Genomic epidemiology of *Plasmodium falciparum* and its *Anopheline* vectors in the context of malaria control on the Bijagós Archipelago of Guinea-Bissau

Sophie Judith Moss

Thesis submitted in accordance with the requirements for the degree of Doctor of Philosophy of the University of London

November 2024

Department of Clinical Research
Faculty of Infectious and Tropical Diseases
London School of Hygiene & Tropical Medicine,
University of London

Funded by the Medical Research Council London Intercollegiate Doctoral Training Partnership

I, Sophie Judith Moss, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed: 

Date: 28th November 2024

Abstract

This thesis uses genomic techniques to investigate the genetic diversity and antimalarial resistance status of the *Plasmodium falciparum* parasite, and the genetic diversity and insecticide resistance status of key *Anopheles gambiae sensu lato* vectors, on the Bijagós Archipelago of Guinea-Bissau. This investigation was conducted in the context of malaria control interventions on this remote Archipelago, where the genomic epidemiology of malaria has not previously been studied. Selective whole-genome amplification was used to investigate population dynamics and drug resistance mutations in whole genome sequence (WGS) data from Bijagós *P. falciparum* isolates. This analysis identified fixation of the *pfdhfr* mutations N51I and S108N, associated with resistance to sulphadoxine-pyrimethamine, and the continued presence of *pfcrt* K76T, associated with chloroquine resistance. *P. falciparum* from the Archipelago clustered with samples from mainland West Africa without forming a separate phylogenetic group. This thesis continues by analysing the insecticide resistance status of the major vector *Anopheles gambiae.* High throughput multiplex amplicon sequencing is used to identify molecular markers of insecticide resistance, revealing four single-nucleotide polymorphisms (SNPs) previously associated with pyrethroid resistance, and an additional eight previously unreported SNPs. This genomic analysis is built upon with phenotypic bioassays in the Bijagós to investigate susceptibility to deltamethrin, and deltamethrin resistance is identified on Bubaque island. WGS data from deltamethrin resistant and susceptible *An. gambiae sensu stricto* is then analysed to identify insecticide resistance associated SNPs and signatures of selection. Genome wide selection scans are conducted, identifying two selective sweeps in the population, which overlap with metabolic genes previously associated with insecticide resistance, including *cyp9k1* and the *cyp6aa/cyp6p* gene cluster. The population dynamics and insecticide resistance status of the malaria vector *Anopheles melas,* which is abundant on the islands, are then analysed. This analysis identifies two separate phylogenetic clusters of *An. melas* and investigates genetic differentiation, signatures of selection, and structural variation within the genome. This thesis generates and analyses the first WGS data for *P. falciparum, An. gambiae s.l.,* and *An. melas* from the Bijagós Archipelago, and includes the first phenotypic testing for deltamethrin resistance on the islands; contributing data and analyses of significance to malaria control policy in the region.

Contents

Acknowledgements

Thank you to my supervisors Anna Last, Susana Campino, Sanjeev Krishna and Taane Clark, and my advisor Alfred Amambua-Ngwa, for your time, energy, and for the opportunity to conduct this research. Thank you to the Medical Research Council for funding this work, and for the training opportunities that I have been given during my PhD scholarship. Thank you to the whole MATAMAL team for the opportunity to conduct fieldwork in the beautiful Bijagós Archipelago, and for the opportunity to work together with the team at MRC Gambia. In particular, thank you to Eunice, Liz, Harry, Hristina, Braima, Julio, Sainey, Mamadou, Ansumane and Rob, for your help and guidance, and to the people of the Bijagós who participated in the MATAMAL study. Thank you to the Campino Clark group at LSHTM, an amazing group of researchers whose friendship and collaboration have made this process a truly enjoyable chapter. A special thank you to Ashley, Emma, Gabbie, Leen, and Holly, for being hugely supportive friends and fantastic scientists; to my Girton 'Gaks' for your endless love and laughter supply over the past ten years; and to Kirsty, who has supported me on the entire journey. A heartfelt thank you to my parents Cathy and Michael, who gave me a love of learning; to my brother Andrew, who taught me that I could do and be anything; and to my grandparents Jim and Peter, whom I love dearly. This journey would not have started without the love and support of my family, for which I will forever be grateful. Finally, thank you to my husband Harry, for your unwavering love and encouragement. Thank you for everything that you do, and especially for proposing halfway through.

I dedicate this PhD to the memory of my two grandmothers, Susie Burrows and Judith Moss, who would burst with pride if they could read this today.

Thesis Publications and Manuscripts

1. **Moss, S.**, Mańko, E., Vasileva, H., Teixeira Da Silva, E., Goncalves, A., Osborne, A., Phelan, J., Rodrigues, A., Djata, P., D'Alessandro, U., Mabey, D., Krishna, S., Last, A., Clark, T.G., Campino, S. Population dynamics and drug resistance mutations in *Plasmodium falciparum* on the Bijagós Archipelago, Guinea-Bissau. *Sci Rep* **13**, 6311 (2023). https://doi.org/10.1038/s41598-023-33176-1
2. **Moss, S.**, Pretorius, E., Ceesay, S., Hutchins, H., Teixeira Da Silva, E., Ousmane Ndiath, M., Jones, R.T., Vasilvea, H., Phelan, J., Acford-Palmer, H., Collins, E., Rodrigues, A., Krishna, S., Clark, T.G., Last, A., Campino, S. Genomic surveillance of *Anopheles* mosquitoes on the Bijagós Archipelago using custom targeted amplicon sequencing identifies mutations associated with insecticide resistance. *Parasites Vectors* **17**, 10 (2024). https://doi.org/10.1186/s13071-023-06085-5
3. **Moss, S.**, Jones, R.T., Pretorius, E., Teixeira Da Silva, E., Higgins, M., Kristan, M., Acford-Palmer, H., Collins, E., Rodrigues, A., Krishna, S., Clark, T.G., Last, A., Campino, S. Phenotypic evidence of deltamethrin resistance and identification of selective sweeps in *Anopheles* mosquitoes on the Bijagós Archipelago, Guinea-Bissau. *Sci Rep* **14**, 22840 (2024). https://doi.org/10.1038/s41598-024-73996-3
4. **Moss, S.**, Pretorius, E., Ceesay, S., Teixeira Da Silva, E., Hutchins, H., Ousmane Ndiath, M., Palmer, H., Collins, E., Higgins, M., Phelan, J., Jones, R.T., Rodrigues, A., Krishna, S., Clark, T.G., Last, A., Campino, S. Whole genome sequence analysis of population structure and insecticide resistance markers in *Anopheles melas* from the Bijagós Archipelago, Guinea-Bissau. *Parasites Vectors* **17**, 396 (2024). https://doi.org/10.1186/s13071-024-06476-2

Additional Publications

During my PhD I have contributed to other manuscripts which are not included in this thesis. These are outlined below:

1. **Moss, S.,** Mańko, E., Krishna, S., Campino, S., Clark, T.G., Last, A. How has mass drug administration with dihydroartemisinin-piperaquine impacted molecular markers of drug resistance? A systematic review. Malar J. 2022 Jun 11;21(1):186. doi: 10.1186/s12936-022-04181-y. PMID: 35690758; PMCID: PMC9188255.
2. Collins, E., Quintana, M. J., Morales, R., **Moss, S.,** Acford-Palmer, H., Higgins, M., Phelan, J., Clark, T.G., Brown, G., Campino, S. Profiling insecticide resistance phenotypes and genotypes in *Aedes aegypti* populations across four regions in Puerto Rico, 22 February 2024, PREPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs-3943783/v1]
3. Hutchins, H., Pretorius, E., Bradley, J., Teixeira Da Silva, E., Vasileva, H., Ousmane Ndiath, M., Jones, R.T., Massire Soumare, H., Nyang, H., Prom, A., Sambou, S., Ceesay, F., Ceesay, S., **Moss, S.,** Mabey, D., Djata, P., Nante J.E., Martins, C., Logan, J.G., Slater, H., Tetteh, K., Drakeley C., D'Alessandro, U., Rodrigues A., Last, A. A cluster randomised placebo-controlled trial of adjunctive ivermectin mass drug administration for malaria control on the Bijagós Archipelago of Guinea-Bissau: The MATAMAL trial 25 March 2024, Published, The Lancet Infectious Diseases: 10.1016/S1473-3099(24)00580-2.

Abbreviations

ACTs - Artemisinin-Based Combination Therapies

AChE- Acetylcholinesterase

AL - Artemether-Lumefantrine

AS+AQ - Artesunate-Amodiaquine

AS+SP - Artesunate and Sulphadoxine-Pyrimethamine

ASMQ - Artesunate-Mefloquine

ASPY - Artesunate-Pyronaridine

CQ - Chloroquine

CNVs - Copy Number Variants

DBS - Dried Blood Spots

DC - Discriminating Concentrations

DESs - Dried Erythrocyte Spots

DDT - Dichlorodiphenyltrichloroethane

DHAP - Dihydroartemisinin-Piperaquine

DHA-PPQ - Dihydroartemisinin-Piperaquine

DRC – Democratic Republic of Congo

FWS - Fixation Within Species

GSTs - Glutathione-S-Transferases

IBD - Identity by Descent

iHS - Integrated Haplotype Score

IRS - Indoor Residual Spraying

ITNs - Insecticide Treated Nets

kdr - Knockdown Resistance

MDA - Mass Drug Administration

NMCP - National Malaria Control Program

NGS - Next-Generation-Sequencing

NMCPs - National Malaria Control Programmes

PBO - Piperonyl Butoxide

PCA - Principal Component Analysis

PDMC - Post-Discharge Malaria Chemoprevention

PMC - Perennial Malaria Chemoprevention

qPCR - Quantitative Polymerase Chain Reaction

RCTs - Randomised Controlled Trials

RDL – Resistant to dieldrin

RDTs - Rapid Diagnostic Tests

SMC - Seasonal Malaria Chemoprevention

SNPs - Single Nucleotide Polymorphisms

SP - Sulphadoxine-Pyrimethamine

SP-AQ - Sulphadoxine-Amodiaquine

SWGA - Selective Whole Genome Amplification

TESs - Therapeutic Efficacy Studies

VGSC - Voltage-Gated Sodium Channel

WGS - Whole-Genome Sequencing

WHO - World Health Organization

XP-EHH - Cross Population Extended Haplotype Homozygosity

CHAPTER ONE

Introduction

The global burden of malaria

Despite understanding the causes of malaria for over a century, this disease still kills over 600,000 people every year[1]. In 2022, there were approximately 249 million malaria cases across 85 malaria endemic countries and areas (**Figure 1**)[1]. An estimated 93.6% of cases and 95.4% of deaths occurred in the World Health Organization (WHO) African Region, with 78.1% of all deaths in the region occurring in children aged under 5[1]. Malaria is caused by protozoan parasites belonging to the genus *Plasmodium.* There are six *Plasmodium* species which are known to infect humans: *P. falciparum, P. vivax, P. malariae, P. ovale curtisi (Poc), P. ovale wallikeri (Pow),* and *P. knowlesi.* The most lethal of these parasites is *P. falciparum,* which predominates in Africa and is responsible for the majority of malaria deaths[1]. The second most dangerous *Plasmodium* species is *P. vivax,* which is the most prevalent *Plasmodium* species outside of Africa and caused 45.7% of malaria cases in the WHO South-East Asia Region in 2022. The WHO Global Technical Strategy 2016-2030 aims to reduce the global malaria burden by 90% by 2030[1]. Considerable progress has been made, and global estimated malaria deaths decreased from 864,000 in 2000 to 586,000 in 2015 (**Figure 2**). However, the decline in mortality rate has slowed since 2015, reaching 576,000 in 2019, and malaria deaths then increased by around 10% to 631,000 in 2020, before declining to 608,000 in 2022. The increase between 2019 and 2020 has been attributed mostly to disruption to malaria control programme activities due to the COVID-19 pandemic[2]. Compared to the goals of the WHO Global Technical Strategy, malaria mortality is off track by 53%[1]. This plateau in progress is attributable to multiple factors, including the evolution of resistance to antimalarial drugs in the *Plasmodium* parasite, and the evolution of insecticide resistance in the *Anopheles* mosquito host.

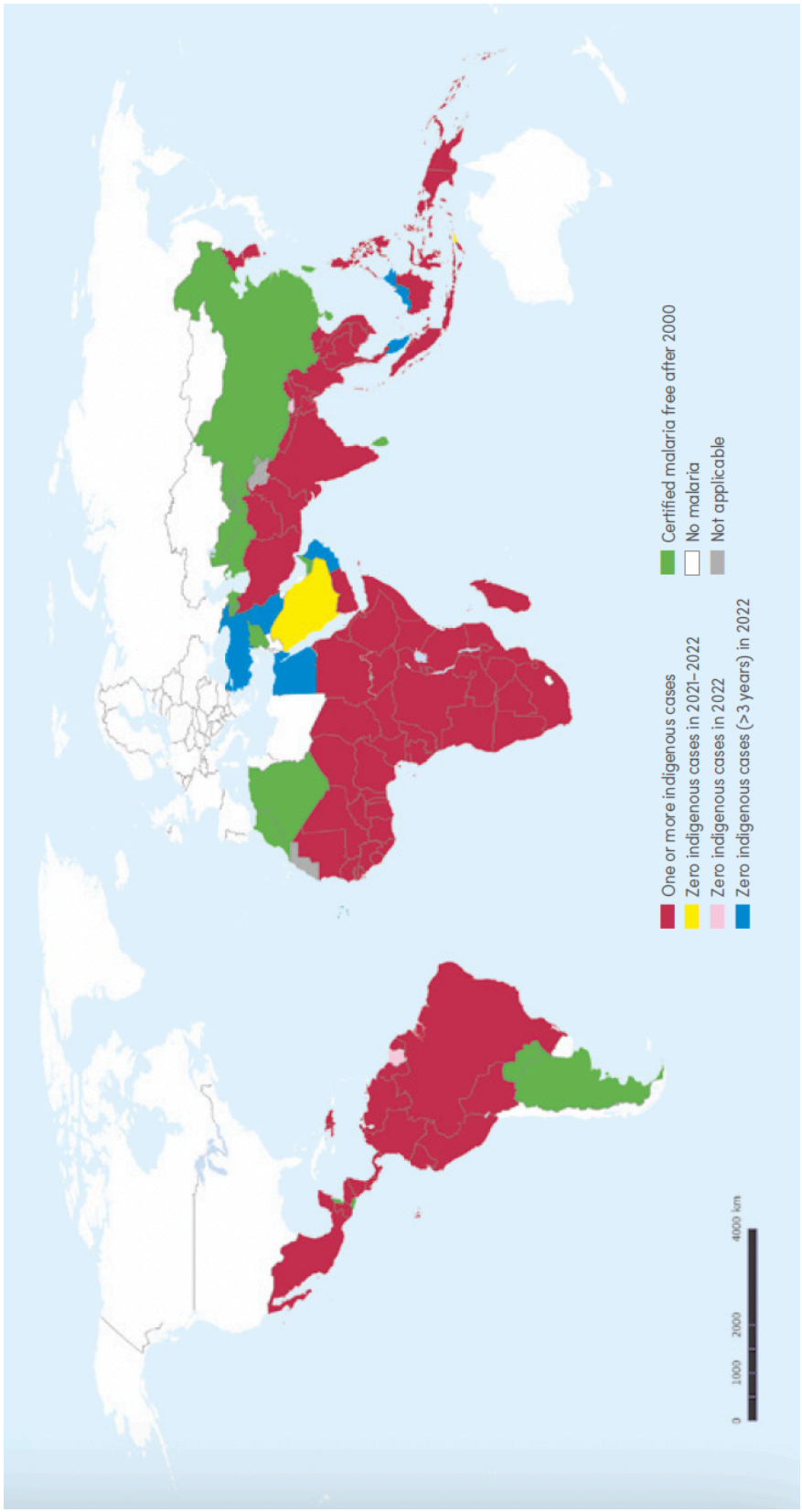


***Figure 1**: Countries with indigenous cases in 2000 and their status by 2022, Fig. 3.2. in the World Malaria Report 2023[1]. Belize was certified as malaria free in 2023. Cabo Verde was certified as malaria free in January 2024. Since 2018, Malaysia has reported zero indigenous malaria cases from human Plasmodium species.*

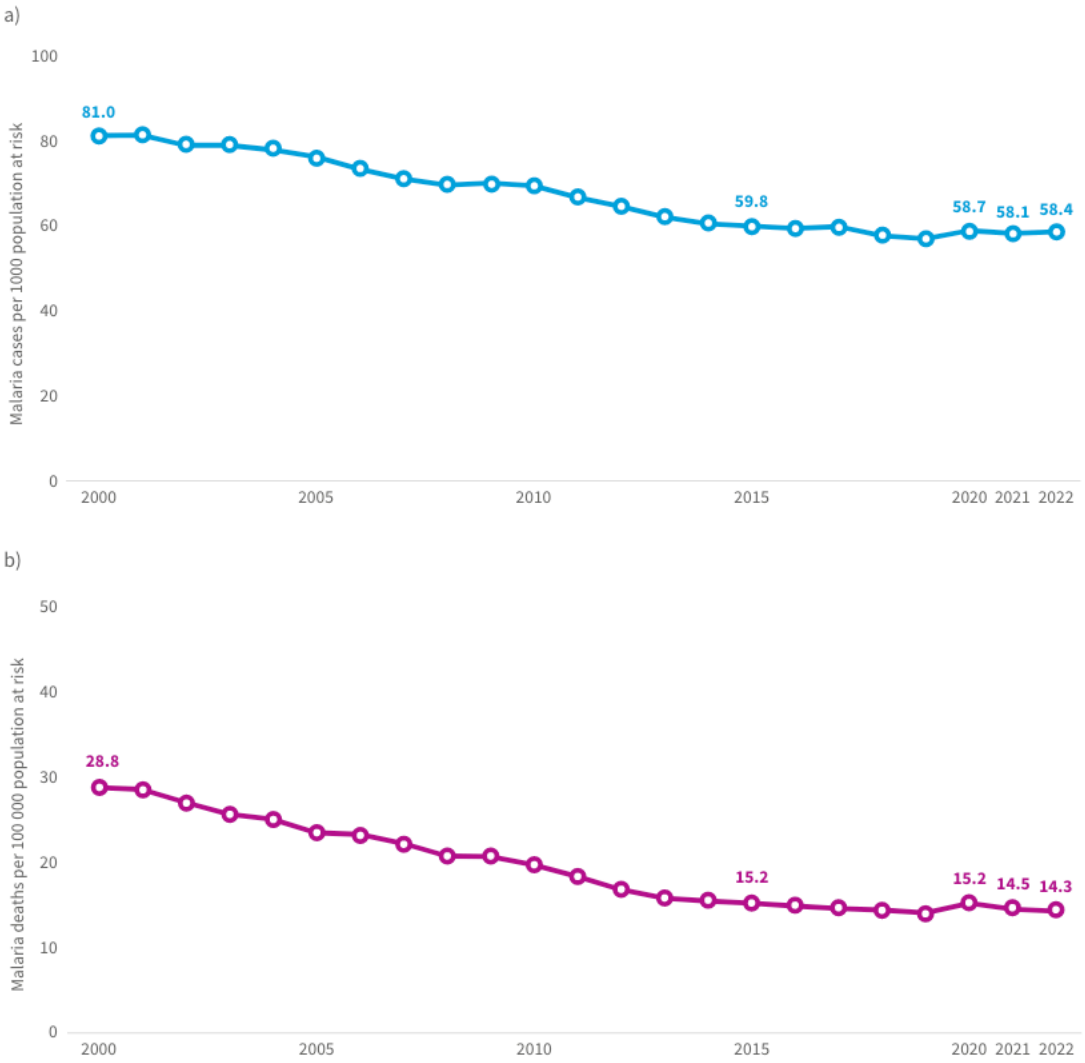


***Figure 2**: Global trends in a) malaria case incidence (cases per 1000 population at risk) and b) mortality rate (deaths per 100,000 population at risk), 2000-2022. Fig. 3.3. in the World Malaria Report 2023*[1]

The life cycle of *Plasmodium falciparum*

Plasmodium's life cycle includes phases in both human and mosquito hosts (**Figure 3**)[3]. The *Plasmodium* parasite is transmitted by female *Anopheles* mosquitoes, which inject sporozoites into a human host on biting. These sporozoites migrate to the liver and infect hepatocytes, where they undergo asexual reproduction within the cells, which then rupture to release merozoites into the bloodstream (**Figure 3**, stages 1-5). The lifecycle differs here for *P. vivax* and *P. ovale,* for which dormant liver hypnozoites are produced[4,5]. If untreated, these dormant stage hypnozoites can persist in the liver and cause malaria relapses by invading the bloodstream weeks or months after initial infection[6,7]. When merozoites have been released into the blood, they infect erythrocytes and multiply, maturing into either trophozoites or gametocytes (**Figure 3**, stages 5-7). Trophozoites

continue to mature into schizonts, which rupture, producing more merozoites which continue to infect other erythrocytes in the erythrocytic cycle. The clinical symptoms of malaria, including cyclical fevers, are caused primarily by this schizont rupture and the destruction of the erythrocytes[8]. Throughout this time, some merozoites develop into gametocytes, which travel in the bloodstream until they are ingested by an *Anopheles* mosquito during a blood meal. When a mosquito ingests gametocytes, the male and female gametocytes fuse to generate zygotes, which are the only developmental stage of the parasite that has a diploid genome[9]. This defines the mosquito as the definitive host for *Plasmodium,* as it is where sexual reproduction takes place[9]. The zygotes then differentiate to form motile, haploid, ookinetes in the mosquito gut. These ookinetes penetrate the midgut wall of the mosquito, where they develop into oocysts. These oocysts subsequently rupture and release sporozoites into the mosquito haemolymph, which migrate to the salivary glands of the mosquito, ready to be injected into a human or other mammalian host during the mosquito's next bloodmeal (**Figure 3**, stages 8-12)[3].

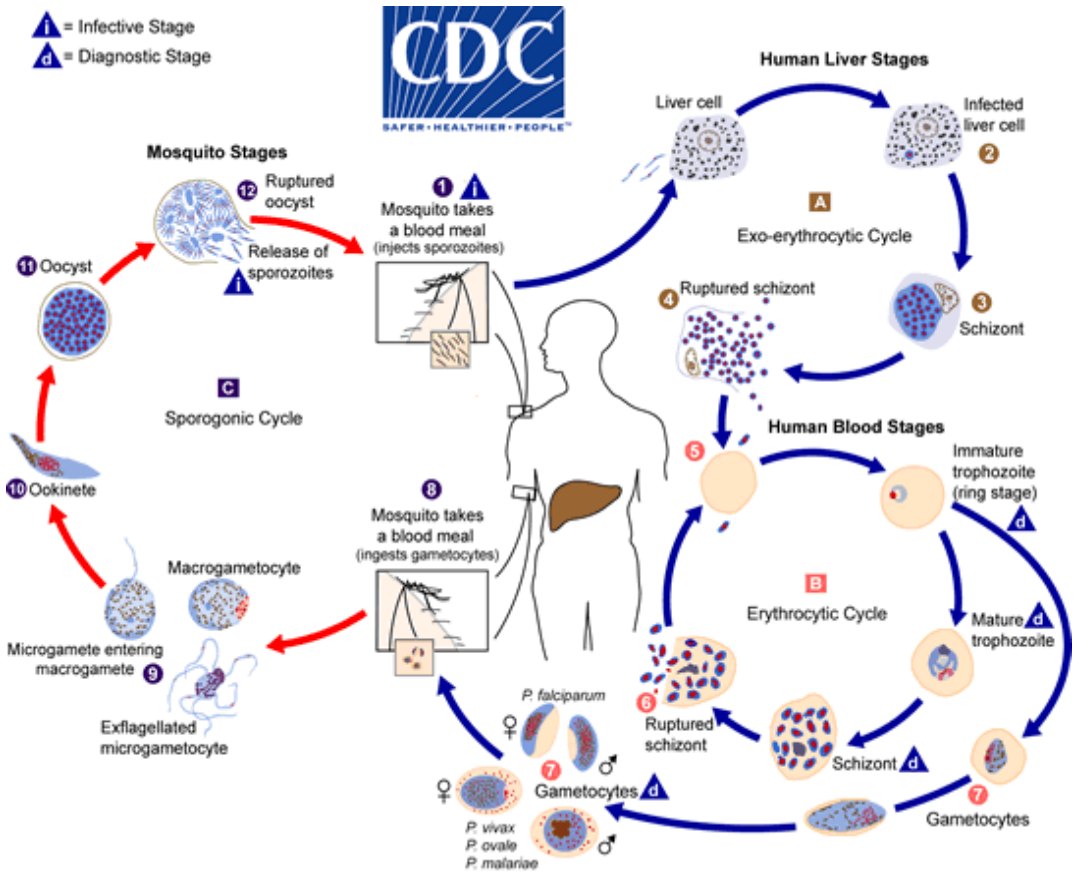


***Figure 3**: The Plasmodium falciparum lifecycle, which includes phases in the mosquito and the human host.[3]*

Anopheles mosquitoes

Plasmodium parasites are transmitted by female *Anopheles* mosquitoes, which require a blood meal for egg development[10]. There are over 450 formally recognised *Anopheles* species, and 41 are considered capable of transmitting malaria at a level of major concern to public health[11]. In Africa, the major malaria vectors are *An. gambiae sensu stricto* (*s.s.*)*, An. coluzzii,* and *An. arabiensis* from the *Anopheles gambiae* species complex, '*Anopheles gambiae sensu lato (s.l.)*'[11,12], and *An. funestus* from the Funestus subgroup[13,14]. The *Anopheles gambiae s.l.* species complex was originally considered to be a single species, but has now been identified to contain at least nine cryptic species which cannot be distinguished morphologically, but can be distinguished genomically[12]. In addition to *An. gambiae s.s., An. coluzzii,* and *An. arabiensis,* there are two other cryptic species in the *Anopheles gambiae s.l.* complex which have been identified to transmit malaria and are considered important vectors: *An. melas,* and *An. merus*[12]*.* Ecological characteristics vary between species, including preferred breeding sites, feeding behaviour, and resting behaviour. For example, *An. melas* and *An. merus* are saltwater tolerant, with *An. melas* present in coastal West Africa, and *An. merus* present in coastal East Africa (**Figure 4**)[13,15]. Whereas, *An. gambiae* and *An. coluzzii* are considered freshwater species, and *An. arabiensis* generally breeds in semi-arid and savannah areas[16]. Furthermore, *An. gambiae* is highly anthropophilic (prefers to bite humans), and endophagic (feeds indoors), whereas *An. arabiensis* is generally considered to be more zoophilic (prefers to bite animals) and exophagic (feeds outdoors), albeit with high intraspecies variability[11]. In addition to these vectors, *An. stephensi*, which was previously confined to certain countries in South Asia and the Arabian Peninsula, has been identified as an invasive malaria vector in Africa[17]. This species is an efficient, urban, vector for malaria and presents a threat to malaria control on the African continent[17], where it has been found in Djibouti (2012)[18], Ethiopia (2016)[19], Sudan (2016)[20], Somalia (2019)[21], Nigeria (2020)[22], Eritrea (2022)[23], Ghana (2022)[24] and Kenya (2022)[17,25].

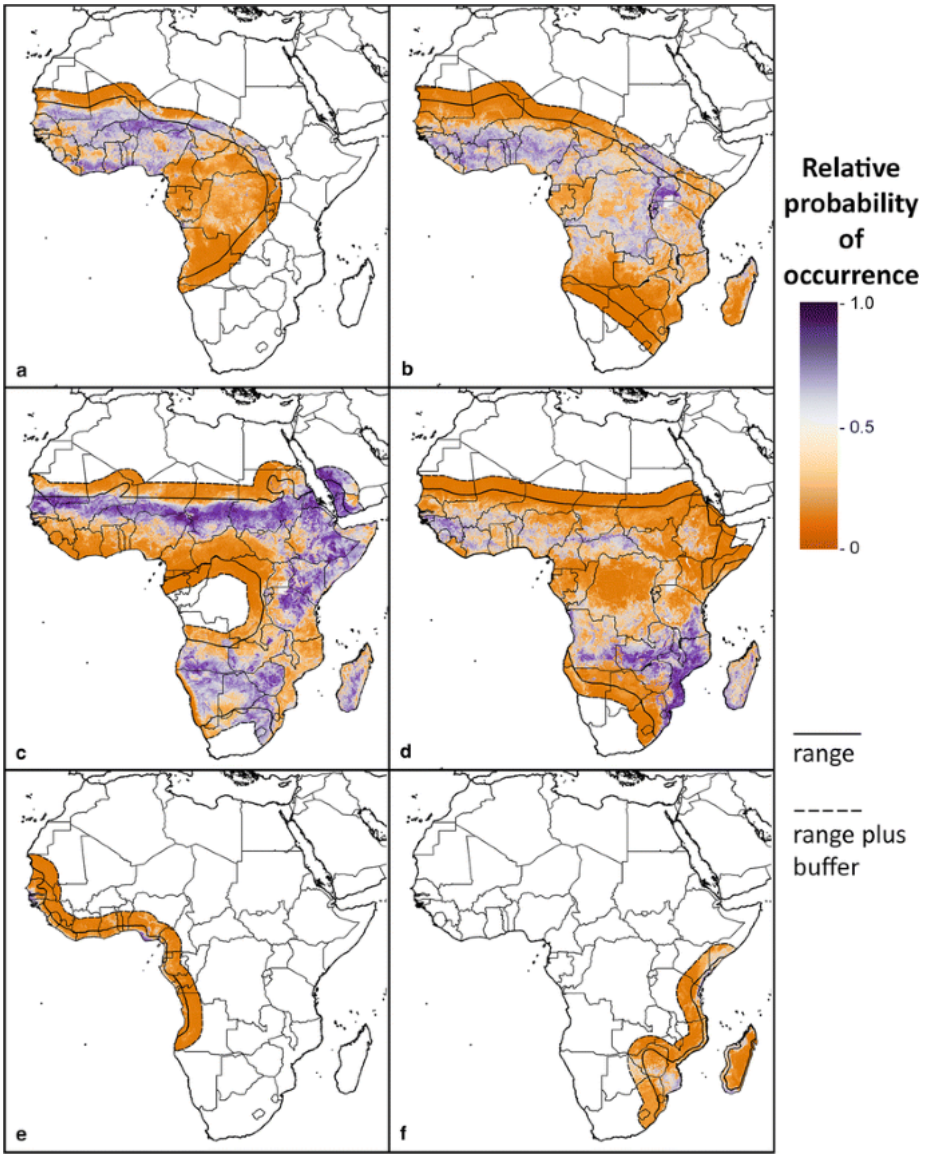


***Figure 4**: Predictive maps for the distribution of key malaria vectors, Fig. 1. in Wiebe et al (2017)[13]. The relative probability of the occurrence of each species is shown within its range, plus a 300km buffer. a) An. coluzzii b) An. gambiae s.s. c) An. arabiensis d) An. funestus e) An melas f) An merus.*

Clinical presentation of malaria

The clinical symptoms of malaria are primarily caused by the rupture of schizonts and destruction of erythrocytes[8]. The disease first presents with non-specific flu-like symptoms, including fever, chills, headache, myalgias, nausea, and vomiting[26–28]. The most characteristic symptom of malaria is fever, which is usually irregular at first and can develop into a cyclic pattern depending on the timing of synchronised schizogony in the erythrocytic cycle (**Figure 3**) [26,27]. This timing varies depending on the infecting species of *Plasmodium.* In *P. ovale* and *P. vivax* infections, schizogony occurs and leads to fever every 48 hours (tertian fever), whereas in *P. malariae* infections, fever occurs every 72 hours (quartan malaria)[26]. In *P. falciparum* infections, fever can occur every 48 hours, but patients

frequently present with irregular and erratic fevers[26,29]. Malaria is commonly classified as asymptomatic, uncomplicated, or complicated (severe)[28]. Severe malaria is often fatal, and almost all cases are caused by *P. falciparum*[8,27,28]. Severe malaria is caused by microvascular obstruction by infected erythrocytes in the capillaries, leading to complications in the central nervous system, respiratory system, pulmonary system and/or hematopoietic system[26,27]. Patients are classified as having severe malaria if they have one more of the clinical or laboratory features in **Table 1**[27,28,30].

***Table 1**: Features of severe falciparum malaria*[27,28,30]

Clinical features of severe malaria
• Impaired consciousness (including unrousable coma) • Prostration (generalised weakness so that the patient is unable to sit, stand or walk without assistance) • Multiple convulsions (more than two episodes within 24h) • Deep breathing and respiratory distress (acidotic breathing) • Acute pulmonary oedema and acute respiratory distress syndrome • Circulatory collapse or shock, systolic blood pressure < 80mm Hg in adults and < 50mm Hg in children • Acute kidney injury • Clinical jaundice plus evidence of other vital organ dysfunction • Abnormal bleeding
Laboratory features of severe malaria • Hypoglycaemia (< 2.2 mmol/l/ or < 40 mg/dl) • Metabolic acidosis (plasma bicarbonate < 15mmol/l) • Severe normocytic anaemia (haemoglobin < 5g/dl, packed cell volume < 15% in children; < 20% in adults) • Haemoglobinuria • Hyperlactataemia (lactate > 5mmol/l) • Renal impairment (serum creatinine > 265µmol/l) • Pulmonary oedema (radiological).

Asymptomatic malaria infections are recognised as *Plasmodium* parasitaemia of any density, without the presence of fever or other acute symptoms, in individuals who have not recently received antimalarial treatment[31,32]. These asymptomatic infections result from partial immunity in individuals, which controls but does not eliminate the *Plasmodium* infection. These infections can persist for several months or even years[32,33], and have also been termed 'chronic infections' due to their association with morbidities including recurrent symptomatic parasitaemia[32], increased risk of invasive bacterial disease[34–36], anaemia[32,37–39], maternal and neonatal mortality[39,40], and cognitive impairment[41,42]. Furthermore, asymptomatic infections present a chronic reservoir of circulating

parasites capable of sustaining malaria transmission, forming a barrier to malaria control and elimination[33,43].

Malaria prevention

The WHO recommends the prevention of malaria through the use of vector control strategies, preventative chemotherapies, and the administration of malaria vaccines[30].

Vector control strategies

Vector control strategies for malaria involve the use of insecticides to combat *Anopheles* mosquitoes. Insecticides can be divided into larvicides, which target mosquito larval stages and are used for spraying mosquito breeding sites, and adulticides, which target adult mosquitoes. Vector control strategies strongly recommended by the WHO include Insecticide Treated Nets (ITNs) and Indoor Residual Spraying (IRS)[30]. There is strong evidence in favour of using ITNs[44–49] and IRS[50–54] for malaria control. A 2018 Cochrane review of 23 different randomised controlled trials (RCTs) identified that the use of ITNs reduced *P. falciparum* malaria prevalence by 17% compared to not using bed nets and by 10% compared to using untreated bed nets[55]. Furthermore, in the range of 15-85% ITN coverage of the population, a 20% increase has been shown to reduce malaria transmission intensity in the community by approximately 50%[56]. This is reflective of the 'community effect', whereby individuals in the same locality benefit from the use of ITNs by others, even if they do not sleep under a net themselves[56]. A subsequent Cochrane review concluded that the addition of IRS with a non-pyrethroid-like insecticide in communities already using ITNs was associated with reduced malaria prevalence when compared to using ITNs alone. Whereas, no additional benefit was identified from the addition of IRS with a pyrethroid-like insecticide[57]. ITNs are credited as the most impactful vector control tool deployed to date, having averted approximately 68% of malaria related deaths in Africa between 2000 and 2015[58]. However, the efficacy of both ITNs and IRS is highly impacted by insecticide resistance in the local vector population[59–61].

Currently, all ITNs contain pyrethroid insecticides[62], which work by targeting the voltage-gated sodium channel (VGSC) and disrupting nervous system function, resulting in paralysis and death[63]. Unfortunately, *Anopheles* mosquitoes have developed widespread resistance to these pyrethroids[64]. In 2017 WHO recommended a new type of ITN; pyrethroid-PBO nets, which showed improved control of malaria transmission in areas of pyrethroid resistance when compared to standard pyrethroid nets[62,65,66]. These nets contain piperonyl butoxide (PBO), which is a synergist that inhibits

specific metabolic enzymes within mosquitoes, enhancing the potency of the pyrethroids[62]. More recently, in 2023, WHO issued a strong recommendation for the deployment of a second new net, pyrethroid-chlorfenapyr ITNs, instead of pyrethroid-only nets in areas of pyrethroid resistance[1]. These pyrethroid-chlorfenapyr ITNs combine a pyrethroid with a pyrrole insecticide that has a separate mode of action, increasing the efficacy of the net[67]. Pyrethroid insecticides target the insect nervous system, whereas pyrroles target respiratory pathways through the disruption of oxidative phosphorylation in the mitochondria[68]. WHO also issued a conditional recommendation for the deployment of these pyrethroid-chlorfenapyr ITNs instead of pyrethroid-PBO nets in areas of known pyrethroid resistance[1]. A third new ITN, pyrethroid-pyriproxyfen nets, were also given a conditional recommendation by WHO for deployment instead of pyrethroid-only nets in areas of pyrethroid resistance. These nets combine a pyrethroid with the insect growth regulator pyriproxyfen, which disrupts mosquito growth and reproduction[69]. However, WHO issued a conditional recommendation *against* the deployment of pyrethroid-pyriproxyfen nets instead of pyrethroid-PBO nets, as the pyrethroid-PBO nets were viewed as more cost-effective in this comparison. Other vector control strategies include larval source management through the larviciding of breeding sites, which is a less frequently used intervention[30]. Additional vector control strategies are under development which do not rely on insecticides. These include the release of sterile male mosquitoes[70], and the release of mosquitoes infected with *Wolbachia*, which is an endosymbiotic bacteria present in multiple species of malaria vector[71]. The presence of *Wolbachia* can reduce the transmission of pathogens inside the mosquito[72].

Preventive chemotherapies

Preventive chemotherapies are used in malaria endemic areas to reduce malaria cases through administering full therapeutic courses of antimalarial drugs at prescheduled times, irrespective of infection status, to treat existing infections and prevent new infections[30]. Preventive chemotherapies that are recommended by WHO include intermittent preventive treatment of malaria in pregnancy (IPTp), perennial malaria chemoprevention (PMC), seasonal malaria chemoprevention (SMC), intermittent preventive treatment of malaria in school-aged children (IPTsc), post-discharge malaria chemoprevention (PDMC), and mass drug administration (MDA), which is used for malaria burden and transmission reduction and mass relapse prevention[30]. Recent WHO recommendations for preventive chemotherapies are more flexible than they have been in past years. WHO no longer specifies strict age groups, transmission intensity thresholds, numbers of doses/cycles, or specific drugs for preventive chemotherapy, giving increased flexibility to National Malaria Control

programmes (NMCPs) to adapt control strategies to their specific settings[30]. The most widely used antimalarial drugs in preventive chemotherapy can be seen in **Table 2.**

***Table 2**: The most widely used antimalarial drugs in preventive chemotherapy regimens recommended by WHO.*

Preventive chemotherapy	**Antimalarial medicine most widely used**[30]
Intermittent preventive treatment of malaria in pregnancy (IPTp)	Sulphadoxine pyrimethamine (SP)
Perennial malaria chemoprevention (PMC) - *formerly intermittent preventive treatment of malaria in infants (IPTi)*	SP
Seasonal malaria chemoprevention (SMC)	Sulphadoxine-amodiaquine (SP-AQ)
Intermittent preventive treatment of malaria in school-aged children (IPTsc)	SP combined with an aminoquinoline (either AQ or piperaquine), and artemisinin-based combination therapy including an aminoquinoline: artesunate-amodiaquine (AS-AQ) or dihydroartemisinin-piperaquine (DHAP).
Post-discharge malaria chemoprevention (PDMC)	SP, artemether-lumefantrine (AL) and DHAP
Mass Drug Administration	Usually, a combination medicine different to that used for first-line malaria treatment, such as SP or DHAP.

Malaria vaccines

In a major stride combatting malaria mortality worldwide, WHO has recommended the use of two circumsporozoite-based malaria vaccines: RTS,S/AS01 (RTS,S) and R21/Matrix-M (R21). The RTS,S vaccine has been delivered in three pilot countries (Ghana, Kenya, and Malawi) since 2019 through the Malaria Vaccine Implementation Programme and has reached nearly 2 million children, resulting in a 13% decrease in all-cause early childhood deaths and a substantial reduction in severe malaria[73]. In October 2021, WHO recommended the use of RTS,S for children living in regions with moderate to high transmission of malaria[30]. Subsequently, in October 2023, WHO recommended a second malaria vaccine, R21/Matrix-M (R21)[74]. R21 was shown to reduce symptomatic malaria cases by 75% during the first 12 months, giving it a similar efficacy to RTS,S, and R21 may be easier and cheaper to produce[74,75]. At the time of writing, at least 28 countries in Africa are planning to introduce vaccines

as part of their national immunisation programmes, and Gavi, the Vaccine Alliance, has approved support to roll out malaria vaccines to 18 countries[76].

Malaria treatment

The emergence of *Plasmodium* resistance to monotherapies led the WHO to recommend artemisinin-based combination therapies (ACTs) as first-line treatment for uncomplicated malaria[1]. These therapies combine a fast-acting artemisinin derivative with a short half-life[77] with a slower acting partner drug to clear any residual parasites following the action of the artemisinin derivative[78]. ACTs used to treat uncomplicated malaria include artemether-lumefantrine (AL), artesunate-amodiaquine (AS+AQ), artesunate-mefloquine (ASMQ), dihydroartemisinin-piperaquine (DHAP), artesunate and sulphadoxine-pyrimethamine (AS+SP), and artesunate-pyronaridine (ASPY)[30]. Severe malaria is usually treated with intravenous or intramuscular artesunate[30].

Antimalarial resistance

The evolutionary arms race

The evolution of resistance to antimalarials is an alarming threat to malaria control efforts[1]. Resistance to older drugs including chloroquine (CQ) and SP has led to decreased efficacy in malaria treatment, and CQ resistance alone has contributed to millions of excess malaria deaths[79,80]. In response to increasing resistance, WHO recommended the use of ACTs as first-line therapy, and by 2005 ACTs were the standard treatment choice for uncomplicated malaria across Africa[81]. Historically, resistance to antimalarials was first detected in Southeast Asia before spreading into Africa[82,83]. ACTs remain highly efficacious for now[84]. However, partial resistance to artemisinin (ART-R) has emerged in Southeast Asia[85], and recent studies have confirmed the emergence of ART-R in Africa, including in Rwanda[86–88], Uganda[80,89,90] , Eritrea[84,91,92] and Tanzania[93]. The presence of ART-R is also suspected in Ethiopia[94], the Democratic Republic of Congo (DRC)[95] and Sudan[96] due to the presence of mutations in the *pfkelch13* gene associated with resistance. Estimates in the scenario of widespread artemisinin partial resistance and partner drug resistance predict an excess of 80,000 deaths per year in Africa and a yearly economic impact of US$ 1 billion for the continent[84]. Strengthening the surveillance of antimalarial drug resistance is essential to address this threat, and is central to the 2022 WHO 'Strategy to respond to antimalarial drug resistance in Africa'[84].

Genomic surveillance of antimalarial resistance

Therapeutic efficacy studies (TESs) are the gold standard for determining the efficacy of antimalarials. These studies involve a diagnosis of malaria, confirmed by testing, followed by the supervised use of a particular drug and follow-up of the patient for a defined period of time[84]. TESs alone are not sufficient to confirm drug resistance. These must be confirmed using other methods, such as[97,98]:

a) Molecular methods to identify molecular markers of drug resistance in the *Plasmodium* genome, including single nucleotide polymorphisms (SNPs) and copy number variants (CNVs)
b) *Ex vivo* and *in vitro* drug assays, where fresh or cultured *Plasmodium* parasites are exposed to drugs within a laboratory environment and their drug sensitivity is measured
c) *In vivo* measurements of drug levels in the blood to distinguish whether treatment failure is due to resistance or due to insufficient antimalarial drug exposure in the patient

Ex vivo, in vivo, and *in vitro* surveillance of antimalarial resistance is expensive and resource intensive, requiring substantial laboratory infrastructure and specialised staff. However, rapid improvements in next-generation-sequencing (NGS) technology have reduced the time taken and cost of genome sequencing, resulting in an explosion in the availability of genome sequence data[97]. In turn, this has propelled the development of global databases of *P. falciparum* data, including open-source databases of whole-genome sequence (WGS) data created by the collaborative network MalariaGEN and the partner collaboration PAMGEN (Pan-African Malaria Genetic Epidemiology Network)[99]. The availability of vast repositories of sequence data enables the surveillance of molecular markers associated with resistance. So whilst TESs remain the gold-standard and the use of *in vivo* and *in vitro* studies remains central to diagnosing resistance, this wealth of genomic data means that genomic surveillance has an increasingly important role in monitoring antimalarial resistance, including the identification of novel molecular markers of drug resistance[100,101].

P. falciparum has three genomes: nuclear, mitochondrial and the apicoplast. The nuclear genome is 23 megabases (mb) long and made up of 14 chromosomes, while the mitochondrial genome is 6 kilobases (kb), and the apicoplast genome is 35kb[102]. WGS data is popular for in depth analysis of resistance markers. However, WGS is expensive and requires high concentrations of parasite DNA to generate. This problem is magnified when analysing parasite DNA from dried blood spots, as these blood spots contain mostly human DNA[103]. One method of preferentially amplifying parasite DNA from dried blood spots is selective whole genome amplification (SWGA)[103]. This technique uses primers which preferentially bind to the parasite DNA and selectively amplify this, instead of human DNA, enabling higher concentrations of parasite DNA to be obtained which can then be used in WGS.

Another lower cost option for sequencing parasite DNA is targeted amplicon sequencing, which can be used for sequencing shorter sections of DNA in multiplex reactions, drastically reducing the cost and time taken to sequence target regions within the genome[104]. Amplicon sequencing can be implemented using long read Oxford Nanopore Technology (ONT) such as the MinIon[105], or short read MiSeq Illumina sequencing[106]. Molecular markers recognised by the WHO as associated with antimalarial resistance can be seen in **Table 3**[98].

***Table 3**: WHO recognised molecular markers of P. falciparum resistance to antimalarial drugs[98]*

Drug	**Loci**	**Polymorphisms associated with drug resistance**
Chloroquine	*Pfcrt*	K76T + different sets of mutations at other codons (including C72S, M74I, N75E, A220S, Q271E, N326S, I356T and R371I)
	Pfmdr1 (in combination with *Pfcrt* mutations only)	N86Y, Y184F, S1034C, N1042D and D1246Y
Amodiaquine	Yet to be validated	Studies show selection for *Pfmdr1* 86Y
Piperaquine	*Pfpm2-3*	*Pfpm2-3* increased copy number
	Pfcrt	Detected *in vivo*: T93S, H97Y, F145I, I218F and C350R
		Detected *in vitro*: T93S, H97Y, F145I, I218F, M343L and G353V
Pyrimethamine	*Pfdhfr*	N51I, C59R, S108N and I164L
Sulphadoxine	*Pfdhps*	S436A/F, A437G, K540E, A581G and A613T/S
Proguanil	*Pfdhfr*	A16V, N51I, C59R, S108N and I164L
Lumefantrine	Yet to be validated	Studies show selection for *pfmdr1* N86
Mefloquine	*Pfmdr1*	*Pfmdr1* increased copy number
Atovaquone	*Pfcytb*	Y268N/S/C
Artemisinin and its derivatives	*PfK13*	Validated markers: F446I, N458Y, M476I, Y493H, R539T, I543T, P553L, R561H, P574L, C580Y Candidate/associated markers: P441L, G449A, C469F/Y, A481V, R515K, P527H, N5371I/D, G538V, V568G, R622I, A675V

Insecticide resistance

Another evolutionary arms race

Along with the evolution of antimalarial resistance, a second evolutionary arms race threatens malaria control and elimination; the evolution of insecticide resistance in *Anopheles* vectors. Vector control using insecticides has been the most successful method of reducing malaria cases and deaths to date, and insecticide resistance threatens the efficacy of these interventions[58]. There are six classes of adulticide insecticides used in vector control: pyrethroids, organochlorines, carbamates, organophosphates, neonicotinoids and pyrroles[1]. Unfortunately, resistance to at least one class of insecticide has been reported in malaria vectors from 78 of the 88 malaria endemic countries that provided data to WHO between 2010 and 2020[1]. Furthermore, 19 countries have confirmed resistance to pyrethroids, organochlorines, carbamates and organophosphates in at least one malaria vector[1]. Alarmingly, resistance to pyrethroids, which are the insecticide class used in ITNs, was reported by 87% of countries[1,62]. WHO developed the Global Plan for Insecticide Resistance of Malaria (GPIRM) in 2012[107] in efforts to combat the threat of insecticide resistance, which was followed by the development of the Global Vector Control Response (GVCR) for 2017-2030[108]. The GVCR recommends that countries develop Insecticide Resistance Monitoring and Management Plans (IRMMPs) involving increased surveillance of insecticide resistance in efforts to tackle this rising threat[109].

Insecticide resistance is caused by four main mechanisms[110,111]:

1. Target-site resistance, which occurs when a genetic mutation alters the protein receptor within the mosquito that the insecticide is targeting, rendering the insecticide unable to bind and exhibit its mode of action.
2. Metabolic resistance, which occurs when mosquito metabolic enzymes are able to break down or sequester insecticides before they can have a toxic effect. There are three families of metabolic enzymes which are strongly associated with insecticide resistance in malaria vectors, including monooxygenases (P450s), esterases, and glutathione-S-transferases (GSTs).
3. Cuticular resistance, where genetic polymorphisms lead to a thickening of the mosquito's cuticle, reducing the ability of the insecticide to penetrate the cuticle and harm the mosquito.
4. Behavioural resistance, where a change in mosquito activity enables it to avoid exposure to insecticides, such as avoidance of ITNs through changing behaviour to bite during the day when people are not protected by ITNs.

The most studied mechanisms of insecticide resistance are target-site resistance and metabolic resistance.

Target-site resistance

Pyrethroids and the organochlorine dichlorodiphenyltrichloroethane (DDT) work by binding to and disrupting the action of the voltage-gated sodium channels of insect nerves, resulting in insect paralysis and death[63,112,113]. The most well characterised SNPs in the *vgsc* gene are the knock-down resistance (kdr) mutations L995F/S, which can cause pyrethroid resistance (position 1014 in *Musca domestica*)[114]. L995F was first identified in West Africa, whereas L995S was first identified in East Africa[115,116]. However, the distribution of these two alleles has been shown to overlap[117]. Other mutations in the *vgsc* gene have been identified which contribute to pyrethroid resistance. These include the N1570Y mutation, which has been found to increase pyrethroid resistance when found in association with L995F[118], the V402L mutation associated with pyrethroid resistance[119], and the I1527T mutation associated with permethrin resistance[120].

Mutations in the gamma-aminobutyric acid (GABA) receptor (*rdl* – resistant to dieldrin) have been associated with resistance to the organochlorine dieldrin. The A296G mutation has been found to confer dieldrin resistance in *An. gambiae*[121], and the A296S mutation has been found to confer dieldrin resistance in *An. arabiensis*[121], *An. stephensi, Aedes aegypti* and *An. funestus*[122]. Organochlorines including DDT and dieldrin have been banned due to environmental and human health concerns[113]. However, widely used pesticides are also known to act on the GABA receptor[123], and the target-site mutations associated with organochlorine resistance can still be identified in mosquito populations[124,125].

Organophosphate and carbamate insecticides work by inhibiting the function of acetylcholinesterase (AChE), which is essential for terminating synaptic transmission by hydrolysing the neurotransmitter acetylcholine[126]. They work by competitive binding to the AChE active site, resulting in continuous nervous transmission and associated muscle or nerve fibre stimulation[127,128]. Mosquito resistance to organophosphate and carbamate insecticides has been associated with the G280S mutation in one of the genes encoding AChE, *ace1*[129] (G119S in *Torpedo californica*[130]), and organophosphate resistance has been associated with *ace1* gene duplications[128].

Metabolic resistance

Metabolically mediated insecticide resistance has been strongly associated with three enzyme families: cytochrome P450 monooxygenases (P450s)[131,132], esterases[111] and glutathione-S-transferases (GSTs)[133]. Increased activity of these enzymes through upregulation or gene duplication increases the mosquito's ability to sequester or breakdown insecticides, leading to insecticide resistance[110]. Pyrethroid resistance in *Anopheles* mosquitoes has been strongly associated with copy number variants (CNVs) in the P450 encoding genes: *cyp6aa1*[134], *cyp6p3*, *cyp6m2*, *cyp6z1* and *cyp9k1*[135,136], and in the GST encoding gene *gste2*[137]. These have been identified as the most important metabolic genes to date, and have all been shown to metabolise insecticides *in vitro*[134,138,139]. Insecticide resistance has been associated with significant overexpression of genes in these metabolic families, including *cyp6aa1*, *cyp6m2*, *cyp6p3, cyp6p4, cyp6z1,* and *cyp9k1* in *An. gambiae s.l.*[135,139–141]. Furthermore, specific SNPs within these gene families have been associated with resistance. This includes L43F in *cyp4j5* and I236M in *cyp6p4,* which have been associated with resistance to pyrethroids in *An. gambiae*[139,142]. Furthermore, the L119F[143] and I114T[137] SNPs in *gste2* have been associated with resistance to pyrethroids and DDT. Metabolic resistance to carbamates and organophosphates has been associated with increased copy number of *ace1*[128]. These *ace1* duplications commonly co-occur with the *ace1* G280S SNP on the same resistance haplotype, which has emerged in *An. gambiae* and then spread into *An. coluzzii* in West Africa through interspecific hybridisation (introgression)[128]. Several transcription factors have been implicated in regulating pyrethroid resistance, including *Maf-S* and *Met,* the silencing of which resulted in significantly reduced expression of some metabolic genes, including *cyp6m2*[144,145].

Genomic surveillance of insecticide resistance

The gold-standard for monitoring the presence of insecticide resistance is the use of phenotypic susceptibility bioassays using WHO tube tests, WHO bottle bioassays, CDC bottle bioassays or WHO susceptibility tests for mosquito larvae[146]. After the presence of insecticide resistance has been confirmed, WHO recommends that intensity bioassays are conducted to determine the intensity level of the resistance, and that synergist-insecticide bioassays using PBO are conducted to identify the presence of metabolic resistance[146]. WHO also recommends the use of molecular and biochemical assays to analyse enzyme expression of metabolic genes and target-site SNPs which may be the mechanistic cause behind phenotypic resistance[110,146].

Genomic surveillance using WGS or target-site sequencing can be used to identify target-site mutations that have previously been associated with resistance[124,147,148], and to identify novel SNPs that may be associated with resistance[63]. Molecular approaches can also be used to elucidate metabolic mechanisms of resistance. For example, RNA sequencing has been used to explore patterns of gene expression[149] and WGS has been used to identify CNVs including metabolic genes which are associated with insecticide resistance[138]. The *Anopheles gambiae* 1000 Genomes Project (Ag1000G; http://www.malariagen.net/ag1000g) provides an open source resource of whole genome sequences from the *Anopheles gambiae s.l.* complex for the investigation of mosquito genome variation and evolution[150]. This has enabled large scale studies of *Anopheles* population genetic variation[138,150–152], which have confirmed that *Anopheles* mosquitoes are among the most genetically diverse eukaryotic organisms known[150]. The *An. gambiae s.s.* genome is ~264 mb long and is made up of 3 chromosomes and the mitochondrial genome[153]. The insecticide resistance status of *Anopheles* mosquitoes on the Bijagós Archipelago of Guinea-Bissau is poorly understood, and genomic data from malaria vectors on the islands was very limited prior to the research that has been conducted for this thesis.

The Bijagós Archipelago

Malaria endemicity

Malaria is a persistent health problem in Guinea-Bissau in West Africa, where there is a population of around 2.1 million people and an estimated 225,200 malaria cases and 1,023 malaria deaths occurred in 2022[1]. The Bijagós is an archipelago of 88 islands and islets situated off the coast of Guinea-Bissau, with a population of around 25,000 people (**Figure 5**)[154]. The islands are separated from mainland Guinea-Bissau by around 50 km of Atlantic Ocean and are primarily accessed by ferry, which takes four hours and sails only twice each week. The islands have a rainy season between June and October/November and a dry season from December to May, with peak malaria transmission occurring at the end of the rainy season in October and November[155]. Cross-sectional malaria prevalence surveys using 18S rRNA quantitative polymerase chain reaction (qPCR) identified the prevalence of *P. falciparum* infection to be 16.7% in November 2017, 17.3% in November 2018, and 10.0% in November 2019[156]. The persistent circulation of subclinical parasitaemia on the Archipelago makes malaria elimination particularly difficult, as asymptomatic malaria cases are a major driver of transmission, and asymptomatic individuals do not seek treatment[31,157].

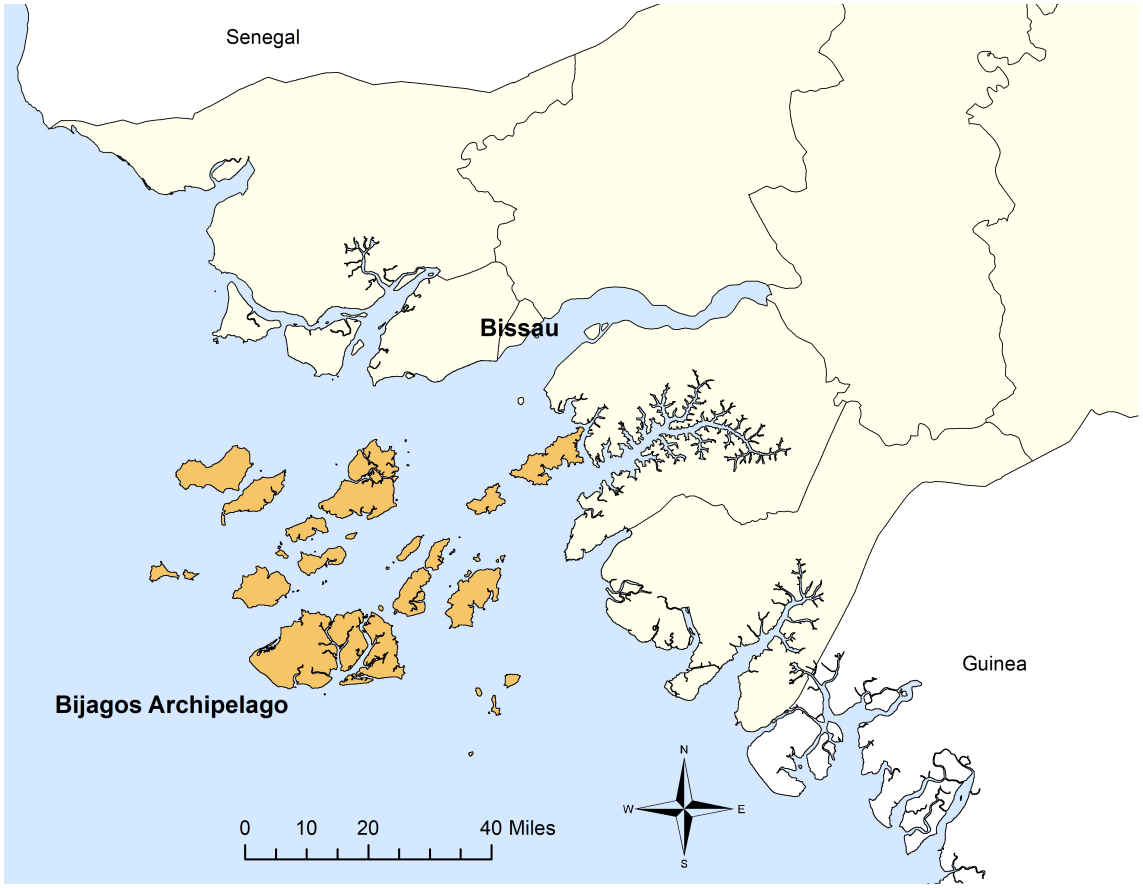


***Figure 5**: Bijagós Archipelago, Guinea-Bissau. Map created by E. Pretorius, LSHTM, using ArcGIS ArcMap10.8.1[125].*

Malaria vectors on the Bijagós Archipelago

A vector surveillance survey on the most populated island of the Archipelago, Bubaque, was conducted in 2017[155]. This survey identified all *Anopheles* species on the island to be members of the *An. gambiae sensu lato (s.l.)* complex. Of the species present, *An. gambiae s.s.* was identified as the primary malaria vector, and *An. melas* was hypothesised to be responsible for low level transmission during the dry season[155]. A larger surveillance study on 16 islands took place during the peak malaria transmission in 2019[158]. This study also identified all *Anopheles* species on the Archipelago to be members of *An. gambiae s.l.* complex and identified 85.2% of trapped mosquitoes to be *An. melas.* This study identified a sporozoite rate of 0.86% and found that all *P. falciparum* circumsporozoite protein (CSP) positive specimens collected were *An. melas,* indicating that this species has a potential role in malaria transmission on the Bijagós[158]. This is supported by previous studies which have identified *An. melas* to have a role in malaria transmission, including in neighbouring The Gambia[159,94], Senegal[160], and Equatorial Guinea[161].

The *Anopheles gambiae* complex is an example of speciation in the presence of gene flow, with considerable introgression of alleles between species[150,162,163]. Guinea-Bissau is a region of uncharacteristically high hybridisation rates between *An. gambiae s.s.* and *An. coluzzii,* which have been recorded consistently at over 20% since 1995 [163,164]. This hybridisation rate is much higher than has been recorded in West and Central Africa, where rates are usually below 1% despite broad sympatry[165], and recent research suggests that these hybrids may actually be a novel putative cryptic taxon, provisionally named the "Bissau molecular form"[166]. This adds further complexity to understanding gene flow in this mosquito population and the impact that this may have on the spread of insecticide resistance alleles.

Malaria control on the Bijagós Archipelago

The National Malaria Control Program (NCMP) of Guinea-Bissau has deployed a combination of parasite and vector control tools to combat malaria, including case detection and treatment with AL, intermittent preventive treatment in pregnancy (IPTp) with SP, and pyrethroid treated ITNs[167]. Prior to the introduction of ACTs, CQ was used routinely in Guinea-Bissau at triple the standard dose, and was often prescribed for children without malaria due to the limited availability of malaria diagnostics[168]. Despite high CQ doses being routinely used for decades, treatment with CQ was well-tolerated in the population[169]. AL was introduced as first-line treatment for malaria in 2008[170]. SP was not used for first-line treatment, but was recommended for second-line treatment between 1996 and 2007[171]. ITNs were distributed in 2011, 2014 and 2017, with planned deployment every 3 years[172]. Approximately ninety percent of distributed nets have been PermaNet® 2.0, which are treated with the pyrethroid deltamethrin[125]. ITN coverage is estimated to be high, at 92%. However, IPTp coverage is low and in some areas non-existent[154]. Other vector control interventions such as indoor residual spraying and larviciding are not used on the Bijagós[154]. There was one pilot round of SMC with SP-AQ delivered on the islands from August to November 2020, but this was not continued[154]. The MDA trial MATAMAL was implemented on the Bijagós during 2021 and 2022. MATAMAL was a cluster-randomised, community-wide MDA with two arms randomised to either combined oral ivermectin and dihydroartemisinin-piperaquine (DHAP), or oral ivermectin-placebo and DHAP[154]. Malaria prevalence on the islands was significantly reduced during the peak transmission season following the MDA intervention in 2022 to a *P. falciparum* prevalence of 5.05% in the control arm and 6.64% in the intervention arm[156]. *P. falciparum* prevalence was quantified using the varATS qPCR, which targets the var gene acidic terminal sequence (varATS, 59 copies per genome) and is more sensitive than the 18S rRNA qPCR (5-8 copies per genome)[173].

Antimalarial resistance on the Bijagós Archipelago

The antimalarial resistance status of *Plasmodium* parasites on the Bijagós Archipelago has not previously been studied, and genomic data for Bijagós *Plasmodium* isolates was absent prior to the research conducted for this thesis. However, studies have been conducted on mainland Guinea-Bissau, where CQ was used routinely before the introduction of AL in 2008[174–178]. A study in Bissau analysed molecular markers from 318 malaria patients between October 2014 and March 2016 using multiplex amplicon sequencing from venous blood samples collected at Bandim Health Centre. This study found a *pfcrt* M74I frequency of 52.9%, N75E of 52.9%, K76T of 52.4%, *pfdhps* V431I frequency of 0.6%, A436S of 35.7%, A437G of 66.8%, K540E of 0.3%, A613S/T of 0.7%, *pfdhfr* N51I frequency of 88%, C59R of 87.1%, S108N of 92.4%, and *pfmdr1* N86Y frequency of 9.7%, N1042D of 0.3% and D1246Y of 1%[104]. No markers in the *pfkelch13* gene previously associated with artemisinin partial resistance were identified[104]. A concurrent study in Bissau analysed molecular markers of antimalarial resistance from 1390 rapid diagnostic tests (RDTs) which tested positive for *Plasmodium* DNA by PCR, collected between May 2014 and April 2017[171]. Molecular markers were analysed in DNA extracted from these RDTs using amplicon sequencing. Similar frequencies of antimalarial resistance mutations were identified. The *pfcrt* c. 72-76 CVMNK wildtype haplotype decreased significantly from 2014 to 2016; this haplotype was found in 67% of samples in 2014, 60% in 2015 and 45% in 2016, with mutant K76T rising significantly in prevalence from 43% in 2014, to 53% in 2015, to 61% in 2016. *Pfmdr1* N86 increased in prevalence from 88% in 2014, to 97% in 2015, to 92% in 2016. The *pfdhfr* IRN triple mutant (N51I, C59R, S108N) was found in 73% of samples in 2014, 92% in 2015, and 85% in 2016. The *pfdhps* A437G mutation was found in 33% of samples in 2014, 53% in 2015, and 58% in 2016. Eighteen SNPs were identified in the *pfkelch13* gene, including two non-synonymous SNPs in the propeller region; R592K and T535M, but no SNPs associated with artemisinin partial resistance were identified. The findings of these studies were consistent with SNP frequencies previously published between 2003 and 2012[170], but also indicated a continued increase in the frequency of *pfmdr1* N86, which may be reflective of the use of AL for malaria treatment in Guinea-Bissau[179–181]. Furthermore, these studies indicate increasing frequency of *pfdhfr/pfdhps* mutations associated with SP resistance, which likely reflects of the use of SP for IPTp[104,170,178]. Importantly, the efficacy of AL as first-line malaria treatment in Guinea-Bissau was measured between 2012 and 2015, and is still high despite the increased prevalence of *pfmdr1* N86, resulting in 94% adequate clinical and parasitological response[182].

Insecticide resistance on the Bijagós Archipelago

Prior to this thesis, the only phenotypic testing for insecticide resistance on the Bijagós Archipelago was published by Ant *et al* (2020)[155], who tested the susceptibility of *An. gambiae s.l.* on Bubaque island in 2017 to two insecticides; permethrin and α-cypermethrin. This study used CDC-bottle bioassays and found full susceptibility to permethrin but moderate resistance to α-cypermethrin. This study investigated the presence of *vgsc* L995F and L995S and identified L995F at a frequency of 36% in *An. gambiae s.s.*, 35% in *An. coluzzii* and 42% in *An. gambiae s.s. - An. coluzzii* hybrids. No L995F alleles were identified in *An. melas* mosquitoes[155]. Synergist assays with pre-exposure to PBO followed by exposure to α-cypermethrin resulted in 100% mortality, suggesting the likely presence of metabolic mechanisms of α-cypermethrin resistance. Furthermore, carrying the L995F mutation was significantly associated with the α-cypermethrin resistant phenotype[155]. The prevalence of other mutations associated with insecticide resistance had not been tested prior to this thesis. However, phenotypic tests for resistance and additional molecular markers of resistance have previously been investigated on mainland Guinea-Bissau. The most recent study was conducted in 2018 in five regions of the mainland (Bafatá, Bissau, Buba, Cacheu, and Gabu) and identified resistance to deltamethrin and permethrin in *An. gambiae s.l.*[172]. This study identified *vgsc* L995F at high frequencies (75.01 - 91.96 %) and L995S at low frequencies (0.00 - 1.25 %) across Buba and Gabu-Bafatá. Whereas, in the Bissau sentinel site, L995S was found at higher frequencies (58.71 %) and L995F was found at lower frequencies (31.29 %). This study also identified the presence of *vgsc* N1570Y mutation across all sites (6.62 % - 17.56 %) and the presence of *ace1* G280S (ranging from 0.00 % in Bissau to 12.4 % in Gabu). This study assessed expression levels of seven cytochrome P450 genes, and identified statistically significant overexpression of *cyp6z1* in the Gabu *An. gambiae* s.s – *An. coluzzii* hybrids[172].

References

1. World malaria report 2023. https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2023.
2. Amratia, P. *et al.* Indirect effects of the COVID-19 pandemic on malaria intervention coverage, morbidity, and mortality in Africa: a geospatial modelling analysis. *The Lancet Infectious Diseases* **21**, 59–69 (2021).
3. Centers for Disease Control and Prevention (CDC), About Malaria. https://www.cdc.gov/malaria/about/biology/index.html.

4. Mikolajczak, S. A. *et al.* Plasmodium vivax liver stage development and hypnozoite persistence in human liver-chimeric mice. *Cell Host Microbe* **17**, 526–535 (2015).
5. Collins, W. E. & Jeffery, G. M. Plasmodium ovale: Parasite and Disease. *Clin Microbiol Rev* **18**, 570–581 (2005).
6. Battle, K. E. *et al.* Geographical variation in Plasmodium vivax relapse. *Malaria Journal* **13**, 144 (2014).
7. White, N. J. Determinants of relapse periodicity in Plasmodium vivax malaria. *Malaria Journal* **10**, 297 (2011).
8. Trampuz, A., Jereb, M., Muzlovic, I. & Prabhu, R. M. Clinical review: Severe malaria. *Crit Care* **7**, 315–323 (2003).
9. Sato, S. Plasmodium—a brief introduction to the parasites causing human malaria and their basic biology. *Journal of Physiological Anthropology* **40**, 1–13 (2021).
10. Hansen, I. A., Attardo, G. M., Rodriguez, S. D. & Drake, L. L. Four-way regulation of mosquito yolk protein precursor genes by juvenile hormone-, ecdysone-, nutrient-, and insulin-like peptide signaling pathways. *Front Physiol* **5**, 103 (2014).
11. Sinka, M. E. Global Distribution of the Dominant Vector Species of Malaria. in *Anopheles mosquitoes - New insights into malaria vectors* (IntechOpen, 2013). doi:10.5772/54163.
12. Loughlin, S. O. The expanding Anopheles gambiae species complex. *Pathogens and Global Health* **114**, 1–1 (2020).
13. Wiebe, A. *et al.* Geographical distributions of African malaria vector sibling species and evidence for insecticide resistance. *Malaria Journal* **16**, 85 (2017).
14. Msugupakulya, B. J. *et al.* Changes in contributions of different Anopheles vector species to malaria transmission in east and southern Africa from 2000 to 2022. *Parasites & Vectors* **16**, 408 (2023).
15. White, G. B. Anopheles gambiae complex and disease transmission in Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **68**, 278–298 (1974).

16. Kirby, M. J. & Lindsay, S. W. Responses of adult mosquitoes of two sibling species, Anopheles arabiensis and A. gambiae s.s. (Diptera: Culicidae), to high temperatures. *Bull Entomol Res* **94**, 441–448 (2004).
17. Al-Eryani, S. M. *et al.* Public health impact of the spread of Anopheles stephensi in the WHO Eastern Mediterranean Region countries in Horn of Africa and Yemen: need for integrated vector surveillance and control. *Malaria Journal* **22**, 187 (2023).
18. Faulde, M. K., Rueda, L. M. & Khaireh, B. A. First record of the Asian malaria vector Anopheles stephensi and its possible role in the resurgence of malaria in Djibouti, Horn of Africa. *Acta tropica* **139**, 39–43 (2014).
19. Carter, T. E. *et al.* First detection of Anopheles stephensi Liston, 1901 (Diptera: culicidae) in Ethiopia using molecular and morphological approaches. *Acta Trop* **188**, 180–186 (2018).
20. Ahmed, A., Khogali, R., Elnour, M.-A. B., Nakao, R. & Salim, B. Emergence of the invasive malaria vector Anopheles stephensi in Khartoum State, Central Sudan. *Parasites & Vectors* **14**, 511 (2021).
21. Ali, S., Samake, J. N., Spear, J. & Carter, T. E. Morphological identification and genetic characterization of Anopheles stephensi in Somaliland. *Parasites & Vectors 2022 15:1* **15**, 1–6 (2022).
22. Initiative, P. M. Researchers on the Frontlines Tracking the Spread of a New Mosquito in Nigeria. *PMI* https://www.pmi.gov/researchers-on-the-frontlines-tracking-the-spread-of-a-new-mosquito-in-nigeria/ (2023).
23. Abubakr, M. *et al.* The Phylodynamic and Spread of the Invasive Asian Malaria Vectors, Anopheles stephensi, in Sudan. *Biology (Basel)* **11**, 409 (2022).
24. Afrane, Y. A. *et al.* First detection of Anopheles stephensi in Ghana using molecular surveillance. *bioRxiv* 2023.12.01.569589 (2023) doi:10.1101/2023.12.01.569589.
25. Partners convening a regional response to the invasion of Anopheles stephensi in Africa: meeting report, 8–10 March 2023. https://www.who.int/publications-detail-redirect/9789240075535.

26. Crutcher, J. M. & Hoffman, S. L. Malaria. in *Medical Microbiology* (ed. Baron, S.) (University of Texas Medical Branch at Galveston, Galveston (TX), 1996).
27. Bartoloni, A. & Zammarchi, L. Clinical Aspects of Uncomplicated and Severe Malaria. *Mediterr J Hematol Infect Dis* **4**, e2012026 (2012).
28. Phillips, M. A. *et al.* Malaria. *Nature reviews. Disease primers* **3**, 17050 (2017).
29. Zekar, L. & Sharman, T. Plasmodium falciparum Malaria. in *StatPearls* (StatPearls Publishing, Treasure Island (FL), 2024).
30. WHO guidelines for malaria, 16 October 2023. (2023).
31. Lindblade, K. A., Steinhardt, L., Samuels, A., Kachur, S. P. & Slutsker, L. The silent threat: asymptomatic parasitemia and malaria transmission. *Expert Rev Anti Infect Ther* **11**, 623–639 (2013).
32. Chen, I. *et al.* "Asymptomatic" Malaria: A Chronic and Debilitating Infection That Should Be Treated. *PLoS Med* **13**, e1001942 (2016).
33. Bousema, T., Okell, L., Felger, I. & Drakeley, C. Asymptomatic malaria infections: detectability, transmissibility and public health relevance. *Nat Rev Microbiol* **12**, 833–840 (2014).
34. Nadjm, B. *et al.* WHO guidelines for antimicrobial treatment in children admitted to hospital in an area of intense Plasmodium falciparum transmission: prospective study. *BMJ* **340**, c1350 (2010).
35. Mabey, D. C. W., Brown, A. & Greenwood, B. M. Plasmodium falciparum Malaria and Salmonella Infections in Gambian Children. *The Journal of Infectious Diseases* **155**, 1319–1321 (1987).
36. Scott, J. A. G. *et al.* Relation between falciparum malaria and bacteraemia in Kenyan children: a population-based, case-control study and a longitudinal study. *Lancet* **378**, 1316–1323 (2011).
37. Tine, R. C. K. *et al.* The association between malaria parasitaemia, erythrocyte polymorphisms, malnutrition and anaemia in children less than 10 years in Senegal: a case control study. *BMC Res Notes* **5**, 565 (2012).

38. Kurtzhals, J. A. *et al.* Anaemia caused by asymptomatic Plasmodium falciparum infection in semi-immune African schoolchildren. *Trans R Soc Trop Med Hyg* **93**, 623–627 (1999).
39. Cottrell, G. *et al.* Submicroscopic Plasmodium falciparum Infections Are Associated With Maternal Anemia, Premature Births, and Low Birth Weight. *Clinical Infectious Diseases* **60**, 1481–1488 (2015).
40. Tagbor, H., Bruce, J., Browne, E., Greenwood, B. & Chandramohan, D. Malaria in pregnancy in an area of stable and intense transmission: is it asymptomatic? *Tropical Medicine & International Health* **13**, 1016–1021 (2008).
41. Fernando, D., de Silva, D., Carter, R., Mendis, K. N. & Wickremasinghe, R. A randomized, double-blind, placebo-controlled, clinical trial of the impact of malaria prevention on the educational attainment of school children. *Am J Trop Med Hyg* **74**, 386–393 (2006).
42. Clarke, S. E. *et al.* Effect of intermittent preventive treatment of malaria on health and education in schoolchildren: a cluster-randomised, double-blind, placebo-controlled trial. *The Lancet* **372**, 127–138 (2008).
43. Bousema, J. T. *et al.* Plasmodium falciparum gametocyte carriage in asymptomatic children in western Kenya. *Malaria journal* **3**, (2004).
44. Alonso, P. L. *et al.* A malaria control trial using insecticide-treated bed nets and targeted chemoprophylaxis in a rural area of The Gambia, West Africa 6. The impact of the interventions on mortality and morbidity from malaria. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **87**, 37–44 (1993).
45. Kroeger, A., Mancheno, M., Alarcon, J. & Pesse, K. Insecticide-impregnated bed nets for malaria control: varying experiences from Ecuador, Colombia, and Peru concerning acceptability and effectiveness. *Am J Trop Med Hyg* **53**, 313–323 (1995).
46. D'Alessandro, U. *et al.* A comparison of the efficacy of insecticide-treated and untreated bed nets in preventing malaria in Gambian children. *Trans R Soc Trop Med Hyg* **89**, 596–598 (1995).

47. Sih, C. *et al.* Efficacy of chlorfenapyr-pyrethroid and piperonyl butoxide-pyrethroid long-lasting insecticidal nets (LLINs) compared to pyrethroid-only LLINs for malaria control in Côte d'Ivoire: a three group, cluster randomised trial. *Trials* **25**, 151 (2024).
48. Protopopoff, N. *et al.* Effectiveness of piperonyl butoxide and pyrethroid-treated long-lasting insecticidal nets (LLINs) versus pyrethroid-only LLINs with and without indoor residual spray against malaria infection: third year results of a cluster, randomised controlled, two-by-two factorial design trial in Tanzania. *Malar J* **22**, 294 (2023).
49. Mosha, J. F. *et al.* Effectiveness of long-lasting insecticidal nets with pyriproxyfen-pyrethroid, chlorfenapyr-pyrethroid, or piperonyl butoxide-pyrethroid versus pyrethroid only against malaria in Tanzania: final-year results of a four-arm, single-blind, cluster-randomised trial. *Lancet Infect Dis* **24**, 87–97 (2024).
50. Ngufor, C. *et al.* Community evaluation of VECTRON^{TM} T500, a broflanilide insecticide, for indoor residual spraying for malaria vector control in central Benin; a two arm non-inferiority cluster randomised trial. *Sci Rep* **13**, 17852 (2023).
51. Bath, D. *et al.* Effectiveness and cost-effectiveness of reactive, targeted indoor residual spraying for malaria control in low-transmission settings: a cluster-randomised, non-inferiority trial in South Africa. *Lancet* **397**, 816–827 (2021).
52. Wagman, J. M. *et al.* Reduced exposure to malaria vectors following indoor residual spraying of pirimiphos-methyl in a high-burden district of rural Mozambique with high ownership of long-lasting insecticidal nets: entomological surveillance results from a cluster-randomized trial. *Malar J* **20**, 54 (2021).
53. Yewhalaw, D. *et al.* An experimental hut study evaluating the impact of pyrethroid-only and PBO nets alone and in combination with pirimiphos-methyl-based IRS in Ethiopia. *Malaria Journal* **21**, 238 (2022).

54. Kenea, O. *et al.* Impact of combining indoor residual spraying and long-lasting insecticidal nets on Anopheles arabiensis in Ethiopia: results from a cluster randomized controlled trial. *Malaria Journal* **18**, 182 (2019).
55. Pryce, J., Richardson, M. & Lengeler, C. Insecticide-treated nets for preventing malaria. *Cochrane Database of Systematic Reviews* (2018) doi:10.1002/14651858.CD000363.pub3.
56. J, L., N, C. & L, P. *How Insecticide-Treated Nets (ITNs) Work: The Biological Mechanisms by Which ITNs Give Personal- and Community-Level Protection against Malaria*. https://zenodo.org/records/6393253 (2022) doi:10.5281/zenodo.6393253.
57. Pryce, J., Medley, N. & Choi, L. Indoor residual spraying for preventing malaria in communities using insecticide-treated nets. *Cochrane Database of Systematic Reviews* (2022) doi:10.1002/14651858.CD012688.pub3.
58. Bhatt, S. *et al.* The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. *Nature* **526**, 207 (2015).
59. Russell, T. L., Beebe, N. W., Cooper, R. D., Lobo, N. F. & Burkot, T. R. Successful malaria elimination strategies require interventions that target changing vector behaviours. *Malaria Journal* **12**, 56 (2013).
60. Strode, C., Donegan, S., Garner, P., Enayati, A. A. & Hemingway, J. The Impact of Pyrethroid Resistance on the Efficacy of Insecticide-Treated Bed Nets against African Anopheline Mosquitoes: Systematic Review and Meta-Analysis. *PLOS Medicine* **11**, e1001619 (2014).
61. Lindsay, S. W., Thomas, M. B. & Kleinschmidt, I. Threats to the effectiveness of insecticide-treated bednets for malaria control: thinking beyond insecticide resistance. *The Lancet Global Health* **9**, e1325–e1331 (2021).
62. Gleave, K., Lissenden, N., Chaplin, M., Choi, L. & Ranson, H. Piperonyl butoxide (PBO) combined with pyrethroids in insecticide-treated nets to prevent malaria in Africa. *Cochrane Database Syst Rev* **5**, CD012776 (2021).

63. Clarkson, C. S. *et al.* The genetic architecture of target-site resistance to pyrethroid insecticides in the African malaria vectors Anopheles gambiae and Anopheles coluzzii. *Molecular Ecology* **30**, 5303–5317 (2021).
64. Malaria Threat Map, World Health Organization https://apps.who.int/malaria/maps/threats/.
65. Protopopoff, N. *et al.* Effectiveness of a long-lasting piperonyl butoxide-treated insecticidal net and indoor residual spray interventions, separately and together, against malaria transmitted by pyrethroid-resistant mosquitoes: a cluster, randomised controlled, two-by-two factorial design trial. *Lancet* **391**, 1577–1588 (2018).
66. Ketoh, G. K. *et al.* Efficacy of two PBO long lasting insecticidal nets against natural populations of Anopheles gambiae s.l. in experimental huts, Kolokopé, Togo. *PLoS One* **13**, e0192492 (2018).
67. Accrombessi, M. *et al.* Efficacy of pyriproxyfen-pyrethroid long-lasting insecticidal nets (LLINs) and chlorfenapyr-pyrethroid LLINs compared with pyrethroid-only LLINs for malaria control in Benin: a cluster-randomised, superiority trial. *The Lancet* **401**, 435–446 (2023).
68. N'Guessan, R. *et al.* Chlorfenapyr: A pyrrole insecticide for the control of pyrethroid or DDT resistant *Anopheles gambiae* (Diptera: Culicidae) mosquitoes. *Acta Tropica* **102**, 69–78 (2007).
69. Grisales, N. *et al.* Pyriproxyfen-treated bed nets reduce reproductive fitness and longevity of pyrethroid-resistant Anopheles gambiae under laboratory and field conditions. *Malaria Journal* **20**, 273 (2021).
70. Yao, F. A. *et al.* Mark-release-recapture experiment in Burkina Faso demonstrates reduced fitness and dispersal of genetically-modified sterile malaria mosquitoes. *Nat Commun* **13**, 796 (2022).
71. Jeffries, C. L. *et al.* Evidence for natural hybridization and novel Wolbachia strain superinfections in the Anopheles gambiae complex from Guinea. *R Soc Open Sci* **8**, 202032.
72. Walker, T. *et al.* Stable high-density and maternally inherited Wolbachia infections in Anopheles moucheti and Anopheles demeilloni mosquitoes. *Current Biology* **31**, 2310-2320.e5 (2021).

73. World Health Organization. Full evidence report on the rtss as01 malaria vaccine for sage mpag (sept2021).pdf. https://cdn.who.int/media/docs/default-source/immunization/mvip/full-evidence-report-on-the-rtss-as01-malaria-vaccine-for-sage-mpag-(sept2021).pdf.
74. Moorthy, V., Hamel, M. J. & Smith, P. G. Malaria vaccines for children: and now there are two. *The Lancet* **0**, (2024).
75. Datoo, M. S. *et al.* Safety and efficacy of malaria vaccine candidate R21/Matrix-M in African children: a multicentre, double-blind, randomised, phase 3 trial. *The Lancet* **403**, 533–544 (2024).
76. WHO recommends R21/Matrix-M vaccine for malaria prevention in updated advice on immunization. https://www.who.int/news/item/02-10-2023-who-recommends-r21-matrix-m-vaccine-for-malaria-prevention-in-updated-advice-on-immunization.
77. Meshnick, S. R., Taylor, T. E. & Kamchonwongpaisan, S. Artemisinin and the antimalarial endoperoxides: from herbal remedy to targeted chemotherapy. *Microbiol Rev* **60**, 301–315 (1996).
78. Dyann F, W. & Pedro L, A. *Malaria Biology in the Era of Eradication*. *Malaria: Biology in the Era of Eradication* (2017).
79. Trape, J. F. The public health impact of chloroquine resistance in Africa. *American Journal of Tropical Medicine and Hygiene* **64**, 12–17 (2001).
80. Conrad Melissa D. *et al.* Evolution of Partial Resistance to Artemisinins in Malaria Parasites in Uganda. *New England Journal of Medicine* **389**, 722–732 (2023).
81. Conrad, M. D. & Rosenthal, P. J. Antimalarial drug resistance in Africa: the calm before the storm? *The Lancet. Infectious diseases* **19**, e338–e351 (2019).
82. Takala-Harrison, S. & Laufer, M. K. Antimalarial drug resistance in Africa: Key lessons for the future. *Annals of the New York Academy of Sciences* **1342**, 62–67 (2015).
83. Moss, S. *et al.* How has mass drug administration with dihydroartemisinin-piperaquine impacted molecular markers of drug resistance? A systematic review. *Malaria journal* **21**, 186 (2022).

84. World Health Organization. Strategy to respond to antimalarial drug resistance in Africa. (2022).
85. Fairhurst, R. M. & Dondorp, A. M. Artemisinin-Resistant Plasmodium falciparum Malaria. *Microbiology spectrum* **4**, (2016).
86. Uwimana, A. *et al.* Emergence and clonal expansion of in vitro artemisinin-resistant Plasmodium falciparum kelch13 R561H mutant parasites in Rwanda. *Nature medicine* (2020) doi:10.1038/s41591-020-1005-2.
87. Uwimana, A. *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* **21**, 1120–1128 (2021).
88. Straimer, J., Gandhi, P., Renner, K. C. & Schmitt, E. K. High Prevalence of Plasmodium falciparum K13 Mutations in Rwanda Is Associated With Slow Parasite Clearance After Treatment With Artemether-Lumefantrine. *J Infect Dis* **225**, 1411–1414 (2022).
89. Balikagala, B. *et al.* Evidence of Artemisinin-Resistant Malaria in Africa. *N Engl J Med* **385**, 1163–1171 (2021).
90. Asua, V. *et al.* Changing Prevalence of Potential Mediators of Aminoquinoline, Antifolate, and Artemisinin Resistance Across Uganda. *J Infect Dis* **223**, 985–994 (2021).
91. Mihreteab, S. *et al.* Increasing Artemisinin-Resistant HRP2-Negative Malaria in Eritrea. *N Engl J Med* **389**, 1191–1202 (2023).
92. Greenwood Brian. Artemisinin-Resistant and HRP-Negative Malaria Parasites in Africa. *New England Journal of Medicine* **389**, 1162–1164 (2023).
93. Ishengoma, D. S. *et al.* Evidence of artemisinin partial resistance in North-western Tanzania: clinical and drug resistance markers study. *medRxiv* 2024.01.31.24301954 (2024) doi:10.1101/2024.01.31.24301954.
94. Emiru, T. *et al.* Evidence for a role of Anopheles stephensi in the spread of drug- and diagnosis-resistant malaria in Africa. *Nat Med* **29**, 3203–3211 (2023).

95. Kahunu, G. M. *et al.* Identification of the PfK13 mutations R561H and P441L in the Democratic Republic of Congo. *International Journal of Infectious Diseases* **139**, 41–49 (2024).

96. Adam, M. *et al.* Antimalarial drug efficacy and resistance in malaria-endemic countries in HANMAT-PIAM_net countries of the Eastern Mediterranean Region 2016-2020: Clinical and genetic studies. *Trop Med Int Health* **28**, 817–829 (2023).

97. Nsanzabana, C., Djalle, D., Guérin, P. J., Ménard, D. & González, I. J. Tools for surveillance of anti-malarial drug resistance: An assessment of the current landscape. *Malaria Journal* **17**, 1–16 (2018).

98. Report on antimalarial drug efficacy, resistance and response: 10 years of surveillance (2010–2019). *Geneva: World Health Organization; 2020. Licence: CC BY-NC-SA 3.0 IGO*.

99. MalariaGEN *et al.* An open dataset of Plasmodium falciparum genome variation in 7,000 worldwide samples. *Wellcome Open Research* **6**, (2021).

100. Plowe, C. V. *et al.* World Antimalarial Resistance Network (WARN) III: Molecular markers for drug resistant malaria. *Malar J* **6**, 121 (2007).

101. Roper, C. *et al.* Molecular surveillance for artemisinin resistance in Africa. *The Lancet Infectious Diseases* **14**, 668–670 (2014).

102. Gardner, M. J. *et al. Genome Sequence of the Human Malaria Parasite Plasmodium Falciparum*. www.nature.com/nature (2002).

103. Oyola, S. O. *et al.* Efficient depletion of host DNA contamination in malaria clinical sequencing. *Journal of Clinical Microbiology* **51**, 745–751 (2013).

104. Nag, S. *et al.* High throughput resistance profiling of Plasmodium falciparum infections based on custom dual indexing and Illumina next generation sequencing-technology. *Scientific Reports* **7**, 1–13 (2017).

105. Jain, M., Olsen, H. E., Paten, B. & Akeson, M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biology* **17**, 239 (2016).

106. Ravi, R. K., Walton, K. & Khosroheidari, M. MiSeq: A Next Generation Sequencing Platform for Genomic Analysis. *Methods Mol Biol* **1706**, 223–232 (2018).

107. World Health Organization. Global Malaria Programme. Global plan for insecticide resistance management in malaria vectors. 131 (2012).

108. Global vector control response 2017–2030. Geneva: World Health Organization; 2017. Licence: CC BY-NC-SA 3.0 IGO.

109. World Health Organization. *Framework for a National Plan for Monitoring and Management of Insecticide Resistance in Malaria Vectors Framework for a National Plan for Monitoring and Management of Insecticide Resistance in Malaria Vectors Global Malaria Programme World Health Organization*. (2017).

110. World Health Organization. Global report on insecticide resistance in malaria vectors: 2010-2016 Global Malaria Programme.

111. Hemingway, J. The molecular basis of two contrasting metabolic mechanisms of insecticide resistance. *Insect Biochemistry and Molecular Biology* **30**, 1009–1015 (2000).

112. Soderlund, D. M. Molecular Mechanisms of Pyrethroid Insecticide Neurotoxicity: Recent Advances. *Arch Toxicol* **86**, 165–181 (2012).

113. Costa, L. G. The neurotoxicity of organochlorine and pyrethroid pesticides. *Handb Clin Neurol* **131**, 135–148 (2015).

114. Silva, A. P. B., Santos, J. M. M. & Martins, A. J. Mutations in the voltage-gated sodium channel gene of anophelines and their association with resistance to pyrethroids – a review. *Parasites & Vectors* **7**, 450 (2014).

115. Ranson, H. *et al.* Identification of a point mutation in the voltage-gated sodium channel gene of Kenyan Anopheles gambiae associated with resistance to DDT and pyrethroids. *Insect molecular biology* **9**, 491–497 (2000).

116. Martinez-Torres, D. *et al.* Molecular characterization of pyrethroid knockdown resistance (kdr) in the major malaria vector Anopheles gambiae s.s. *Insect Molecular Biology* **7**, 179–184 (1998).

117. Kabula, B. *et al.* Co-occurrence and distribution of East (L1014S) and West (L1014F) African knock-down resistance in Anopheles gambiae sensu lato population of Tanzania. *Trop Med Int Health* **19**, 1–11 (2014).

118. Jones, C. M. *et al.* Footprints of positive selection associated with a mutation (N1575Y) in the voltage-gated sodium channel of Anopheles gambiae. *Proc Natl Acad Sci U S A* **109**, 6614–6619 (2012).

119. Williams, J., Cowlishaw, R., Sanou, A., Ranson, H. & Grigoraki, L. In vivo functional validation of the V402L voltage gated sodium channel mutation in the malaria vector An. gambiae. *Pest Manag Sci* **78**, 1155–1163 (2022).

120. Collins, E. *et al.* The relationship between insecticide resistance, mosquito age and malaria prevalence in Anopheles gambiae s.l. from Guinea. *Scientific reports* **9**, 8846 (2019).

121. Du, W. *et al.* Independent mutations in the Rdl locus confer dieldrin resistance to Anopheles gambiae and An. arabiensis. *Insect Molecular Biology* **14**, 179–183 (2005).

122. Wondji, C. S. *et al.* Identification and distribution of a GABA receptor mutation conferring dieldrin resistance in the malaria vector Anopheles funestus in Africa. *Insect Biochem Mol Biol* **41**, 484–491 (2011).

123. Casida, J. E. & Durkin, K. A. Novel GABA receptor pesticide targets. *Pesticide Biochemistry and Physiology* **121**, 22–30 (2015).

124. Acford-Palmer, H. *et al.* Identification of two insecticide resistance markers in Ethiopian Anopheles stephensi mosquitoes using a multiplex amplicon sequencing assay. *Sci Rep* **13**, 5612 (2023).

125. Moss, S. *et al.* Genomic surveillance of Anopheles mosquitoes on the Bijagós Archipelago using custom targeted amplicon sequencing identifies mutations associated with insecticide resistance. *Parasites Vectors* **17**, 10 (2024).

126. Weill, M. *et al.* Insecticide resistance in mosquito vectors. *Nature* **423**, 136–137 (2003).

127. Fukuto, T. R. Mechanism of action of organophosphorus and carbamate insecticides. *Environ Health Perspect* **87**, 245–254 (1990).

128. Grau-Bové, X. *et al.* Resistance to pirimiphos-methyl in West African Anopheles is spreading via duplication and introgression of the Ace1 locus. *PLoS Genetics* **17**, (2021).

129. Weill, M. *et al.* The unique mutation in ace-1 giving high insecticide resistance is easily detectable in mosquito vectors. *Insect molecular biology* **13**, 1–7 (2004).

130. Greenblatt, H. M. *et al.* The complex of a bivalent derivative of galanthamine with torpedo acetylcholinesterase displays drastic deformation of the active-site gorge: implications for structure-based drug design. *J Am Chem Soc* **126**, 15405–15411 (2004).

131. Vontas, J., Katsavou, E. & Mavridis, K. Cytochrome P450-based metabolic insecticide resistance in Anopheles and Aedes mosquito vectors: Muddying the waters. *Pestic Biochem Physiol* **170**, 104666 (2020).

132. Scott, J. G. Cytochromes P450 and insecticide resistance. *Insect Biochemistry and Molecular Biology* **29**, 757–777 (1999).

133. Enayati, A. A., Ranson, H. & Hemingway, J. Insect glutathione transferases and insecticide resistance. *Insect Mol Biol* **14**, 3–8 (2005).

134. Ibrahim, S. S. *et al.* Pyrethroid Resistance in the Major Malaria Vector Anopheles funestus is Exacerbated by Overexpression and Overactivity of the P450 CYP6AA1 Across Africa. *Genes* **9**, 140 (2018).

135. Vontas, J. *et al.* Rapid selection of a pyrethroid metabolic enzyme CYP9K1 by operational malaria control activities. *Proceedings of the National Academy of Sciences* **115**, 4619–4624 (2018).

136. Hearn, J. *et al.* Multi-omics analysis identifies a CYP9K1 haplotype conferring pyrethroid resistance in the malaria vector Anopheles funestus in East Africa. *Mol Ecol* **31**, 3642–3657 (2022).

137. Mitchell, S. N. *et al.* Metabolic and Target-Site Mechanisms Combine to Confer Strong DDT Resistance in Anopheles gambiae. *PLOS ONE* **9**, e92662 (2014).

138. Lucas, E. R. *et al.* Whole-genome sequencing reveals high complexity of copy number variation at insecticide resistance loci in malaria mosquitoes. *Genome Res* **29**, 1250–1261 (2019).

139. Njoroge, H. *et al.* Identification of a rapidly-spreading triple mutant for high-level metabolic insecticide resistance in Anopheles gambiae provides a real-time molecular diagnostic for antimalarial intervention deployment. *Molecular Ecology* **31**, 4307–4318 (2022).

140. Ingham, V. & Nagi, S. Genomic Profiling of Insecticide Resistance in Malaria Vectors: Insights into Molecular Mechanisms. *Res Sq* rs.3.rs-3910702 (2024) doi:10.21203/rs.3.rs-3910702/v1.

141. Meiwald, A. *et al.* Association of Reduced Long-Lasting Insecticidal Net Efficacy and Pyrethroid Insecticide Resistance With Overexpression of CYP6P4, CYP6P3, and CYP6Z1 in Populations of Anopheles coluzzii From Southeast Côte d'Ivoire. *J Infect Dis* **225**, 1424–1434 (2022).

142. Weetman, D. *et al.* Candidate-gene based GWAS identifies reproducible DNA markers for metabolic pyrethroid resistance from standing genetic variation in East African Anopheles gambiae. *Sci Rep* **8**, 2920 (2018).

143. Riveron, J. M. *et al.* A single mutation in the GSTe2 gene allows tracking of metabolically based insecticide resistance in a major malaria vector. *Genome biology* **15**, (2014).

144. Ingham, V. A., Pignatelli, P., Moore, J. D., Wagstaff, S. & Ranson, H. The transcription factor Maf-S regulates metabolic resistance to insecticides in the malaria vector Anopheles gambiae. *BMC Genomics* **18**, 669 (2017).

145. Ingham, V. A., Wagstaff, S. & Ranson, H. Transcriptomic meta-signatures identified in Anopheles gambiae populations reveal previously undetected insecticide resistance mechanisms. *Nat Commun* **9**, 5282 (2018).

146. World Health Organization. Manual for monitoring insecticide resistance in mosquito vectors and selecting appropriate interventions, 2022.

147. Campos, M. *et al.* High-throughput barcoding method for the genetic surveillance of insecticide resistance and species identification in Anopheles gambiae complex malaria vectors. *Scientific Reports* **12**, 13893 (2022).

148. Lucas, E. R. *et al.* A high throughput multi-locus insecticide resistance marker panel for tracking resistance emergence and spread in Anopheles gambiae. *Sci Rep* **9**, 13335 (2019).

149. Messenger, L. A. *et al.* A whole transcriptomic approach provides novel insights into the molecular basis of organophosphate and pyrethroid resistance in Anopheles arabiensis from Ethiopia. *Insect Biochemistry and Molecular Biology* **139**, 103655 (2021).

150. Genetic diversity of the African malaria vector Anopheles gambiae. *Nature* **552**, 96–100 (2017).

151. Lucas, E. R. *et al.* Genome-wide association studies reveal novel loci associated with pyrethroid and organophosphate resistance in Anopheles gambiae and Anopheles coluzzii. *Nat Commun* **14**, 4946 (2023).

152. Anopheles gambiae 1000 Genomes Consortium. Genome variation and population structure among 1142 mosquitoes of the African malaria vector species Anopheles gambiae and Anopheles coluzzii. *Genome Res* **30**, 1533–1546 (2020).

153. Holt, R. A. *et al.* The genome sequence of the malaria mosquito Anopheles gambiae. *Science* **298**, 129–149 (2002).

154. Hutchins, H. *et al.* Protocol for a cluster randomised placebo-controlled trial of adjunctive ivermectin mass drug administration for malaria control on the Bijagós Archipelago of Guinea-Bissau: the MATAMAL trial. *BMJ Open* **13**, e072347 (2023).

155. Ant, T. *et al.* A survey of Anopheles species composition and insecticide resistance on the island of Bubaque , Bijagos Archipelago , Guinea - Bissau. *Malaria Journal* 1–9 (2020) doi:10.1186/s12936-020-3115-1.

156. Harry Hutchins *et al.* UNDER REVIEW: A cluster randomised placebo-controlled trial of adjunctive ivermectin mass drug administration for malaria control on the Bijagos Archipelago of Guinea-Bissau: The MATAMAL trial.

157. McGregor, D. *et al.* The epidemiology of plasmodium falciparum malaria in the bijagos islands of guinea-bissau. *American Journal of Tropical Medicine and Hygiene* **104**, 2117–2122 (2021).

158. Pretorius Elizabeth. UNDER REVIEW - A survey of indoor and outdoor biting behaviour, species composition and circumsporozoite rate of malaria vectors in the Bijagós Archipelago, Guinea-Bissau.

159. Bryan, J. H., Petrarca, V., Di Deco, M. A. & Coluzzi, M. Adult behaviour of members of the Anopheles gambiae complex in the Gambia with special reference to An. melas and its chromosomal variants. *Parassitologia* **29**, 221–249 (1987).

160. Sy, O. *et al.* Residual malaria transmission and the role of Anopheles arabiensis and Anopheles melas in central Senegal. *Journal of Medical Entomology* **60**, 546–553 (2023).

161. Overgaard, H. J. *et al.* Malaria transmission after five years of vector control on Bioko Island, Equatorial Guinea. *Parasites & Vectors* **5**, 253 (2012).

162. Barrón, M. G. *et al.* A new species in the major malaria vector complex sheds light on reticulated species evolution. *Sci Rep* **9**, 14753 (2019).

163. Oliveira, E. *et al.* High levels of hybridization between molecular forms of Anopheles gambiae from Guinea Bissau. *Journal of Medical Entomology* **45**, 1057–1063 (2008).

164. Marsden, C. D. *et al.* Asymmetric introgression between the M and S forms of the malaria vector, Anopheles gambiae, maintains divergence despite extensive hybridization. *Mol Ecol* **20**, 4983–4994 (2011).

165. Vicente, J. L. *et al.* Massive introgression drives species radiation at the range limit of Anopheles gambiae. *Sci Rep* **7**, 46451 (2017).

166. Torre, A. D. *et al.* Speciation within the Anopheles gambiae complex: high-throughput whole genome sequencing reveals evidence of a putative new cryptic taxon in 'far-west' Africa. Preprint at https://doi.org/10.21203/rs.3.rs-3914444/v1 (2024).

167. Hutchins, H. *et al.* A survey of knowledge , attitudes and practices regarding malaria and bed nets on Bubaque Island , Guinea - Bissau. *Malaria Journal* 1–15 (2020) doi:10.1186/s12936-020-03469-1.

168. Ursing, J., Kofoed, P. E., Rodrigues, A., Bergqvist, Y. & Rombo, L. Chloroquine is grossly overdosed and overused but well tolerated in Guinea-Bissau. *Antimicrobial Agents and Chemotherapy* **53**, 180–185 (2009).

169. Ursing, J. *et al.* High-Dose Chloroquine for Uncomplicated Plasmodium falciparum Malaria Is Well Tolerated and Causes Similar QT Interval Prolongation as Standard-Dose Chloroquine in Children. *Antimicrobial agents and chemotherapy* **64**, 1–12 (2020).

170. Jovel, I. T., Kofoed, P. E., Rombo, L., Rodrigues, A. & Ursing, J. Temporal and seasonal changes of genetic polymorphisms associated with altered drug susceptibility to chloroquine, lumefantrine, and quinine in Guinea-Bissau between 2003 and 2012. *Antimicrobial Agents and Chemotherapy* **59**, 872–879 (2015).

171. Nag, S. *et al.* Proof of concept: used malaria rapid diagnostic tests applied for parallel sequencing for surveillance of molecular markers of anti-malarial resistance in Bissau, Guinea-Bissau during 2014-2017. *Malaria journal* **18**, 252 (2019).

172. Silva, R., Mavridis, K., Vontas, J., Rodrigues, A. & Costa, H. Monitoring and molecular profiling of contemporary insecticide resistance status of malaria vectors in Guinea – Bissau. *Acta Tropica* **206**, 105440 (2020).

173. Hofmann, N. *et al.* Ultra-Sensitive Detection of Plasmodium falciparum by Amplification of Multi-Copy Subtelomeric Targets. *PLoS medicine* **12**, e1001788 (2015).

174. Ursing, J. *et al.* Chloroquine-susceptible and -resistant Plasmodium falciparum strains survive high chloroquine concentrations by becoming dormant but are eliminated by prolonged exposure. *J Antimicrob Chemother* **77**, 1005–1011 (2022).

175. Ursing, J. *et al.* Similar Efficacy and Tolerability of Double-Dose Chloroquine and Artemether-Lumefantrine for Treatment of Plasmodium falciparum Infection in Guinea-Bissau : A Randomized Trial. **203**, 109–116 (2011).

176. Kofoed, P.-E. *et al.* Different doses of amodiaquine and chloroquine for treatment of uncomplicated malaria in children in Guinea-Bissau: implications for future treatment recommendations. *Transactions of The Royal Society of Tropical Medicine and Hygiene* **101**, 231–238 (2007).

177. Ursing, J., Kofoed, P.-E., Rodrigues, A. & Rombo, L. No Seasonal Accumulation of Resistant P. falciparum when High-Dose Chloroquine Is Used. *PLoS One* **4**, e6866 (2009).

178. Kofoed, P.-E. *et al.* Genetic markers of resistance to pyrimethamine and sulfonamides in Plasmodium falciparum parasites compared with the resistance patterns in isolates of Escherichia coli from the same children in Guinea-Bissau. *Tropical Medicine & International Health* **9**, 171–177 (2004).

179. Sisowath, C. *et al.* In vivo selection of Plasmodium falciparum pfmdr1 86N coding alleles by Artemether-Lumefantrine (Coartem). *Journal of Infectious Diseases* **191**, 1014–1017 (2005).

180. Baraka, V. *et al.* In Vivo Selection of Plasmodium falciparum Pfcrt and Pfmdr1 Variants by Artemether-Lumefantrine and Dihydroartemisinin- Piperaquine in Burkina Faso. **59**, 734–737 (2015).

181. Humphreys, G. S. *et al.* Amodiaquine and Artemether-Lumefantrine Select Distinct Alleles of the Plasmodium falciparum mdr1 Gene in Tanzanian Children Treated for Uncomplicated Malaria. **51**, 991–997 (2007).

182. Ursing, J., Rombo, L., Rodrigues, A. & Kofoed, P. Artemether-Lumefantrine versus Dihydroartemisinin-Piperaquine for Treatment of Uncomplicated Plasmodium falciparum

Malaria in Children Aged Less than 15 Years in Guinea-Bissau – An Open- Label Non-Inferiority Randomised Clinical. **22**, 1–11 (2016).

183. Deitz, K. C. *et al.* Genome-Wide Divergence in the West-African Malaria Vector Anopheles melas. *G3 (Bethesda)* **6**, 2867–2879 (2016).

CHAPTER TWO

Thesis Structure, Hypotheses and Objectives

Structure and Hypotheses

In this thesis I use genomic techniques to ask, what is the genetic diversity and antimalarial resistance status of the *P. falciparum* parasite, and what is the genetic diversity and insecticide resistance status of *An. gambiae s.l.* vectors, on the Bijagós Archipelago of Guinea-Bissau? I ask these questions in the context of malaria control interventions on this remote Archipelago, where the genomic epidemiology of malaria has not previously been studied. This thesis provides data on the genetic diversity of *P. falciparum* and *An. gambiae s.l.,* in addition to antimalarial and insecticide resistance data for the first time in this region. In doing so, this work presents a baseline of data relevant to the NMCP and future researchers for optimising malaria control interventions, which can be built upon during future genomic surveillance and malaria control efforts.

Chapter 3 (Published paper)

In **Chapter 3,** I use dried blood spots (DBS) collected as part of a cross-sectional malaria prevalence survey which was conducted on five islands in the Bijagós Archipelago during 2018. These DBS samples were collected from asymptomatic individuals, and *P. falciparum* positivity was detected using 18S rRNA qPCR. In this chapter, I identify population dynamics and drug resistance mutations in *P. falciparum* within these DBS samples. I hypothesised that antimalarial resistance mutations associated with resistance to AL, CQ, and SP would be present due to the current or previous use of these drugs in this region. I use selective whole-genome amplification to enable the sequencing of parasite data from very low parasitaemia, asymptomatic infections. I generate a baseline drug-resistance profile which paves the way for future research into the impact of MDA on molecular markers of antimalarial resistance. Furthermore, this contributes an open-source WGS resource, as the first generation of WGS data of *P. falciparum* from the Bijagós Archipelago. I identified fixation of the *pfdhfr* mutations N51I and S108N, associated with resistance to sulphadoxine-pyrimethamine, and the continued presence of *pfcrt* K76T, associated with chloroquine resistance. I further show that *P. falciparum* from the Archipelago cluster with samples from mainland West Africa without forming a separate phylogenetic group.

Chapter 4 (Published paper)

In **Chapter 4**, I expand focus to the malaria vector *Anopheles gambiae.* I investigate the prevalence of molecular markers associated with insecticide resistance in *An. gambiae s.l.* mosquitoes from 13 islands across the Bijagós. These mosquitoes were collected using CDC light traps in October and November 2019. I hypothesised that molecular markers associated with resistance to pyrethroids would be widespread on the Archipelago due to the use of pyrethroid ITNs on the islands and the presence of pyrethroid resistance on mainland Guinea-Bissau[172,64]. To enable a larger sample size, I use multiplex amplicon sequencing, a more rapid and cost-efficient method of gene sequencing. The mainstay for malaria control on the Bijagós is ITNs impregnated with the pyrethroid deltamethrin, and the prevalence of molecular markers of resistance to pyrethroids on the islands is previously understudied. Using rapid multiplex amplicon sequencing, I identify that across the 13 islands sampled, the prevalence of insecticide resistance mutations was low. However, I find four mutations previously associated with insecticide resistance at low prevalence and an additional eight non-synonymous mutations which were previously unknown in Bijagós *An. gambiae s.l.* mosquitoes. Notably, I identify the presence of molecular markers of resistance to deltamethrin, which is important as the mainstay of vector control on the islands is the use of pyrethroid ITNs.

Chapter 5 (Published paper)

In **Chapter 5,** I build on my surveillance of molecular markers in **Chapter 4** and conduct phenotypic bioassays during fieldwork to investigate whether *Anopheles* mosquitoes on this isolated Archipelago show phenotypic resistance to deltamethrin. I hypothesised that deltamethrin resistance would be present on the islands due to the spread of pyrethroid resistance in West Africa and the use of pyrethroid ITNs on the islands. I conclude that deltamethrin resistance is present on the main island of Bubaque, and then use whole genome sequencing to search for signatures of selection among the population of mosquitoes tested. I identify six SNPs previously associated with, or putatively associated with, insecticide resistance, and twenty additional non-synonymous SNPs in insecticide-resistance associated genes. I use genome wide selection scans using Garud's H_{12} statistic and identify two selective sweeps in the *An. gambiae s.s.* population. Both selective sweeps overlap with metabolic genes previously associated with insecticide resistance, including *cyp9k1* and the *cyp6aa/cyp6p* gene cluster. This chapter presents the first phenotypic testing for deltamethrin resistance and the first WGS data for *Anopheles* mosquitoes from the Bijagós Archipelago, contributing data of significance for vector control policy in the region.

Chapter 6 (Published paper)

In **Chapter 6**, I analyse the insecticide resistance status and genetic diversity of the malaria vector *An. melas,* which may have a role in the transmission of malaria on the Bijagós Archipelago, particularly during the dry season. I hypothesise that *An. melas* harbour low frequencies of insecticide resistance mutations common to other species in the *An. gambiae s.l.* complex, and that we may detect genetic separation between *An. melas* from different islands. I find complete absence of insecticide resistance SNPs common to *An. gambiae s.l.,* and no evidence of selective sweeps across insecticide resistance genes. Furthermore, I detect no genetic separation between *An. melas* from different islands, but clustering of this species into two groups which are independent of island origin. This chapter provides the first WGS analysis of *An. melas* from the Bijagós Archipelago, and the first population genetic analysis for this mosquito species from individual mosquito WGS data. This provides novel insights into insecticide resistance genetics within this species, including the identification of 26 novel SNPs in the *An. melas* population which may be associated with insecticide resistance. *An. melas* is part of the *An. gambiae s.l.* complex, which is a model for speciation under intense gene flow[183]. Understanding the genetic diversity and population structure of this species aids our understanding of how this malaria vector is evolving, and knowledge of insecticide resistance genomics can be used to help to inform future vector control interventions.

Chapter 7: Discussion and conclusions

In the **final chapter**, I discuss the potential incorporation of genome surveillance into malaria control programmes.

Thesis Objectives

1. To generate and analyse WGS data from DBS samples collected prior to implementation of MDA with DHAP on the Bijagós, to investigate population dynamics and molecular markers of drug resistance in *P. falciparum* infections.
2. To use multiplex amplicon sequencing to identify molecular markers associated with insecticide resistance in *An. gambiae s.l.* complex mosquitoes collected during the peak malaria transmission season in October/November 2019.
3. To conduct phenotypic testing of deltamethrin susceptibility in *An. gambiae s.l.* complex mosquitoes on the main island of Bubaque.
4. To generate and analyse WGS data from deltamethrin susceptible and resistant *An. gambiae s.s.* mosquitoes, to investigate genetic diversity and signatures of selection.
5. To generate and analyse WGS data for *An. melas* mosquitoes from across the Archipelago, to investigate the genetic diversity, insecticide resistance status and signatures of selection within this population.

***Table 4**: Published and submitted manuscripts included in this thesis.*

Thesis Chapter	Manuscript Title	Journal	Authors	Status
Chapter 3	Population dynamics and drug resistance mutations in *Plasmodium falciparum* on the Bijagós Archipelago, Guinea-Bissau. *Sci Rep* **13**, 6311 (2023). https://doi.org/10.1038/s41598-023-33176-1	Scientific Reports	**Moss S**, Mańko E, Vasileva H, Da Silva ET, Goncalves A, Osborne A, Phelan J, Rodrigues A, Djata P, D'Alessandro U, Mabey D, Krishna S, Last A, Clark TG, Campino S.	Published 18th April 2023
Chapter 4	Genomic surveillance of Anopheles mosquitoes on the Bijagós Archipelago using custom targeted amplicon sequencing identifies mutations associated with insecticide resistance. Parasit Vectors. 2024 Jan 4;17(1):10. doi: 10.1186/s13071-023-06085-5. PMID: 38178249; PMCID: PMC10768400.	Parasites & Vectors	**Moss S**, Pretorius E, Ceesay S, Hutchins H, Da Silva ET, Ndiath MO, Jones RT, Vasileva H, Phelan J, Acford-Palmer H, Collins E, Rodrigues A, Krishna S, Clark TG, Last A, Campino S.	Published 4th January 2024
Chapter 5	Phenotypic evidence of deltamethrin resistance and identification of selective sweeps in Anopheles mosquitoes on the Bijagós Archipelago, Guinea-Bissau.	Scientific Reports	**Moss S**, Jones R, Pretorius E, Da Silva ET, Higgins M, Kristan M, Acford-Palmer H, Collins LE, Rodrigues A, Krishna S, Clark TG, Last A, Campino S.	Published 1st October 2024
Chapter 6	Whole genome sequence analysis of population dynamics and insecticide resistance markers in *Anopheles melas* from the Bijagós Archipelago, Guinea-Bissau	Parasites & Vectors	**Moss S**, Pretorius E, Ceesay S, Da Silva ET, Hutchins H, Ndiath MO, Acford-Palmer H, Collins LE, Higgins M, Phelan J, Jones RT, Vasileva H, Rodrigues A, Krishna S, Clark TG, Last A, Campino S	Published 18th September 2024

CHAPTER THREE

Published Paper: Population dynamics and drug resistance mutations in *Plasmodium falciparum* on the Bijagós Archipelago, Guinea-Bissau.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	lsh2005123	**Title**	Miss
First Name(s)	Sophie		
Surname/Family Name	Moss		
Thesis Title	Genomic epidemiology of *Plasmodium falciparum* and its *Anopheline* vectors in the context of malaria control on the Bijagós Archipelago of Guinea-Bissau		
Primary Supervisor	Dr Anna Last		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	Scientific Reports		
When was the work published?	18th April 2023		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	N/A		
Have you retained the copyright for the work?*	**Yes**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	

Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I devised the research question and designed the study. I conducted all *Plasmodium* DNA extraction and SWGA amplification, preparation for sequencing and all bioinformatic analyses. I wrote the manuscript, which was commented on by my co-authors.

SECTION E

Student Signature	[redacted]
Date	2nd April 2024

Supervisor Signature	Anna Last Digitally signed by Anna Last [redacted] Date: 2024.05.22 18:22:23 +01'00'
Date	22nd May 2024

scientific reports

Check for updates

OPEN

Population dynamics and drug resistance mutations in *Plasmodium falciparum* on the Bijagós Archipelago, Guinea-Bissau

Sophie Moss[1]✉, Emilia Mańko[1], Hristina Vasileva[1], Eunice Teixeira Da Silva[1,2], Adriana Goncalves[1], Ashley Osborne[1], Jody Phelan[1], Amabelia Rodrigues[3], Paulo Djata[4], Umberto D'Alessandro[5], David Mabey[1], Sanjeev Krishna[6,7,8], Anna Last[1,10], Taane G. Clark[1,9,10] & Susana Campino[1,10]✉

Following integrated malaria control interventions, malaria burden on the Bijagós Archipelago has significantly decreased. Understanding the genomic diversity of circulating *Plasmodium falciparum* malaria parasites can assist infection control, through identifying drug resistance mutations and characterising the complexity of population structure. This study presents the first whole genome sequence data for *P. falciparum* isolates from the Bijagós Archipelago. Amplified DNA from *P. falciparum* isolates sourced from dried blood spot samples of 15 asymptomatic malaria cases were sequenced. Using 1.3 million SNPs characterised across 795 African *P. falciparum* isolates, population structure analyses revealed that isolates from the archipelago cluster with samples from mainland West Africa and appear closely related to mainland populations; without forming a separate phylogenetic cluster. This study characterises SNPs associated with antimalarial drug resistance on the archipelago. We observed fixation of the PfDHFR mutations N51I and S108N, associated with resistance to sulphadoxine-pyrimethamine, and the continued presence of PfCRT K76T, associated with chloroquine resistance. These data have relevance for infection control and drug resistance surveillance; particularly considering expected increases in antimalarial drug use following updated WHO recommendations, and the recent implementation of seasonal malaria chemoprevention and mass drug administration in the region.

Despite knowing the causes of malaria for over a century, this disease is still a major public health problem. The World Health Organization (WHO) Global Technical Strategy aims to reduce the global malaria burden by 90% by 2030[1]. Considerable progress has been made. However, the decline in malaria mortality has slowed since 2014, and malaria deaths increased to an estimated 627,000 in 2020. This was a 12% rise from 2019, with most deaths caused by the parasite *Plasmodium falciparum*[1]. To meet the WHO target by 2030, novel strategies and tools for malaria elimination are needed.

Recently updated WHO recommendations suggest that antimalarials will be increasingly deployed for the control and elimination of *P. falciparum,* on a mass scale. These recommendations included expanding the use of intermittent preventive treatment in pregnancy with sulphadoxine-pyrimethamine (IPTp-SP) to all pregnant

[1]Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, UK. [2]Ministry of Public Health, Bissau, Guinea-Bissau. [3]Bandim Health Project, Bissau, Guinea-Bissau. [4]National Malaria Control Programme, Ministry of Public Health, Bissau, Guinea-Bissau. [5]MRC the Gambia Unit, Fajara, The Gambia. [6]Clinical Academic Group, Institute for Infection and Immunity, and St. George's University Hospitals NHS Foundation Trust, St. George's University of London, London, UK. [7]Centre de Recherches Médicales de Lambaréné (CERMEL), Lambaréné, Gabon. [8]Institut Für Tropenmedizin Universitätsklinikum Tübingen, Tübingen, Germany. [9]Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK. [10]These authors contributed equally: Anna Last, Taane G. Clark and Susana Campino. ✉email: sophie.moss@lshtm.ac.uk; susana.campino@lshtm.ac.uk

women, broader use of perennial malaria chemoprevention (PMC) and seasonal malaria chemoprevention (SMC), and expanded guidance for the use of mass drug administration (MDA)[2]. As a result of these recommendations, the mass use of antimalarials is likely to increase in a variety of transmission settings at the discretion of relevant control programmes. Molecular surveillance is a useful tool for evaluating the impact of mass antimalarial drug use through monitoring malaria prevalence, parasite population dynamics, and genomic markers of drug resistance.

Endemic island communities are a unique landscape within which to test malaria interventions, including MDA, due to their isolated geography. Understanding the population dynamics and genetic variation of *P. falciparum* populations on islands provides crucial information for the evaluation of these interventions, as well as routine malaria surveillance. Improved whole genome sequencing technology, including the development of Next Generation Sequencing (NGS), has dramatically reduced the time and cost of genomic sequencing through enabling the sequencing of millions of reads in parallel[3]. The subsequent increase in the generation and availability of genome sequence data allows for high resolution analysis of the genomic architecture of *P. falciparum* populations.

The Bijagós Archipelago is a group of 88 islands and islets situated off the coast of Guinea-Bissau, West Africa. The islands are isolated from mainland Guinea-Bissau; separated by 70 km of sea and accessible by ferry, which takes four hours and sails once each week. Currently, malaria interventions in the Bijagós include enhanced case finding, intermittent preventive treatment in pregnancy (IPTp) and long-lasting insecticidal nets (LLINs)[4]. Despite good coverage and adherence to these interventions, including maximal control efforts using LLINs, malaria case incidence and mortality both increased by between 5 and 25%, from 2015 to 2020[1]. Although *P. falciparum* populations have been studied in mainland Guinea-Bissau, there has been no previous research conducted on the genomic architecture of *P. falciparum* parasites on the Bijagós Archipelago. This is vital for researchers to be able to understand the impact of malaria interventions within the region. Eighteen of the islands within the archipelago are inhabited year-round, with a total population of around 25,000 people[5]. The majority of the population live in forest villages and are largely hunter gatherers with some subsistence farming and fishing[4]. The islands have a rainy season between June and November and a dry season from December to May, with peak malaria transmission occurring at the end of the rainy season in October and November[6]. Malaria prevalence on the Bijagós is seasonal and stable, with low to moderate transmission intensity. Most malaria cases are asymptomatic. A malaria prevalence survey was conducted on the most populated island within the archipelago in August 2017, during the start of the malaria transmission season. This was a cross-sectional survey of the population, which identified the prevalence of *P. falciparum* infection to be 16.9% by qPCR[5]. Persistent circulation of subclinical parasitaemia makes reducing malaria transmission very difficult in this setting. This is because asymptomatic malaria cases are a major driver of malaria transmission, and asymptomatic individuals do not generally seek treatment[5].

One of the main challenges of whole genome sequencing (WGS) clinical malaria parasite isolates from dried blood spots is obtaining sufficient high-quality parasite DNA from infected individuals. This is particularly challenging when there is low parasitaemia in asymptomatic infections, and when working with dried blood spot (DBS) clinical samples, as the majority of DNA extracted from DBS is human. To address this, selective whole genome amplification (SWGA) can be used to selectively amplify *P. falciparum* DNA directly from DBS samples. This can result in better sequencing coverage of low-density *P. falciparum* infections. This higher sequencing coverage is necessary for robust population genomic analyses[7]. Using the SWGA method, we generated 15 whole genome sequences of *P. falciparum* from 15 dried blood spot samples. These DBS samples were collected from asymptomatic individuals of all ages during a cross sectional malaria prevalence survey conducted on five islands on the archipelago in 2018. This is the first WGS data from *P. falciparum* parasites on these islands. These data provide insights into the genetic variation and population dynamics of the *P. falciparum* parasites, including genomic markers associated with antimalarial drug resistance. The genomic variation within the Bijagós Archipelago is contextualised with approximately 800 isolates from other regions in Africa, using publicly available WGS data[8]. Importantly, these data provide a baseline assessment of genomic variation in 2018, prior to the subsequent implementation of SMC and MDA programmes in the region.

Results

Whole genome sequence data. The filtering steps for *P. falciparum* isolates for inclusion in genomic analyses can be seen in Supplementary Fig. 4. DNA was extracted from 2500 dried blood spots. qPCR was then used to test for the presence of *P. falciparum* in the extracted DNA. This identified 145 samples which were positive for *P. falciparum* with a CT value < 30. *P. falciparum* DNA was successfully amplified using Selective Whole Genome Amplification to sufficient levels for whole genome sequencing in 48 samples. Downstream filtering for high-quality whole genome sequence data resulted in 15 samples being carried forwards for bioinformatic analysis (please refer to the Methods section for additional detail).

A pairwise nucleotide matrix of 1,326,629 filtered SNPs was generated using whole genome sequences from 795 *P. falciparum* isolates, collected from 10 countries across Africa (Supplementary Table 1). This included isolates from Central Africa (Gabon, n = 57), West Africa (Ivory Coast n = 70, The Gambia n = 148, Bijagós Archipelago Guinea-Bissau (n = 15), East Africa (Kenya n = 138, Uganda n = 17), Southeast Africa (Malawi n = 149, Madagascar n = 24), South Central Africa (DRC n = 150) and Horn of Africa (Eritrea n = 27). The F_{WS} metric was used to investigate within-sample diversity. Most samples appeared monoclonal, with "high" F_{WS} estimates of ≥ 0.95[9] (Supplementary Table 2). The samples from the Bijagós had a mean F_{WS} of 0.965, with 86.7% monoclonality ($F_{WS} \geq 0.95$), which was higher than isolates from other countries, including The Gambia and Ivory Coast, which exhibited 67.6% and 61.4% monoclonality respectively. This high F_{WS} score indicates low multiplicity of infection within samples from the Bijagós. This may be due to low transmission intensity overall, or be because

people generally live in small, more isolated communities, and supports the assumption that our population analyses are robust.

***P. falciparum* isolates from the Bijagós Archipelago cluster within West Africa.** Principal component analysis (PCA) was conducted using the previously described pairwise genetic distance matrices to investigate the population structure of the *P. falciparum* isolates. Isolates from the Bijagós Archipelago were found to cluster with samples from West Africa when compared to other regions of Africa (Fig. 1). A SNP based Neighbour-Joining tree was generated to further investigate where the Bijagós *P. falciparum* isolates were situated phylogeographically. The samples from the Bijagós (labelled as Guinea-Bissau in purple), clustered with other West African countries. The samples from the Bijagós, The Gambia and Ivory Coast appear to be mixed

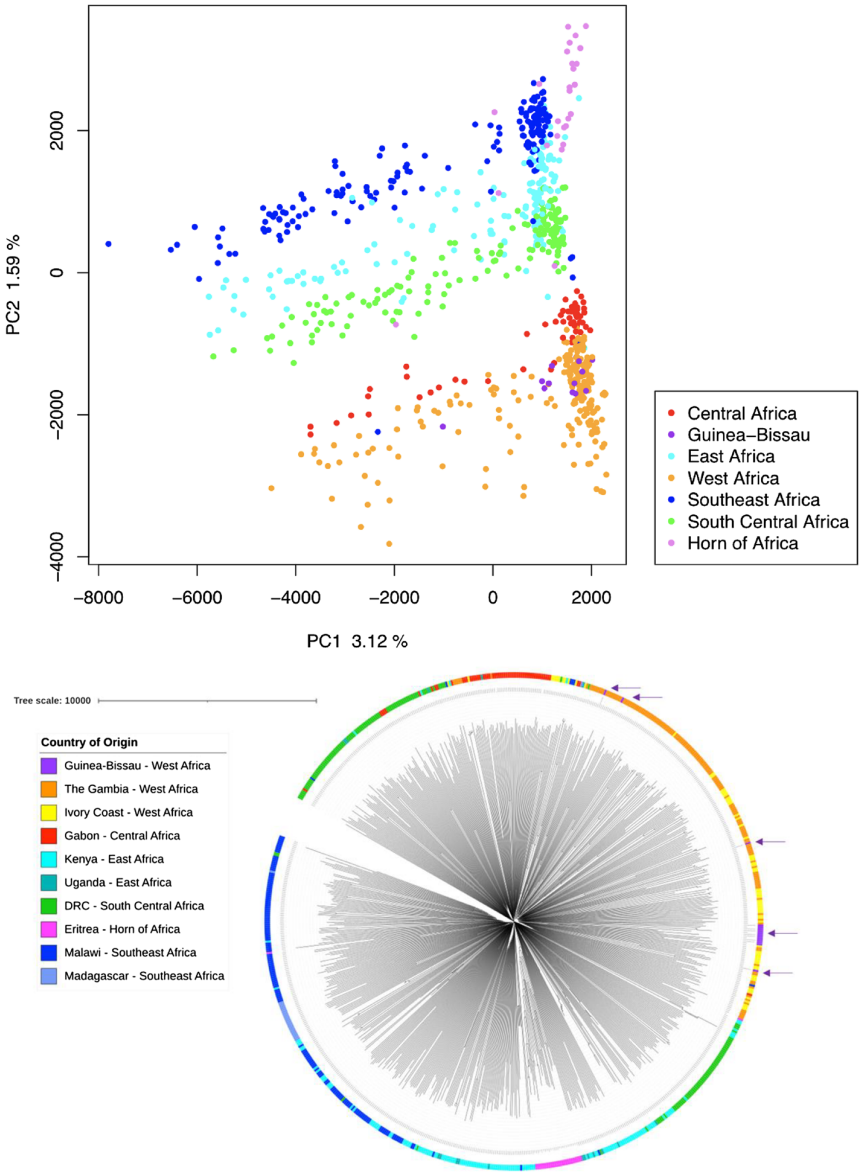


Figure 1. Population structure analysis: (top) Principal component analysis demonstrating that the isolates from the Bijagós (in purple) cluster with other isolates from West Africa (in orange and yellow). Generated using pairwise genetic distance matrices containing 1,326,629 SNPs from 795 isolates. (bottom) Neighbour-Joining tree for the samples and SNPs described above. Isolates were sampled from Central Africa (Gabon, n = 57), West Africa (Ivory Coast n = 70, The Gambia n = 148, Bijagós Archipelago Guinea-Bissau (n = 15), East Africa (Kenya n = 138, Uganda n = 17), Southeast Africa (Malawi n = 149, Madagascar n = 24), South Central Africa (DRC n = 150) and Horn of Africa (Eritrea n = 27). Bijagós isolates (indicated with purple arrows) are found interspersed with samples from West Africa (orange and yellow).

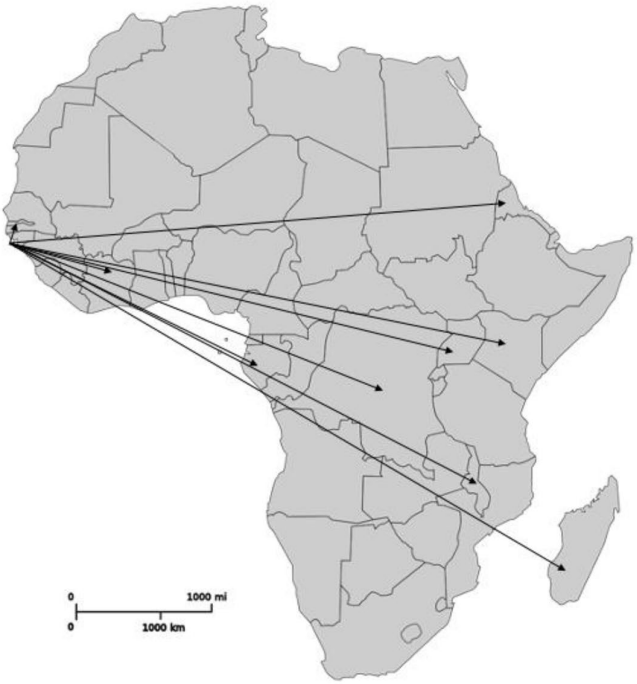
and closely related, without forming distinct clusters. This suggests a high degree of mixing between parasites from the Bijagós with those from neighbouring West African countries, reflecting a high degree of regional transmission of *P. falciparum* across West African borders, which has been previously reported[10].

Population differentiation. Population differentiation due to genetic structure was measured using Fixation index (FST) statistics[11], where SNPs with values of 1 indicate perfect differentiation between populations. SNPs that differentiated the Bijagós Archipelago from other African countries or regions were identified. SNPs with a threshold of $F_{ST} > 0.75$, $F_{ST} > 0.85$, and $F_{ST} > 0.9$ were identified to investigate patterns of differentiation between isolates from the Bijagós and the rest of Africa.

The isolates from the Bijagós were compared separately to each of the other countries within the dataset. The number of SNPs with $F_{ST} > 0.75$ for each country comparison are described in Fig. 2. There was an expected trend towards greater SNP differences with increased geographical distance. The largest degree of population differentiation was found between the Bijagós and Eritrea, with 151 SNPs with $F_{ST} > 0.75$.

SNPs with an $F_{ST} > 0.85$ were identified by comparing isolates from the Bijagós with those from other countries in West Africa (N = 233), East Africa (N = 155), Southeast Africa (N = 172), South Central Africa (N = 150), and the Horn of Africa (N = 27), respectively (Table 1). When comparing the isolates from the Bijagós with samples from other African regions, there were 15 SNPs with $F_{ST} > 0.9$. This included 2 SNPs when comparing isolates from the Bijagós with East Africa, and 13 SNPs when comparing with the Horn of Africa (Table 1). SNPs with $F_{ST} > 0.9$ included those in the rhoptry-associated protein 1, RAP1 (PF3D7_1410400), and reticulocyte binding protein homologue 1 (PfRh) (PF3D7_0402300). The RAP1 protein is part of the rhoptry-associated protein (RAP) complex, which is important for parasitophorous vacuole membrane structure and intraerythrocytic parasite growth[12]. The PfRh protein is part of the reticulocyte binding-like (RBL) family, binds to erythrocytes and plays a key role in merozoite invasion[13].

Genetic relatedness. DNA segments that are identical between individuals and have been inherited from a common ancestor without recombination are termed identical by descent (IBD)[14]. IBD analysis was conducted to infer pairwise identity by descent between *Plasmodium* isolates within populations. Genomic relatedness was determined by analysing the proportion, or fraction, of the genome that was IBD between the *P. falciparum* isolates. Cumulative IBD across the genome was calculated using sliding window analysis for each population using windows of 10,000 bases (Supplementary Table 3). Isolates from the Bijagós exhibited median IBD fractions of 0.04, ranging between 0.001 and 0.217. This was slightly higher than the other countries in West Africa; Ivory Coast had a median IBD of 0.003, while The Gambia had a median IBD of 0.011. Higher IBD within a population indicates that a higher proportion of alleles at the same locus were inherited from a common ancestor. The higher IBD score for the Bijagós isolates suggests reduced outcrossing within the population, in comparison with other West African countries, which is likely to reflect their isolated island status.



Region	Country	N of SNPs with $F_{ST} > 0.75$	Approx. distance from Bijagós Archipelago (km)
West Africa	The Gambia	20	181
	Ivory Coast	3	1145
Central Africa	Gabon	2	3276
South Central Africa	DRC	35	4625
East Africa	Uganda	14	5285
	Kenya	43	5960
Horn of Africa	Eritrea	151	5850
Southeast Africa	Malawi	59	6176
	Madagascar	44	7548

Figure 2. Map showing the proximity of the Bijagós Archipelago to the other countries within this dataset, and the number of SNPs with $F_{ST} > 0.75$ for each country comparison.

Chr	Position	Gene	F_{ST} value	% Alternate allele – Bijagós population	% Alternate allele – Population 2	Population 2
10	331,746	PF3D7_1008100: PHD finger protein PHD1	0.892	50.0	0.5	West Africa
12	1,265,899	PF3D7_1230800: pre-mRNA-splicing regulator WTAP, putative	0.864	40.0	0	West Africa
6	697,009	PF3D7_0616800 malate:quinone oxidoreductase	0.854	0	91.0	East Africa
6	1,110,905	PF3D7_0627700 transportin	0.855	50	0	East Africa
6	1,115,439	PF3D7_0627800 acetyl-CoA synthetase	0.888	64.3	0.6	East Africa
7	**732,862**	eukaryotic translation initiation factor 3 subunit I, putative	**0.922**	**83.3**	**0.6**	**East Africa**
9	1,178,903	high molecular weight rhoptry protein 2	0.873	92.9	5.2	East Africa
9	1,402,365	PF3D7_0935500 *Plasmodium* exported protein, unknown function	0.869	73.3	1.6	East Africa
10	331,746	PF3D7_1008100: PHD finger protein PHD1	0.853	50.0	0.7	East Africa
10	658,448	serine/threonine protein kinase, FIKK family	0.865	81.8	3.9	East Africa
11	1,376,369	PF3D7_1135100: protein phosphatase PPM8, putative	0.870	13.3	94.2	East Africa
14	**421,792**	**rhoptry-associated protein 1**	**0.914**	**21.4**	**98.7**	**East Africa**
6	1,115,439	PF3D7_0627800 acetyl-CoA synthetase	0.887	64.3	0.6	Southeast Africa
7	732,862	eukaryotic translation initiation factor 3 subunit I, putative	0.893	83.3	1.7	Southeast Africa
10	331,746	PF3D7_1008100: PHD finger protein PHD1	0.899	50.0	0	Southeast Africa
12	1,265,899	PF3D7_1230800: pre-mRNA-splicing regulator WTAP, putative	0.886	40.0	0	Southeast Africa
12	1,265,899	PF3D7_1230800: pre-mRNA-splicing regulator WTAP, putative	0.882	12.5	0	South Central Africa
2	318,919 and **319,674**	**PF3D7_0207900 serine repeat antigen 2**	**0.929**	0	**94.4**	**Horn of Africa**
2	**367,271**	**6-cysteine protein P230p**	**0.934**	**0**	**94.7**	**Horn of Africa**
3	204,230	PF3D7_0304100 inner membrane complex protein 1e, putative	0.865	0	89.5	Horn of Africa
3	729,186	PF3D7_0317700: CPSF (cleavage and polyadenylation specific factor), subunit A, putative	0.859	0	87.5	Horn of Africa
4	**138,554** and 138,623	**reticulocyte binding protein homologue 1**	**0.942**	**0**	**95.8**	**Horn of Africa**
5	608,869	PF3D7_0514600: ribose-5-phosphate isomerase, putative	0.888	0	91.3	Horn of Africa
5	1,042,252	PF3D7_0525100 acyl-CoA synthetase	0.865	78.6	0	Horn of Africa
7	732,865	eukaryotic translation initiation factor 3 subunit I, putative	0.882	100	8.7	Horn of Africa
7	**894,384**	**PF3D7_0720700: phosphoinositide-binding protein PX1**	**0.923**	**100**	**3.7**	**Horn of Africa**
8	585,854	PF3D7_0811600: conserved protein, unknown function	0.895	0	92.6	Horn of Africa
8	**1,031,401**	**histone acetyltransferase GCN5**	**0.948**	**92.9**	**0**	**Horn of Africa**
8	**1,056,829**	**conserved *Plasmodium* protein, unknown function**	**0.948**	**100**	**3.7**	**Horn of Africa**
8	1,239,302	PF3D7_0828800: GPI-anchored micronemal antigen	0.879	0	88.9	Horn of Africa
8	**1,345,148**	***Plasmodium* exported protein, unknown function**	**1.000**	**100**	**0**	**Horn of Africa**
9	138,846	PF3D7_0801900: lysine-specific histone demethylase, putative	0.877	0	88.5	Horn of Africa
9	778,075	PF3D7_0918900: gamma-glutamylcysteine synthetase	0.870	0	90.5	Horn of Africa
9	**1,178,903 and 1,179,069**	**high molecular weight rhoptry protein 2**	**0.970**	**92.9**	**0**	**Horn of Africa**

Continued

Chr	Position	Gene	F_{ST} value	% Alternate allele – Bijagós population	% Alternate allele – Population 2	Population 2
9	**1,316,936**	**PF3D7_0933100: conserved *Plasmodium* protein, unknown function**	**0.925**	**0**	**92.3**	**Horn of Africa**
9	**1,406,576**	**PF3D7_0935600: gametocytogenesis-implicated protein**	**0.916**	**100**	**4.3**	**Horn of Africa**
9	1,407,002	PF3D7_0935600: gametocytogenesis-implicated protein	0.884	0	90.5	Horn of Africa
9	1,437,289 and 1,438,135	PF3D7_0936300: ring-exported protein 3	0.884	92.3	3.8	Horn of Africa
10	**658,448**	**serine/threonine protein kinase, FIKK family**	**0.908**	**81.8**	**0**	**Horn of Africa**
11	354,078	PF3D7_1108000: IWS1-like protein, putative	0.890	0	91.3	Horn of Africa
11	812,065	PF3D7_1121400: WD repeat-containing protein, putative	0.894	0	92.0	Horn of Africa
11	1,050,233	PF3D7_1126800: alternative splicing factor SR-MG, putative	0.862	0	87.0	Horn of Africa
11	1,376,369	PF3D7_1135100: protein phosphatase PPM8, putative	0.894	13.3	100	Horn of Africa
11	**1,507,877**	**guanylyl cyclase alpha**	**0.925**	**0**	**92.3**	**Horn of Africa**
12	**666,002**	**perforin-like protein 2**	**0.943**	**0**	**95.7**	**Horn of Africa**
12	1,998,804	PF3D7_1248700: conserved protein, unknown function	0.876	0	88.5	Horn of Africa
13	2,107,756	PF3D7_1352700: intron-binding protein aquarius, putative	0.851	0	88.9	Horn of Africa
14	812,013	PF3D7_1419400: conserved *Plasmodium* membrane protein, unknown function	0.878	0	90.5	Horn of Africa
14	1,227,066	PF3D7_1431200: OST-HTH associated domain protein, putative	0.875	6.7	95.2	Horn of Africa
14	1,989,372	PF3D7_1448500: conserved *Plasmodium* protein, unknown function	0.888	7.7	96.3	Horn of Africa
14	2,004,493	PF3D7_1449000: gamete egress and sporozoite traversal protein, putative	0.899	13.3	100	Horn of Africa
14	2,340,921	PF3D7_1457100: conserved *Plasmodium* protein, unknown function	0.876	0	88.5	Horn of Africa
14	2,343,679	PF3D7_1457200: thioredoxin 1	0.879	0	88.9	Horn of Africa
14	2,357,358	PF3D7_1457400: conserved *Plasmodium* protein, unknown function	0.897	0	92.6	Horn of Africa
14	2,364,543	PF3D7_1457600: conserved *Plasmodium* protein, unknown function	0.869	0	92.6	Horn of Africa
14	2,364,847	PF3D7_1457600: conserved *Plasmodium* protein, unknown function	0.854	0	88.9	Horn of Africa
14	2,367,254	PF3D7_1457700: large ribosomal subunit nuclear export factor, putative	0.851	0	88.5	Horn of Africa
14	2,469,318	PF3D7_1460500: conserved *Plasmodium* protein, unknown function	0.894	0	91.7	Horn of Africa
14	2,535,998	PF3D7_1462400: conserved *Plasmodium* protein, unknown function	0.899	0	92.6	Horn of Africa
14	2,845,773	PF3D7_1469400: nucleoside transporter 3, putative	0.859	0	87.5	Horn of Africa

Table 1. Genetic markers of population differentiation between the isolates from the Bijagós Archipelago and isolates from countries in the other African regions, with $F_{ST} > 0.85$. SNPs with $F_{ST} > 0.9$ have been highlighted in bold. Only SNPs where all calls in the Bijagós population were either reference, alternate, or mixed, were included. Sample sizes: Bijagós Archipelago N = 15, West Africa N = 233, East Africa N = 155, Southeast Africa N = 172, South Central Africa N = 150, Horn of Africa N = 27.

Genomic regions under recent positive selection. WGS data was analysed for signatures of recent positive selection. Single population iHS and cross-population XP-EHH metrics were calculated, which are statistics derived from extended haplotype homozygosity. Within the isolates from the Bijagós, there were seven SNPs which exhibited significantly high iHS scores, where $(-\log_{10}[1 - 2 \mid \Phi_{iHS} - 0.5 \mid]) > 4.0$ (Supplementary Table 4). These SNPs were located within two surface-associated interspersed (SURFIN) pseudogenes, SURFIN 1.2 and SURFIN 13.1. Cross-population analysis was conducted between isolates from the Bijagós and other regional populations within Africa. This analysis identified seven candidate regions under directional selection based on XP-EHH values (Supplementary Table 5). When comparing isolates from the Bijagós with those from all other African regions, there were four common candidate regions under directional selection. These regions encoded cytoadherence linked asexual proteins 3.1 and 3.2, the ABC transporter B family member 4 (putative), and the CX3CL1-binding protein 2. When comparing isolates from the Bijagós to those from Central Africa, South Central Africa, Southeast Africa and West Africa, there were two additional common candidate regions under directional selection. These regions encode the 26S protease regulatory subunit 8 (putative) and a conserved *Plasmodium* protein of unknown function (PF3D7_1248700). When comparing isolates from the Bijagós with isolates from Southeast Africa and West Africa, two additional common candidate regions were identified, which encode merozoite surface protein 3 and methionine tRNA ligase. When comparing with East Africa, one additional candidate region was found, encoding Plasmepsin X. When comparing with South Central Africa, two additional candidate regions were identified, encoding the 26S protease regulatory subunit 6A (putative), and structural maintenance of chromosomes protein 1 (putative) (Supplementary Table 5).

Population admixture. Admixture analysis is a method of investigating the geographical origins of samples, based on their ancestry. Admixture analysis was conducted on 387,395 SNPs using ADMIXTURE software. Using a cross-validation approach, five ancestral populations were inferred (K value = 5). Figure 3 (top) shows these five ancestral populations (K = 1 to K = 5) as different colours. The figure includes each individual *P. falciparum* isolate along the x-axis, and the ancestral populations of each isolate along the y-axis, represented by colour. This analysis suggested that isolates from the Bijagós contained high proportions of ancestral genome fragments associated with West African *P. falciparum* populations (Fig. 3).

PCA was carried out using the previously described pairwise genetic distance matrix containing 1,326,629 SNPs from 795 isolates. Each *P. falciparum* isolate on the PCA was coloured according to its dominant ancestral population (of K = 1 to K = 5). This is shown in Fig. 3 (bottom). The overlapping of dominant ancestry reinforced high proportions of shared ancestral genome fragments for the samples across Africa (Fig. 3).

Identification of mutations associated with drug resistance. Mutations associated with resistance to antimalarial drugs were analysed using the WGS data from the Bijagós isolates (n = 15) and isolates from West Africa (n = 218). Comparison of the prevalence of drug resistance mutations between these populations was carried out to understand similarities and differences between these isolates (Table 2).

The K76T mutation, associated with resistance to chloroquine[15], was found at a slightly lower frequency (28.6%) in isolates from the Bijagós compared to isolates from The Gambia and Ivory Coast in mainland West Africa (51.4%) (non-significant $P > 0.05$). Overall, PfDHFR mutations associated with resistance to pyrimethamine[16] were found at higher prevalence in isolates from the Bijagós than those from mainland West Africa. N51I was found at significantly higher prevalence in the Bijagós (100%) than in mainland West Africa (76.6%) ($P < 0.05$). S108N was found at 100% prevalence in the Bijagós compared to 81.1% in mainland West Africa (non-significant $P > 0.05$). C59R was found at 92.9% prevalence compared to 77.5% in mainland West Africa (non-significant $P > 0.05$). Of the 15 Bijagós samples sequenced, one was a double mutant with N51I-S108N, and 13 were triple mutants with N51I-C59R-S108N. The PfDHPS A437G mutation, associated with resistance to sulphadoxine[17], was found in 60% of isolates from the Bijagós. K540E was not found in isolates from the Bijagós or mainland West Africa. A581G was not found in isolates from the Bijagós but was found at a prevalence of 0.9% in isolates from mainland West Africa (non-significant difference $P > 0.05$). Of the 15 Bijagós samples sequenced, two were double mutant with S436A-A437G.

PfMDR1 mutations N86Y and Y184F, associated with resistance to chloroquine[18,19], were found in the Bijagós isolates at frequencies of 7.7% and 53.8%, respectively. The prevalence of N86Y and Y184F was slightly higher in isolates from mainland West Africa, at 15.6% and 60.1%, respectively (non-significant difference $P > 0.05$). The mutation D1246Y, also associated with chloroquine resistance[18,19], was not found in isolates from the Bijagós, but was found in 0.9% of isolates from mainland West Africa (non-significant difference $P > 0.05$). The N458Y mutation in PfKELCH13, associated with artemisinin partial resistance[20], was not found in isolates from the Bijagós or mainland West Africa. There was not sufficient genomic coverage within the Bijagós sample WGS data of other PfKELCH13 marker positions to report these reliably. Mutations which could not be reported on are listed in Supplementary Table 7. The PfEXO E415G mutation, associated with resistance to piperaquine, and the PfAP2μ S160N/T mutation, which may be associated with artemisinin-resistance, were not found in either population[21].

Discussion

This study presents *Plasmodium falciparum* sequence data from the Bijagós Archipelago, Guinea-Bissau. Investigating the genomic architecture of malaria parasites in endemic countries increases the understanding of parasite genetic variation and population dynamics. This includes the prevalence of genomic mutations associated with drug resistance. Knowledge of drug resistance markers and population dynamics can assist in malaria control and surveillance, which are key to reducing the global malaria burden in line with the World Health Organization (WHO) Global Technical Strategy[1]. Genomic surveillance of drug resistance markers is particularly

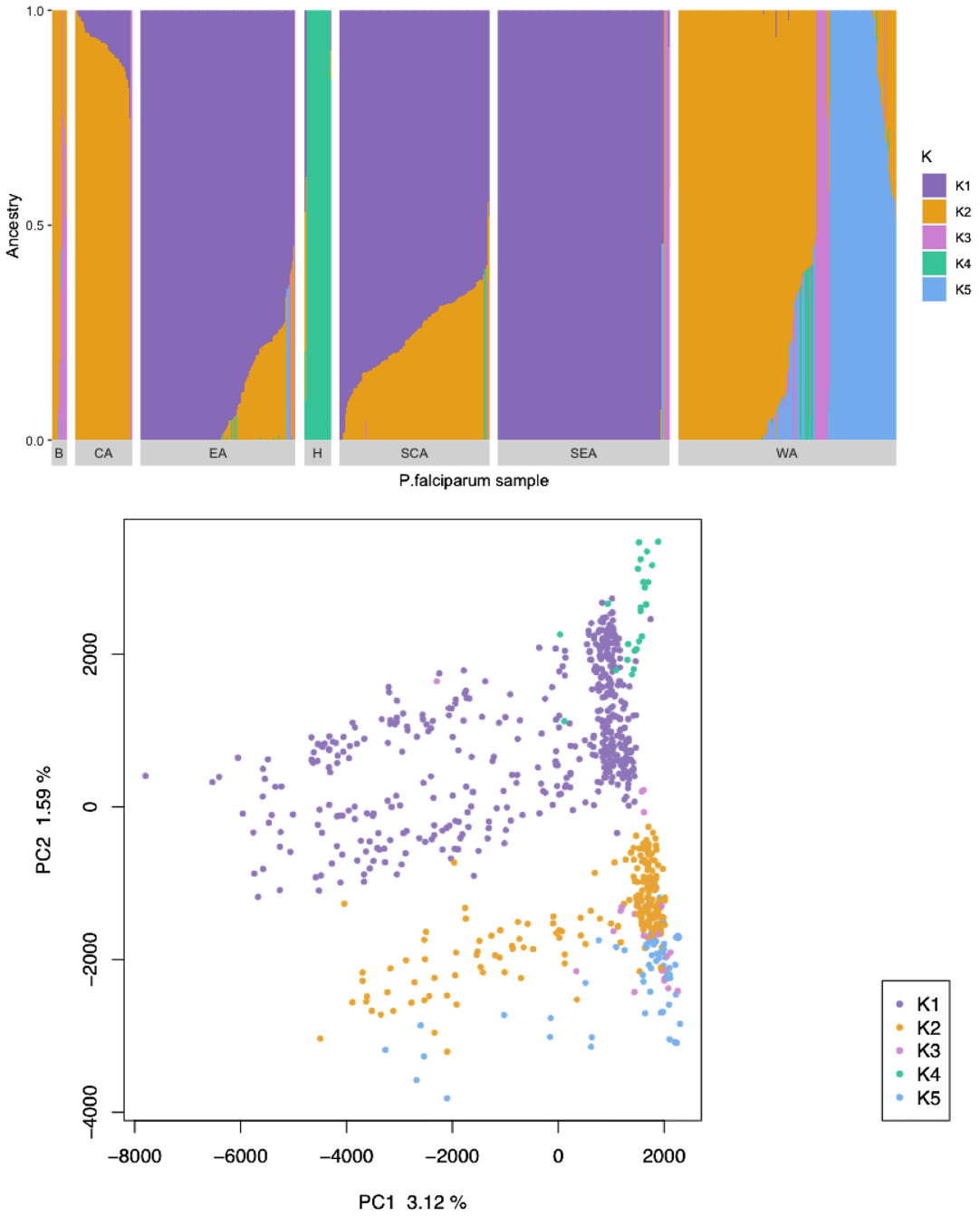


Figure 3. (top) Ancestry analysis of *P. falciparum* isolates from (B) the Bijagós Archipelago, (CA) Central Africa, (EA) East Africa, (H) Horn of Africa, (SCA) South Central Africa, (SEA) Southeast Africa and (WA) West Africa. Individual isolates are plotted along the x-axis, with ancestry coefficients on the y-axis. The number of ancestral populations, K, used in this analysis was K = 5. (bottom) Principal component analysis generated using pairwise genetic distance matrices. This is coloured by the maximum K value for each isolate.

important considering the likely increase in the mass use of antimalarials following recently expanded WHO recommendations[2].

The Bijagós Archipelago is a region with low-moderate malaria transmission intensity, exhibiting stable and seasonal malaria cases. Many malaria cases in the region are asymptomatic, which presents a difficult challenge to control programmes, as asymptomatic cases sustain a reservoir of infections[22]. Consequently, this reservoir is capable of continually driving malaria transmission, which makes eliminating malaria in this area extremely difficult[23,24]. Furthermore, drug resistance is hypothesised to spread faster in low transmission settings than

Gene	Position	Mut	Ref allele	Alt allele	n/N mutation Bijagós	n/N mutation W. Africa	% Mut in Bijagós	% Mut in W. Africa	*P* value (X^2)
pfcrt	403,625	K76T	A	C	4/14	112/218	28.6	51.4	0.098
pfdhfr	748,239	N51I	A	T	14/14	167/218	100	76.6	0.040
pfdhfr	748,262	C59R	T	C	13/14	169/218	92.9	77.5	0.158
pfdhfr	748,410	S108N	G	A	14/14	176/218	100	81.1	0.070
pfdhps	549,993	K540E	A	G	0/15	0/218	0	0	-
pfdhps	549,685	A437G	C	G	9/15	162/218	60.0	74.3	0.225
pfdhps	550,117	A581G	C	G	0/15	2/218	0	0.9	0.709
pfdhps	549,681	S436A	T	G/C	3/15	64/218	20.0	29.4	0.439
pfmdr1	958,145	N86Y	A	T	1/13	34/218	7.7	15.6	0.440
pfmdr1	958,440	Y184F	A	T	7/13	131/218	53.8	60.1	0.656
pfmdr1	961,625	D1246Y	G	T	0/13	2/218	0	0.9	0.729
pfkelch13	1,725,626	N458Y	T	A/C	0/14	0/218	0	0	-
pfexo	2,504,560	E415G	A	G	0/13	0/218	0	0	-
pfap2μ	718,433	S160N/T	G	A/C	0/14	12/218	0	5.5	0.367

Table 2. Frequency of mutations associated with resistance to different anti-malarial drugs, from Bijagós *P. falciparum* isolates (n = 15, sample collection in 2018) and *P. falciparum* isolates from West Africa (Total n = 218, Ivory Coast n = 70—sample collection in 2013, The Gambia n = 148—sample collection in 2008 and 2014).

in high transmission settings due to a higher proportion of infections being exposed to antimalarial drugs. This increases the probability of treatment failure and the subsequent transmission of resistant parasites in the population[25].

The data analyses presented in this study characterise asymptomatic malaria infections. These infections are more difficult to study due to the low quantities of parasite DNA in clinical samples, particularly in dried blood spots (DBS). This means that despite the importance of understanding asymptomatic cases, most research into genomic variation and drug resistance is conducted on parasites from symptomatic malaria infections. This study made use of Selective Whole Genome Amplification (SWGA) to amplify parasite DNA from isolates collected using DBS, from asymptomatic infections within the Bijagós[7]. These advances in molecular methods, along with improved Next Generation Sequencing technology, allow for feasible routine surveillance of parasite genetic variation and drug resistance markers from dried blood spot samples, which can be collected during routine cross-sectional surveys.

Following successful amplification of low-density asymptomatic *P. falciparum* infections obtained in the Bijagós, population analyses were conducted. The Bijagós *P. falciparum* genomic variation was then contextualised with the wider African *P. falciparum* population. Principal component analysis (PCA) indicated that the malaria samples from the Bijagós cluster with samples from other countries in West Africa, including The Gambia and Ivory Coast. A relatively low percentage of the variance between malaria samples from different countries was explained within the PCA. This, along with the absence of SNPs with $F_{ST} = 1$, reflects the high degree of mixing between *P. falciparum* isolates in the Bijagós Archipelago and those on the mainland. Despite this relatively small sample size of 15 from the Bijagós, these results corroborate the high degree of mixing between *P. falciparum* isolates demonstrated across the African continent[10]. The samples from the Bijagós had a mean F_{WS} of 0.965, with 86.7% monoclonality (with Fws ≥ 0.95). This was the highest percentage of monoclonality when compared with samples from other countries across the African continent. The high percentage of monoclonal infections may be due to people in the Bijagós living in small, relatively isolated, communities, which could reduce the likelihood of simultaneous infection with parasites of different genotypes. This monoclonality may also reflect the low to moderate transmission intensity of *P. falciparum* in the Bijagós. Despite the high degree of population mixing, identity by descent analysis revealed that isolates from the Bijagós had slightly higher IBD fractions than isolates from the other West African countries, which may reflect their more isolated island status. Admixture analysis gave further insight into the ancestral origins of the Bijagós Archipelago isolates amongst African populations, suggesting Bijagós isolates obtained large proportions of their ancestral genome fragments from West African *P. falciparum* lineages.

Fixation index (F_{ST}) analyses were used to measure population differentiation due to genetic structure. F_{ST} values range from 0 to 1, where a value of 0 for a particular SNP indicates that two populations have no difference in allele frequency, whereas a value of 1 indicates complete differentiation in allele frequency between the two populations [11]. Therefore, a value of 0 would indicate a high degree of population mixing. Whereas, higher F_{ST} values at a particular SNP indicate greater population differentiation. Fifteen SNPs with a high $F_{ST} > 0.9$ were identified when comparing isolates from the Bijagós with those from other African regions. The greatest amount of population differentiation was found between samples from the Bijagós and Eritrea. Interestingly, SNPs in the Bijagós samples were not fixed at either allele for many of the loci with high F_{ST} values, indicating that the population is in transition for many of these alleles. Overall, there were no SNPs with $F_{ST} = 1$ in any of the regional comparisons. This indicates that there were no SNPs specific to the Bijagós parasite population. It

is likely that a larger sample size would be needed to reveal these SNPs with $F_{ST} = 1$, which could then be used to generate a molecular barcode for the *P. falciparum* samples from the Bijagós able to predict geographic origin[26].

Regions under selective pressure included seven loci within the surface-associated interspersed (SURFIN) pseudogenes, SURFIN 1.2 and SURFIN 13.1. SURFINs are variable erythrocyte surface antigens and are targets of naturally acquired immunity against malaria[27]. SURFIN 1.2 has been associated with chloroquine sensitivity[27], and it has been suggested that both SURFIN 1.2 and SURFIN 13.1 are involved in cell surface adhesion[27,28]. However, additional phenotypic experimentation is required to confirm the role of these two pseudogenes. Cross-population analysis identified seven candidate regions under directional selection based on XP-EHH values. These candidate regions included genes encoding Plasmepsin X (PMX), the ABC transporter B family member 4 (ABCB4), and *P. falciparum* merozoite surface protein 3 (PfMSP3). Plasmepsin X is essential for asexual parasite development and is considered a multi-stage antimalarial target[29]. ABCB4 is a member of the ABC (ATP-binding cassette) B family. This is a group of transport proteins which includes the PfMDR1 protein, and polymorphisms in the ABC transporters have been involved in anti-malarial drug resistance[30]. PfMSP3 is a merozoite surface protein and is a potential vaccine candidate[31].

This study reported on genomic markers associated with drug resistance in the Bijagós Archipelago. Anti-malarial treatments administered on the islands include artemether-lumefantrine and IPTp with sulphadoxine-pyrimethamine (SP). Intravenous artesunate or quinine are only used in severe, or complicated, cases. The PfCRT K76T mutation, associated with resistance to chloroquine, had a frequency of 28.6% on the islands, despite chloroquine not being included in malaria treatment policy since 2006 in Guinea-Bissau. This was a slightly lower prevalence than observed in isolates from mainland West Africa, but this difference was not significant ($P > 0.05$). Persistence of the K76T mutation may indicate that it has not yet been lost due to fitness-cost. The PfDHFR mutations associated with resistance to pyrimethamine, N51I and S108N, were found to be fixed (100% prevalence) in the Bijagós isolates, and C59R was found close to fixation at 92.9% prevalence. These mutations were found at a higher prevalence in the Bijagós isolates than in the isolates from mainland West Africa, with the N51I mutation at significantly higher prevalence in the Bijagós ($P = 0.04$). This difference may be partially explained by the collection years of these isolates, as isolates from the Bijagós were collected in 2018, whereas isolates from mainland West Africa were collected earlier, between 2008 and 2014. Therefore, the Bijagós isolates were collected following additional years of selection pressure from the use of SP in the region (Supplementary Table 6). Thirteen of the Bijagós isolates sequenced exhibited the triple mutant N51I-C59R-S108N haplotype. In addition, the PfDHPS A437G and S436A mutations, associated with resistance to sulphadoxine, were found at 60% and 20% prevalence, respectively. Although this is a small sample set of 15 isolates, the presence of mutations associated with resistance to SP warrants further investigation with a larger sample size. Encouragingly, the PfEXO E415G mutation associated with piperaquine resistance, and the PfAP2μ S160N/T mutation, which may be associated with artemisinin-resistance, were not found in the Bijagós isolates. Limited sequencing coverage of the PfKELCH13 gene made investigating the majority of validated or candidate resistance mutations in this gene unreliable. However, there was sufficient coverage of the N458Y genome position to identify reliably that this mutation was absent in the Bijagós isolates. The artemisinin-combination therapy dihydroartemisinin-piperaquine (DHA-PPQ) was not used on the islands prior to this study. However, there has since been MDA with DHA-PPQ across the archipelago.

Despite the limitations in generating sequence data from asymptomatic isolates collected using DBS, this study reported high quality whole genome sequence data for 15 Bijagós isolates. This is a small sample size which has limitations for generalisation. However, the population genomic analyses results are consistent with those from neighbouring African countries, indicating that these results are robust. Furthermore, this study further demonstrates the capacity of SWGA methods in generating WGS data from low-density parasite concentrations. This dataset forms the only dataset of WGS data for the Bijagós. Therefore, it forms an important baseline in the context of evaluating subsequent MDA and SMC interventions that have been implemented in the region. The Bijagós WGS data was contextualised with publicly available WGS sequence data from isolates across Africa. Information about the symptomatic or asymptomatic status of these infections is not publicly available, so comparisons concerning patient symptom status cannot be made using this dataset. Additional sequencing of *P. falciparum* isolates is required to reveal a more comprehensive picture of regional dynamics, particularly concerning the prevalence of drug resistance markers. This should ideally include the analysis of samples collected from both asymptomatic and symptomatic infections, and across different years, to understand temporal changes in resistance marker frequency. High throughput, low-cost, methods, such as amplicon sequencing, have the potential to achieve this. In turn, this information can be used to inform programmatic decisions in malaria control, working towards malaria elimination.

Materials and methods

Dried blood spot collection. Dried blood spots were collected onto filter paper from residents of the Bijagós Archipelago during a cross-sectional survey in 2018. Dried blood spots were sampled as part of this mapping survey, regardless of whether the participant had symptoms of malaria.

This study was conducted in accordance with the Declaration of Helsinki and granted ethical approval by the Comite Nacional de Eticas de Saude (Guinea-Bissau) and the London School of Hygiene and Tropical Medicine Research Ethics Committee (UK). Written or thumbprint informed consent was obtained from all study participants.

DNA extraction and qPCR for malaria positivity. Following collection, dried blood spots were punched for DNA extraction using 3 mm diameter punches. Three punches were taken per sample for DNA

extraction. DNA was extracted using the PureLink™ Pro 96 Genomic DNA Purification Kit (ThermoFisher), (cat No: K182104A). DNA extract was quantified for *Plasmodium falciparum* 18S gene using qPCR.

***Plasmodium* genome amplification and sequencing.** DNA extracts which had a qPCR CT value < 30 were tested using the Qubit for DNA concentration. Using this CT threshold, n = 145 samples were selected for further analysis. DNA concentration in the extracted dried blood spot samples was very low, of the range 0.011 ng/µl to 0.268 ng/µl. Selective whole genome amplification (SWGA) was used to amplify the *Plasmodium* DNA prior to sequencing, using a protocol adapted from Oyola et al.[7]. Each SWGA reaction used 30 µl of extracted DNA, which was combined with 5 µl of *phi29* DNA Polymerase Reaction Buffer (New England Biolabs), 0.5 µl of Recombinant Albumin (New England Biolabs), 0.5 µl of SWGA primers (250 mM) (Roche), 5 µl dNTPs (New England Biolabs) and 6 µl of water, to a total volume of 50 µl per reaction. SWGA reactions were then run in a thermocycler using the following programme steps: 35 °C for 5 min, 34 °C for 10 min, 33 °C for 15 min, 32 °C for 20 min, 31 °C for 25 min, 30 °C for 16 h, 65 °C for 10 min, and held at 4 °C. SWGA reactions increased the DNA concentration of samples to > 300 ng DNA in of the 48 of the 145 (33.1%) samples amplified.

Whole genome sequencing. The SWGA product from these samples was purified using KAPA pure beads (Roche). Beads were equilibrated to room temperature and mixed with SWGA product at a 1:1 volume ratio. The KAPA pure beads Genomic DNA Purification protocol steps were then followed (https://rochesequencingstore.com/wp-content/uploads/2022/07/KAPA-Pure-Beads-Technical-Data-Sheet.pdf). DNA was then quantified using the Qubit dsDNA HS Kit (Thermo Fisher). Samples contained > 300 ng DNA and were sequenced at Eurofins Genomics using the Illumina Novaseq 6000, producing a minimum of 3.75 M paired reads (250 bp reads) per sample.

Bioinformatics. FastQ files from WGS were used to produce VCF files for each sample. To do this, FastQ files were trimmed using *trimmomatic* software (version 0.39, using the parameters LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20 MINLEN:36) to remove poor quality sequences[32]. The trimmed data was then aligned to the reference *P. falciparum* 3D7 genome (version 3) using *bwa-mem* software (version 0.7.17-r1188, default parameters), to produce a BAM file for each sample. The *samtools* (version 1.12) functions fixmate and markdup were applied to the resulting BAM files[33]. Following this, the *GATK*'s BaseRecalibrator and ApplyBQSR functions were applied for calibration and correction of the BAM files using the *P. falciparum* genetic crosses 1.0 dataset (http://www.malariagen.net/data/pf-crosses-1.0). SNPs and indels were called with *GATK's* HaplotypeCaller (version 4.1.4.1) using the option -ERC GVCF[34]. Validated VCFs were imported into GenomicsDB using *GATK*'s function GenomicsDBImport, and a combined VCF was created using *GATK*'s GenotypeGVCFs function. This VCF file contains all the genotype calls for all the variant positions across all samples.

For further analysis, only variants in core regions of the genome were included and assigned a quality score using GATK's Variant Quality Score Recalibration (VQSR). VariantRecalibrator was run using the previously mentioned genetic crosses dataset as a training set with the following parameters used for SNPs: -an QD -an FS -an SOR -an DP -maxGaussians 8 and -mq-cap-for-logit-jitter-transform 70. Following this, GATK's ApplyVQSR was run using the parameter -truth-sensitivity-filter-level 99.0 to obtain a Variant Quality Score Log-Odds (VQSLOD). Variants with a VQSLOD score < 0, representing variants more likely to be false than true, were filtered out.

Additionally, isolates with more than 40% of SNPs missing genotype data (177/2129 isolates) were excluded from downstream analysis to ensure that samples which passed quality control had uniform sequencing coverage across the whole genome. SNPs were then processed to replace the genotype call in the VCF for a mixed call whenever the secondary MAF was at least 20% in a given SNP on each individual sample. A subset of only biallelic SNPs was also obtained from this VCF. The filtered VCF was then converted into a binary matrix consisting of genotypes (0 = reference allele; 1 = mutant allele; 0.5 = mixed allele) across all SNPs and isolates. The final matrix consisted of 1952 isolates and 1,326,629 high-quality SNPs. This bioinformatics pipeline is described in detail by Benavente et al.[35]. An annotated VCF of all extracted variants was made using *snpEff* software (version 5.1), which annotates the variants[36]. Fifteen of the Bijagós samples had sufficient quality data for downstream bioinformatic analysis, with over ~ 40% of the genome covered by > 5 reads.

Population genomic analysis. Whole Genome Sequence data from the samples from the Bijagós were compared with publicly available data from mainland Guinea-Bissau and other African countries, through the MalariaGEN Pf3k project and ongoing LSHTM studies (https://www.malariagen.net/parasite/pf3k). This included whole genome sequences from 795 isolates, from 10 different countries across Africa (Supplementary Table 1). This included isolates from Central Africa (Gabon, n = 57), West Africa (Ivory Coast n = 70, The Gambia n = 148, Bijagós Archipelago Guinea-Bissau (n = 15), East Africa (Kenya n = 138, Uganda n = 17), Southeast Africa (Malawi n = 149, Madagascar n = 24), South Central Africa (DRC n = 150) and Horn of Africa (Eritrea n = 27). A binary matrix of pairwise genetic distances was constructed from this WGS data. To investigate population structure, Principal Component Analysis was conducted using the filtered bi-allelic nucleotide matrix, using the R package *amap* (version 0.8–18), and a neighbour-joining (N-J) tree was constructed using the R package *ape* (version 5.5). Neighbour-joining trees were visualised in iTOL[37]. Fixation index statistics (F_{ST}) were calculated to infer nucleotide diversity and population differentiation using VCFtools (version 0.1.16, https://vcftools.github.io/documentation.html#fst). F_{ST} statistic outputs were only included in downstream analysis if a particular SNP had coverage in at least 2/3 of both of the total sample populations that were being compared.

The F_{WS} metric was used to quantify the extent of multiplicity of infection (MOI), by assessing within host diversity of *Plasmodium* in comparison to local population diversity[38]. This was calculated using the moimix R package (version 0.0.2.9001, https://bahlolab.github.io/moimix/)[39]. This indicated the frequency of inbreeding

and outcrossing in the population. High F_{WS} (> 0.95) is highly indicative of monoclonality, whereas low F_{WS} indicates high levels of mixing between the parasites in the population, and subsequently a poorly defined population sub-structure.

IBD analysis was conducted using *hmmIBD* software (version 2.0.4) (default parameters). This software implements a hidden Markov model for detecting genomic regions that are identical by descent (IBD) for pairs of haploid samples[14]. Additional filtering steps were performed on the binary SNP matrix for IBD analysis: SNPs were selected with minor allele frequency > 0.01, samples were filtered for F_{WS} > 0.95 and mixed calls were filtered to alternate calls. IBD analysis enabled inference of pairwise identity by descent between *Plasmodium* isolates within populations. Genomic relatedness was determined by analysing the proportion, or fraction, of the genome that was IBD between *Plasmodium* isolates. Cumulative IBD across the genome was calculated using sliding window analysis (using windows of 10,000 bases) for each population. The median and range of IBD values were calculated. Thirteen of the Bijagós samples passed F_{WS} filtering (≥ 0.95) to be included in IBD analysis, and despite this small sample size, the raised IBD results appear robust.

Whole genome sequence data was interrogated for signatures of selection, using the rehh package to calculate two different statistics derived from extended haplotype homozygosity: iHS and XP-EHH[40]. These statistics are based on the decay of haplotype homozygosity, due to recombination, in the absence of selection on a particular allele[41]. Integrated haplotype score (iHS) identifies signatures of recent selection in a particular population. A negative iHS suggests that selection has favoured a derived allele, whereas a positive iHS suggests selection for the ancestral allele[41]. Cross-population extended haplotype homozygosity (XP-EHH) compares selection between two populations, through comparing the lengths of haplotypes associated with the same allele in different populations[42]. Positive XP-EHH indicates selection occurring in population A, whereas negative XP-EHH indicates selection occurring in population B. Highly variable genes (e.g., *PfEMP1*) were removed from this analysis due to the increased likelihood of them indicating false signals of selection pressure, due to an increased risk of alignment errors.

Admixture analysis was conducted using the ADMIXTURE software (version 1.3)[43]. Analysis was conducted on 387,395 SNPs with minor allele frequencies of at least 0.001. This is a program used to estimate ancestry in unrelated individuals, from autosomal SNP genotype sets. PLINK (version 1.9) was used to convert the filtered VCF file to a bed file. ADMIXTURE was then used to estimate ancestry using this bed file and a specified K value. K is the estimated number of ancestral populations. The optimum K value for ancestral admixture coefficients was estimated through cross-validation of 1–10 dimensions of eigenvalue decay and found to be K = 5. Cross-validation was repeated using 10 randomised seed inputs and results were averaged across seeds to find the optimum K value. The output was visualised in R (v3.5.1). SNPs associated with resistance to antimalarial drugs were analysed in whole genome sequence data using the dplyr package (v1.0.10) in R.

Data availability

The processed datasets are available at NCBI: http://www.ncbi.nlm.nih.gov/bioproject/905557. Bioproject: PRJNA905557. The following sample SRR accession numbers can be used to find individual sample sequence data using the SRA database search in NCBI, https://www.ncbi.nlm.nih.gov/sra/: SRR22412636, SRR22412629, SRR22412635, SRR22412622, SRR22412623, SRR22412624, SRR22412625, SRR22412626, SRR22412627, SRR22412628, SRR22412630, SRR22412631, SRR22412632, SRR22412633, and SRR22412634.

Received: 23 November 2022; Accepted: 8 April 2023
Published online: 18 April 2023

References

1. World Malaria Report 2021. *Geneva: World Health Organization; 2021. Licence: CC BY-NC-SA 3.0 IGO.*
2. Geneva: World Health Organization; 2022 (WHO/UCN/GMP/2022.01 Rev.2). License: CC BY-NC-SA 3.0 IGO. *WHO Guidelines for malaria - 3 June 2022*. http://apps.who.int/bookorders. (2022).
3. Nsanzabana, C., Djalle, D., Guérin, P. J., Ménard, D. & González, I. J. Tools for surveillance of anti-malarial drug resistance: An assessment of the current landscape. *Malar J* **17**, 1–16 (2018).
4. Hutchins, H. *et al.* A survey of knowledge, attitudes and practices regarding malaria and bed nets on Bubaque Island Guinea-Bissau. *Malar. J.* https://doi.org/10.1186/s12936-020-03469-1 (2020).
5. McGregor, D. *et al.* The epidemiology of plasmodium falciparum malaria in the Bijagós islands of Guinea-Bissau. *Am. J. Trop. Med. Hyg.* **104**, 2117–2122 (2021).
6. Ant, T. *et al.* A survey of Anopheles species composition and insecticide resistance on the island of Bubaque, Bijagós Archipelago, Guinea-Bissau. *Malar. J.* **19**, 1–10 (2020).
7. Oyola, S. O. *et al.* Whole genome sequencing of Plasmodium falciparum from dried blood spots using selective whole genome amplification. *Malar. J.* https://doi.org/10.1186/s12936-016-1641-7 (2016).
8. MalariaGEN Pf3k Project. https://www.malariagen.net/parasite/pf3k.
9. Auburn, S. *et al.* Characterization of within-host plasmodium falciparum diversity using next-generation sequence data. *PLoS ONE* **7**, 32891 (2012).
10. Amambua-Ngwa, A. *et al. Major subpopulations of Plasmodium falciparum in sub-Saharan Africa*. https://www.science.org.
11. Weir, B. S. & Cockerham, C. C. Estimating F-statistics for the analysis of population structure. *Evolution (N Y)* **38**, 1358 (1984).
12. Ghosh, S. *et al.* The Plasmodium rhoptry associated protein complex is important for parasitophorous vacuole membrane structure and intraerythrocytic parasite growth. *Cell Microbiol.* **19**, e12733 (2017).
13. Triglia, T., Tham, W. H., Hodder, A. & Cowman, A. F. Reticulocyte binding protein homologues are key adhesins during erythrocyte invasion by *Plasmodium falciparum*. *Cell Microbiol.* **11**, 1671–1687 (2009).
14. Schaffner, S. F., Taylor, A. R., Wong, W., Wirth, D. F. & Neafsey, D. E. HmmIBD: Software to infer pairwise identity by descent between haploid genotypes. *Malar. J.* **17**, 1–4 (2018).
15. Roepe, P. D. Molecular and physiologic basis of quinoline drug resistance in *Plasmodium falciparum* malaria. *Future Microbiol.* 4, 441–455. Preprint at https://doi.org/10.2217/fmb.09.15 (2009).

16. Basco, L. K. & Ringwald, P. Molecular epidemiology of malaria in Yaounde, Cameroon. vi. Sequence variations in the *Plasmodium falciparum* dihydrofolate reductase-thymidylate synthase gene and in vitro resistance to pyrimethamine and cycloguanil. *Am. J. Trop. Med. Hyg.* 62 (2000).
17. Hyde, J. E. Drug-resistant malaria - An insight. *FEBS J.* 274, 4688–4698. Preprint at https://doi.org/10.1111/j.1742-4658.2007.05999.x (2007).
18. Reed, M. B., Saliba, K. J., Caruana, S. R., Kirk, K. & Cowman, A. F. Pgh1 modulates sensitivity and resistance to multiple antimalarials in *Plasmodium falciparum*. *Nature* **403**, 906–909 (2000).
19. Pickard, A. L. *et al.* Resistance to antimalarials in Southeast Asia and genetic polymorphisms in pfmdr1. *Antimicrob. Agents Chemother.* **47**, 2418–2423 (2003).
20. Ménard, D. *et al.* A Worldwide Map of *Plasmodium falciparum* K13-propeller polymorphisms. *N Engl J Med* **374**, 2453–2464 (2016).
21. Amato, R. *et al.* Genetic markers associated with dihydroartemisinin–piperaquine failure in *Plasmodium falciparum* malaria in Cambodia: A genotype–phenotype association study. *Lancet Infect. Dis.* **17**, 164–173 (2017).
22. Bousema, J. T. *et al. Plasmodium falciparum* gametocyte carriage in asymptomatic children in western Kenya. *Malar. J.* **3**, (2004).
23. Zoghi, S. *et al.* Survey for asymptomatic malaria cases in low transmission settings of Iran under elimination programme. *Malar. J.* **11**, 1–10 (2012).
24. Andolina, C. *et al.* Sources of persistent malaria transmission in a setting with effective malaria control in eastern Uganda: A longitudinal, observational cohort study. *Lancet Infect. Dis.* **21**, 1568–1578 (2021).
25. Klein, E. Y., Smith, D. L., Boni, M. F. & Laxminarayan, R. Clinically immune hosts as a refuge for drug-sensitive malaria parasites. *Malar. J.* **7**, (2008).
26. Diez Benavente, E. *et al.* A molecular barcode to inform the geographical origin and transmission dynamics of *Plasmodium vivax* malaria. *PLoS Genet* **16**, e1008576 (2020).
27. Kanoi, B. N. *et al.* Global repertoire of human antibodies against *Plasmodium falciparum* RIFINs, SURFINs, and STEVORs in a malaria exposed population. *Front. Immunol.* **11**, 893 (2020).
28. Prajapati, S. K. *et al.* The transcriptome of circulating sexually committed *Plasmodium falciparum* ring stage parasites forecasts malaria transmission potential. *Nat. Commun.* **11**, (2020).
29. Richardson, L. W. *et al.* Substrate peptidomimetic inhibitors of *P. falciparum* Plasmepsin X with potent antimalarial activity. *ChemMedChem* **17**, e202200306 (2022).
30. Koenderink, J. B., Kavishe, R. A., Rijpma, S. R. & Russel, F. G. M. The ABCs of multidrug resistance in malaria. *Trends Parasitol.* **26**, 440–446 (2010).
31. Alves, K. C. S., Guimarães, J. M., de Almeida, M. E. M. & Mariúba, L. A. M. *Plasmodium falciparum* merozoite surface protein 3 as a vaccine candidate: A brief review. *Rev. Inst. Med. Trop. Sao Paulo* **64**, (2022).
32. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
33. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, (2021).
34. McKenna, A. *et al.* The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
35. Benavente, E. D. *et al.* Distinctive genetic structure and selection patterns in *Plasmodium vivax* from South Asia and East Africa. *Nat. Commun.* **12**, 1–11 (2021).
36. Cingolani, P., Platts, A., Wang le L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., Ruden, D, M. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) **6**(2), 80–92 (2012). PMID: 22728672
37. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL): An online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**, 127–128 (2007).
38. Manske, M. *et al.* Analysis of Plasmodium falciparum diversity in natural infections by deep sequencing. *Nature* **487**, 375 (2012).
39. Lee, S., Harrison, A., Tessier, N., Tavul, L., Miotto, O., Siba, P., Kwiatkowski, D., Müller, I., Barry, A. E., Bahlo, M. Assessing clonality in malaria parasites using massively parallel sequencing data, 2016, in preparation.
40. Gautier, M., Klassmann, A. & Vitalis, R. rehh 2.0: A reimplementation of the R package rehh to detect positive selection from haplotype structure. *Mol. Ecol. Resour.* **17**, 78–90 (2017).
41. Saetre, G.-P., Ravinet, M. *Evolutionary Genetic - Concepts, Analysis and Practice* (OUP Oxford, 2019).
42. Tang, K., Thornton, K. R. & Stoneking, M. A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biol.* **5**, e171 (2007).
43. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).

Acknowledgements

We wish to thank the communities and study participants of the Bijagós Archipelago, Guinea-Bissau.

Author contributions

This study was designed by SM. AL was Principal Investigator of the DTNMaPa mapping survey, during which dried blood spots were collected by the mapping team, led by AG and ETS. HV conducted DNA extraction of dried blood spots and qPCR for malaria positivity for dried blood spot samples, under the supervision of AL. SM conducted all *Plasmodium* DNA amplification under the supervision of SC. SM performed all bioinformatic analysis with the support of EM, AO, and JP, and under the supervision of SC and TGC. SM wrote the first draft of the manuscript, which was reviewed and contributed to by all authors. The final manuscript was read and approved by all authors.

Funding

SM is funded by Medical Research Council UK (Grant No. MR/N013638/1). EM is funded by MRC MR/X005895/1. TGC and SC are funded by Medical Research Council UK grants (Ref. no. MR/M01360X/1, MR/N010469/1, MR/R025576/1, MR/R020973/1, and MR/X005895/1). AL, HV and ETS are funded by Joint Global Health Trials Scheme (MRC, Wellcome Trust, UKRI, NIHR, Grant no. MR/S005013/1). SK is co-chair of the Guidelines Development Group for Antimalarials of the World Health Organization; views here are personal and do not represent those of the committee. AO is funded by a LSHTM PhD studentship funded by the WISE programme of MEXT. The DTNMaPa mapping survey was funding by MR/P023843/1 MRC Global Challenges Research Foundation.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-33176-1.

Correspondence and requests for materials should be addressed to S.M. or S.C.

Reprints and permissions information is available at www.nature.com/reprints.

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

Supplementary Information

Supplementary Table 1: Study isolates (n=795) across 10 different African countries.

Region	Country	Sample number (N)
Central Africa	Gabon	57
West Africa	Ivory Coast	70
	The Gambia	148
	Guinea-Bissau (Bijagós Archipelago*)	15
East Africa	Kenya	138
	Uganda	17
Southeast Africa	Malawi	149
	Madagascar	24
South Central Africa	Democratic Republic of the Congo	150
Horn of Africa	Eritrea	27

* 15 Bijagós samples were from 5 different islands: N=7 Bubaque, N=4 Canhabaque, N=2 Formosa, N=1 Orango, N=1 Soga.

Supplementary Table 2: F_{WS} metric assessing multiplicity of infection by assessing within-host heterozygosity (Hw) in comparison to local population heterozygosity (Hs). N = sample size.

Country	Mean F_{WS}	N ≥ 0.95	% N ≥ 0.95
Democratic Republic of Congo	0.820	56	37.3
Eritrea	0.951	22	78.6
Gabon	0.874	31	54.4
Gambia	0.908	100	67.6
Bijagós Archipelago, Guinea-Bissau	0.965	13	86.7
Ivory Coast	0.888	43	61.4
Kenya	0.863	76	55.1
Madagascar	0.943	19	79.2
Malawi	0.772	49	33.1
Uganda	0.788	7	41.2

Supplementary Table 3: Median identity by descent (IBD) fractions and ranges for each population of P. falciparum isolates, determined using hmmIBD software.

Country	Median IBD fraction	Range
Democratic Republic of Congo	0.001	0.000 – 0.213
Eritrea	0.118	0.027 – 0.320
Gabon	0.005	0.000 – 0.244
The Gambia	0.011	0.004 – 0.191
Guinea-Bissau (Bijagós)	0.040	0.001 – 0.217
Ivory Coast	0.003	0.002 – 0.089
Kenya	0.012	0.008 – 0.161

Madagascar	0.006	0.000 – 0.114
Malawi	0.003	0.001 – 0.169
Uganda	0.067	0.016 – 0.187

Supplementary Table 4: Integrated haplotype score (iHS) statistic scores for isolates from the Bijagós archipelago, N = 15. 7 SNPs within genes exhibited significantly high iHS scores. There were 5 positions within the SURFIN 13.1 gene; the position with the highest iHS value is displayed in the table below.

Chr	Gene	Number of positions	Position	iHS value	Log P value	Ref	Alt
1	Surface-associated interspersed protein 1.2 (SURFIN 1.2).	1	513013	-4.35	4.87	A	G
13	Surface-associated interspersed protein 13.1 (SURFIN 13.1).	5	106212	4.31	4.79	G	T

Supplementary Table 5: Candidate regions under selection from cross-population analysis using XP-EHH scores, between isolates from the Bijagós and samples from other African regions

Population comparison	Chromosomal region	Gene ID	Gene product
Bijagós vs Central Africa	Chr3, 120000-150000	PF3D7_0302500(CLAG3.1); PF3D7_0302600(ABCB4); PF3D7_0302200(CLAG3.2)	Cytoadherence linked asexual protein 3.1; ABC transporter B family member 4, putative; Cytoadherence linked asexual protein 3.2
Bijagós vs Central Africa	Chr 12, 2000000 - 2030000	PF3D7_1248700(N/A); PF3D7_1248900(RPT6)	Conserved Plasmodium protein, unknown function; 26S protease regulatory subunit 8, putative
Bijagós vs Central Africa	Chr 13, 90000-120000	PF3D7_1301700(CBP2)	CX3CL1-binding protein 2

Bijagós vs East Africa	Chr 3, 120000-150000	PF3D7_0302500(CLAG3.1); PF3D7_0302600(ABCB4); PF3D7_0302200(CLAG3.2)	Cytoadherence linked asexual protein 3.1; ABC transporter B family member 4, putative; Cytoadherence linked asexual protein 3.2
Bijagós vs East Africa	Chr 8, 410000 – 440000	PF3D7_0808200(PMX)	Plasmepsin X
Bijagós vs East Africa	Chr 13, 90000 - 120000	PF3D7_1301700(CBP2)	CX3CL1-binding protein 2
Bijagós vs South Central Africa	Chr 2, 120000 – 150000	PF3D7_0302500(CLAG3.1); PF3D7_0302600(ABCB4); PF3D7_0302200(CLAG3.2)	cytoadherence linked asexual protein 3.1; ABC transporter B family member 4, putative; cytoadherence linked asexual protein 3.2
Bijagós vs South Central Africa	Chr 11, 1170000 - 1200000	PF3D7_1130700(SMC1); PF3D7_1130400(RPT5)	structural maintenance of chromosomes protein 1, putative; 26S protease regulatory subunit 6A, putative
Bijagós vs South Central Africa	Chr 12, 2000000 - 2030000	PF3D7_1248700(N/A); PF3D7_1248900(RPT6)	conserved Plasmodium protein, unknown function; 26S protease regulatory subunit 8, putative
Bijagós vs South Central Africa	Chr 13, 90000 - 120000	PF3D7_1301700(CBP2)	CX3CL1-binding protein 2
Bijagós vs Southeast Africa	Chr3, 120000-150000	PF3D7_0302500(CLAG3.1); PF3D7_0302600(ABCB4); PF3D7_0302200(CLAG3.2)	Cytoadherence linked asexual protein 3.1; ABC transporter B family member 4, putative; Cytoadherence linked asexual protein 3.2
Bijagós vs Southeast Africa	Chr 10, 1380000 - 1410000	PF3D7_1035400(MSP3); PF3D7_1034900(MRScyt)	merozoite surface protein 3; methionine--tRNA ligase
Bijagós vs Southeast Africa	Chr 12, 2000000 - 2030000	PF3D7_1248700(N/A); PF3D7_1248900(RPT6)	Conserved Plasmodium protein, unknown function; 26S rotease regulatory subunit 8, putative
Bijagós vs Southeast Africa	Chr 13, 90000 - 120000	PF3D7_1301700(CBP2)	CX3CL1-binding protein 2
Bijagós vs West Africa	Chr3 120000-150000	PF3D7_0302500(CLAG3.1); PF3D7_0302600(ABCB4); PF3D7_0302200(CLAG3.2)	cytoadherence linked asexual protein 3.1; ABC transporter B family member 4, putative; cytoadherence linked asexual protein 3.2

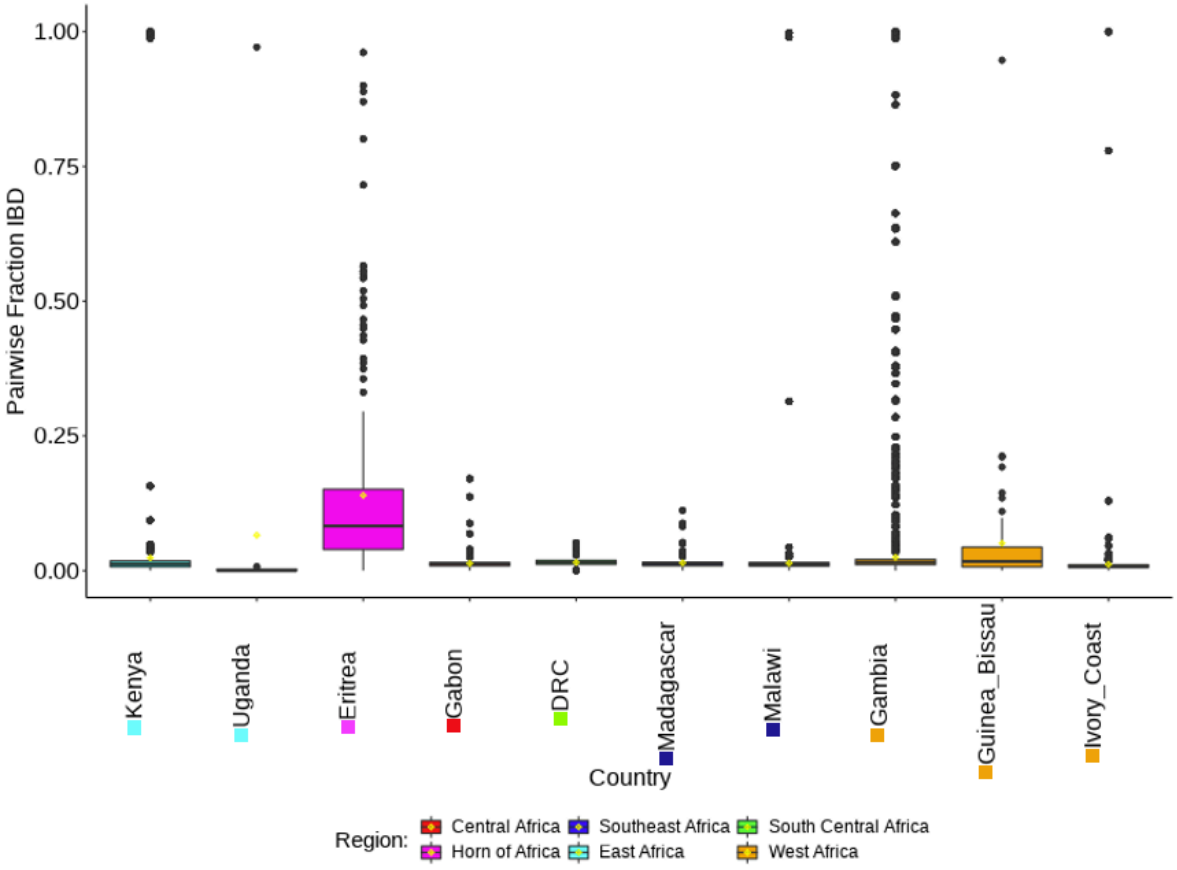
Bijagós vs West Africa	Chr 10, 1380000 - 1410000	PF3D7_1035400(MSP3); PF3D7_1034900(MRScyt)	merozoite surface protein 3; methionine--tRNA ligase
Bijagós vs West Africa	Chr 12, 2000000 - 2030000	PF3D7_1248700(N/A); PF3D7_1248900(RPT6)	conserved Plasmodium protein, unknown function; 26S protease regulatory subunit 8, putative
Bijagós vs West Africa	Chr 13, 90000 - 120000	PF3D7_1301700(CBP2)	CX3CL1-binding protein 2

Supplementary Table 6: Year and location of sample collection for publicly available genomes (malariagen.net)

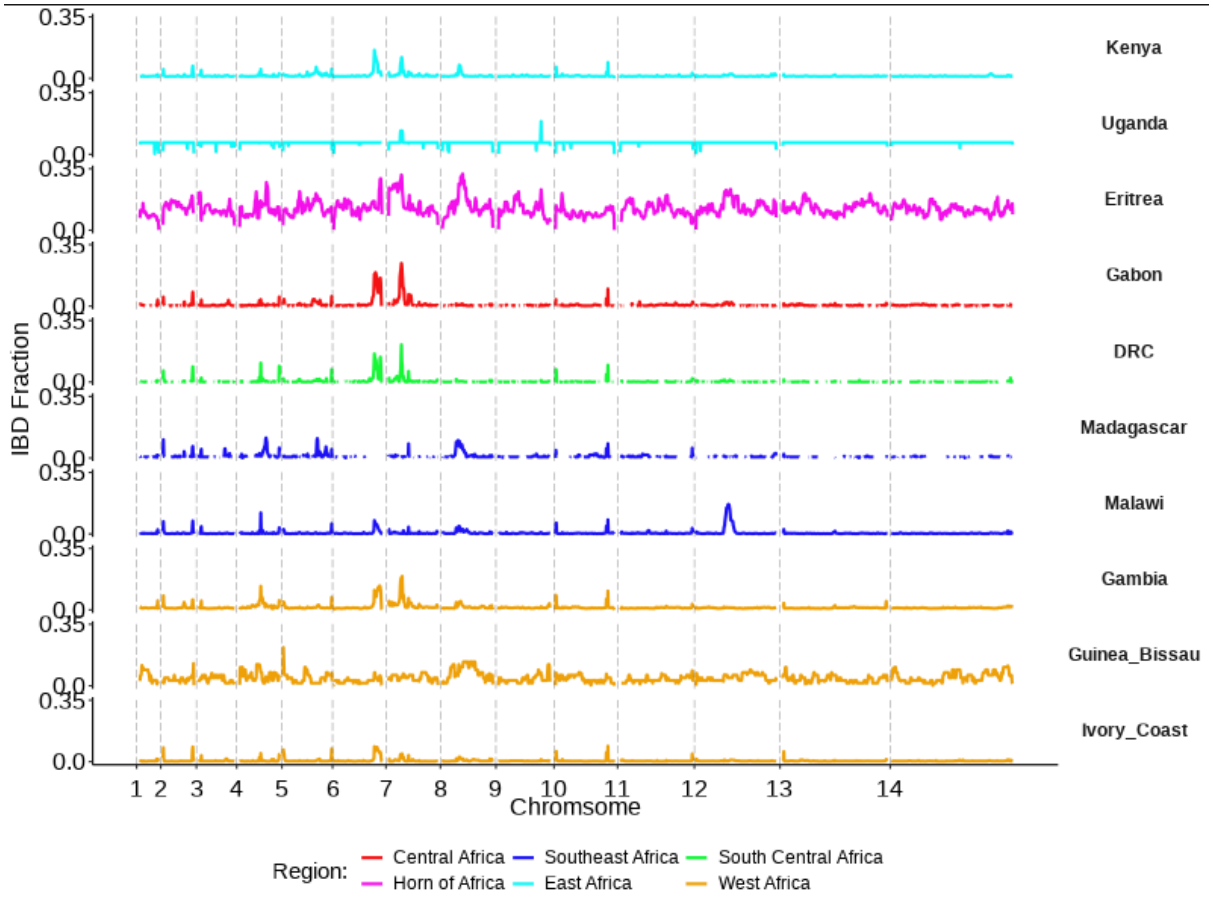
Country of origin	**Site**	**Year(s) of collection**
Democratic Rep. of Congo	Kinshasa	2011 and 2012
Eritrea	Not specified	Not specified
Gabon	Not specified	Not specified
The Gambia	Brikama and Basse	2008 and 2014
Ivory Coast	Abobo, Koumassi and Yopougon	2013
Kenya	Kilifi, Kosumu and Kombewa	2007 and 2014
Madagascar	Maevatanana, Antsohihy and Farafangana	2012, 2013 and 2014
Malawi	Chikwawa and Zomba	2011
Uganda	Apac and not specified	2009 and 2010

*Supplementary Table 7: Pfkelch13 mutations which could not be reported on reliably due to poor sequencing coverage (*https://apps.who.int/iris/bitstream/handle/10665/274362/WHO-CDS-GMP-2018.18-eng.pdf*).*

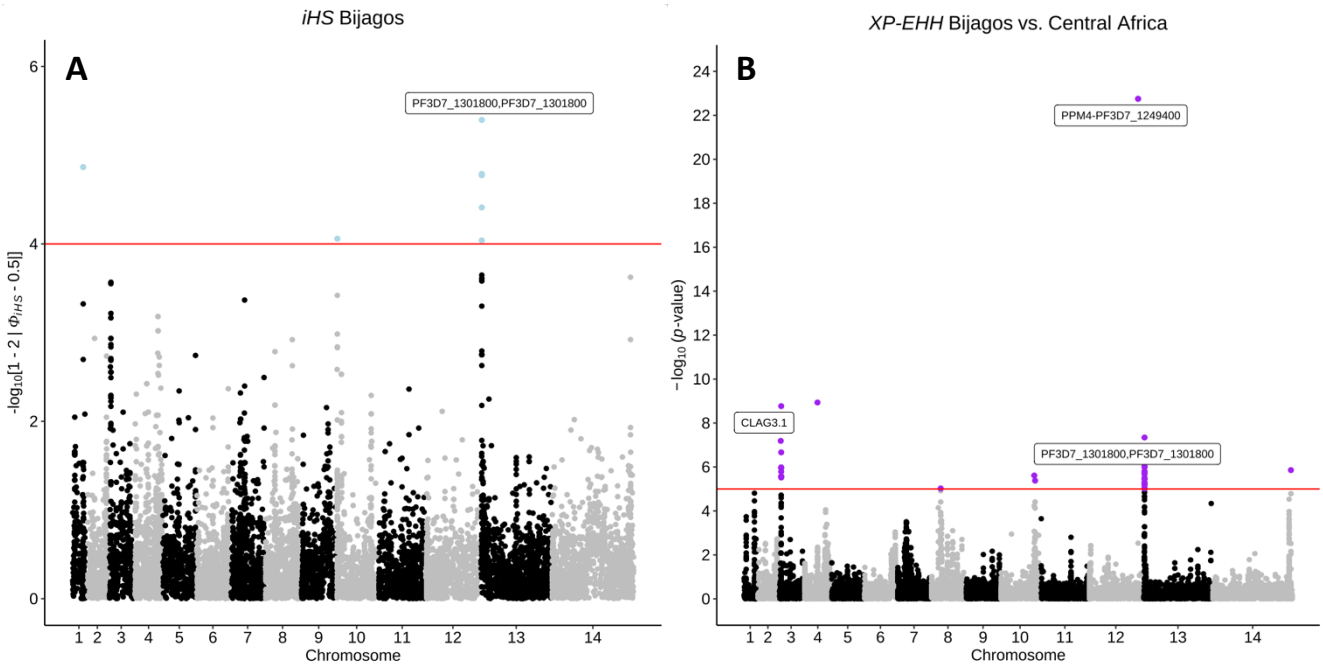
PfKELCH13 mutations which could not be reported on reliably due to poor sequencing coverage of the PfKELCH13 gene.
Validated: F446I, M476I, Y493H, R539T, I543T, P553L, R561H, C580Y, C469Y
Candidate/associated: P441L, G449A, C469F, A481V, P527H, N537I/D, G538V, V568G, P574L, F673I, A675V



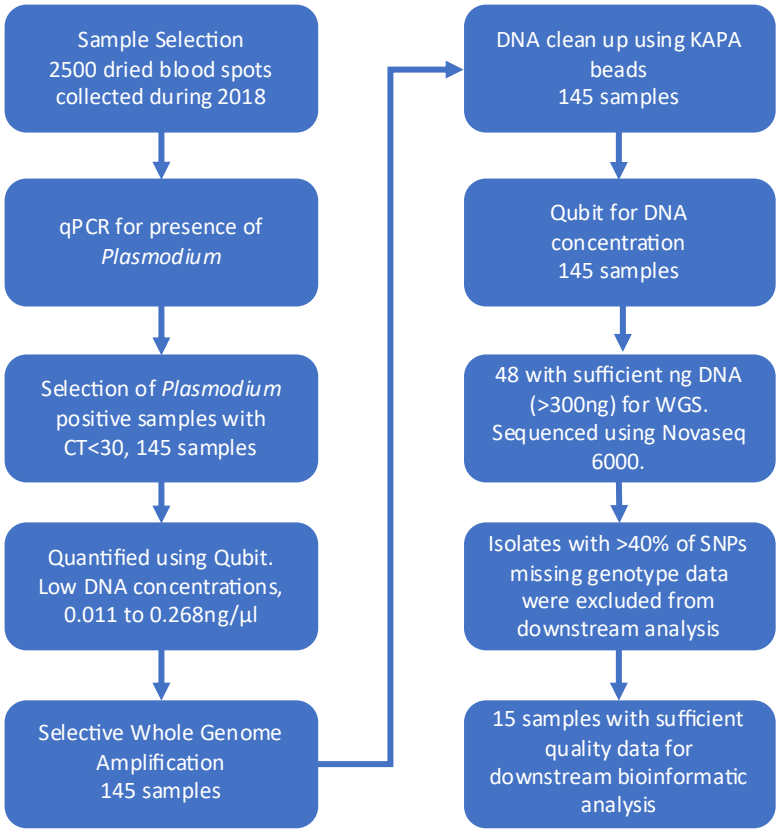
Supplementary Figure 1: A boxplot showing the pairwise identity by identity by descent (IBD) fraction for P. falciparum isolates from each of the countries analysed.



Supplementary Figure 2: Identity by descent (IBD) fractions along the genome for P. falciparum isolates from each of the countries analysed.



Supplementary Figure 3: Manhattan plots showing A) iHS scores for isolates from the Bijagós, Guinea-Bissau, B) XP-EHH scores from cross-population analysis between isolates from the Bijagós and isolates from West Africa, from The Gambia and Ivory Coast



Supplementary Figure 4: Filtering steps for P. falciparum isolates for inclusion in genomic analyses.

CHAPTER FOUR

Published paper: Genomic surveillance of *Anopheles* mosquitoes on the Bijagós Archipelago using custom targeted amplicon sequencing identifies mutations associated with insecticide resistance.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	lsh2005123	**Title**	Miss
First Name(s)	Sophie		
Surname/Family Name	Moss		
Thesis Title	Genomic epidemiology of *Plasmodium falciparum* and its *Anopheline* vectors in the context of malaria control on the Bijagós Archipelago of Guinea-Bissau		
Primary Supervisor	Dr Anna Last		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	Parasites & Vectors		
When was the work published?	4th January 2024		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	N/A		
Have you retained the copyright for the work?*	**Yes**	Was the work subject to academic peer review?	**Yes**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	

Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I devised the research question and designed the study. I conducted primer modification, multiplex amplicon PCR assays and post-PCR sample processing. I prepared samples for sequencing and conducted all bioinformatic analyses. I wrote the manuscript, which was commented on by my co-authors.

SECTION E

Student Signature	[redacted]
Date	2nd April 2024

Supervisor Signature	Anna Last Digitally signed by Anna Last [redacted] Date: 2024.05.22 18:20:04 +01'00'
Date	22nd May 2024

Moss *et al. Parasites & Vectors* (2024) 17:10
https://doi.org/10.1186/s13071-023-06085-5

Parasites & Vectors

RESEARCH Open Access

Genomic surveillance of *Anopheles* mosquitoes on the Bijagós Archipelago using custom targeted amplicon sequencing identifies mutations associated with insecticide resistance

Sophie Moss[1*], Elizabeth Pretorius[1], Sainey Ceesay[2], Harry Hutchins[1], Eunice Teixeira da Silva[3,4], Mamadou Ousmane Ndiath[2], Robert T. Jones[1], Hristina Vasileva[1], Jody Phelan[1], Holly Acford-Palmer[1], Emma Collins[1], Amabelia Rodrigues[4], Sanjeev Krishna[5,6,7], Taane G. Clark[1,8†], Anna Last[1†] and Susana Campino[1†]

Abstract

Background Insecticide resistance is reducing the efficacy of vector control interventions, consequently threatening efforts to control vector-borne diseases, including malaria. Investigating the prevalence of molecular markers of resistance is a useful tool for monitoring the spread of insecticide resistance in disease vectors. The Bijagós Archipelago (Bijagós) in Guinea-Bissau is a region of stable malaria transmission where insecticide-treated nets are the mainstay for malaria control. However, the prevalence of molecular markers of insecticide resistance in malaria vectors is not well understood.

Methods A total of 214 *Anopheles* mosquitoes were analysed from 13 islands across the Bijagós. These mosquitoes were collected using CDC light traps in November 2019, during the peak malaria transmission season. High-throughput multiplex amplicon sequencing was used to investigate the prevalence of 17 different molecular markers associated with insecticide resistance in four genes: *vgsc*, *rdl*, *ace1* and *gste2*.

Results Of the 17 screened mutations, four were identified in mosquitoes from the Bijagós: *vgsc* L995F (12.2%), N1570Y (6.2%) and A1746S (0.7%) and *rdl* A269G (1.1%). This study is the first to report the L995F knock-down resistance (kdr)-west allele in *Anopheles melas* on the Archipelago. An additional eight non-synonymous single-nucleotide polymorphisms were identified across the four genes which have not been described previously. The prevalences of the *vgsc* L995F and N1570Y mutations were higher on Bubaque Island than on the other islands in this study; Bubaque is the most populous island in the archipelago, with the greatest population mobility and connection to continental Guinea-Bissau.

Conclusions This study provides the first surveillance data for genetic markers present in malaria vectors from islands across the Bijagós Archipelago. Overall prevalence of insecticide resistance mutations was found to be low. However,

†Taane G. Clark, Anna Last and Susana Campino are joint last authors.

*Correspondence:
Sophie Moss
sophie.moss@lshtm.ac.uk
Full list of author information is available at the end of the article

the identification of the *vgsc* L995F and N1570Y mutations associated with pyrethroid resistance warrants further monitoring. This is particularly important as the mainstay of malaria control on the islands is the use of pyrethroid insecticide-treated nets.

Keywords Insecticide resistance, Vector control, Molecular monitoring, *Anopheles* mosquitoes

Background

Malaria still causes over 600,000 deaths every year, with around 95% of global cases occurring in Africa. Most of these cases are caused by the *Plasmodium falciparum* parasite, which is transmitted by *Anopheles* mosquitoes [1]. Vector control interventions are estimated to have reduced clinical malaria cases in Africa by over 600 million between 2000 and 2015 [2]. Of these interventions, insecticide-treated nets (ITNs) are estimated to have had the most impact, averting 68% of clinical malaria cases [2]. However, evolving insecticide resistance threatens the efficacy of vector control interventions. This presents a threat to the control of all vector-borne diseases, including malaria [3]. Unfortunately, malaria vectors from all WHO malaria endemic regions have evolved resistance to all four classes of insecticides: pyrethroids, organochlorines, organophosphates and carbamates [4]. Consequently, the WHO recommends that countries routinely monitor insecticide resistance in malaria vectors to inform effective vector control policy [4, 5].

The WHO standard methodology for evaluating phenotypic resistance in vectors uses WHO tube tests, WHO bottle bioassays or CDC bottle bioassays [6]. These methods are time-consuming, and direct comparison between the results of CDC bottle bioassays and the WHO methods is not recommended. The WHO recognises that monitoring genetic mutations associated with insecticide resistance is valuable to resistance surveillance [4]. Furthermore, understanding the molecular mechanisms behind the evolution of resistance requires greater knowledge of the genomics underpinning resistance. Pyrethroid and dichlorodiphenyltrichloroethane (DDT) resistance has been associated with target-site mutations in the voltage-gated sodium channel gene (*vgsc*), also known as knock-down resistance, or kdr, mutations. This includes L995F (kdr-west) [7] and L995S (kdr-east) [8] (position 1014 in *Musca domestica*). In addition, metabolic resistance to pyrethroids and DDT has been associated with mutations in the glutathione s-transferase epsilon 2 gene (*gste2*), including L119F [9] and I114T [10]. Resistance to organophosphate and carbamate insecticides has been associated with the G280S mutation (formerly known as G119S) in the acetylcholinesterase 1 gene (*ace1*) [11], and organophosphate resistance has been linked with *ace1* gene duplications [12]. Mutations in the gamma-aminobutyric acid (GABA) receptor (*rdl*), including A296G have been associated with resistance to the organochlorine insecticide dieldrin [13].

The Bijagós Archipelago (Bijagós) is located 70 km off the coast of Guinea-Bissau, West Africa and includes 88 islands and islets, of which 18 are inhabited year-round. Malaria remains a key public health problem on the Bijagós, where malaria mortality increased between 2015 and 2021 [1]. The main component of malaria control in Guinea-Bissau is the use of ITNs, which have been distributed widely since 2011. Of the distributed ITNs in the Bijagós, 90% are PermaNet® (Amabelia Rodrigues, personal communication), which are treated with deltamethrin, a pyrethroid insecticide. Coverage and use of ITNs on the Bijagós is reportedly high, with > 90% of 2018 survey participants reporting sleeping under a bed net [14]. No other vector control interventions are implemented on the islands, for example, indoor residual spraying, space spraying, or larviciding. The Bijagós Archipelago has a rainy season from June to October/November, and a dry season from December to May, with peak malaria transmission occurring in October and November. Two entomological surveys on the archipelago have previously been conducted in 2009 and 2017 [15, 16]. The major malaria vector has been identified as *Anopheles gambiae* sensu stricto (*A. gambiae* s.s.), which is the most prevalent Anopheline species during the onset of the rainy season in June/July. This contrasts to the end of the rainy season in November, and during the remainder of the dry season, where *Anopheles melas* has been found to be the predominant Anopheline species [16]. Guinea-Bissau is a particularly interesting location to study resistance evolution due to its high levels of *Anopheles gambiae* s.s.-*Anopheles coluzzii* hybridisation of > 20%, which may indicate the evolution of a novel hybrid form [17]. This unusually high level of recombination between species may provide the potential for the rapid spread of insecticide resistance alleles between *An. gambiae* s.s. and *An. coluzzii* populations.

The status of insecticide resistance in malaria vectors on the Bijagós Archipelago is largely unknown. Ant et al. [16] conducted an entomological survey of insecticide resistance on Bubaque island in 2017. This survey included CDC-bottle bioassays of *Anopheles gambiae* sensu lato (*A. gambiae* s.l.) mosquitoes, which suggested moderate resistance to α-cypermethrin and full susceptibility to permethrin. In the same study, Ant et al. [16] also

investigated the presence of the *vgsc* L995F and L995S mutations, and found moderate frequencies of L995F in *An. gambiae* s.s. (36%), *An. coluzzii* (35%), and *An. gambiae/An. coluzzii* hybrids (42%). However, the presence of other mutations associated with insecticide resistance has not previously been investigated. Furthermore, the study by Ant et al. [16] was conducted with mosquitoes from Bubaque island only.

There have been recent advances in monitoring molecular markers of insecticide resistance with multiplex amplicon sequencing [18]. This method provides a high-throughput alternative to previous methods, such as Sanger sequencing or whole-genome sequencing, and uses primer barcoding to enable targeted sequencing of multiple genetic loci in many samples simultaneously, minimising required resources and lowering sequencing costs [18, 19]. In the present study, we aimed to conduct high-throughput molecular monitoring of insecticide resistance on the Bijagós, using mosquito samples collected from islands across the archipelago. Custom-targeted amplicon sequencing was used to investigate a panel of molecular markers associated with insecticide resistance across the *ace1, gste2, rdl* and *vgsc* genes. This approach was conducted using multiplex PCR assays and dual-indexing barcodes [18] to enable high-throughput sequencing of multiple loci across multiple samples.

Methods

Mosquito sampling and speciation

Mosquitoes were collected using indoor CDC miniature light traps (model 512; John W. Hock Company, Gainesville, FL, USA) using previously described methodology [20]. Mosquitoes were separated by genus, and *Anopheles* mosquitoes were morphologically identified using previously described keys [21]. All collected specimens were found to belong to *Anopheles gambiae* s.l. A subsample was then sent to the Medical Research Council Unit The Gambia (MRCG) at London School of Hygiene & Tropical Medicine for molecular analysis. Mosquito DNA was extracted and each sample was identified to species using PCR–restriction fragment length polymorphism (RFLP) [22].

DNA extraction

DNA was extracted using the QIAamp® 96 DNA QIAcube® HT KIT (Qiagen, Hilden, Germany) with the QIA cube Extractor Robot, following the manufacturer instructions. DNA was eluted in 80 µl AE buffer and stored at − 20 °C.

Species identification

Mosquitoes were identified to species using PCR–RFLP, based on the protocol of Fanello et al. [22] which uses primers to amplify the intergenic spacer (IGS) region to differentiate the members of the *An. gambiae* complex. The following primers were used: Universal F: GTG TGCCCCTTCCTCGATGT; *An. gambiae* R: CTGGTT TGGTCGGCACGTTT; *Anopheles arabiensis* R: AAG TGTCCTTCTCCATCCTA; *An. melas* R: TGACCA ACCCACTCCCTTGA. Amplified PCR products were then digested using the HhaI enzyme to differentiate *An. gambiae* s.s. and *An. coluzzii* specimens. The band sizes of the PCR products were visualised by electrophoresis using the QIAxcel capillary electrophoresis system (Qiagen). The band sizes of the PCR products were analysed to distinguish species: *An. gambiae* s.s. (257 and 110 bp), *An. arabiensis* (292 bp), *An. melas* (435 bp), *An. coluzzii* (367 bp) and *An. coluzzii/An. gambiae* s.s. hybrid (257, 110 and 367 bp).

Primer design

Primers were adapted from Campos et al. [18] to amplify seven regions of the *Anopheles* genome covering four different genes and 17 different single-nucleotide polymorphisms (SNPs) previously associated with insecticide resistance. Assays were designed to generate amplicons of approximately 500 bp. Primers were modified where possible to improve binding to the *An. melas* genome, using sequences downloaded from VectorBase (www.vectorbase.org). The primers used to amplify these regions are provided in Additional file 1: Table S2). The reverse primer for *gste2* was modified from Campos et al. [18] from TTCCAAATGCTTCCAAATTT to GGCTAGCAC AAACTTGC. To enable multiplex amplicon sequencing, primers were further adapted in line with previously published methods to incorporate barcodes for multiplexing [18]. Briefly, 5′ barcode tags (8 bp long) were added to each forward and reverse primer (Additional file 1: Table S3). A total of 10 unique forward barcodes and 10 unique reverse barcodes were used. To enable multiplex sequencing, each mosquito sample was assigned a unique barcode combination prior to PCR amplification, resulting in the sample specific barcode being incorporated into the amplified PCR product.

Multiplex PCR reactions

The Thermo Fisher Scientific (Waltham, MA, USA) multiplex primer analyzer was used to identify which primers were most suitable for multiplexing for amplicon generation. This resulted in multiplex 1 (*ace1* and *vgsc-3*), multiplex 2 (*rdl* and *vgsc-2*), multiplex 3 (*vgsc-4* and *gste2*) and a simplex reaction (*vgsc-1*). Each PCR reaction had a final reaction volume of 25 µl, containing 0.25 µl Q5 Hot-start DNA polymerase (New England Biolabs Ltd., Hitchin, UK), 5 µl of Q5 buffer (New England Biolabs Ltd.), 1 µl of DNA template (2–10 ng), 0.5 µl of 10 mM dNTPs, an

average of 0.63 µl of each forward and reverse primer at 10 µM and 15.75 µl of nuclease-free H_2O. PCR amplifications were conducted using the following reaction conditions: hot-start activation of Q5 Hot-start DNA polymerase for 30 s at 98 °C; followed by 35 cycles of denaturation at 98 °C for 10 s, annealing at 57 °C for 60 s and elongation 72 °C for 60 s; with a final elongation step of 72 °C for 2 min.

Amplicon purification and next generation sequencing

Following the gene-specific multiplex PCR reactions, PCR products containing amplified amplicons were run in 1% agarose gels to detect the presence and the size of PCR products. The PCR products from different mosquito samples containing different barcodes were pooled. Pools were purified using KAPA pure beads (Roche Diagnostics, Indianapolis, IN, USA) to remove excess primers and PCR reagents at a ratio of 0.8:1 (µl of beads to µl of DNA). Purified pools were quantified using a Qubit before adjustment to a final concentration of 20 ng/µl in a final volume of 25 µl. Pooled samples were sent for sequencing at Genewiz GmbH (Leipzig, Germany) on the Illumina MiSeq platform paired-end (2 × 250 bp) configuration (Illumina Inc., San Diego, CA, USA), and a minimum of 50,000 reads were obtained per pool of amplicons. To obtain high sequencing coverage, a maximum of 200 amplicons were pooled per experiment, at a cost of < US$0.5 per amplicon.

Bioinformatics analysis

Raw pooled FASTA sequences were de-multiplexed to separate sequences from individual mosquitoes through identification of sample-specific barcodes. Sequences were then processed to identify SNPs and insertions/deletions (INDELs) using a custom python script (https://github.com/LSHTMPathogenSeqLab/amplicon-seq). FASTA sequences were trimmed using Trimmomatic software [23] (version 0.39, using the parameters LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20 MINLEN:36) to remove poor quality sequences. Paired-end reads were mapped to the *An. gambiae* reference genome (Anopheles_gambiae.AgamP4.dna.toplevel.fa) using BWA-MEM software (version 0.7.17-r1188, default parameters) to produce a BAM file for each mosquito. SNPs were called with Freebayes (version v1.3.6, haplotype-length-1) and GATK's HaplotypeCaller (version 4.1.4.1, default parameters). Variants called by both variant callers were combined and filtered using BCFtools (version 1.16) for a minimum allele depth of 10, minimum read depth of 30 and phred score of > 30 per base. SnpEff software (version 5.1d) was used to annotate variants. Species-specific variants were removed to ensure these did not confound the analysis. Variants were genotyped based on the percentage of alternative allele reads within the total depth coverage: (i) homozygous reference (< 25% alternate allele calls); (ii) heterozygous (25–75% alternate allele calls); and (iii) homozygous alternate (> 75% alternate allele reads), as per Campos et al. [18]. SNPs were retained for analysis if they were present in more than one sample.

Statistical tests

When comparing the allelic frequencies of alleles between island groups or species, statistical significance was investigated using a Chi-square (χ^2)-test in R (df = 1) (version 4.1.0; R Foundation for Statistical Computing, Vienna, Austria).

Results

Mosquito species and geographic location

Mosquitoes were collected during a cross-sectional survey conducted on 16 islands of the Bijagós Archipelago during the peak malaria transmission season in November 2019. Of 409 mosquitoes collected, 82.2% were *An. melas,* 9.5% were *An. coluzzii-gambiae* hybrids, 6.82% were *An. gambiae* s.s. and 1.47% were *An. coluzzii* (Soubrier et al., unpublished data). A subset of $n = 214$ *Anopheles* mosquitoes were used for analysis in this study. These mosquitoes were collected from 13 islands across the archipelago: Bubaque, Canhabaque (Roxa), Canogo, Caravela, Formosa, Tchedega, Meneque, Nhago (Ponta), Orangozinho, Rubane, Soga, Uno and Uracane (Fig. 1). The 214 mosquitoes included in this study are listed by the island on which they were collected and by species in Table 1. *Anopheles melas* was the most abundant species collected (113/214), with the majority of *An. melas* collected from Bubaque, Tchedega and Soga (70/113); these three islands accounted for 46% of all samples collected (99/214).

Amplicon sequencing coverage and detection of genetic variants

The presence of SNPs was investigated in each target gene. The SNPs targeted for amplification are listed in Additional file 1: Table S1. Amplicons were designed to cover these target SNPs and the surrounding gene region with a length averaging 500 bp, enabling the detection of SNPs in the neighbouring gene region (Additional file 1: Table S2). The average coverage obtained for each amplicon was very high (range: 210.43 to 632.91). This depth of coverage led to the identification of 155 high-quality SNPs across the seven amplicons. Of the 155 SNPs detected, 115 (74.2%) were in exons (12 non-synonymous SNPs and 103 synonymous SNPs), 28 (18.1%) were in introns, nine were upstream of exons and three were in splice regions (Table 2).

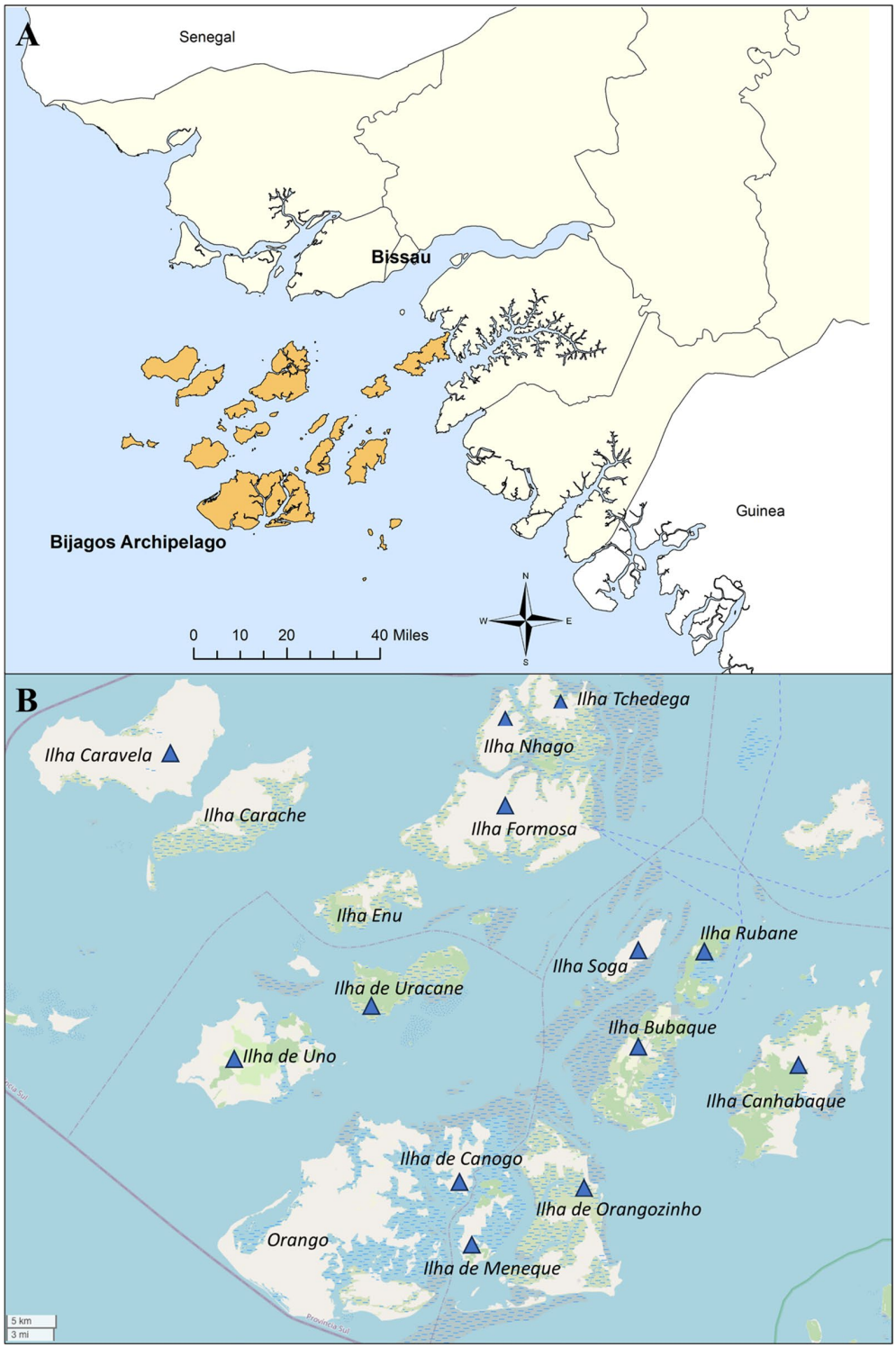


Fig. 1 Mosquito sample collection sites. **a** Location of Bijagós Archipelago, created using ArcGIS ArcMap 10.8.1. **b** Mosquitoes were collected from the 13 islands labelled with blue triangles. Map from OpenStreetMap 2023-05-06

Table 1 Mosquitoes included in analysis and their island of origin in the Bijagós Archipelago

Anopheles species	Island													*Total*
	Bubaque	Canhabaque (Roxa)	Canogo	Caravela	Formosa	Tchedega	Meneque	Nhago (Ponta)	Orangozinho	Rubane	Soga	Uno	Uracane	
An. melas	24	2	1	10	0	24	2	7	0	5	22	12	4	*113*
An. gambiae sensu stricto	3	0	2	0	6	3	4	0	11	5	2	4	0	*40*
An. coluzzii	2	0	1	0	1	0	1	0	4	1	0	1	0	*11*
An. coluzzii/An. gambiae hybrids	12	3	5	1	4	3	1	0	9	3	4	4	1	*50*
Total	41	5	9	11	11	30	8	7	24	14	28	21	5	214

Species included *An. melas*, *An. gambiae* s.s., *An. coluzzii*, and *An. coluzzii–An. gambiae hybrids*

Table 2 Distribution of single-nucleotide polymorphisms detected in each gene

Gene[a]	Average coverage of amplicon (bp)	Synonymous (*n*)	Non-synonymous (*n*)	Intron (*n*)	Splice region (*n*)	Upstream gene variant (*n*)	Total (*n*)
vgsc-DIS6	600.81	1	2	16	0	0	19
vgsc-DIIS6	323.46	3	1	2	0	0	6
vgsc-DIIIS6	597.83	5	1	5	1	0	12
vgsc-DIVS5	210.43	5	1	5	1	0	12
rdl	253.54	2	1	0	0	0	3
gste2	336.72	22	4	0	1	9	36
ace1	632.91	65	2	0	0	0	67
Total		103	12	28	3	9	155

[a] Four amplicons are in the voltage-gated sodium channel gene (*vgsc* I-IV, D = domain, S = subunit). One amplicon is in each of the gamma-aminobutyric acid (GABA) receptor gene (*rdl*), glutathione s-transferase epsilon 2 gene (*gste2*) and acetylcholinesterase 1 gene (*ace1*)

The position and frequency of detected non-synonymous variants is summarised Table 3.

Two non-synonymous SNPs were identified in the *vgsc*-DIS6 amplicon: Phe389Leu (F389L) and Glu429Lys (E429K). F389L was found at an allelic frequency of 0.6%, and E429K was found at an allelic frequency of 20.9%. These SNPs have not previously been associated with insecticide resistance. One non-synonymous SNP was identified in the *vgsc*-DIIS6 amplicon, namely the Leu995Phe (L995F) mutation, which is also known as the kdr-west allele. This mutation was found at an allelic frequency of 12.2% in the total population of mosquitoes sampled, and has previously been associated with resistance to pyrethroids and DDT [7]. One non-synonymous mutation, Asn1570Tyr (N1570Y), was identified in the *vgsc*-DIIIS6 amplicon, at an allelic frequency of 6.2%. The N1570Y mutation has been associated with increased levels of pyrethroid resistance when associated with L995F [24]. One non-synonymous SNP was identified in the *vgsc*-DIVS5 amplicon, Ala1746Ser (A1746S), with an allelic frequency of 0.7%. A1746S has previously been detected in *Anopheles gambiae* s.l. mosquitoes in Guinea and Ivory Coast [18, 25], and has been identified as a variant with a potential functional role in pyrethroid resistance [25]. One non-synonymous SNP, Ala296Gly (A296G), was identified in the *rdl* amplicon at an allelic frequency of 1.1%. A296G has previously been associated with resistance to dieldrin [13]. Two non-synonymous mutations were identified in the *ace1* amplicon, namely Pro203Gln (P203Q) at 0.6% and Glu323Asp (E323D) at

Table 3 Position and frequency of detected non-synonymous single-nucleotide polymorphisms in mosquitoes sampled from the Bijagós Archipelago

Chromosome	Amplicon	Position	*N*[a]	Mutation	Genotype frequencies (%)			Alternate allele frequency (%)
					Homozygous (reference)	Heterozygous	Homozygous alternate	
2L	*vgsc*-I	2391191	159	F389L	98.7	1.3	0.0	0.6
		2391309	160	E429K	58.1	41.9	0.0	20.9
	vgsc-II	2422652	135	L995F[b*]	82.2	11.1	6.7	12.2
	vgsc-III	2429745	170	N1570Y[b]	88.8	10.0	1.2	6.2
	vgsc-IV	2430424	139	A1746S[b]	98.6	1.4	0.0	0.7
2L	*rdl*	25429236	89	A296G[b]	97.8	2.2	0.0	1.1
2R	*ace1*	3491844	175	P203Q	98.9	1.1	0.0	0.6
		3492205	175	E323D	98.3	1.1	0.6	1.1
3R	*gste2*	28597858	182	I187F	98.9	1.1	0.0	0.5
		28597896	182	P174L	48.4	2.7	48.9	50.3
		28597905	182	G171D	89.6	9.3	1.1	5.8
		28597956	189	T154S	69.3	21.7	9.0	19.8

[a] N is the total number of alleles called for each position

[b] This mutation was one of the 17 targeted insecticide resistance single-nucleotide polymorphisms

1.1%. These two SNPs have not previously been associated with insecticide resistance. Four non-synonymous SNPs were identified in the *gste2* amplicon: Ile187Phe (I187F), Pro174Leu (P174L), Gly171Asp (G171D) and Thr154Ser (T154S); these mutations were found at allelic frequencies of 0.5%, 50.3%, 5.8%, and 19.8%, respectively, and have not previously been associated with insecticide resistance.

Insecticide resistance SNP distribution per island

Overall, four of the targeted SNPs (Additional file 1: Table S1) were identified in the samples collected on the Bijagós, including three mutations in the *vgsc* gene (L995F, N1570Y and A1746S) and one mutation in the *rdl* gene (A296G). The frequency of these mutations per island was investigated (Table 4). The L995F mutation was found at an overall frequency across the archipelago of 12.2% and was present in mosquitoes from 10 of the 13 islands. The N1570Y mutation was present at a frequency of 6.2% (present in 11/13 islands) and the A1746S mutation was present at a frequency of 0.7% (present in 2/13 islands). The *rdl* A269G mutation had a frequency of 1.1% (present in 2/13 islands).

Five of the islands were represented by > 20 mosquitoes: Bubaque ($n=41$), Tchedega ($n=30$), Orangozinho ($n=24$), Soga ($n=28$) and Uno ($n=21$). The allelic frequencies of the four insecticide resistance mutations identified were mapped to island location (Fig. 2), and differences assessed. Bubaque had a significantly higher allelic frequency of *vgsc* L995F than Tchedega ($P=0.023$), Soga ($P=0.001$) and Uno ($P=0.008$), and a slightly higher L994F frequency than Orangozinho, but this increase was not significant ($P=0.110$). Bubaque had a significantly higher allelic frequency of *vgsc* N1570Y than Tchedega ($P=0.016$) and Uno ($P=0.031$). Bubaque had a slightly higher frequency of *vgsc* N1570Y than Orangozhino and Soga, but these differences were not significant. There was no statistical significance between the allelic frequency of the *vgsc* A1746S mutation between islands. The *rdl* A296G mutation was found on Formosa and Tchedega islands only.

The number of different samples which had one or multiple insecticide resistance SNPs were investigated (Fig. 3). A small number of mosquitoes were found to have more than one resistance-associated mutation. This included five mosquitoes from Bubaque island, which had two *vgsc* mutations (L995F and N1570Y, or L995F and A1746S), and one mosquito from Orangozinho which had two *vgsc* mutations (L995F and N1570Y). All other mosquitoes had only one of the four identified insecticide resistance mutations.

Table 4 Allelic frequency of insecticide resistance single-nucleotide polymorphisms per island

Island	N^a	Allelic frequency of insecticide resistance mutation (%)[b]			
		vgsc L995F	*vgsc* N1570Y	*vgsc* A1746S	*rdl* A296G
Bubaque	41	26.0	11.0	1.9	–
Canhabaque	5	33.0	12.5	–	–
Canogo	9	14.3	6.3	–	–
Caravela	11	8.3	5.6	–	–
Formosa	11	31.3	16.7	–	5.6
Tchedega	30	3.3	2.1	–	10.0
Meneque	8	20.0	14.3	–	–
Nhago	7	–	–	–	–
Orangozinho	24	7.7	2.9	–	–
Rubane	14	11.1	4.2	5.6	–
Soga	28	–	2.2	–	–
Uno	21	–	–	–	–
Uracane	5	12.5	12.5	–	–

[a] Total number of mosquitos sampled from each island

[b] The proportion of alleles that harbour any of the four identified target mutations

Distribution of insecticide resistance SNPs per mosquito species

The distribution of insecticide resistance SNPs between different *Anopheles* species samples was investigated (Table 5). The total sample set contained 214 mosquitoes: 113 *An. melas*, 40 *An. gambiae* s.s., 11 *An. coluzzii* and 50 *An. coluzzii-An. gambiae* hybrid mosquitoes.

The *vgsc* L995F mutation was found in 24 mosquitoes, and across species (3 *An. melas*, 7 *An. gambiae* s.s., 2 *An. coluzzii*, 12 *An. coluzzii-An. gambiae* hybrids). The *vgsc* N1570Y mutation was found in 18 mosquitoes (2 *An. melas*, 4 *An. gambiae* s.s., 2 *An. coluzzii* and 10 *An. coluzzii-gambiae* hybrids). The *vgsc* mutation A1746S was found in two mosquitoes (one *An. melas* and one *An. gambiae* s.s). Similarly, the *rdl* A296G mutation was found in only two mosquitoes (1 *An. gambiae* s.s. and 1 *An. coluzzii-gambiae* hybrid). Interestingly, *An. melas* had significantly lower allelic frequencies of *vgsc* L995F and N1570Y than the other mosquito species ($P<0.05$). There was no statistically significant difference between the allelic frequencies of different species for *vgsc* A1746S or *rdl* A296G.

Discussion

Vector control interventions, in particular ITNs, are the most effective method of reducing deaths due to malaria to date [2]. The emergence and spread of insecticide resistance is a growing threat to the utility of these interventions. Monitoring the prevalence of genomic polymorphisms associated with insecticide

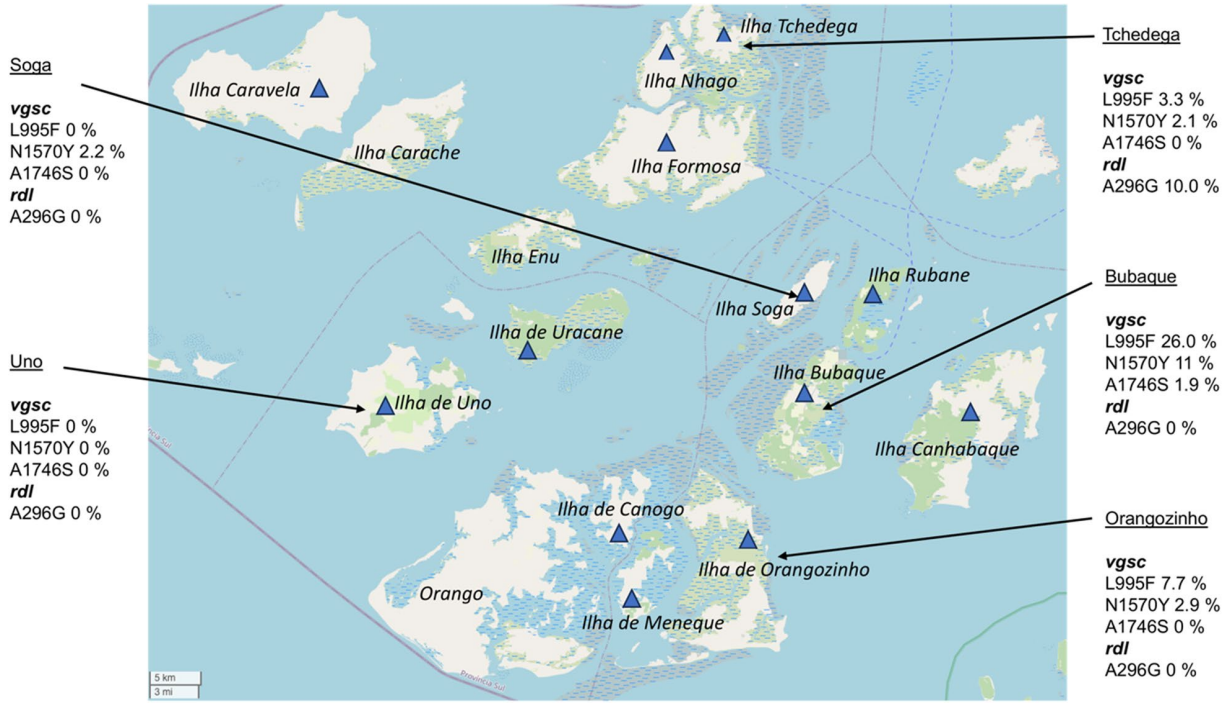


Fig. 2 Allelic frequencies of the insecticide resistance mutations identified, mapped to island geographic location. Only islands with a sample size > 20 sampled mosquitoes were included in this analysis. *rdl*, Gamma-aminobutyric acid receptor gene; *vgsc*, voltage-gated sodium channel gene

resistance is an important tool for monitoring the distribution and spread of insecticide resistance, and to complement phenotypic bioassays of resistance [18, 26, 27]. Monitoring polymorphisms can provide early warning signals to disease control programmes that resistance markers are changing. In turn, this information may inform evidence-based decisions in vector control policy. This is particularly useful in situations where resistance is present at low frequency and may be difficult to detect using phenotypic bioassays.

Multiplex amplicon sequencing is a high-throughput, low-cost and resource-efficient method of investigating the frequency of SNPs in sample populations. This technique has been used widely to investigate the frequency of insecticide resistance mutations in mosquito populations [18, 26–28]. We employed this technique to conduct the first investigation of multiple genomic markers of insecticide resistance across islands of the Bijagós Archipelago. We investigated the presence of genetic markers, including 17 SNPs associated with insecticide resistance in the *vgsc, gste2, rdl* and *ace1* genes. Four of the 17 targeted SNPs were identified in the Bijagós mosquito population. These were *vgsc* L995F and N1570Y, which are associated with DDT and pyrethroid resistance [7, 8], *vgsc* A1746S, which has a potentially functional role in pyrethroid resistance [25], and the *rdl* A296G mutation associated with dieldrin resistance [13].

The *vgsc* L995F (kdr-west) mutation was found at an allelic frequency of 12.2% across the archipelago and a frequency of 26.0% on the island of Bubaque. This was lower than previously identified by Ant et al. in a 2017 survey on Bubaque island, which found frequencies of 35% for *An. coluzzii,* 36% for *An. gambiae* s.s. 42% for *An. gambiae* s.s.-*An. coluzzii* hybrids [16]. However, our study methods had the capacity to distinguish between homozygous and heterozygous calls, which is likely to explain this discrepancy in allele frequency. A separate survey in Guinea-Bissau in 2018 by Silva et al. [29] also found higher frequencies of the L995F mutation (31.29%) in the nation's capital, Bissau. We found the *vgsc* N1570Y mutation at an allelic frequency of 6.2% across the archipelago. This mutation has been referred to as the 'super-kdr' mutation and is predicted to intensify L995F-mediated pyrethroid resistance and/or compensate for the deleterious fitness effects of L995F [24]. The N1570Y mutation has been found previously in Bissau at a similar frequency of 6.62% [29]. The *vgsc* A1746S mutation, which has previously been identified in Guinea and Ivory Coast [25], was found in the present study at 0.7% allelic frequency. Two additional non-synonymous SNPs were found in the *vgsc* gene, Phe389Leu (F389L) (0.6%) and

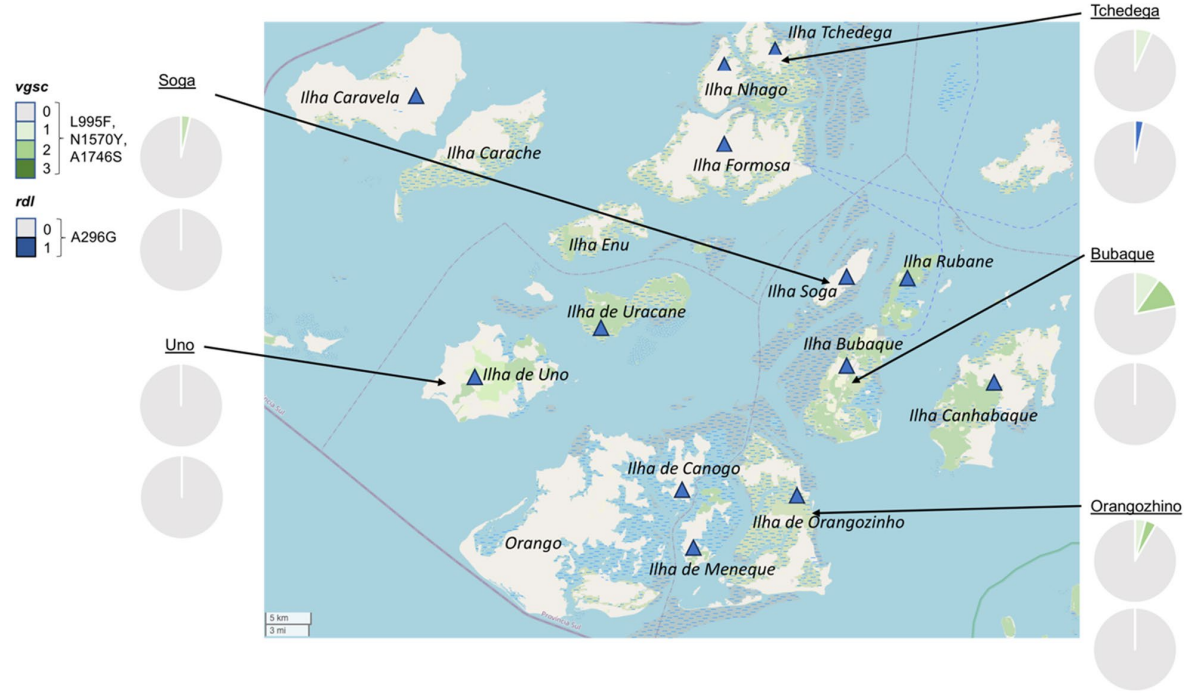


Fig. 3 Distribution of cumulative insecticide resistance mutations across five islands on the Bijagós Archipelago. Only regions with > 20 sampled mosquitoes are shown (Bubaque, Orangozinho, Tchedega, Soga and Uno). The top pie chart for each island represents the percentage of samples with 0, 1, 2 or 3 of the targeted *vgsc* mutations, with the *vgsc* mutations shown in green. The lower pie chart for each island represents the percentage of samples with 0 or 1 *rdl* mutation, with the *rdl* mutations shown in blue. *rdl*, Gamma-aminobutyric acid receptor gene; *vgsc*, voltage-gated sodium channel gene

Table 5 Distribution of the four identified insecticide resistance single-nucleotide polymorphisms by *Anopheles* species

Anopheles species	*N*[a]	Allelic frequency of insecticide resistance mutation (%)			
		vgsc L995F	*vgsc* N1570Y	*vgsc* A1746S	*rdl* A296G
An. melas	113	2.14	1.12	0.76	0
An. gambiae sensu stricto	40	14.29	6.45	1.67	1.67
An. coluzzii	11	25.0	14.29	–	–
An. coluzzii-An. gambiae hybrids	50	21.1	13.1	–	1.25

[a] *N* is the total number of mosquitoes collected from each species

E429K (20.9%); these mutations have not previously been reported.

The A296G mutation in the *rdl* gene associated with resistance to dieldrin [13] was found at an allelic frequency of 1.1%. No additional non-synonymous SNPs were identified in this gene. There were two non-synonymous SNPs identified in the *ace1* gene: P203Q (0.6%) and E323D (1.1%); these SNPs have not been reported previously. The G280S mutation (formerly G119S) is associated with resistance to carbamates and organophosphates [30], and was not found in the Bijagós mosquito population. This mutation was also not found by Silva et al. [29] in Bissau, but it was found at other sites on the mainland, including Gabu, the easternmost region of Guinea-Bissau. Four non-synonymous mutations were found in the *gste2* amplicon which have not previously been reported: I187F (0.5%), P174L (50.3%), G171D (5.8%) and T154S (19.8%). The I114T and L119V mutations associated with pyrethroid [9] and DDT resistance [9, 10], and the F120L mutation associated with DDT resistance [31], were not identified in the Bijagós mosquito population. A notably larger number of SNPs were

identified in the *ace1* and *gste2* amplicons than in the *vgsc* or *rdl* amplicons. This may be due to copy number variations, which are well-reported in the *An. gambiae* complex to result in multiple copies of *ace1* and *gste2* alleles [32, 33]. The presence of multiple copies may allow the accumulation of mutations in one copy of an allele without leading to deleterious fitness effects due to the presence of additional allele copies, enabling the accumulation of greater numbers of mutations over time.

We compared the frequency of insecticide resistance SNPs on five islands: Bubaque, Tchedega, Orangozinho, Soga and Uno. The frequency of L995F was significantly higher on Bubaque Island than on Tchedega, Soga or Uno, and slightly higher than on Orangozinho. Bubaque is the most populated island, home to around 28% of the total population of the archipelago (Ministério da Saúde Pública, Guinea-Bissau). It is also the site of the main ferry port to Bissau. Therefore, it is likely that there is a higher density of pyrethroid ITNs on Bubaque, resulting in greater selection pressure for the L995F mutation. Furthermore, the ferry link to mainland Guinea-Bissau provides a regular opportunity for 'hitch-hiking' mosquitoes carrying resistance alleles to be transported to Bubaque. This hitch-hiking may result in a larger number of resistance alleles in the Bubaque mosquito population for positive selection to act upon. The sample size of mosquitoes from each island was limited and should be increased in future work to enable better comparisons of allele frequency between islands.

We compared the frequency of insecticide resistance SNPs between species and found that *An. melas* had significantly lower allelic frequencies of *vgsc* L995F and N1570Y than *An. gambiae* s.s., *An. coluzzii* and *An. coluzzii-An. gambiae* hybrids ($P < 0.05$). Our study identified the L995F (kdr-west) allele in *An. melas* for the first time on the archipelago. *Anopheles melas* is described as an opportunistic vector which is highly anthropophilic and zoophilic, whereas *An. gambiae* s.s. has been described as a mostly anthropophilic species with a preference for human hosts [34]. This suggests that *An. melas* may be more likely to feed on animal hosts and encounter fewer ITNs than *An. gambiae* s.s. In addition, *Anopheles melas* also have higher tolerance for saline environments and are often found in salt marshes or mangroves [34]. This may also mean that *An. melas* encounter ITNs less frequently. In this case, *An. melas* would be under weaker selection pressure for *vgsc* resistance-conferring mutations than the other species. Furthermore, the extreme rates of *An. gambiae-An. coluzzii* hybridisation found on the Bijagós (> 20%) may facilitate increased propagation of *vgsc* mutations between *An. gambiae* s.s. and *An. coluzzii*, enabling these mutations to reach higher frequencies within these populations. However, a relatively small number of mosquitoes were sampled overall in this study, including only $n = 11$ *An. coluzzii* samples, making it difficult to draw definitive conclusions from these inter-species comparisons.

Overall, this study demonstrates the usability of high-throughput multiplex amplicon sequencing in screening insecticide resistance mutations in *Anopheles gambiae* s.l. mosquitoes. Four resistance-associated SNPs were identified in the Bijagós mosquito population. An additional eight non-synonymous SNPs were identified which have not previously been described. Phenotypic assays would be needed to determine whether these SNPs contribute to resistance. Importantly, insecticide resistance is only one component of effective vector control. Coverage, uptake and durability of ITNs, along with longevity of the insecticide, are all important factors which need to be measured [35].

Limitations of this study include the relatively small sample of mosquitoes used in the analyses and that only loci that have been previously associated with insecticide resistance were surveyed. However, this is an easily scalable method, and additional genes can be added to the multiplex assay as these are identified. Multiplex amplicon sequencing allows the rapid screening of many loci, resulting in high-quality genomic data which can complement phenotypic assays of insecticide resistance and inform vector control policy.

Conclusions

This study demonstrates the utility of multiplex amplicon sequencing for rapidly assessing the prevalence of insecticide resistance markers in mosquito populations. Furthermore, we provide the first assessment of molecular markers of insecticide resistance from multiple islands across the Bijagós Archipelago. Overall prevalence of these markers was low, but the presence of *vgsc* mutations L995F and N1570Y warrants further attention, particularly as vector control on the islands is limited to pyrethroid ITNs. The identification of additional molecular markers of insecticide resistance is required. This may be achieved through the combination of phenotypic bioassay data with genomic data, generated using either whole-genome sequencing or using high-throughput methods such as the multiplex amplicon sequencing technique used in this study.

Abbreviations

ace1	Acetylcholinesterase 1
CDC	Centers for Disease Control and Prevention
DDT	Dichlorodiphenyltrichloroethane
gste2	Glutathione S-Transferase Epsilon 2
IGS	Intergenic Spacer
INDELs	Insertions/Deletions
ITNs	Insecticide-Treated Nets

kdr	Knock-Down Resistance
RFLP	Restriction Fragment Length Polymorphism
rdl	Gamma-Aminobutyric Acid Receptor
SNP	Single Nucleotide Polymorphism
vgsc	Voltage-Gated Sodium Channel
WHO	World Health Organization

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13071-023-06085-5.

Additional file 1: Table S1. Target SNPs investigated which have been associated with insecticide resistance in the Culicidae family [18]. Genome and amino acid positions are labelled according to the position in *Anopheles gambiae* P4 genome. **Table S2.** Primers for insecticide resistance amplicon sequencing for the *Anopheles gambiae* sensu latu complex, adapted from Campos et al. [18]. **Table S3.** Barcodes BC1–BC10 were concatenated to the 5′ end of forward primers; BC11–BC20 were concatenated to the 5′ end of reverse primers.

Acknowledgements
We thank the extended MATAMAL entomological field team for their work collecting these mosquito samples. Thank you to Luis Ie, Tome Mendes, Ansumane Cassama and Malam Djassi, who were field assistants for this survey.

Author contributions
EP and RTJ conducted the mosquito collection on the Bijagós Archipelago during a cross-sectional survey, which was coordinated with HH and ES. The cross-sectional survey was designed by AL, EP and RTJ. SCe conducted DNA extraction and speciation at the Medical Research Council, The Gambia (MRCG) under the supervision of MON. HV assisted with sample organisation and transport. SM conducted primer modification, multiplex amplicon PCR assays and post-PCR sample processing. SM conducted the bioinformatic analysis of sequence data under the supervision of TGC and SCa, and with guidance from JP, HAF and EC. SM wrote the first draft of this manuscript. All authors reviewed and approved the final manuscript.

Funding
SM is funded by Medical Research Council UK (Grant No. MR/N013638/1). TGC and SC are funded by UKRI grants (Ref. no. MR/X005895/1, BB/X018156/1). AL, EP, ES, HH, HV, RTJ and SC are funded by Joint Global Health Trials Scheme (MRC, Wellcome Trust, UKRI, NIHR, Grant no. MR/S005013/1). HAP is funded by a LIDO-DTP PhD studentship. EC is funded by a Medical Research Council LID PhD Studentship.

Availability of data and materials
The raw sequence data generated and analysed during the current study are available in the European Nucleotide Archive (Project ID: PRJEB62886, Accession Numbers: ERS15575388–ERS15575673).

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Faculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, London, UK. [2]Medical Research Council, The Gambia (MRCG), Fajara, Gambia. [3]Ministério de Saúde Pública, Bissau, Guinea-Bissau. [4]Projecto de Saúde Bandim, Bissau, Guinea-Bissau. [5]Clinical Academic Group, Institute for Infection and Immunity, St. George's University Hospitals NHS Foundation Trust–St. George's University of London, London, UK. [6]Centre de Recherches Médicales de Lambaréné (CERMEL), Lambaréné, Gabon. [7]Institut Für Tropenmedizin Universitätsklinikum Tübingen, Tübingen, Germany. [8]Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London, UK.

Received: 28 June 2023 Accepted: 6 December 2023
Published online: 04 January 2024

References

1. WHO. World malaria report 2022. 2022. https://www.who.int/teams/global-malaria-programme. Accessed 1st June 2023.
2. Bhatt S, Weiss DJ, Cameron E, Bisanzio D, Mappin B, Dalrymple U, et al. The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. Nature. 2015;526:207. https://doi.org/10.1038/NATURE15535.
3. Ranson H, Lissenden N. Insecticide resistance in African *Anopheles* mosquitoes: a worsening situation that needs urgent action to maintain malaria control. Trends Parasitol. 2016;32:187–96. https://doi.org/10.1016/j.pt.2015.11.010.
4. WHO. Global report on insecticide resistance in malaria vectors: 2010–2016. 2017. https://www.who.int/publications/i/item/9789241514057. Accessed 1st June 2023.
5. WHO. Framework for a national plan for monitoring and management of insecticide resistance in malaria vectors 2017. https://www.who.int/publications/i/item/9789241512138. Accessed 1st June 2023.
6. WHO. Manual for monitoring insecticide resistance in mosquito vectors and selecting appropriate interventions. 2022. https://www.who.int/publications/i/item/9789240051089. Accessed 1st June 2023.
7. Martinez-Torres D, Chandre F, Williamson MS, Darriet F, Bergé J. B, Devonshire AL, et al. Molecular characterization of pyrethroid knockdown resistance (kdr) in the major malaria vector *Anopheles gambiae* s.s. Insect Mol Biol. 1998;7:179–84. https://doi.org/10.1046/j.1365-2583.1998.72062.x.
8. Ranson H, Jensen B, Vulule JM, Wang X, Hemingway J, Collins FH. Identification of a point mutation in the voltage-gated sodium channel gene of Kenyan *Anopheles gambiae* associated with resistance to DDT and pyrethroids. Insect Mol Biol. 2000;9:491–7. https://doi.org/10.1046/J.1365-2583.2000.00209.X.
9. Riveron JM, Yunta C, Ibrahim SS, Djouaka R, Irving H, Menze BD, et al. A single mutation in the GSTe2 gene allows tracking of metabolically based insecticide resistance in a major malaria vector. Genome Biol. 2014;15:R27. https://doi.org/10.1186/GB-2014-15-2-R27.
10. Mitchell SN, Rigden DJ, Dowd AJ, Lu F, Wilding CS, Weetman D, et al. Metabolic and target-site mechanisms combine to confer strong DDT resistance in *Anopheles gambiae*. PLoS ONE. 2014;9:e92662. https://doi.org/10.1371/JOURNAL.PONE.0092662.
11. Weill M, Malcolm C, Chandre F, Mogensen K, Berthomieu A, Marquine M, et al. The unique mutation in *ace-1* giving high insecticide resistance is easily detectable in mosquito vectors. Insect Mol Biol. 2004;13:1–7. https://doi.org/10.1111/J.1365-2583.2004.00452.X.
12. Grau-Bové X, Lucas E, Pipini D, Rippon E, van 'tHof AE, Constant E, et al. Resistance to pirimiphos-methyl in West African *Anopheles* is spreading via duplication and introgression of the *Ace1* locus. PLoS Genet. 2021. https://doi.org/10.1371/journal.pgen.1009253.
13. Du W, Awololoa TS, Howell P, Koekemoer LL, Brooke BD, Benedict MQ, et al. Independent mutations in the Rdl locus confer dieldrin resistance to *Anopheles gambiae* and *An. arabiensis*. Insect Mol Biol. 2005;14:179–83. https://doi.org/10.1111/J.1365-2583.2005.00544.X.
14. Hutchins H, Power G, Ant T, Teixeira da Silva E, Goncalves A, Rodrigues A, et al. A survey of knowledge, attitudes and practices regarding malaria and bed nets on Bubaque Island, Guinea-Bissau. Malar J. 2020;19:412. https://doi.org/10.1186/s12936-020-03469-1.
15. Marsden CD, Cornel A, Lee Y, Sanford MR, Norris LC, Goodell PB, et al. An analysis of two island groups as potential sites for trials of transgenic mosquitoes for malaria control. Evol Appl. 2013;6:706–20. https://doi.org/10.1111/eva.12056.

16. Ant T, Foley E, Tytheridge S, Johnston C, Goncalves A, Ceesay S, et al. A survey of *Anopheles* species composition and insecticide resistance on the island of Bubaque, Bijagos Archipelago, Guinea-Bissau. Malar J. 2020;19:27. https://doi.org/10.1186/s12936-020-3115-1.
17. Vicente JL, Clarkson CS, Caputo B, Gomes B, Pombi M, Sousa CA, et al. Massive introgression drives species radiation at the range limit of *Anopheles gambiae*. Sci Rep. 2017;7:46451. https://doi.org/10.1038/srep46451.
18. Campos M, Phelan J, Spadar A, Collins E, Gonçalves A, Pelloquin B, et al. High-throughput barcoding method for the genetic surveillance of insecticide resistance and species identification in *Anopheles gambiae* complex malaria vectors. Sci Rep. 2022;12:13893. https://doi.org/10.1038/S41598-022-17822-8.
19. Nag S, Dalgaard MD, Kofoed PE, Ursing J, Crespo M, Andersen LO, et al. High throughput resistance profiling of *Plasmodium falciparum* infections based on custom dual indexing and Illumina next generation sequencing-technology. Sci Rep. 2017;7:1–13. https://doi.org/10.1038/s41598-017-02724-x.
20. Mboera LE, Kihonda J, Braks MA, Knols BG. Short report: Influence of centers for disease control light trap position, relative to a human-baited bed net, on catches of *Anopheles gambiae* and *Culex quinquefasciatus* in Tanzania. Am J Trop Med Hyg. 1998;59:595–6. https://doi.org/10.4269/ajtmh.1998.59.595.
21. Coetzee M. Key to the females of Afrotropical *Anopheles* mosquitoes (Diptera: Culicidae). Malar J. 2020;19:70. https://doi.org/10.1186/s12936-020-3144-9.
22. Fanello C, Santolamazza F, Torre AD. Simultaneous identification of species and molecular forms of the *Anopheles gambiae* complex by PCR-RFLP. Med Vet Entomol. 2002;16:461–4. https://doi.org/10.1046/J.1365-2915.2002.00393.X.
23. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20. https://doi.org/10.1093/BIOINFORMATICS/BTU170.
24. Jones CM, Liyanapathirana M, Agossa FR, Weetman D, Randon H, Donnely MJ, Wilding CS. Footprints of positive selection associated with a mutation (N1575Y) in the voltage-gated sodium channel of *Anopheles gambiae*. Proc Natl Acad Sci USA. 2012;109:6614–9. https://doi.org/10.1073/pnas.1201475109.
25. Clarkson CS, Miles A, Harding NJ, O'Reilly AO, Weetman D, Kwiatkowski D, et al. The genetic architecture of target-site resistance to pyrethroid insecticides in the African malaria vectors *Anopheles gambiae* and *Anopheles coluzzii*. Molecular Ecology. 2021;30:5303-5317. https://doi.org/10.1111/mec.15845.
26. Acford-Palmer H, Phelan JE, Tadesee FG, Kristan M, Collins E, Spadar A, et al. Identification of two insecticide resistance markers in Ethiopian *Anopheles stephensi* mosquitoes using a multiplex amplicon sequencing assay. Sci Rep. 2023;13:5612. https://doi.org/10.1038/s41598-023-32336-7.
27. Collins EL, Phelan JE, Hubner M, Spadar A, Campos M, Ward D, et al. A next generation targeted amplicon sequencing method to screen for insecticide resistance mutations in *Aedes aegypti* populations reveals a rdl mutation in mosquitoes from Cabo Verde. PLoS Negl Trop Dis. 2022;16:e0010935. https://doi.org/10.1371/journal.pntd.0010935.
28. Ni R, Liu N, Li M, Qian W, Qiu X. Identification and phylogenetic analysis of voltage-gated sodium channel haplotypes in the malaria vector *Anopheles sinensis* using a high-throughput amplicon sequencing approach. Parasit Vectors. 2021;14:499. https://doi.org/10.1186/s13071-021-05009-5.
29. Silva R, Mavridis K, Vontas J, Rodrigues A, Costa H. Monitoring and molecular profiling of contemporary insecticide resistance status of malaria vectors in Guinea-Bissau. Acta Trop. 2020;206:105440. https://doi.org/10.1016/j.actatropica.2020.105440.
30. Assogba BS, Djogbénou LS, Saizonou J, Milesi P, Djossou L, Djegbe I, et al. Phenotypic effects of concomitant insensitive acetylcholinesterase (ace-1R) and knockdown resistance (kdrR) in *Anopheles gambiae*: a hindrance for insecticide resistance management for malaria vector control. Parasit Vectors. 2014;7:548. https://doi.org/10.1186/s13071-014-0548-9.
31. Pontes FJS, Maia RT, Lima MCP, Ayres CFJ, Soares TA. The role of the conformational dynamics of glutathione *S*-transferase epsilon class on insecticide resistance in *Anopheles gambiae*. J Braz Chem Soc. 2016;27:1602-15. https://doi.org/10.5935/0103-5053.20160040.
32. Lucas ER, Miles A, Harding NJ, Clarkson CS, Lawniczak MKN, Kwiatkowski DP, et al. Whole-genome sequencing reveals high complexity of copy number variation at insecticide resistance loci in malaria mosquitoes. Genome Res. 2019;29:1250–61. https://doi.org/10.1101/gr.245795.118.
33. Weetman D, Mitchell SN, Wilding CS, Birks DP, Yawson AE, Essandoh J, et al. Contemporary evolution of resistance at the major insecticide target site gene *Ace-1* by mutation and copy number variation in the malaria mosquito *Anopheles gambiae*. Mol Ecol. 2015;24:2656–72. https://doi.org/10.1111/mec.13197.
34. Sinka ME, Bangs MJ, Manguin S, Coetzee M, Mbogo CM, Hemingway J, et al. The dominant *Anopheles* vectors of human malaria in Africa, Europe and the Middle East: occurrence data, distribution maps and bionomic précis. Parasit Vectors. 2010;3:117. https://doi.org/10.1186/1756-3305-3-117.
35. Lindsay SW, Thomas MB, Kleinschmidt I. Threats to the effectiveness of insecticide-treated bednets for malaria control: thinking beyond insecticide resistance. Lancet Glob Health. 2021;9:e1325–31. https://doi.org/10.1016/S2214-109X(21)00216-3.

Publisher's Note

Supplementary Information

Supplementary Table 1: Target SNPs investigated which have been associated with insecticide resistance in the *Culicidae* family [147]. Genome positions and amino acid positions are labelled according to the position in *Anopheles gambiae* P4 genome.

Gene	Chromosome	Position	SNP	Amino Acid change
VGSC (DI-S6)	AgamP4_2L	2391228	G>C, T	V402L
VGSC (DII-S6)	AgamP4_2L	2422575	T >C	S970P
VGSC (DII-S6)	AgamP4_2L	2422643	A>G	I992M
VGSC (DII-S6)	AgamP4_2L	2422651	T >C	L995S
VGSC (DII-S6)	AgamP4_2L	2422652	A>T	L995F
VGSC (DII-S6)	AgamP4_2L	2422657	T>G	V997G
VGSC (DIII-S6)	AgamP4_2L	2429617	T>C	I1527T
VGSC (DIII-S6)	AgamP4_2L	2429623	T>G	F1529C
VGSC (DIII-S6)	AgamP4_2L	2429745	A>T	N1570Y
VGSC (DIV-S5)	AgamP4_2L	2430424	G>T	A1746S
VGSC (DIV-S5)	AgamP4_2L	2430460	G>T	D1758Y
GSTE2	AgamP4_3R	28598166	T>C	I114T
GSTE2	AgamP4_3R	28598062	C>G	L119V
GSTE2	AgamP4_3R	28598057	C>G,A	F120 L
RDL	AgamP4_2L	25429236	C>G	A296G
RDL	AgamP4_2L	25429235	G>T	A296S
ACE1	AgamP4_2R	3492074	G>A	G280S

Supplementary Table 2: Primers for insecticide resistance amplicon sequencing for the *Anopheles gambiae sensu latu complex*, adapted from Campos *et al* (2022) [147]

Target Gene	Amplicon	Forward primer	Reverse Primer	Chromosome	Position*	Product Size (bp)*
vgsc	VGSC-I	ATTCGTTATTCTTCAGATGAACT	ATTCTCACCCGAAGTGC	AgamP4_2L	2390813-2391328	517
	VGSC-II	GTTTTGCTAGCCTAATTGC	TGTCGGTTGAACGGATGCTATT	AgamP4_2L	2422417-2422919	503
	VGSC-III	TTCATGGGAAAATTCACCAA	AATTAGTGCTCCAAACACAAAC	AgamP4_2L	2429356-2429845	490
	VGSC-IV	CGAGCCATGGAATTTGT	TGATGTGATCCAGTTACAGA	AgamP4_2L	2430093-2430593	501
ace-1	ACE1	CTGGTGGTCAACACGGA	GAACAGTCCCGCATTGC	AgamP4_2R	3491732-3492229	498

gste2	GSTE2	GCCCGGATGAGATTCAT	GGCTAGCACAAACTTGC	AgamP4_3R	28597778 - 28598182-	404
rdl	RDL	CATTGCAATCATCACCATCA	CCAGCAGACTGGCAAATACC	AgamP4_2L	25428861-25429373	500

*Position and product size in *An. gambiae P4.*

Supplementary Table 3: Barcodes, BC1 to BC10 were concatenated to the 5' end of forward primers. BC11 to BC20 were concatenated to the 5' end of reverse primers.

Name	Barcode	Name	Barcode
BC1N	CTATCACG	BC11N	ATGGCTAG
BC2N	TCCAGTGT	BC12N	GACTTGGT
BC3N	GATCAGTA	BC13N	TCGATCAC
BC4N	AGTGTCGG	BC14N	ACACGTCA
BC5N	GTAGCGCT	BC15N	CAATGTGC
BC6N	CATCTAAC	BC16N	GGGACTAC
BC7N	TACAGATC	BC17N	ACGTACTG
BC8N	CGTCTTGT	BC18N	TGATTGCC
BC9N	TATGATCA	BC19N	AACTCTAC
BC10N	GGTAGCTT	BC20N	TGACTCAA

CHAPTER FIVE

Published paper: Phenotypic evidence of deltamethrin resistance and identification of selective sweeps in *Anopheles* mosquitoes on the Bijagós Archipelago, Guinea-Bissau.

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	lsh2005123	**Title**	Miss
First Name(s)	Sophie		
Surname/Family Name	Moss		
Thesis Title	Genomic epidemiology of *Plasmodium falciparum* and its *Anopheline* vectors in the context of malaria control on the Bijagós Archipelago of Guinea-Bissau		
Primary Supervisor	Dr Anna Last		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	N/A		
When was the work published?	N/A		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	N/A		
Have you retained the copyright for the work?*	**N/A**	Was the work subject to academic peer review?	**N/A**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	Scientific Reports

Please list the paper's authors in the intended authorship order:	Sophie Moss, Robert T. Jones, Elizabeth Pretorius, Eunice Teixeira da Silva, Matthew Higgins, Mojca Kristan, Holly Acford-Palmer, Emma L. Collins, Amabelia Rodrigues, Sanjeev Krishna, Taane G. Clark, Anna Last, Susana Campino
Stage of publication	**Submitted**

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I devised the research question and designed the study. I conducted phenotypic testing for deltamethrin resistance in the Archipelago, all DNA extraction and species identification, all preparation for sequencing, and all bioinformatic analyses. I wrote the manuscript, which was commented on by my co-authors

SECTION E

Student Signature	[redacted]
Date	2nd April 2024

Supervisor Signature	Anna Last Digitally signed by Anna Last [redacted] Date: 2024.05.22 18:21:41 +01'00'
Date	22nd May 2024

scientific reports

Check for updates

OPEN

Phenotypic evidence of deltamethrin resistance and identification of selective sweeps in *Anopheles* mosquitoes on the Bijagós Archipelago, Guinea-Bissau

Sophie Moss[1]✉, Robert T. Jones[1], Elizabeth Pretorius[1], Eunice Teixeira da Silva[2,3], Matthew Higgins[1], Mojca Kristan[1], Holly Acford-Palmer[1], Emma L. Collins[1], Amabelia Rodrigues[2,3], Sanjeev Krishna[4,5,6], Taane G. Clark[1,7,8], Anna Last[1,8] & Susana Campino[1,8]

Vector control in the Bijagós Archipelago of Guinea-Bissau currently relies on pyrethroid insecticide-treated nets. However, data on insecticide resistance in Guinea-Bissau is limited. This study identified deltamethrin resistance in the *Anopheles gambiae* sensu lato complex on Bubaque island using WHO tube tests in November 2022. Whole genome sequencing of *An. gambiae* sensu stricto mosquitoes identified six single nucleotide polymorphisms (SNPs) previously associated with, or putatively associated with, insecticide resistance: T791M, L995F, N1570Y, A1746S and P1874L in the *vgsc* gene, and L119V in the *gste2* gene. Twenty additional non-synonymous SNPs were identified in insecticide-resistance associated genes. Four of these SNPs were present at frequencies over 5% in the population: T154S, I126F and G26S in the *vgsc* gene and A65S in *ace1*. Genome wide selection scans using Garud's H_{12} statistic identified two selective sweeps: one in chromosome X and one in chromosome 2R. Both selective sweeps overlap with metabolic genes previously associated with insecticide resistance, including *cyp9k1* and the *cyp6aa/cyp6p* gene cluster. This study presents the first phenotypic testing for deltamethrin resistance and the first whole genome sequence data for *Anopheles gambiae* mosquitoes from the Bijagós, contributing data of significance for vector control policy in this region.

Insecticide Treated Nets (ITNs) are the most effective method of reducing malaria cases to date, accounting for an estimated 68% of the reduction in malaria cases between 2000 and 2015[1]. Currently, all ITNs contain pyrethroid insecticides[2]. However, the evolution of resistance to pyrethroids in *Anopheles* mosquitoes threatens the efficacy of these ITNs. The World Health Organization (WHO) recommends that countries monitor insecticide resistance in disease vectors as part of the development of insecticide resistance monitoring and management plans (IRMMPs)[3], assisting national control programmes to make evidence-based decisions about which insecticides to use. Of the 88 countries that reported insecticide resistance data to WHO between 2010 and 2020, 87% reported resistance to pyrethroids in at least one malaria vector, and 33% of countries reported resistance to all four insecticide classes commonly used in vector control[4]. The gold standard for monitoring insecticide resistance is phenotypic testing using WHO tube tests or bottle bioassays[5], which are time and resource intensive. However, the identification and monitoring of molecular markers associated with resistance could supplement these phenotypic bioassays. Furthermore, the rapid development of sequencing technologies makes whole genome sequencing (WGS) a promising approach for identifying and monitoring the spread of molecular markers on a large scale. Previous studies using WGS data have identified regions of the *Anopheles*

[1]Faculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, London, UK. [2]Projecto de Saúde Bandim, Bissau, Guinea-Bissau. [3]Ministério de Saúde Pública, Bissau, Guinea-Bissau. [4]Clinical Academic Group, Institute for Infection and Immunity, and St, George's University Hospitals NHS Foundation Trust, St. George's University of London, London, UK. [5]Centre de Recherches Médicales de Lambaréné (CERMEL), Lambaréné, Gabon. [6]Institut für Tropenmedizin Universitätsklinikum Tübingen, Tübingen, Germany. [7]Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London, UK. [8]These authors jointly supervised this work: Taane G. Clark, Anna Last, and Susana Campino. ✉email: sophie.moss@lshtm.ac.uk

gambiae genome associated with pyrethroid insecticide resistance. These include studies conducted by the *Anopheles gambiae* 1000 Genomes Consortium (Ag1000G)[6,7]. Key regions of the genome that have previously been associated with pyrethroid resistance include the cytochrome P450 genes, with the *cyp6aa1* and *cyp9k1* genes demonstrating the strongest signals of association with deltamethrin resistance[6,8]. In addition, the *Tep* family of immune related genes has been found to be associated with deltamethrin resistance[6].

The most well understood mechanisms of insecticide resistance are metabolic and target-site resistance. Metabolic resistance occurs when mosquitoes evolve the ability to rapidly detoxify insecticides, which is attributed primarily to three different enzyme families; the cytochrome P450s, esterases and glutathione-S-transferases (GSTs)[9]. Whereas, target-site resistance occurs through modifications to the binding sites of insect protein receptors, resulting in conformational changes which impede insecticide binding. Other mechanisms that are less well studied include behavioural resistance, such as mosquitoes evolving to bite during the day when people are not protected by ITNs, and cuticular resistance, which is the evolution of a thickened or altered cuticle that is more difficult for insecticides to penetrate[5]. Pyrethroid resistance in *Anopheles gambiae* is associated with target-site mutations in the voltage-gated sodium channel (*vgsc*) gene, also known as knock-down or "kdr" mutations. These mutations include the L995F[10] (kdr-west) allele (position 1014 in *Musca domestica*), which is widespread in West and Central Africa, and the L995S[11] (kdr-east) allele, which is widespread in East Africa[4]. Another well characterised *vgsc* mutation, N1570Y, has been associated with increased levels of pyrethroid resistance when associated with L995F[12]. In addition, L119V in the glutathione S-transferase epsilon class 2 (*gste2)* gene, has been associated with increased resistance to the pyrethroid permethrin[13]. Metabolically mediated pyrethroid resistance has been associated with copy number variations (CNVs) in multiple genes involved in the metabolism of insecticides. This includes CNVs in the cytochrome P450 genes *cyp6aa1*[14], *cyp6p3*, *cyp6m2*, *cyp6z1*, *cyp9k1*[15,16], and the *gste2* gene, all of which have been shown to metabolise insecticides[14,17].

Malaria is a persistent public health problem on the Bijagós Archipelago of Guinea-Bissau, where the peak prevalence of *Plasmodium falciparum* parasitaemia is up to 15%[18]. ITNs are the only vector control intervention currently used on the islands. Of these ITNs, 90% are PermaNet® 2.0, which are impregnated with the pyrethroid insecticide deltamethrin[19]. However, there is little data available about insecticide resistance on the Bijagós or on mainland Guinea-Bissau[20]. A survey on Bubaque island in 2017 identified *An. gambiae* s.s. as the primary malaria vector during the rainy season in June and July, and a majority of *An. melas* during November and December[21]. The Bijagós is also a site of extremely high hybridisation rates between *An. gambiae* s.s. and *An. coluzzii*. These hybrids have been consistently recorded at proportions over 20% since 1995 [22,23], which is much higher than found elsewhere in West and Central Africa, where rates of hybridisation are usually below 1%[24]. A recent study suggests that these hybrid mosquitoes may be a novel, cryptic taxon provisionally named the "Bissau molecular form"[25]. Previously, we identified a low prevalence of molecular markers associated with insecticide resistance on the Archipelago, using multiplex amplicon sequencing of *An. gambiae* sensu lato (s.l.) mosquitoes collected in 2019 from 13 different islands. Of 17 mutations screened, we identified three *vgsc* mutations previously or putatively associated with pyrethroid resistance: L995F, N1570Y, A1746S, and the *rdl* mutation A269G which is associated with resistance to the organochlorine dieldrin[19]. The present study builds on this research by: (1) using WHO tube tests to conduct insecticide resistance bioassays for deltamethrin resistance, which have not been conducted on the Bijagós islands previously, and (2) generating whole-genome sequence data for deltamethrin resistant and susceptible individuals of the major vector *Anopheles gambiae* s.s., to identify molecular markers in insecticide resistance genes. This study contributes the first whole genome sequencing data of *Anopheles gambiae* s.s. vectors from the Bijagós.

Results

Phenotypic insecticide resistance assay results

Phenotypic tests were conducted in November 2022, investigating the status of resistance to discriminating concentrations and 5-times (5x) intensity concentrations of deltamethrin in *Anopheles gambiae* s.l. mosquitoes. WHO tube tests were conducted according to WHO guidelines[5]. Combined assay results are provided in Table 1. Mosquito mortality in control tests 24 h post-exposure was ≤ 20% for each test conducted. Corrected treatment mortality (%) was calculated using Abbott's formula when control mortality was ≥ 5%, in line with WHO guidelines[5].

The total number of mosquitoes exposed to deltamethrin at the discriminating concentration of 0.05% over all replicates was 101, which meets the WHO recommended optimal sample size for this bioassay[5]. Corrected treatment mortality at this concentration (0.05%) was 45.35% (Table 1), which confirmed resistance to deltamethrin[5]. Intensity of resistance was then investigated with 0.25% deltamethrin, which is five times (5x)

Treatment	24 h post-treatment exposure			
	Alive	Knocked down	Mortality (%)	Corrected treatment mortality (%)
Control ($n=66$)	61	5	7.58	
Deltamethrin 0.05% ($n=101$)	51	50	49.50	45.36
Control ($n=37$)	33	4	10.81	
Deltamethrin 0.25% ($n=45$)	3	42	93.33	92.52

Table 1. Results of WHO tube tests measuring resistance to deltamethrin. Discriminating concentration testing using 0.05% deltamethrin and 5x intensity concentration testing with 0.25% deltamethrin. Tests were conducted in November 2022 during the peak malaria transmission season.

the discrimination concentration. Corrected treatment mortality at this concentration (0.25%) was 92.52% (Table 1), indicating moderate to high intensity resistance according to WHO guidelines[5]. Due to logistical constraints during fieldwork, the sample size for the intensity resistance testing was $n=45$, which is lower than the WHO recommended sample size for intensity testing. Therefore, although these tests confirm the presence of deltamethrin resistance, additional tests are required to confirm the resistance intensity[5].

Whole genome sequencing of *Anopheles gambiae* s.s. mosquitoes

Whole genome sequencing was conducted with *Anopheles gambiae* s.s mosquitoes for 23 resistant and 10 susceptible mosquitoes (tested with 0.05% deltamethrin during WHO tube tests), and 9 control mosquitoes (exposed to oil instead of insecticide). Variants were called and filtered for quality, resulting in a total of 16,452,859 high quality variants for analysis.

Allele frequencies of insecticide resistance SNPs

Allele frequencies of 27 candidate SNPs associated with resistance across the *vgsc*, *gste2*, *rdl*, and *ace1* genes were investigated (Supplementary Data 1). Six of these SNPs were identified in the Bijagós whole genome sequence data (Table 2). The *vgsc* T791M and A1746S mutations were identified in susceptible and control mosquitoes at frequencies between 5.0 and 6.3%, but were absent in resistant mosquitoes. The *vgsc* L995F mutation was found at similar frequencies in both resistant, susceptible, and control mosquitoes, at frequencies between 15.0 and 16.9%. The N1570Y mutation was identified at a frequency of 10.9% in resistant mosquitoes and 6.3% in control mosquitoes, with no susceptible mosquitoes carrying this mutation. The P1874L allele was identified in resistant, susceptible and control mosquitoes at frequencies of 4.3%, 10.0% and 6.3% respectively. The *gste2* L119V mutation was identified in resistant mosquitoes at a frequency of 6.8% and was not identified in any susceptible or control mosquitoes. Additional genotype data is provided per mosquito in Supplementary Data 1. Odds ratios with 95% confidence intervals were calculated and no significant associations were found between the resistance phenotype and these SNPs in this sample set, which is likely due to the small sample size (Supplementary Data 3).

Linkage disequilibrium of *vgsc* insecticide-resistance mutations

Linkage disequilibrium (LD) was calculated for each pair of variants using the method of Rogers and Huff[26] (Fig. 1). R^2 values are shown in Supplementary Data 1. The alternate alleles T791M and A1746S were in complete LD with an R^2 value of 1. The alternate alleles L995F and N1570Y had an R^2 value of 0.372, indicating that these alleles were in moderate LD. The following alternate alleles were in weak LD: T791M and L995F ($R^2=0.112$), L995F and A1746S ($R^2=0.112$), L995F and P1874L ($R^2=0.301$). The T791M mutation was only ever identified in association with L995F. However, as the L995F allele was identified at a higher frequency and in samples without the presence of T791M, these two alleles were not found in high LD. Other SNP combinations showed very low or absent LD ($R^2 \leq 0.023$). Note that the sample sizes of alternate SNPs used to calculate LD were small (Table 2).

Allele frequencies of other non-synonymous SNPs

The four genes associated with insecticide resistance that were investigated here (*vgsc*, *gste2*, *rdl* and *ace1)* were also investigated for additional non-synonymous SNPs that have not previously been associated with insecticide resistance. An additional 20 non-synonymous SNPs were identified in insecticide resistance genes (Table 3). SNPs with an allele frequency above 5% have been highlighted in bold. Odds ratios with 95% confidence intervals were calculated for these SNPs, but no significant associations were found between the resistance phenotype and these SNPs in this sample set (Supplementary Data 3).

Investigation of copy number variant (CNV) alleles associated with metabolic resistance: detection using soft-clipped reads

Copy number variants (CNVs) in genes associated with metabolic insecticide resistance were also investigated. We analysed CNVs that were previously identified by the *Anopheles gambiae* 1000 Genomes Consortium[17]. As whole genome amplification (WGA) was used to amplify the mosquito DNA, we could not reliably identify

Chromosome	Gene	Position	SNP	Resistant mosquitoes ($n=23$)		Susceptible mosquitoes ($n=10$)		Control mosquitoes ($n=9$)	
				Alternate allele frequency (%)	Total alleles called	Alternate allele frequency (%)	Total alleles called	Alternate allele frequency (%)	Total alleles called
2 L	*vgsc*	2,416,980	T791M	0	36	5.6	18	5.6	18
2 L	*vgsc*	2,422,652	L995F	15.9	44	15.0	20	16.7	18
2 L	*vgsc*	2,429,745	N1570Y	10.9	46	0	20	6.3	16
2 L	*vgsc*	2,430,424	A1746S	0	46	5.0	20	6.3	16
2 L	*vgsc*	2,430,881	P1874L	4.3	46	10.0	20	6.3	16
3R	*gste2*	28,598,062	L119V	6.8	44	0	18	0	18

Table 2. Candidate resistance SNPs identified in *Anopheles gambiae s.s.* mosquitoes. The table includes the alternate allele frequencies and total number of alleles called, arranged by insecticide resistance phenotype.

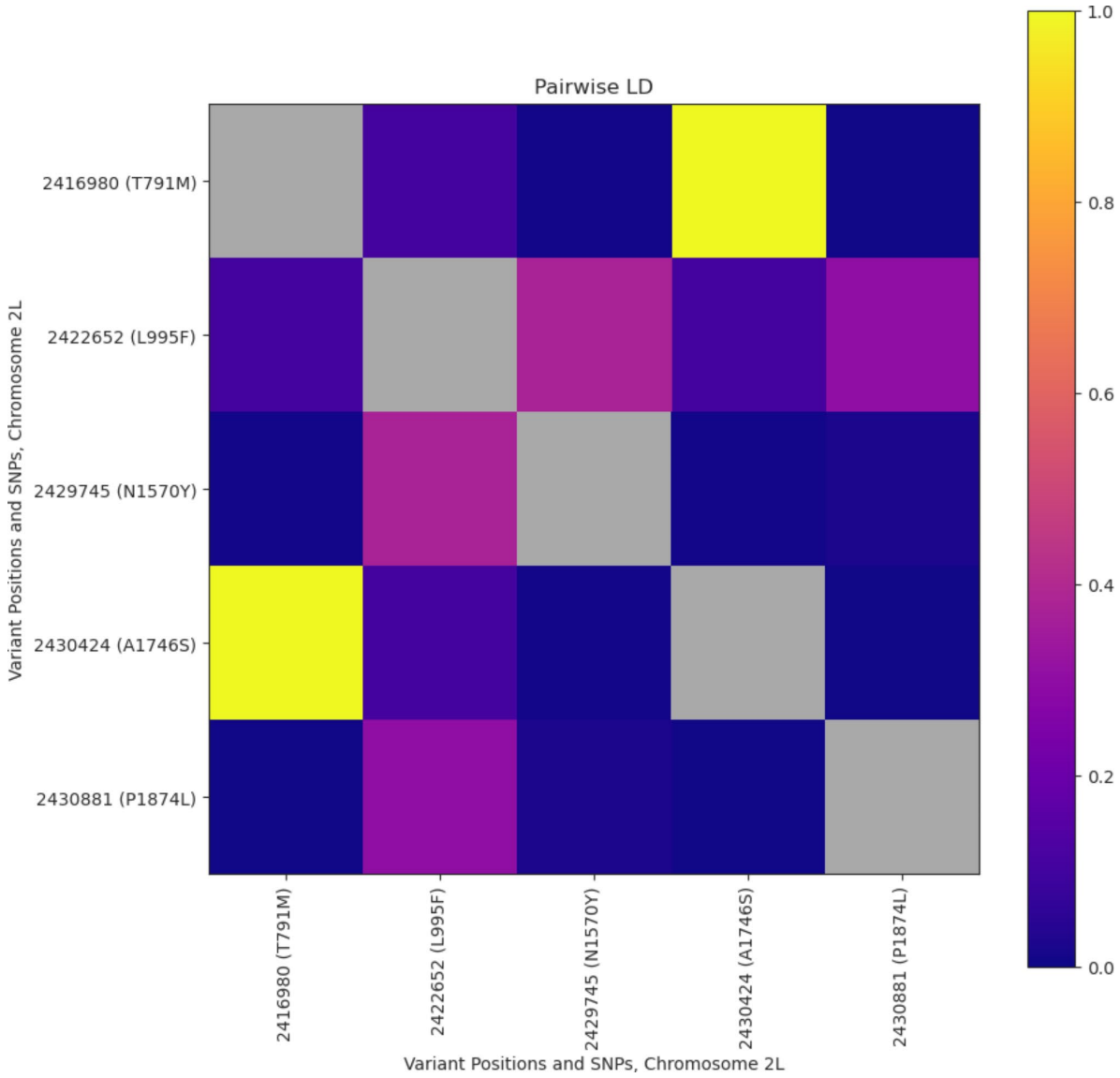


Fig. 1. Pairwise linkage disequilibrium (Rogers and Huff) between insecticide-resistance SNPs in the *vgsc* gene. A value of 1 indicates that the two alleles are in complete linkage, which means that these two alleles are only ever found in combination with each other.

CNVs using read coverage data or quantitative PCR. Instead, we screened for these CNVs by identifying soft-clipped reads that were able to detect the CNV[17]. The investigated CNVs are listed in Supplementary Data 2 and included CNV alleles in cytochrome-P450 genes: *cyp6aa – cyp6p* region, *cyp6m – cyp6z* region and *cyp9k1*, and within the *gstu - gste* cluster region. This analysis indicated the putative presence of four CNV alleles of interest in our Bijagós mosquitoes: Cyp6aap_Dup7, Cyp6aap_Dup11, Cyp9k1_Dup12, and Gstue_Dup3. Three of these CNV alleles were identified within the susceptible mosquito population: Cyp6aap_Dup11 and Gsteu_Dup3 in one susceptible mosquito, and Cyp9k1_Dup12 in one susceptible mosquito. One CNV allele was present in one resistant mosquito, Cyp6aap_Dup7. CNV allele counts can be found in Supplementary Data 2.

Windowed measures of differentiation and selection

The fixation index (F_{ST}) is a measure of genetic differentiation between populations[6]. F_{ST} was calculated between resistant and susceptible mosquitoes over each chromosome in 1Kbp windows. This metric was used to identify regions of differentiation between the resistant and susceptible mosquitoes. F_{ST} scores were plotted along each chromosome, and significant peaks in F_{ST} were identified (Supplementary Data 1), which encompassed 51

Chrom	Gene	Position	Ref	Alt	Amino acid Change	Resistant mosquitoes ($n=23$)		Susceptible mosquitoes ($n=10$)		Control mosquitoes ($n=9$)	
						Heterozygous	Homozygous Alternate	Heterozygous	Homozygous Alternate	Heterozygous	Homozygous Alternate
2 L	*vgsc*	2,431,232	G	A	G1991E	1	0	0	0	0	0
	vgsc	2,431,330	A	T	T2024S	1	0	0	0	0	0
	vgsc	2,431,394	C	T	P2045L	0	0	0	0	1	0
3R	*gste2*	28,597,772	C	A	K215N	0	1	0	0	0	0
	gste2	28,597,891	C	T	E176K	2	0	1	0	1	0
	gste2	**28,597,956**	**G**	**C**	**T154S**	**13**	**4**	**5**	**1**	**5**	**1**
	gste2	**28,598,041**	**T**	**A**	**I126F**	**3**	**0**	**1**	**0**	**1**	**0**
	gste2	**28,598,505**	**C**	**T**	**G26S**	**2**	**0**	**1**	**0**	**0**	**0**
	gste2	28,598,568	C	G	V5L	1	0	0	0	1	0
	gste2	28,598,573	T	A	N3I	1	0	0	0	1	0
2 L	*rdl*	25,382,837	G	A	V63I	1	0	0	0	0	0
	rdl	25,433,554	C	T	P473S	1	0	0	0	0	0
2R	*ace1*	3,489,253	G	A	G14E	1	0	0	0	0	0
	ace1	3,489,358	C	G	A49G	1	0	0	0	0	0
	ace1	3,489,391	C	T	A60V	0	0	1	0	0	0
	ace1	**3,489,405**	**G**	**T**	**A65S**	**14**	**5**	**7**	**2**	**4**	**1**
	ace1	3,489,475	C	T	S88L	0	0	0	0	1	0
	ace1	3,491,883	C	T	T216I	1	0	0	0	0	0
	ace1	3,493,410	G	A	S648N	0	0	0	0	1	0
	ace1	3,493,714	G	T	G1991E	0	0	1	0	0	0

Table 3. Additional non-synonymous SNPs identified in insecticide resistance genes *vgsc*, *gste2*, *rdl and ace1*. Includes the number of mosquitoes and their phenotype per SNP. SNPs with an allele frequency over 5% in the sample population are highlighted in bold. Ref = reference alt = alternate.

protein-coding genes (Table 4). The top five significant F_{ST} values for each chromosome are summarised in Table 4, with more detail on F_{ST} significance criteria available in Supplementary Data 1.

Haplotype clusters

Each window of the genome with a 'peak' in F_{ST} was treated as a putative window of interest. For each of these windows, hierarchical clustering of haplotypes was performed to identify high frequency haplotypes which may be under directional selection, making them candidates for association with insecticide resistance. We identified clusters of ≥ 20 highly similar haplotypes in two windows of chromosome 3L (7295325–7305324) and (7296325–7306324), one window of chromosome 3R (52339701–52349700) and in 10 windows of chromosome X. There were no significant differences in the haplotypes present between the resistant and the susceptible mosquitoes (Supplementary Data 1). A larger sample size is likely necessary to reveal associations between haplotype clusters and the resistance phenotype.

Genome wide selection scan: H12 for recent selective sweeps

We calculated Garud's H_{12} statistic in 1000 SNP windows across the genome to detect signatures of recent selective sweeps (Fig. 2)[6,27]. H_{12} is a statistic designed to identify both hard and soft selective sweeps[28]. This metric was calculated for susceptible and resistant mosquitoes. We found the signatures of two recent selective sweeps in the population; one on Chromosome X (Fig. 2) and one on Chromosome 2R (Fig. 3), but no sweeps that were unique to resistant mosquitoes. These peaks indicate selection in the mosquito population as a whole, but do not distinguish whether these sweeps are associated with resistance to deltamethrin. Selective sweeps were not identified in chromosomes 2L, 3L or 3R, where no H_{12} values greater than 0.2 were found (Supplementary Data 1).

Protein coding genes identified in the peaks of H_{12} are summarised in Supplementary Data 1. Overlapping genes were identified if they were in a 1000 SNP window with a H_{12} value above 0.2. This identified 55 different genes in chromosome X and 12 different genes in chromosome 2R. These genes included the cytochrome P450 genes *cyp9k1*, *cyp6aa1*, *cyp6aa2*, *coeae60*, *cyp6p15p*, *cyp6p3*, *cyp6p5*, *cyp6p4*, *cyp6p1*, *cyp6p2*, and *cyp6ad1*.

Discussion

Monitoring and managing the evolution of insecticide resistance is a vital component in the prevention and elimination of vector-borne diseases[3]. Improvements in sequencing technology have enabled the generation of whole genome sequence (WGS) data for vectors on a large scale[29]. This data can be used to identify molecular markers of insecticide resistance, which can supplement the results of phenotypic bioassays. This study used WHO tube tests[5] to investigate the presence of deltamethrin resistance in mosquitoes on Bubaque island in the

Chromosome	Gene	Description (Vector Base)	F_{ST} value
2L	AGAP004835	Protein coding gene, unspecified product	0.177
	AGAP029530	Protein coding gene - unspecified product	0.150
	AGAP005510	Oxysterol binding protein-like 9	0.153
	AGAP007133	Protein coding gene - unspecified product	0.140
	AGAP007576	Protein coding gene - unspecified product	0.196
2R	AGAP001535	Histone-lysine N-methyltransferase ASH1L	0.156
	AGAP001786	Protein coding gene - unspecified product	0.184
	AGAP002748	Protein kinase C	0.163
	AGAP029576	Protein coding gene – unspecified product	0.155
	AGAP003115	Protein coding gene - unspecified product	0.199
3L	AGAP010621	Glycine transporter	0.198
	AGAP029077	Protein coding gene - unspecified product	0.174
	AGAP029564	Ig-like domain-containing protein	0.293
	AGAP029790	Protein coding gene - unspecified product	0.195
	AGAP028685	DUF4806 domain-containing protein	0.195
3R	AGAP010407	Elongator complex protein 4	0.311
	AGAP010884	Protein coding gene – unspecified product	0.133
	AGAP011540	Dynein intermediate chain 2, axonemal	0.156
	AGAP029516	Protein coding gene - unspecified product	0.128
	AGAP029760	Clustered mitochondria protein homolog	0.135
X	AGAP000433	Ras-related protein Rab-39B	0.145
	AGAP029672	Protein tweety homolog	0.150
	AGAP029671	Protein coding gene - unspecified product	0.150
	AGAP000863	Lachesin	0.125
	AGAP000932	Protein coding gene - unspecified product	0.127

Table 4. Genes identified within F_{ST} windows of significance, between resistant and susceptible *Anopheles gambiae* s.s. whole genome sequence data.

Bijagós Archipelago. This was followed by WGS of *Anopheles gambiae* s.s. mosquitoes to investigate molecular markers associated with resistance.

Phenotypic bioassays revealed deltamethrin resistance in the mosquito population according to WHO guidelines[5], and intensity testing indicated that resistance is of moderate to high intensity. These are important findings, particularly considering the reliance on pyrethroid ITNs for vector control in the Bijagós. This data can be incorporated into future decisions on which vector-based control tools to implement. Pyrethroid resistance has become widespread across malaria endemic countries[20], and a number of updated ITN recommendations have been issued in recent years[4]. In 2017, WHO recommended the use of pyrethroid-PBO ITNs, which contain the synergist piperonyl butoxide (PBO), which inhibits cytochrome P450 metabolic enzymes within mosquitoes and enhances the potency of the pyrethroid net component[2]. This was followed in 2023 by a strong recommendation for the deployment of pyrethroid-chlorfenapyr ITNs in areas of pyrethroid resistance[4]. These pyrethroid-chlorfenapyr ITNs combine a pyrethroid with a pyrrole insecticide with a separate mode of action, increasing the efficacy of the net[4]. A third type of ITN, pyrethroid-pyriproxyfen nets, were also given a conditional recommendation by WHO for deployment instead of pyrethroid-only nets in areas of pyrethroid resistance. These nets combine a pyrethroid with the insect growth regulator pyriproxyfen, which disrupts mosquito growth and reproduction[30]. Given that this study identified resistance to pyrethroids on the most populated island of the Bijagós, these recommendations have increased relevance in the area. Future studies would benefit from conducting additional WHO intensity bioassays with 5x and 10x deltamethrin concentrations to gather more information about the intensity of resistance. In addition, future studies would benefit from synergist-insecticide bioassays to further investigate metabolic resistance in this population[31]. Furthermore, it would be informative to collect insecticide bioassay data from other islands in the archipelago, as one limitation of this study is that experiments were conducted on Bubaque island only.

WGS analysis of *Anopheles gambiae* s.s. mosquitoes revealed the presence of six SNPs that have previously been associated with, or putatively associated with, insecticide resistance. These were the *vgsc* SNPs: T791M, L995F, N1570Y, A1746S, and P1874L, and the *gste2* SNP L119V. Two mutations: *vgsc* N1570Y and *gste2* L119V, were identified in resistant mosquitoes but not in susceptible mosquitoes. Previous research has shown that the N1570Y mutation increases pyrethroid resistance levels when in combination with L995F[12]. Analysis of linkage disequilibrium identified that the T791M and the A1746S mutations were in complete linkage disequilibrium. Twenty additional non-synonymous SNPs were identified across the four insecticide resistance genes which have not previously been associated with insecticide resistance. Of these, four mutations were present at > 5% frequency, including T154S, I126F and G26S in the *vgsc* gene and A65S in *ace1*. All four of these mutations were present in both resistant and susceptible mosquitoes. One of these SNPs, T154S, was previously reported

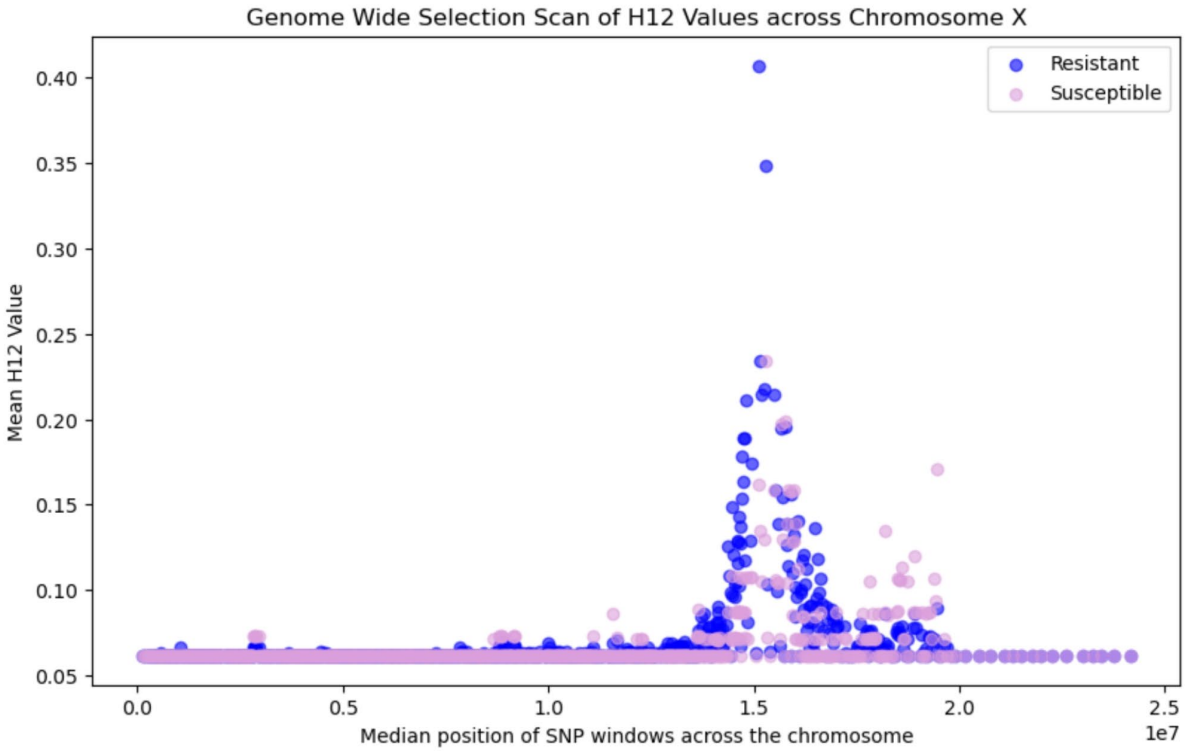


Fig. 2. Genome Wide Selection Scan: H_{12} values across chromosome X for resistant (blue) and susceptible (pink) mosquitoes. The peak in H_{12} values indicates a selective sweep over this portion of the chromosome.

in mosquitoes from the Bijagós[19], Guinea, Ivory Coast and the laboratory strain Kisumu originally isolated from Kenya[32], warranting further investigation. The other SNPs at frequency of > 5% have not previously been reported in the literature to our knowledge.

In our previous study, multiplex amplicon sequencing of > 200 mosquitoes collected in 2019 identified four mutations previously associated with, or putatively associated with, insecticide resistance: *vgsc* L995F, N1570Y, A1746S and *rdl* A296G[19]. In contrast, this study did not identify the *rdl* A296G mutation. This SNP was previously identified at very low frequency, so its absence may be explained by the smaller sample set of mosquitoes used in this study[19]. Of note, the mosquitoes sequenced in this study were collected over a smaller geographical area than those collected in 2019[19].

The presence of copy number variants (CNVs) in metabolic genes offers the possibility of elevated transcription of genes associated with enzymatic detoxification of insecticides, and has been strongly linked with insecticide resistance in *Anopheles* mosquitoes[33]. A number of CNVs of interest have previously been identified in *Anopheles gambiae* s.l[17]. Because we used WGA genomic data, we could not analyse CNVs using read coverage data or quantitative PCR. Therefore, we used a rapid screening method to indicate the likely presence of these specific CNVs through the analysis of soft clipped reads at previously identified breakpoints[17]. The only CNV allele indicated in this study that was unique to resistant mosquitoes was Cyp6aap_Dup7. This soft-clipping screen is only indicative and to confirm the presence or absence of CNVs in future studies, WGS should be conducted without prior WGA, allowing CNVs to be verified through the analysis of read coverage data. This is a small sample set of deltamethrin resistant mosquitoes, but the indicative presence of Cyp6aap_Dup7 is particularly interesting; this CNV contains two cytochrome P450 genes adjacent to the *cyp6* gene cluster, *cyp6aa1* and *cyp6aa2*, and contains part of the carboxylesterase *coeae60* gene[17]. Cyp6aap_Dup7 has been found previously in *An. gambiae* s.l. in Burkina Faso, Côte d'Ivoire, Ghana and Guinea[17] and *cyp* gene amplifications have previously been identified in mosquitoes from mainland Guinea-Bissau[7]. Copy number of *cyp6aa1* has previously been associated with deltamethrin resistance in *An. coluzzii* in Ghana[6], and copy number of *cyp6aa2* has been associated with deltamethrin resistance in *An. coluzzii* in Côte d'Ivoire[34]. Furthermore, transcription of *cyp6aap1* was higher in pyrethroid resistant *An. coluzzii* compared to susceptible colonies in Burkina Faso[35], and CYP6AA1 was overexpressed in pyrethroid resistant *An. gambiae* s.s. compared to susceptible colonies in Tanzania[36]. CYP6AA1 has been shown to metabolize pyrethroids in *Drosophila melanogaster*, and modelling indicates that it should be able to bind to permethrin and deltamethrin in *An. gambiae*[14]. One limitation of our approach is that we were only able to investigate CNVs which had previously been identified by the Ag1000G project[17], and only those that consistently identified with soft-clipped reads during our verification step. We are

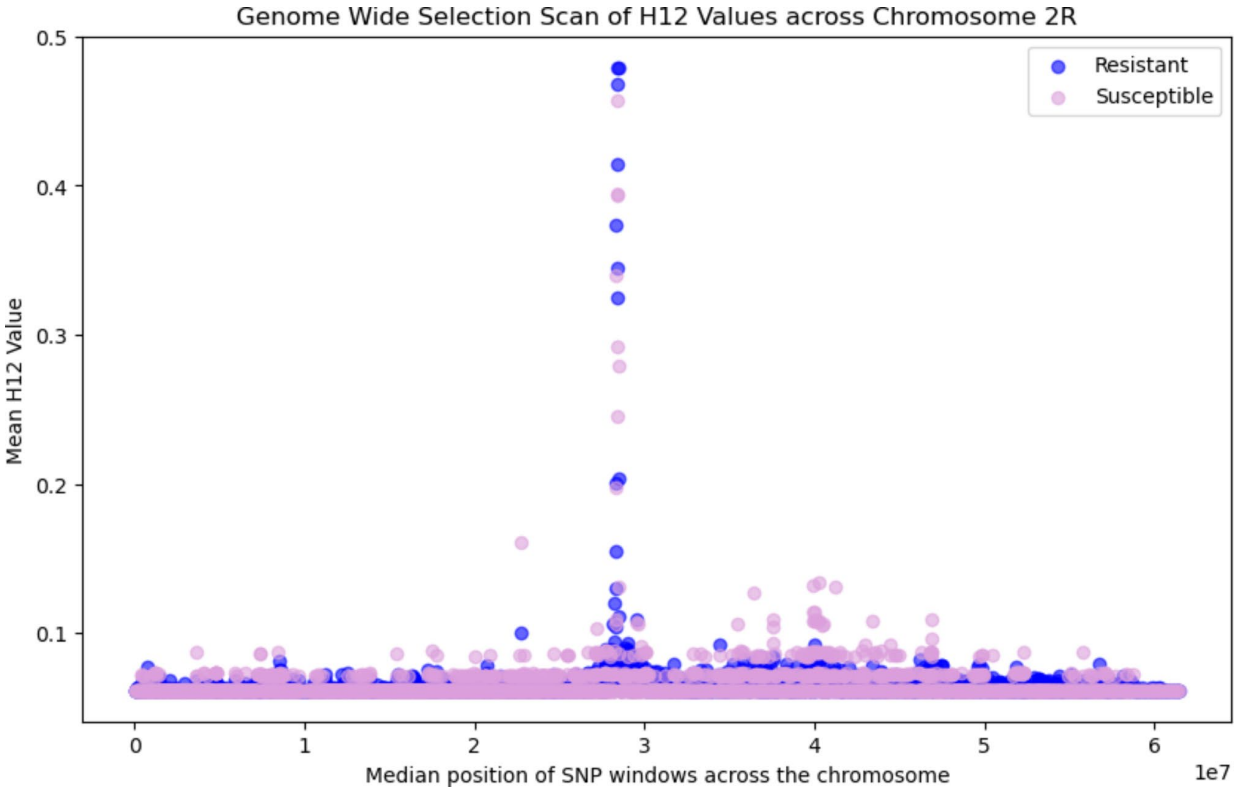


Fig. 3. Genome Wide Selection Scan: H_{12} values across chromosome 2R for resistant (blue) and susceptible (pink) mosquitoes. The peak in H_{12} values indicates a selective sweep over this portion of the chromosome.

only able to indicate the putative presence of CNV alleles using this screening method, and other CNV alleles associated with resistance may be present in the population.

Along with investigating SNPs and CNV alleles, we analysed the WGS data for signatures of differentiation and selection. Genes identified with significant fixation index (F_{ST}) scores included the *CPFL4* gene (Cuticular protein 4 from CPFL family - AGAP010905, F_{ST} = 0.122) on chromosome 3L. This gene family encodes structural cuticular proteins, for which differential expression has been associated with insecticide resistance in *An. arabiensis*[37]. We identified two selective sweeps in the *Anopheles gambiae* s.s. mosquito genomes. These sweeps are not associated with resistance in our small sample set, but they do indicate selection within the mosquito population. The selective sweep on chromosome X included the *cyp9k1* gene, which is a cytochrome P450 gene able to metabolise deltamethrin[15,16]. This gene has previously been associated with deltamethrin resistance in *An. gambiae* s.l[6]., *An. funestus*[16] and *An. coluzzii*[38], and has been found to be up-regulated in mosquitoes that are resistant to pyrethroids[39] and DDT[40]. This study identified the putative presence of one CNV allele containing the *cyp9k1* gene in the Bijagós population (Cyp9k1_Dup12)[17]. The selective sweep identified in chromosome 2R overlapped with several metabolic genes associated with insecticide resistance from the *cyp6aa/cyp6p* gene cluster, including *cyp6aa1*, *cyp6aa2*, *cyp6p1*, *cyp6p2*, *cyp6p3*, *cyp6p4*, *cyp6p5*, *cyp6p15p* and *cyp6ad1*, and the carboxylesterase gene *coeae60*. Overexpression of the *coeae60* gene has previously been associated with permethrin resistance in *An. coluzzii*[41]. Multiple studies have identified associations between these genes from the Cyp6 subfamily and pyrethroid resistance; upregulation of *cyp6m2* and *cyp6p3* has been identified in pyrethroid resistant *An. gambiae*[42] and *An. coluzzii*[38], expression of *cyp6p3* has also been shown to have a role in carbamate resistance, and expression of *cyp6p4* has been associated with pyrethroid resistance in *An. arabiensis*[42]. Of note, the putative CNV allele that we found indicated within our resistant mosquito population (Cyp6aap_Dup7) contains the genes *cyp6aa1*, *cyp6aa2*, and part of *coeae60*[17], associated with metabolic insecticide resistance. The indicative presence of CNV alleles and selective sweeps over metabolic genes indicates that metabolic resistance is a key component of insecticide resistance on the Archipelago. However, synergist assays are required to further understand the extent of metabolic resistance.

This study provides evidence of deltamethrin resistance and the first whole genome sequence data analysis of *An. gambiae* s.s. mosquitoes from the Bijagós Archipelago. Our genome-wide selection scan using H_{12} has identified two selective sweeps, both of which contain metabolic resistance genes that have previously been associated with metabolic resistance to pyrethroids. These selective sweeps indicate that the use of pyrethroid-based ITNs in the Bijagós has likely resulted in selection in metabolic genes. Future studies should include synergist-insecticide bioassays to investigate the presence of metabolic resistance in this mosquito population. The small sample sizes in this study have made drawing statistically significant conclusions between the presence

of SNPs, sweeps, haplotype clusters and the resistance phenotype difficult. However, our work provides a baseline for future studies with larger sample sizes to measure robust statistical significance. Furthermore, this study provides mortality data for deltamethrin resistance which can contribute towards evidence-based decision making in the selection of future vector control measures.

Methods

Mosquito sampling and insecticide resistance assays

Mosquito larvae were collected by dipping in three in-land pools close to Bubaque village on Bubaque island in November 2022. All larvae were collected in vials and reared in larval trays using their native water where possible. Upon eclosion, adult mosquitoes were maintained in 35 × 35 × 35 cm insect rearing cages (BugDorm, Taichung, Taiwan) with 10% sucrose solution. WHO tube tests were conducted with non-bloodfed females between 3 and 5 days old as per WHO guidelines[5]. Susceptibility to deltamethrin was measured at discriminating doses of 0.05% and at 5x concentration (0.25%) using deltamethrin-treated paper, which was procured from Universiti Sains Malaysia, Penang, Malaysia. Deltamethrin 0.05% concentration tests were conducted over 5 replicates, with an average of 20 mosquitoes introduced per tube. Deltamethrin 0.25% concentration tests were conducted over 3 replicates, with an average of 15 mosquitoes introduced per tube. Control tests were conducted using control test papers treated with oil as opposed to insecticide, which were procured from Universiti Sains Malaysia, Penang.

DNA extraction and species identification

DNA was extracted using Dynabeads™ (ThermoFisher Scientific Inc), following the standard protocol. Species identification was conducted using the Bass (2008) qPCR protocol[43] to distinguish *An. gambiae (An. gambiae* s.s, *An. coluzzii* and *An. gambiae – An. coluzzii* hybrids) from *An. melas.* The Santolamazza (2008) SINE200 PCR[44] was then used to distinguish between *An. gambiae* s.s, *An. coluzzii* and *An. gambiae – An. coluzzii* hybrids. DNA was quantified using the Qubit dsDNA HS Kit (Thermo Fisher Scientific Inc).

Whole genome sequencing

Mosquitoes identified as *Anopheles gambiae* s.s. were selected for whole genome sequencing. A small quantity of genetic material was available for whole genome sequencing, so samples were processed using Whole Genome Amplification (WGA) prior to sequencing. WGA was conducted using the REPLI-g® mini kit (Qiagen) following the standard protocol. This included 23 resistant, 10 susceptible and 9 control mosquitoes. Control mosquitoes were used in phenotypic bioassays as control mosquitoes for survival, so the resistance phenotype for controls is unknown. DNA was sequenced at Eurofins Genomics GmbH, Germany, using the Illumina Novaseq 6000 (2 × 150 bp paired-end read configuration).

Bioinformatic analysis

Raw FastQ files were trimmed using *trimmomatic* software (version 0.39, using the parameters LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20 MINLEN:36) to remove poor quality sequences[45]. Trimmed data was aligned to the *Anopheles gambiae* reference genome (Anopheles_gambiae.AgamP4.dna.toplevel.fa), using *bwa-mem* software (version 0.7.17-r1188, default parameters) to produce a BAM file for each sample. The *samtools* (version 1.12) functions fixmate and markdup were applied to the resulting BAM files[46]. SNPs were called using *GATK*'s HaplotypeCaller (version 4.1.4.1) using the option -ERC GVCF. Validated VCFs were merged into a database using *GATK*'s GenomicsDBImport function, and a combined VCF was created using *GATK*'s GenotypeGVCFs function[47]. This combined VCF was filtered for high quality variants using *bcftools* (version 1.17)[46]. Samples were retained if 40% of the genome had > 10-fold coverage. *GATK* VariantFiltration was then used to retain SNPs with high quality following *GATK* filter recommendations: QD > 5.0, QUAL > 30.0, SOR < 3.0, FS < 60.0, MQ > 40.0, MQRankSum > -12.5 and ReadPosRankSum > -8.0. *Bcftools* was used to retain SNPs with DP > 5 and QQ > 20, and to remove variants with a high proportion of missing genotypes (> 20%). The filtered VCF was phased using *beagle* (version 5.2)[48].

Linkage disequilibrium (LD) was calculated using the method of Rogers and Huff[26], using the *allel.rogers_huff_r* function in scikit-allel (https://scikit-allel.readthedocs.io/en/stable/), to provide an R^2 value for each combination of target SNPs in the *vgsc* gene.

Due to low sample availability in this study, samples underwent whole genome amplification (WGA) prior to whole genome sequencing. Therefore, we could not reliably identify CNVs using read coverage data, as coverage can be distorted by WGA. Instead, we used a rapid screening method to investigate the presence of specific CNV alleles of interest in the Bijagós mosquitoes by identifying soft clipping. This involved computing the proportion of reads that had been soft-clipped at CNV breakpoints previously described by Lucas et al. (2019)[17]. In doing so, we screened for 40 of these previously identified CNV alleles by analysing the proportion of soft-clipped reads in mapped reads with mapping quality score ≥ 10 (Supplementary Data 2). Normalised clipping was computed from BAM files as the proportion of reads at each position that had been soft clipped. Soft-clipping screening was conducted for a 'verification set' of samples available from The *Anopheles gambiae* 1000 Genomes Project phase 3 data resource, accessed via the MalariaGEN Ag3 API client, for which CNV data is available. Screening for soft-clipping at known breakpoints in each CNV of interest[17] was able to correctly identify the presence of CNVs, indicated by a high proportion of soft clipped reads present at the start, end, or both ends of the known CNV. A threshold of proportion soft clipped was set at 19.5% from the soft-clipping proportions computed for the verification set of samples (Supplementary Data 2).

F_{ST} was calculated between resistant and susceptible mosquitoes for non-overlapping 1 Kbp windows of the genome using the *allel.windowed_patterson_fst* function in scikit-allel. As described in Lucas et al. (2023)[6], peaks in F_{ST} could have been caused by extended haplotype homozygosity in a region due to a selective sweep, even if

that sweep were not due to phenotype, because the non-independence of SNPs within that window would lead to increased F_{ST} variance compared to other regions of the genome[6]. To identify peaks that were associated with phenotype, we performed 200 simulations in which the phenotype of the mosquitoes was randomly permuted and F_{ST} was recalculated. F_{ST} windows of interest were then kept if their F_{ST} value was higher than the 99th centile of these simulations. Additionally, we only kept F_{ST} windows of interest if the F_{ST} value was higher than three times the minimum distribution, as in Lucas et al. 2023[6], to reduce noise.

Haplotype clusters were determined by hierarchical clustering on pairwise genetic distance between haplotypes. Dendrograms were generated using the *scipy_cluster_hierarchy* function in SciPy (version 1.11.1), linkage method = single and metric = hamming. Clusters with > 20 similar haplotypes were included in Supplementary Data 1.

To detect regions of the genome undergoing selective sweeps, which may be due to selection pressure for insecticide resistance, we performed a genome wide selection scan using Garud's H_{12} statistic[28,29]. Garud's H_{12} was calculated using the *garuds_h* function in scikit-allel (https://scikit-allel.readthedocs.io/en/stable/), using phased biallelic SNPs in windows of 1000 SNPs. 200 Iterations of H_{12} were computed for samples, and the mean value for each 1Kbp window of SNPs was plotted.

Data availability

The processed datasets are available at ENA Project PRJEB71957, https://www.ebi.ac.uk/ena/browser/view/PRJEB71957. All code used to analyse the data can be found at https://github.com/sophiemoss/anopheles_popgen.

Received: 15 February 2024; Accepted: 23 September 2024
Published online: 01 October 2024

References

1. Bhatt, S. et al. The effect of malaria control on Plasmodium Falciparum in Africa between 2000 and 2015. *Nature*. **526**, 207 (2015).
2. Gleave, K., Lissenden, N., Chaplin, M., Choi, L. & Ranson, H. Piperonyl butoxide (PBO) combined with pyrethroids in insecticide-treated nets to prevent malaria in Africa. *Cochrane Database Syst. Rev.* **5**, CD012776 (2021).
3. World Health Organization. Framework for a national plan for monitoring and management of insecticide resistance in malaria vectors global malaria programme (2017). https://www.who.int/publications/i/item/9789241512138. Accessed 15 January 2024.
4. World Health Organization. World Malaria Report 2023. https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2023. Accessed 15 Jan 2024.
5. World Health Organization. Manual for monitoring insecticide resistance in mosquito vectors and selecting appropriate interventions (2022). https://www.who.int/publications/i/item/9789240051089. Accessed 15 Jan 2024.
6. Lucas, E. R. et al. Genome-wide association studies reveal novel loci associated with pyrethroid and organophosphate resistance in Anopheles gambiae and Anopheles coluzzii. *Nat. Commun.* **14**, 4946 (2023).
7. Anopheles gambiae 1000 Genomes Consortium. Genome variation and population structure among 1142 mosquitoes of the African malaria vector species Anopheles gambiae and Anopheles coluzzii. *Genome Res.* **30**, 1533–1546 (2020).
8. Njoroge, H. et al. Identification of a rapidly-spreading triple mutant for high-level metabolic insecticide resistance in Anopheles gambiae provides a real-time molecular diagnostic for antimalarial intervention deployment. *Mol. Ecol.* **31**, 4307–4318 (2022).
9. Hemingway, J. & Ranson, H. Insecticide Resistance in Insect vectors of Human Disease. *Annu. Rev. Entomol.* **45**, 371–391 (2000).
10. Martinez-Torres, D. et al. Molecular characterization of pyrethroid knockdown resistance (kdr) in the major malaria vector Anopheles gambiae s.s. *Insect Mol. Biol.* **7**, 179–184 (1998).
11. Ranson, H. et al. Identification of a point mutation in the voltage-gated sodium channel gene of Kenyan Anopheles gambiae associated with resistance to DDT and pyrethroids. *Insect Mol. Biol.* **9**, 491–497 (2000).
12. Jones, C. M. et al. Footprints of positive selection associated with a mutation (N1575Y) in the voltage-gated sodium channel of Anopheles gambiae. *Proc. Natl. Acad. Sci. U. S. A.* 109, 6614–6619 (2012).
13. Lucas, E. R. et al. A high throughput multi-locus insecticide resistance marker panel for tracking resistance emergence and spread in Anopheles gambiae. *Sci. Rep.* **9**, 13335 (2019).
14. Ibrahim, S. S. et al. Pyrethroid Resistance in the Major Malaria Vector Anopheles Funestus is exacerbated by overexpression and overactivity of the P450 CYP6AA1 across Africa. *Genes*. **9**, 140 (2018).
15. Vontas, J. et al. Rapid selection of a pyrethroid metabolic enzyme CYP9K1 by operational malaria control activities. *Proc. Natl. Acad. Sci.* **115**, 4619–4624 (2018).
16. Hearn, J. et al. Multi-omics analysis identifies a CYP9K1 haplotype conferring pyrethroid resistance in the malaria vector Anopheles Funestus in East Africa. *Mol. Ecol.* **31**, 3642–3657 (2022).
17. Lucas, E. R. et al. Whole-genome sequencing reveals high complexity of copy number variation at insecticide resistance loci in malaria mosquitoes. *Genome Res.* **29**, 1250–1261 (2019).
18. Hutchins, H. et al. Protocol for a cluster randomised placebo-controlled trial of adjunctive ivermectin mass drug administration for malaria control on the Bijagós Archipelago of Guinea-Bissau: the MATAMAL trial. *BMJ Open.* **13**, e072347 (2023).
19. Moss, S. et al. Genomic surveillance of Anopheles mosquitoes on the Bijagós Archipelago using custom targeted amplicon sequencing identifies mutations associated with insecticide resistance. *Parasit. Vectors*. **17**, 10 (2024).
20. Malaria Threat Map. World Health Organization. https://apps.who.int/malaria/maps/threats/. Accessed 15 Jan 2024.
21. Ant, T. et al. A survey of Anopheles species composition and insecticide resistance on the island of Bubaque, Bijagos Archipelago, Guinea – bissau. *Malar. J.* 1–9. https://doi.org/10.1186/s12936-020-3115-1 (2020).
22. Oliveira, E. et al. High levels of hybridization between molecular forms of Anopheles gambiae from Guinea Bissau. *J. Med. Entomol.* **45**, 1057–1063 (2008).
23. Marsden, C. D. et al. Asymmetric introgression between the M and S forms of the malaria vector, Anopheles gambiae, maintains divergence despite extensive hybridization. *Mol. Ecol.* **20**, 4983–4994 (2011).
24. Vicente, J. L. et al. Massive introgression drives species radiation at the range limit of Anopheles gambiae. *Sci. Rep.* **7**, 46451 (2017).
25. Torre, A. D. et al. Speciation within the Anopheles gambiae complex: high-throughput whole genome sequencing reveals evidence of a putative new cryptic taxon in 'far-west' Africa. Preprint at (2024). https://doi.org/10.21203/rs.3.rs-3914444/v1
26. Rogers, A. R. & Huff, C. Linkage disequilibrium between loci with unknown phase. *Genetics*. **182**, 839–844 (2009).
27. Garud, N. R., Messer, P. W., Buzbas, E. O. & Petrov, D. A. Recent selective sweeps in North American Drosophila melanogaster Show signatures of Soft sweeps. *PLOS Genet.* **11**, e1005004 (2015).
28. Garud, N. R. & Rosenberg, N. A. Enhancing the mathematical properties of new haplotype homozygosity statistics for the detection of selective sweeps. *Theor. Popul. Biol.* **102**, 94–101 (2015).
29. Anopheles Gambiae 1000 Genomes Consortium. Genetic diversity of the African malaria vector Anopheles gambiae. *Nature* **552**, 96–100 (2017).

30. Grisales, N. et al. Pyriproxyfen-treated bed nets reduce reproductive fitness and longevity of pyrethroid-resistant Anopheles gambiae under laboratory and field conditions. *Malar. J.* **20**, 273 (2021).
31. Silva, R., Mavridis, K., Vontas, J., Rodrigues, A. & Costa, H. Monitoring and molecular profiling of contemporary insecticide resistance status of malaria vectors in Guinea – Bissau. *Acta Trop.* **206**, 105440 (2020).
32. Campos, M. et al. High-throughput barcoding method for the genetic surveillance of insecticide resistance and species identification in Anopheles gambiae complex malaria vectors. *Sci. Rep.* **12**, 13893 (2022).
33. Weetman, D., Djogbenou, L. S. & Lucas, E. Copy number variation (CNV) and insecticide resistance in mosquitoes: evolving knowledge or an evolving problem? *Curr. Opin. Insect Sci.* **27**, 82–88 (2018).
34. Widespread occurrence of. Copy number variants and fixation of pyrethroid target site resistance in Anopheles gambiae (s.l.) from southern Côte d'Ivoire. *Curr. Res. Parasitol. Vector-Borne Dis.* **3**, 100117 (2023).
35. Kwiatkowska, R. M. et al. Dissecting the mechanisms responsible for the multiple insecticide resistance phenotype in Anopheles gambiae s.s., M form, from Vallée Du Kou, Burkina Faso. *Gene.* **519**, 98–106 (2013).
36. Matowo, J. et al. Expression of pyrethroid metabolizing P450 enzymes characterizes highly resistant Anopheles vector species targeted by successful deployment of PBO-treated bednets in Tanzania. *PLOS ONE.* **17**, e0249440 (2022).
37. Simma, E. A. et al. Genome-wide gene expression profiling reveals that cuticle alterations and P450 detoxification are associated with deltamethrin and DDT resistance in Anopheles arabiensis populations from Ethiopia. *Pest Manag Sci.* **75**, 1808–1818 (2019).
38. Bamou, R. et al. Increased prevalence of insecticide resistance in Anopheles coluzzii populations in the city of Yaoundé, Cameroon and influence on pyrethroid-only treated bed net efficacy. *Parasite* **28**, 8 (2021).
39. Ngufor, C. et al. Insecticide resistance profile of Anopheles gambiae from a phase II field station in Cové, southern Benin: implications for the evaluation of novel vector control products. *Malar. J.* **14**, 464 (2015).
40. Tene, B. F. et al. Resistance to DDT in an urban setting: common mechanisms implicated in both M and S forms of Anopheles gambiae in the City of Yaoundé Cameroon. *PLOS ONE.* **8**, e61408 (2013).
41. Main, B. J., Everitt, A., Cornel, A. J., Hormozdiari, F. & Lanzaro, G. C. Genetic variation associated with increased insecticide resistance in the malaria mosquito, Anopheles coluzzii. *Parasit. Vectors*. **11**, 225 (2018).
42. Donnelly, M. J., Isaacs, A. T. & Weetman, D. Identification validation, and application of molecular diagnostics for insecticide resistance in malaria vectors. *Trends Parasitol.* **32**, 197–206 (2016).
43. Bass, C., Williamson, M. S. & Field, L. M. Development of a multiplex real-time PCR assay for identification of members of the Anopheles gambiae species complex. *Acta Trop.* **107**, 50–53 (2008).
44. Santolamazza, F. et al. Insertion polymorphisms of SINE200 retrotransposons within speciation islands of Anopheles gambiae molecular forms. *Malar. J.* **7**, 163 (2008).
45. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. **30**, 2114–2120 (2014).
46. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *GigaScience* **10** (2021). https://doi.org/10.1093/gigascience/giaboo8.
47. McKenna, A. et al. The genome analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
48. Browning, B. L., Tian, X., Zhou, Y. & Browning, S. R. Fast two-stage phasing of large-scale sequence data. *Am. J. Hum. Genet.* **108**, 1880–1890 (2021).

Acknowledgements

Thank you to Ansumane Cassama for your assistance with mosquito larvae collection and rearing. We wish to thank the communities of the Bijagós Archipelago, Guinea-Bissau.

Author contributions

SM and RJ conducted insecticide resistance testing with guidance from EP. MK assisted with species identification. SM conducted species identification assays, DNA extractions and bioinformatic analysis, with guidance from MH, HAP, EC, and TGC. ETS and AR assisted with logistics within Guinea-Bissau. SK, TGC, AL and SC supervised the project.

Funding

SM is funded by Medical Research Council UK (Grant No. MR/N013638/1). TGC, SC and MH are funded by UKRI grants (BBSRC BB/X018156/1; MRC IAA2129, MR/R026297/1, and MR/X005895/1; EPSRC EP/Y018842/1). AL, EP, ES and RTJ are funded by Joint Global Health Trials Scheme (MRC, Wellcome Trust, UKRI, NIHR, Grant no. MR/S005013/1). HAP is funded by a LIDO-DTP PhD studentship. EC is funded by a Medical Research Council LID PhD Studentship.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-73996-3.

Correspondence and requests for materials should be addressed to S.M.

Supplementary Data 1

Supplementary Table 1: Insecticide resistance SNPs of interest

Position	**Chromosome**	**SNP**	**Gene**
2390177	2L	R254K	*vgsc*
2391228	2L	V402L	*vgsc*
2399997	2L	D466H	*vgsc*
2400071	2L	M490I	*vgsc*
2402466	2L	G531V	*vgsc*
2407967	2L	Q697P	*vgsc*
2416980	2L	T791M	*vgsc*
2422651	2L	L995S	*vgsc*
2422652	2L	L995F	*vgsc*
2429556	2L	V1507I	*vgsc*
2429617	2L	I1527T	*vgsc*
2429745	2L	N1570Y	*vgsc*
2429897	2L	E1597G	*vgsc*
2429915	2L	K1603T	*vgsc*
2430424	2L	A1746S	*vgsc*
2430817	2L	V1853I	*vgsc*
2430863	2L	I1868T	*vgsc*
2430880	2L	P1874S	*vgsc*
2430881	2L	P1874L	*vgsc*
2431061	2L	A1934V	*vgsc*
2431079	2L	I1940T	*vgsc*
28598166	3R	I114T	*gste2*
28598057	3R	F120L	*gste2*
28598062	3R	L119V	*gste2*
25429236	2L	A296G	*rdl*
25429235	2L	A296S	*rdl*
3492074	2R	G280S*	*ace1*

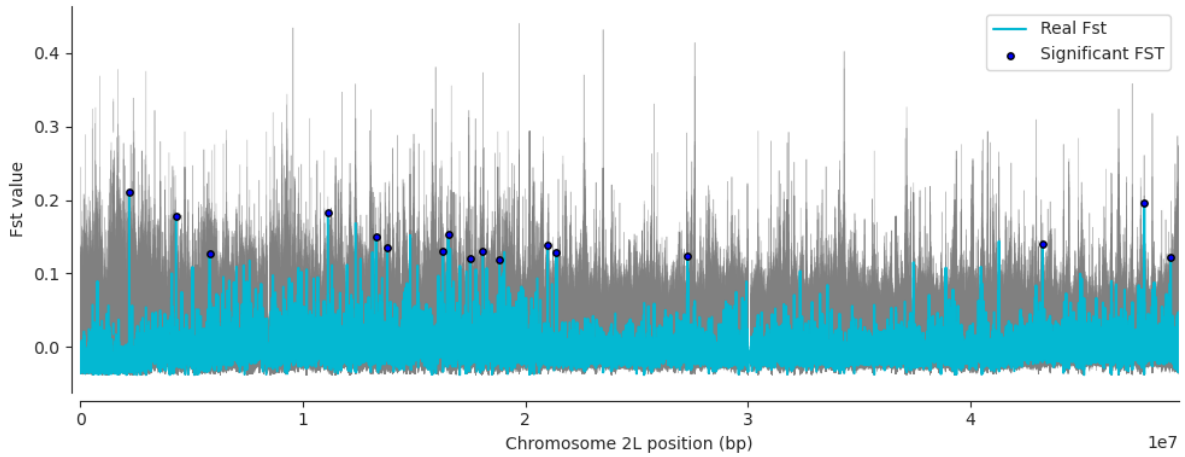
Supplementary Table 2: Linkage disequilibrium R2 values for pairwise combinations of insecticide resistance mutations in vgsc

R^2 value	2416980 (T791M)	2422652 (L995F)	2429745 (N1570Y)	2430424 (A1746S)	2430881 (P1874L)
2416980 (T791M)	0.000	0.112	0.008	1.000	0.007
2422652 (L995F)	0.112	0.000	0.372	0.112	0.301
2429745 (N1570Y)	0.008	0.372	0.000	0.008	0.023
2430424 (A1746S)	1.000	0.112	0.008	0.000	0.007

2430881 (P1874L)	0.007	0.301	0.023	0.007	0.000

Supplementary Information: F_{ST}

F_{ST} was calculated between resistant and susceptible mosquitoes. This was calculated over 1000bp windows for each chromosome using the scikit-allel, *allel.windowed_patterson_fst* function. F_{ST} values can be seen plotted in blue. To identify significant F_{ST} values we followed the methodology in Lucas *et al* (2023)[2]: we found the difference between the smallest F_{ST} value and the mode of the distribution, and considered an outlier to be any value more than three times this distance away from the mode. We also calculated 200 permutations of $F_{ST,}$ which can be seen plotted in grey. F_{ST} values were considered significant if they were greater than the 99th percentile of the 200 permutations for that window. 1000bp windows with significant F_{ST} values are highlighted with dark blue dots.



Supplementary Figure 1: FST values between resistant and susceptible mosquitoes across Chromosome 2L

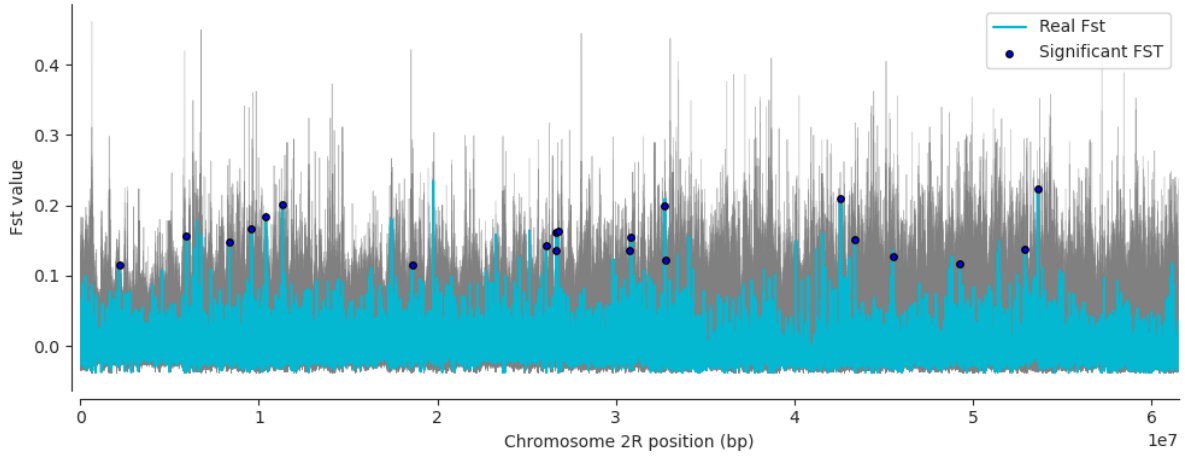
There were 16 provisional windows of interest identified on chromosome 2L, with F_{ST} values greater than the 99th percentile of 200 permuted F_{ST} values.

Supplementary Table 3: Provisional windows of interest identified on chromosome 2L, where significant FST values were calculated between resistant and susceptible mosquitoes.

Window (chromosome 2L)	**'Real' F_{ST} Value**	**99th Percentile of 200 permutated F_{ST} values**	**Feature(s) in overlapping window**
2194206 -2195205	0.211	0.107	-
4299206 - 4300205	0.177	0.074	AGAP004835 – protein coding gene, unspecified product
5811206 - 5812205	0.127	0.084	-
11122206 - 11123205	0.183	0.176	-
13283206 - 13284205	0.150	0.101	AGAP029530 - protein coding gene - unspecified product
13785206 - 13786205	0.135	0.132	-

16287206 - 16288205	0.130	0.118	AGAP005505 - RNA-binding protein 45
16529206 - 16530205	0.153	0.085	AGAP005510 - oxysterol binding protein-like 9
17512206 - 17513205	0.120	0.102	-
18079206 - 18080205	0.129	0.090	AGAP005637 - aldehyde oxidase
18832206 - 18833205	0.119	0.064	-
20991206 - 20992205	0.139	0.103	-
21390206 - 21391205	0.129	0.080	-
27276206 - 27277205	0.123	0.086	AGAP029236 - putative metabotropic glutamate receptor 1
43230206 - 43231205	0.140	0.122	AGAP007133 – protein coding gene - unspecified product
47794206 - 47795205	0.196	0.168	AGAP007576 – protein coding gene - unspecified product
48981206 - 48982205	0.121	0.048	AGAP029465 – protein coding gene -unspecified product

Chromosome 2R



Supplementary Figure 2: FST values between resistant and susceptible mosquitoes across Chromosome 2R

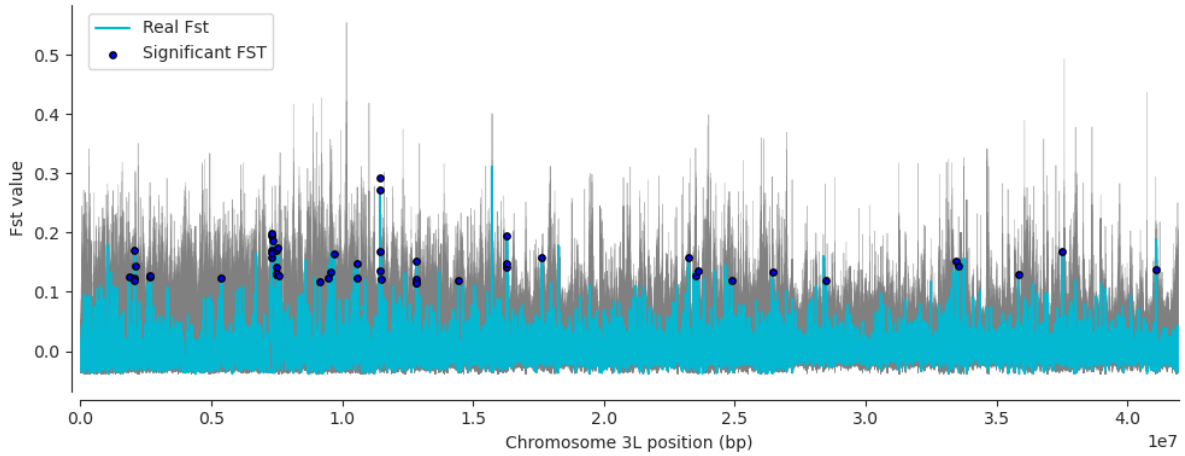
There were 21 provisional windows of interest identified on chromosome 2R, with F_{ST} values greater than the 99th percentile of 200 permuted F_{ST} values.

Supplementary Table 4: Provisional windows of interest identified on chromosome 2R, where significant FST values were calculated between resistant and susceptible mosquitoes.

Window (chromosome 2R)	**'Real' F_{ST} Value**	**99th Percentile of 200 permutated F_{ST} values**	**Feature(s) in overlapping window**
2178086 – 2179085	0.116	0.088	AGAP001273 – karyopherin (importin) alpha 4
5946086 -5947085	0.156	0.082	AGAP001535 – histone-lysine N-methyltransferase ASH1L

8353086 – 8354085	0.148	0.116	AGAP001683 – calcium/calmodulin-dependent serine protein kinase
9578086 – 9579085	0.166	0.074	-
10394086 – 10395085	0.184	0.112	AGAP001786 – protein coding gene – unspecified product
11333086 – 11334085	0.201	0.085	-
18600086 – 18601085	0.115	0.049	AGAP002288 - hydroxymethylglutaryl-CoA reductase (NADPH)
26085086 – 26086085	0.142	0.121	AGAP002748 -protein kinase C
26645086 – 26646085	0.161	0.046	AGAP002748 -protein kinase C
26649086 – 26650085	0.137	0.052	AGAP002748 -protein kinase C
26754086 – 26755085	0.163	0.115	AGAP002748 -protein kinase C
30763086 – 30764085	0.136	0.128	AGAP002997 – tumor necrosis factor alpha-induced protein 8-like protein
30859086 – 30860085	0.155	0.089	AGAP029576 – protein coding gene – unspecified product
32748086 – 32749085	0.199	0.120	AGAP003115 – protein coding gene – unspecified product
32755086 – 32756085	0.123	0.073	-
42603086 – 42604085	0.209	0.082	-
43388086 – 43389085	0.151	0.096	AGAP003795 – protein coding gene – unspecified product AGAP003796 – cyclin-dependent kinases regulatory subunit 1
45526086 – 45527085	0.127	0.091	-
49260086 – 49261085	0.116	0.109	AGAP004059 – Ca-activated cl channel protein
52913086 – 52914085	0.138	0.071	AGAP004232 – pellino
53629086 – 53630085	0.223	0.178	-

Chromosome 3L



Supplementary Figure 3: FST values between resistant and susceptible mosquitoes across Chromosome 3L

There were 51 provisional windows of interest identified on chromosome 3L, with F_{ST} values greater than the 99th percentile of 200 permuted F_{ST} values.

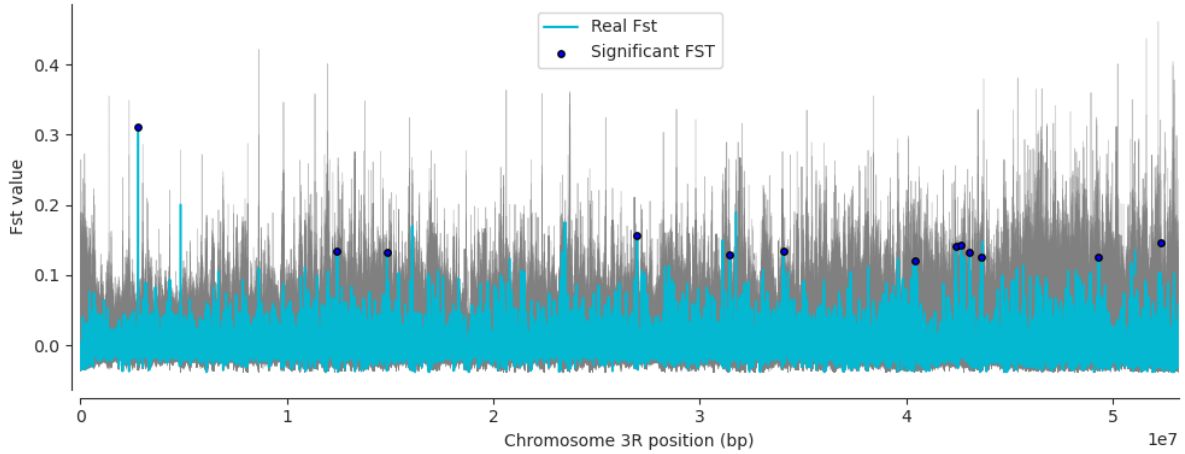
Supplementary Table 5: Provisional windows of interest identified on chromosome 3L, where significant FST values were calculated between resistant and susceptible mosquitoes.

Window (chromosome 3L)	**'Real' F_{ST} Value**	**99th Percentile of 200 permutated F_{ST} values**	**Feature(s) in overlapping window**
1849825 - 1850824	0.126	0.085	-
2056825 - 2057824	0.123	0.119	-
2066825 - 2067824	0.170	0.099	-
2068825 - 2069824	0.119	0.100	-
2078825 - 2079824	0.144	0.128	-
2659825 - 2660824	0.125	0.081	-
2675825 - 2676824	0.127	0.111	-
5379825 - 5380824	0.123	0.113	-
7299825 - 7300824	0.171	0.119	-
7300825 - 7301824	0.195	0.136	-
7322825 - 7323824	0.158	0.152	AGAP010621 -glycine transporter
7323825 - 7324824	0.198	0.152	AGAP010621 -glycine transporter
7324825 - 7325824	0.166	0.165	AGAP010621 -glycine transporter
7328825 - 7329824	0.186	0.179	-
7483825 - 7484824	0.141	0.072	-
7484825 - 7485824	0.130	0.079	-
7485825 - 7486824	0.132	0.062	-
7486825 - 7487824	0.171	0.098	-
7553825 - 7554824	0.174	0.155	AGAP029077 – protein coding - unspecified product
7597825 - 7598824	0.128	0.086	-
9128825 - 9129824	0.116	0.109	AGAP010737 – protein coding - unspecified product
9444825 - 9445824	0.124	0.099	-
9554825 - 9555824	0.134	0.111	-
9696825 - 9697824	0.164	0.129	AGAP029992 – protein coding - unspecified product
10576825 -10577824	0.147	0.083	AGAP010800 -nuclear protein NHN1
10579825 - 10580824	0.124	0.067	AGAP010800 -nuclear protein NHN1
11424825 - 11425824	0.135	0.089	AGAP029564 - Ig-like domain-containing protein

11428825 - 11429824	0.293	0.244	AGAP029564 - Ig-like domain-containing protein
11444825 - 11445824	0.271	0.188	AGAP029564 - Ig-like domain-containing protein
11465825 - 11466824	0.167	0.165	-
11475825 - 11476824	0.122	0.087	-
12825825 - 12826824	0.115	0.071	-
12829825 - 12830824	0.122	0.113	AGAP010905 - cuticular protein 4 from CPFL family
12830825 - 12831824	0.152	0.130	
12845825 - 12846824	0.114	0.078	
14423825 - 14424824	0.120	0.115	AGAP010978 - nuclear pore complex protein Nup62
16288825 - 16289824	0.141	0.114	AGAP029789 - DUF4806 domain-containing protein AGAP029790 – protein coding gene - unspecified product
16290825 - 16291824	0.195	0.119	AGAP029790 – protein coding gene - unspecified product AGAP028685 - DUF4806 domain-containing protein
16296825 - 16297824	0.148	0.094	AGAP028688 – protein coding gene - unspecified product
17622825 - 17623824	0.159	0.129	-
23222825 - 23223824	0.158	0.144	-
23501825 - 23502824	0.128	0.103	-
23591825 - 23592824	0.137	0.079	AGAP011395 - patched 1
24875825 - 24876824	0.119	0.115	-
26433825 - 26434824	0.133	0.067	AGAP011520 - CS domain-containing protein
28495825 - 28496824	0.119	0.059	-
33442825 - 33443824	0.151	0.081	-
33547825 - 33548824	0.143	0.084	-
35824825 - 35825824	0.129	0.105	AGAP011980 – protein coding gene - unspecified product
37509825 - 37510824	0.168	0.124	AGAP012087 - 7SK snRNA methylphosphate capping enzyme

			AGAP012088 - stromal membrane-associated protein
41074825 - 41075824	0.138	0.091	AGAP012358 - ATP-dependent RNA helicase DDX52/ROK1

Chromosome 3R



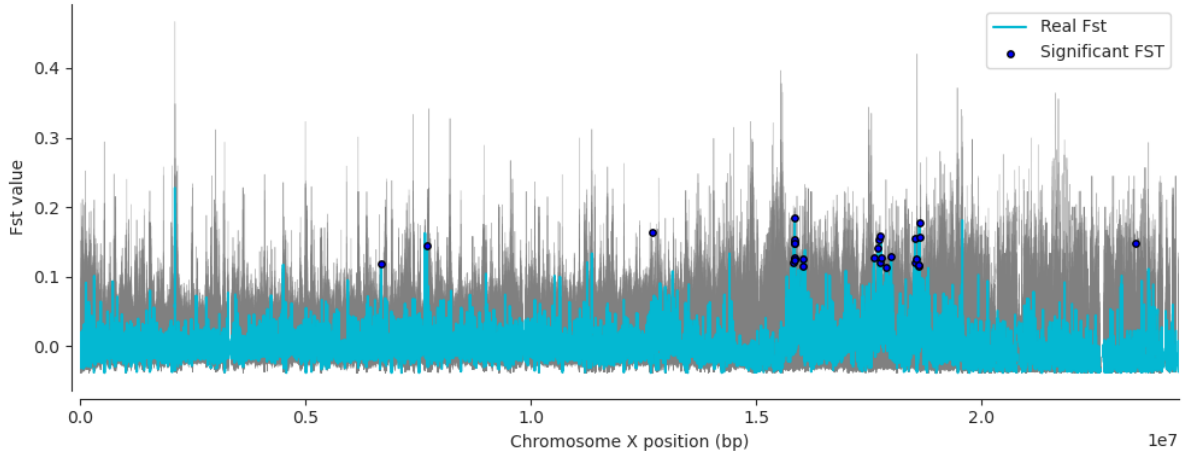
Supplementary Figure 4: FST values between resistant and susceptible mosquitoes across Chromosome 3R

There were 13 provisional windows of interest identified on chromosome 3R, with F_{ST} values greater than the 99th percentile of 200 permuted F_{ST} values.

Supplementary Table 6: Provisional windows of interest identified on chromosome 3R, where significant FST values were calculated between resistant and susceptible mosquitoes.

Window (chromosome 3R)	**'Real' F_{ST} Value**	**99th Percentile of 200 permutated F_{ST} values**	**Feature(s) in overlapping window**
2791201 - 2792200	0.311	0.107	AGAP010407 - Elongator complex protein 4
12436201 - 12437200	0.133	0.112	AGAP010884 – protein coding gene – unspecified product
14843201 - 14844200	0.131	0.078	-
26950201 - 26951200	0.156	0.101	AGAP011540 - dynein intermediate chain 2, axonemal
31449201 - 31450200	0.128	0.108	AGAP029516 – protein coding gene - unspecified product
34047201 - 34048200	0.135	0.085	AGAP029760 - clustered mitochondria protein homolog
40445201 - 40446200	0.120	0.099	-
42389201 - 42390200	0.141	0.086	-
42641201 - 42642200	0.143	0.095	-
43054201 - 43055200	0.132	0.078	-
43618201 - 43619200	0.126	0.098	-
49311201 - 49312200	0.125	0.055	-
52344201 - 52345200	0.146	0.140	-

Chromosome X



Supplementary Figure 5: FST values between resistant and susceptible mosquitoes across Chromosome X

There were 30 provisional windows of interest identified on chromosome X, with F_{ST} values greater than the 99th percentile of 200 permuted F_{ST} values.

Supplementary Table 7: Provisional windows of interest identified on chromosome X, where significant FST values were calculated between resistant and susceptible mosquitoes.

Window (chromosome X)	**'Real' F_{ST} Value**	**99th Percentile of 200 permutated F_{ST} values**	**Feature(s) in overlapping window**
6676973 - 6677972	0.119	0.072	-
7695973 - 7696972	0.145	0.112	AGAP000433 - Ras-related protein Rab-39B
12708973 - 12709972	0.164	0.090	-
15836973 - 15837972	0.121	0.095	AGAP000859 – protein coding – unspecified product
15842973 - 15843972	0.125	0.094	AGAP029672 - protein tweety homolog
15843973 - 15844972	0.128	0.112	AGAP029672 - protein tweety homolog
15846973 - 15847972	0.150	0.104	AGAP029672 - protein tweety homolog AGAP029671 – protein coding gene - unspecified product
15854973 - 15855972	0.148	0.112	-
15855973 - 15856972	0.154	0.121	-
15856973 - 15857972	0.185	0.165	-
15857973 - 15858972	0.149	0.120	-
15867973 - 15868972	0.124	0.117	AGAP000861- protein coding gene - unspecified product
16030973 - 16031972	0.125	0.105	AGAP000863 – lachesin
16043973 - 16044972	0.115	0.081	-

17620973 - 17621972	0.127	0.044	AGAP000932 - unspecified product
17703973 - 17704972	0.141	0.062	-
17738973 - 17739972	0.154	0.051	-
17745973 - 17746972	0.121	0.069	-
17766973 - 17767972	0.158	0.070	-
17772973 - 17773972	0.127	0.078	-
17900973 - 17901972	0.114	0.098	-
17999973 - 18000972	0.130	0.073	-
18523973 - 18524972	0.120	0.105	-
18531973 - 18532972	0.155	0.146	-
18567973 - 18568972	0.125	0.075	-
18621973 - 18622972	0.115	0.060	-
18623973 - 18624972	0.118	0.068	-
18624973 - 18625972	0.157	0.079	-
18627973 - 18628972	0.178	0.058	-
23417973 - 23418972	0.147	0.089	-

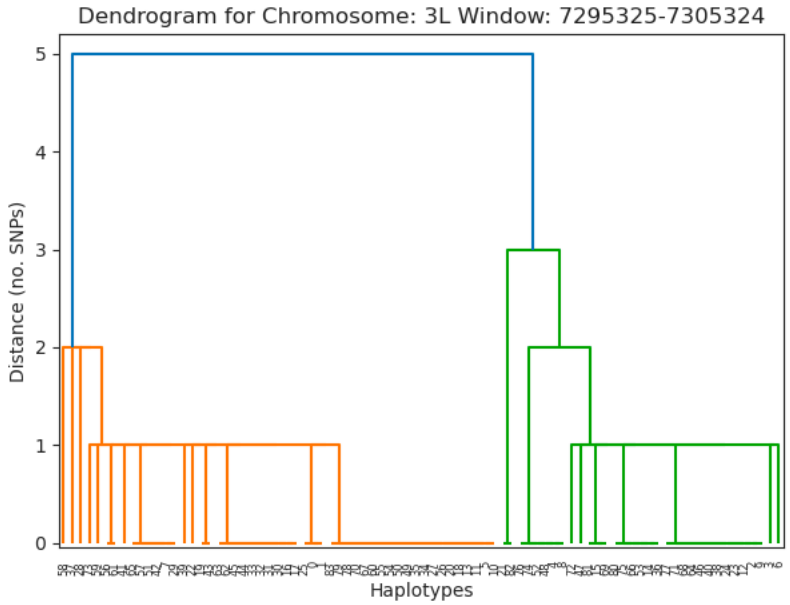
Haplotype association

We investigated the presence of swept haplotypes within each putative window of interest identified during the F_{ST} analysis. Hierarchical clustering was performed on haplotypes in each window of interest using pairwise genetic distances between haplotypes. This hierarchical clustering identified clusters of > 20 haplotypes that are highly similar to each other, which we refer to as 'haplotype clusters'. Dendrograms were generated for each window of interest and the tree was cut at 1% of the maximum distance identified for each window. Dendrograms and genomic positions are included below. Genomic positions given are relative to each chromosome. Associations between haplotype clusters and insecticide resistance phenotype were explored. Mosquitoes are diploid and so have two haplotypes. For each haplotype cluster, mosquitoes were given a score of 2 if both of their haplotypes matched with a haplotype in the cluster, a score of 1 if 1 of their haplotypes matched with a haplotype in the cluster, or a score of 0 if neither of their haplotypes matched with haplotypes in the cluster. These scores were then aggregated by phenotype for each window and can be seen in the tables below. Sample sizes were not large enough to compute robust regression models to determine if there was a significant association between these haplotype clusters and resistance phenotype. However, indicative analysis does not reveal any of these haplotype clusters to be shared by a majority of resistant mosquitoes (see tables below). We conclude that a larger sample size is required to reveal associations between haplotype clusters and the resistance haplotype.

Chromosome 3L

Supplementary Table 8: Window 7295325 – 7305324 included one cluster of ≥20 haplotypes.

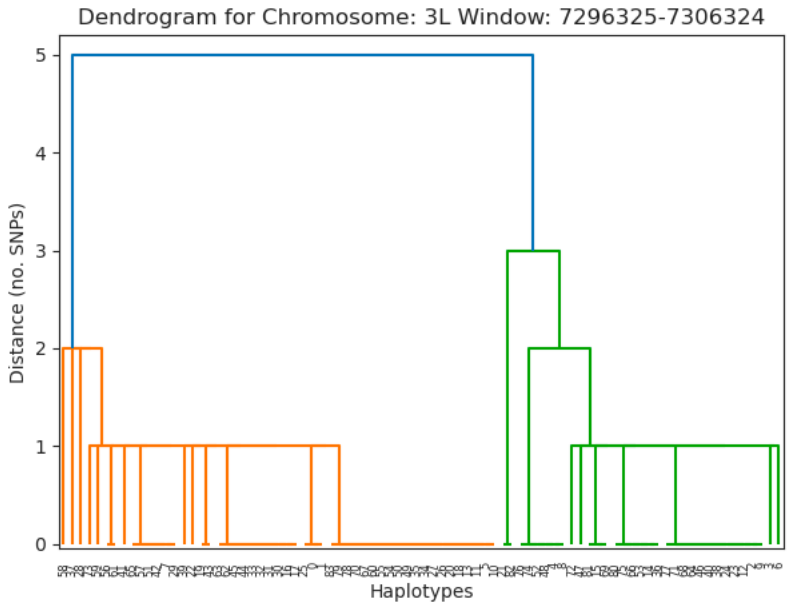
Window 7295325 – 7305324 haplotype cluster		
Phenotype	Score	% of mosquito haplotypes
Resistant	11	11/46 (23.9%)
Susceptible	5	5/20 (25.0%)
Control	4	4/18 (22.2%)



Supplementary Figure 6: Dendrogram for Chromosome 3L, Window 7295325 - 7305324

Supplementary Table 9: Window 7296325-7306324 included one cluster of ≥ 20 haplotypes.

Window 7296325-7306324 haplotype cluster		
Phenotype	Score	% of mosquito haplotypes
Resistant	11	11/46 (23.9%)
Susceptible	5	5/20 (25.0%)
Control	4	4/18 (22.2%)

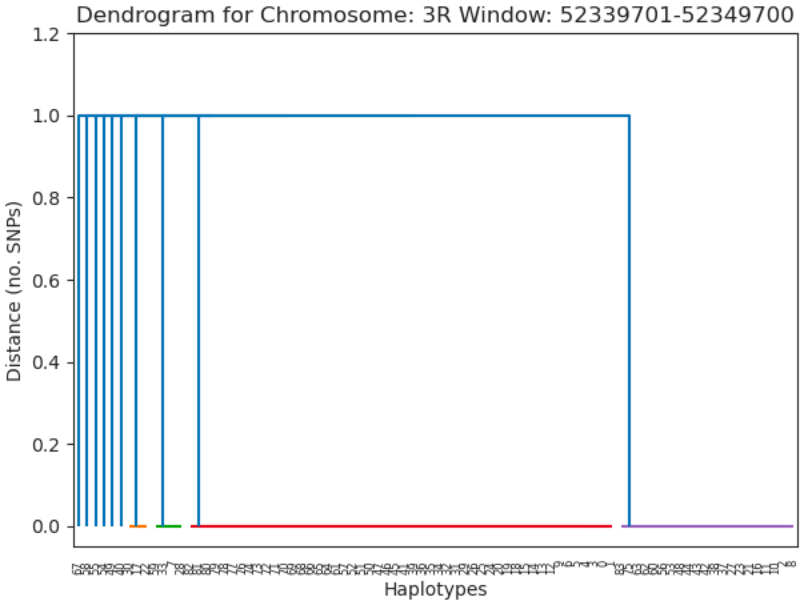


Supplementary Figure 7: Dendrogram for Chromosome 3L, Window 7296325 - 7306324

Chromosome 3R

Supplementary Table 10: Window 52339701 – 52349700 included two clusters of ≥20 haplotypes.

Window 52339701 – 52349700 haplotype cluster (red)		
Phenotype	Score	% of mosquito haplotypes
Resistant	22	22/46 (47.8%)
Susceptible	17	17/20 (85.0%)
Control	11	11/18 (61.1%)
Window 52339701 – 52349700 haplotype cluster (purple)		
Phenotype	Score	% of mosquito haplotypes
Resistant	14	14/46 (30.4%)
Susceptible	2	2/20 (10.0%)
Control	5	5/18 (27.8%)

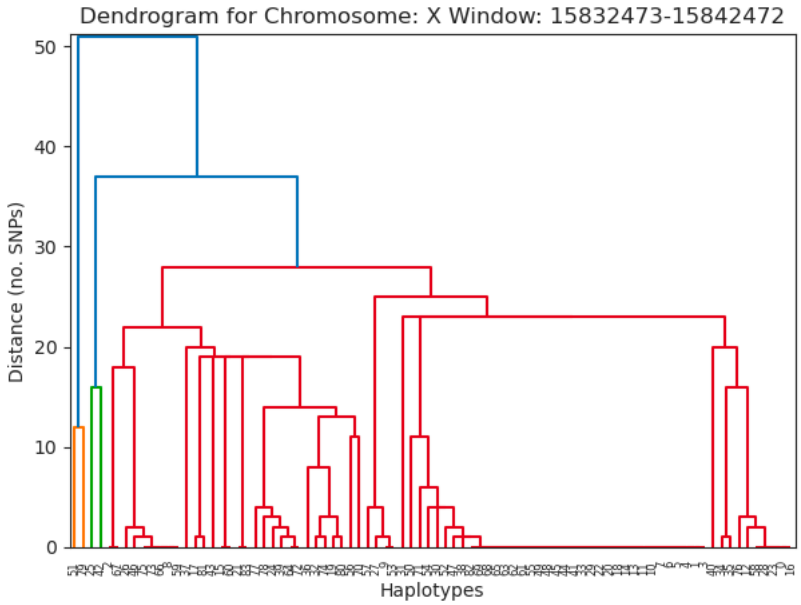


Supplementary Figure 8: Dendrogram for Chromosome 3R, Window 52339701-52349700

Chromosome X

Supplementary Table 11: Window 15832473-15842472 included one cluster of ≥20 haplotypes.

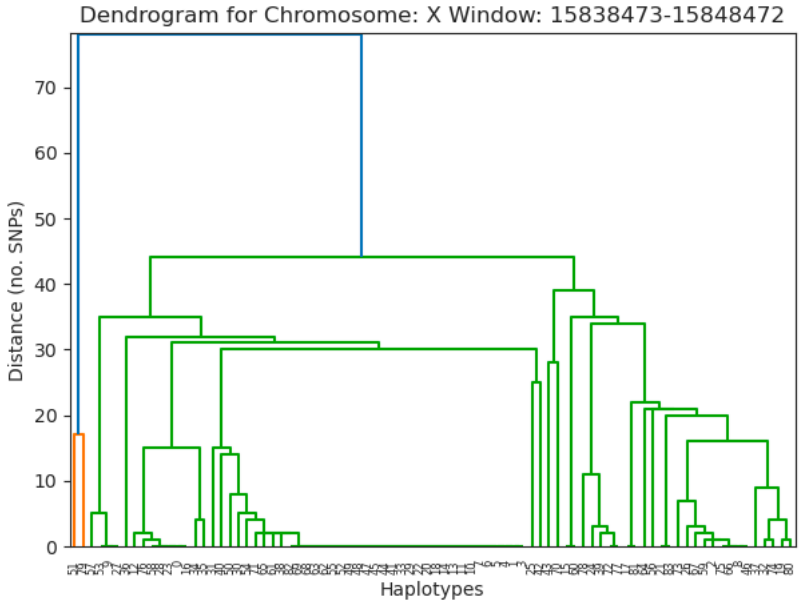
Window 15832473-15842472 haplotype cluster		
Phenotype	Score	% of mosquito haplotypes
Resistant	14	14/46 (30.4%)
Susceptible	4	4/20 (20.0%)
Control	10	10/18 (55.6%)



Supplementary Figure 9: Dendrogram for Chromosome X, Window 15832473 - 15842472

Supplementary Table 12: Window 15838473-15848472 had one cluster of ≥ 20 haplotypes.

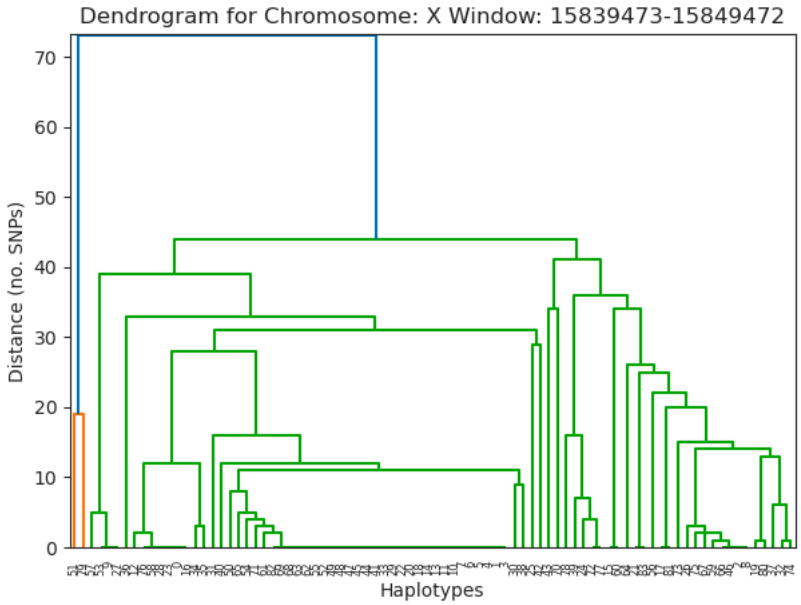
Window 15838473-15848472 haplotype cluster		
Phenotype	Score	% of mosquito haplotypes
Resistant	15	15/46 (32.6%)
Susceptible	3	3/20 (15.0%)
Control	10	10/18 (55.6%)



Supplementary Figure 10: Dendrogram for Chromosome X, Window 15838473 - 15848472

Supplementary Table 13: Window 15839473-15849472 had one cluster with ≥ 20 haplotypes.

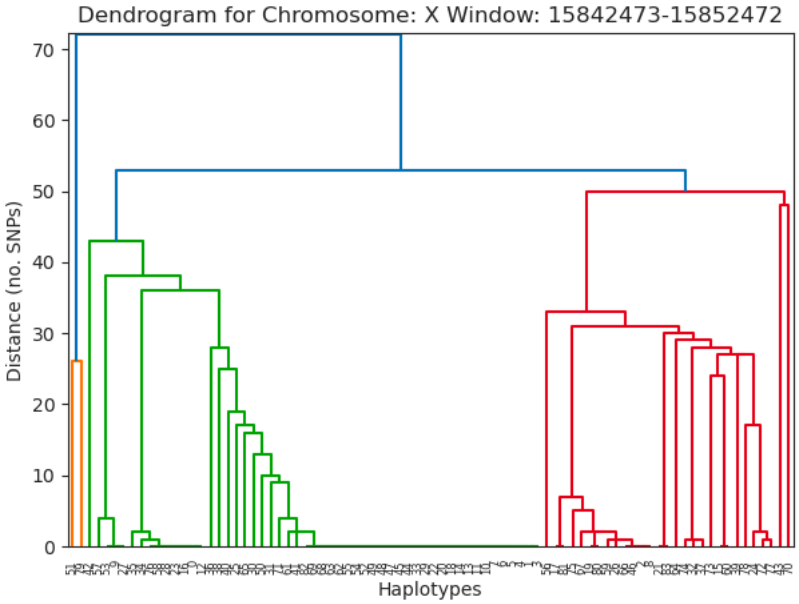
Window 15839473-15849472 haplotype cluster		
Phenotype	Score	% of mosquito haplotypes
Resistant	15	15/46 (32.6%)
Susceptible	3	3/20 (15.0%)
Control	10	10/18 (55.6%)



Supplementary Figure 11: Dendrogram for Chromosome X, Window 15839473-15849472

Supplementary Table 14: Window 15842473-15852472 had one cluster with ≥ 20 haplotypes

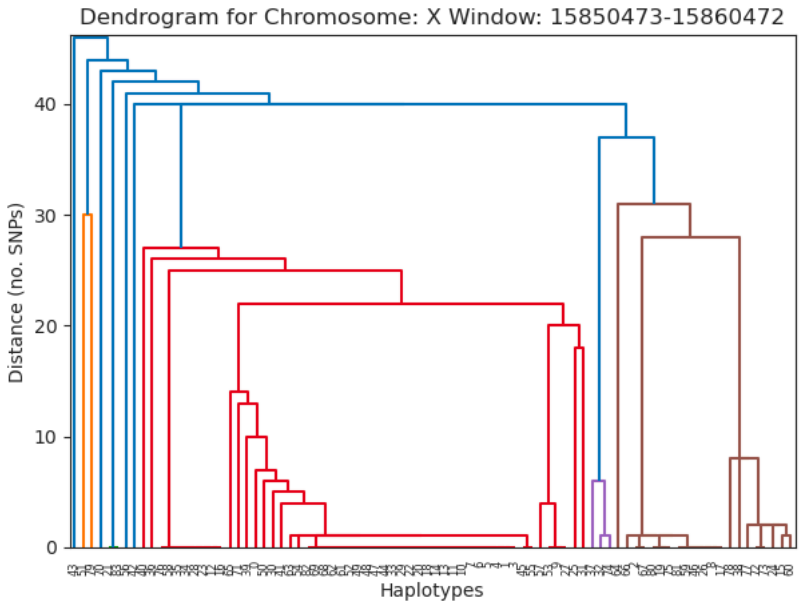
Window 15842473-15852472 haplotype cluster		
Phenotype	Score	% of mosquito haplotypes
Resistant	15	15/46 (32.6%)
Susceptible	3	3/20 (15.0%)
Control	10	10/18 (55.6%)



Supplementary Figure 12: Dendrogram for Chromosome X, Window 15842473-15852472

Supplementary Table 15: Window 15850473-15860472 had one cluster with ≥ 20 haplotypes.

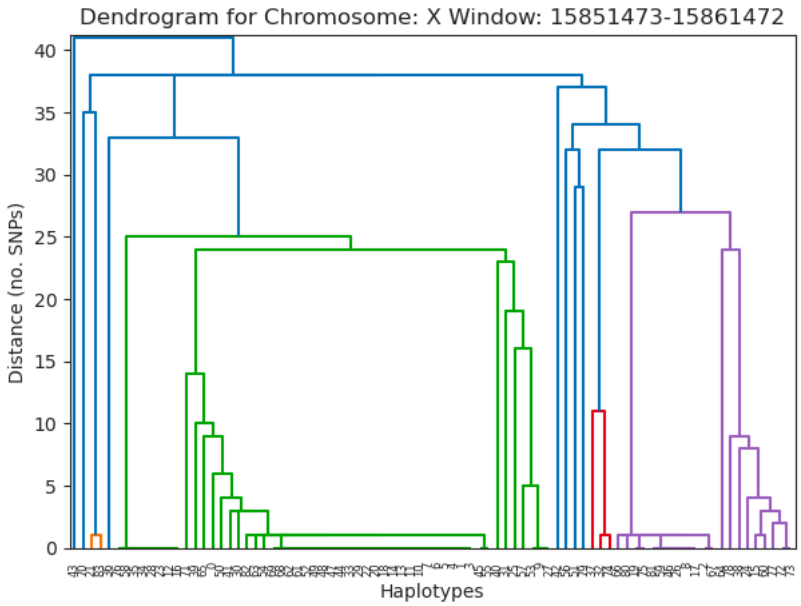
Window 15850473-15860472 haplotype cluster		
Phenotype	Score	% of mosquito haplotypes
Resistant	12	12/46 (26.1%)
Susceptible	3	3/20 (15.0%)
Control	10	10/18 (55.6%)



Supplementary Figure 13: Dendrogram for Chromosome X, Window 15850473 - 15860472

Supplementary Table 16: Window 15851473-15861472 had one cluster with ≥ 20 haplotypes.

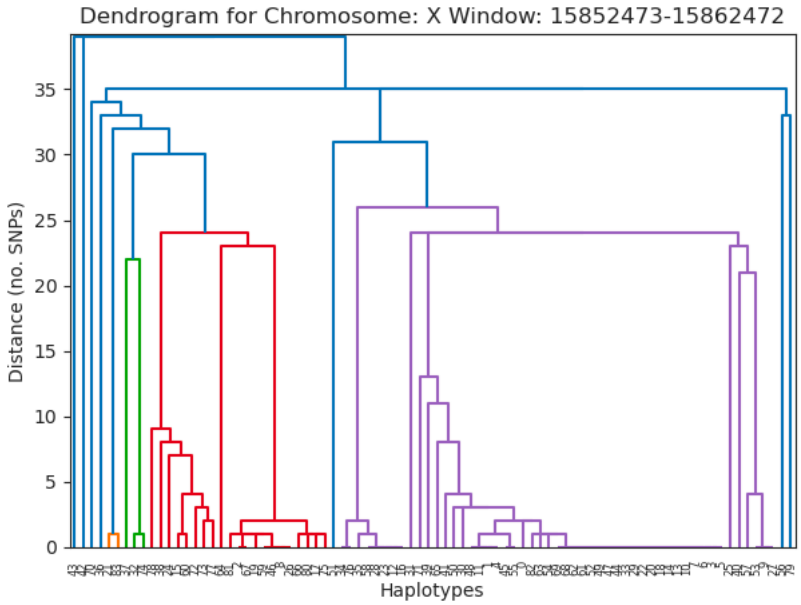
Window 15851473-15861472 haplotype cluster		
Phenotype	Score	% of mosquito haplotypes
Resistant	12	12/46 (26.1%)
Susceptible	2	2/20 (10.0%)
Control	10	10/18 (55.6%)



Supplementary Figure 14: Dendrogram for Chromosome X, Window 15851473 - 15861472

Supplementary Table 17: Window 15852473-15862472 had one cluster with ≥ 20 haplotypes.

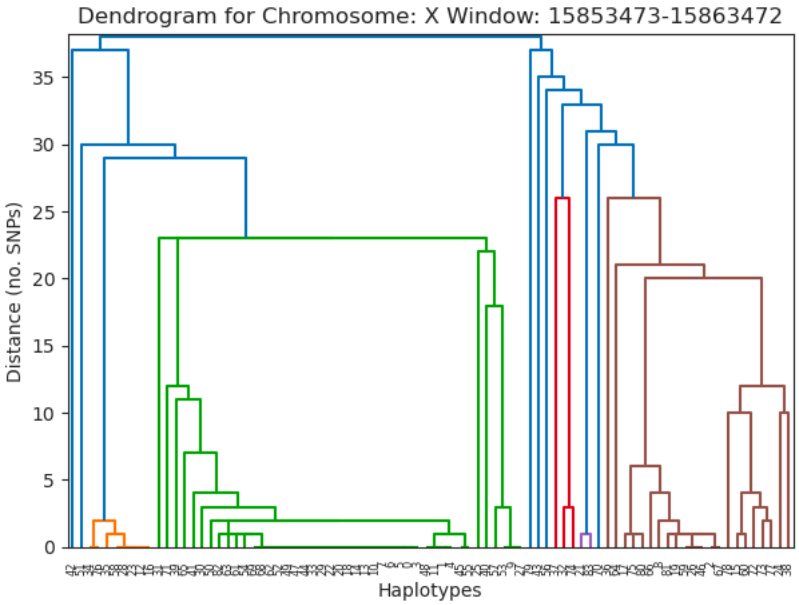
Window 15852473-15862472 haplotype cluster		
Phenotype	Score	% of mosquito haplotypes
Resistant	11	11/46 (23.9%)
Susceptible	2	2/20 (10.0%)
Control	7	7/18 (38.9%)



Supplementary Figure 15: Dendrogram for Chromosome X, Window 15852473-15862472

Supplementary Table 18: Window 15853473-15863472 had one cluster with ≥ 20 haplotypes.

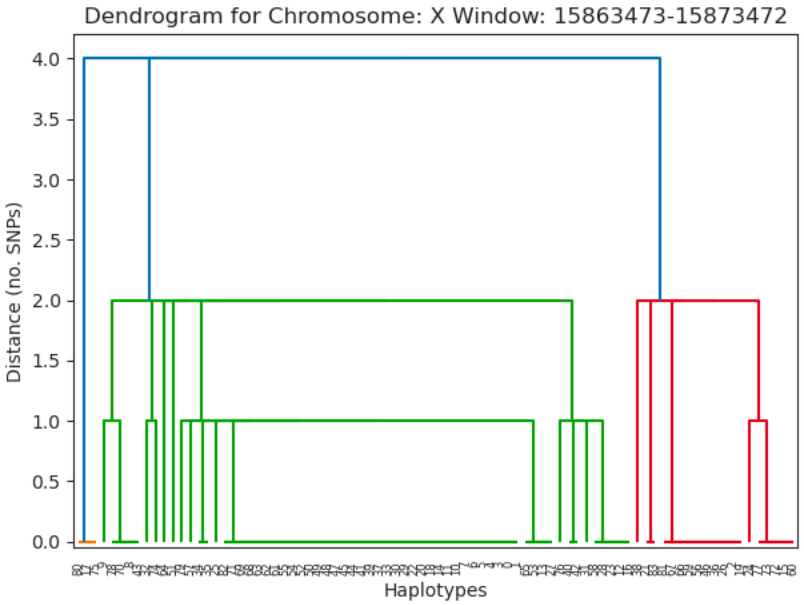
Window 15853473-15863472 haplotype cluster		
Phenotype	Score	% of mosquito haplotypes
Resistant	10	10/46 (21.7%)
Susceptible	2	2/20 (10.0%)
Control	8	8/18 (44.4%)



Supplementary Figure 16: Dendrogram for Chromosome X, Window 15853473-15863472

Supplementary Table 19: Window 15863473-15873472 had one cluster with ≥ 20 haplotypes.

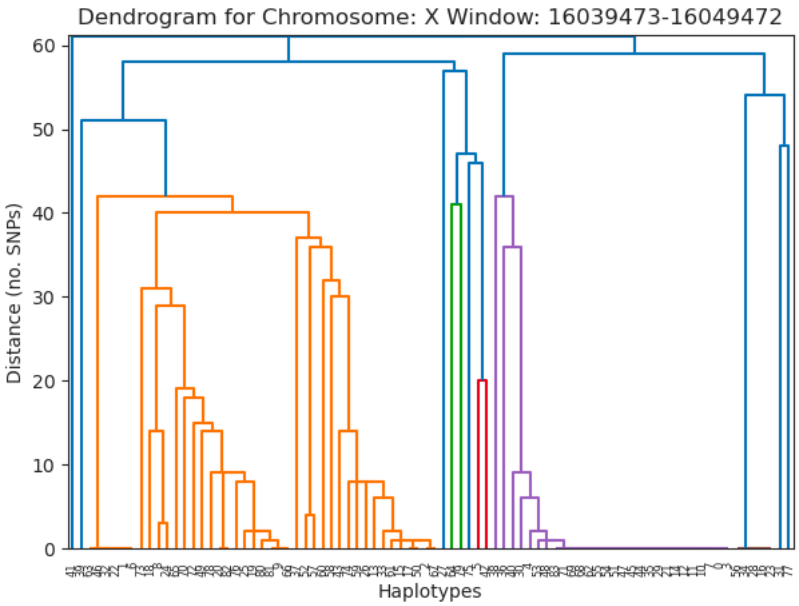
Window 15863473-15873472 haplotype cluster		
Phenotype	Score	% of mosquito haplotypes
Resistant	21	21/46 (45.7%)
Susceptible	4	4/20 (20.0%)
Control	10	10/18 (55.6%)



Supplementary Figure 17: Dendrogram for Chromosome X, Window 15863473 - 15873472

Supplementary Table 20: Window 16039473-16049472 had one cluster with ≥ 20 haplotypes.

Window 16039473-16049472 haplotype cluster		
Phenotype	Score	% of mosquito haplotypes
Resistant	10	10/46 (21.7%)
Susceptible	4	4/20 (20.0%)
Control	7	7/18 (38.9%)



Supplementary Figure 18: Dendrogram for Chromosome X, Window 16039473 - 16049472

Supplementary Data 2: Soft Clipping

We analysed soft-clipped reads to detect the presence of CNV alleles previously identified by the *Anopheles gambiae* 1000 Genomes Consortium[1]. This included the following CNV alleles:

CNV Allele[1]
Cyp6aap_Dup1
Cyp6aap_Dup2
Cyp6aap_Dup3
Cyp6aap_Dup4
Cyp6aap_Dup5
Cyp6aap_Dup6
Cyp6aap_Dup7
Cyp6aap_Dup8
Cyp6aap_Dup9
Cyp6aap_Dup10
Cyp6aap_Dup11
Cyp6aap_Dup12
Cyp6aap_Dup13
Cyp6aap_Dup14
Cyp6m_Dup1
Cyp6z_Dup1
Cyp9k1_Dup1
Cyp9k1_Dup2
Cyp9k1_Dup3
Cyp9k1_Dup4
Cyp9k1_Dup5
Cyp9k1_Dup6
Cyp9k1_Dup7
Cyp9k1_Dup8
Cyp9k1_Dup9
Cyp9k1_Dup10
Cyp9k1_Dup11
Cyp9k1_Dup12
Cyp9k1_Dup13
Cyp9k1_Dup14
Cyp9k1_Dup15
Cyp9k1_Dup16
Gstue_Dup1
Gstue_Dup3
Gstue_Dup4
Gstue_Dup5
Gstue_Dup6
Gstue_Dup7
Gstue_Dup8
Gstue_Dup9

Bijagós soft clipping: The proportion of reads at the start or end of a CNV that have been soft-clipped, for each *An. gambiae s.s.* mosquito from the Bijagós sample set. Breakpoint positions defined in Lucas et al (2019) supplementary information. Soft-clipping proportions above the verification threshold have been highlighted in green to indicate the possible presence of this CNV allele.

Sample	Cyp6aap_Dup11_Pos_Range_Start	Cyp6aap_Dup11_Pos_Range_End
c12	0.00	0.00
c14	0.00	2.50
c16	0.00	2.22
c17	0.00	0.00
c2	5.88	0.00
c4	0.00	0.00
c5	0.00	0.00
c7	0.00	0.00
c9	0.00	0.00
res10	0.00	0.00
res11	0.00	0.00
res12	0.00	0.00
res15	0.00	0.00
res16	0.00	0.00
res17	0.00	0.00
res24	0.00	2.56
res29	0.00	0.00
res35	0.00	0.00
res36	0.00	0.00
res3	0.00	1.85
res45	0.00	0.00
res46	0.00	0.00
res49	15.38	1.30
res50	0.00	2.40
res52	0.00	1.18
res54	0.00	5.88
res55	3.13	0.97
res56	0.00	1.52
res57	0.00	0.00
res6	0.00	0.00
res8	0.00	0.00
res9	0.00	0.00
sus18	0.00	0.00
sus25	0.00	0.00
sus2	10.53	1.47
sus30	33.33	3.13
sus33	0.00	0.00
sus35	0.00	0.00
sus3	0.00	0.00
sus40	0.00	2.70
sus4	0.00	0.00
sus7	0.00	1.49

Sample	Cyp6aap_Dup7_Pos_Range_Start	Cyp6aap_Dup7_Pos_Range_End
c12	0.00	0.00
c14	3.57	0.00
c16	0.00	0.00
c17	0.00	0.00
c2	0.00	0.00
c4	0.00	0.00
c5	0.00	0.00
c7	0.00	0.00
c9	0.00	0.00
res10	0.00	0.00
res11	0.00	0.00
res12	0.00	0.00
res15	0.00	0.00
res16	0.00	0.00
res17	0.00	0.00
res24	0.00	0.00
res29	0.00	0.00
res35	0.00	0.00
res36	0.00	0.00
res3	0.00	0.00
res45	0.00	0.00
res46	0.00	0.00
res49	0.00	0.00
res50	0.00	14.29
res52	0.00	0.00
res54	0.00	0.00
res55	0.00	0.00
res56	0.00	0.00
res57	0.00	0.00
res6	0.00	0.00
res8	0.00	0.00
res9	0.00	20.00
sus18	0.00	0.00
sus25	0.00	0.00
sus2	0.00	0.00
sus30	0.00	0.00
sus33	0.00	0.00
sus35	0.00	0.00
sus3	0.00	0.00
sus40	0.00	0.00
sus4	0.00	0.00
sus7	0.00	0.00

Sample	Cyp9k1_Dup12_Pos_Range_Start	Cyp9k1_Dup12_Pos_Range_End
c12	0.00	0.00
c14	0.00	0.00
c16	0.00	0.00
c17	0.00	0.00
c2	0.00	0.00
c4	0.00	0.00
c5	0.00	0.00
c7	0.00	0.00
c9	0.00	0.00
res10	0.00	0.00
res11	0.00	0.00
res12	0.00	0.00
res15	0.00	0.00
res16	0.00	0.00
res17	0.00	0.00
res24	0.00	0.00
res29	0.00	0.00
res35	0.00	0.00
res36	0.00	0.00
res3	0.00	0.00
res45	0.00	6.67
res46	0.00	0.00
res49	0.00	0.00
res50	0.00	0.00
res52	0.00	0.00
res54	0.00	0.00
res55	0.00	5.88
res56	0.00	0.00
res57	0.00	0.00
res6	0.00	0.00
res8	0.00	0.00
res9	0.00	0.00
sus18	0.00	0.00
sus25	0.00	0.00
sus2	0.00	0.00
sus30	0.00	0.00
sus33	0.00	0.00
sus35	0.00	0.00
sus3	0.00	22.22
sus40	0.00	0.00
sus4	0.00	0.00
sus7	0.00	0.00

Sample	Gstue_Dup3_Pos_Range_Start	Gstue_Dup3_Pos_Range_End
c12	0.00	0.00
c14	0.00	0.00
c16	0.00	0.00
c17	0.00	0.00
c2	9.09	0.00
c4	0.00	0.00
c5	0.00	0.00
c7	0.00	0.00
c9	0.00	0.00
res10	0.00	0.00
res11	0.00	0.00
res12	0.00	0.00
res15	0.00	0.00
res16	0.00	0.00
res17	0.00	0.00
res24	0.00	0.00
res29	0.00	0.00
res35	0.00	0.00
res36	0.00	0.00
res3	0.00	0.00
res45	0.00	0.00
res46	0.00	0.00
res49	0.00	0.00
res50	0.00	0.00
res52	0.00	0.00
res54	0.00	0.00
res55	0.00	0.00
res56	0.00	0.00
res57	0.00	0.00
res6	0.00	0.00
res8	0.00	0.00
res9	0.00	0.00
sus18	0.00	0.00
sus25	0.00	0.00
sus2	0.00	0.00
sus30	0.00	25.00
sus33	0.00	0.00
sus35	0.00	0.00
sus3	0.00	0.00
sus40	0.00	0.00
sus4	0.00	0.00
sus7	0.00	0.00

Supplementary Data 3: Odds Ratios

Allele based Odds Ratio calculations

Table 1 (from main text):

Chromosome	Gene	Position	SNP	Resistant mosquitoes				Susceptible mosquitoes				Control mosquitoes			
				Homozygous reference	Heterozygous	Homozygous alternate	No call	Homozygous reference	Heterozygous	Homozygous alternate	No call	Homozygous reference	Heterozygous	Homozygous alternate	No call
2L	*vgsc*	2416980	T791M	18	0	0	5	8	1	0	1	8	1	0	0
2L	*vgsc*	2422652	L995F	15	7	0	0	7	3	0	0	6	3	0	0
2L	*vgsc*	2429745	N1570Y	18	5	0	0	10	0	0	0	7	1	0	1
2L	*vgsc*	2430424	A1746S	23	0	0	0	9	1	0	0	7	1	0	1
2L	*vgsc*	2430881	P1874L	21	2	0	0	8	2	0	0	7	1	0	1
3R	*gste2*	28598062	L119V	19	3	0	1	9	0	0	1	9	0	0	0

Odds Ratio calculations:

T791M

	Resistant phenotype	Susceptible phenotype
Carries allele	0	1
Does not carry allele	18	8

Using the Haldane-Anscombe correction.
OR = (0.5*8.5)/(1.5*18.5) = 0.153
ln(OR) = -1.876
SE(ln(OR)) = sqrt(1/0.5 + 1/1.5 + 1/18.5 + 1/8.5) = 1.685
ln(95% Confidence Interval) = -1.876 ± (1.96 x 1.685) = -1.565 ± 3.303 = -4.868 to 1.738
95% CI = e^(-4.868) to e^(1.738)
95% CI = 0.007 to 5.686
No statistical significance.

L995F

	Resistant phenotype	Susceptible phenotype
Carries allele	7	3
Does not carry allele	15	7

Odds Ratio = 1.086
95% Confidence Interval (0.172, 8.4)
No statistical significance.

N1570Y

	Resistant phenotype	Susceptible phenotype
Carries allele	5	0
Does not carry allele	18	10

Using the Haldane-Anscombe correction.
OR = (5.5*10.5)/(0.5*18.5) = 6.243
ln(OR) = 1.831
SE = sqrt(1/5.5 + 1/0.5 + 1/18.5 + 1/10.5) = 1.527
ln(95% CI) = 1.831 ± (1.96*1.527)
ln(95% CI) = 1.831 ± 2.99 = -1.159 to 4.821
95% CI = (0.314, 124.09)
No statistical significance

A1746S

	Resistant phenotype	Susceptible phenotype
Carries allele	0	1
Does not carry allele	23	9

Using the Haldane-Anscombe correction.
OR = (0.5*9.5)/(1.5*23.5) = 0.135
ln(OR) = -2.002
SE = sqrt(1/0.5 + 1/1 + 1/23 + 1/9) = 1.776
ln(95% CI) = -2.002 ± (1.96*1.776)
ln(95% CI) = -2.002 ± 3.481 = -5.483 to 1.479
95% CI = (0.004, 4.389)
No statistical significance

P1874L

	Resistant phenotype	Susceptible phenotype
Carries allele	2	2
Does not carry allele	21	8

Odds Ratio = 0.394
95% Confidence Interval (0.025, 6.287)
No statistical significance.

L119V

	Resistant phenotype	Susceptible phenotype
Carries allele	3	0
Does not carry allele	19	9

Using the Haldane-Anscombe correction.
OR = (3.5*9.5)/(0.5*19.5) = 3.41
ln(OR) = 1.227
SE = sqrt(1/3.5 + 1/0.5 + 1/19.5 + 1/9.5) = 1.56
ln(95% CI) = 1.227± (1.96*1.56)
ln(95% CI) = 1.227 ± 3.06 = (-1.833,4.287)
95% CI = (0.160, 72.75)
No statistical significance

Table 3 (from main text, includes only SNPs with an allele frequency over 5% in the sample population):

Gene	Position	Amino acid Change	Resistant mosquitoes				Susceptible mosquitoes				Control mosquitoes			
			Homozygous reference	Heterozygous	Homozygous Alternate	No call	Homozygous reference	Heterozygous	Homozygous Alternate	No call	Homozygous reference	Heterozygous	Homozygous Alternate	No call
gste2	28597956	T154S	5	13	4	1	2	5	1	2	3	5	1	0
gste2	28598041	I126F	19	3	0	1	8	1	0	1	8	1	0	0
gste2	28598505	G26S	20	2	0	1	8	1	0	1	9	0	0	0
ace1	3489405	A65S	4	14	5	0	1	7	2	0	4	4	1	0

Odds ratio calculations:

***gste2* T154S**

	Resistant phenotype	Susceptible phenotype
Carries allele	17	6
Does not carry allele	5	2

Odds Ratio = 1.129
95% Confidence Interval (0.0857, 9.606)
No statistical significance.

***gste2* I126F**

	Resistant phenotype	Susceptible phenotype
Carries allele	3	1
Does not carry allele	19	8

Odds Ratio = 1.254
95% Confidence Interval (0.0846,74.492)
No statistical significance.

***gste2* G26S**

	Resistant phenotype	Susceptible phenotype
Carries allele	2	1
Does not carry allele	20	8

Odds Ratio = 0.081
95% Confidence Interval (0.0369,53.029)
No statistical significance.

***ace1* A65S**

	Resistant phenotype	Susceptible phenotype
Carries allele	19	9
Does not carry allele	4	1

Odds Ratio = 0.537
95% Confidence Interval (0.009,6.556)
No statistical significance.

CHAPTER SIX

Published paper: Whole genome sequence analysis of population structure and insecticide resistance markers in *Anopheles melas* from the Bijagós Archipelago, Guinea-Bissau

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
T: +44 (0)20 7299 4646
F: +44 (0)20 7299 4656
www.lshtm.ac.uk

RESEARCH PAPER COVER SHEET

Please note that a cover sheet must be completed for each research paper included within a thesis.

SECTION A – Student Details

Student ID Number	lsh2005123	**Title**	Miss
First Name(s)	Sophie		
Surname/Family Name	Moss		
Thesis Title	Genomic epidemiology of *Plasmodium falciparum* and its *Anopheline* vectors in the context of malaria control on the Bijagós Archipelago of Guinea-Bissau		
Primary Supervisor	Dr Anna Last		

If the Research Paper has previously been published please complete Section B, if not please move to Section C.

SECTION B – Paper already published

Where was the work published?	N/A		
When was the work published?	N/A		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	N/A		
Have you retained the copyright for the work?*	**N/A**	Was the work subject to academic peer review?	**N/A**

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	Parasites & Vectors

Please list the paper's authors in the intended authorship order:	Sophie Moss, Elizabeth Pretorius, Sainey Ceesay, Eunice Teixeira Da Silva, Harry Hutchins, Mamadou Ousmane Ndiath, Holly Acford-Palmer, Emma Collins, Matthew Higgins, Jody Phelan, Robert T. Jones, Hristina Vasileva, Amabelia Rodrigues, Sanjeev Krishna, Taane G. Clark, Anna Last, Susana Campino
Stage of publication	**Submitted**

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I devised the research question and designed the study. I prepared all samples for sequencing. I conducted all bioinformatic analyses of sequence data. I wrote the manuscript, which was commented on by my co-authors.

SECTION E

Student Signature	[redacted]
Date	22nd May 2024

Supervisor Signature	Anna Last Digitally signed by Anna Last [redacted] Date: 2024.05.22 18:20:51 +01'00'
Date	22nd May 2024

Moss *et al. Parasites & Vectors* (2024) 17:396
https://doi.org/10.1186/s13071-024-06476-2

Parasites & Vectors

RESEARCH Open Access

Whole genome sequence analysis of population structure and insecticide resistance markers in *Anopheles melas* from the Bijagós Archipelago, Guinea-Bissau

Sophie Moss[1*], Elizabeth Pretorius[1], Sainey Ceesay[2], Eunice Teixeira da Silva[3,4], Harry Hutchins[1], Mamadou Ousmane Ndiath[2], Holly Acford-Palmer[1], Emma L. Collins[1], Matthew Higgins[1], Jody Phelan[1], Robert T. Jones[1], Hristina Vasileva[1,2], Amabelia Rodrigues[3,4], Sanjeev Krishna[5,6,7], Taane G. Clark[1,8†], Anna Last[1†] and Susana Campino[1†]

Abstract

Background *Anopheles melas* is an understudied malaria vector with a potential role in malaria transmission on the Bijagós Archipelago of Guinea-Bissau. This study presents the first whole-genome sequencing and population genetic analysis for this species from the Bijagós. To our knowledge, this also represents the largest population genetic analysis using WGS data from non-pooled *An. melas* mosquitoes.

Methods WGS was conducted for 30 individual *An. melas* collected during the peak malaria transmission season in 2019 from six different islands on the Bijagós Archipelago. Bioinformatics tools were used to investigate the population structure and prevalence of insecticide resistance markers in this mosquito population.

Results Insecticide resistance mutations associated with pyrethroid resistance in *Anopheles gambiae* s.s. from the Bijagós were absent in the *An. melas* population, and no signatures of selective sweeps were identified in insecticide resistance-associated genes. Analysis of structural variants identified a large duplication encompassing the cytochrome-P450 gene *cyp9k1*. Phylogenetic analysis using publicly available mitochondrial genomes indicated that *An. melas* from the Bijagós split into two phylogenetic groups because of differentiation on the mitochondrial genome attributed to the cytochrome C oxidase subunits COX I and COX II and the NADH dehydrogenase subunits 1, 4, 4L and 5.

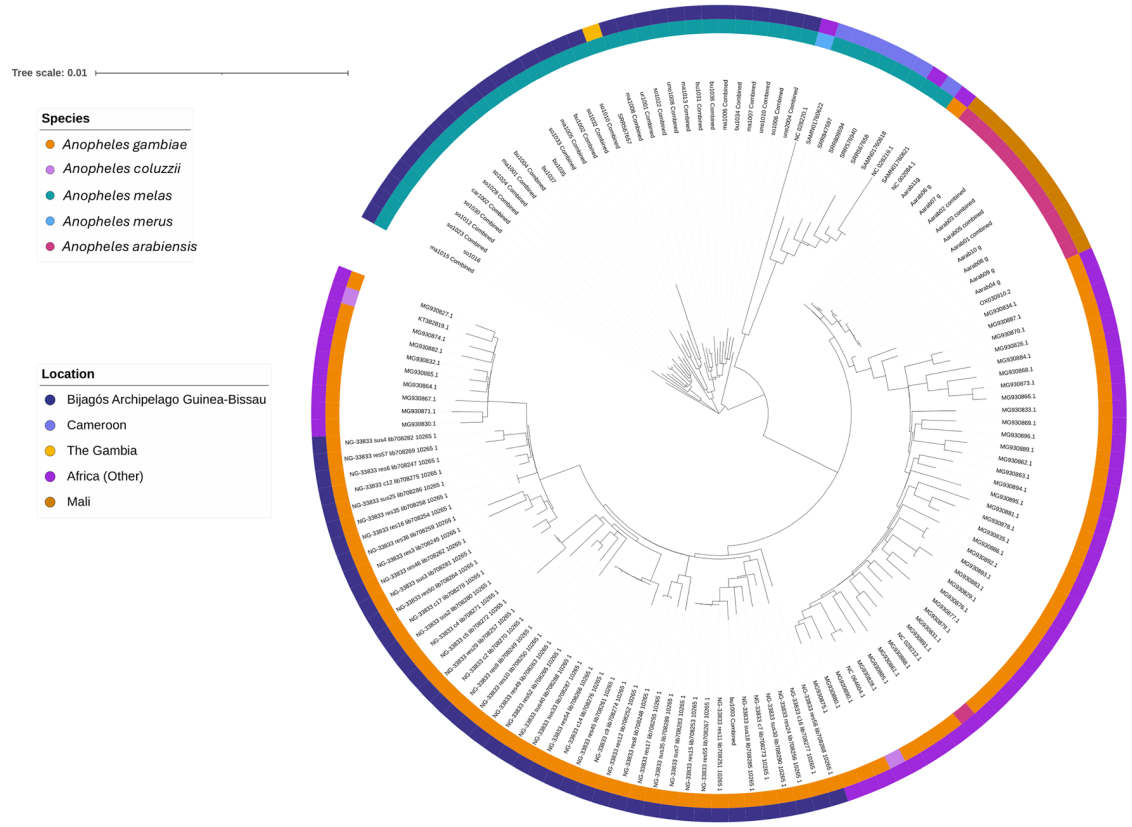
Conclusions This study identified an absence of insecticide-resistant SNPs common to *An. gambiae* in the *An. melas* population, but did identify structural variation over insecticide resistance-associated genes. Furthermore, this study presents novel insights into the population structure of this malaria vector using WGS analysis. Additional studies are required to further understand the role of this vector in malaria transmission.

Keywords Anopheles melas, Whole-genome sequencing, Guinea-Bissau, Population structure, Insecticide resistance, Malaria

†Taane G. Clark, Anna Last and Susana Campino are joint last authors.

*Correspondence:
Sophie Moss
sophie.moss@lshtm.ac.uk
Full list of author information is available at the end of the article

Graphical abstract



Background

Malaria is a persistent public health problem in Guinea-Bissau, West Africa, where the population of 2.1 million people experienced an estimated 225,200 malaria cases and 1023 malaria deaths in 2022 [1]. The Bijagós Archipelago (Bijagós) is a group of 88 islands and islets located approximately 50 km off the coast of Guinea-Bissau and is a designated UNESCO Biosphere Reserve [2]. The Archipelago is home to approximately 25,000 people, who live on 19 permanently inhabited islands [3]. Malaria transmission on the Bijagós is highly seasonal and stable [4], and *Plasmodium falciparum* prevalence on the Archipelago can peak at up to 15% at the end of the rainy season in November [5].

A survey in 2017 on Bubaque Island, the most populous island of the Archipelago, identified all *Anopheles* species on Bubaque to be members of the *Anopheles gambiae* sensu lato (s.l.) complex. Of the species present, *Anopheles gambiae* sensu stricto (s.s.) was identified as the primary malaria vector and *Anopheles melas* as having a role in low level transmission during the dry season [4]. *Anopheles melas* is a saltwater-tolerant species able to sustain population numbers during the dry season by laying eggs in brackish water, giving this species a particular advantage during the dry season when freshwater oviposition sites have dried up [6]. The Bijagós Archipelago has an abundance of mangroves and mud flats [2], which are commonly associated with the presence of *An. melas* [7, 8]. A larger survey conducted on 16 of the inhabited islands of the Archipelago was conducted between October and December 2019 during the peak malaria transmission season. This survey used indoor and outdoor light-traps and identified 85.2% of trapped mosquitoes to be *An. melas* (Pretorius et al. 2024, in review). A sub-sample of mosquitoes were investigated for *Plasmodium falciparum* sporozoite positivity using circumsporozoite protein (CSP) ELISA, which calculated a sporozoite rate of 0.86% and identified all CSP-positive specimens collected as *An. melas* (Pretorius et al. 2024, in review). This study indicates that *An. melas* may be important in malaria transmission on the Bijagós, particularly regarding residual transmission during the dry

season (Pretorius et al. 2024, in review). This is supported by previous studies which have identified *An. melas* to have a role in malaria transmission, including in Senegal [9], The Gambia [10, 11] and Equatorial Guinea [12].

Vector control on the Bijagós relies on the use of insecticide-treated nets (ITNs) impregnated with pyrethroid insecticides, which are distributed every 3 years and have high estimated coverage and usage of around 90% [13, 14]. Pyrethroid ITNs are the most successful vector control intervention developed to date, having prevented approximately 68% of malaria deaths in Africa between 2000 and 2015 [15]. Alarmingly, resistance to pyrethroids is highly prevalent worldwide [16]. Of all countries that reported resistance data to WHO between 2010 and 2020, 87% declared pyrethroid resistance in at least one malaria vector [1]. Resistance to pyrethroids has been associated with single-nucleotide polymorphisms (SNPs) in the voltage-gated sodium channel gene (*vgsc*) of *An. gambiae*, including L995F and L995S, also known as the *kdr* west and *kdr* east alleles [17, 18], and the N1570Y mutation [19]. In addition, target site mutations in the *gste2* gene have been associated with pyrethroid resistance, including L119V and I114T [20]. Resistance to pyrethroids has been associated with copy number variants (CNVs) encompassing genes in three major enzyme families: cytochrome-P450s, esterases and glutathione-S-transferases (GSTs) [21–24].

Insecticide resistance is a growing threat to the control of malaria. This threat has propelled the understanding of *Anopheles* genetic variation through international collaborations such as the Anopheles Gambiae 1000 Genomes Project [25]. Genomics research has focused on the key malaria vectors *An. gambiae* s.s. and *Anopheles coluzzii*, but few studies have investigated *An. melas*. On the Bijagós, a previous study on Bubaque Island investigated the presence of the *kdr* east and west alleles in *An. gambiae* s.l. mosquitoes using targeted PCR sequencing [4], and a subsequent study across 13 islands investigated the prevalence of known insecticide resistance mutations using high-throughput multiplex-amplicon sequencing [26]. This study identified four mutations associated with insecticide resistance in *An. melas* at low prevalence. This included three mutations in the *vgsc* gene, L995F, N1570Y and A1746S, one mutation in the *rdl* gene, A296G, and no known insecticide resistance mutations in the *ace1* or *gste2* genes [26]. However, no previous studies using whole-genome sequence (WGS) data from *An. melas* on the Bijagós Archipelago have been conducted, and population structure and signatures of selection have not previously been investigated. Furthermore, to our knowledge, only two studies analysing WGS data from *An. melas* have previously been published. This includes a study of Pool-seq WGS data from *An. melas* in The Gambia, Cameroon and Equatorial Guinea, where the DNA from several individual mosquitoes was pooled prior to sequencing [27] and the Anopheles 16 Genomes Project, which produced the reference genome assembly for *An. melas* [28]. The previous study using Pool-seq WGS data identified three genetically distinct clusters of *An. melas* on the West African coast, *An. melas* West, *An. melas* South and *An. melas* Bioko, which ranged respectively from The Gambia to Northwest Cameroon, Southeast Cameroon to Angola and Bioko Island in Equatorial Guinea [27, 29]. Genetic differentiation between these clusters was high and mostly distributed evenly across the genome, with elevated differentiation along the X chromosome, indicative of allopatric divergence [27].

Here, we generate and analyse WGS data from 30 individual mosquitoes of this little-known species, collected during the 2019 vector survey on the Bijagós Archipelago, combined with WGS data from seven *An. melas* specimens made available through the *Anopheles* 16 Genomes Project [28]. WGS data are a valuable resource which we utilise to investigate the genetic diversity of this mosquito population, the presence of SNPs associated with insecticide resistance and genomic signatures of selection in this species.

Methods

Mosquito sampling

Mosquitoes were collected from the Bijagós Archipelago during October and November 2019 using CDC indoor and outdoor light traps (Model 512; John W. Hock Co., Gainesville, FL, USA) using previously described methodology [30]. This included mosquitoes from six different islands across the Archipelago highlighted with purple triangles in Fig. 1: Soga, Bubaque, Tchedega (Maio), Uno, Caravela and Uracane. Collected mosquitoes were separated by genus, and *Anopheles* mosquitoes were morphologically identified using previously described keys [31]. All mosquitoes were identified as belonging to the *An. gambiae* s.l. complex.

DNA extraction

DNA extraction was conducted following the manufacturer's instructions, using the QIAamp® 96 DNA QIAcube® HT Kit (Qiagen) with the QIAcube Extraction Robot. DNA was eluted in 80 µl AE buffer and stored at – 20 ℃. No additional processing of DNA, for example whole-genome amplification or selective genome amplification, was conducted prior to sequencing.

Species identification

Female mosquitoes were then identified to species at the Medical Research Council Unit The Gambia at London School of Hygiene & Tropical Medicine using

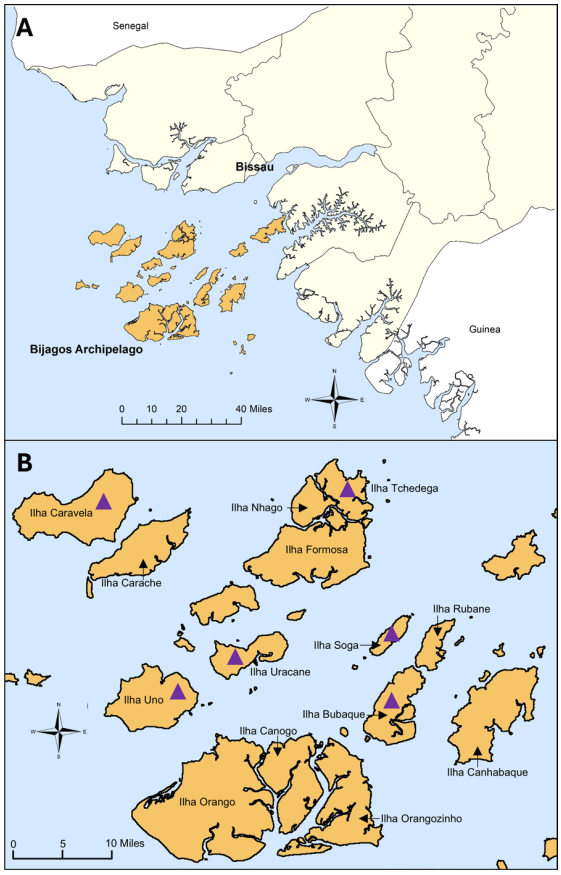


Fig. 1 Mosquito sample collection sites. **A** Location of Bijagós Archipelago, created using ArcGIS ArcMap 10.8.1. **B** Mosquitoes were collected from the six islands labelled with purple triangles. Map from OpenStreetMap 2023-05-06

PCR-RFLP to distinguish between members of the *An. gambiae* s.l. complex, based on the protocol of Fanello et al. [32]. This used primers to amplify the intergenic spacer (IGS) region, which differs in size between members of the *An. gambiae* complex. The following primers were used: Universal F: GTGTGCCCCTTCCTC GATGT; *Anopheles gambiae* R: CTGGTTTGGTCG GCACGTTT; *An. arabiensis* R: AAGTGTCCTTCT CCATCCTA; *An. melas* R: TGACCAACCCACTCC CTTGA. Amplified PCR products were then digested using the HhaI enzyme to differentiate *An. gambiae* s.s. and *An. coluzzii* specimens. The band sizes of the PCR products were visualised using electrophoresis with the QIAxcel capillary electrophoresis system (Qiagen). The band sizes of PCR product were analysed to distinguish species: *An. gambiae* s.s. (257 and 110 bp), *An. arabiensis* (292 bp), *An. melas* (435 bp), *An. coluzzii* (367 bp) and *An. coluzzii/An. gambiae* s.s. hybrid (257 bp, 110 bp and 367 bp).

Whole-genome sequencing and bioinformatic analysis

The DNA of 30 *An. melas* mosquitoes from six different islands across the archipelago was whole-genome sequenced at Eurofins Genomics using the Illumina Novaseq 6000 (2×150 bp configuration). This included 11 *An. melas* from Soga, 7 from Bubaque, 7 from Tchedega (Maio), 3 from Uno, 1 from Caravela and 1 from Uracane. Average read depth for all samples across the whole genome was 32.61. Over 55% of the genome for each sample had a read depth ≥ 20. Publicly available *An. melas* data were included for population analyses. This included one sample from The Gambia and six samples from Cameroon, made available through the Anopheles 16 Genomes Project [28, 33]. Mosquitoes were phenotypically identified as female during sampling, and this was called again using the modal coverage ratio between chromosome 3R and chromosome X (with a ratio between 0.4 and 0.6 = male and ratio between 0.8 and 1.2 = female, with other ratios leading to sample exclusion, as previously described [34]).

Raw WGS data were trimmed using *trimmomatic* (version 0.39) [35] before aligning to the *Anopheles gambiae* (AgamP4) reference genome and the *Anopheles melas* (AmelC2) reference genome, using *bwa-mem* software (default parameters)[36]. The AgamP4 alignment was taken forward as the AgamP4 reference genome is of better quality than the *An. melas* reference genome, and the percentage of reads which mapped to AgamP4 was higher (average 92.8%) than the percentage of reads which mapped to AmelC2 (average 79.3%). Furthermore, the AmelC2 reference genome is not a chromosomal level assembly, consisting of over 22,000 scaffolds with a mean N50 value of 18,103, compared to a AgamP4, which is a chromosomal level assembly with mean N50 value of 49,364,325. Therefore, mapping of *An. melas* WGS data to the AgamP4 reference genome also allowed interpretation of genetic variation in the context of chromosome location. Finally, mapping of *An. melas* data to the *An. gambiae* reference genome has been conducted in previously published studies with amplicon [26] and WGS data [27] and is an accepted method. Mapping and coverage statistics from the resulting bam files were calculated using *samtools* [37]. Variants were called for each sample using *GATK's* HaplotypeCaller (v 4.1.4.1) to generate a VCF for each sample [38]. A combined, genotyped VCF was created for the *An. melas* samples from the Bijagós, The Gambia and Cameroon using GATK's GenomicsDBImport and GenotypeGVCFs function [38]. The multi-sample VCF was filtered using *bcftools* (v 1.17) and *GATK's* VariantFiltration to include chromosomal variants with the following parameters: QD > 5.0, QUAL > 30.0, SOR < 3.0, FS < 60.0, MQ > 40.0, MQRankSum > − 12.5, ReadPosRankSum > -8.0. Reads

were subsequently filtered to remove reads with DP < 5.0 or GQ < 20.0, and variants were filtered to remove those with > 20.0% missing genotypes or MAF < 0.01. The final filtered VCF contained 6,767,012 variants.

Population genetic analysis

A distance matrix was generated using PLINK (v 1.90b6.21), and principal component analyses were computed in R (v 4.3.1). A maximum likelihood tree was made using RAxML-NG (v 1.2.0) with the mitochondrial FASTA sequences for the *An. melas* samples and 43 *An. gambiae* samples from the Bijagós Archipelago [26]. The resulting maximum likelihood tree was visualised using iTOL [39]. Admixture analysis was computed using ADMIXTURE (v 1.3.0) [40]. The estimated number of ancestral populations (optimum K-value) was computed through cross-validation of 1–10 dimensions of eigenvalue decay (k = 4).

Nucleotide diversity (π) was computed for the $N = 30$ Bijagós *An. melas* mosquitoes in 20,000-bp windows on chromosome 3 L using phased filtered variants and the scikit-allel function *allel.windowed_diversity*. Tajima's D was calculated using the function *allel.windowed_tajima_d*. Analysis was conducted with chromosome 3 L only because of the presence of large chromosomal inversions in the other chromosomes in the *An. gambiae* s.l. complex [41].

Genetic divergence between *An. melas* from the Bijagós and Cameroon was investigated using the fixation index, F_{ST}. The *windowed_weir_cockerham_fst* function in scikit-allel was used to compute F_{ST} in 1-kbp windows over each chromosome (https://scikit-allel.readthedocs.io/en/stable/). Signatures of selection were investigated using three different complementary statistics: H12 [42], iHS and XP-EHH [43, 44]. Garud's H_{12} was computed using the *moving_garud_h* function in scikit-allel, using phased biallelic SNPs in windows of 1000 SNPs. Two hundred iterations of H_{12} were calculated, and the mean value for each window was plotted. iHS was computed using phased biallelic SNPs using the *allel.ihs* function in scikit-allel (https://scikit-allel.readthedocs.io/en/stable/). Raw iHS scores were standardized using the *allel.standardize_by_allele_count* function, and *p*-values were computed and plotted. XP-EHH was calculated using phased biallelic SNPs using the *allel.xpehh* function in scikit-allel (https://scikit-allel.readthedocs.io/en/stable/). XP-EHH scores were standardised using the *allel.standardize_by_allele_count* function and plotted.

Identification of SNPs associated with insecticide resistance

The filtered VCF was analysed to identify SNPs in four genes previously associated with resistance: *vgsc, rdl, gste2* and *ace1*. Identified variants were annotated using SnpEff (v 5.1d). The *vgsc* G2042C non-synonymous (NS) SNP was identified in 100% of allele calls in *An. melas* from all locations. This mutation has not been reported in this study as a novel NS SNP with possible association with resistance, as the identified presence in 100% of *An. melas* may have resulted from alignment to the AgamP4 genome, reflecting a species-specific mutation between our *An. melas* samples and the reference *An. gambiae*.

Identification of structural variants

DELLY software was used to identify large structural variants (SVs) [45]. Individual bcf files were created for each sample from their bam files using DELLY [45], which were then merged and filtered to remove samples with average genome read depth of < 20× and SVs with > 20% missingness. Filtered SVs were retained for analysis.

Results

Whole-genome sequence data and genetic diversity

WGS data from 30 *An. melas* from the Bijagós Archipelago were combined with an additional seven *An. melas* mosquitoes with publicly available WGS data, which were downloaded and incorporated for analysis. This included *An. melas* from Cameroon ($N = 6$) and The Gambia ($N = 1$). The combined dataset ($N = 37$) contained 6,767,012 high-quality SNPs after filtering. Average sequencing depth across the core genome ranged from 23.1 to 55.61-fold coverage, with an average of 32.61-fold coverage across all samples. The mitochondrial genome was sequenced to very high depth, averaging 2600-fold coverage and ranging from a minimum of 420- up to 5930-fold coverage.

Nucleotide diversity (π) and Tajima's D were computed in 20,000-bp windows across chromosome 3L. This resulted in mean π = 0.003 (SD = 0.001), indicating that for any pair of mosquitoes, 0.3% of nucleotides would differ, implying low nucleotide diversity in the population. Mean Tajima's D for chromosome 3L was calculated as D = − 1.531 (SD = 0.710), indicating an excess of rare alleles. Tajima's D was much higher at the start of the chromosome, with mean D = 1.518 (SD = 1.334) for the first 50 windows (1Mbp) of the chromosome, suggesting balancing selection in this region (Supplementary Data 1).

Population genetic and ancestry analysis

A maximum likelihood tree was generated using mitochondrial sequences from the Bijagós, combined with publicly available mitochondrial sequences from additional *Anopheles* mosquitoes from across Africa in the *Anopheles gambiae* s.l. complex: *An. gambiae* s.s., *An. melas, An. merus, An. arabiensis* and *An. coluzzii* (Fig. 2).

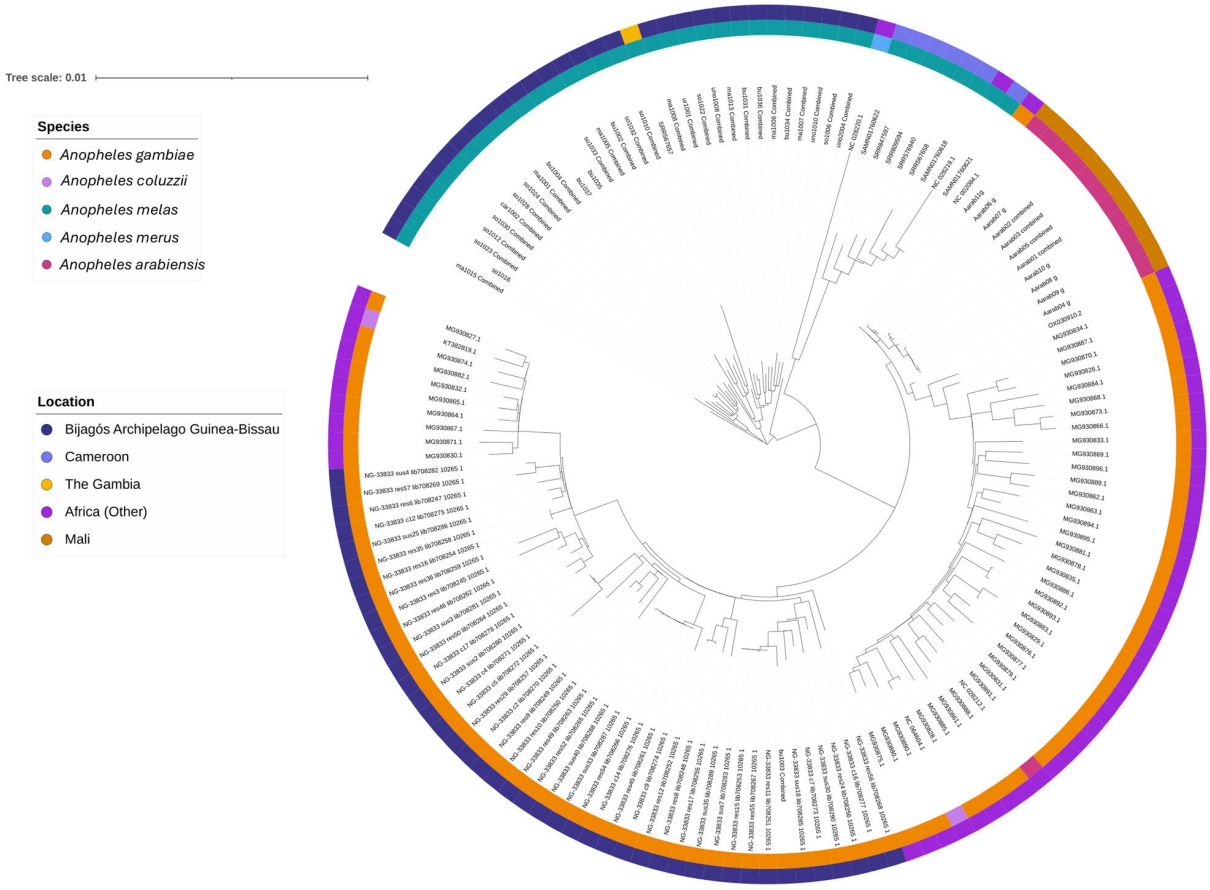


Fig. 2 Maximum likelihood tree of whole mitochondrial sequences: *Anopheles melas* and *An. gambiae* from the Bijagós Archipelago with other species from the *An. gambiae* sensu lato species complex

As expected, *An. melas* from the Bijagós group with the other *An. melas* specimens from The Gambia and Cameroon.

A second maximum likelihood tree was computed using whole mitochondrial sequences of *An. melas* ($N=37$) and previously published *An. gambiae* samples from the Bijagós ($N=43$) ([26], in review) (Fig. 3). The *An. melas* samples from the Bijagós group with the *An. melas* individual from The Gambia, and are situated next to the *An. melas* samples from Cameroon. No clustering of *An. melas* from different islands on the Bijagós was identified in the maximum likelihood tree. However, the Bijagós *An. melas* appear to form two distinct clusters, labelled A and B (Fig. 3). These clusters do not correspond to different islands of the Bijagós Archipelago, with mosquito specimens from several islands appearing in each cluster.

Principal component analysis (PCA) was conducted per chromosome to further investigate the relationship between the Bijagós *An. melas* mosquitoes (Fig. 4). For all chromosomes, PCA indicated that *An. melas* from the Bijagós is genetically distinct from *An. melas* samples from Cameroon and The Gambia. Comparisons with the Gambia should be treated with caution as this analysis includes only one sample. This geographic separation depicted by PC1 explains a large amount of variation for all chromosomes, ranging from 28.7% in chromosome 2 L to 59.8% in the mitochondrial genome. The clustering of *An. melas* from the Bijagós into two phylogenetic groups (A and B in Fig. 3) is supported by the PCA analysis for the mitochondrial genome. However, PCA analysis of the chromosomes 2 L, 2R and 3 L shows *An. melas* from the Bijagós clustering into one group, and chromosome X and 3R indicate some divergence but not clearly between groups A and B. Furthermore, little variance was explained by PC2 for the X and 3R chromosomes compared to the mitochondrial genome, 1.8% for chromosome X and 5.5% for chromosome 3R, compared with 15.7% for the mitochondrial genome (Fig. 4).

Fig. 3 Maximum likelihood tree using *Anopheles melas* mitochondrial sequences from the Bijagós Archipelago, The Gambia and Cameroon and *An. gambiae* mitochondrial sequences from the Bijagós Archipelago. *An. melas* from the Bijagós split into two groups, labelled A and B. Support values can be seen in Supplementary Data 1

The clustering of Bijagós *An. melas* into groups A and B was further investigated using fixation index (F_{ST}) analysis across 100-bp and 1000-bp windows. The highest F_{ST} values were on the mitochondrial genome, with a mean F_{ST} value of 0.476, compared to a mean F_{ST} value between 0.038 and 0.040 on the other chromosomes (Table 1). This indicates that most of the population differentiation is due to genetic differentiation between *An. melas* groups A and B in regions on the mitochondrial genome.

Genes underlying peaks in F_{ST} were identified. There were seven genomic windows with an $F_{ST} \geq 0.5$. These windows of the genome include the protein-coding genes detailed in Table 2. This includes genes encoding the

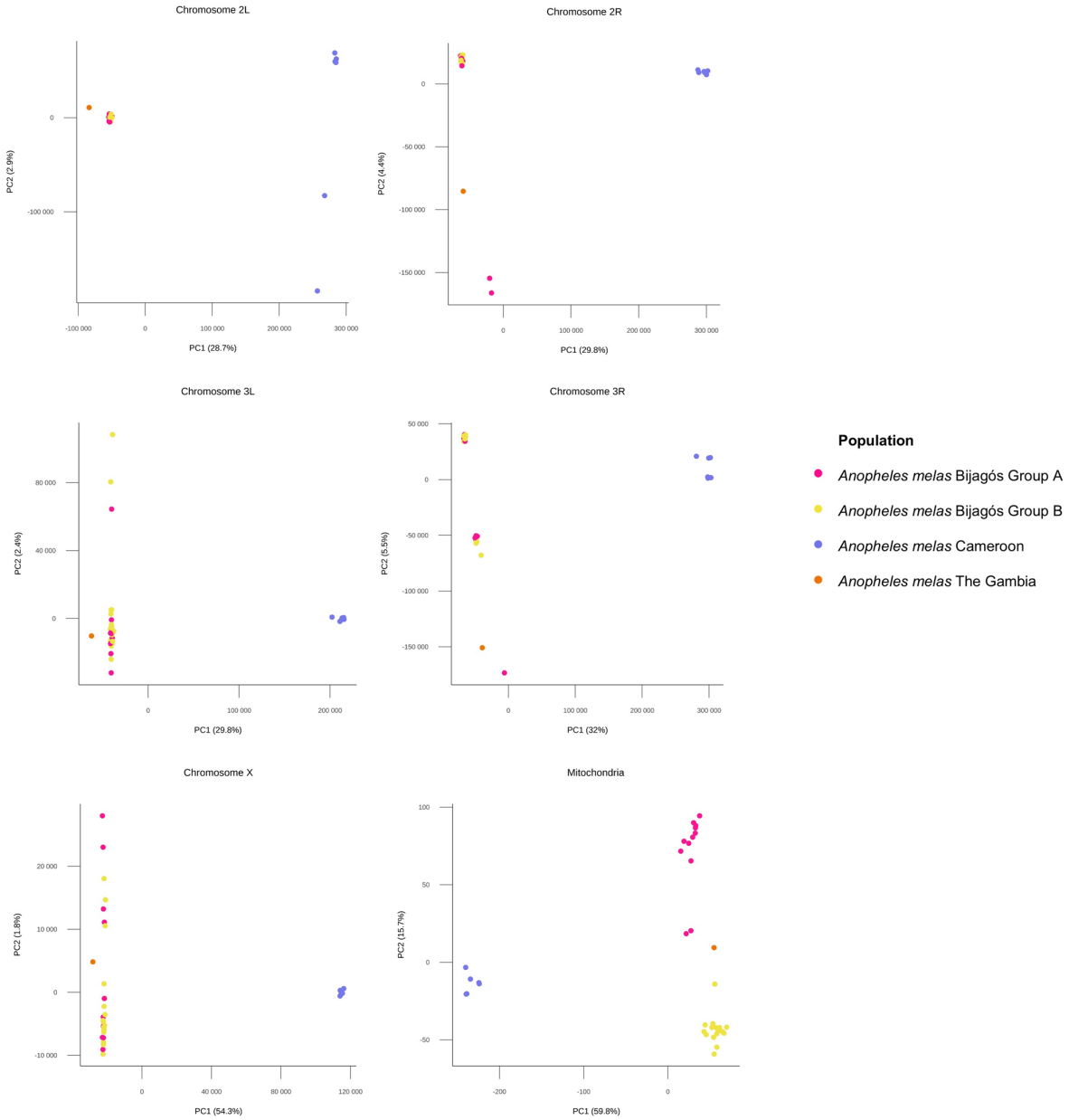


Fig. 4 Principal component analysis for *An. melas* from the Bijagós Archipelago, Guinea-Bissau ($N=30$), Cameroon ($N=6$) and The Gambia ($N=1$)

cytochrome C oxidase subunits COX I and COX II and the NADH dehydrogenase subunits 1, 4, 4L and 5.

Additional PCA was conducted to investigate the relationship between the genomes of the *An. melas* specimens from the Bijagós, Cameroon and The Gambia and *An. gambiae* s.s. mosquitoes from the Bijagós to investigate the possibility of hybridisation between *An. melas* and *An. gambiae* s.s. (Fig. 5). This mitochondrial PCA indicates that *An. melas* from the Bijagós separates from *Anopheles gambiae* s.s. from the Bijagós, and this relationship is also reflected in the PCA analyses of all other individual chromosomes (Supplementary Data 1). This gives no indication of hybridization between the *An. melas* and *An. gambiae* s.s. from the Archipelago.

Admixture analysis was conducted with WGS data from the combined sample set of $N=37$ *An. melas*

Table 1 F_{ST} calculated between the two clusters of *Anopheles melas*, A and B, from the Bijagós Archipelago, calculated for each chromosome and the mitochondrial genome

Chromosome	Mean (median) F_{ST} for each chromosome (100-bp windows)	Mean (median) F_{ST} for each chromosome (1000-bp windows)
2L	0.038 (0.023)	0.021 (0.014)
2R	0.039 (0.024)	0.025 (0.016)
3L	0.038 (0.024)	0.023 (0.014)
3R	0.040 (0.025)	0.025 (0.016)
X	0.038 (0.022)	0.026 (0.016)
Mt	0.476 (0.528)	0.331 (0.441)

available globally (Fig. 6). This admixture analysis indicated an optimum of K = 4 ancestral groups. *Anopheles melas* from the Bijagós and The Gambia had a mixture of three different K ancestries (K = 1, 2 and 4), whilst *An. melas* from Cameroon had a clearly separate ancestry (K = 3).

Population differentiation between *An. melas* from the Bijagós ($N = 30$) and *An. melas* from Cameroon ($N = 6$) was investigated using the fixation index (F_{ST}) statistic (Table 3).

Fifty-nine protein coding genes were identified as overlapping windows with $F_{ST} \geq 0.9$ on chromosome

Table 2 Protein coding genes in regions of high Fst (Fst ≥ 0.5) in the mitochondrial genome, comparing the two clusters of *Anopheles melas* from the Bijagós Archipelago

Window position in mitochondrial genome (100-bp windows)	Highest F_{ST} value	Gene ID	Description
1567–1666	0.799	AGAP028364	Cytochrome C oxidase subunit (*cox1*)
1667–1766	0.528		
3467–3566	0.630	AGAP028366	Cytochrome C oxidase subunit II (*cox2*)
7067–7166	0.758	AGAP028380	NADH dehydrogenase subunit 5 (*nadh5*)
8667–8766	0.548	AGAP028382	NADH dehydrogenase subunit 4 (*nadh4*)
9667–9766	0.678	AGAP028383	NADH dehydrogenase subunit 4L (*nadh4l*)
12,367–12,466	0.748	AGAP028389	NADH dehydrogenase subunit 1 (*nadh1*)

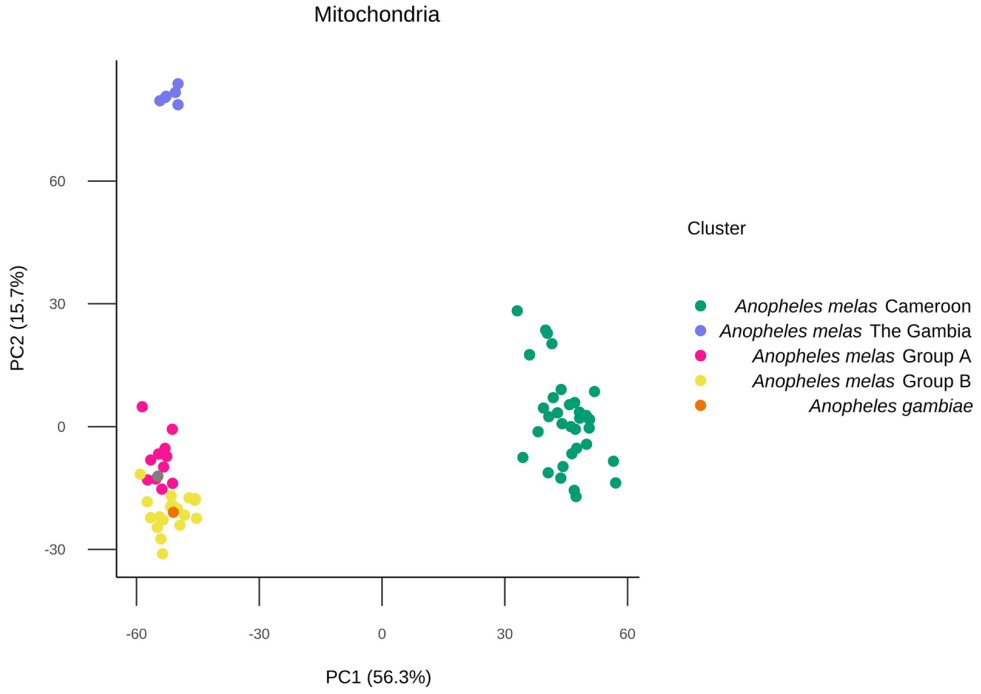


Fig. 5 Principal components analysis comparing *Anopheles melas* and *An. gambiae* s.s. from the Bijagós Archipelago. Includes additional *An. melas* samples from Cameroon and The Gambia

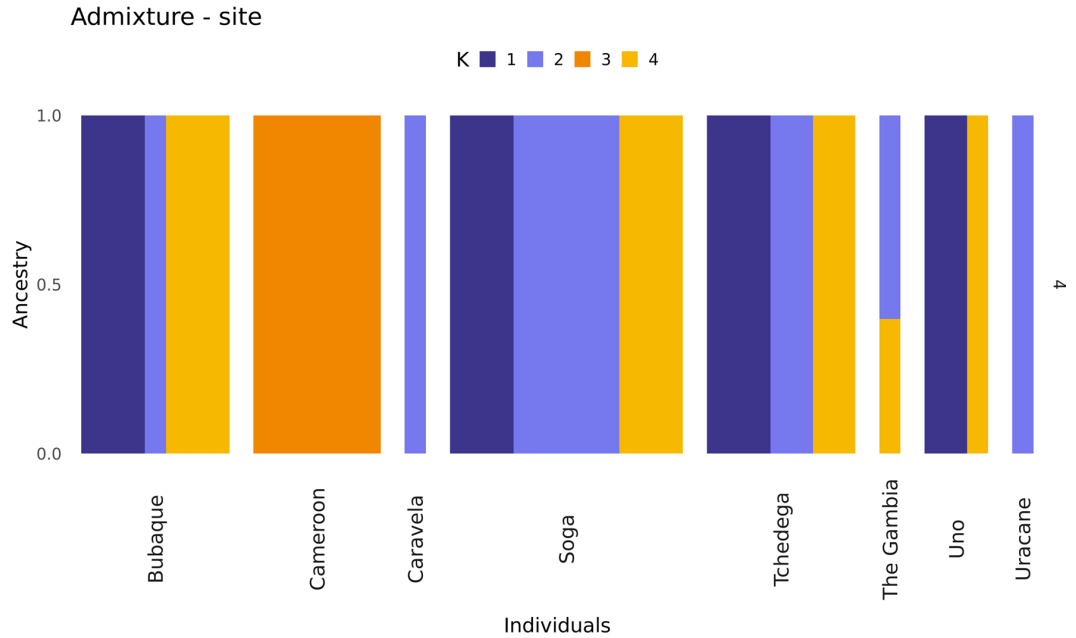


Fig. 6 Admixture based on geographic location, K = 4. *N* = 1 The Gambia, *N* = 6 Cameroon, *N* = 30 Bijagós [*N* = 11 from Soga, 7 from Bubaque, 7 from Tchedega (Maio), 3 from Uno, 1 from Caravela and 1 from Uracane]

2L, 72 on chromosome 2R, 51 on chromosome 3L, 56 on chromosome 3R and 56 on chromosome X (Supplementary Data 1). No genomic windows with $F_{ST} \geq 0.9$ were identified on the mitochondrial genome. Consequently, windows with $F_{ST} \geq 0.6$ were investigated, and 15 protein coding genes were identified (Supplementary Data 1). This included cytochrome C oxidase subunits II (*cox2*) and III (*cox3*) and NADH dehydrogenase subunits ND2 (*nadh2)*, ND3 (*nadh3*), ND4 (*nadh4*), ND4L (*nadh4l*) and ND5 (*nadh5*).

Table 3 Mean (median) F_{ST} for each chromosome, comparing *Anopheles melas* from the Bijagós and Cameroon

Chromosome	Mean (median) F_{ST} for each chromosome (1-kbp windows)
2L	0.257 (0.213)
2R	0.261 (0.216)
3L	0.270 (0.228)
3R	0.273 (0.233)
X	0.314 (0.245)
Mt	0.357 (0.305)

Selection analysis identified signals of selection within *An. melas* populations

Genome-wide selection scans of filtered variants were performed to identify signals of directional selection within and between populations of *An. melas*. Three different selection metrics were calculated: integrated haplotype score (iHS) was used to identify regions of the genome under selection within the *An. melas* Bijagós population. The cross-population haplotype homozygosity metric (XP-EHH) was used to identify regions of the genome under selection when comparing the Bijagós and Cameroon *An. melas* populations. Finally, Garud's H12 was computed as an additional method to identify both hard and soft selective sweeps within the Bijagós *An. melas* population. In a single population analysis of *An. melas* from the Bijagós using the iHS metric [44], 194 loci were identified as having significant iHS scores ($iHS \geq 4$) (Fig. 7). This included 102 SNPs in 29 different protein coding genes: 3 in chromosome 2L, 13 in chromosome 2R, 7 in chromosome 3L, 4 in chromosome 3R and 2 in chromosome X (Table 4). None of the protein coding genes identified with significant iHS scores have previously been implicated in insecticide resistance.

Cross-population analysis using XP-EHH was conducted between *An. melas* from the Bijagós and Cameroon (Fig. 8). More positive scores indicate positive selection in the Bijagós *An. melas* population (significant

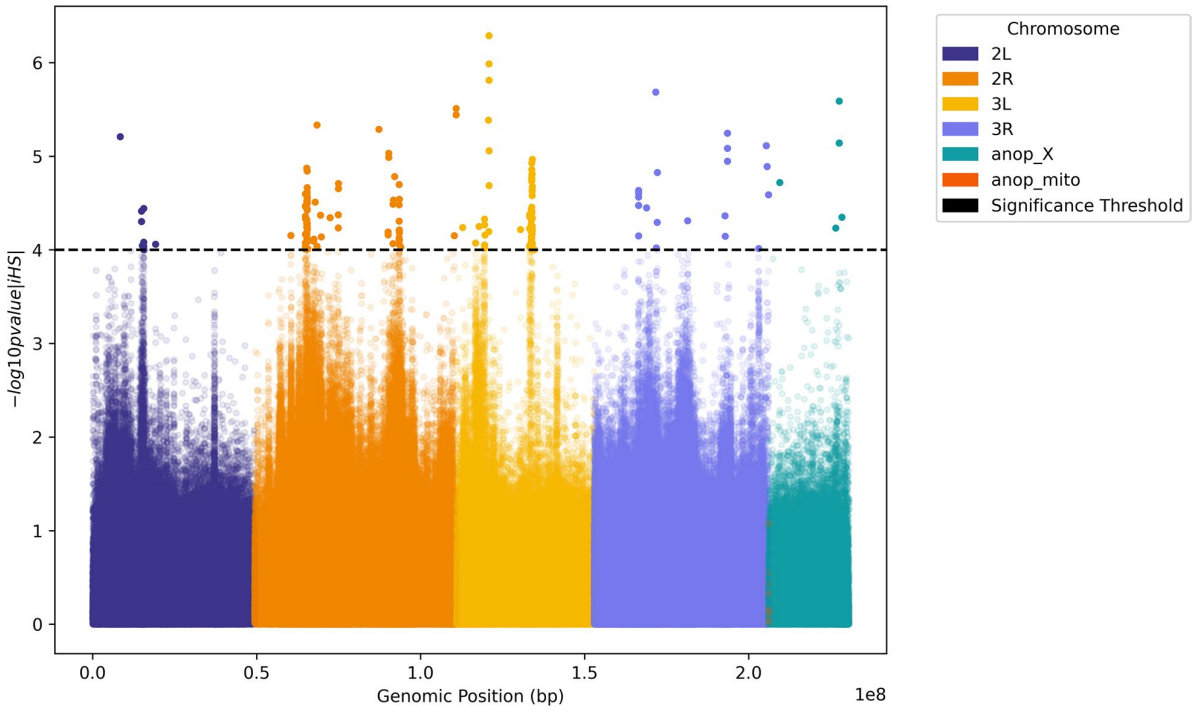


Fig. 7 Within population selection analysis using iHS scores for *Anopheles melas* from the Bijagós Archipelago

at XP-EHH ≥ 5), whereas more negative scores indicate positive selection in the Cameroon *An. melas* population (significant at XP-EHH ≤ − 5). Protein coding genes containing SNPs under positive selection in the Bijagós population (Table 5) and the Cameroon population (Table 6) are detailed.

Garud's H_{12} was used to identify signatures of recent positive selection in the Bijagós *An. melas* population [42]. H_{12} was computed per chromosome, and no clear peaks of selection were identified on any chromosome (Supplementary Data 1).

Detection of structural variants

Structural variants (SVs) were identified in the Bijagós *An. melas* mosquito population using DELLY software [45] and were discovered in relation to the AgamP4 (*An. gambiae)* reference genome. A total of 113,121 SVs were identified across the whole genome following quality control filtering. Of these 113,121 filtered SVs, 116 were identified in genes associated with insecticide resistance or in gene families associated with insecticide metabolism. This included 48 deletions, 38 inversions, 22 insertions and 8 duplications.

SVs were annotated using SnpEff [46]. Of the 48 identified deletions, four were annotated as having high impact. This included one deletion (40 bp) in chromosome 2R, which was found at 3% allelic frequency and resulted in a frameshift in *cyp6p15p* and the up- or downstream modification of *cyp6aa2, coeae60, cyp6p3* and *cyp6p5*. Two high-impact deletions on chromosome 2 L (766,776 bp) and 3R (714,267 bp) resulted in feature ablation of multiple genes and were found in all *An. melas* samples from the Bijagós and Cameroon, indicating that these deletions may be species specific to *An. melas* and could have been identified as a result of aligning to the AgamP4 reference genome. The fourth high-impact deletion was identified in chromosome X (505,172 bp) at 47.9% allelic frequency in the Bijagós population and 28.6% in the Cameroon population, which resulted in the deletion of multiple genes including AGAP000817, AGAP000816 and AGAP013474 (protein coding genes with unspecified products). Two deletions were annotated as having moderate impact. The first was found at 1.7% allelic frequency in the *cyp6m4* gene on chromosome 3R (575 bp). The second was on chromosome 2R (38 bp) in AGAP013202 and was identified in all *An. melas* samples from the Bijagós and Cameroon so may be *An. melas* specific.

Of the 22 insertions identified, 5 were in introns, 8 were in intergenic regions, 4 were downstream gene variants, and 5 were upstream gene variants. None of these

Table 4 Significant iHS scores (iHS ≥ 4) in protein coding genes in *Anopheles melas* from the Bijagós Archipelago

Chromosome	iHS score	Protein coding gene
2L	4.36	AGAP005394—MFS domain-containing protein
	4.19	AGAP005449—E3 ubiquitin-protein ligase CBL
	4.00	AGAP005450—protein coding gene—unspecified product
2R	4.15	AGAP001824—protein coding gene—unspecified product
	4.50	AGAP002118—zinc finger protein 622
	4.14	AGAP002119—dual-specificity tyrosine-(Y)-phosphorylation regulated kinase
	4.54	AGAP002123—axin
	4.09	AGAP002235—GATA-binding protein 4/5/6
	4.11	AGAP002243—ankyrin repeat and FYVE domain-containing protein 1
	4.51	AGAP002274—protein coding gene—unspecified product
	4.69	AGAP002299—XK-related protein
	4.37	AGAP002336—Ig-like domain-containing protein
	4.34	AGAP002573—GTPase-activating Rap/Ran-GAP domain-like protein 3
	4.49	AGAP002677—coiled-coil domain-containing protein lobo homolog
	4.78	AGAP029573—protein coding gene—unspecified product
	4.15	AGAP004670—protein coding gene—unspecified product
3L	4.24	AGAP010344—solute carrier family 26
	4.07	AGAP010536—nucleolar complex protein 2
	5.39	AGAP029471—protein coding gene—unspecified product
	6.03	AGAP029721—fibrinogen C-terminal domain-containing protein
	4.22	AGAP011223—fibrinogen C-terminal domain-containing protein
	4.30	AGAP011379—Frizzled receptor
	4.46	AGAP011384—protein coding gene—unspecified product
3R	4.45	AGAP008712—solute carrier organic anion transporter family member
	4.02	AGAP008826—protein coding gene—unspecified product
	4.15	AGAP009716—cadherin
	5.11	AGAP010295—Ca_chan_IQ domain-containing protein
X	4.23	AGAP001046—Abl interactor 2
	4.35	AGAP001064—transmembrane emp24 domain-containing protein 10 precursor

insertions were identified as having high impact, and 14 were identified in all of the Bijagós *An. melas* isolates, indicating that they may be species specific insertions.

Of the eight identified duplications, five were annotated as having high impact. Two of these duplications were identified on chromosome 2R at 2% allelic frequency. The first leads to bidirectional gene fusion between *cyp6p4* and the solute carrier family 8 sodium/calcium exchanger (AGAP002859), and the second leads to bidirectional gene fusion between AGAP002876 (encoding a DNA glycosylase) and AGAP002877 (encoding a tetratricopeptide repeat protein). Two other high-impact duplications were identified on chromosome 3R. The first was found at 2% allelic frequency and leads to gene fusion between *cyp6m2* and *cyp6m3*, whereas the second occurred at 4% allelic frequency and results in a premature stop codon in *gste1*. Finally, a large duplication on chromosome X between genomic positions 14,939,828 and 15,316,733 (376,905 bp) was identified to have high impact. All Bijagós *An. melas* and six of the seven Cameroon *An. melas* were heterozygous for this duplication, which spans multiple protein coding genes including the cytochrome P450 *cyp9k1*. In addition, one moderate impact duplication was identified at 1.7% allelic frequency in *cyp9k1* on chromosome X.

Of the 38 identified inversions, 8 were annotated as having high impact. This included three different inversions in the *vgsc* gene on chromosome 2L found at 10%, 4% and 2% allelic frequency, respectively, two different inversions in the AGAP009189 gene on chromosome 3R found at 2% allelic frequency, one inversion in *gste8* on chromosome 3R found at 4% allelic frequency and one inversion in the zf-C3Hc3H domain-containing protein (AGAP029190) gene in chromosome 2R found at 6% frequency. One large, high-impact, inversion in chromosome 2L spanned from genomic positions 2,161,433 to 2,666,161 and led to bidirectional gene fusion between AGAP004717 and AGAP029667

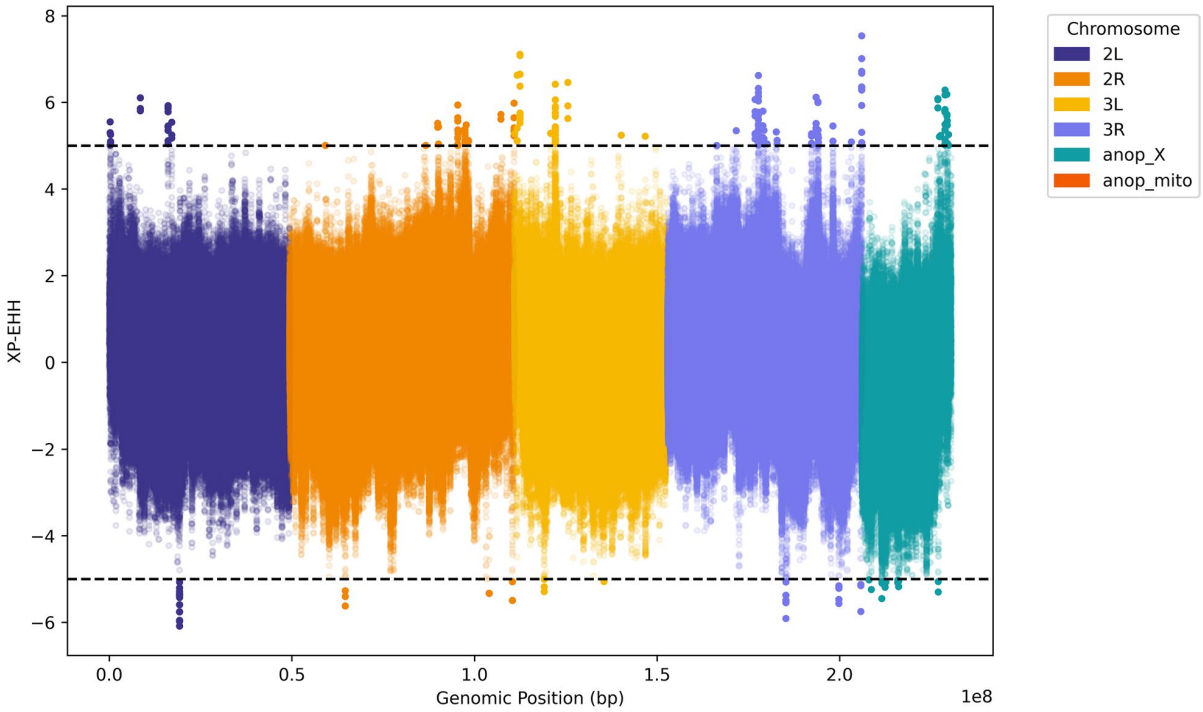


Fig. 8 Cross population selection analysis using XP-EHH metric between *Anopheles melas* from the Bijagós and Cameroon

Table 5 Protein coding genes containing SNPs under positive selection in the Bijagós *Anopheles melas* population, relative to the Cameroon *An. melas* population, with XP-EHH scores ≥ 5

Chromosome	XP-EHH score	Protein coding gene
2L	5.54	AGAP005483—transcription factor grauzone
	5.32	AGAP005562—RING-box protein 2
2R	5.00	AGAP001773—putative allatostatin receptor 2
	5.36	AGAP003997—casein kinase 1, gamma
	5.20	AGAP004032—alpha-mannosidase
3L	5.28	AGAP010310—eukaryotic peptide chain release factor subunit
	5.56	AGAP010324—Groucho
	5.29	AGAP029721—fibrinogen C-terminal domain-containing protein
	5.45	AGAP010814—TEP6 thioester-containing protein 6
3R	5.52	AGAP028085—LITAF domain-containing protein
	5.39	AGAP009115—phosphatidylinositol phospholipase C, beta
	5.08	AGAP009253—protein coding gene—unspecified product
	5.17	AGAP009723—cadherin
	5.28	AGAP029826—Tox-SGS domain-containing protein
X	5.41	AGAP001073—protein coding gene—unspecified product

(protein coding genes with unspecified products). All *An. melas* Bijagós and Cameroon mosquitoes were homozygous alternate for this inversion, indicating that this could be a species-specific inversion in *An. melas* identified through alignment to the AgamP4 reference genome. One additional moderate impact inversion was identified on chromosome 3L at 15% allelic frequency, impacting the DNA-directed RNA polymerase III subunit RPC1 (AGAP004703), compass component SPP1 (AGAP004704), arginyl-tRNA synthetase

Table 6 Protein coding genes containing SNPs under positive selection in the Cameroon *Anopheles melas* population, relative to the Bijagós *An. melas* population, with XP-EHH scores ≤ − 5

Chromosome	XP-EHH score	Protein coding gene
2R	− 5.35	AGAP013143—Rho GTPase-activating protein 26
3L	− 5.15	AGAP010663—female reproductive tract protease GLEANR_2575
3R	− 5.48	AGAP009415—lysophosphatidate acyltransferase
	− 5.30	AGAP009970—Cullin-associated NEDD8-dissociated protein 1
X	− 5.02	AGAP000116—Rab11 family-interacting protein 1/2/5
	− 5.24	AGAP013158—protein coding gene—unspecified product
	− 5.45	AGAP000304—protein twisted gastrulation
	− 5.07	AGAP000311—UNC93-like protein MFSD11
	− 5.19	AGAP000356—SCP domain-containing protein
	− 5.06	AGAP000547—rabenosyn-5
	− 5.11	AGAP000567—protein coding gene—unspecified product

Table 7 Non-synonymous SNPs identified in the Bijagós *Anopheles melas* mosquito population in resistance genes

Gene	Chromosome	Position[a]	Average read depth	SNP[b]	Transcript	Homozygous reference	Heterozygous	Homozygous alternate	Allelic frequency
vgsc	2L	2,358,197	31.37	S14C	AGAP004707-RD	21	9	0	15%
		2,390,341	59.67	T309A	AGAP004707-RD	29	1	0	2%
		2,390,449	63.67	G317D	AGAP004707-RD	28	2	0	3%
		2,390,472	61.37	S325T	AGAP004707-RD	29	1	0	2%
		2,390,737	56.73	L377Q	AGAP004707-RD	29	1	0	2%
		2,391,309	69.53	E429K	AGAP004707-RD	1	29	0	48%
		2,400,071	33.63	†M490I	AGAP004707-RD	29	1	0	2%
		2,416,868	33.47	L754M	AGAP004707-RD	28	2	0	3%
		2,424,377	31.57	N1123D	AGAP004707-RD	28	2	0	3%
		2,429,991	33.37	E1628D	AGAP004707-RD	29	1	0	2%
		2,431,184	36.30	R1975Q	AGAP004707-RD	29	1	0	2%
		2,431,396	33.33	G2046S	AGAP004707-RD	0	0	30	100%
gste2	3R	28,597,858	34.53	I187F	AGAP009194-RA	29	1	0	2%
		28,597,879	34.47	H180Y	AGAP009194-RA	29	1	0	2%
		28,597,905	35.83	G171D	AGAP009194-RA	23	7	0	12%
		28,598,032	36.73	D129N	AGAP009194-RA	29	1	0	2%
		28,598,526	28.37	E19K	AGAP009194-RA	29	1	0	2%
		28,598,573	29.83	N3S	AGAP009194-RA	4	11	15	68%
rdl	2L	25,433,558	28.60	P474Q	AGAP006028-RA	29	1	0	2%
		25,433,561	27.70	P475Q	AGAP006028-RA	29	1	0	2%
ace1	2R	3,489,216	34.23	E2K	AGAP001356-RA	29	1	0	2%
		3,489,310	34.50	P33Q	AGAP001356-RA	29	1	0	2%
		3,489,391	36.37	A60V	AGAP001356-RA	29	1	0	2%
		3,493,397	28.50	P644T	AGAP001356-RA	26	3	1	8%
		3,493,401	28.77	N645S	AGAP001356-RA	28	1	0	2%
		3,493,715	28.47	A714V	AGAP001356-RA	28	1	0	2%
		3,493,759	31.03	L729F	AGAP001356-RA	29	1	0	2%
		3,493,771	31.00	V733I	AGAP001356-RA	28	2	0	3%

[a] SNP positions correspond to the AgamP4 reference genome. [b]Codon numbering according to transcript in the AgamP4 reference genome. M490I has been highlighted with a † symbol

(AGAP004708) and five protein coding genes with unspecified products.

Identification of non-synonymous SNPs in resistance genes

We investigated the presence of target site mutations in the *vgsc, gste2, rdl* and *ace1* genes, which have previously been associated with insecticide resistance (Supplementary Data 1). In total, we identified 28 non-synonymous mutations in these resistance genes (Table 7). Of these, the only mutation that has been reported previously is *vgsc* M490I, which we identified at an allelic frequency of 2% and is highlighted with a † symbol in Table 7. This mutation was previously identified in *An. gambiae* samples from Kenya, where it was found to potentially be under selection [47]. To our knowledge, none of the other missense mutations identified here have previously been reported in the *An. gambiae* s.l. complex. The *vgsc* G2046S mutation was fixed in the Bijagós *An. melas* population (100% allelic frequency) and was identified in one *An. melas* sample from The Gambia (100% allelic frequency) but was not found in any *An. melas* samples from Cameroon (0% allelic frequency).

Discussion

Anopheles melas is highly abundant on the Bijagós Archipelago of Guinea-Bissau and may have a role in malaria transmission [4] (Pretorius et al. 2024, in review). However, the population structure and insecticide resistance status of this malaria vector are not well understood. This study used WGS data from 30 *An. melas* from across the Archipelago to investigate genetic diversity, population structure and signatures of selection in insecticide resistance genes within this vector population.

Maximum likelihood trees generated using the whole mitochondrial genome showed that *An. melas* from the Bijagós split into two groups. Mosquito samples were collected from six different islands in the Archipelago, but this split was not associated with sampling island, and analyses indicated that the clustering of *An. melas* into two groups was due to genetic differentiation on the mitochondrial genome. The protein coding genes underlying the peaks in F_{ST} on the mitochondrial genome included the cytochrome C oxidase subunits *cox1* and *cox2* and the NADH dehydrogenase subunits *nadh1*, *nadh4*, *nadh4L* and *nadh5*. Both *cox1* and *nadh4* have previously been used to investigate phylogenetic relationships within multiple vector complexes and between cryptic species, including the *An. gambiae* s.l. complex [48–50]. Further investigation with a larger sample size of *An. gambiae* s.l. complex mosquitoes from the Bijagós Archipelago will help to understand the clusters observed.

The computed maximum likelihood trees indicate that *An. melas* from the Bijagós are closely related to *An. melas* from The Gambia and Cameroon. The level of genetic differentiation between *An. melas* from the Bijagós and Cameroon was higher than previously identified for other species of *An. gambiae*. Average F_{ST} across chromosomes 3L and 3R between the Bijagós and Cameroon *An. melas* was 0.27. However, F_{ST} across these chromosomes was previously identified at a tenfold lower score of 0.028 between mainland Guinea-Bissau and Cameroon *An. gambiae* [25]. Higher than expected levels of genetic differentiation have previously been identified between *An. melas* populations using whole-genome data. A previous study by Deitz et al. found that divergence between large *An. melas* clusters along the West African coast were due to high levels of differentiation across the whole genome, indicative of allopatric separation [27]. Nucleotide diversity was similar to that previously identified for *An. melas* from Bioko Island, Equatorial Guinea ($\pi = 0.0034$) [27], and was lower than the average nucleotide diversity calculated for *An. gambiae* sampled from 15 locations across Africa ($\pi = 0.015$) [25]. Ancestry analysis using a combined database of these *An. melas* samples indicated $K = 4$ ancestral populations, with *An. melas* samples from the Bijagós and The Gambia sharing ancestries $K = 1$, 2 and 4 and Cameroon *An. melas* samples sharing a distinct $K = 3$ ancestry. There was no clear distinction between the ancestries of *An. melas* from different islands on the Archipelago, indicating historical gene flow between the islands, despite the geographical distance between islands being greater than the distance *An. gambiae* s.l. are known to disperse [51]. This is supported by a previous study, which identified extensive gene flow between *Anopheles* s.s. on the Bijagós Archipelago and mainland Guinea-Bissau [52]. Fixation index analysis identified 59 protein coding genes with high genetic differentiation between *An. melas* from the Bijagós and Cameroon, including the *cyp307a1* gene (AGAP001309) on chromosome X, which is a member of the cytochrome P450 gene family associated with metabolic resistance to insecticides [53, 54]. This analysis was conducted with a large sample size disparity, with 30 samples from the Bijagós vs. 6 samples from Cameroon. Additional WGS data are required to further investigate this genetic differentiation.

Genome-wide selection scans were computed to identify signatures of selection across the genome. Within-population analysis of Bijagós *An. melas* using the iHS statistic identified signatures of directional selection in 29 protein coding genes, none of which have previously been associated with insecticide resistance. Cross-population analysis between *An. melas* from the Bijagós and Cameroon using the XP-EHH metric identified

15 protein coding genes within the Bijagós population undergoing positive selection compared to the Cameroon population. This included the gene encoding *tep6*, a thioester-containing protein in the same family as *tep1*, which is implicated in *An. gambiae* resistance to parasite infection [55]. This analysis also identified 11 protein coding genes undergoing positive selection in the Cameroon population compared to the Bijagós population, none of which have previously been associated with insecticide resistance. Genome-wide selection scans using the H_{12} statistic did not identify any clear selective sweeps in the Bijagós *An. melas* genome, in contrast to our previous study of *An. gambiae* s.s. collected during 2022 from Bubaque Island, where H_{12} analysis identified two distinct selective sweeps on chromosomes X and 2R spanning multiple cytochrome-P450 genes involved in insecticide metabolism ([26], in review). The absence of selective sweeps in insecticide resistance-associated genes in the *An. melas* genome suggests that this species may be under less selective pressure from insecticides than *An. gambiae* on these islands.

Structural variants (SVs) were analysed in the Bijagós *An. melas* population in relation to the *An. gambiae* AgamP4 reference genome. One deletion identified in chromosome 2R at 3% allelic frequency resulted in a frameshift in *cyp6p15p* and modification of *cyp6aa2, coeae60, cyp6p3* and *cyp6p5*. These *cyp6p* genes are associated with metabolic insecticide resistance in mosquitoes [53, 54, 56, 57]. Another variant resulted in the duplication of multiple protein coding genes including the cytochrome P450 *cyp9k1*, which metabolises deltamethrin and has been associated with pyrethroid resistance in *An. coluzzii* populations following vector control interventions [21, 22]. Furthermore, *cyp9k1* duplications have previously been identified in *An. gambiae* s.l. complex mosquitoes from mainland Guinea-Bissau ([23], Supplementary S7) [23]. Copy number variants resulting in gene duplication are under positive selection in the *An. gambiae* s.l. complex [23], and CNVs leading to increased expression of metabolic genes have been shown to increase insecticide metabolism, leading to insecticide resistance [56, 58–60]. Eight high-impact inversions were identified, including three inversions in the *vgsc* gene associated with resistance to DDT and pyrethroids [17, 18, 56, 61], and one inversion in the *gste8* gene, which is in the same gene family as *gste2*, which encodes a DDT-detoxifying enzyme [62].

Analysis of non-synonymous SNPs in insecticide resistance genes identified the *vgsc* M490I mutation in the Bijagós *An. melas* population at low frequency, which has previously been reported in *An. gambiae* in Kenya as under possible directional selection [47]. No other SNPs previously associated with insecticide resistance were found. In our previous study of amplicon data from *An. melas* from the Archipelago collected in the same year, three SNPs previously associated with insecticide resistance were identified. These were *vgsc* L995F (2.14% allelic frequency), N1570Y (1.12% allelic frequency) and A1746S (0.76% allelic frequency) [26]. However, these SNPs were identified at very low frequency and found at significantly lower frequency in *An. melas* than in *An. gambiae* s.s. Therefore, the absence of these SNPs in our dataset is not unexpected.

The absence of insecticide resistance-associated SNPs in *An. melas* further suggests that this species is under less insecticide resistance selection pressure than *An. gambiae* on the Bijagós Archipelago. This is supported by a previous study in Equatorial Guinea, where no insecticide resistance mutations were identified in *An. melas* [12]. This may be because *An. melas* is biting people outdoors, circumventing exposure to the insecticides in ITNs, or because *An. melas* may be feeding mostly on non-human hosts [6]. In an entomological survey on the Bijagós in 2019, a greater proportion of *Plasmodium*-positive *An. melas* were caught in the outdoor than indoor traps. However, *Plasmodium*-positive *An. melas* were also caught in indoor traps, indicating human-host seeking indoors and outdoors (Pretorius et al. 2024, in review). As ITNs are the only vector control method used in the Bijagós, outdoor biting would reduce selection pressure for resistance evolution in this species. Whilst maintaining susceptibility to insecticides is positive, preferential outdoor biting by *An. melas* may present further issues for vector control on the islands as conventional ITNs and IRS may not be as effective. This is supported by studies in Equatorial Guinea which identified high levels of outdoor biting by *An. melas* [63]. Alternatively, the absence of insecticide resistance-associated mutations and selective sweeps in *An. melas* found in this study may be because this species is evolving separate molecular mechanisms of resistance to *An. gambiae* s.s. Notably, *Anopheles* mosquitoes are among the most genetically diverse eukaryotic organisms known [25], and it is plausible that different species may evolve unique molecular pathways for resistance. Further investigation using additional sampling, phenotypic bioassays, synergist-insecticide bioassays and 'omics studies should be undertaken to understand the resistance status and molecular mechanisms of resistance in this understudied malaria vector.

The main limitations of this study are the absence of phenotypic insecticide resistance data for *An. melas* and that the current *An. melas* reference genome necessitated aligning our *An. melas* WGS data to the *An. gambiae* (AgamP4) reference genome. The AgamP4 reference genome is of higher quality, and 92.8% of *An. melas* reads mapped successfully to AgamP4 compared to

79.3% mapping to the poorer quality AmelC2 reference genome, giving us confidence in our approach. However, future analyses would benefit from a chromosome-level reference genome assembly for *An. melas*, particularly as all structural variants were discovered in relation to the AgamP4 reference. Furthermore, the genes discussed in this study have been annotated from the *An. gambiae* reference genome, and though these genes are likely to have very similar functions in *An. melas*, they could play different roles.

Conclusions

In conclusion, using WGS data, this study identifies two separate phylogenetic clusters of *An. melas* on the Bijagós Archipelago of Guinea-Bissau because of genetic differentiation on the mitochondrial genome. Structural variants encompassing genes that could be involved in metabolic insecticide resistance were identified. However, common SNPs associated with insecticide resistance in *An. gambiae* s.s. were absent in the *An. melas* population, and there were no clear signatures of selection in known insecticide-resistance genes. This suggests that *An. melas* may experience less selective pressure for insecticide resistance evolution than *An. gambiae*, potentially through biting outdoors and circumventing selection pressure from ITNs or because they are feeding primarily on non-human hosts. Further investigations using larger data sets and phenotypic bioassays are required.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13071-024-06476-2.

Additional file 1

Acknowledgements
We acknowledge the communities of the Bijagós Archipelago, where these mosquito samples were collected, and all who took part or assisted in the cross-sectional survey.

Author contributions
EP coordinated mosquito collection on the Bijagós Archipelago during a cross-sectional survey with HH, ES and RTJ. The cross-sectional survey was designed by AL, EP and RTJ. SCe conducted DNA extraction and species identification at the Medical Research Council, The Gambia (MRCG), under the supervision of MON. SM designed this study and conducted all bioinformatic analysis of sequence data under the supervision of TGC and SCa, with guidance from HAF, EC, MH and JP. HV assisted with sample logistics and transport. SM wrote the first draft of this manuscript. All authors reviewed and approved the final manuscript.

Funding
SM and EC are funded by Medical Research Council UK (grant no. MR/N013638/1). TGC, SC, HAP and MH are funded by Medical Research Council UK grants (ref. no. MR/M01360X/1, MR/N010469/1, MR/R025576/1, MR/R020973/1 and MR/X005895/1). AL, EP, ETS, SC, RTJ and HV are funded by Joint Global Health Trials Scheme (MRC, Wellcome Trust, UKRI, NIHR, grant no. MR/S005013/1).

Availability of data and materials
The raw sequence data generated and analysed during this study are available in the European Nucleotide Archive (project ID: PRJEB75927, accession numbers ERS20101387 to ERS20101416).

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Faculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, London WC1E 7HT, UK. [2]Medical Research Council Unit The Gambia at the London School of Hygiene and Tropical Medicine, Banjul, The Gambia. [3]Projecto de Saúde Bandim, Bissau, Guinea-Bissau. [4]Ministério de Saúde Pública, Bissau, Guinea-Bissau. [5]Clinical Academic Group, Institute for Infection and Immunity, and St. George's University Hospitals NHS Foundation Trust, St. George's University of London, London, UK. [6]Centre de Recherches Médicales de Lambaréné (CERMEL), Lambaréné, Gabon. [7]Institut Für Tropenmedizin Universitätsklinikum Tübingen, Tübingen, Germany. [8]Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London WC1E 7HT, UK.

Received: 22 May 2024 Accepted: 3 September 2024
Published online: 18 September 2024

References

1. World Health Organization. World Malaria Report 2023. https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2023. Accessed 15 Jan 2024.
2. UNESCO. Boloma Bijagós Biosphere Reserve, Guinea-Bissau. https://www.unesco.org/en/mab/bolama-bijagos. Accessed 28 Mar 2024
3. National Institute of Statistics, Guinea-Bissau. Guinea-Bissau 2009 Census. https://guinebissau.opendataforafrica.org/amthtjd/guinea-bissau-2009-census. Accessed 28 Mar 2024
4. Ant T, Foley E, Tytheridge S, Johnston C, Goncalves A, Ceesay S, et al. A survey of *Anopheles* species composition and insecticide resistance on the island of Bubaque, Bijagós Archipelago, Guinea-Bissau. Malar J. 2020;19:1–9.
5. Hutchins H, Bradley J, Pretorius E, Teixeira da Silva E, Vasileva H, Jones RT, et al. Protocol for a cluster randomised placebo-controlled trial of adjunctive ivermectin mass drug administration for malaria control on the Bijagós Archipelago of Guinea-Bissau: the MATAMAL trial. BMJ Open. 2023;13:e072347.
6. Sinka ME, Bangs MJ, Manguin S, Coetzee M, Mbogo CM, Hemingway J, et al. The dominant *Anopheles* vectors of human malaria in Africa, Europe and the Middle East: occurrence data, distribution maps and bionomic précis. Parasit Vectors. 2010;3:117.
7. Thomson RCM. Studies on the breeding places and control of *Anopheles gambiae* and *A. gambiae* var. melas in Coastal Districts of Sierra Leone. Bull Entom Res. 1946;36:185–252.
8. Giglioli ME. Oviposition by *Anopheles melas* and its effect on egg survival during the dry season in the Gambia. West Africa Ann Entomol Soc Am. 1965;58:885–91.
9. Sy O, Sarr PC, Assogba BS, Nourdine MA, Ndiaye A, Konaté L, et al. Residual malaria transmission and the role of *Anopheles arabiensis* and *Anopheles melas* in central Senegal. J Med Entomol. 2023;60:546–53.

10. Caputo B, Nwakanma D, Jawara M, Adiamoh M, Dia I, Konate L, et al. *Anopheles gambiae* complex along The Gambia river, with particular reference to the molecular forms of *An. gambiae* s.s. Malar J. 2008;7:182.
11. Bryan JH, Petrarca V, Di Deco MA, Coluzzi M. Adult behaviour of members of the *Anopheles gambiae* complex in the Gambia with special reference to *An. melas* and its chromosomal variants. Parassitologia. 1987;29:221–49.
12. Overgaard HJ, Reddy VP, Abaga S, Matias A, Reddy MR, Kulkarni V, et al. Malaria transmission after five years of vector control on Bioko Island, Equatorial Guinea. Parasit Vectors. 2012;5:253.
13. McGregor D, da Silva ET, Grignard L, Goncalves A, Vasileva H, Mabey D, et al. The epidemiology of *Plasmodium falciparum* malaria in the Bijagós islands of Guinea-Bissau. Am J Trop Med and Hyg. 2021;104:2117–22.
14. Hutchins H, Power G, Ant T, Teixeira E, Goncalves A, Rodrigues A, et al. A survey of knowledge, attitudes and practices regarding malaria and bed nets on Bubaque Island, Guinea—Bissau. Malar J. 2020;19:1–15.
15. Bhatt S, Weiss DJ, Cameron E, Bisanzio D, Mappin B, Dalrymple U, et al. The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. Nature. 2015;526:207.
16. World Health Organization. Malaria Threat Map. https://apps.who.int/malaria/maps/threats/. Accessed 1 Mar 2024.
17. Ranson H, Jensen B, Vulule JM, Wang X, Hemingway J, Collins FH. Identification of a point mutation in the voltage-gated sodium channel gene of Kenyan *Anopheles gambiae* associated with resistance to DDT and pyrethroids. Insect Mol Biol. 2000;9:491–7.
18. Martinez-Torres D, Chandre F, Williamson MS, Darriet F, Bergé JB, Devonshire AL, et al. Molecular characterization of pyrethroid knockdown resistance (kdr) in the major malaria vector *Anopheles gambiae* s.s. Insect Mol Biol. 1998;7:179–84.
19. Jones CM, Liyanapathirana M, Agossa FR, Weetman D, Ranson H, Donnelly MJ, et al. Footprints of positive selection associated with a mutation (N1575Y) in the voltage-gated sodium channel of *Anopheles gambiae*. Proc Natl Acad Sci U S A. 2012;109:6614–9.
20. Lucas ER, Rockett KA, Lynd A, Essandoh J, Grisales N, Kemei B, et al. A high throughput multi-locus insecticide resistance marker panel for tracking resistance emergence and spread in *Anopheles gambiae*. Sci Rep. 2019;9:13335.
21. Vontas J, Grigoraki L, Morgan J, Tsakireli D, Fuseini G, Segura L, et al. Rapid selection of a pyrethroid metabolic enzyme CYP9K1 by operational malaria control activities. PNAS. 2018;115:4619–24.
22. Hearn J, Djoko Tagne CS, Ibrahim SS, Tene-Fossog B, Mugenzi LMJ, Irving H, et al. Multi-omics analysis identifies a CYP9K1 haplotype conferring pyrethroid resistance in the malaria vector *Anopheles funestus* in East Africa. Mol Ecol. 2022;31:3642–57.
23. Lucas ER, Miles A, Harding NJ, Clarkson CS, Lawniczak MKN, Kwiatkowski DP, et al. Whole-genome sequencing reveals high complexity of copy number variation at insecticide resistance loci in malaria mosquitoes. Genome Res. 2019;29:1250–61.
24. Ibrahim SS, Amvongo-Adjia N, Wondji MJ, Irving H, Riveron JM, Wondji CS. Pyrethroid resistance in the major malaria vector *Anopheles funestus* is exacerbated by overexpression and overactivity of the P450 CYP6AA1 across Africa. Genes. 2018;9:140.
25 Anopheles gambiae 1000 Genomes Consortium. Genetic diversity of the African malaria vector *Anopheles gambiae*. Nature. 2017;552:96–100.
26. Moss S, Pretorius E, Ceesay S, Hutchins H, da Silva ET, Ndiath MO, et al. Genomic surveillance of *Anopheles* mosquitoes on the Bijagós Archipelago using custom targeted amplicon sequencing identifies mutations associated with insecticide resistance. Parasit Vectors. 2024;17:10.
27 Deitz KC, Athrey GA, Jawara M, Overgaard HJ, Matias A, Slotman MA. Genome-wide divergence in the West-African malaria vector *Anopheles melas*. G3. 2016;6:2867–79.
28. Neafsey DE, Waterhouse RM, Abai MR, Aganezov SS, Alekseyev MA, Allen JE, et al. Mosquito genomics. Highly evolvable malaria vectors: the genomes of 16 Anopheles mosquitoes. Science. 2015;347:1258522.
29. Deitz KC, Athrey G, Reddy MR, Overgaard HJ, Matias A, Jawara M, et al. Genetic isolation within the malaria mosquito *Anopheles melas*. Mol Ecol. 2012;21:4498–513.
30. Mboera LE, Kihonda J, Braks MA, Knols BG. Short report: Influence of Centers for Disease Control light trap position, relative to a human-baited bed net, on catches of *Anopheles gambiae* and *Culex quinquefasciatus* in Tanzania. Am J Trop Med Hyg. 1998;59:595–6.
31. Coetzee M. Key to the females of Afrotropical *Anopheles* mosquitoes (Diptera: Culicidae). Malar J. 2020;19:70.
32. Fanello C, Santolamazza F, Della TA. Simultaneous identification of species and molecular forms of the *Anopheles gambiae* complex by PCR-RFLP. Med Vet Entomol. 2002;16:461–4.
33 Neafsey DE, Christophides GK, Collins FH, Emrich SJ, Fontaine MC, Gelbart W, et al. The evolution of the Anopheles 16 genomes project. G3. 2013;3:1191–4.
34. Lucas ER, Nagi SC, Egyir-Yawson A, Essandoh J, Dadzie S, Chabi J, et al. Genome-wide association studies reveal novel loci associated with pyrethroid and organophosphate resistance in *Anopheles gambiae* and *Anopheles coluzzii*. Nat Commun. 2023;14:4946.
35. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.
36. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv. 2013;16:1303.
37 Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. GigaScience. 2021. https://doi.org/10.1093/gigascience/giab008.
38. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–303.
39. Letunic I, Bork P. Interactive tree of life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics. 2007;23:127–8.
40. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19:1655–64.
41. Anopheles gambiae 1000 Genomes Consortium. Genome variation and population structure among 1142 mosquitoes of the African malaria vector species *Anopheles gambiae* and *Anopheles coluzzii*. Genome Res. 2020;30:1533–46.
42. Garud NR, Rosenberg NA. Enhancing the mathematical properties of new haplotype homozygosity statistics for the detection of selective sweeps. Theor Popul Biol. 2015;102:94–101.
43. Tang K, Thornton KR, Stoneking M. A new approach for using genome scans to detect recent positive selection in the human genome. PLoS Biol. 2007;5:e171.
44 Gautier M, Klassmann A, Vitalis R. rehh 2.0: a reimplementation of the R package rehh to detect positive selection from haplotype structure. Mol Ecol Resour. 2017;17:78–90.
45. Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics. 2012;28:i333–9.
46. Cingolani P, Platts A, le Wang L, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly. 2012;6:80–92.
47. Clarkson CS, Miles A, Harding NJ, O'Reilly AO, Weetman D, Kwiatkowski D, et al. The genetic architecture of target-site resistance to pyrethroid insecticides in the African malaria vectors *Anopheles gambiae* and *Anopheles coluzzii*. Mol Eco. 2021;30:5303–17.
48. Rubinoff D, Holland BS. Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. Syst Biol. 2005;54:952–61.
49. Campos M, Phelan J, Spadar A, Collins E, Gonçalves A, Pelloquin B, et al. High-throughput barcoding method for the genetic surveillance of insecticide resistance and species identification in *Anopheles gambiae* complex malaria vectors. Sci Rep. 2022;12:13893.
50. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol Mar Biol Biotechnol. 1994;3:294–9.
51. Kaufmann C, Briegel H. Flight performance of the malaria vectors *Anopheles gambiae* and *Anopheles atroparvus*. J Vector Ecol. 2004;29:140–53.
52. Marsden CD, Cornel A, Lee Y, Sanford MR, Norris LC, Goodell PB, et al. An analysis of two island groups as potential sites for trials of transgenic mosquitoes for malaria control. Evol Appl. 2013;6:706–20.
53. Scott JG. Cytochromes P450 and insecticide resistance. Insect Biochem Mol Biol. 1999;29:757–77.
54. Vontas J, Katsavou E, Mavridis K. Cytochrome P450-based metabolic insecticide resistance in *Anopheles* and *Aedes* mosquito vectors: muddying the waters. Pestic Biochem Physiol. 2020;170:104666.

55. Hamid-Adiamoh M, Jabang AMJ, Opondo KO, Ndiath MO, Assogba BS, Amambua-Ngwa A. Distribution of *Anopheles gambiae* thioester-containing protein 1 alleles along malaria transmission gradients in The Gambia. Malar J. 2023;22:89.
56. Edi CV, Djogbénou L, Jenkins AM, Regna K, Muskavitch MAT, Poupardin R, et al. CYP6 P450 enzymes and ACE-1 duplication produce extreme and multiple insecticide resistance in the malaria mosquito *Anopheles gambiae*. PLoS Genet. 2014;10:e1004236.
57. Müller P, Warr E, Stevenson BJ, Pignatelli PM, Morgan JC, Steven A, et al. Field-caught permethrin-resistant *Anopheles gambiae* overexpress CYP6P3, a P450 that metabolises pyrethroids. PLoS Genet. 2008;4:e1000286.
58. Grigoraki L, Lagnel J, Kioulos I, Kampouraki A, Morou E, Labbé P, et al. Transcriptome profiling and genetic study reveal amplified carboxylesterase genes implicated in temephos resistance, in the Asian tiger mosquito *Aedes albopictus*. PLoS Negl Trop Dis. 2015;9:e0003771.
59. Schmidt JM, Good RT, Appleton B, Sherrard J, Raymant GC, Bogwitz MR, et al. Copy number variation and transposable elements feature in recent, ongoing adaptation at the Cyp6g1 locus. PLoS Gen. 2010;6:e1000998.
60. Itokawa K, Komagata O, Kasai S, Masada M, Tomita T. *Cis*-acting mutation and duplication: history of molecular evolution in a P450 haplotype responsible for insecticide resistance in *Culex quinquefasciatus*. Insect Biochem Mol Biol. 2011;41:503–12.
61. Williams J, Cowlishaw R, Sanou A, Ranson H, Grigoraki L. In vivo functional validation of the V402L voltage gated sodium channel mutation in the malaria vector *Anopheles gambiae*. Pest Manag Sci. 2022;8:1155–63.
62. Mitchell SN, Rigden DJ, Dowd AJ, Lu F, Wilding CS, Weetman D, et al. Metabolic and target-site mechanisms combine to confer strong DDT resistance in *Anopheles gambiae*. PLoS ONE. 2014;9:e92662.
63. Reddy MR, Overgaard HJ, Abaga S, Reddy VP, Caccone A, Kiszewski AE, et al. Outdoor host seeking behaviour of *Anopheles gambiae* mosquitoes following initiation of malaria vector control on Bioko Island, Equatorial Guinea. Malar J. 2011;10:184.

Publisher's Note

Supplementary Data

Supplementary Table 1: Single Nucleotide Polymorphisms (SNPs) with $F_{ST} \geq 0.9$ for chromosome 2L: comparison between An. melas from the Bijagós Archipelago with An. melas from Cameroon.

Chromosome	Window (Genomic bps)	F_{ST} Value	Gene ID	Protein coding gene
2L	2897239-2898238	0.905	AGAP029113	protein coding gene - unspecified product
	4429239-4430238	0.937	AGAP004847	Class B Scavenger Receptor (CD36 domain)
	4956239-4957238	0.937	AGAP004886	protein coding gene - unspecified product
	6891239-6892238	0.937	AGAP029910	protein coding gene - unspecified product
	7332239-7333238	0.937	AGAP004970	protein coding gene - unspecified product
			AGAP004971	protein coding gene - unspecified product
	7838239-7839238, 7873239-7874238	0.905	AGAP005008	protein coding gene - unspecified product
	9956239-9957238	0.902	AGAP005093	protein coding gene - unspecified product
	10159239-10160238	0.937	AGAP005106	protein coding gene - unspecified product
	10335239-10336238	0.937	AGAP005127	RNA-binding protein 15
	12182239-12183238	0.937	AGAP005197	E2 ubiquitin-conjugating enzyme, bendless
	12220239-12221238	0.937	AGAP005205	peptidoglycan recognition protein (long)
	12448239-12449238	0.937	AGAP005219	protein coding gene - unspecified product
	12672239-12673238	0.905	AGAP005235	protein coding gene - unspecified product
	12999239-13000238	0.905	AGAP005246	serine protease inhibitor (serpin) 10
	13330239-13331238	0.937	AGAP029530	protein coding gene - unspecified product
	13509239-13510238	0.937	AGAP005273	LRRCT domain-containing protein
	13842239-13843238	0.906	AGAP005300	protein coding gene - unspecified product
	14259239-14260238	0.937	AGAP005339	chitinase
	14448239-14449238	0.937	AGAP005351	DEAD box polypeptide 5
	14643239-14644238	0.937	AGAP005359	F-box and WD-40 domain protein 7
	14682239-14683238	0.937	AGAP005368	protein coding gene - unspecified product
	16464239-16465238	0.937	AGAP005508	protein coding gene - unspecified product
	16692239-16693238	0.905	AGAP005534	protein coding gene - unspecified product
	16962239-16963238	0.937	AGAP005549	Pre-mRNA-splicing factor rse1
	16992239-16993238	0.937	AGAP005553	cellular retinaldehyde binding protein
	16997239-16998238	0.937	AGAP005554	protein coding gene - unspecified product
	20056239-20057238	0.937	AGAP029489	protein coding gene - unspecified product
	20304239-20305238	0.905	AGAP005759	DUF2236 domain-containing protein
	22662239-22663238	0.937	AGAP005866	glutamate decarboxylase
	24444239-24445238	0.905	AGAP005978	zinc finger and SCAN domain containing 2

			AGAP005979	protein coding gene - unspecified product
	25392239-25393238	0.937	AGAP006028	GABA-gated chloride channel subunit
	25525239-25526238	0.937	AGAP006038	serine/arginine repetitive matrix protein 2
	25592239-25593238	0.937	AGAP006045	protein yorkie
	25970239-25971238	0.905	AGAP006063	protein coding gene - unspecified product
	26504239-26505238	0.905	AGAP006089	protein coding gene - unspecified product
	30252239-30253238	0.905	AGAP006347	potassium voltage-gated channel KQT-like subfamily, invertebrate
	30424239-30425238	0.937	AGAP006350	defective proboscis extension response 11
	32692239-32693238	0.902	AGAP006475	protein coding gene - unspecified product
	35674239-35675238	0.937	AGAP006652	ubiquitin carboxyl-terminal hydrolase 10
	37076239-37077238	0.937	AGAP006702	protein coding gene - unspecified product
	38676239-38677238	0.937	AGAP006805	Lariat debranching enzyme
	40223239-40224238	0.905	AGAP006954	CLIP-domain serine protease
	45424239-45425238	0.937	AGAP007327	protein coding gene - unspecified product
	45781239-45782238	0.905	AGAP028466	protein coding gene - unspecified product
	46057239-46058238	0.905	AGAP007361	translocation protein SEC63
	46122239-46123238	0.937	AGAP007369	Ras-related protein Rap-1b precursor
	46467239-46468238	0.905	AGAP007416	MH2 domain-containing protein
	46728239-46729238	0.937	AGAP007465	protein coding gene - unspecified product
	46949239-46950238	0.937	AGAP007500	Lachesin
	47187239-47188238	0.921	AGAP007523	myosin heavy chain
	47638239-47639238	0.905	AGAP007563	protein coding gene - unspecified product
	47756239-47757238	0.937	AGAP007567	membrane bound O-acyltransferase domain containing 2a
	47786239-47787238	0.937	AGAP007573	RAD54-like protein 2
	48026239-48027238	0.937	AGAP007587	protein coding gene - unspecified product
	48549239-48550238	0.937	AGAP007640	protein coding gene - unspecified product
	48883239-48884238	0.937	AGAP007667	sugar transporter ERD6-like 4
	49188239-49189238	0.937	AGAP007712	A-kinase anchor protein 13

Supplementary Table 2: Single Nucleotide Polymorphisms (SNPs) with $F_{ST} \geq 0.9$ for chromosome 2R: comparison between An. melas from the Bijagós Archipelago with An. melas from Cameroon.

Chromosome	Window (Genomic bps)	F_{ST} Value	Gene ID	Protein coding gene
2R	644736-645735	0.905	AGAP001157	UBR-type domain-containing protein
	748736-749735	0.905	AGAP001179	nuclear pore complex protein Nup88
	2467736-2468735	0.905	AGAP012950	protein coding gene - unspecified product
	5358736-5359735	0.905	AGAP001472	SAP domain-containing protein
	7212736-7213735	0.937	AGAP001635	sodium-coupled monocarboxylate transporter 2
	8460736-8461735	0.937	AGAP001690	protein coding gene - unspecified product
	9205736-9206735	0.937	AGAP001743	estrogen-related receptor ERR
	10340736-10341735	0.905	AGAP001786	protein coding gene - unspecified product
	11033736-11034735	0.921	AGAP001824	protein coding gene - unspecified product
	11482736-11483735	0.937	AGAP029451	ATP-binding cassette transporter (ABC transporter) family G member 3
	11983736-11984735	0.905	AGAP001905	zinc finger RNA-binding protein
	14010736-14011735	0.937	AGAP002019	COMPASS component SWD3
	16660736-16661735	0.937	AGAP029491	protein coding gene - unspecified product
	16871736-16872735	0.937	AGAP002172	POU domain transcription factor, class 4
	18039736-18040735	0.937	AGAP002238	protein coding gene - unspecified product
	18263736-18264735	0.906	AGAP002261	Ras GTPase-activating protein 3
	19415736-19416735	0.937	AGAP002315	autophagy related gene
	19950736-19951735	0.937	AGAP002329	carnitine O-octanoyltransferase
	20116736-20117735	0.905	AGAP002336	Ig-like domain-containing protein
	22059736-22060735	0.937	AGAP002503	4-coumarate:CoA ligase
			AGAP002502	translation initiation factor 4G
	22085736-22086735	0.937	AGAP002505	voltage-dependent calcium channel alpha 1, invertebrate
	23047736-23048735	0.905	AGAP002573	GTPase-activating Rap/Ran-GAP domain-like protein 3
	23135736-23136735	0.905	AGAP002577	voltage-dependent p/q type calcium channel
	23177736-23178735	0.937	AGAP002578	protein coding gene - unspecified product
	23276736-23277735	0.903	AGAP002581	protein coding gene - unspecified product
	23291736-23292735	0.937	AGAP002586	glycogen(starch) synthase
	24200736-24201735	0.937	AGAP012958	IRF-2BP1_2 domain-containing protein
	25093736-25094735	0.905	AGAP002649	protein coding gene - unspecified product
	25186736-25187735	0.937	AGAP002654	protein coding gene - unspecified product
	25886736-25887735	0.937	AGAP002707	defective proboscis extension response 18
	28401736-28402735	0.905	AGAP002859	solute carrier family 8 (sodium/calcium exchanger)
	29658736-29659735	0.937	AGAP002924	protein tweety homolog

	29838736-29839735	0.937	AGAP002935	Eukaryotic translation initiation factor 3 subunit F
	32937736-32938735	0.937	AGAP003124	dihydropyrimidinase
	33575736-33576735	0.905	AGAP013292	Rho-GAP domain-containing protein
	33590736-33591735	0.905	AGAP003180	protein coding gene - unspecified product
			AGAP029060	protein coding gene - unspecified product
	34787736-34788735	0.905	AGAP003276	splicing factor, arginine/serine-rich 17
	34917736-34918735	0.937	AGAP003283	atrial natriuretic peptide receptor A
	35926736-35927735	0.937	AGAP003318	protein coding gene - unspecified product
	36885736-36886735	0.905	AGAP003366	myotubularin-related protein 13
	37843736-37844735	0.937	AGAP003445	tuberous sclerosis 1
	37869736-37870735	0.937	AGAP003449	rootletin
	37966736-37967735	0.937	AGAP003463	protein coding gene - unspecified product
	39069736-39070735	0.937	AGAP003516	protein coding gene - unspecified product
	40460736-40461735	0.905	AGAP003604	protein coding gene - unspecified product
	40512736-40513735	0.937	AGAP003613	palmitoyltransferase ZDHHC3
	44341736-44342735	0.937	AGAP003845	bromodomain-containing protein 8
	45584736-45585735	0.905	AGAP003887	Knickkopf
	46351736-46352735	0.905	AGAP003925	kinesin family member 1/13/14
	47409736-47410735	0.905	AGAP003976	protein coding gene - unspecified product
	47718736-47719735	0.937	AGAP003997	casein kinase 1, gamma
	48605736-48606735, 48610736-48611735	0.905	AGAP004046	Mpv17-like protein
	49938736-49939735	0.905	AGAP004106	protein coding gene - unspecified product
			AGAP004107	protein coding gene - unspecified product
	50126736-50127735	0.937	AGAP004112	protein kinase domain-containing protein
	51445736-51446735	0.937	AGAP004184	protein coding gene - unspecified product
	51641736-51642735	0.937	AGAP004192	heat shock 70kDa protein 5
	51964736-51965735	0.905	AGAP004215	protein coding gene - unspecified product
	53248736-53249735	0.905	AGAP004249	Methyltransf_11 domain-containing protein
	54363736-54364735	0.905	AGAP004310	protein coding gene - unspecified product
	54375736-54376735	0.937	AGAP004311	phosphopantothenate-cysteine ligase
	55236736-55237735	0.937	AGAP004369	protein coding gene - unspecified product
	55696736-55697735	0.905	AGAP004402	solute carrier family 17 member 7
	56153736-56154735, 56178736 -56179735	0.919	AGAP004443	glycogen synthase kinase 3 beta
	56563736-56564735	0.937	AGAP004453	GPCR Dopamine Family 3
	56740736-56741735	0.937	AGAP013052	protein coding gene - unspecified product
			AGAP004474	protein coding gene - unspecified product
	56805736-56806735	0.905	AGAP004483	Lethal giant larvae

Supplementary Table 3: Single Nucleotide Polymorphisms (SNPs) with $F^{ST} \geq 0.9$ for chromosome 3L: comparison between An. melas from the Bijagós Archipelago with An. melas from Cameroon.

Chromosome	Window (Genomic bps)	F_{ST} Value	Gene ID	Protein coding gene
3L	2813051-2814050	0.905	AGAP010410	protein coding gene - unspecified product
	3062051-3063050, 3064051-3065050	0.937	AGAP010422	Limbic system-associated membrane protein
	3316051-3317050	0.937	AGAP010435	protein coding gene - unspecified product
	3893051-3894050	0.916	AGAP010456	protein coding gene - unspecified product
	5011051-5012050	0.937	AGAP010490	rabconnectin
	6257051-6258050, 6259051-6260050	0.937	AGAP010551	STE20-like kinase
	7928051-7929050	0.921	AGAP010642	WH2 domain-containing protein
	7931051-7932050	0.937	AGAP010643	protein coding gene - unspecified product
	8613051-8614050	0.937	AGAP028154	V-type proton ATPase proteolipid subunit
	10390051-10391050	0.905	AGAP029197	Fatty acyl-CoA reductase
	12417051-12418050	0.937	AGAP010883	kinesin-like protein costa
	12418051-12419050	0.937	AGAP010884	protein coding gene - unspecified product
	12759051-12760050	0.937	AGAP028617	Sortilin-related receptor
	12896051-12897050	0.937	AGAP010910	phosphatidylinositol glycan, class Z
	13496051-13497050	0.905	AGAP010939	autophagy related gene
	16550051-16551050	0.905	AGAP011094	protein coding gene - unspecified product
	17204051-17205050	0.937	AGAP011113	potassium voltage-gated channel KQT-like subfamily member 1
	20503051-20504050	0.905	AGAP011279	solute carrier family 6 (neurotransmitter transporter, amino acid/orphan)
	21595051-21596050	0.937	AGAP011349	GABA-gated chloride channel
	21754051-21755050	0.937	AGAP011355	protein coding gene - unspecified product
	23138051-23139050	0.905	AGAP011384	protein coding gene - unspecified product
	24403051-24404050	0.902	AGAP011421	IQ motif and SEC7 domain-containing protein
	24589051-24590050	0.905	AGAP011433	protein coding gene - unspecified product
	24760051-24761050	0.905	AGAP011446	far upstream element-binding protein
	25469051-25470050, 25660051-25661050	0.937	AGAP029048	Electroneutral potassium-chloride cotransporter 1
	27994051-27995050	0.901	AGAP011573	protein coding gene - unspecified product
	29931051-29932050	0.937	AGAP011618	DOMON domain-containing protein
	31421051-31422050	0.905	AGAP029516	protein coding gene - unspecified product
	31497051-31498050	0.937	AGAP011695	homeobox protein SIX1
	32456051-32457050	0.937	AGAP029606	protein coding gene - unspecified product
	33225051-33226050	0.937	AGAP011779	TAK1-associated binding protein 2
	33453051-33454050	0.937	AGAP011809	protein coding gene - unspecified product

	33951051-33952050	0.906	AGAP029637	protein coding gene - unspecified product
	34665051-34666050	0.937	AGAP011893	Potentail helicase MOV-10
	35229051-35230050	0.937	AGAP011929	protein coding gene - unspecified product
	36011051-36012050	0.937	AGAP029590	protein coding gene - unspecified product
	36050051-36051050	0.937	AGAP012000	fibrinogen and fibronectin
	36522051-36523050	0.905	AGAP012023	DBF4-type domain-containing protein
	36683051-36684050	0.916	AGAP028133	protein coding gene - unspecified product
	37106051-37107050	0.937	AGAP029616	RRM domain-containing protein
	37207051-37208050	0.937	AGAP029469	Non-specific serine/threonine protein kinase
	37403051-37404050	0.905	AGAP029447	protein coding gene - unspecified product
	37463051-37464050	0.902	AGAP012079	Miranda
	37486051-37487050	0.905	AGAP029653	protein coding gene - unspecified product
	37671051-37672050	0.937	AGAP012103	Insulin-like growth factor 2 mRNA-binding protein 2
	37766051-37767050	0.937	AGAP012113	ubiquitin carboxyl-terminal hydrolase L3
	37857051-37858050	0.937	AGAP012123	F-box and leucine-rich repeat protein 16
	38092051-38093050	0.937	AGAP012154	solute carrier family 15 member 1
	38570051-38571050	0.937	AGAP012193	ubiquitin-conjugating enzyme E2 F
	38574051-38575050	0.905	AGAP012194	DnaJ (Hsp40) homolog, subfamily B
	39339051-39340050	0.937	AGAP012252	classical protein kinase C

Supplementary Table 4: Single Nucleotide Polymorphisms (SNPs) with $F^{ST} \geq 0.9$ for chromosome 3R: comparison between An. melas from the Bijagós Archipelago with An. melas from Cameroon.

Chromosome	Window (Genomic bps)	F_{ST} Value	Gene ID	Protein coding gene
3R	528366-529365	0.937	AGAP007771	Nucleic_acid_bd domain-containing protein
	1245366-1246365, 1309366-1310365	0.937	AGAP007803	protein coding gene - unspecified product
	1543366-1544365	0.937	AGAP007836	sodium-coupled monocarboxylate transporter 2
	1553366-1554365	0.921	AGAP007838	Sodium-coupled monocarboxylate transporter 2
	2085366-2086365	0.905	AGAP007849	laminin, alpha 1/2
	2420366-2421365	0.937	AGAP007877	protein coding gene - unspecified product
	2585366-2586365	0.937	AGAP007901	Ras-related protein Rab-5C
	2914366-2915365	0.905	AGAP007925	Ralgapb
	2939366-2940365	0.905	AGAP007928	Turtle protein, isoform
	3296366-3297365	0.905	AGAP007955	protein coding gene - unspecified product
	5111366-5112365	0.937	AGAP008077	protein coding gene - unspecified product
	8799366-8800365	0.905	AGAP008304	3',5'-cyclic-nucleotide phosphodiesterase
	9612366-9613365	0.937	AGAP008347	Putative glycoprotein hormone rk-like receptor
	10075366-10076365	0.937	AGAP008364	thioester-containing protein 15 (TEP15)
	10214366-10215365	0.905	AGAP008384	protein coding gene - unspecified product
	10555366-10556365	0.937	AGAP008418	Rab-GAP TBC domain-containing protein
	13746366-13747365	0.937	AGAP008626	Androgen-induced 1
	13834366-13835365	0.906	AGAP008640	protein coding gene - unspecified product
	13961366-13962365	0.905	AGAP008646	cGMP-specific 3',5'-cyclic phosphodiesterase
	15993366-15994365	0.937	AGAP008717	hydroxymethylglutaryl-CoA lyase
	18219366-18220365	0.937	AGAP008814	transcription factor 4/12
	19492366-19493365	0.937	AGAP028027	protein coding gene - unspecified product
	20311366-20312365	0.905	AGAP008889	ATP-binding cassette transporter (ABC transporter) family G member 1
	20661366-20662365	0.937	AGAP008912	Rac GTPase-activating protein 1
	22484366-22485365	0.905	AGAP008979	protein coding gene - unspecified product
	25851366-25852365	0.937	AGAP009112	TPR_REGION domain-containing protein
	27722366-27723365	0.937	AGAP009158	protein coding gene - unspecified product
	31271366-31272365	0.937	AGAP009325	protein coding gene - unspecified product
	31800366-31801365	0.937	AGAP009364	protein coding gene - unspecified product
	32994366-32995365	0.905	AGAP009424	Calponin-homology (CH) domain-containing protein
	33476366-33477365	0.937	AGAP009432	Daxx-like protein

	34725366-34726365	0.925	AGAP009493	nicotinic acetylcholine receptor subunit alpha 9
	35330366-35331365	0.937	AGAP009522	protein coding gene - unspecified product
	35605366-35606365	0.916	AGAP009537	cytochrome c
	36058366-36059365	0.937	AGAP029897	protein coding gene - unspecified product
	36546366-36547365	0.905	AGAP009580	protein coding gene - unspecified product
	37167366-37168365	0.905	AGAP009625	protein coding gene - unspecified product
	37243366-37244365	0.921	AGAP009626	Kv channel-interacting protein 4 isoform 2
	37472366-37473365	0.937	AGAP009641	kelch-like protein diablo
	37663366-37664365	0.905	AGAP029517	protein coding gene - unspecified product
	41751366-41752365	0.937	AGAP029589	protein coding gene - unspecified product
	41865366-41866365	0.937	AGAP029558	Protein Wnt
	42190366-42191365	0.937	AGAP028645	Aryl hydrocarbon receptor nuclear translocator
	43190366-43191365, 43201366-43202365	0.937	AGAP009777	mothers against decapentaplegic homolog 2/3
	43366366-43367365	0.905	AGAP009793	protein coding gene - unspecified product
	44349366-44350365	0.937	AGAP009853	gustatory receptor 5
	44516366-44517365	0.937	AGAP009860	protein coding gene - unspecified product
	44741366-44742365	0.937	AGAP029461	protein coding gene - unspecified product
	45275366-45276365	0.937	AGAP029826	Tox-SGS domain-containing protein
	45444366-45445365	0.905	AGAP009932	protein coding gene - unspecified product
	46152366-46153365	0.905	AGAP029448	protein coding gene - unspecified product
	47386366-47387365	0.905	AGAP009995	anoctamin 10
	47594366-47595365	0.905	AGAP010012	Leucine-rich repeat and calponin-like proteiny domain-containing protein 3
	47851366-47852365	0.905	AGAP029234	protein coding gene - unspecified product
	49308366-49309365	0.937	AGAP010135	NCK adaptor protein
	50375366-50376365	0.903	AGAP010189	Arf-GAP domain-containing protein

Supplementary Table 5: Single Nucleotide Polymorphisms (SNPs) with $F^{ST} \geq 0.9$ for chromosome X: comparison between An. melas from the Bijagós Archipelago with An. melas from Cameroon.

Chromosome	Window (Genomic bps)	F_{ST} Value	Gene ID	Protein coding gene
X	211895-212894	0.905	AGAP000017	protein coding gene - unspecified product
	447895-448894	0.937	AGAP000035	protein coding gene - unspecified product
	708895-709894	0.937	AGAP000045	G_PROTEIN_RECEP_F1_2 domain-containing protein
	1470895-1471894	0.905	AGAP029109	protein coding gene - unspecified product
	1744895-1745894	0.905	AGAP000104	protein coding gene - unspecified product
	1835895-1836894	0.937	AGAP000112	NAD+ synthase (glutamine-hydrolysing)
	2011895-2012894	0.937	AGAP000123	C-type lectin (CTL)
	2052895-2053894	0.905	AGAP013373	protein coding gene - unspecified product
	2185895-2186894	0.905	AGAP000143	beta-mannosidase
	2553895-2555894	0.937	AGAP013158	protein coding gene - unspecified product
	2804895-2805894	0.937	AGAP000159	Kinesin-like protein
	2828895-2829894	0.937	AGAP000161	Adaptor-related protein complex 3, delta 1 subunit
	2910895-2911894	0.937	AGAP000171	E3 ubiquitin-protein ligase
	3421895-3422894	0.905	AGAP000205	protein coding gene - unspecified product
	4174895-4175894	0.905	AGAP000223	protein coding gene - unspecified product
	4231895-4232894	0.937	AGAP000225	protein coding gene - unspecified product
	4373895-4374894	0.937	AGAP000232	regulator of G-protein signaling
	4414895-4415894, 4420895-4421894	0.937	AGAP000235	Thymosin
	4436895-4437894	0.905	AGAP000236	3',5'-cyclic-nucleotide phosphodiesterase
	4555895-4556894	0.937	AGAP000243	Rhophilin-2
	4937895-4938894	0.905	AGAP000264	protein coding gene - unspecified product
	5318895-5319894	0.905	AGAP000297	protein coding gene - unspecified product
	5808895-5809894	0.937	AGAP000331	low density lipoprotein-related protein 2
	6273895-6274894	0.937	AGAP000351	neuropeptide Y receptor 1
	6988895-6989894	0.905	AGAP000383	protein coding gene - unspecified product
	7196895-7197894	0.905	AGAP000393	Trimeric intracellular cation channel type B
	7206895-7207894	0.937	AGAP013009	protein coding gene - unspecified product
	7250895-7251894	0.937	AGAP000399	squid
	7338895-7339894	0.937	AGAP000410	histone deacetylase 4/5
	7560895-7561894	0.905	AGAP000421	protein coding gene - unspecified product
	7881895-7882894	0.937	AGAP000449	protein coding gene - unspecified product
	8295895-8296894, 8335895-8336894	0.937	AGAP000477	serine/threonine kinase 32

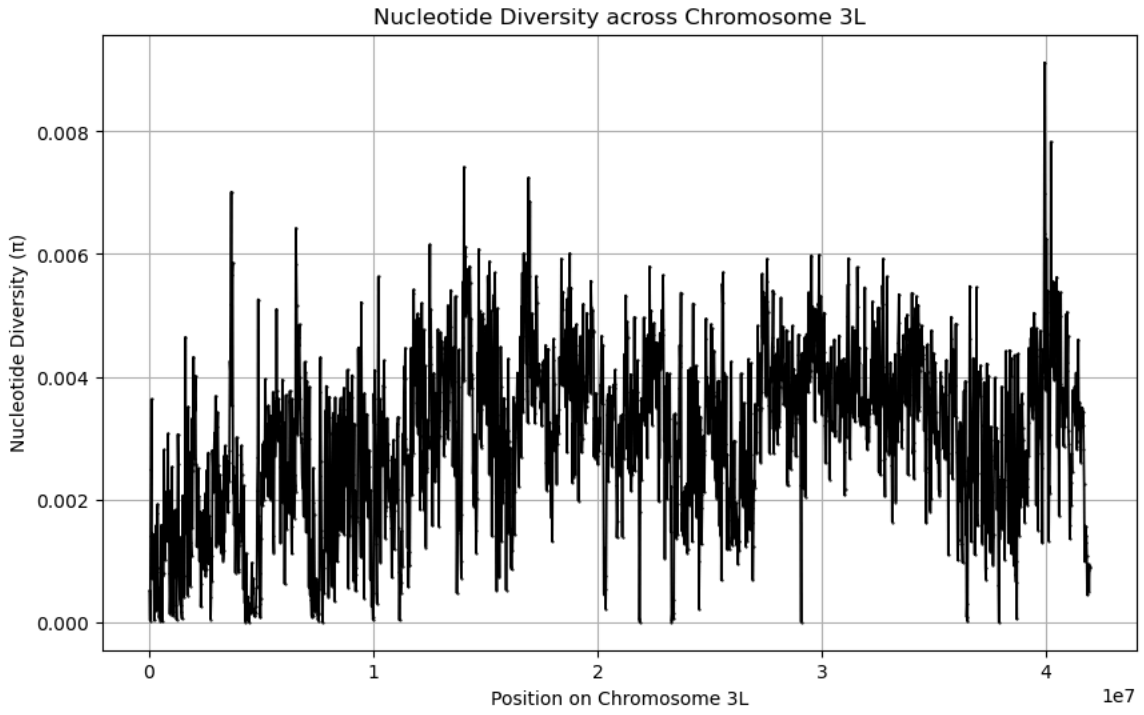
	8696895-8697894	0.937	AGAP029670	protein coding gene - unspecified product
	9252895-9253894	0.937	AGAP000519	diacylglycerol kinase (ATP dependent)
	9414895-9415894, 9430895-9431894	0.905	AGAP000522	Ig-like domain-containing protein
	9717895-9718894	0.937	AGAP013055	protein coding gene - unspecified product
	10571895-10582894	0.905	AGAP000591	decaprenyl-diphosphate synthase subunit 1
	10663895-10664894	0.919	AGAP000596	lysophosphatidylcholine acyltransferase / lyso-PAF acetyltransferase
	10692895-10693894	0.937	AGAP000601	LRRNT domain-containing protein
	11035895-11036894, 11117895-11118894	0.905	AGAP000606	putative GPCR class a orphan receptor 19
	11236895-11237894	0.905	AGAP000620	protein coding gene - unspecified product
	11788895-11789894	0.937	AGAP000663	protein coding gene - unspecified product
	12170895-12171894	0.905	AGAP000685	protein coding gene - unspecified product
	14279895-14280894	0.912	AGAP000785	Synaptic vesicle protein
	14327895-14328894	0.937	AGAP000787	protein coding gene - unspecified product
	15416895-15417894	0.905	AGAP000834	protein coding gene - unspecified product
	16030895-16031894	0.905	AGAP000863	Lachesin
	16434895-16435894	0.937	AGAP000874	protein coding gene - unspecified product
	17022895-17023894	0.937	AGAP000900	Protein-tyrosine sulfotransferase
	17087895-17088894	0.905	AGAP029488	protein coding gene - unspecified product
	17455895-17456894	0.926	AGAP000926	inositol hexakisphosphate/diphosphoinositol-pentakisphosphate kinase
	17615895-17616894	0.905	AGAP000932	protein coding gene - unspecified product
	17823895-17824894	0.937	AGAP029105	protein coding gene - unspecified product
	18750895-18751894	0.937	AGAP000974	protein coding gene - unspecified product
	20011895-20012894	0.937	AGAP001039	cytochrome P450 – CYP307A1

Supplementary Table 6: Single Nucleotide Polymorphisms (SNPs) with $F_{ST} \geq 0.6$ for the mitochondrial genome: comparison between An. melas from the Bijagós Archipelago with An. melas from Cameroon.

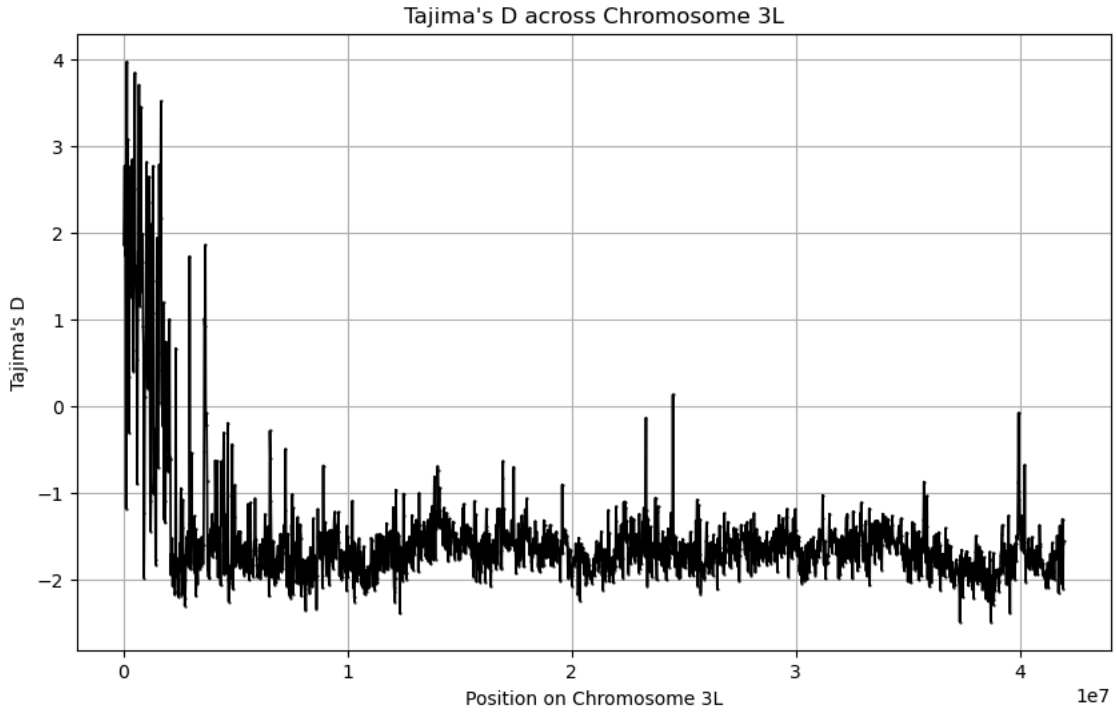
Chromosome	Window (Genomic bps)	F_{ST} Value	Gene ID	Protein coding gene
Mt	1172-2171	0.661	AGAP028360	NADH dehydrogenase subunit 2
			AGAP028361	Protein coding gene - unspecified product
			AGAP028362	Protein coding gene - unspecified product
			AGAP028363	Protein coding gene - unspecified product
			AGAP028364	cytochrome c oxidase subunit I
	5172-6171	0.750	AGAP028371	cytochrome c oxidase subunit III
			AGAP028372	tRNA-Gly for anticodon UCC
			AGAP028373	NADH dehydrogenase subunit 3
			AGAP028374	Protein coding gene - unspecified product
			AGAP028375	Protein coding gene - unspecified product
			AGAP028376	Protein coding gene - unspecified product
			AGAP028377	Protein coding gene - unspecified product
	7172-8171	0.856	AGAP028380	NADH dehydrogenase subunit 5
			AGAP028381	Protein coding gene - unspecified product
			AGAP028382	NADH dehydrogenase subunit 4

Supplementary Table 7: Non-synonymous SNPs associated/putatively associated with insecticide resistance which were investigated in the An. melas population

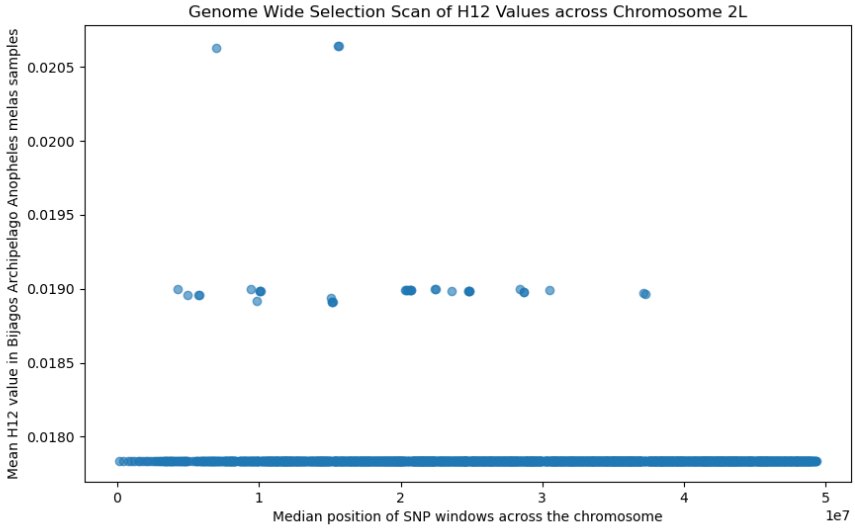
Gene	Chromosome	Average depth of coverage	Position	SNP	Frequency of SNP in *Anopheles melas*
vgsc	2L	24.98	2390177	R254K	0%
			2391228	V402L	0%
			2399997	D466H	0%
			2400071	**M490I**	**2%**
			2402466	G531V	0%
			2407967	Q697P	0%
			2416980	T791M	0%
			2422651	L995S	0%
			2422652	L995F	0%
			2429556	V1507I	0%
			2429617	I1527T	0%
			2429745	N1570Y	0%
			2429897	E1597G	0%
			2429915	K1603T	0%
			2430424	A1746S	0%
			2430817	V1853I	0%
			2430863	I1868T	0%
			2430880	P1874S	0%
			2430881	P1874L	0%
			2431061	A1934V	0%
			2431079	I1940T	0%
gste2	3R	22.21	28598166	I114T	0%
			28598057	F120L	0%
			28598062	L119V	0%
rdl	2L	19.70	25429236	A296G	0%
			25429235	A296S	0%
ace1	2R	21.45	3492074	G280S*	0%



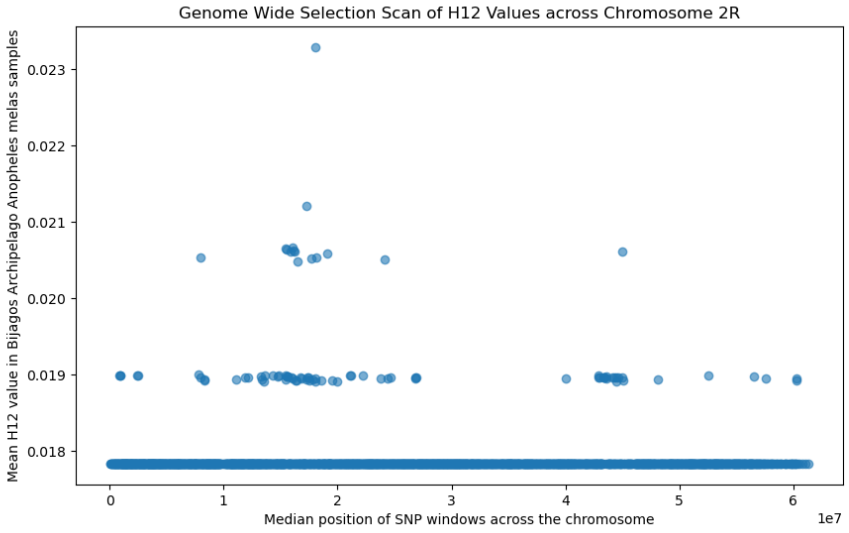
Supplementary Figure 1: Nucleotide diversity across chromosome 3L, Bijagós An. melas mosquitoes, N = 30



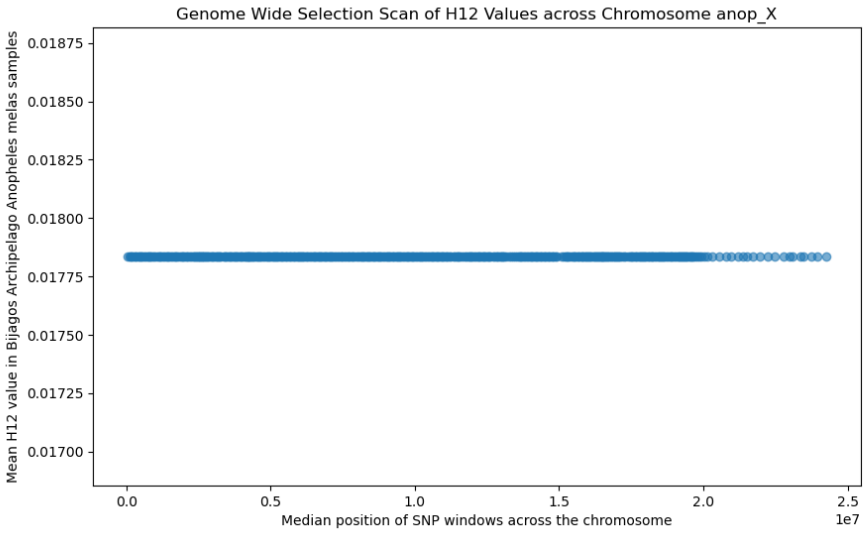
Supplementary Figure 2: Tajima's D computed across chromosome 3L, Bijagós An. melas mosquitoes, N = 30



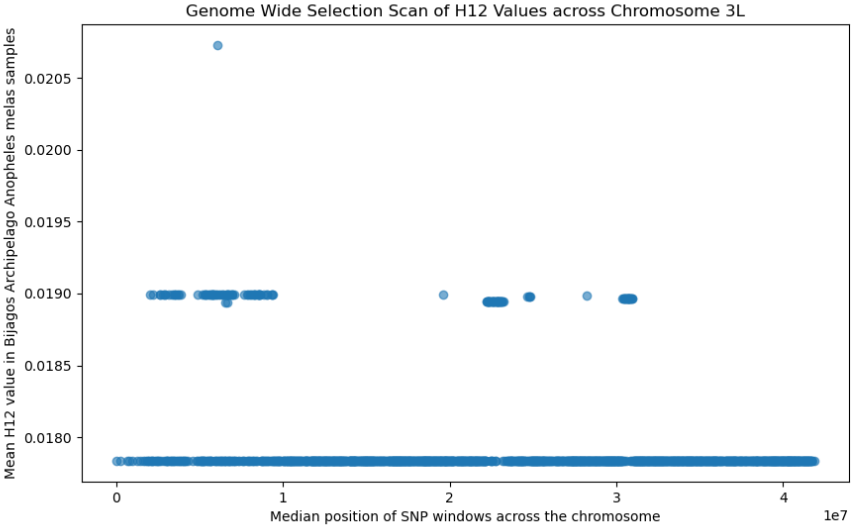
Supplementary Figure 3: Genome wide selection scan of H12 values across Chromosome 2L, Bijagós An. melas, N=30



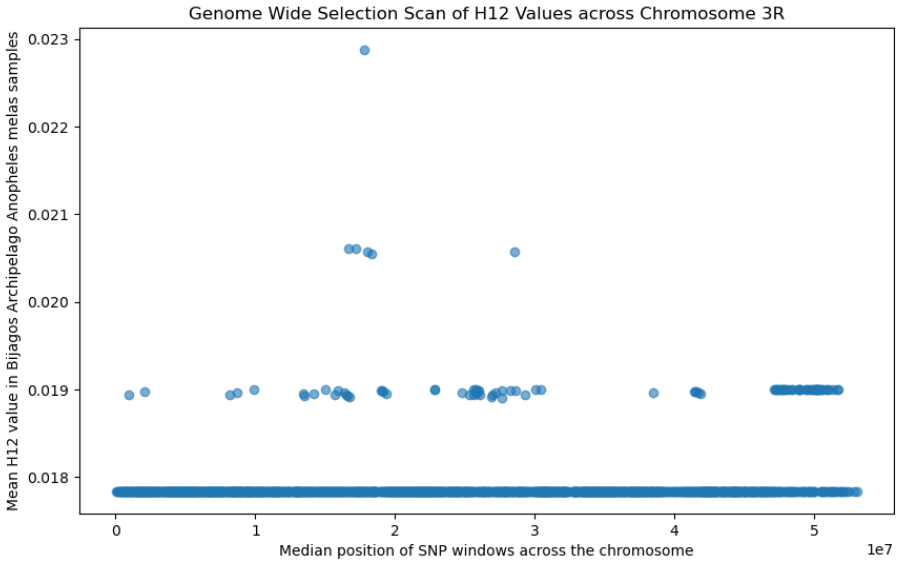
Supplementary Figure 4: Genome wide selection scan of H12 values across Chromosome 2R, Bijagós An. melas, N=30



Supplementary Figure 5: Genome wide selection scan of H12 values across Chromosome X, Bijagós An. melas, N=30



Supplementary Figure 6: Genome wide selection scan of H12 values across Chromosome 3L, Bijagós An. melas, N=30



Supplementary Figure 7: Genome wide selection scan of H12 values across Chromosome 2R, Bijagós An. melas, N=30

CHAPTER SEVEN

Discussion

Malaria control on the Bijagós Archipelago of Guinea-Bissau is currently comprised of a combination of tools targeting the *Plasmodium* parasite and the *Anopheles* vectors. These include pyrethroid-treated ITNs, IPTp with SP, and case detection and treatment with AL[1]. Despite good coverage and use of these interventions, malaria transmission persists on the Bijagós, and innovative malaria control approaches are required to reach elimination[2]. As interventions are implemented, an evolutionary arms race ensues between the *Plasmodium* parasite and antimalarials, and between the *Anopheles* vectors and insecticides[3]. Understanding and monitoring the evolution of drug resistance and insecticide resistance is a crucial component in effective malaria control, and resistance data can be used to make evidence-based decisions on which control tools to implement, and where[4,5]. Prior to this thesis, there have been no previous studies of antimalarial resistance in *Plasmodium* parasites on the Bijagós, and previous insecticide resistance studies in *Anopheles* vectors were limited to the main island of Bubaque and to a very limited set of insecticides and associated molecular markers[6]. The aim of this thesis was to harness genome sequencing and surveillance tools to explore the genetic diversity and resistance status of the *Plasmodium* parasite and *Anopheles* vectors on the Archipelago, to better inform future malaria surveillance activities and control interventions in this region. This section provides a brief overview of the findings of this thesis in relation to its objectives, including limitations, lessons learned, and recommendations for future studies. For full details, please see the chapters highlighted below for each objective.

Plasmodium falciparum on the Bijagós Archipelago

Objective 1: To generate and analyse WGS data from DBS samples collected prior to implementation of MDA with DHAP on the Bijagós, to investigate population dynamics and molecular markers of drug resistance in *P. falciparum* infections.

Published Research Paper [7] *- Chapter 3*

The MATAMAL MDA trial using DHAP and adjunctive ivermectin was implemented in 2021 and 2022[2]. This thesis used DBS samples collected during a cross-sectional survey in 2018 prior to the start of the MATAMAL trial to investigate the genetic diversity and drug resistance profile of *P. falciparum* on the Archipelago. High quality WGS data was generated for a small sample of *P. falciparum* infections

from the Bijagós and analysed in a combined dataset with 795 open-source *P. falciparum* samples from across the African continent[8]. Population structure analyses identified that Bijagós *P. falciparum* clustered with samples from mainland West Africa, without forming a separate phylogenetic group. The Bijagós *P. falciparum* samples had 86.7% monoclonality ($F_{WS} \geq 0.95$) and slightly higher Identity-by-Descent (IBD) fractions than *P. falciparum* from the other West African countries analysed. These results are likely to reflect reduced outcrossing within the *P. falciparum* population and reduced likelihood of simultaneous infection with *P. falciparum* of different genotypes, indicative of the isolated nature of the Archipelago and its low to moderate malaria transmission intensity.

I hypothesised that antimalarial resistance mutations associated with resistance to SP, CQ, and AL would be present due to the use of these drugs in the region. In line with my hypothesis, we observed widespread molecular markers of resistance to SP, which are likely to reflect the continued use of IPTp with SP. This included fixation of the SP resistance-associated mutations *pfdhfr* N51I and S108N, 92.9% allelic frequency of C59R, and high allelic frequency of *pfdhps* A437G (60.0%) and S436A (20.0%). 13 of the Bijagós samples carried the triple mutant haplotype *pfdhfr* N51I-C59R-S108N. These allelic frequencies were broadly similar to previously published data from mainland Guinea-Bissau, in a 2014-2016 study by Nag *et al* (2017)[11]. Nag *et al* found a slightly lower frequency of *pfdhfr* N51I (88%), S108N (92.4%) and C59R (87.1%), and a slightly higher frequency of *pfdhps* A437G (66.8%)[11]. In this thesis, we identified the *pfcrt* K76T mutation associated with chloroquine resistance at a frequency of 28.6% on the Archipelago. Prior to the introduction of AL in 2008[9], malaria was routinely treated with triple the standard dose of CQ[10]. Between 2014 and 2016, the K76T mutation was identified at 52-61% prevalence on mainland Bissau using multiplex amplicon sequencing by Nag *et al* (2017)[11] and targeted sequencing of DNA extracted from RDTs by Nag *et al* (2019)[12]. This indicates that the K76T mutation has been maintained in Guinea-Bissau despite chloroquine not being included in malaria treatment policy since 2008. Whereas, the withdrawal of CQ in other settings has resulted in the reversion to wild-type K76 and the recovery of CQ susceptibility[13,14,15]. The maintenance of the K76T allele in Guinea-Bissau may be due to off-label use of CQ, which often involves the sale of low-quality CQ products[16]. The extent of off-label CQ use in the Bijagós or in mainland Guinea-Bissau is not well understood, but is seen widely in other malaria endemic countries[16]. The first-line drug for malaria treatment in the Bijagós is AL[2]. The *Pfmdr1* N86 allele associated with resistance to lumefantrine[17,18] was identified at a frequency of 92.3% in our study, which is similar to the allele frequency identified in mainland Bissau by previous studies between 2014 and 2016 (88-92%)[11,12]. Despite this high frequency, the efficacy of AL for first-line treatment in Guinea-Bissau is still high (94%), as identified in a randomised controlled trial conducted

between 2012 and 2015[19]. The *pfexo* E415G mutation associated with resistance to piperaquine[20] was not identified. Finally, mutations associated with partial-resistance to artemisinin were not identified in the Bijagós *P. falciparum* samples, in line with previous studies conducted on mainland Bissau between 2014 and 2016[11,12].

This research provides the first WGS data from *P. falciparum* on the Archipelago. Furthermore, this research provides a baseline genetic diversity and resistance profile of resistance-associated SNPs prior to the implementation of MDA with DHAP in the region. This work can be built upon in the future, following the conclusion of the MDA trial, to better understand the impact of MDA and other malaria control interventions on antimalarial resistance and population dynamics of *P. falciparum* on the Archipelago. Changes in the frequencies of antimalarial resistance markers following interventions can be monitored over time, providing early warning signals of the emergence of antimalarial resistance, and allowing for the identification of novel genomic markers of resistance.

Anopheles mosquitoes on the Bijagós Archipelago

Objective 2: To use multiplex amplicon sequencing to identify molecular markers associated with insecticide resistance in *An. gambiae s.l.* complex mosquitoes collected during the peak malaria transmission season in October/November 2019.

Published Research Paper [21] *– Chapter 4*

Although WGS was successfully utilised in **Chapter 3** to amplify *P. falciparum* DNA, this technique is expensive and can be cost-prohibitive. To overcome this limitation when analysing *Anopheles gambiae*, I demonstrated the utility of high-throughput multiplex amplicon sequencing to target ~500 bp stretches of the genome. Multiplex barcodes enabled DNA from different mosquito specimens and target regions to be pooled together, necessitating fewer PCR reactions and reducing the cost of sequencing[11,22–24]. This thesis used high-throughput multiplex amplicon sequencing to screen *An. gambiae s.l.* complex mosquitoes from the Bijagós for mutations associated with insecticide resistance. These mosquitoes were collected from islands across the Archipelago during a vector surveillance survey during the peak malaria transmission season in October and November 2019[25]. I hypothesised that insecticide resistance mutations associated with pyrethroid resistance would be widespread in *An. gambiae s.l.* on the Archipelago. Seventeen mutations previously associated with resistance were investigated across four key insecticide resistance genes. Four of these screened mutations were present in the Bijagós population: *vgsc* L995F (12.2% allelic

frequency), N1570Y (6.2%) and A1746S (0.7%) associated with pyrethroid resistance, and *rdl* A269G (1.1%) associated with resistance to dieldrin. An additional eight non-synonymous SNPs were identified which have not been reported previously and may be involved in insecticide resistance. This study was also the first to report the *vgsc* L995F (kdr)-west allele in *An. melas*, which was identified at an allelic frequency of 2.14%. Inter-island comparisons were conducted between islands where more than 20 mosquito specimens were collected. These included the islands of Bubaque, Orangozinho, Soga, Tchedega (Maio), and Uno. The *vgsc* L995F and N1570Y mutations were found at a significantly higher prevalence on Bubaque than on the other islands studied. Bubaque is the most populated island in the Archipelago and the site of the main port connection to mainland Bissau. The higher frequency of mutations associated with pyrethroid resistance on Bubaque island may be due to the presence of more ITNs on this island than the other islands, as it is more populated. This may have resulted in increased selection pressure for pyrethroid resistance in *An. gambiae s.l.* populations on Bubaque compared to elsewhere. In addition, the frequency of insecticide resistance mutations could be impacted by "hitchhiking" mosquitoes travelling between the ports on Bubaque and mainland Bissau, which may drive the spread of resistance mutations from the mainland.

Objective 3: To conduct phenotypic testing of deltamethrin susceptibility in *An. gambiae s.l.* complex mosquitoes on the main island of Bubaque.

Published Research Paper – Chapter 5

The identification of mutations associated with pyrethroid resistance in **Objective 2, Chapter 4,** coupled with the absence of phenotypic testing for deltamethrin resistance prior to this thesis, necessitated conducting insecticide susceptibility testing on the Archipelago. In **Chapter 5**, I reared *An. gambiae s.l.* mosquitoes on Bubaque island and tested them for phenotypic resistance to deltamethrin using WHO tube tests[5]. I hypothesised that deltamethrin resistance would be present on the Archipelago due to the use of pyrethroid ITNs on the islands and the presence of pyrethroid resistance on mainland Guinea-Bissau[26]. In line with my hypothesis, this insecticide susceptibility testing identified a corrected treatment mortality of 45.35% at discriminating concentrations (DC) of deltamethrin (0.05%)[5]. The WHO thresholds for confirming resistance to an insecticide are corrected treatment mortality < 90%, confirming the presence of deltamethrin resistance in this vector population. Subsequent testing with 5x DC concentration was conducted to identify the intensity of deltamethrin resistance. Intensity concentration bioassays resulted in 92.52% mortality, indicating moderate to high level resistance. This was the first phenotypic testing for deltamethrin resistance to

be conducted on the Bijagós Archipelago, providing valuable data to the NMCP for making evidence-based decisions about which vector control tools to implement on these islands. Due to the presence of pyrethroid resistance, the use of pyrethroid-chlorfenapyr ITNs instead of pyrethroid-only ITNs may be beneficial. However, additional phenotypic susceptibility studies may be required to generate the required evidence for this policy change.

Objective 4: To generate and analyse WGS data from deltamethrin susceptible and resistant *An. gambiae s.s.* mosquitoes, to investigate genetic diversity and signatures of selection.

Published Research Paper – Chapter 5

Following the identification of phenotypic deltamethrin resistance in **Objective 3, Chapter 5**, *An. gambiae s.s.* complex mosquitoes identified as deltamethrin resistant and deltamethrin susceptible were processed and their DNA was whole genome sequenced. In contrast to the multiplex amplicon sequencing technique used in **Chapter 4**, whole genome sequencing was used to identify signatures of selection across the genome, and to aim to identify novel markers of deltamethrin resistance. Insecticide resistance mutations were investigated, and six SNPs previously associated with insecticide resistance were identified in the mosquito population. This included T791M, L995F, N1570Y, A1746S and P1874L in the *vgsc* gene, and L119V in the *gste2* gene. No SNPs were found in 100% association with the resistant or susceptible phenotype. However, the *vgsc* N1570Y and L119V *gste2* SNPs associated with pyrethroid resistance were identified only in resistant mosquitoes and not in susceptible mosquitoes. An additional 20 non-synonymous SNPs were identified in insecticide-resistance associated genes which have not been reported previously. Four of these novel SNPs were present at frequencies ≥ 5% in the population: T154S, I126F and G26S in the *vgsc* gene and A65S in the *ace1* gene. Notably, three of the four mutations identified in **Chapter 4** using multiplex amplicon sequencing of *An. gambiae s.l.* mosquitoes were also identified in this *An. gambiae s.s.* Bubaque population using WGS. These were the *vgsc* mutations associated with pyrethroid resistance: L995F, N1570Y, and A1746S. Whereas, the *rdl* A269G mutation associated with dieldrin resistance that was identified in the *An. gambiae s.l.* population in **Chapter 4** was not identified in *An. gambiae s.s*. in this chapter. This may be because A269G was identified at a low allelic frequency of 1.1% in **Chapter 4** using amplicon sequencing, and this low frequency may explain its absence in the sample of *An. gambiae s.s.* mosquitoes analysed with WGS in this chapter.

Anopheles mosquitoes have been identified as some of the most genetically diverse eukaryotic organisms ever sequenced[27], and known SNPs associated with resistance are still unable to explain the majority of the variance in phenotypic resistance seen in *Anopheles* vector populations[28,29]. In addition to SNPs, insecticide resistance has been associated with CNVs in metabolic resistance genes including cytochrome P450s[30,31], and pyrethroid resistance in *Anopheles* has been strongly associated with CNVs in P450 encoding genes[28,30,32–35]. In this chapter, genome wide selection scans using Garud's H_{12} statistic were used to investigate the presence of selective sweeps in the resistant and susceptible *An. gambiae s.s.* mosquitoes across each chromosome. This analysis identified two selective sweeps: one in chromosome X and one in chromosome 2R. These selective sweeps overlapped with metabolic genes previously associated with insecticide resistance, including the cytochrome P450s *cyp9k1* and the *cyp6aa/cyp6p* gene cluster[32,33,36]. These selective sweeps were identified in both resistant and susceptible mosquito populations and were not associated with the resistance phenotype in our small sample set. However, they do indicate selection within the mosquito population in genes known to metabolise insecticides[28,32,33,36]. **Chapter 5** presents the first phenotypic testing for deltamethrin resistance and the first WGS data for *Anopheles* mosquitoes from the Archipelago, contributing further data of significance for vector control policy in this region.

Objective 5: To generate and analyse WGS data for *An. melas* mosquitoes from across the Archipelago, to investigate the genetic diversity, insecticide resistance status and signatures of selection within this population.

Published Research Paper – Chapter 6

An. melas is the most abundantly reported *Anopheles* species on the Bijagós Archipelago and persists throughout the dry season when population numbers of other *An. gambiae s.l.* species dwindle[25]. This species may have a role in malaria transmission on the Archipelago[6,25]. However, the genetic diversity and insecticide resistance status of *An. melas* on the Bijagós has not previously been studied. In **Chapter 6,** I generated and analysed WGS data for 30 *An. melas* mosquitoes collected from across the Bijagós Archipelago. These data provide unique insights into this species which may be relevant to malaria control in this and other similar coastal settings in West Africa.

I hypothesised that *An. melas* would harbour low frequencies of insecticide resistance mutations common to other species in the *An. gambiae s.l.* complex, and that we would detect genetic separation between *An. melas* from different islands. Maximum likelihood trees indicated that the

population of *An. melas* from the Archipelago clustered into two distinct groups. However, these groups were not determined by the geographic location of samples. This is suggestive of historical gene flow between *An. melas* populations on the different islands, despite their geographical distance being greater than the known dispersal ranges of *An. gambiae s.l.* mosquitoes[37]. Closer analysis using fixation statistics and principal components analysis (PCA) revealed that the differentiation of *An. melas* into two distinct groups was mainly localised to the mitochondrial genome. Whereas the nuclear genomes were relatively homogenous between groups. Furthermore, genetic differentiation between the two groups localised to the cytochrome C oxidase subunits *cox1* and *cox2*, and the NADH dehydrogenase subunits *nadh1*, *nadh4*, *nadh4L* and *nadh5*. These gene families have been used previously to investigate phylogenetic relationships within the *An. gambiae s.l.* complex[22,38,39], and genetic separation between the two groups of *An. melas* is intriguing. Guinea-Bissau is a region of extreme hybridisation, with rates of *An. gambiae s.s. – An. coluzzii* hybridisation > 20% consistently reported[40]. However, the *An. melas* grouping was not explained by hybridisation in this analysis. This study warrants WGS with a larger sample size of *An. melas* to further investigate the existence of this grouping, but hints towards the possible existence of a sub-species of *An. melas* in Guinea-Bissau.

Analysis of genetic variation in *An. melas* found that SNPs commonly associated with insecticide resistance in *An. gambiae s.s.* and *An. coluzzii* were absent in the *An. melas* population. Only one of the target insecticide resistance mutations was identified in *An. melas*: *vgsc* M490I. An additional 27 non-synonymous SNPs were identified in the resistance genes *vgsc, gste2, rdl,* and *ace1* which have not previously been reported. Signatures of selection were investigated across insecticide resistance genes using a genome-wide selection scan with Garud's H_{12} statistic, and no peaks indicating selection were observed. This absence of selection signatures contrasted starkly to the two peaks of selection that were identified using H_{12} analysis in chromosomes X and 2R of *Anopheles gambiae s.s.* in **Chapter 5**. Within population and between population selective signatures were also investigated in *An. melas* using the iHS and XP-EHH statistics. No significant iHS or XP-EHH scores were identified in loci associated with insecticide resistance. Overall, these data indicate an apparent absence of selection pressure from insecticides in the *An. melas* population. There are a number of possible explanations for this absence of selection signatures. 1) It may be because *An. melas* come into contact with ITNs less frequently than other *An. gambiae s.l.* species through outdoor biting or circumventing exposure to ITNs, which is supported by previous research[25,41]. 2) Alternatively, *An. melas* may be evolving resistance in response to insecticide selection pressure, but through different molecular mechanisms to other more extensively studied *An. gambiae s.l.* species. Consequently,

this may be resulting in different target SNPs in *An. melas* which could even be in different insecticide resistance genes. 3) The sample size of 30 was relatively small and we may have seen signatures of selection in the *An. melas* genome using a larger sample size. In **Chapter 5**, using a larger sample size of 113 *An. melas,* we identified three of the *An. gambiae s.l.* target insecticide resistance mutations at low allelic frequencies: *vgsc* L995F (2.14%), N1570Y (1.12%) and A1746S (0.76%). This suggests that these alleles are circulating in the *An. melas* population, albeit at very low frequency, and that we may have identified them in this WGS study with a larger sample size. 4) Finally, *An. melas* could be feeding mostly on non-human hosts and therefore have little exposure to ITNs. Overall, the apparent absence of insecticide resistance mutations and absence of selective sweeps in insecticide resistance genes in *An. melas* is encouraging. However, if this species is playing a role in malaria transmission and is circumventing contact with insecticides, this presents a potential challenge to vector control as conventional ITNs and IRS may be less effective for controlling this species. More research is required to understand the contribution of *An. melas* to malaria transmission and whether it is phenotypically resistant to key insecticides.

Key Study Limitations

The limitations of this thesis have been discussed in detail in the individual chapters, and key limitations are summarised here.

In **Chapter 3**, *P. falciparum* WGS data was generated from DBS samples collected during a cross-sectional survey during the peak malaria transmission season in 2018. The DBS samples were collected from survey participants with asymptomatic malaria, making acquisition of *Plasmodium* DNA difficult. To overcome this challenge, SWGA was used to selectively amplify the *Plasmodium* DNA within the total DNA extraction from the DBS[42–44]. This enabled the acquisition of sufficient quantities of parasite DNA to generate high quality WGS data, enabling the analysis of population dynamics and drug resistance mutations in these parasitic infections. Although successful, data were generated for a small set of *P. falciparum* infections, making it challenging to draw conclusions about the *P. falciparum* population on the Archipelago as a whole. Furthermore, the use of SWGA to amplify DNA can distort genome coverage data, which meant that analysing the copy number of genes associated with antimalarial drug resistance was not feasible, and coverage of mutations associated with artemisinin-resistance was poor. In future studies, a larger number of *P. falciparum* samples from the Archipelago should be used to generate drug resistance profiles. For example, using targeted multiplex amplicon sequencing[11,44] or direct WGS from dried erythrocyte spots (DESs)

collected from leukocyte-depleted venous blood samples[45]. These techniques may allow for more reliable acquisition of parasite genome sequence data in this resource limited setting.

In **Chapter 4,** multiplex amplicon sequencing was used to investigate the prevalence of insecticide resistance markers in *Anopheles gambiae s.l.* from 13 islands across the Archipelago. The key limitation of this study was that inter-island comparisons were limited to 5 islands due to the small sample sizes of mosquitoes sampled from the remaining islands. In the future, multiplex amplicon sequencing could be scaled up to include more samples from each of the different islands, including multiple sites per island, to enable more robust comparisons of allele frequencies of insecticide resistance markers between different locations in the Archipelago.

In **Chapter 5**, insecticide susceptibility bioassays were conducted to test whether the *An. gambiae s.l.* vector population on Bubaque island exhibited deltamethrin resistance. The recommended sample size for discriminating concentration testing was reached[5], confirming the presence of deltamethrin resistance on the islands. However, due to logistical constraints during fieldwork, the sample size for testing with 5x and 10x discriminating concentrations was not reached. Therefore, moderate to high intensity resistance was indicated, but further phenotypic tests are required to confirm this finding.

In **Chapter 6**, the genetic diversity and insecticide resistance status of *An. melas* was investigated. The key limitation of this chapter is that the *An. melas* sequencing reads were aligned to the *An. gambiae s.l.* reference, as the quality of the *An. melas* reference genome was not sufficient for population genetic analyses. This meant that structural variants and gene annotations were made in reference to the *An. gambiae s.l.* reference. Future studies of the population dynamics within this species would benefit from an improved reference genome for *An. melas*.

Future Studies

This thesis has investigated the genetic diversity and antimalarial resistance status of *P. falciparum*, and the genetic diversity and insecticide resistance status of *An. gambiae s.l.* vectors, on the Bijagós Archipelago. As a result of this work, there are several areas of research which should be further explored.

Chapter 3 of this thesis investigated the population dynamics and insecticide resistance markers in *P. falciparum* from the Bijagós for the first time, using DBS samples collected before the start of the MATAMAL MDA trial. Following the conclusion of the MATAMAL MDA trial, additional DBS samples

have been collected as part of routine malaria prevalence surveys[2]. These DBS samples could be used to compare the prevalence of antimalarial resistance mutations in the *P. falciparum* population at the end of the trial with the baseline resistance profile generated in **Chapter 3**. This would enable investigation of the impact of the MATAMAL MDA with DHAP on molecular markers of antimalarial drug resistance. If candidate resistance markers were identified, these could be investigated using *in vitro* functional characterisation to study their impact on drug resistance. Future work with larger sample sizes also presents the opportunity to investigate genetic differentiation between *P. falciparum* from different islands of the Archipelago, which would require high quality genomic data to be generated from enough samples from different islands.

Chapter 4 of this thesis uses multiplex-amplicon sequencing to investigate the prevalence of molecular markers of insecticide resistance in *An. gambiae s.l.* on the islands. **Chapter 5** includes susceptibility testing for deltamethrin resistance in *An. gambiae s.l.* using WHO tube tests on Bubaque island, followed by WGS of mosquitoes to investigate the genomics of deltamethrin resistance. Future studies would benefit from continuing with intensity concentration bioassays to determine the intensity of deltamethrin resistance. In addition, susceptibility testing should be expanded to test additional insecticides which may be used in the region in the future, such as the pyrrole insecticide chlorfenapyr[5], and synergist-insecticide bioassays should be used to identify the presence of metabolic resistance[5]. Currently, the infrastructure required to conduct larval rearing and insecticide susceptibility bioassays does not exist on islands other than Bubaque. However, investment in the development of a limited number of sentinel sites over a small number of islands would enable the collection of extensive insecticide susceptibility data. For example, sentinel sites could be set up on 4 of the 18 permanently inhabited islands, deliberately selected to be geographically far apart, and routine insecticide susceptibility testing could be conducted. This would generate insecticide resistance data across different timepoints from different geographic locations across the Archipelago. This could include the collection of phenotypic bioassay data, intensity of resistance data, synergist insecticide bioassay data, and genomic data. Genomic data could be generated using high throughput multiplex amplicon sequencing to routinely investigate the prevalence of known resistance mutations, and using WGS, which could be used to further unveil the population structure of *An. gambiae s.l.* on the islands[35]. The phenotypic and genomic insecticide resistance data collected could be used to inform surveillance and potentially implementation of future vector control tools on the Archipelago. For example, to input into an Insecticide Resistance Monitoring and Management plan, as recommended by WHO[46,47]. Furthermore, this data could be analysed longitudinally to monitor the evolution of insecticide resistance in response to the

implementation of novel vector control tools. Importantly, knowledge of insecticide resistance status is only one component of effective vector control. There is also a need to measure additional factors including the coverage, usage, and durability of ITNs on the islands, along with longevity of insecticides[48,49].

Chapter 6 of this thesis studies the lesser-known vector species *An. melas.* Further studies need to be conducted to understand the contribution of this species to malaria transmission, and to understand the phenotypic resistance status of this vector. Considering the recommendations for longitudinal genomic and phenotypic resistance surveillance in *Anopheles* vectors on the Archipelago, these studies should include provision for molecular characterisation of *Anopheles gambiae s.l.* species. This would enable phenotypic and genomic resistance data to be species disaggregated, which would allow us to understand the potentially differing insecticide resistance statuses of the different species within this cryptic vector complex. This would require large mosquito sample size collections to be able to catch enough of each species to ensure robust sample sizes during phenotypic testing[35].
Finally, research should be conducted into the level of migration of *An. gambiae s.l.* mosquitoes and *P. falciparum* parasites between mainland Guinea-Bissau and the Bijagós Archipelago. This could be done by investigating phylogenetic relationships between mainland and island populations, as has been done previously on the Comoros islands in the Indian Ocean[50]. Levels of migration will be important to understand to be able to predict the likelihood of reintroduction of *Plasmodium* to the Bijagós Archipelago, in the event of successful malaria elimination.

Historically, genomics research involving data derived from malaria-endemic communities has occurred in the context of large-scale export of samples from countries in the Global South to wealthier institutions in the Global North, with analysis conducted in the majority by researchers of European descent[51–54]. This pattern reinforces existing inequalities in Global Health, and demonstrates the need for greater equity in genomics research[54]. The genomics community has made some progress in increasing equity[53], but the future calls for the funders of health research to invest in sustainable capacity strengthening in the Global South, and for increased knowledge sharing between institutions in the Global North and Global South. Capacity strengthening in the Bijagós Archipelago could include the investment of funding by research partners and programmes into the development of sentinel sites on the islands, where phenotypic and genomic testing into the presence of antimalarial resistance and insecticide resistance could take place. Technological advances including the development of portable and more resource-efficient methods for genome

sequencing are encouraging and will facilitate the use of these technologies in resource-limited settings. Emerging technologies including portable sequencing devices such as the MinION enable long read sequencing in remote settings, and may increase the availability of sequencing technologies in resource-limited settings[55]. Furthermore, these portable devices present the possibility of generating real-time insights into drug resistance markers in patients with symptomatic malaria, which could then inform treatment regimens[55].

As novel malaria control tools are implemented and sequencing technologies continue to improve, genomic techniques can be used to understand the impact of malaria control interventions on the genetic diversity, population structure, and resistance profile of the *P. falciparum* parasite and *An. gambiae s.l.* mosquitoes. By understanding the impact of malaria interventions on both the parasite and vectors in specific settings, the selection and development of malaria control tools can be better optimised towards achieving elimination of this disease, and winning the evolutionary arms race.

References

1. Hutchins, H. *et al.* A survey of knowledge , attitudes and practices regarding malaria and bed nets on Bubaque Island , Guinea - Bissau. *Malaria Journal* 1–15 (2020) doi:10.1186/s12936-020-03469-1.
2. Hutchins, H. *et al.* Protocol for a cluster randomised placebo-controlled trial of adjunctive ivermectin mass drug administration for malaria control on the Bijagós Archipelago of Guinea-Bissau: the MATAMAL trial. *BMJ Open* **13**, e072347 (2023).
3. Moss, S. *et al.* How has mass drug administration with dihydroartemisinin-piperaquine impacted molecular markers of drug resistance? A systematic review. *Malaria journal* **21**, 186 (2022).
4. World malaria report 2023. https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2023.
5. World Health Organization. Manual for monitoring insecticide resistance in mosquito vectors and selecting appropriate interventions, 2022.

6. Ant, T. *et al.* A survey of Anopheles species composition and insecticide resistance on the island of Bubaque , Bijagos Archipelago , Guinea - Bissau. *Malaria Journal* 1–9 (2020) doi:10.1186/s12936-020-3115-1.
7. Moss, S. *et al.* Population dynamics and drug resistance mutations in Plasmodium falciparum on the Bijagós Archipelago, Guinea-Bissau. *Sci Rep* **13**, 1–14 (2023).
8. MalariaGEN *et al.* An open dataset of Plasmodium falciparum genome variation in 7,000 worldwide samples. *Wellcome Open Research* **6**, (2021).
9. Jovel, I. T., Kofoed, P. E., Rombo, L., Rodrigues, A. & Ursing, J. Temporal and seasonal changes of genetic polymorphisms associated with altered drug susceptibility to chloroquine, lumefantrine, and quinine in Guinea-Bissau between 2003 and 2012. *Antimicrobial Agents and Chemotherapy* **59**, 872–879 (2015).
10. Ursing, J. *et al.* Chloroquine-susceptible and -resistant Plasmodium falciparum strains survive high chloroquine concentrations by becoming dormant but are eliminated by prolonged exposure. *J Antimicrob Chemother* **77**, 1005–1011 (2022).
11. Nag, S. *et al.* High throughput resistance profiling of Plasmodium falciparum infections based on custom dual indexing and Illumina next generation sequencing-technology. *Scientific Reports* **7**, 1–13 (2017).
12. Nag, S. *et al.* Proof of concept: used malaria rapid diagnostic tests applied for parallel sequencing for surveillance of molecular markers of anti-malarial resistance in Bissau, Guinea-Bissau during 2014-2017. *Malaria journal* **18**, 252 (2019).
13. Mohammed, A. *et al.* Trends in chloroquine resistance marker, Pfcrt-K76T mutation ten years after chloroquine withdrawal in Tanzania. *Malaria Journal* **12**, 415 (2013).
14. Kublin, J. G. *et al.* Reemergence of Chloroquine-Sensitive Plasmodium falciparum Malaria after Cessation of Chloroquine Use in Malawi. *The Journal of Infectious Diseases* **187**, 1870–1875 (2003).

15. Mwai, L. *et al.* Chloroquine resistance before and after its withdrawal in Kenya. *Malar J* **8**, 106 (2009).
16. SAWADOGO, C. W. *et al.* Quality of chloroquine tablets available in Africa. *Ann Trop Med Parasitol* **105**, 447–453 (2011).
17. Humphreys, G. S. *et al.* Amodiaquine and Artemether-Lumefantrine Select Distinct Alleles of the Plasmodium falciparum mdr1 Gene in Tanzanian Children Treated for Uncomplicated Malaria . **51**, 991–997 (2007).
18. Sisowath, C. *et al.* In vivo selection of Plasmodium falciparum pfmdr1 86N coding alleles by Artemether-Lumefantrine (Coartem). *Journal of Infectious Diseases* **191**, 1014–1017 (2005).
19. Ursing, J., Rombo, L., Rodrigues, A. & Kofoed, P. Artemether-Lumefantrine versus Dihydroartemisinin-Piperaquine for Treatment of Uncomplicated Plasmodium falciparum Malaria in Children Aged Less than 15 Years in Guinea-Bissau – An Open- Label Non-Inferiority Randomised Clinical. **22**, 1–11 (2016).
20. Amato, R. *et al.* Genetic markers associated with dihydroartemisinin–piperaquine failure in Plasmodium falciparum malaria in Cambodia: a genotype–phenotype association study. *The Lancet Infectious Diseases* **17**, 164–173 (2017).
21. Moss, S. *et al.* Genomic surveillance of Anopheles mosquitoes on the Bijagós Archipelago using custom targeted amplicon sequencing identifies mutations associated with insecticide resistance. *Parasites Vectors* **17**, 10 (2024).
22. Campos, M. *et al.* High-throughput barcoding method for the genetic surveillance of insecticide resistance and species identification in Anopheles gambiae complex malaria vectors. *Scientific Reports* **12**, 13893 (2022).
23. Acford-Palmer, H. *et al.* Identification of two insecticide resistance markers in Ethiopian Anopheles stephensi mosquitoes using a multiplex amplicon sequencing assay. *Sci Rep* **13**, 5612 (2023).

24. Collins, E. L. *et al.* A next generation targeted amplicon sequencing method to screen for insecticide resistance mutations in Aedes aegypti populations reveals a rdl mutation in mosquitoes from Cabo Verde. *PLoS Negl Trop Dis* **16**, e0010935 (2022).
25. Pretorius Elizabeth. UNDER REVIEW - A survey of indoor and outdoor biting behaviour, species composition and circumsporozoite rate of malaria vectors in the Bijagós Archipelago, Guinea-Bissau.
26. Silva, R., Mavridis, K., Vontas, J., Rodrigues, A. & Costa, H. Monitoring and molecular profiling of contemporary insecticide resistance status of malaria vectors in Guinea – Bissau. *Acta Tropica* **206**, 105440 (2020).
27. Genetic diversity of the African malaria vector Anopheles gambiae. *Nature* **552**, 96–100 (2017).
28. Lucas, E. R. *et al.* Whole-genome sequencing reveals high complexity of copy number variation at insecticide resistance loci in malaria mosquitoes. *Genome Res* **29**, 1250–1261 (2019).
29. Donnelly, M. J., Isaacs, A. T. & Weetman, D. Identification, Validation, and Application of Molecular Diagnostics for Insecticide Resistance in Malaria Vectors. *Trends Parasitol* **32**, 197–206 (2016).
30. Vontas, J., Katsavou, E. & Mavridis, K. Cytochrome P450-based metabolic insecticide resistance in Anopheles and Aedes mosquito vectors: Muddying the waters. *Pestic Biochem Physiol* **170**, 104666 (2020).
31. Scott, J. G. Cytochromes P450 and insecticide resistance. *Insect Biochemistry and Molecular Biology* **29**, 757–777 (1999).
32. Ibrahim, S. S. *et al.* Pyrethroid Resistance in the Major Malaria Vector Anopheles funestus is Exacerbated by Overexpression and Overactivity of the P450 CYP6AA1 Across Africa. *Genes* **9**, 140 (2018).
33. Hearn, J. *et al.* Multi-omics analysis identifies a CYP9K1 haplotype conferring pyrethroid resistance in the malaria vector Anopheles funestus in East Africa. *Mol Ecol* **31**, 3642–3657 (2022).

34. Mitchell, S. N. *et al.* Metabolic and Target-Site Mechanisms Combine to Confer Strong DDT Resistance in Anopheles gambiae. *PLOS ONE* **9**, e92662 (2014).
35. Lucas, E. R. *et al.* Genome-wide association studies reveal novel loci associated with pyrethroid and organophosphate resistance in Anopheles gambiae and Anopheles coluzzii. *Nat Commun* **14**, 4946 (2023).
36. Vontas, J. *et al.* Rapid selection of a pyrethroid metabolic enzyme CYP9K1 by operational malaria control activities. *Proceedings of the National Academy of Sciences* **115**, 4619–4624 (2018).
37. Kaufmann, C. & Briegel, H. Flight performance of the malaria vectors Anopheles gambiae and Anopheles atroparvus. *J Vector Ecol* **29**, 140–153 (2004).
38. Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates.
39. Rubinoff, D. & Holland, B. S. Between Two Extremes: Mitochondrial DNA is neither the Panacea nor the Nemesis of Phylogenetic and Taxonomic Inference. *Systematic Biology* **54**, 952–961 (2005).
40. Vicente, J. L. *et al.* Massive introgression drives species radiation at the range limit of Anopheles gambiae. *Sci Rep* **7**, 46451 (2017).
41. Reddy, M. R. *et al.* Outdoor host seeking behaviour of Anopheles gambiae mosquitoes following initiation of malaria vector control on Bioko Island, Equatorial Guinea. *Malaria Journal* **10**, 184 (2011).
42. Oyola, S. O. *et al.* Whole genome sequencing of Plasmodium falciparum from dried blood spots using selective whole genome amplification. *Malaria Journal* 1–12 (2016) doi:10.1186/s12936-016-1641-7.
43. Ibrahim, A. *et al.* Selective whole genome amplification of Plasmodium malariae DNA from clinical samples reveals insights into population structure. *Scientific Reports* **10**, 1–11 (2020).
44. Osborne, A. *et al.* Characterizing the genomic variation and population dynamics of Plasmodium falciparum malaria parasites in and around Lake Victoria, Kenya. *Scientific Reports* **11**, (2021).

45. Nag, S. *et al.* Direct whole - genome sequencing of Plasmodium falciparum specimens from dried erythrocyte spots. *Malaria Journal* 1–8 (2018) doi:10.1186/s12936-018-2232-6.
46. World Health Organization. *Framework for a National Plan for Monitoring and Management of Insecticide Resistance in Malaria Vectors Framework for a National Plan for Monitoring and Management of Insecticide Resistance in Malaria Vectors Global Malaria Programme World Health Organization*. (2017).
47. World Health Organization. Global Malaria Programme. Global plan for insecticide resistance management in malaria vectors. 131 (2012).
48. Lindsay, S. W., Thomas, M. B. & Kleinschmidt, I. Threats to the effectiveness of insecticide-treated bednets for malaria control: thinking beyond insecticide resistance. *The Lancet Global Health* **9**, e1325–e1331 (2021).
49. Hiruy, H. N. *et al.* Durability of long-lasting insecticidal nets (LLINs) in Ethiopia. *Malar J* **22**, 109 (2023).
50. Ditter, R. E. *et al.* Anopheles gambiae on remote islands in the Indian Ocean: origins and prospects for malaria elimination by genetic modification of extant populations. *Sci Rep* **13**, 20830 (2023).
51. Vries, J. de & Pepper, M. Genomic sovereignty and the African promise: mining the African genome for the benefit of Africa. *Journal of Medical Ethics* **38**, 474–478 (2012).
52. de Vries, J. *et al.* Ethical issues in human genomics research in developing countries. *BMC Medical Ethics* **12**, 5 (2011).
53. Atutornu, J., Milne, R., Costa, A., Patch, C. & Middleton, A. Towards equitable and trustworthy genomics research. *EBioMedicine* **76**, 103879 (2022).
54. Fatumo, S. *et al.* A roadmap to increase diversity in genomic studies. *Nat Med* **28**, 243–250 (2022).
55. Jain, M., Olsen, H. E., Paten, B. & Akeson, M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biology* **17**, 239 (2016).