet environments with intermediate humidity, moderate population densities, and landscapes characterized by low to moderate tree cover.

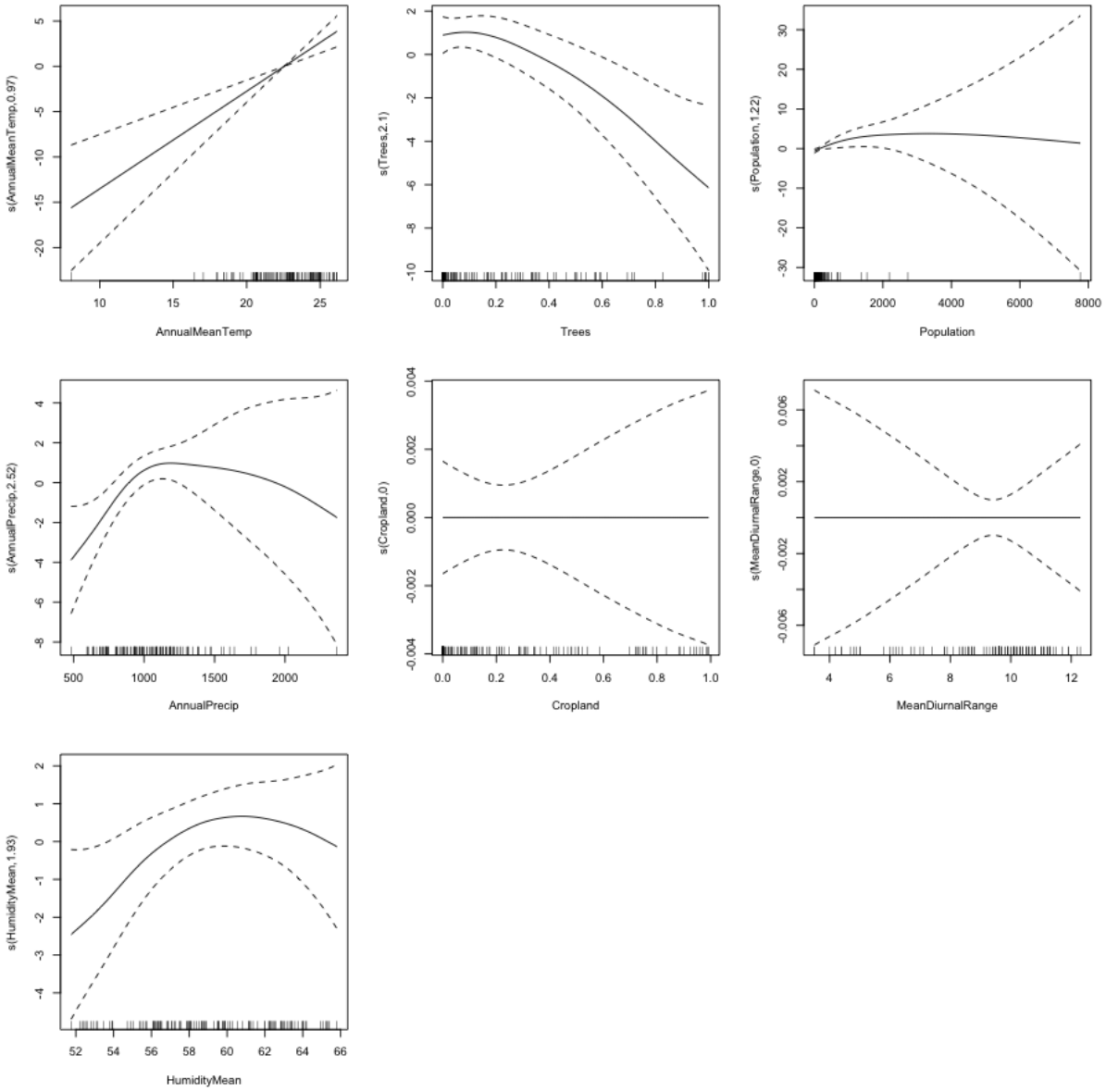


Figure 20. Partial dependence plots for environmental predictors of habitat suitability for *Anopheles funestus.* Each plot shows the partial effect of a single predictor on the predicted habitat suitability of *An. funestus* as estimated by the Generalized Additive Model (GAM). The solid line represents the smoothed effect (cantered to a mean of zero), while the dashed lines indicate ±2 standard errors (confidence intervals). Rug marks along the x-axis show the distribution of observed values for each predictor, indicating regions with strong model support.

4.6 Predicted Species Distribution

The predicted distributions of *An. funestus*, *An. gambiae*, and *An. arabiensis* across Tanzania exhibit distinct spatial patterns, closely reflecting each species' ecological preferences and landscape-level constraints. The maps clearly delineate areas of high predicted probability of

occurrence, demonstrating strong associations with major hydrological features, lowland corridors, and anthropogenically modified environments. Notably, both large and smaller water bodies including Lakes Victoria, Tanganyika, Nyasa, Rukwa, Eyasi, Natron, Sulunga, Kitangiri, Nyagamoma, and Sagara as well as artificial reservoirs like Mtera Dam (represented as white gaps on the maps), are consistently flanked by areas of high predicted mosquito presence. This spatial pattern likely reflects their co-location in low-altitude zones and their influence on surrounding environmental conditions such as increased humidity and the support of irrigation schemes which in turn create favourable breeding habitats. Conversely, closed-canopy forests, high-elevation mountain ranges, and regions characterized by temperate Köppen climate zones (Cwa, Cwb, Cwc, and Csa) consistently correspond to areas of reduced predicted presence across all three *Anopheles* species, as clearly depicted in the maps.

An. funestus shows a predicted distribution concentrated along the Indian Ocean coastal belt, particularly from Tanga through Dar es Salaam to Lindi and Mtwara, while notably avoiding densely populated urban zones such as the Dar es Salaam city center. Additional predicted presence is observed in the Rufiji-Kilombero floodplain and along the edges of the Nyerere (Selous) Reserve and along lakeshores, especially around Lake Victoria, Rukwa, Eyasi, Natron, Sulunga, Tanganyika and the eastern shore of Lake Nyasa (from Ludewa to Mbamba Bay). In contrast, central Tanzania, including Dodoma, Singida, and Tabora and regions characterized by temperate Köppen climate zones (Cwa, Cwb, Cwc, and Csa) or arid steppe climates (BSh and BSk) exhibit sparse or absent predicted presence. Dense forested reserves such as Kigosi, Moyowosi and montane blocks like the Ngorongoro Crater and the East Usambara Mountains also appear largely unoccupied, reflecting the species' limited distribution in cooler, high-altitude, and closed-canopy forest environments (see figure 21).

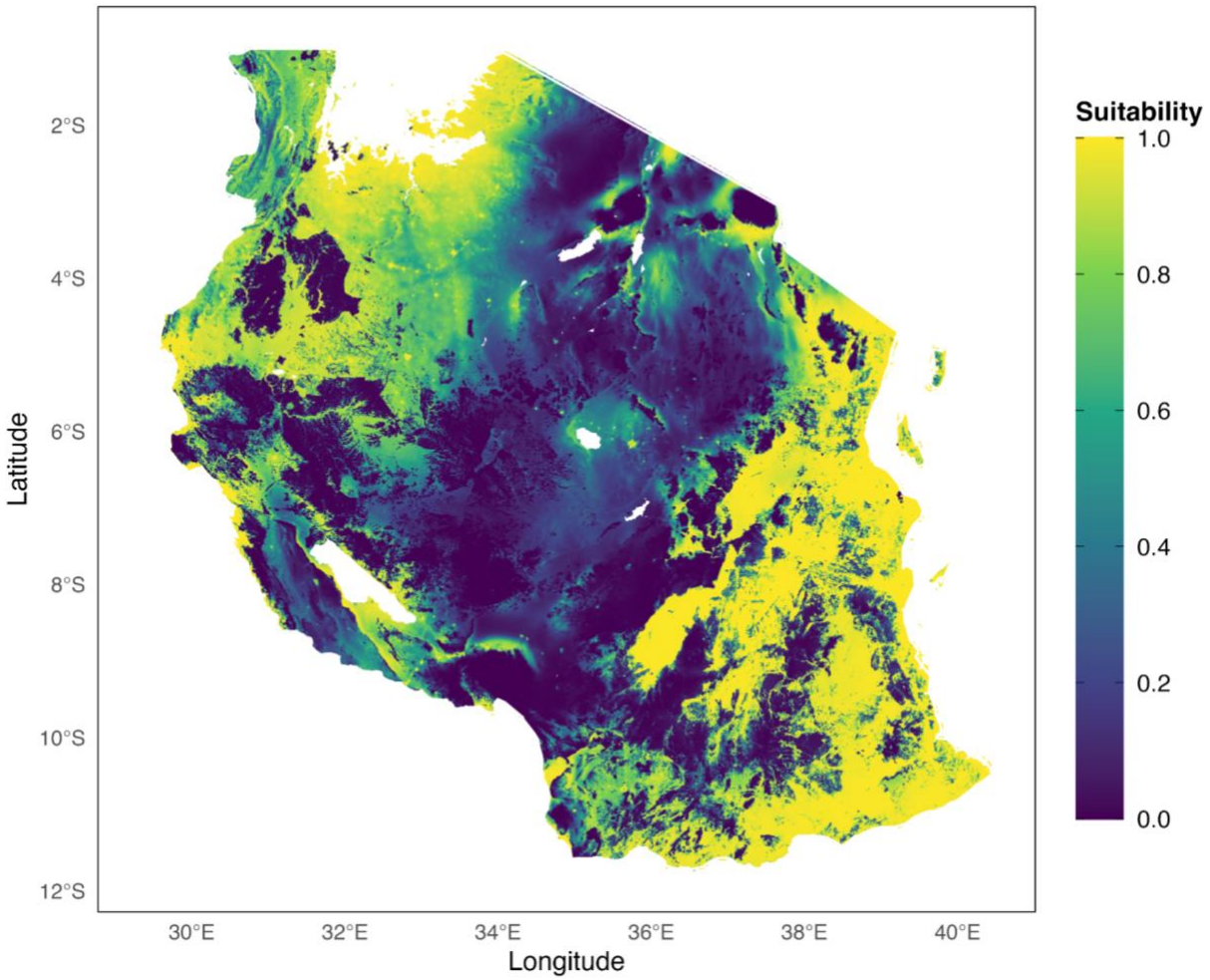


Figure 21. Predicted habitat suitability for *Anopheles funestus* in Tanzania. Modelled distribution of An. funestus with habitat suitability values scaled from low (dark purple) to high (yellow) on the Viridis colour map.

The predicted distribution of *An. gambiae* was slightly broader than that of *An. funestus*, though it generally followed the same spatial trends. It extended across the coastal lowlands, eastern and southeastern Tanzania, and penetrated moderately into central regions. However, unlike *An. funestus*, *An. gambiae* maintained a high predicted presence even in densely populated urban areas, including the core of Dar es Salaam. Notable zones of high predicted presence include the entire coastline, the eastern shore of Lake Nyasa (from Ludewa to Mbamba Bay), areas surrounding Lake Victoria, Rukwa, Eyasi, Natron, Sulunga, Tanganyika and the Zanzibar archipelago, while still avoiding rainforest regions. Like *funestus*, *An. gambiae* appears less frequently and mostly in patches in parts of central Tanzania, particularly around localized water bodies such as the Mtera Reservoir and Lake Sulunga. Generally, the predictions show a clear decline in densely forested and mountainous areas, including the Uluguru and Usambara

ranges and the Ngorongoro Highlands, indicating limited presence in closed-canopy systems. Narrow but persistent predicted presence is also observed along cropland-dominated foothills around Mount Kilimanjaro and Mount Meru. Similar to *An. funestus*, *An. gambiae* shows avoidance of arid steppe zones (BSh and BSk) and temperate Köppen climate regions (Cwa, Cwb, Cwc, and Csa) (see figure 22).

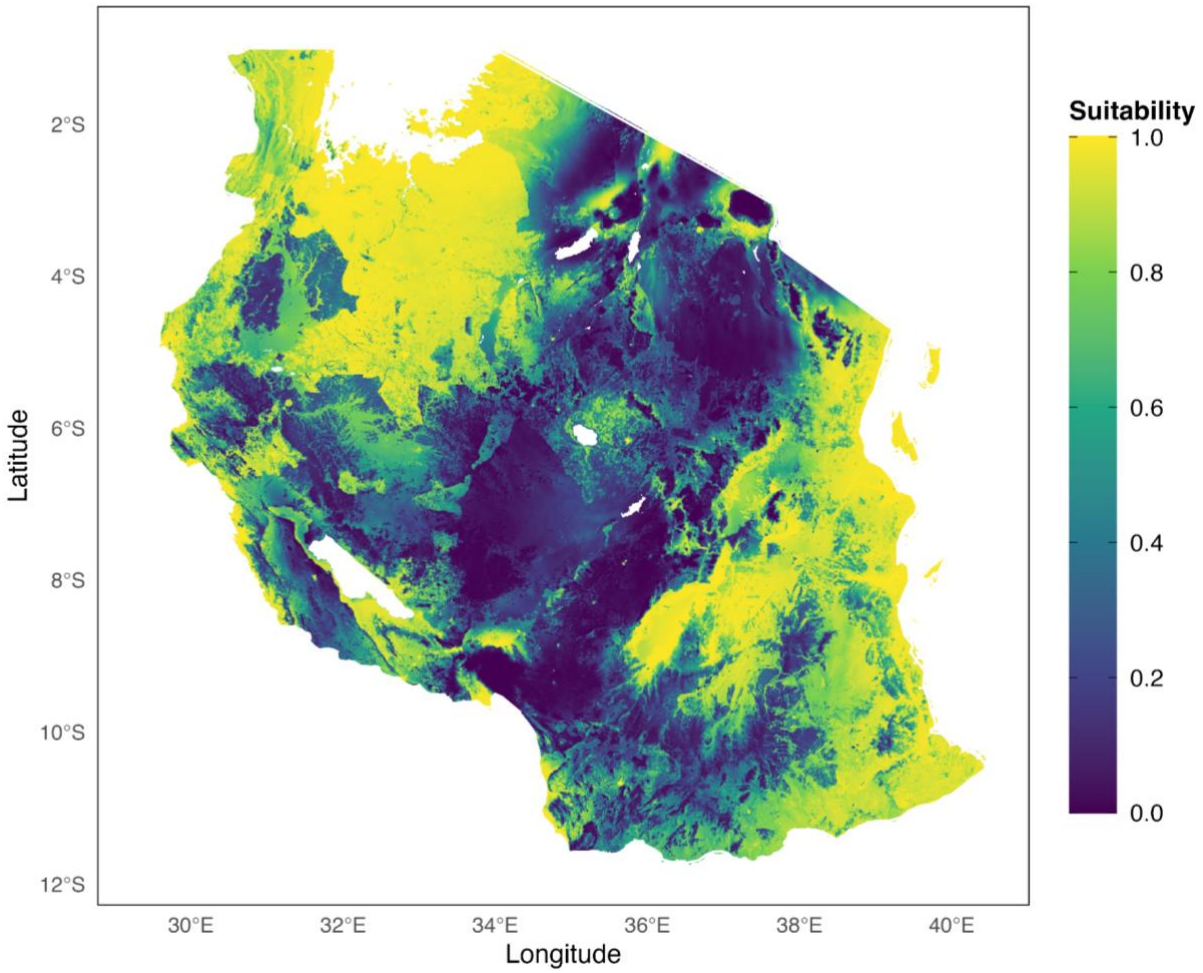


Figure 22. Predicted habitat suitability for *Anopheles gambiae* in Tanzania. Modelled distribution of An. gambiae with habitat suitability values scaled from low (dark purple) to high (yellow) on the Viridis colour map.

In contrast, *An. arabiensis* demonstrates the widest predicted distribution footprint across the country, occupying nearly all areas predicted for *An. funestus* and *An. gambiae* and extending well beyond them. The model indicates extensive predicted presence across the semi-arid and agriculturally dominated landscapes of central and northern Tanzania, including Dodoma, Singida, Manyara, Tabora, and the entire Lake Victoria basin. High presence probabilities also surround inland lakes and water-retaining landscapes such as Lakes Rukwa, Eyasi, and Natron, as well as the Mtera Reservoir, and remain robust along the Lake Tanganyika shore and the

Indian Ocean coastal plain. Compared to the other two species, *An. arabiensis* shows greater spatial reach into drier and more anthropogenic environments, maintaining predicted occurrence even in sparsely vegetated zones. However, like *An. funestus* and *An. gambiae*, it consistently avoids dense forest blocks, high-elevation mountain ranges, and temperate Köppen climate regions (Cwa, Cwb, Cwc, and Csa), aligning with its known ecological intolerance to cool, shaded habitats. While *An. arabiensis* dominates across many arid regions, the model also reveals a patchy distribution in some zones, with areas of lower predicted probability interspersed among areas of high presence, suggesting local variability in habitat suitability or vector adaptation (see Figure 23).

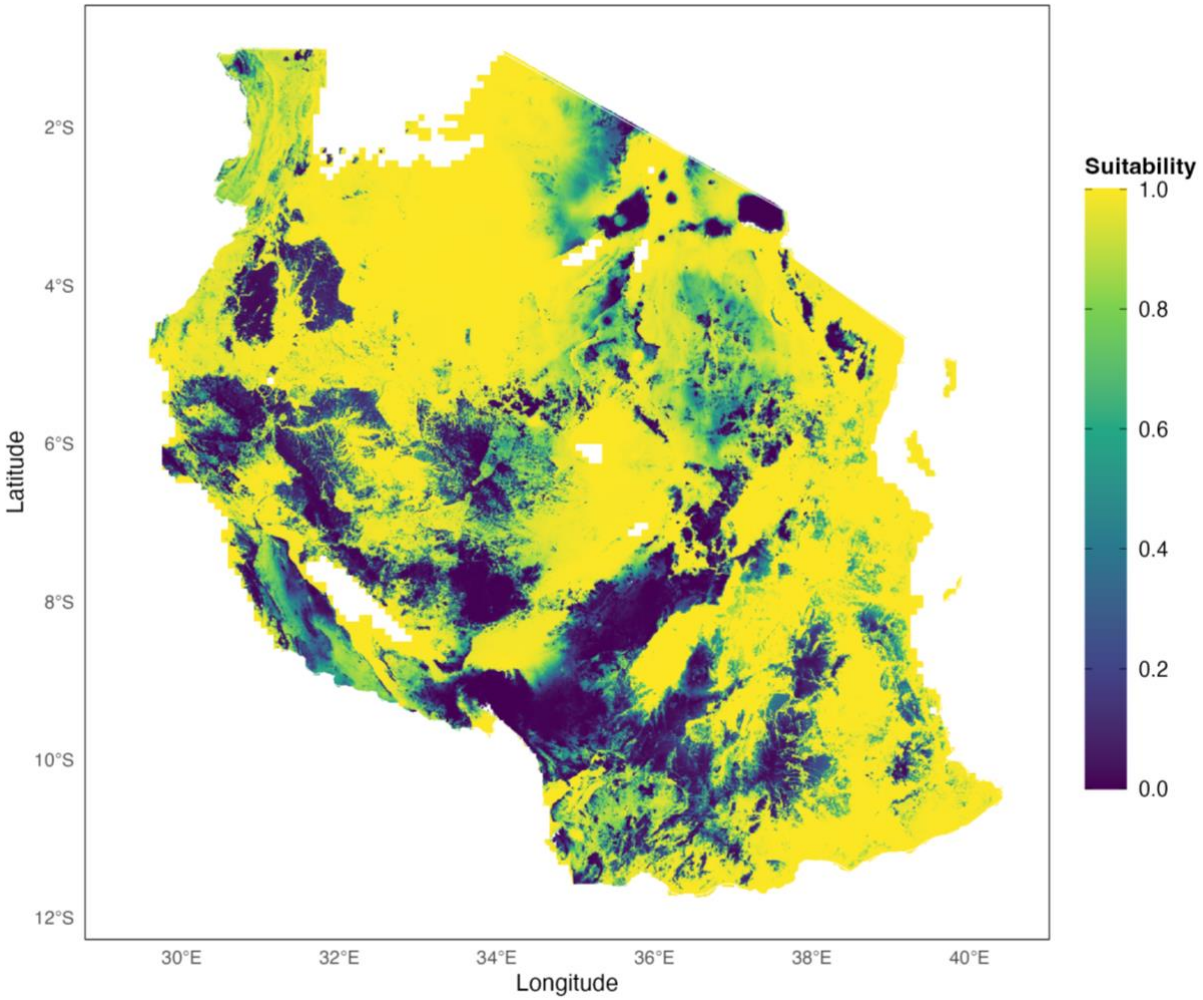


Figure 23. Predicted habitat suitability for *Anopheles arabiensis* in Tanzania. Modelled distribution of An. arabiensis with habitat suitability values scaled from low (dark purple) to high (yellow) on the Viridis colour map.

Together, the three vectors reveal an ecologically stratified yet partially overlapping landscape of malaria risk. Coastal lowlands, lake littorals, and floodplains form consistent hotspots for

all species, while the Southern Highlands, forested reserves, and crater highlands consistently suppress suitability. *An. funestus* and *An. gambiae* concentrate in humid lowlands and cropland peripheries, whereas *An. arabiensis* dominates open, semi-arid zones and thrives near water bodies especially in low-canopy or anthropogenic landscapes, while avoiding densely forested surroundings.

4.7 Discussion

This study revealed pronounced spatial heterogeneity in malaria vector distributions across Tanzania. The three dominant species, *An. arabiensis*, *An. gambiae*, and *An. funestus*, responded differently to climatic, ecological, and anthropogenic gradients, reflecting their distinct ecological niches. The models highlighted fine-scale species-environment relationships that define where each vector is most likely to persist and where their distributions may overlap.

Among the modelled ecological predictors, annual mean temperature emerged as a dominant factor for all three species, with suitability increasing linearly across the data ranges, highlighting their adaptation to warm tropical climates [192]. Precipitation showed a unimodal response, peaking at 1,500–2,500 mm/year for *An. arabiensis*, *An. gambiae*, and *An. funestus*, indicating an optimal moisture range for larval habitats [193]. All three species were associated with moderately vegetated areas, with peak suitability occurring at canopy cover around 20–40%. Species-specific predictors further distinguished their niches: cattle density positively influenced *An. arabiensis*, aligning with its zoophilic behaviour [158], while *An. gambiae* was associated with cropland cover and population density, reflecting its anthropophilic tendencies [194] in human-modified and agricultural landscapes [195]. *An. funestus* preferred moderate humidity ~60% and low population density below ~2,000 persons/km², consistent with its reliance on stable, vegetated water bodies in rural areas [164,165]. These environmental preferences not only reinforce established ecological traits for each species but also delineate the distinct niche boundaries that govern their spatial segregation and potential overlap.

The spatial predictions revealed that *An. funestus* and *An. gambiae* share a strong preference for warm, humid lowland environments, with both species showing high predicted presence along Tanzania's lakeshores, coastal belt, southern lowlands, and northwestern regions. Their distributions were closely associated with ecological zones featuring permanent or semi-

permanent, water bodies, aligning with known breeding requirements and reflecting a shared physiological dependence on stable aquatic habitats and intolerance to desiccation stress [164]. Consequently, both species were consistently absent from arid steppe zones (Köppen BSh and BSk) and temperate highland regions (Cwa, Cwb, Cwc, and Csa), highlighting climatic thresholds that confine their distributions to low-to-mid elevation corridors [81]. However, *An. gambiae* exhibited notably higher predicted presence in densely populated urban centres such as Dar es Salaam, underscoring its greater adaptability to anthropogenic habitats compared to the more rural distribution of *An. funestus*, though such boundaries may shift under warming scenarios or land-use changes [189,196]. In contrast, *An. arabiensis* demonstrated the widest spatial footprint across the predicted space. It occupied nearly all regions suitable for *An. funestus* and *An. gambiae*, and extended further into semi-arid landscapes, particularly in central and northern regions of the country. The model suggested that *An. arabiensis* tolerated lower humidity and sparser vegetation than the other two vectors, consistent with its ecological generalism and ability to exploit ephemeral or man-made water bodies [157,195,197,198]. Despite its broad range, the species also showed exclusion from high-elevation zones, temperate and dense forested areas, likely due to its intolerance of cooler, shaded environments [81].

These distributions delineate malaria risk zones that closely correspond with reports from the NMCP [62]. Notably, they reveal finer-scale hotspots such as areas surrounding water bodies like Mtera Dam, Lake Surunga, and the foothills of Njombe along the Lake Nyasa shoreline that are often overlooked when using regional averages to map malaria prevalence. Overlapping vector presence in the coastal belt, Great Lakes shores, and Rufiji–Kilombero valley corresponds to high-transmission hotspots, with high malaria prevalence in children in areas like the Lake Victoria basin [63]. Peri-urban areas, including Dar es Salaam, sustain transmission due to *An. gambiae* and *An. arabiensis* breeding in man-made sites [199]. Conversely, highlands and dense forests show low predicted occurrences, aligning with historically low or unstable transmission [200]. These patterns reflect the ecological limits of each species, with *An. funestus* tied to permanent waters, *An. gambiae* thriving in humid, human-modified areas, and *An. arabiensis* dominates drier, open landscapes.

Our findings align with regional studies, validating the model's robustness. The restriction of *An. funestus* to low elevations matches observations by Kulkarni et al. (2010) [201], who noted its absence above ~1,900 m. *An. gambiae*'s prevalence in coastal and lowland areas, with

patchy inland presence, concur with continental maps [81,196]. *An. arabiensis*' extension into semi-arid zones reflects its documented resilience in dry environments [157,202]. Specific predictions, such as vector presence in the Kilombero Valley's rice fields, align with surveys showing *An. arabiensis* and *An. funestus* breeding in distinct rice growth stages [195]. Recent studies, like Matowo et al. (2021) [162], confirm high *An. funestus* suitability along Lake Victoria, reinforcing our model's accuracy.

The spatial and ecological distinctions among *An. funestus*, *An. gambiae*, and *An. arabiensis* have direct implications for vector-specific interventions under Tanzania's NMCP. In regions with high predicted co-occurrence particularly the northwestern lake zone and southeastern coastal belt an Integrated Vector Management framework is warranted to address the species-specific ecological niches and behavioural differences that drive localized transmission dynamics [32,33,35]. For species that show high predicted occurrence in urban areas, scaling up insecticide-treated net coverage and improving drainage infrastructure are critical. These interventions address its anthropophilic behaviour and adaptability to man-made larval habitats such as domestic containers, necessitating sustained and consistent implementation [203]. In semi-arid and central regions where *An. arabiensis* dominates, the results support among others the use of endectocides through insecticide-treated livestock, given its positive association to high cattle density tendencies observed in the model [161]. Furthermore seasonal larviciding in irrigation channels and temporary pools during peak rainfall periods may suppress its breeding in ephemeral habitats [158]. In low-risk highland districts like Iringa and Njombe, predictive maps suggest that malaria risk remains constrained by cooler temperatures, but warming trends could destabilize this balance. Surveillance systems triggered by temperature thresholds (e.g., >20°C during December–March rains) could offer early-warning capacity to prevent outbreaks [196].

The models also provide a basis for climate-informed planning. With projected temperature increases under climatic change scenarios, these vectors are likely to expand their range into highland regions such as Arusha and Njombe by 2040 [196]. Pre-emptive interventions, informed by consistent climate monitoring, and larval site surveillance may prevent the establishment of stable transmission zones. Moreover, low-risk forested areas like Udzungwa, Selous game reserves etc, could become ecologically permissive for vectors if deforestation continues to reduce canopy cover and create sunlit pools, potentially enabling these species to colonize the newly modified habitats [84]. Integrating vector suitability maps with climate and

land-use change projections will be critical for anticipating range shifts and optimizing the spatial targeting of NMCP interventions.

These results are not without limitations. Ecological suitability does not necessarily imply realized presence, as local absences may arise from interspecific competition, predation, dispersal barriers, or suppression following vector control interventions such as larval source management or indoor residual spraying (IRS). Conversely, some areas predicted as suitable may remain unoccupied due to limited colonization, habitat fragmentation, or recent environmental change.

Although the habitat suitability models exhibited strong predictive performance and ecological coherence, their interpretation warrants measured caution due to the potential for circularity in environmental modelling. Pseudo-absences were defined from areas considered environmentally unsuitable for *Anopheles* occurrence, based on climatic and ecological thresholds similar to those later used as predictors. This methodological overlap, while methodologically necessary, can introduce partial dependence between data generation and model fitting.

The present framework, however, incorporated several safeguards to mitigate this risk. First, pseudo-absences were drawn from extreme environmental ranges, well beyond known physiological and ecological tolerance limits to minimize overlap with potential habitat conditions. Each pseudo-absence was further screened to ensure it was not spatially proximate to presence points or located within plausible dispersal ranges of *Anopheles* populations, thereby reducing the risk of environmental or geographic colocalization (Figure 17). Second, all predictors underwent rigorous collinearity screening, objective model selection, and comprehensive diagnostic validation to ensure that retained variables contributed genuine explanatory power rather than artefactual circularity. Third, model generalizability was tested using spatial block cross-validation, which evaluates predictive strength across geographically independent subsets. Collectively, these procedures strengthen the interpretive reliability of the results and limit the influence of circularity on model outcomes.

Accordingly, while the models should not be interpreted as establishing direct causality, they provide robust evidence of biologically consistent associations between environmental gradients and *Anopheles* distributions. By combining physiologically defined pseudo-absences

with true presence data, this integrative framework advances beyond single-method approaches, bridging the mechanistic precision of physiological mapping with the flexibility of correlative modelling. The models thus maintain strong inferential validity while transparently acknowledging the methodological constraints inherent to ecological prediction. Despite these caveats, the framework demonstrated high internal consistency and ecological realism, aligning with known vector bionomics, observed spatial distributions, and malaria prevalence patterns across Tanzania. It therefore represents a robust and scalable decision-support tool for malaria control, particularly suited to regional and national planning where fine-resolution risk stratification is essential for targeted intervention.

Chapter Five: Population Structure of *Anopheles arabiensis* in Tanzania: The Implication of Using ANOSPP in Malaria Vectors Surveillance

Chapter summary

Understanding the population structure and regional connectivity of malaria vectors is crucial for tracking the spread of traits like insecticide resistance and for predicting how genetically engineered mosquitoes will spread. Despite this importance, detailed genetic insights across Tanzania remain limited. This chapter addresses that gap by analysing *Anopheles arabiensis* population structure across the country and beyond using short amplicon sequences generated by the ANOSPP amplicon sequencing platform. Samples were collected from 21 Tanzania districts with *An. arabiensis* presence. Genetic differentiation among districts was generally low, but with significant isolation by distance patterns in the country. To place these findings in a broader continental context, Tanzanian data were combined with samples from seven other African countries, revealing three major genetic clusters: (i) Tanzania, Uganda, and eastern Democratic Republic of Congo; (ii) Madagascar; and (iii) West Africa, with further substructure within Nigeria. These results validate the ANOSPP platform for scalable population genetics and offer initial insights into gene flow and vector connectivity, supporting more spatially informed malaria control strategies.

5.1 Background on Population Dynamics of *Anopheles arabiensis*

Compared to the other major African malaria vectors, *An. gambiae* and *An. funestus*, *An. arabiensis* typically exhibits lower chromosomal inversion diversity, yet demonstrates remarkable ecological versatility and behavioural flexibility[204–206]. These traits are likely shaped by a combination of recent selective sweeps and historical demographic expansions [207]. This adaptive plasticity is mirrored in genetic studies that reveal extensive gene flow and large effective population sizes (Ne) across broad spatial scales. For instance, Donnelly et al. (1999) [208] reported low levels of genetic differentiation between Tanzania and Mozambique *An. arabiensis* populations, suggesting recent range expansion and consistently high Ne. Supporting this, Kent et al. (2007) [209] analysed populations in southern Zambia over three transmission seasons, including one characterized by severe drought, and found no evidence of genetic bottlenecks, significant allele frequency shifts, or reductions in heterozygosity. The populations remained effectively panmictic across 2,000 km², with

negligible differentiation even between sites 80 km apart, indicative of large, interconnected demes resilient to temporal environmental fluctuations.

However, this broad-scale genetic homogeneity often masks subtle but meaningful patterns of cryptic population structure at finer spatial resolutions. In Tanzania, Maliti et al. (2014) [87] examined populations from five coastal districts, the Kilombero Valley, and the islands of Zanzibar (Unguja and Pemba). The study reported overall low genetic differentiation (mean FST ≈ 0.015), consistent with high levels of gene flow. However, a significant isolation-by-distance pattern was detected (Mantel test r = 0.46, p = 0.0008), indicating a spatial genetic gradient not evident from regional differentiation estimates alone. Temporal consistency of this structure suggests stable, underlying patterns of limited dispersal. Further fine-scale studies in southern Tanzania's Kilombero Valley highlight even deeper microgeographic structuring. Ng'habi et al. (2011) [86] detected at least two genetically distinct *An. arabiensis* clusters coexisting within the same villages. These subpopulations, while sympatric, exhibited restricted gene flow likely driven by fine-scale ecological heterogeneity or behavioural divergence, rather than overt physical barriers. Expanding on these regional patterns, recent analyses reinforce the findings of broad connectivity across much of Tanzania, while also acknowledging the influence of geographic and topographic barriers. Mwinyi et al. (2025) [88] report sustained gene flow across diverse ecological zones, indicative of demographic stability. However, the presence of topographic barriers such as the Rift Valley in East Africa contributes to regional genetic subdivision [210], while geographic isolation by large water bodies between West Africa and surrounding islands (e.g., Madagascar, Reunion, Mauritius) promotes more pronounced genetic differentiation [211].

Understanding the population structure of *An. arabiensis* is critical not only for evolutionary biology but also for malaria control planning. Patterns of gene flow influence how interventions in one region may affect populations elsewhere. Genomic data can be used to identify loci under selection, inform resistance spread, and predict the spread of gene drive. However, the presence of structural variants like chromosomal inversions can limit the uniform effectiveness of such interventions, necessitating population-specific designs [212]. Genomic surveillance also offers early detection of resistance evolution potentially before phenotypic failure is observed. For instance, Seck et al. (2025) [213] documented significant population structure and reduced genetic diversity in vector competence loci in *An. gambiae s.l,* across 19 countries, highlighting the need for context-sensitive, genomically informed vector control strategies. As

An. arabiensis continues to expand across ecologically diverse landscapes, integrating population genetic insights with ecological surveillance will be essential for designing sustainable, precision-targeted vector control strategies that remain robust in the face of environmental change and evolutionary adaptation.

5. 2 Results

5.2.1 Variant Calling, Filtering, and Genotype Imputation

After quality control, 59 of the 62 ANOSPP amplicon targets were retained, excluding three that were consistently missing or poorly amplified across specimens. Across these 59 targets, the concatenated callable base-pair sites comprised 9,473 base-pair sites for all 1,425 specimens. Within this sequence, 1,394 sites (14.7%) were polymorphic, the majority of which (1,362; 97.7%) were biallelic, while 32 sites (2.3%) exhibited more than two alleles. To enable population-genomic analyses, a biallelic genotype matrix was constructed from the filtered variant dataset. This ensured a standardized representation of genetic variation across samples and reduced noise from poorly resolved sites. **A** ≤10% per-site missingness filter was applied, retaining 8,796 positions and reducing overall missingness to 0.66%**.** Monomorphic sites (invariant sites across the specimens) were subsequently excluded, producing 1,271 biallelic polymorphic loci in the final dataset**.** Remaining missing genotypes were imputed using the MissForest [135] random-forest algorithm (out-of-bag NRMSE = 0.38), yielding a complete 1,425 × 1,271 genotype matrix. This matrix served as the foundation for downstream analyses including PCA, DAPC, FST estimation, and spatial-genetic modelling.

5.2.2 Genetic Clustering and Population Structure at the Country Level

5.2.2.1 PCA and DAPC at the Country Level

Principal component analysis (PCA) revealed no geographic population structure across Tanzania. Along PC1 (6% variance explained), the only visible pattern was the presence of three parallel bands, a configuration that is characteristic of segregating chromosomal inversions, rather than discrete population clusters. PC2 (3.87%) showed no spatial separation among districts (Figure 24A). To identify the genomic basis of this signal, loci were ranked by their squared loadings on PC1. A single amplicon, target 17, accounted for >90% of the variance. Because target 17 lies within the well-characterized 2Rb inversion region in *An.*

arabiensis and is adjacent to target 18 (also within 2Rb), both targets were removed. After their exclusion, the banded pattern persisted (Figure 24C).

A second loading screen revealed that targets 11 and 46 now contributed >90% of the remaining signal. Target 46 lies within the 3Ra inversion region; therefore, all 3Ra-associated amplicons (targets 46–50) were removed to test whether this inversion alone explained the structure. However, the PCA pattern still remained as in in figure Figure 24C. At this point, target 11 emerged as the dominant contributor (>90%) to the signal. When target 11 was removed, together with all 2Rb and 3Ra loci, the banding collapsed entirely, and no structure remained (Figure 24E).

When examined in isolation, the 3Ra-linked targets produced only a weak and diffuse pattern, consistent with the low frequency of the 3Ra inversion previously documented in Tanzanian *An. arabiensis* populations, about 2% in the cytogenetic survey of Mnzava & Di Deco (1990) conducted across several locations [214] and 8–17% in the Kilombero Valley reported by Main et al. (2016) (~8–17%)[215]. In contrast, 2Rb-linked targets generated a strong signal, in line with the consistently high frequencies of the 2Rb inversion reported in the same studies (~55% across several Tanzanian locations and >80% in the Kilombero Valley)[214,215].

Target 11 lies upstream of, but not within the 2La inversion, which is fixed in *An. arabiensis*. The strong loading of target 11 remains unexplained; it may reflect either residual influence from the nearby 2La region (which is unexpected) or a previously uncharacterized local structural polymorphism. Resolving this would require higher-resolution approaches and is beyond the scope of the present study.

This absence of geographic structure was confirmed using a supervised method. Discriminant Analysis of Principal Components (DAPC), using district as a grouping factor and retaining 45 PCs (selected via cross-validation and a-score optimization), also showed no evidence of clustering (Figure 24B, D, F).

Together, these results indicate high genetic connectivity and weak spatial differentiation of *An. arabiensis* across Tanzania, in agreement with previous findings from Tanzania[87,88].

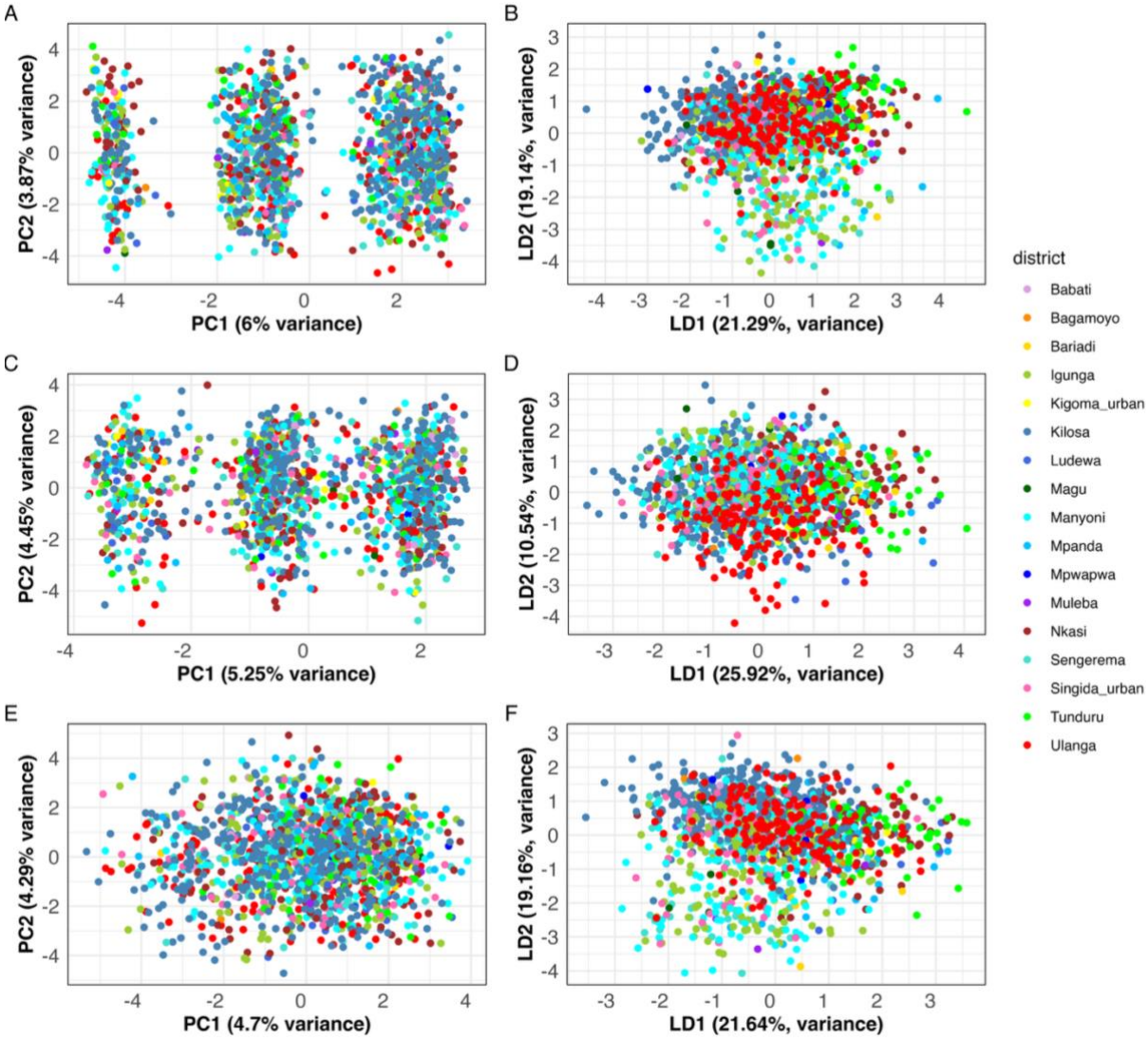


Figure 24. Population structure of *Anopheles arabiensis* across Tanzania inferred from PCA and DAPC. (A) PCA of all polymorphic biallelic sites shows three parallel bands along PC1, a pattern characteristic of segregating chromosomal inversions rather than discrete population clusters. (B) DAPC using district as the grouping factor shows no evidence of geographic clustering. (C) PCA after removing 2Rb-associated targets 17 and 18 retains the same banded pattern observed in (A). (D) The corresponding DAPC still shows no clustering by district. (E) PCA after additionally removing target 11 and all 3Ra and 2Rb associated targets causes the banding pattern to collapse, leaving no residual structure. (F) DAPC on this final filtered dataset likewise shows no detectable clustering, consistent with very weak population structure across Tanzania.

5.2.3 Genetic Differentiation at the Country Level

Generally, the genetic differentiation was relatively low across all sampled areas. Overall genetic differentiation across districts was low (FST, 0.0049). Pairwise FST values were generally low (0 to 0.016), with most comparisons showing non-significant differentiation after bootstrapping (Figure 25).

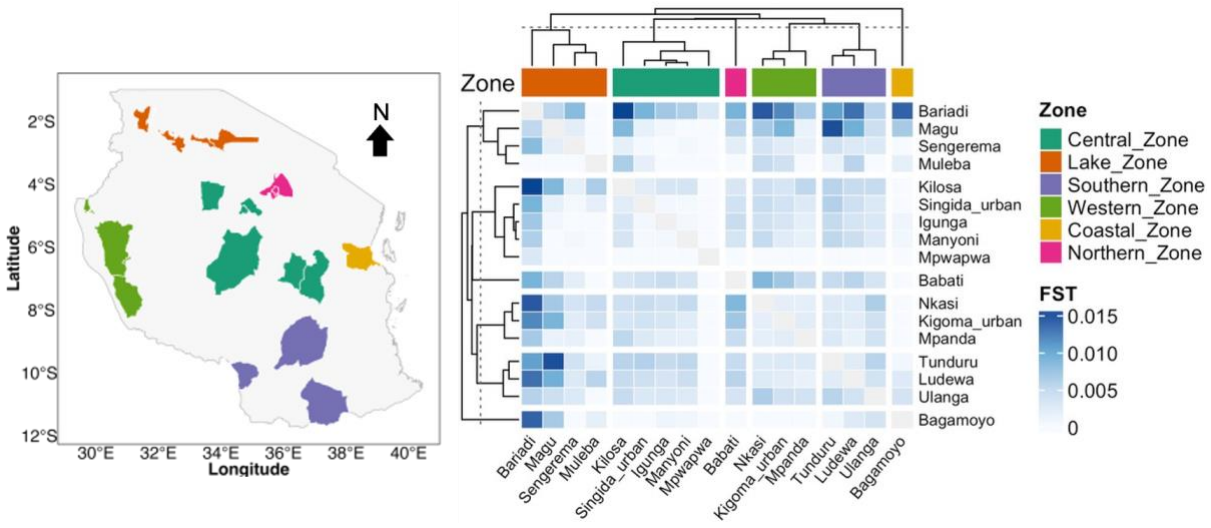


Figure 25. District-level pairwise FST heatmap. District colours on the map denote administrative zones. The heatmap shows uniformly low genetic differentiation among *Anopheles arabiensis* populations (FST = 0–0.016; white to dark blue). Hierarchical clustering reveals no strong district-level structure, consistent with high gene flow and weak population differentiation across Tanzania.

5.2.4 Isolation by Distance at the Country Level

To examine isolation by distance (IBD), Rousset's genetic distance ($a = F_{ST}/(1–F_{ST})$) was used with local populations defined at a 1 km radius, consistent with the typical reported dispersal range of *An. arabiensis* [110,216]. Because the IBD relationship is known to be reliable only within an intermediate spatial band and becomes diluted by broad-scale heterogeneity and rare long-distance migration, we restricted analyses to pairs separated by ≤100 km. This distance represents a biologically coherent scale: it is above the species' average flight range (<1 km) yet avoids the confounding influence of broad-scale heterogeneity and windborne migration events documented over hundreds of kilometres [217]. Including more distant pairs reduced the correlation between genetic and geographic distance, consistent with previous work showing that large-scale heterogeneity flattens or obscures local IBD signals [218,219]. Within this ≤100 km window, a clear signal of IBD across all samples was found (masked Mantel r = 0.377, p = 0.003; Figure 26). Ordinary least squares regression of Rousset's a on log-

transformed geographic distance gave a slope of 0.0295 (95% CI 0.0225–0.0365), while a mixed-effects MLPE model accounting for non-independence of population pairs yielded a standardized slope of 0.157 (95% CI 0.070–0.244). Back-transformation to the original scale corresponded to an increase of ~0.0123 (95% CI 0.0055 – 0.0191) per unit of log-distance, consistent with weak but significant isolation by distance over short to intermediate scales.

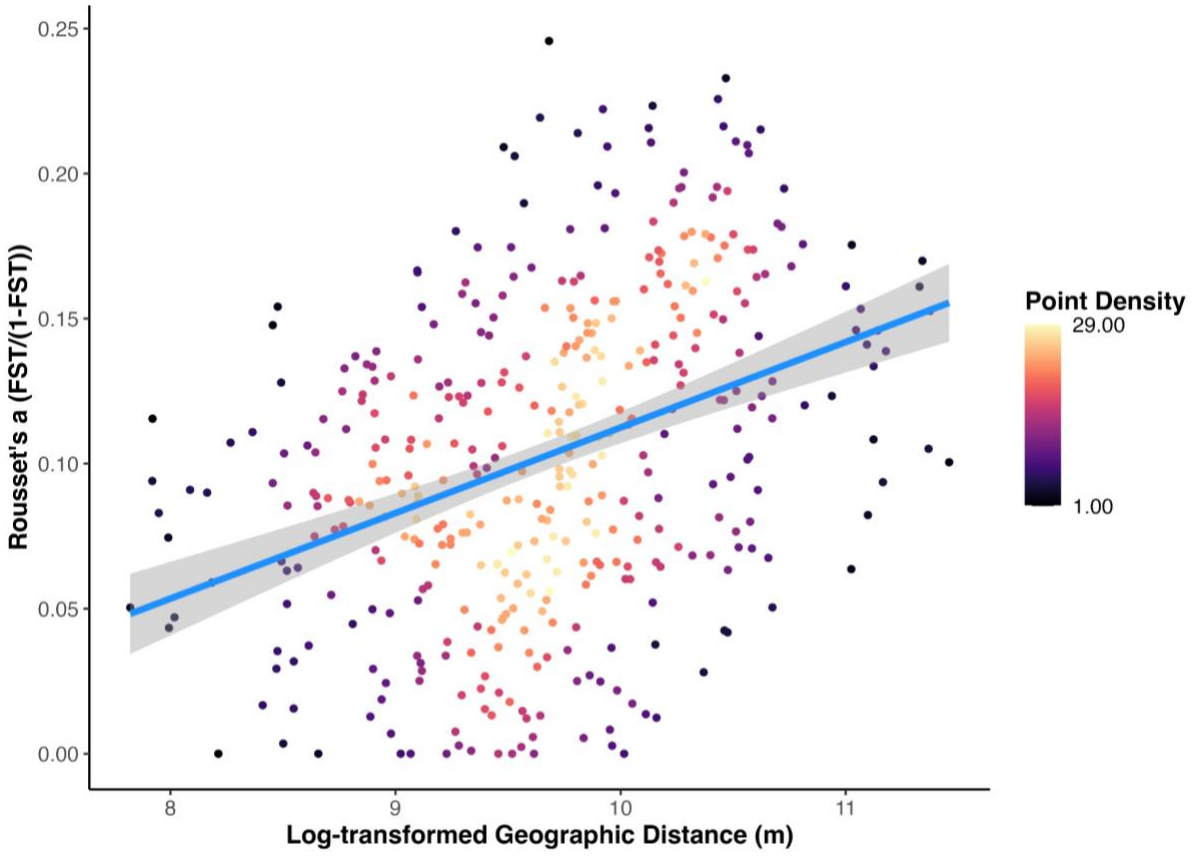


Figure 26. Isolation by distance pattern in *Anopheles arabiensis*. Genetic distance (Rousset's a) plotted against log-transformed geographic distance for all sample pairs, with populations defined at a 1 km radius.

5.2.5 Continental Patterns of Genetic Structure: Tanzania in Context

To assess the position of Tanzanian populations of *An. arabiensis* within the broader continental genetic landscape of Africa, ANOSPP datasets from 2,367 specimens collected across multiple African countries (representing West, East, and Central Africa, as well as Madagascar) were also analysed. The same data processing procedures and analytical methods applied to the Tanzanian data (as articulated in section 5.2.1) were used here.

5.2.5.1 PCA at the Continental Level

Principal component analysis (PCA) of the ANOSPP-derived dataset revealed an elongated, continuous distribution of genetic variation across the first two axes (Figure 27). No abrupt separations were evident among countries or regions. Tanzanian samples overlapped extensively with those from Uganda and the Democratic Republic of Congo (DRC), whereas Madagascar and West African populations were positioned toward opposite extremes of the primary axis of variation.

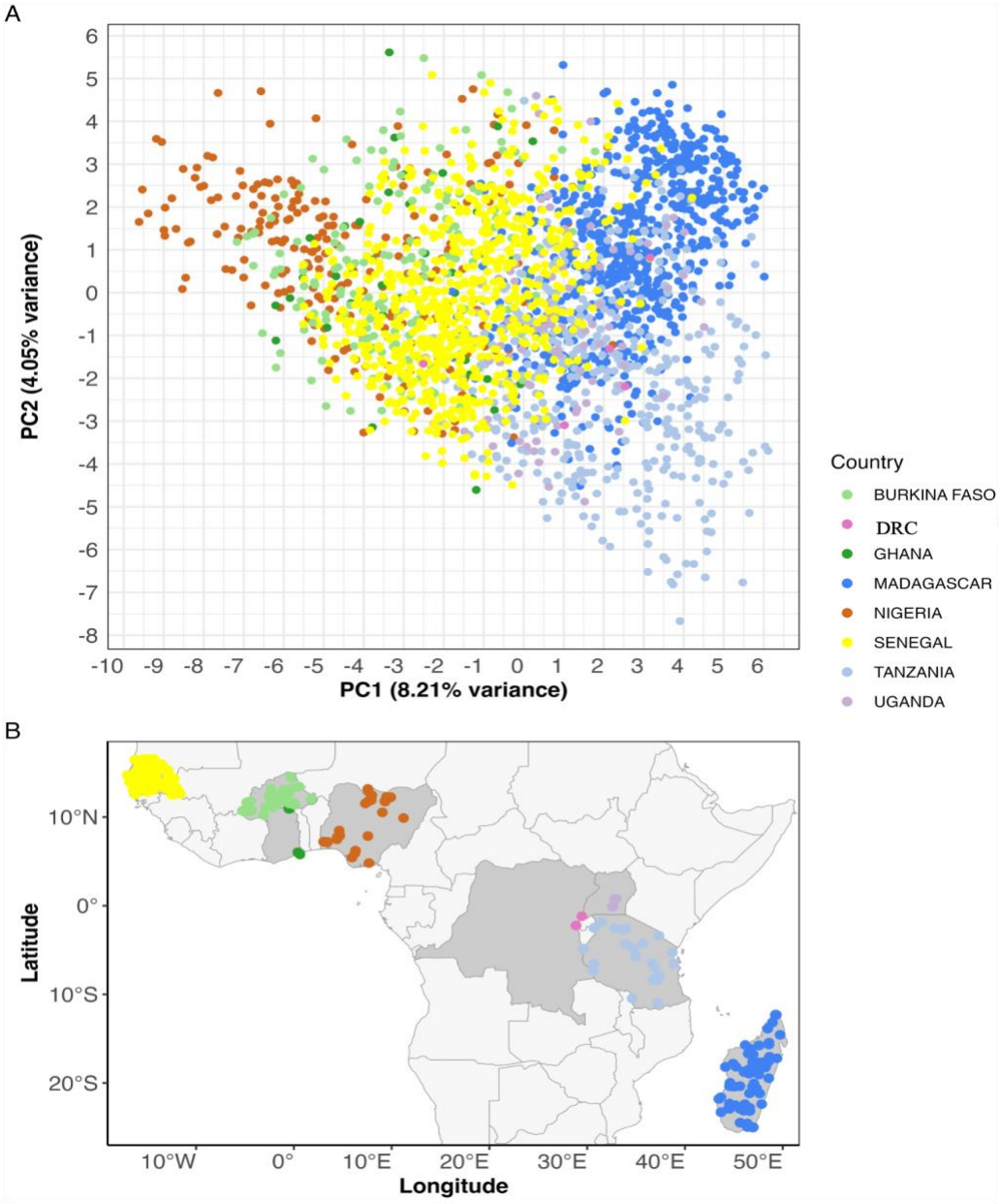


Figure 27. Population structure of *An. arabiensis* using PCA across sub-Saharan Africa. (A) PCA scores for PC1 vs PC2; each point is an individual colour by country. Percent variance explained is indicated on the axes. (B) Sampling map showing collection locations; symbols are coloured by country to match panel A.

5.2.5.2 DAPC at the Continental Level

A supervised Discriminant Analysis of Principal Components (DAPC) was performed using the *adegenet* package in R [139], with countries predefined as populations. Seventy principal components were retained based on cross-validation, selected to minimize root mean squared error and optimize the α-score to prevent overfitting. The analysis revealed three major genetic clusters (Figure 28A-B): one corresponding to Madagascar, a second to Eastern Africa (Tanzania, Uganda, and the Democratic Republic of Congo), and a third to Western Africa (Nigeria, Ghana, Burkina Faso, and Senegal). Visualization of the first three linear discriminants in three-dimensional space uncovered additional substructure within Nigeria, where a subset of individuals diverged from the broader Western African cluster. Therefore, a further focused DAPC restricted to Western Africa confirmed the presence of two genetically distinct Nigerian subclusters. The first cluster (West–C1) comprised all populations from southern Nigeria along with two northern populations located along major roads. The second cluster (West–C2) included the remaining northern Nigerian populations, which grouped with individuals from other West African countries (Figure 28C–D).

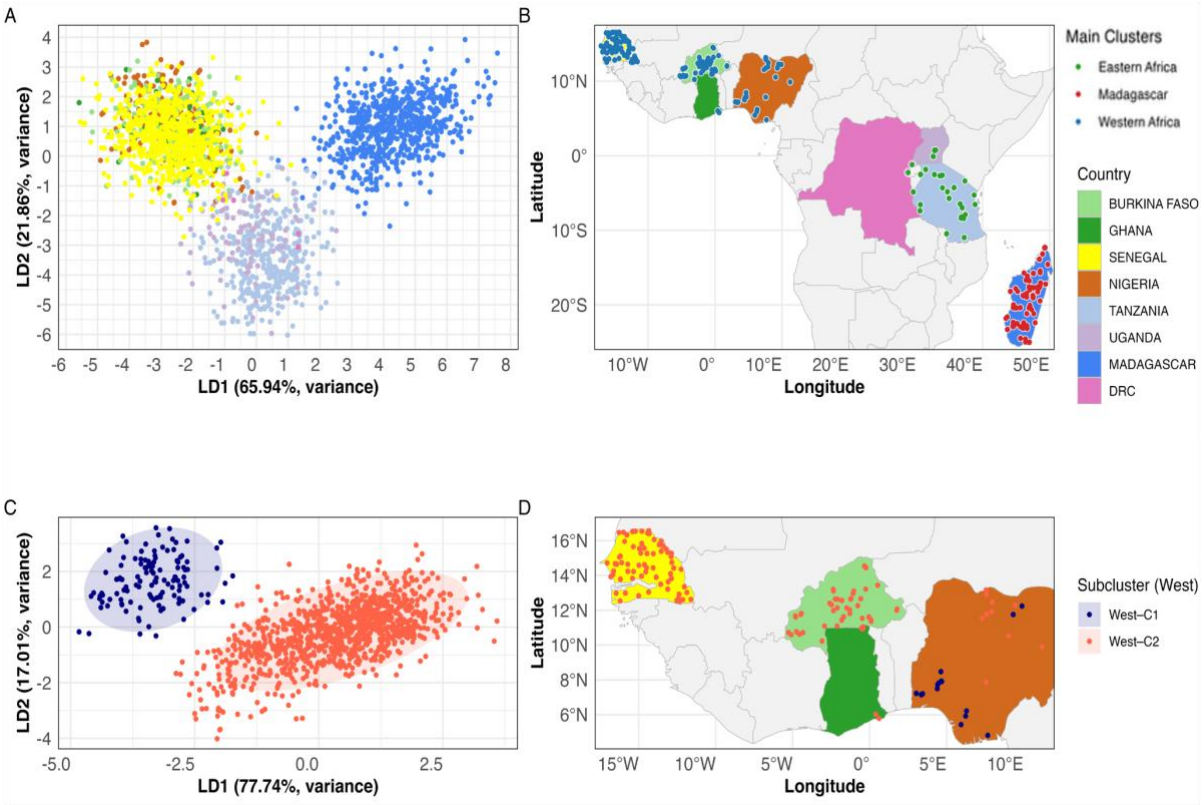


Figure 28. Population structure of *An. arabiensis* using DAPC Across Sub-Saharan Africa. (A) a DAPC plot showing three major genetic clusters (West African, East African, and Madagascar), colour representing the country. (B) A map of Africa depicting the geographic locations of these three main clusters where the block colour represents the country,

and the dots colour represent the cluster (Main Clusters in legend). (C) A focused DAPC on the Western African samples identified two sub-clusters, West–C1(dark blue) and West–C2 (dark orange). (D) A map showing the geographic distribution of the two western sub-clusters (West–C1 and West–C2).

5.2.5.3 Genetic Differentiation at Continental Level

Pairwise Weir and Cockerham's FST values indicated generally low genetic differentiation across the continent (Figure 29). The highest relative differentiation was observed between Madagascar and mainland populations, with FST values reaching up to 0.057. This was followed by differentiation between western and eastern African populations. In contrast, eastern African populations showed minimal genetic differentiation (FST ≤ 0.013), suggesting high levels of gene flow and connectivity within the region. Similarly, western African populations exhibited low levels of differentiation, with the lowest observed between Ghana and Burkina Faso (FST = 0.002), highlighting substantial genetic homogeneity across this subregion.

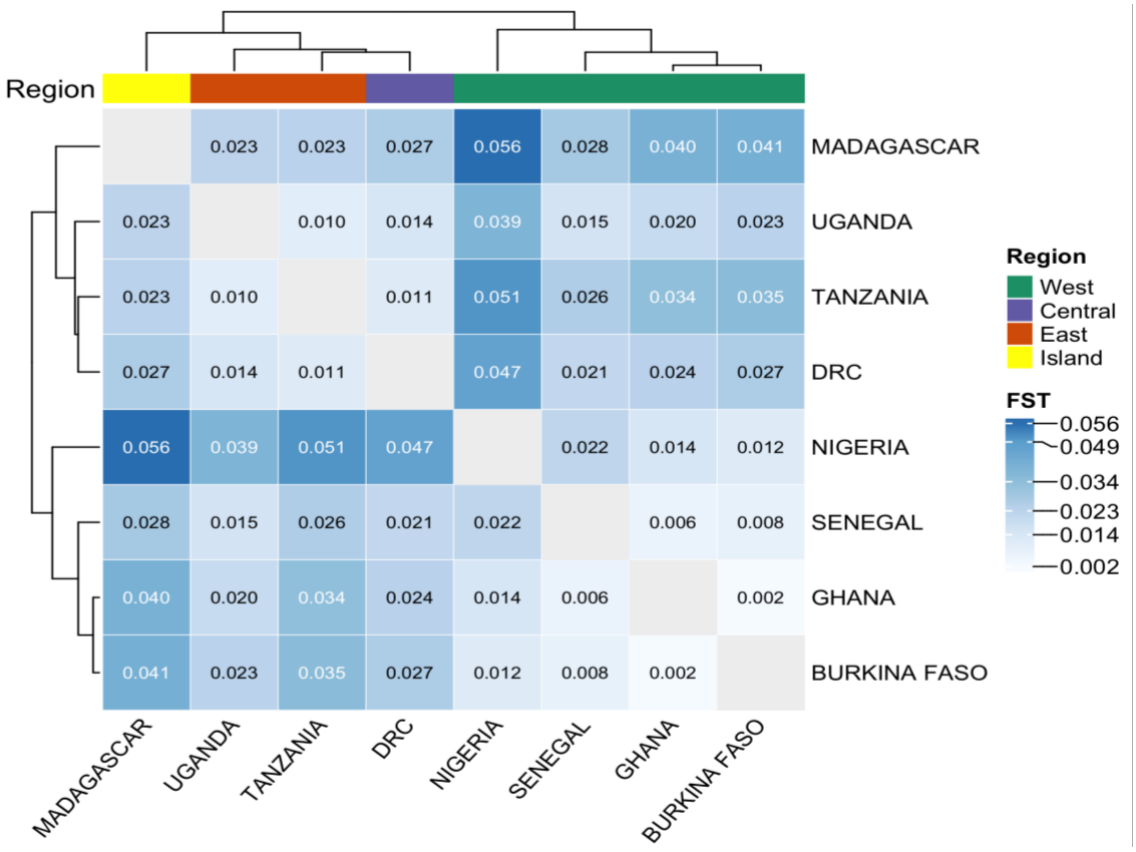


Figure 29. Pairwise Genetic Differentiation (FST) across Sub-Saharan Africa populations. Genetic differentiation was generally low, with the highest values between Madagascar and mainland populations (up to 0.057). Minimal differentiation was observed

within eastern and western Africa, particularly between Ghana and Burkina Faso (FST = 0.002).

5.2.5.4 Isolation by Distance at Continental Level

Given the continuous genetic patterns observed in the PCA, isolation by distance (IBD) was evaluated at the continental scale using two complementary approaches. First, individual-level structure was assessed using Euclidean distances in DAPC space, showing a strong correlation with geographic distance (Mantel r = 0.767, p = 0.0013; Figure 30A). Second, pairwise FST values were transformed into Rousset's a (FST / (1 - FST)), a linearized measure of genetic differentiation expected to increase with geographic distance under stepping-stone IBD models, which likewise revealed a strong positive correlation between genetic and geographic distances (Mantel r = 0.679, p = 0.0061; Figure 30B).Together, these results demonstrate that genetic similarity decreases with increasing geographic separation, consistent with a stepping-stone model of isolation by distance operating across both population and individual scales.

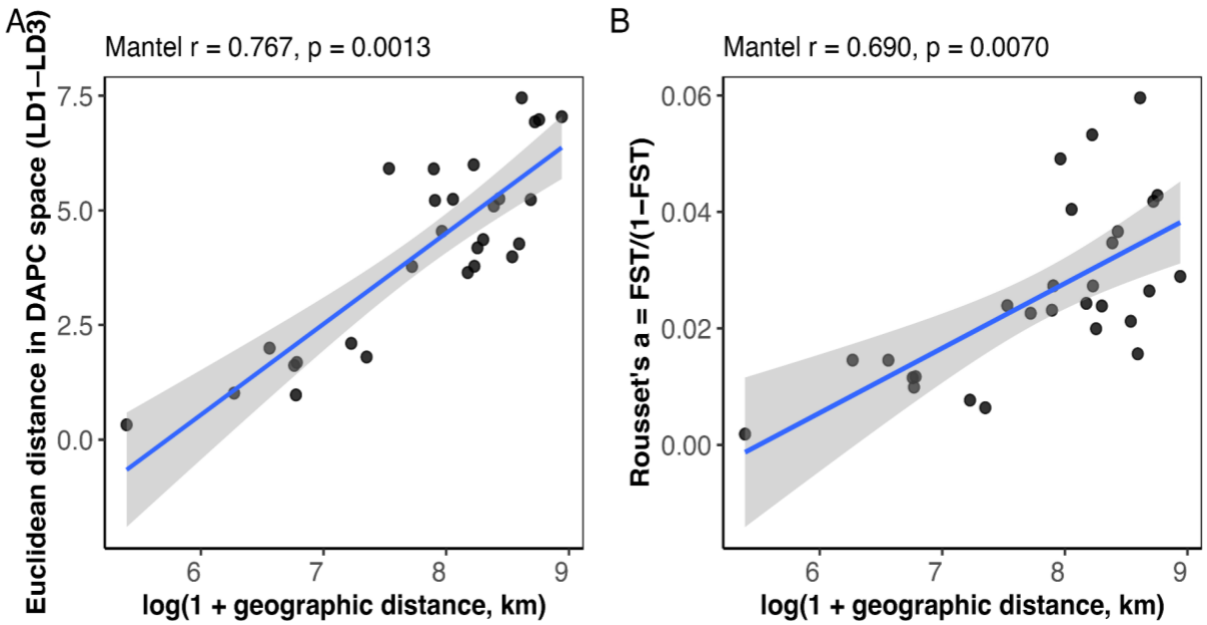


Figure 30. Isolation by distance across populations. (A) Euclidean genetic distances in DAPC space and (B) Rousset's genetic distance both show significant positive correlations with log-transformed geographic distance.

5.2.5.5 Isolation by Resistance at Continental Level (Effect of Natural Barrier to Gene Flow)

Given that Madagascar exhibited a relatively high FST compared to other countries, isolation by resistance using the ocean as a resistance surface (barrier to gene flow) was tested, while controlling for geographic distance. The analysis revealed that oceanic separation was strongly associated with Rousset's genetic distance (Mantel $r = 0.689$, $p = 0.001$), and this relationship remained significant after controlling for geographic distance (partial Mantel $r = 0.443$, $p = 0.001$). MLPE models supported the influence of oceanic resistance, both without ($\beta = 1.093 \pm 0.001$, $t = 1141.1$) and with geographic distance included ($\beta = 1.061 \pm 0.001$, $t = 1222.5$), explaining a large portion of the variation (marginal $R^2 = 0.598$; conditional $R^2 = 0.635$). These findings indicate that the ocean acts as a major barrier to gene flow between Madagascar and the mainland. In continental Africa, the Central African rainforest was evaluated as a potential resistance surface. The initial Mantel test showed a positive correlation with genetic distance (Mantel $r = 0.656$, $p = 0.011$), but this effect decreased and was not statistically significant after accounting for geographic distance (partial Mantel $r = 0.307$, $p = 0.172$), suggesting overlap with spatial structure. In contrast, MLPE models identified separate effects of both rainforest resistance ($\beta = 0.897 \pm 0.001$, $t = 1023.3$) and geographic distance ($\beta = 1.213 \pm 0.001$, $t = 1462.6$), together explaining a notable share of genetic variation (marginal $R^2 = 0.573$; conditional $R^2 = 0.618$). This suggests that, although related to spatial distance, rainforest resistance independently limits gene flow.

5.2.5.6 Isolation by Environment (Effect of Climate/Environmental Condition to Gene Flow)

Within West Africa, where DAPC revealed population sub-structuring, genetic differentiation was low within clusters (FST = 0.0154 in West–C1 and 0.0049 in West–C2) but higher between clusters (pairwise FST = 0.0565). Overall genetic differentiation across the region was 0.0167. Across the region, genetic distance correlated with geographic distance (Mantel r = 0.452, p = 0.001), mean diurnal temperature range (MDR) (Mantel r = 0.385, p = 0.001), and precipitation (Mantel r = 0.258, p = 0.001). After controlling for geographic distance, both precipitation (partial Mantel r = 0.139, p = 0.001) and MDR (partial Mantel r = 0.265, p = 0.001) remained significant. MLPE models confirmed independent positive effects of precipitation ($\beta = 0.0158 \pm 0.0012$, t = 12.69), MDR ($\beta = 0.4514 \pm 0.0013$, t = 339.25), and geographic distance ($\beta = 0.4459 \pm 0.0012$, t = 387.22), explaining a moderate proportion of genetic variation (marginal $R^2 = 0.294$; conditional $R^2 = 0.359$). These results indicate that both precipitation patterns and

temperature variability, which are the potent measure of climate, contribute to genetic structuring in addition to isolation by distance.

5.6 Discussion

This study demonstrates that *Anopheles arabiensis* populations across Tanzania exhibit high genetic connectivity, with very weak differentiation among districts. The absence of discrete genetic clusters in both PCA and DAPC analyses indicates that populations form a largely continuous gene pool, consistent with extensive gene flow across the country. Although overall structure was minimal, a scale-dependent isolation-by-distance (IBD) pattern observed: genetic distance increased with geographic distance for population pairs separated by up to 100 km, but this relationship weakened when more distant pairs were included. Such attenuation is expected in heterogeneous landscapes where occasional long-distance dispersal or passive movement disrupts fine-scale spatial patterns[217–219]. This pattern aligns with findings from Maliti et al. (2014)[87], who similarly reported low overall differentiation in *An. arabiensis* but a detectable IBD signal at local spatial scales. Together, these results support a stepping-stone model of spatial connectivity, rather than strict panmixia, whereby gene flow is highest among neighbouring populations but remains sufficient across broader regions to maintain country-wide genetic homogeneity.

At the continental scale, genetic differentiation followed a generally continuous rather than fragmented pattern: similarity declined gradually with distance, consistent with isolation by distance under a stepping-stone model [87,210]. Superimposed on this background, ecological barriers and environmental conditions further structured populations. Oceanic separation produced the clearest resistance signal, with Madagascar consistently distinct from mainland populations, reflecting the rarity of transoceanic dispersal [211]. The Central African rainforest also acted as a resistance barrier, dividing western and eastern populations; while partially correlated with distance, mixed-effects models supported an independent contribution [220]. In West Africa, differentiation was generally low, but two Nigerian clusters emerged, with southern populations diverging from northern Nigeria and neighbouring countries. Climatic variables, especially climatic mean diurnal temperature range (MDR) and precipitation, remained significantly associated with genetic distance even after controlling for geographical distance, implicating isolation by environment [221,222]. By contrast, the East African Rift showed no excess differentiation, consistent with *An. arabiensis* ecology, which favours

movement along river valleys, tolerance of varied climates, and human-mediated dispersal [87].

These results align with previous findings that *An. arabiensis* maintains large effective population sizes, broad ecological adaptability, and relatively low regional differentiation across southern and eastern Africa, permitting extensive gene flow [87,88,208,209]. Within Tanzania, low mean FST values and significant isolation by distance parallel earlier observations [87,88]. The present findings, quantify rainforest and oceanic resistance alongside distance in a unified framework, and demonstrate that climatic temperature fluctuations and rainfall, which are potent measures of climate, independently influence West African population structure. Collectively, the evidence suggests that hard barriers such as oceans and dense rainforest shape gene flow in some regions, while climatic variation filters genetic connectivity in others [88,223].

The low genetic differentiation observed across much of Africa highlights extensive connectivity among *An. arabiensis* populations, implying that adaptive alleles, such as those linked to insecticide resistance or host-seeking behaviour, can spread readily across regions. This underscores the need for regional coordination in surveillance and intervention planning. At the same time, isolation-by-resistance signals, such as the ocean around Madagascar and the Central African rainforest, demonstrate that ecological barriers can constrain gene flow, creating potential refugia for susceptible populations or limiting the spread of adaptive variants. For malaria control, this means large-scale programs must anticipate the rapid dissemination of adaptive traits, while fine-scale genomic architecture and ecological barriers may modulate intervention effectiveness, including, but not limited to the use of genetically modified mosquitoes.

The results presented here highlight the broader value of the ANOSPP panel beyond its initial role in mosquito species identification [96–98]. By revealing both continental-scale connectivity and local genomic structuring in *An. arabiensis*, the panel demonstrates sufficient resolution to capture processes that are central to malaria control, such as the potential for spread of adaptive alleles, ecological barriers to gene flow, and signatures of inversions. Importantly, this was achieved at much lower cost compared to whole-genome sequencing, making ANOSPP a practical option for national programs and cross-border surveillance efforts. Adoption of the ANOSPP panel could therefore accelerate the integration of genomic

data into malaria vector control strategies, enabling public health stakeholders to monitor species distribution, assess the potential impact of ecological barriers, and evaluate the population-genetic context for emerging interventions such as genetically modified mosquitoes. In addition to its utility for routine surveillance, the ANOSPP panel can act as a genomic "search engine," highlighting populations or genomic regions where deeper investigation is warranted. For example, localized signals of isolation by environment, or elevated differentiation between island and mainland populations can flag candidate targets for follow-up using whole-genome sequencing or ecological and behavioural assays. This tiered approach allows malaria control programs and researchers to allocate resources efficiently using ANOSPP to scan broadly for signals of divergence, then focusing advanced methods on the most biologically and operationally relevant cases.

This study is not without limitations. Although missing data were minimized through stringent site filtering, some genotypes required imputation to ensure full specimen representation across loci. Comparative evaluation of zero, mean, and random-forest (*missForest*)[135] imputation methods showed consistent clustering and ordination structures, indicating that imputation did not introduce artificial genetic patterns. Instead, *missForest* [135] enhanced within-cluster coherence and clarified existing population boundaries without altering the underlying topology, confirming that the results are robust to the imputation process.

Beyond data completeness, the inherent genomic scope of the amplicon panel also imposes limitations. Amplicon-based assays survey only a small fraction of the genome, leaving uncharacterized adaptive variants in unlinked regions, soft sweeps, and structural variants undetected. Spatial and temporal sampling was uneven, which may constrain our ability to quantify barrier strengths or detect subtle temporal shifts, while seasonal variation could bias apparent population structure.

Chapter Six: Conclusions and Future Work

This thesis provides the first comprehensive genomic and ecological overview of *Anopheles* mosquitos across Tanzania, placing these findings within a wider African context to inform malaria control and surveillance. Ecological surveys revealed substantial heterogeneity in mosquito community composition, abundance, co-occurrence patterns and parasite distributions across districts, underscoring the complexity of transmission dynamics. By applying the ANOSPP molecular platform alongside traditional morphology, the study demonstrated clear gains in taxonomic resolution, and for the first time in Tanzania, *Plasmodium caprae* in *An. arabiensis*. These results refine the ecological baseline necessary for surveillance and highlight the multifaceted nature of vector–parasite interactions.

Building on these observations, species distribution models integrated climatic and land-cover variables to generate high-resolution predictions of habitat suitability for Tanzania's three primary malaria vectors. The resulting maps aligned with known transmission zones while revealing species-specific ecological preferences, providing a practical decision-support tool for prioritizing interventions. These models demonstrate the value of incorporating ecological and environmental data into surveillance, advancing malaria control from coarse, reactive approaches to spatially precise and predictive systems.

At the population-genomic level, *An. arabiensis* in Tanzania showed strong overall connectivity with significant isolation by distance at the country scale, reflecting widespread gene flow over tens of kilometres. When extended to the continental scale, patterns reflected both geographic and ecological barriers: Madagascar was sharply separated from the mainland by oceanic isolation, while the Central African rainforest divided eastern and western populations, with additional substructure in West Africa associated with climatic gradients. These results reveal a layered reality in which broad connectivity facilitates cohesion across landscapes, while localized genomic architecture preserves ecological adaptation.

The implications for malaria control are significant. Extensive connectivity suggests that adaptive alleles such as those linked to insecticide resistance could spread rapidly across regions, underscoring the need for coordinated, cross-border surveillance and intervention planning. Looking to the future, the same connectivity that allows natural adaptive alleles to

spread could also accelerate the dispersal of engineered constructs, such as gene drive, should genetically modified mosquitoes be released, highlighting the necessity of regional governance and monitoring frameworks. This study also demonstrates the value of ANOSPP as a scalable genomic surveillance tool. Beyond species identification, the platform can function as a "search engine," pinpointing populations, genomic regions, or ecological contexts that warrant deeper exploration with whole-genome sequencing or functional assays. Its cost-effectiveness and ability to capture both ecological and genetic signals make it a strong candidate for routine surveillance programs.

Limitations of this study should be acknowledged. The data represent a temporal snapshot, limiting inference about seasonal turnover and long-term dynamics. The ANOSPP amplicon panel, while efficient, provides extremely limited genomic coverage compared to whole-genome approaches, restricting resolution of signals. Sampling gaps, particularly in West Africa, constrain the generalizability of continental patterns. Species distribution models also carry inherent uncertainties due to environmental predictor resolution and model assumptions. These constraints shape interpretation and inform priorities for future research. Future work should therefore prioritize increasing temporal resolution to capture seasonal and interannual dynamics of connectivity, adaptation, and turnover. Long-read sequencing and targeted assays will be crucial for mapping inversion breakpoints and monitoring their field frequencies. Denser sampling across ecological transitions, would help disentangle demographic from environmental drivers of structure. Integration of genomic data with ecological surveys and movement networks offers a powerful pathway to forecast resistance spread and optimize sentinel site placement, ensuring surveillance systems are adaptive and anticipatory.

By integrating ecological surveys, predictive spatial modelling, and genomic analysis, this thesis advances a multi-layered framework for malaria vector surveillance in Tanzania and beyond. It shows that ANOSPP can replicate and extend established insights into *An. arabiensis* while uncovering new dimensions of connectivity, local structuring, and parasite associations. The findings emphasize that effective malaria control requires attention to both broad-scale connectivity and fine-scale variation, moving surveillance from reactive monitoring to predictive, integrated systems. This contribution strengthens the scientific foundation for precise, efficient, and adaptive vector control, supporting the long-term goal of malaria elimination.

References

1. World Health Organization. World malaria report 2023 [Internet]. Geneva, Switzerland: World Health Organization; 2023 Nov. Available from: https://cdn.who.int/media/docs/default-source/malaria/world-malaria-reports/world-malaria-report-2023.pdf

2. Bakken L, Iversen PO. The impact of malaria during pregnancy on low birth weight in East-Africa: a topical review. Malaria Journal [Internet]. 2021;20:1–9. Available from: https://doi.org/10.1186/s12936-021-03883-z

3. Chua CLL, Khoo SKM, Ong JLE, Ramireddi GK, Yeo TW, Teo A. Malaria in pregnancy: From placental infection to its abnormal development and damage. Front Microbiol [Internet]. 2021;12:777343. Available from: http://dx.doi.org/10.3389/fmicb.2021.777343

4. Garrison A, Boivin MJ, Fiévet N, Zoumenou R, Alao JM, Massougbodji A, et al. The effects of malaria in pregnancy on neurocognitive development in children at 1 and 6 years of age in Benin: A prospective mother-child cohort. Clin Infect Dis [Internet]. 2022;74:766–75. Available from: https://pubmed.ncbi.nlm.nih.gov/34297062/

5. Schantz-Dunn J, Nour NM. Malaria and pregnancy: a global health perspective. Rev Obstet Gynecol [Internet]. 2009;2:186–92. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC2760896/

6. Ashley EA, Phyo AP. Drugs in development for malaria. Drugs [Internet]. 2018;78:861–79. Available from: http://dx.doi.org/10.1007/s40265-018-0911-9

7. Flegg JA, Kandanaarachchi S, Guerin PJ, Dondorp AM, Nosten FH, Otienoburu SD, et al. Spatio-temporal spread of artemisinin resistance in Southeast Asia. PLoS Comput Biol [Internet]. 2024;20:e1012017. Available from: https://scholar.google.com/citations?view_op=view_citation&hl=en&citation_for_view=peoal7wAAAAJ:yaBp1wUtcLsC

8. Plewes K, Leopold SJ, Kingston HWF, Dondorp AM. Malaria: What's new in the management of malaria? Infect Dis Clin North Am [Internet]. 2019;33:39–60. Available from:

https://www.sciencedirect.com/science/article/pii/S0891552018300898

9. Uwimana A, Umulisa N, Venkatesan M, Svigel SS, Zhou Z, Munyaneza T, et al. Association of Plasmodium falciparum kelch13 R561H genotypes with delayed parasite clearance in Rwanda: an open-label, single-arm, multicentre, therapeutic efficacy study. Lancet Infect Dis [Internet]. 2021;21:1120–8. Available from: http://dx.doi.org/10.1016/S1473-3099(21)00142-0

10. Amato R, Pearson RD, Almagro-Garcia J, Amaratunga C, Lim P, Suon S, et al. Origins of the current outbreak of multidrug-resistant malaria in southeast Asia: a retrospective genetic study. Lancet Infect Dis [Internet]. 2018;18:337–45. Available from: https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(18)30068-9/fulltext

11. Amato A, Miotto O, Woodrow C, Almagro-Garcia J, Sinha I, Campino S, et al. Genomic epidemiology of artemisinin resistant malaria. Elife [Internet]. 2016;5. Available from: https://riip.hal.science/pasteur-01971950/

12. Ariey F, Witkowski B, Amaratunga C, Beghain J, Langlois A-C, Khim N, et al. A molecular marker of artemisinin-resistant Plasmodium falciparum malaria. Nature [Internet]. 2014;505:50–5. Available from: http://dx.doi.org/10.1038/nature12876

13. Ariey F. Relevance of K13 mutations for malaria control and elimination program. Malar J [Internet]. 2014;13:O40. Available from: http://dx.doi.org/10.1186/1475-2875-13-s1-o40

14. Hodoameda P, Duah-Quashie NO, Quashie NB. Assessing the roles of molecular markers of antimalarial drug resistance and the host pharmacogenetics in drug-resistant malaria. J Trop Med [Internet]. 2022;2022:3492696. Available from: http://dx.doi.org/10.1155/2022/3492696

15. Hodoameda P. P. Falciparum and its molecular markers of resistance to antimalarial drugs. In: Tyagi RK, editor. Plasmodium Species and Drug Resistance [Internet]. IntechOpen; 2021. Available from: https://www.intechopen.com/chapters/77461

16. Mihreteab S, Anderson K, Pasay C, Smith D, Gatton ML, Cunningham J, et al. Epidemiology of mutant Plasmodium falciparum parasites lacking histidine-rich protein 2/3

genes in Eritrea 2 years after switching from HRP2-based RDTs. Sci Rep [Internet]. 2021;11:21082. Available from: https://www.nature.com/articles/s41598-021-00714-8

17. Kong A, Wilson SA, Ah Y, Nace D, Rogier E, Aidoo M. HRP2 and HRP3 cross-reactivity and implications for HRP2-based RDT use in regions with Plasmodium falciparum hrp2 gene deletions. Malar J [Internet]. 2021;20:207. Available from: http://dx.doi.org/10.1186/s12936-021-03739-6

18. Gatton ML, Smith D, Pasay C, Anderson K, Mihreteab S, Valdivia HO, et al. Comparison of prevalence estimates of pfhrp2 and pfhrp3 deletions in Plasmodium falciparum determined by conventional PCR and multiplex qPCR and implications for surveillance and monitoring. Int J Infect Dis [Internet]. 2024;144:107061. Available from: https://www.sciencedirect.com/science/article/pii/S1201971224001322

19. Rogier E, Battle N, Bakari C, Seth M, Nace D, Herman C, et al. Plasmodium falciparum pfhrp2 and pfhrp3 gene deletions among patients enrolled at 100 health facilities throughout Tanzania: February to July 2021. Scientific Reports [Internet]. 2023;14. Available from: https://www.nature.com/articles/s41598-024-58455-3

20. Ngasala B, Chacky F, Mohamed A, Molteni F, Nyinondi S, Kabula B, et al. Evaluation of malaria rapid diagnostic test performance and pfhrp2 deletion in Tanzania school surveys, 2017. Am J Trop Med Hyg [Internet]. 2024;110:887–91. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC11066367/

21. Churcher TS, Lissenden N, Griffin JT, Worrall E, Ranson H. The impact of pyrethroid resistance on the efficacy and effectiveness of bednets for malaria control in Africa. Elife [Internet]. 2016;5:1–26. Available from: http://dx.doi.org/10.7554/eLife.16090

22. Ranson H, Lissenden N. Insecticide Resistance in African Anopheles Mosquitoes: A Worsening Situation that Needs Urgent Action to Maintain Malaria Control. Trends Parasitol [Internet]. 2016;32:187–96. Available from: http://dx.doi.org/10.1016/j.pt.2015.11.010

23. Corbel V, N'Guessan R, Brengues C, Chandre F, Djogbenou L, Martin T, et al. Multiple insecticide resistance mechanisms in Anopheles gambiae and Culex quinquefasciatus from

Benin, West Africa. Acta Tropica [Internet]. 2007;101:207–16. Available from: http://dx.doi.org/10.1016/j.actatropica.2007.01.005

24. Donnelly MJ, Corbel V, Weetman D, Wilding CS, Williamson MS, Black WC IV. Does kdr genotype predict insecticide-resistance phenotype in mosquitoes? Trends in Parasitology [Internet]. 2009;25:213–9. Available from: http://dx.doi.org/10.1016/j.pt.2009.02.007

25. Govella NJ, Chaki PP, Killeen GF. Entomological surveillance of behavioural resilience and resistance in residual malaria vector populations. Malar J [Internet]. 2013;12:1–9. Available from: http://dx.doi.org/10.1186/1475-2875-12-124

26. Govella NJ, Ferguson H. Why use of interventions targeting outdoor biting mosquitoes will be necessary to achieve malaria elimination. Front Physiol [Internet]. 2012;3 JUN:1–5. Available from: http://dx.doi.org/10.3389/fphys.2012.00199

27. Ibrahim SS, Ndula M, Riveron JM, Irving H, Wondji CS. The P450 CYP6Z1 confers carbamate/pyrethroid cross-resistance in a major African malaria vector beside a novel carbamate-insensitive N485I acetylcholinesterase-1 mutation. Molecular ecology [Internet]. 2016;25:3436–52. Available from: http://dx.doi.org/10.1111/mec.13673

28. Ibrahim SS, Muhammad A, Hearn J, Weedall GD, Nagi SC, Mukhtar MM, et al. Molecular drivers of insecticide resistance in the Sahelo - Sudanian populations of a major malaria vector Anopheles coluzzii. BMC Biol [Internet]. 2023;1–24. Available from: https://doi.org /10.1186/s12915-023-01610-5

29. Ingham VA, Wagstaff S, Ranson H. Transcriptomic meta-signatures identified in Anopheles gambiae populations reveal previously undetected insecticide resistance mechanisms. Nat Commun [Internet]. 2018;9. Available from: http://dx.doi.org/ 10.1038/s41467-018-07615-x

30. Liu N. Insecticide resistance in mosquitoes: Impact, mechanisms, and research directions. Annual Review of Entomology [Internet]. 2015. p. 537–59. Available from: http://dx.doi.org/10.1146/annurev-ento-010814-020828

31. Mmbaga AT, Lwetoijera DW. Current and future opportunities of autodissemination of pyriproxyfen approach for malaria vector control in urban and rural Africa. Wellcome Open Res [Internet]. 2023;8:119. Available from: http://dx.doi.org/10.12688/well comeopenres. 19131.2

32. Ferguson HM, Dornhaus A, Beeche A, Borgemeister C, Gottlieb M, Mulla MS, et al. Ecology: a prerequisite for malaria elimination and eradication. PLoS Med [Internet]. 2010;7:e1000303. Available from: https://journals.plos.org/plosmedicine/article? id=10.1371/ journal.pmed.1000303

33. World Health Organisation. Handbook for Integrated Vector Management [Internet]. World Health Organisation; 2012. Available from: https://www.ivcc.com/wp-content/uploads /2019/07/Handbook-for-Integrated-Vector-Management.pdf

34. Coulibaly MB, Traoré SF, Touré YT. Considerations for Disrupting Malaria Transmission in Africa Using Genetically Modified Mosquitoes, Ecology of Anopheline Disease Vectors, and Current Methods of Control. Genetic Control of Malaria and Dengue [Internet]. 2016;55–67. Available from: http://dx.doi.org/10.1016/B978-0-12-800246-9.00003-X

35. Okumu F, Gyapong M, Casamitjana N, Castro MC, Itoe MA, Okonofua F, et al. What Africa can do to accelerate and sustain progress against malaria. PLOS Glob Public Health [Internet]. 2022;2:e0000262. Available from: http://dx.doi.org/10.1371/journal.pgph.0000262

36. Stanton MC, Kalonde P, Zembere K, Hoek Spaans R, Jones CM. The application of drones for mosquito larval habitat identification in rural environments: a practical approach for malaria control? Malar J [Internet]. 2021;20:244. Available from: http://dx.doi.org/10.1186/s12936-021-03759-2

37. Hardy A, Haji K, Abbas F, Hassan J, Ali A, Yussuf Y, et al. Cost and quality of operational larviciding using drones and smartphone technology. Malar J [Internet]. 2023 [cited 2025 July 31];22:286. Available from: http://dx.doi.org/10.1186/s12936-023-04713-0

38. Carrasco-Escobar G, Manrique E, Ruiz-Cabrejos J, Saavedra M, Alava F, Bickersmith S, et al. High-accuracy detection of malaria vector larval habitats using drone-based multispectral

imagery. PLoS Negl Trop Dis [Internet]. 2019;13:e0007105. Available from: https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0007105

39. Hardy A, Oakes G, Hassan J, Yussuf Y. Improved use of drone imagery for malaria vector control through Technology-assisted digitizing (TAD). Remote Sens (Basel) [Internet]. 2022;14:317. Available from: https://www.mdpi.com/2072-4292/14/2/317

40. Herren JK, Mbaisi L, Mararo E, Makhulu EE, Mobegi VA, Butungi H, et al. A microsporidian impairs Plasmodium falciparum transmission in Anopheles arabiensis mosquitoes. Nat Commun [Internet]. 2020;11:2187. Available from: https://www.nature.com /articles/s41467-020-16121-y

41. Bukhari T, Pevsner R, Herren JK. Microsporidia: a promising vector control tool for residual malaria transmission. Front Trop Dis [Internet]. 2022;3. Available from: http://dx.doi.org/10.3389/fitd.2022.957109

42. Jain H, Sinha AK. Modeling the effect of Wolbachia to control malaria transmission. Expert Syst Appl [Internet]. 2023;221:119769. Available from: https://www.sciencedirect. com /science/article/pii/S0957417423002701

43. Vandana V, Dong S, Sheth T, Sun Q, Wen H, Maldonado A, et al. Wolbachia infection-responsive immune genes suppress Plasmodium falciparum infection in Anopheles stephensi. PLoS Pathog [Internet]. 2024;20:e1012145. Available from: https://journals.plos.org/ plospathogens/article?id=10.1371/journal.ppat.1012145

44. Cansado-Utrilla C, Zhao SY, McCall PJ, Coon KL, Hughes GL. The microbiome and mosquito vectorial capacity: rich potential for discovery and translation. Microbiome [Internet]. 2021;9:111. Available from: http://dx.doi.org/10.1186/s40168-021-01073-2

45. Harbach RE, Kitching IJ. The phylogeny of Anophelinae revisited: inferences about the origin and classification of *Anopheles* (Diptera: Culicidae). Zool Scr [Internet]. 2016;45:34–47. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/zsc.12137

46. Marinotti O, Cerqueira GC, de Almeida LGP, Ferro MIT, Loreto EL da S, Zaha A, et al.

The genome of Anopheles darlingi, the main neotropical malaria vector. Nucleic Acids Res [Internet]. 2013;41:7387–400. Available from: http://dx.doi.org/10.1093/nar/gkt484

47. Anopheles gambiae 1000 Genomes Consortium. Genetic diversity of the African malaria vector Anopheles gambiae. Nature [Internet]. 2017;552:96–100. Available from: https://www.nature.com/articles/nature24995

48. White MT, Griffin JT, Churcher TS, Ferguson NM, Basáñez M-G, Ghani AC. Modelling the impact of vector control interventions on Anopheles gambiae population dynamics. Parasit Vectors [Internet]. 2011;4:153. Available from: http://dx.doi.org/10.1186/1756-3305-4-153

49. Coetzee M. Key to the females of Afrotropical Anopheles mosquitoes (Diptera: Culicidae). Malar J [Internet]. 2020;19:70. Available from: http://dx.doi.org/10.1186/s12936-020-3144-9

50. Gillies. M, Coetzee M. A supplement to the Anophelinae of Africa south of the Sahara (Afrotropical Region). 1987

51. Rattanarithikul R, Panthusiri P. Illustrated keys to the medically important mosquitos of Thailand. Southeast Asian J Trop Med Public Health [Internet]. 1994;25 Suppl 1:1–66. Available from: https://www.ncbi.nlm.nih.gov/pubmed/7831585

52. Cohuet A, Simard F, Toto J-C, Kengne P, Coetzee M, Fontenille D. Species identification within the Anopheles funestus group of malaria vectors in Cameroon and evidence for a new species. Am J Trop Med Hyg [Internet]. 2003;69:200–5. Available from: http://dx.doi.org/10.4269/ajtmh.2003.69.200

53. 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 [Internet]. 1993;49:520–9. Available from: http://dx.doi.org/10.4269/ajtmh.1993.49.520

54. Fanello C, Santolamazza F, della Torre A. Simultaneous identification of species and molecular forms of the Anopheles gambiae complex by PCR-RFLP. Med Vet Entomol [Internet]. 2002;16:461–4. Available from: http://dx.doi.org/10.1046/j.1365-2915.2002.00393.x

55. Wilkins EE, Howell PI, Benedict MQ. IMP PCR primers detect single nucleotide polymorphisms for Anopheles gambiae species identification, Mopti and Savanna rDNA types, and resistance to dieldrin in Anopheles arabiensis. Malar J [Internet]. 2006;5:125. Available from: http://dx.doi.org/10.1186/1475-2875-5-125

56. Erlank E, Koekemoer LL, Coetzee M. The importance of morphological identification of African anopheline mosquitoes (Diptera: Culicidae) for malaria control programmes. Malar J [Internet]. 2018;17:43. Available from: http://dx.doi.org/10.1186/s12936-018-2189-5

57. Ayala D, Goff GL, Robert V, de Jong P, Takken W. Population structure of the malaria vector Anopheles funestus (Diptera: Culicidae) in Madagascar and Comoros. Acta Trop [Internet]. 2006;97:292–300. Available from: http://dx.doi.org/10.1016/j. actatropica. 2005.12.002

58. Sallum MAM, Bergo ES, Flores DC, Forattini OP. Systematic studies on Anopheles galvaoi Causey, Deane & Deane from the subgenus Nysssorhynchus blanchard (Diptera: Culicidae). Mem Inst Oswaldo Cruz [Internet]. 2002;97:1177–89. Available from: http://dx.doi.org/10.1590/s0074-02762002000800020

59. Stevenson J, St Laurent B, Lobo NF, Cooke MK, Kahindi SC, Oriango RM, et al. Novel vectors of malaria parasites in the western highlands of Kenya. Emerg Infect Dis [Internet]. 2012;18:1547–9. Available from: http://dx.doi.org/10.3201/eid1809.120283

60. Besansky NJ, Krzywinski J, Lehmann T, Simard F, Kern M, Mukabayire O, et al. Semipermeable species boundaries between Anopheles gambiae and Anopheles arabiensis: evidence from multilocus DNA sequence variation. Proc Natl Acad Sci U S A [Internet]. 2003;100:10818–23. Available from: http://dx.doi.org/10.1073/pnas.1434337100

61. Lemmon AR, Emme SA, Lemmon EM. Anchored hybrid enrichment for massively high-throughput phylogenomics. Syst Biol [Internet]. 2012;61:727–44. Available from: http://dx.doi.org/10.1093/sysbio/sys049

62. Ministry of Health (MoH) [Tanzania Mainland], Ministry of Health (MoH) [Zanzibar], National Bureau of Statistics (NBS), Office of the Chief Government Statistician (OCGS), and

ICF. Tanzania Demographic and Health Survey and Malaria Indicator Survey 2022 Key Indicators Report. 2022. Report No.: 1.

63. NATIONAL MALARIA CONTROL PROGRAMME. The 2021 School Malaria and Nutrition Survey (SMNS) Report. MINISTRY OF HEALTH; 2022 July. Report No.: 1.

64. Mapua SA, Samb B, Nambunga IH, Mkandawile G, Bwanaly H, Kaindoa EW, et al. Entomological survey of sibling species in the Anopheles funestus group in Tanzania confirms the role of Anopheles parensis as a secondary malaria vector. Parasit Vectors [Internet]. 2024;17:261. Available from: http://dx.doi.org/10.1186/s13071-024-06348-9

65. Mwalimu CD, Kiware S, Nshama R, Derua Y, Machafuko P, Gitanya P, et al. Dynamics of malaria vector composition and Plasmodium falciparum infection in mainland Tanzania: 2017-2021 data from the national malaria vector entomological surveillance. Malar J [Internet]. 2024;23:29. Available from: http://dx.doi.org/10.1186/s12936-024-04849-7

66. Wilkes TJ, Matola YG, Charlwood JD. Anopheles rivulorum, a vector of human malaria in Africa. Med Vet Entomol [Internet]. 1996;10:108–10. Available from: http://dx.doi.org/10.1111/j.1365-2915.1996.tb00092.x

67. Gillies MT. The role of secondary vectors of malaria in North-East Tanganyika. Trans R Soc Trop Med Hyg [Internet]. 1964;58:154–8. Available from: https://academic.oup.com/trstmh/article-lookup/doi/10.1016/0035-9203(64)90004-5

68. Temu EA, Minjas JN, Tuno N, Kawada H, Takagi M. Identification of four members of the Anopheles funestus (Diptera: Culicidae) group and their role in Plasmodium falciparum transmission in Bagamoyo coastal Tanzania. Acta Trop [Internet]. 2007 [cited 2025 May 30];102:119–25. Available from: http://dx.doi.org/10.1016/j.actatropica.2007.04.009

69. Kopya E, Ndo C, Djamouko-Djonkam L, Nkahe L, Awono-Ambene P, Njiokou F, et al. *Anopheles leesoni* Evans 1931, a Member of the *Anopheles funestus* Group, Is a Potential Malaria Vector in Cameroon. Adv Entomol [Internet]. 2022;10:99–109. Available from: http://dx.doi.org/10.4236/ae.2022.101008

70. Antonio-Nkondjio C, Kerah CH, Simard F, Awono-Ambene P, Chouaibou M, Tchuinkam T, et al. Complexity malaria vector- ial system Cameroon: contribution secondary vectors malaria transmission. J Med Entomol. 2006;43:1215–21.

71. Mwangangi JM, Muturi EJ, Muriu SM, Nzovu J, Midega JT, Mbogo C. The role of Anopheles arabiensis and Anopheles coustani in indoor and outdoor malaria transmission in Taveta District, Kenya. Parasit Vectors [Internet]. 2013;6:114. Available from: http://dx.doi.org/10.1186/1756-3305-6-114

72. Nepomichene TNJJ, Tata E, Boyer S. Malaria case in Madagascar, probable implication of a new vector, Anopheles coustani. Malar J [Internet]. 2015;14:475. Available from: http://dx.doi.org/10.1186/s12936-015-1004-9

73. Tedrow RE, Rakotomanga T, Nepomichene T, Howes RE, Ratovonjato J, Ratsimbasoa AC, et al. Anopheles mosquito surveillance in Madagascar reveals multiple blood feeding behavior and Plasmodium infection. PLoS Negl Trop Dis [Internet]. 2019;13:1–21. Available from: http://dx.doi.org/10.1371/journal.pntd.0007176

74. Andrianinarivomanana TM, Randrianaivo FT, Andriamiarimanana MR, Razafimamonjy MR, Velonirina HJS, Puchot N, et al. Colonization of Anopheles coustani, a neglected malaria vector in Madagascar. Parasite [Internet]. 2024;31:31. Available from: https://pmc.ncbi .nlm.nih.gov/articles/PMC11186460/

75. Goupeyou-Youmsi J, Rakotondranaivo T, Puchot N, Peterson I, Girod R, Vigan-Womas I, et al. Differential contribution of Anopheles coustani and Anopheles arabiensis to the transmission of Plasmodium falciparum and Plasmodium vivax in two neighbouring villages of Madagascar. Parasit Vectors [Internet]. 2020;13:430. Available from: https://parasites andvectors.biomedcentral.com/articles/10.1186/s13071-020-04282-0

76. Sendor R, Mitchell CL, Chacky F, Mohamed A, Mhamilawa LE, Molteni F, et al. Similar Prevalence of Plasmodium falciparum and Non–P. falciparum Malaria Infections among Schoolchildren, Tanzania. Emerging Infectious Diseases [Internet]. 2023;29:1143–53. Available from: https://doi.org/10.3201/eid2906.221016

77. Walter Reed Biosystematics Unit (WRBU) | Smithsonian Institution. Mosquitoes [Internet]. Walter Reed Biosystematics Unit (WRBU) | Smithsonian Institution. 2021 [cited 2024 Oct 30]. Available from: https://tinyurl.com/yzvfp76e

78. Takola E, Schielzeth H. Hutchinson's ecological niche for individuals. Biol Philos [Internet]. 2022;37. Available from: http://dx.doi.org/10.1007/s10539-022-09849-y

79. Chase JM. Ecological niche theory. 2011; Available from: http://dx.doi.org/10.7208/9780226736877-006

80. Morrow KH. Niches and niche models. Br J Philos Sci [Internet]. 2024; Available from: http://dx.doi.org/10.1086/730329

81. 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 [Internet]. 2010;3:117. Available from: http://dx.doi.org/10.1186/1756-3305-3-117

82. Sendor R, Mitchell C, Chacky F, Mohamed A, Mhamilawa LE, Molteni F, et al. LOW PREVALENCE OF PLASMODIUM MALARIAE AND P. VIVAX, AND HIGH PREVALENCE OF P. OVALE DETECTED AMONG TANZANIAN SCHOOL CHILDREN WITHIN THE 2017 SCHOOL MALARIA PARASITEMIA SURVEY …. 2021;105:73–73. Available from: https://scholar.google.com/citations?view_op=view_citation&hl=en&citation_for_view=qLxBXeoAAAAJ:e5wmG9Sq2KIC

83. Kotepui M, Kotepui KU, De Jesus Milanez G, Masangkay FR. Plasmodium spp. mixed infection leading to severe malaria: a systematic review and meta-analysis. Sci Rep [Internet]. 2020;10:11068. Available from: http://dx.doi.org/10.1038/s41598-020-68082-3

84. Vythilingam I, Jeyaprakasam NK. Deforestation and non-human primate malarias will be a threat to malaria elimination in the future: Insights from Southeast Asia. Acta Trop [Internet]. 2024;257:107280. Available from: http://dx.doi.org/10.1016/j.actatropica.2024.107280

85. Odero JO, Nambunga IH, Bwanary H, Mkandawile G, Paliga JM, Mapua SA, et al. Distinct

genetic populations and resistance backgrounds of the malaria vector*anopheles funestus* in Tanzania [Internet]. bioRxiv. 2025. p. 2025.01. 21.634154. Available from: https://www.biorxiv.org/content/10.1101/2025.01.21.634154.abstract

86. Ng'habi KR, Knols BGJ, Lee Y, Ferguson HM, Lanzaro GC. Population genetic structure of Anopheles arabiensis and Anopheles gambiae in a malaria endemic region of southern Tanzania. Malar J [Internet]. 2011;10:289. Available from: http://dx.doi.org/10.1186/1475-2875-10-289

87. Maliti D, Ranson H, Magesa S, Kisinza W, Mcha J, Haji K, et al. Islands and stepping-stones: comparative population structure of Anopheles gambiae sensu stricto and Anopheles arabiensis in Tanzania and implications for the spread of insecticide resistance. PLoS One [Internet]. 2014;9:e110910. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0110910

88. Mwinyi SH, Bennett KL, Nagi SC, Kabula B, Matowo J, Weetman D, et al. Genomic analysis reveals a new cryptic taxon within the Anopheles gambiae complex with a distinct insecticide resistance profile in the coast of east Africa. Mol Ecol [Internet]. 2025;e17762. Available from: http://dx.doi.org/10.1111/mec.17762

89. Maliti DV, Govella NJ, Killeen GF, Mirzai N, Johnson PCD, Kreppel K, et al. Development and evaluation of mosquito-electrocuting traps as alternatives to the human landing catch technique for sampling host-seeking malaria vectors. Malar J [Internet]. 2015;14:1–15. Available from: http://dx.doi.org/10.1186/s12936-015-1025-4

90. Githu V, Baravuga ME, Mbarawa A, Msuya HM, Mlacha YP, Chaki PP, et al. Comparative evaluation of different versions of exposure - free mosquito electrocuting traps and barrier screen trap for monitoring outdoor densities and biting time phenotypes by malaria and filariasis vectors in Tanzania. Parasit Vectors [Internet]. 2022;15:1–8. Available from: https://doi.org/10.1186/s13071-022-05549-4

91. Meza FC, Kreppel KS, Maliti DF, Mlwale AT, Mirzai N, Killeen GF, et al. Mosquito electrocuting traps for directly measuring biting rates and host-preferences of Anopheles arabiensis and Anopheles funestus outdoors. Malar J [Internet]. 2019;18:1–11. Available from:

https://doi.org/10.1186/s12936-019-2726-x

92. Maia MF, Robinson A, John A, Mgando J, Simfukwe E, Moore SJ. Comparison of the CDC Backpack aspirator and the Prokopack aspirator for sampling indoor- and outdoor-resting mosquitoes in southern Tanzania. Parasit Vectors [Internet]. 2011;4:124. Available from: http://dx.doi.org/10.1186/1756-3305-4-124

93. T.r. B, T.l. R, L.j. R, H. B, N.w. B, R.d. C, et al. Barrier screens: A method to sample blood-fed and host-seeking exophilic mosquitoes. Malar J [Internet]. 2013;12:1–9. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed11&NEWS=N&AN=2013115588

94. Convention on Biological Diversity United Nation. Nagoya protocol on access to genetic resources and the fair and equitable sharing of benefits arising from their utilization to the convention on biological diversity [Internet]. Convention on Biological Diversity. 2011 [cited 2024 Oct 30]. Available from: https://www.cbd.int/abs/text

95. Korlevi P, Mcalister E, Lawniczak MKN, Mayho M, Makunin A, Flicek P. GBE A Minimally Morphologically Destructive Approach for DNA Retrieval and Whole-Genome Shotgun Sequencing of Pinned. 2021;13:1–13. Available from: http://dx.doi.org/10.1093/gbe/evab226

96. Makunin A, Korlević P, Park N, Goodwin S, Waterhouse RM, Wyschetzki KV, et al. A targeted amplicon sequencing panel to simultaneously identify mosquito species and Plasmodium presence across the entire Anopheles genus. 2022;28–44. Available from: http://dx.doi.org/10.1111/1755-0998.13436

97. Boddé M, Makunin A, Ayala D, Bouafou L, Diabaté A, Ekpo UF, et al. High- resolution species assignment of Anopheles mosquitoes using k- mer distances on targeted sequences. eLife [Internet]. 2022;11:1–40. Available from: http://dx.doi.org/10.7554/eLife.78775

98. Boddé M, Makunin A, Teltscher F, Akorli J, Andoh NE, Bei A, et al. Improved species assignments across the entire Anopheles genus using targeted sequencing. Front Genet [Internet]. 2024;15:1456644. Available from: http://dx.doi.org/10.3389/fgene.2024.1456644

99. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol [Internet]. 2013;30:772–80. Available from: http://dx.doi.org/10.1093/molbev/mst010

100. Wong TKF, Ly-Trong N, Ren H, Baños H, Roger AJ. IQ-TREE 3: Phylogenomic Inference Software using Complex Evolutionary Models. 2025; Available from: https://ecoevorxiv.org/repository/view/8916/

101. Karger DN, Conrad O, Böhner J, Kawohl T, Kreft H, Soria-Auza RW, et al. Climatologies at high resolution for the earth's land surface areas. Sci Data [Internet]. 2017;4:170122. Available from: http://dx.doi.org/10.1038/sdata.2017.122

102. Buchhorn M, Lesiv M, Tsendbazar N-E, Herold M, Bertels L, Smets B. Copernicus Global Land Cover layers—collection 2. Remote Sens (Basel) [Internet]. 2020;12:1044. Available from: http://dx.doi.org/10.3390/rs12061044

103. WorldPop, Bondarenko M. Individual countries 1km population density (2000-2020) [Internet]. University of Southampton; 2020. Available from: http://dx.doi .org/10.5258/SOTON/WP00674

104. Gilbert M, Nicolas G, Cinardi G, Van Boeckel TP, Vanwambeke SO, Wint GRW, et al. Global distribution data for cattle, buffaloes, horses, sheep, goats, pigs, chickens and ducks in 2010. Sci Data [Internet]. 2018;5:180227. Available from: http://dx. doi.org/10. 1038/sdata.2018.227

105. Didan K. MODIS/Terra Vegetation Indices Monthly L3 Global 1km SIN Grid V061 [Internet]. NASA Land Processes Distributed Active Archive Center; 2021. Available from: http://dx.doi.org/10.5067/MODIS/MOD13A3.061

106. USGS-U.S. Geological Survey, LP DAAC-Land Processes Distributed Active Archive Center. AppEEARS [Internet]. [cited 2025 July 9]. Available from: https://appeears.earthdatacloud.nasa.gov

107. Beck HE, Zimmermann NE, McVicar TR, Vergopolan N, Berg A, Wood EF. Present and

future Köppen-Geiger climate classification maps at 1-km resolution. Sci Data [Internet]. 2018;5:180214. Available from: https://www.nature.com/articles/sdata2018214

108. Hijmans RJ. terra: Spatial Data Analysis [Internet]. 2023. Available from: https://CRAN.R-project.org/package=terra

109. Mission NSR. Shuttle Radar Topography Mission (SRTM) Global [Internet]. OpenTopography. OpenTopography; 2013 [cited 2025 July 14]. Available from: http://dx.doi.org/10.5069/G9445JDF

110. Takken W, Charlwood JD, Billingsley PF, Gort G. Dispersal and survival of Anopheles funestus and A. gambiae s.l. (Diptera: Culicidae) during the rainy season in southeast Tanzania. Bull Entomol Res [Internet]. 1998;88:561–6. Available from: https://www.cambridge.org/core/journals/bulletin-of-entomologicalresearch/article/dispersal-and-survival-of-anopheles-funestus-and-a-gambiae-sl-diptera-culicidae-during-the-rainy-season-in-southeast-tanzania/7CA1C2409629A551BFD3654756686406

111. Thomson MC, Connor SJ, Quiñones ML, Jawara M, Todd J, Greenwood BM. Movement of Anopheles gambiae s.l. malaria vectors between villages in The Gambia. Med Vet Entomol [Internet]. 1995;9:413–9. Available from: https://pubmed.ncbi.nlm.nih.gov/8541594/

112. Midega JT, Mbogo CM, Mwnambi H, Wilson MD, Ojwang G, Mwangangi JM, et al. Estimating dispersal and survival of Anopheles gambiae and Anopheles funestus along the Kenyan coast by using mark-release-recapture methods. J Med Entomol [Internet]. 2007;44:923–9. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC2705338/

113. Charlwood JD, Graves PM, Birley MH. Capture-recapture studies with mosquitoes of the group of Anopheles punctulatus Dönitz (Diptera: Culicidae) from Papua New Guinea. Bull Entomol Res [Internet]. 1986;76:211–27. Available from: https://www.cambridge.org/core/journals/bulletin-of-entomological-research/article/capturerecapture-studies-with-mosquitoes-of-the-group-of-anopheles-punctulatus-donitz-diptera-culicidae-from-papua-new-guinea/7919EA893FD4DE371F4CECD104BFDC6C

114. Hijmans RJ. raster: Geographic Data Analysis and Modeling [Internet]. 2023. Available

from: https://CRAN.R-project.org/package=raster

115. Wickham H, Henry L, Hester J, Bryan J, Woo K, Yutani H, et al. tidyverse: Easily Install and Load the Tidyverse [Internet]. 2023. Available from: https://CRAN.R-project.org/package=tidyverse

116. Oksanen J, Blanchet FG, Kindt R, Legendre P, O'Hara B, Simpson GL, et al. Package 'vegan': Community ecology package. R package version 2 0-10 [Internet]. 2013; Available from: https://cran.r-project.org/package=vegan

117. Pebesma E. Simple features for R: Standardized support for spatial vector data. R J [Internet]. 2018;10:439. Available from: http://dx.doi.org/10.32614/rj-2018-009

118. South A. rnaturalearth: World Map Data from Natural Earth [Internet]. 2017. Available from: https://github.com/ropensci/rnaturalearth

119. Wickham H. Ggplot2: Elegant graphics for data analysis [Internet]. 2nd edn. Cham, Switzerland: Springer International Publishing; 2016. Available from: https://link.springer.com/book/10.1007/978-3-319-24277-4

120. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics [Internet]. 2016;32:2847–9. Available from: https://doi.org/10.1093/bioinformatics/btw313

121. Ryan SJ, McNally A, Johnson LR, Mordecai EA, Ben-Horin T, Paaijmans K, et al. Mapping physiological suitability limits for malaria in Africa under climate change. Vector Borne Zoonotic Dis [Internet]. 2015 [cited 2025 July 25];15:718–25. Available from: http://dx.doi.org/10.1089/vbz.2015.1822

122. Chapman D, Pescott OL, Roy HE, Tanner R. Improving species distribution models for invasive non-native species with biologically informed pseudo-absence selection. J Biogeogr [Internet]. 2019;46:1029–40. Available from: http://dx.doi.org/10.1111/jbi.13555

123. Marcelino VR, Verbruggen H. Ecological niche models of invasive seaweeds. J Phycol

[Internet]. 2015;51:606–20. Available from: https://pubmed.ncbi.nlm.nih.gov/26986785/

124. Kearney M, Porter W. Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. Ecol Lett [Internet]. 2009;12:334–50. Available from: http://dx.doi.org/10.1111/j.1461-0248.2008.01277.x

125. Gamliel I, Buba Y, Guy-Haim T, Garval T, Willette D, Rilov G, et al. Incorporating physiology into species distribution models moderates the projected impact of warming on selected Mediterranean marine species. Ecography (Cop) [Internet]. 2020;43:1090–106. Available from: http://dx.doi.org/10.1111/ecog.04423

126. Sinka ME, Bangs MJ, Manguin S, Rubio-Palis Y, Chareonviriyaphap T, Coetzee M, et al. A global map of dominant malaria vectors. Parasit Vectors [Internet]. 2012;5:69. Available from: http://dx.doi.org/10.1186/1756-3305-5-69

127. Wood SN. mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation. 2025.

128. Hartig F. DHARMa: residual diagnostics for hierarchical (multi-level/mixed) regression models. CRAN: Contributed Packages [Internet]. 2016; Available from: https://cir.nii.ac.jp/crid/1360583646752175744

129. Valavi R, Elith J, Lahoz-Monfort JJ, Guillera-Arroita G. blockCV: an R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models [Internet]. bioRxiv. bioRxiv; 2018. Available from: http://dx.doi.org/10.1101/357798

130. Harrell FE Jr. rms: Regression Modeling Strategies. R package version 6.2-0. 2021. 2023

131. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods [Internet]. 2016;13:581–3. Available from: http://dx.doi.org/10.1038/nmeth.3869

132. Zamyatin A, Avdeyev P, Liang J, Sharma A, Chen C, Lukyanchikova V, et al.

Chromosome-level genome assemblies of the malaria vectors Anopheles coluzzii and Anopheles arabiensis. Gigascience [Internet]. 2021;10:giab017. Available from: https://academic.oup.com/gigascience/article/10/3/giab017/6170950

133. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics [Internet]. 2010;26:841–2. Available from: http://dx.doi.org/10.1093/bioinformatics/btq033

134. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics [Internet]. 2011;27:2156–8. Available from: http://dx.doi.org/10.1093/bioinformatics/btr330

135. Stekhoven DJ, Bühlmann P. MissForest--non-parametric missing value imputation for mixed-type data. Bioinformatics [Internet]. 2012;28:112–8. Available from: https://academic.oup.com/bioinformatics/article-abstract/28/1/112/219101

136. Poland J, Endelman J, Dawson J, Rutkoski J, Wu S, Manes Y, et al. Genomic Selection in Wheat Breeding using Genotyping-by-Sequencing. Plant Genome [Internet]. 2012;5:103–13. Available from: http://dx.doi.org/10.3835/plantgenome2012.06.0006

137. Poland JA, Rife TW. Genotyping-by-sequencing for plant breeding and genetics. Plant Genome [Internet]. 2012;5:92–102. Available from: http://dx.doi.org/10.3835/plantgenome2012.05.0005

138. Xavier A, Muir WM, Rainey KM. Impact of imputation methods on the amount of genetic variation captured by a single-nucleotide polymorphism panel in soybeans. BMC Bioinformatics [Internet]. 2016;17:55. Available from: http://dx.doi.org/10.1186/s12859-016-0899-7

139. Jombart T. *adegenet*: a R package for the multivariate analysis of genetic markers. Bioinformatics [Internet]. 2008;24:1403–5. Available from: http://dx.doi.org/10.1093/bioinformatics/btn129

140. Smouse PE, Long JC, Sokal RR. Multiple regression and correlation extensions of the

mantel test of matrix correspondence. Syst Zool [Internet]. 1986;35:627. Available from: http://dx.doi.org/10.2307/2413122

141. Wang L. A method for key updating of IBE with wildcards. Adv Mat Res [Internet]. 2013;765–767:1003–6. Available from: http://dx.doi.org/10.4028/www.scientific.net/amr.765-767.1003

142. Weir BS, Cockerham CC. Estimating f-statistics for the analysis of population structure. Evolution [Internet]. 1984;38:1358–70. Available from: https://pubmed.ncbi.nlm.nih.gov/28563791/

143. Lobo NF, St. Laurent B, Sikaala CH, Hamainza B, Chanda J, Chinula D, et al. Unexpected diversity of Anopheles species in Eastern Zambia: Implications for evaluating vector behavior and interventions using molecular tools. Sci Rep [Internet]. 2015;5:1–10. Available from: http://dx.doi.org/10.1038/srep17952

144. Laurent BS, Cooke M, Krishnankutty SM, Asih P, Mueller JD, Kahindi S, et al. Molecular characterization reveals diverse and unknown malaria vectors in the western Kenyan highlands. Am J Trop Med Hyg [Internet]. 2016;94:327–35. Available from: http://dx.doi.org/10.4269/ajtmh.15-0562

145. Killeen GF, Marshall JM, Kiware SS. Measuring, manipulating exploiting behaviours adult mosquitoes optimize malaria vector control impact BMJ. BMJ Global Health. 2017;2.

146. Yap NJ, Hossain H, Nada-Raja T, Ngui R, Muslim A, Hoh B-P, et al. Natural Human Infections with Plasmodium cynomolgi, P. inui, and 4 other Simian Malaria Parasites, Malaysia. Emerg Infect Dis [Internet]. 2021;27:2187–91. Available from: http://dx.doi.org/10.3201/eid2708.204502

147. Putaporntip C, Kuamsab N, Seethamchai S, Pattanawong U, Rojrung R, Yanmanee S, et al. Cryptic Plasmodium inui and Plasmodium fieldi Infections Among Symptomatic Malaria Patients in Thailand. Clin Infect Dis [Internet]. 2022 [cited 2025 May 28];75:805–12. Available from: http://dx.doi.org/10.1093/cid/ciab1060

148. Raja TN, Hu TH, Kadir KA, Mohamad DSA, Rosli N, Wong LL, et al. Naturally Acquired Human Plasmodium cynomolgi and P. knowlesi Infections, Malaysian Borneo. Emerg Infect Dis [Internet]. 2020;26:1801–9. Available from: http://dx.doi.org/10.3201/eid2608.200343

149. Mewara A, Sreenivasan P, Khurana S. Primate malaria of human importance. Trop Parasitol [Internet]. 2023;13:73–83. Available from: http://dx.doi.org/10.4103/tp.tp_79_22

150. Mayagaya VS, Nkwengulila G, Lyimo IN, Kihonda J, Mtambala H, Ngonyani H, et al. The impact of livestock on the abundance, resting behaviour and sporozoite rate of malaria vectors in southern Tanzania. Malar J [Internet]. 2015;14:17. Available from: http://dx.doi.org/10.1186/s12936-014-0536-8

151. Lyons CL, Coetzee M, Chown SL. Stable and fluctuating temperature effects on the development rate and survival of two malaria vectors, Anopheles arabiensis and Anopheles funestus. Parasit Vectors [Internet]. 2013;6:104. Available from: http://dx.doi.org /10.1186/1756-3305-6-104

152. Klinkenberg E, McCall P, Wilson MD, Amerasinghe FP, Donnelly MJ. Impact of urban agriculture on malaria vectors in Accra, Ghana. Malar J [Internet]. 2008;7:151. Available from: http://dx.doi.org/10.1186/1475-2875-7-151

153. Nguyen AHL, Pattaradilokrat S, Kaewlamun W, Kaneko O, Asada M, Kaewthamasorn M. Myzomyia and Pyretophorus series of Anopheles mosquitoes acting as probable vectors of the goat malaria parasite Plasmodium caprae in Thailand. Sci Rep [Internet]. 2023;13:145. Available from: http://dx.doi.org/10.1038/s41598-022-26833-4

154. Tu HLC, Nugraheni YR, Tiawsirisup S, Saiwichai T, Thiptara A, Kaewthamasorn M. Development of a novel multiplex PCR assay for the detection and differentiation of Plasmodium caprae from Theileria luwenshuni and Babesia spp. in goats. Acta Trop [Internet]. 2021;220:105957. Available from: http://dx.doi.org/10.1016/j.actatropica.2021.105957

155. Kaewthamasorn M, Takeda M, Saiwichai T, Gitaka JN, Tiawsirisup S, Imasato Y, et al. Genetic homogeneity of goat malaria parasites in Asia and Africa suggests their expansion with domestic goat host. Sci Rep [Internet]. 2018;8:5827. Available from:

http://dx.doi.org/10.1038/s41598-018-24048-0

156. Hamza AM, El Rayah EA. A qualitative evidence of the breeding sites of anopheles arabiensis Patton (Diptera: Culicidae) in and around Kassala town, eastern Sudan. Int J Insect Sci [Internet]. 2016;8:65–70. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC4982522/

157. Lindsay SW, Parson L, Thomas CJ. Mapping the ranges and relative abundance of the two principal African malaria vectors, Anopheles gambiae sensu stricto and An. arabiensis, using climate data. Proc Biol Sci [Internet]. 1998;265:847–54. Available from: http://dx.doi.org/10.1098/rspb.1998.0369

158. Mahande A, Mosha F, Mahande J, Kweka E. Feeding and resting behaviour of malaria vector, Anopheles arabiensis with reference to zooprophylaxis. Malar J [Internet]. 2007;6:100. Available from: http://dx.doi.org/10.1186/1475-2875-6-100

159. Killeen GF, Smith TA. Exploring the contributions of bed nets, cattle, insecticides and excitorepellency to malaria control: a deterministic model of mosquito host-seeking behaviour and mortality. Trans R Soc Trop Med Hyg [Internet]. 2007;101:867–80. Available from: http://dx.doi.org/10.1016/j.trstmh.2007.04.022

160. Pathak AK, Shiau JC, Freitas RCS, Kyle DE. Blood meals from 'dead-end' vertebrate hosts enhance transmission potential of malaria-infected mosquitoes. One Health [Internet]. 2023;17:100582. Available from: https://pubmed.ncbi.nlm.nih.gov/38024285/

161. Lynch CM, Churcher TS. Endectocides as a complementary intervention in the malaria control program: a systematic review. Parasites & Vectors. 2021;14.

162. Matowo NS, Martin J, Kulkarni MA, Mosha JF, Lukole E, Isaya G, et al. An increasing role of pyrethroid-resistant Anopheles funestus in malaria transmission in the Lake Zone, Tanzania. Sci Rep [Internet]. 2021;11:13457. Available from: http://dx.doi.org/10.1038/s41598-021-92741-8

163. Debrah I, Afrane YA, Amoah L, Ochwedo KO, Mukabana WR, Zhong D, et al. Larval

ecology and bionomics of *Anopheles funestus* in highland and lowland sites in western Kenya [Internet]. bioRxiv. bioRxiv; 2021. Available from: http://dx.doi.org/10.1101 /2021.08.04.455104

164. Nambunga IH, Ngowo HS, Mapua SA, Hape EE, Msugupakulya BJ, Msaky DS, et al. Aquatic habitats of the malaria vector Anopheles funestus in rural south-eastern Tanzania. Malar J [Internet]. 2020;19:219. Available from: http://dx.doi.org/10.1186/s12936-020-03295-5

165. Kahamba NF, Finda M, Ngowo HS, Msugupakulya BJ, Baldini F, Koekemoer LL, et al. Using ecological observations to improve malaria control in areas where Anopheles funestus is the dominant vector. Malar J [Internet]. 2022;21:1–15. Available from: https://doi.org/10.1186/s12936-022-04198-3

166. Bouafou L, Makanga BK, Rahola N, Boddé M, Ngangué MF, Daron J, et al. Host preference patterns in domestic and wild settings: Insights into Anopheles feeding behavior. Evol Appl [Internet]. 2024;17:e13693. Available from: http://dx.doi.org/10.1111/eva.13693

167. Hay SI, Sinka ME, Okara RM, Kabaria CW, Mbithi PM, Tago CC, et al. Developing global maps of the dominant anopheles vectors of human malaria. PLoS Med [Internet]. 2010;7:e1000209. Available from: https://journals.plos.org/plosmedicine/article? id=10.1371/ journal.pmed.1000209

168. Ayala D, Costantini C, Ose K, Kamdem GC, Antonio-Nkondjio C, Agbor J-P, et al. Habitat suitability and ecological niche profile of major malaria vectors in Cameroon. Malar J [Internet]. 2009;8:307. Available from: http://dx.doi.org/10.1186/1475-2875-8-307

169. Dida GO, Anyona DN, Abuom PO, Akoko D, Adoka SO, Matano A-S, et al. Spatial distribution and habitat characterization of mosquito species during the dry season along the Mara River and its tributaries, in Kenya and Tanzania. Infect Dis Poverty [Internet]. 2018;7:2. Available from: http://dx.doi.org/10.1186/s40249-017-0385-0

170. Kavishe DR, Walsh KA, Msoffe RV, Duggan LM, Tarimo LJ, Butler F, et al. Comparative attrraction of *Anopheles quadriannulatus* and *Anopheles arabiensis* to humans estimated by comparing their relative abundance in samples of mosquito larvae and adults collected across

an ecologically heterogeneous landscape in southern Tanzania. Med Vet Entomol [Internet]. 2025; Available from: https://scholar.google.com/citations?view_op=view_citation&hl=en&citation_for_view=aG1a5W0AAAAJ:LkGwnXOMwfcC

171. Csardi G, Nepusz T. The igraph software package for complex network research. InterJournal, Complex Systems [Internet]. 2006;1695. Available from: https://igraph.org

172. Hijmans RJ, Williams E, Vennes C. geosphere: Spherical Trigonometry [Internet]. 2019. Available from: https://CRAN.R-project.org/package=geosphere

173. Bayoh MN, Lindsay SW. Temperature-related duration of aquatic stages of the Afrotropical malaria vector mosquito Anopheles gambiae in the laboratory. Med Vet Entomol [Internet]. 2004;18:174–9. Available from: http://dx.doi.org/10.1111/j.0269-283x.2004.00495.x

174. Bayoh MN, Lindsay SW. Effect of temperature on the development of the aquatic stages of *Anopheles gambiae* sensu stricto (Diptera: Culicidae). Bull Entomol Res [Internet]. 2003;93:375–81. Available from: http://dx.doi.org/10.1079/ber2003259

175. Charlwood JD. Some like it hot: a differential response to changing temperatures by the malaria vectors Anopheles funestus and An. gambiae s.l. PeerJ [Internet]. 2017;5:e3099. Available from: http://dx.doi.org/10.7717/peerj.3099

176. Agyekum TP, Arko-Mensah J, Botwe PK, Hogarh JN, Issah I, Dwomoh D, et al. Effects of Elevated Temperatures on the Growth and Development of Adult Anopheles gambiae (s.l.) (Diptera: Culicidae) Mosquitoes. J Med Entomol [Internet]. 2022;59:1413–20. Available from: https://academic.oup.com/jme/article-abstract/59/4/1413/6572592

177. Agyekum TP, Botwe PK, Arko-Mensah J, Issah I, Acquah AA, Hogarh JN, et al. A systematic review of the effects of temperature on Anopheles mosquito development and survival: Implications for malaria control in a future warmer climate. Int J Environ Res Public Health [Internet]. 2021;18:7255. Available from: https://www.mdpi.com/1660-4601/18/14/7255

178. Christiansen-Jucht C, Parham PE, Saddler A, Koella JC, Basáñez M-G. Temperature during larval development and adult maintenance influences the survival of Anopheles gambiae s.s. Parasit Vectors [Internet]. 2014;7:489. Available from: http://dx.doi.org/10.1186/s13071-014-0489-3

179. Yamana TK, Eltahir EAB. Incorporating the effects of humidity in a mechanistic model of Anopheles gambiae mosquito population dynamics in the Sahel region of Africa. Parasit Vectors [Internet]. 2013;6:235. Available from: http://dx.doi.org/10.1186/1756-3305-6-235

180. Mordecai EA, Paaijmans KP, Johnson LR, Balzer C, Ben-Horin T, de Moor E, et al. Optimal temperature for malaria transmission is dramatically lower than previously predicted. Ecol Lett [Internet]. 2013;16:22–30. Available from: http://dx.doi.org/10.1111/ele.12015

181. Beck-Johnson LM, Nelson WA, Paaijmans KP, Read AF, Thomas MB, Bjørnstad ON. The effect of temperature on Anopheles mosquito population dynamics and the potential for malaria transmission. PLoS One [Internet]. 2013;8:e79276. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0079276

182. Hinne IA, Attah SK, Mensah BA, Forson AO, Afrane YA. Larval habitat diversity and Anopheles mosquito species distribution in different ecological zones in Ghana. Parasit Vectors [Internet]. 2021;14:193. Available from: http://dx.doi.org/10.1186/s13071-021-04701-w

183. Mangani C, Frake AN, Chipula G, Mkwaila W, Kakota T, Mambo I, et al. Proximity of residence to irrigation determines malaria risk and Anopheles abundance at an irrigated agroecosystem in Malawi. Am J Trop Med Hyg [Internet]. 2021;106:283–92. Available from: http://dx.doi.org/10.1061/(ASCE)HE.19435584.0001436

184. Keiser J, De Castro MC, Maltese MF, Bos R, Tanner M, Singer BH, et al. Effect of irrigation and large dams on the burden of malaria on a global and regional scale. Am J Trop Med Hyg [Internet]. 2005;72:392–406. Available from: https://pubmed.ncbi.nlm.nih.gov/15827275/

185. Getachew D, Balkew M, Tekie H. Anopheles larval species composition and

characterization of breeding habitats in two localities in the Ghibe River Basin, southwestern Ethiopia. Malar J [Internet]. 2020;19:65. Available from: http://dx.doi.org/10.1186/s12936-020-3145-8

186. Nyasvisvo DS, Nhiwatiwa T, Sithole R, Sande S. Characterization of Anopheles mosquito breeding habitats for malaria vector control in Mazowe and Shamva districts, Zimbabwe. J Vector Borne Dis [Internet]. 2025 [cited 2025 July 14];62:154–64. Available from: http://dx.doi.org/10.4103/JVBD.JVBD_85_24

187. Adoha CJ, Sovi A, Padonou GG, Yovogan B, Akinro B, Accrombessi M, et al. Diversity and ecological niche model of malaria vector and non-vector mosquito species in Covè, Ouinhi, and Zangnanado, Southern Benin. Sci Rep [Internet]. 2024;14:16944. Available from: https://www.nature.com/articles/s41598-024-67919-5

188. Mathenge EM, Misiani GO, Oulo DO, Irungu LW, Ndegwa PN, Smith TA, et al. Comparative performance of the Mbita trap, CDC light trap and the human landing catch in the sampling of Anopheles arabiensis, An. funestus and culicine species in a rice irrigation in western Kenya. Malar J [Internet]. 2005;4:1–6. Available from: http://dx.doi.org/10.1186/1475-2875-4-7

189. Kulkarni MA, Desrochers RE, Kajeguka DC, Kaaya RD, Tomayer A, Kweka EJ, et al. 10 years of environmental change on the slopes of Mount Kilimanjaro and its associated shift in malaria vector distributions. Front Public Health [Internet]. 2016;4:281. Available from: https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2016.00281/full

190. Akpan GE, Adepoju KA, Oladosu OR, Adelabu SA. Dominant malaria vector species in Nigeria: Modelling potential distribution of Anopheles gambiae sensu lato and its siblings with MaxEnt. PLoS One [Internet]. 2018;13:e0204233. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0204233

191. Wood SN. Generalized additive models: An introduction with R [Internet]. Chapman and Hall/CRC; 2017. Available from: http://dx.doi.org/10.1201/9781315370279

192. Lindsay SW, Martens WJ. Malaria in the African highlands: past, present and future. Bull

World Health Organ [Internet]. 1998;76:33–45. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC2305628/

193. Paaijmans KP, Wandago MO, Githeko AK, Takken W. Unexpected high losses of Anopheles gambiae larvae due to rainfall. PLoS One [Internet]. 2007;2:e1146. Available from: http://dx.doi.org/10.1371/journal.pone.0001146

194. Takken W, Verhulst NO. Host preferences of blood-feeding mosquitoes. Annu Rev Entomol [Internet]. 2013;58:433–53. Available from: https://www.researchgate.net/publication/231610434_Host_Preferences_of_Blood-Feeding_Mosquitoes

195. Mutero CM, Blank H, Konradsen F, van der Hoek W. Water management for controlling the breeding of Anopheles mosquitoes in rice irrigation schemes in Kenya. Acta Trop [Internet]. 2000;76:253–63. Available from: https://pubmed.ncbi.nlm.nih.gov/10974166/

196. Tonnang HEZ, Kangalawe RYM, Yanda PZ. Predicting and mapping malaria under climate change scenarios: the potential redistribution of malaria vectors in Africa. Malar J [Internet]. 2010;9:111. Available from: http://dx.doi.org/10.1186/1475-2875-9-111

197. Kibret S, McCartney M, Lautze J, Nhamo L, Yan G. The impact of large and small dams on malaria transmission in four basins in Africa. Sci Rep [Internet]. 2021;11:13355. Available from: https://www.nature.com/articles/s41598-021-92924-3

198. Kibret S, Lautze J, McCartney M, Nhamo L, Wilson GG. Malaria and large dams in sub-Saharan Africa: future impacts in a changing climate. Malar J [Internet]. 2016;15:448. Available from: http://dx.doi.org/10.1186/s12936-016-1498-9

199. Geissbühler Y, Kannady K, Chaki PP, Emidi B, Govella NJ, Mayagaya V, et al. Microbial larvicide application by a large-scale, community-based program reduces malaria infection prevalence in urban Dar es Salaam, Tanzania. PLoS One [Internet]. 2009;4:e5107. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0005107

200. Bødker R, Akida J, Shayo D, Kisinza W, Msangeni HA, Pedersen EM, et al. Relationship between altitude and intensity of malaria transmission in the Usambara Mountains, Tanzania.

J Med Entomol [Internet]. 2003;40:706–17. Available from: https://academic.oup.com/jme/article-abstract/40/5/706/864875

201. Kulkarni MA, Desrochers RE, Kerr JT. High resolution niche models of malaria vectors in northern Tanzania: a new capacity to predict malaria risk? PLoS One [Internet]. 2010;5:e9396. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0009396

202. Kelly-Hope LA, Hemingway J, McKenzie FE. Environmental factors associated with the malaria vectors Anopheles gambiae and Anopheles funestus in Kenya. Malar J [Internet]. 2009;8:268. Available from: http://dx.doi.org/10.1186/1475-2875-8-268

203. Mathania MM, Munisi DZ, Silayo RS. Spatial and temporal distribution of Anopheles mosquito's larvae and its determinants in two urban sites in Tanzania with different malaria transmission levels. Parasite Epidemiol Control [Internet]. 2020;11:e00179. Available from: https://www.sciencedirect.com/science/article/pii/S2405673120300489

204. Marsden CD, Lee Y, Kreppel K, Weakley A, Cornel A, Ferguson HM, et al. Diversity, differentiation, and linkage disequilibrium: prospects for association mapping in the malaria vector Anopheles arabiensis. G3 (Bethesda) [Internet]. 2014;4:121–31. Available from: http://dx.doi.org/10.1534/g3.113.008326

205. Hemming-Schroeder E, Zhong D, Machani M, Nguyen H, Thong S, Kahindi S, et al. Ecological drivers of genetic connectivity for African malaria vectors Anopheles gambiae and An. arabiensis. Sci Rep [Internet]. 2020;10:19946. Available from: https://www.nature.com/articles/s41598-020-76248-2

206. Ayala D, Ullastres A, González J. Adaptation through chromosomal inversions in Anopheles. Front Genet [Internet]. 2014;5:129. Available from: http://dx.doi.org/10.3389/fgene.2014.00129

207. Crawford JE, Riehle MM, Guelbeogo WM, Gneme A, Sagnon N, Vernick KD, et al. Reticulate speciation and barriers to introgression in the Anopheles gambiae species complex. Genome Biol Evol [Internet]. 2015;7:3116–31. Available from: https://pubmed.

ncbi.nlm.nih.gov/26615027/

208. Donnelly MJ, Cuamba N, Charlwood JD, Collins FH, Townson H. Population structure in the malaria vector, Anopheles arabiensis patton, in East Africa. Heredity (Edinb) [Internet]. 1999;83 (Pt 4):408–17. Available from: http://dx.doi.org/10.1038/sj.hdy.6885930

209. Kent RJ, Mharakurwa S, Norris DE. Spatial and temporal genetic structure of Anopheles arabiensis in Southern Zambia over consecutive wet and drought years. Am J Trop Med Hyg [Internet]. 2007;77:316–23. Available from: http://dx.doi.org/10.4269/ajtmh.2007.77.316

210. Donnelly MJ, Townson H. Evidence for extensive genetic differentiation among populations of the malaria vector Anopheles arabiensis in Eastern Africa. Insect Mol Biol [Internet]. 2000;9:357–67. Available from: http://dx.doi.org/10.1046/j.1365-2583. 2000.00197.x

211. Simard F, Fontenille D, Lehmann T, Girod R, Brutus L, Gopaul R, et al. High amounts of genetic differentiation between populations of the malaria vector Anopheles arabiensis from West Africa and eastern outer islands. The American journal of tropical medicine and hygiene [Internet]. 1999;60:1000–9. Available from: https://www.researchgate.net/profile/Romain-Girod/publication/12896661_High_amounts_of_genetic_differentiation_between_populations_of_the_malaria_vector_Anopheles_arabiensis_from_West_Africa_and_Eastern_outer_islands/links/004635175aaed950fa000000/High-amounts-of-genetic-differentiation-between-populations-of-the-malaria-vector-Anopheles-arabiensis-from-West-Africa-and-Eastern-outer-islands.pdf

212. Barnes KG, Irving H, Chiumia M, Mzilahowa T, Coleman M, Hemingway J, et al. Restriction to gene flow is associated with changes in the molecular basis of pyrethroid resistance in the malaria vector Anopheles funestus. Proc Natl Acad Sci U S A [Internet]. 2017;114:286–91. Available from: http://dx.doi.org/10.1073/pnas.1615458114

213. Seck F, Diop MF, Mané K, Diallo A, Dieng I, Namountougou M, et al. Reduced genetic diversity of key fertility and vector competency related genes in Anopheles gambiae s.L. Across sub-Saharan Africa. Genes (Basel) [Internet]. 2025;16. Available from: http://dx.doi.org/10.3390/genes16050543

214. Mnzava AEP, Deco MAD. Chromosomal inversion polymorphism in anopheles gambiae and anopheles arabiensis in Tanzania. Int J Trop Insect Sci [Internet]. 1990;11:861–3. Available from: https://www.cambridge.org/core/journals/international-journal-of-tropical-insect-science/article/chromosomal-inversion-polymorphism-in-anopheles-gambiae-and-anopheles-arabiensis-in-tanzania/0D9339CA01CBA98FF43E6C28E497E6C0

215. Main BJ, Lee Y, Ferguson HM, Kreppel KS, Kihonda A, Govella NJ, et al. The genetic basis of host preference and resting behavior in the major African malaria vector, Anopheles arabiensis. PLoS Genet [Internet]. 2016;12:e1006303. Available from: https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1006303

216. Costantini C, Li SG, Della Torre A, Sagnon N, Coluzzi M, Taylor CE. Density, survival and dispersal of Anopheles gambiae complex mosquitoes in a west African Sudan savanna village. Med Vet Entomol [Internet]. 1996 [cited 2025 June 25];10:203–19. Available from: http://dx.doi.org/10.1111/j.1365-2915.1996.tb00733.x

217. Huestis DL, Dao A, Diallo M, Sanogo ZL, Samake D, Yaro AS, et al. Windborne long-distance migration of malaria mosquitoes in the Sahel. Nature [Internet]. 2019;574:404–8. Available from: https://pubmed.ncbi.nlm.nih.gov/31578527/

218. Taylor C, Touré YT, Carnahan J, Norris DE, Dolo G, Traoré SF, et al. Gene flow among populations of the malaria vector, Anopheles gambiae, in Mali, West Africa. Genetics [Internet]. 2001;157:743–50. Available from: https://academic.oup.com/genetics/article-abstract/157/2/743/6048215

219. Lehmann T, Licht M, Elissa N, Maega BTA, Chimumbwa JM, Watsenga FT, et al. Population structure of Anopheles gambiae in Africa. J Hered [Internet]. 2003;94:133–47. Available from: http://dx.doi.org/10.1093/jhered/esg024

220. Pinto J, Egyir-Yawson A, Vicente J, Gomes B, Santolamazza F, Moreno M, et al. Geographic population structure of the African malaria vector Anopheles gambiae suggests a role for the forest-savannah biome transition as a barrier to gene flow. Evol Appl [Internet]. 2013 [cited 2025 May 23];6:910–24. Available from: http://dx.doi.org/10.1111/eva.12075

221. Sougoufara S, Doucouré S, Backé Sembéne PM, Harry M, Sokhna C. Challenges for malaria vector control in sub-Saharan Africa: Resistance and behavioral adaptations in Anopheles populations. J Vector Borne Dis [Internet]. 2017;54:4–15. Available from: http://dx.doi.org/10.4103/0972-9062.203156

222. Kamdem C, Tene Fossog B, Simard F, Etouna J, Ndo C, Kengne P, et al. Anthropogenic habitat disturbance and ecological divergence between incipient species of the malaria mosquito Anopheles gambiae. PLoS One [Internet]. 2012 [cited 2025 Aug 17];7:e39453. Available from: http://dx.doi.org/10.1371/journal.pone.0039453

223. McCann RS, Courneya J-P, Donnelly M, Laufer MK, Mzilahowa T, Stewart K, et al. Variation in spatial population structure in the Anopheles gambiae species complex [Internet]. bioRxiv. 2024. Available from: http://dx.doi.org/10.1101/2024.05.26.595955

Appendix A. Supplementary Table

Supplementary Table A1. Presence-only, georeferenced occurrence records of *An. arabiensis*, *An. funestus*, and *An. gambiae* by data source and year (2011–2024) used in SDM for Tanzania; totals by species and overall are shown.

Data Source	Species	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	Source/Species Total
Deogratius Kavishe (IHI) [170],	*An. arabiensis*											27	41			68
	An. funestus												2			2
	An. gambiae												1			1
Dhibiti Malaria Project (IHI)	*An. arabiensis*													23	82	105
	An. gambiae													20	41	61
Mr. Edmond Bernad (NIMR)	*An. arabiensis*												2			2
	An. funestus												5			5
	An. gambiae													5		5
Dr. Fedros Okumu (IHI)	*An. funestus*								1	23						24
Data from this study	*An. arabiensis*									21	128	72	65			286
	An. funestus									10	64	59	41			174
	An. gambiae										7	57	25	17		106
Matowo et al 2021 [162]	*An. arabiensis*								1							1
	An. funestus											2				2
	An. gambiae											1				1
Mwalimu et al 2024 [65] (NMCP_MVES)	*An. arabiensis*							4	12	3	7	12	5	1		44
	An. funestus							1	8	4	6	3	6	2		30
	An. gambiae				2	6		2	6		2	6	4			20
Nambunga et al. 2020 [164]	*An. funestus*												2			2
	An. arabiensis	27	8	20	1	6	4	5	1	1						73

WHO Threat Map Metadata	*An. gambiae*	17	5	4				1	1							28
Total by Species	*An. arabiensis*	27	8	20	1	6	4	9	14	25	135	111	113	24	82	579
	An. funestus	0	0	0	0	0	0	1	9	37	70	64	56	2	0	239
	An. gambiae	17	5	4	2	6	0	3	7	0	9	64	30	42	41	222
Overall Grand Total		44	13	24	3	12	4	13	30	62	214	239	199	68	123	1040