A genomic investigation into three malaria vectors: *An. darlingi*, *An. funestus*, and *An. stephensi*

Holly Acford-Palmer

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

University of London
April 2024

Department of Infection Biology
Faculty of Infectious and Tropical Diseases
London School of Hygiene and Tropical Medicine

Research Group Affiliations:
Professor Susana Campino, Professor Taane Clark & Dr Jody Phelan

Funded by the Biotechnology and Biological Sciences Research Council

I, Holly Acford-Palmer, 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

Abstract

Different *Anopheles spp.* are able to transmit the *Plasmodium* parasite, the causative agent of malaria. Malaria results in over 600,000 deaths a year, predominantly in children and the immunocompromised. To combat malaria, disease control relies heavily on strategies targeting the vectors populations, including the application of insecticides for indoor residual spraying, and treated bed-nets. These interventions are believed to be responsible for >65% of the reduction in malaria cases observed over the last 15 years. However, resistance to commonly used insecticides is rapidly rising, threatening the World Health Organization targets of reducing global malaria.

The detection of insecticide resistance in mosquitoes relies either on time-consuming phenotypic bioassays or expensive and limited-target molecular assays. To address this, I have developed three species specific multi-target amplicon sequencing (amp-seq) panels for high-throughput surveillance of insecticide resistance in *An. darlingi, An. funestus,* and *An. stephensi.* The assays were validated using field isolates for each species, Ethiopian *An. stephensi,* Brazilian *An. darlingi*, and *An. funestus* from the Democratic Republic of Congo. This work resulted in the detection of known SNPs associated with insecticide resistance, including a pyrethroid resistance associated 2bp deletion in the *CYP6P9a* gene in *An. funestus,* alongside the *kdr*-L1014F and *rdl*-A296S SNPs in *An. stephensi.* Several putatively novel missense SNPs were also identified in genes associated with resistance for all three species.

The availability of genomic data for many *Anopheles spp* is limited but can be used to provide insights into ongoing selective pressure, potentially due to insecticides, as well as unravelling population dynamics. This thesis expands the available sequence data for *An. stephensi* and *An. darlingi* to provide insights into the genomic landscape of these vectors. I generated whole genome sequencing (WGS) data and performed population genetics analysis on *An. darlingi* (n=31) and *An. stephensi* (n=72) isolates. For *An. stephensi,* the study included isolates from Ethiopia (n=27), India (field 21; colony 16) and Pakistan (n=8). An ancestral analysis revealed shared ancestry between Ethiopian and Indian field isolates. Further, insecticide resistance linked mutations were identified, including the *kdr*- L1014F and *rdl*-A296S in the Ethiopian isolates, and the *rdl*-A296S and *rdl*-V327I in the Indian isolates. For *An. darlingi,* the study included samples from Rondônia state in Brazil (colony 8, field 23). No known insecticide resistance associated mutations were identified in the Brazilian *An. darlingi,* either by amp-seq or WGS. To find other mechanisms, population genetic tests of selection were applied, and identified candidate regions for insecticide resistance, such as *CYP4c1, CYP4c3, and CYP307a1.*

Overall, these investigations have increased our understanding of *An. darlingi* and *An. stephensi* genomic diversity and provide baseline data and analysis for much needed larger studies. The high-throughput sequencing-based assays developed will inform insecticide resistance surveillance. Their utility was demonstrated through their role in identifying putative novel mutations involved in insecticide susceptibility, which can be followed up in functional studies. Overall, genomics-based approaches, such as those developed here, have the potential to inform control strategies across a range of vector borne diseases. Through the integration of low-cost and high-throughput approaches within vector control programs, it will assist with the urgent need to disrupt transmission, and thereby reduce the high burden of disease.

Acknowledgements

Firstly, I'd like to thank my supervisors Susana, Taane and Jody, for all the help and support they have given over the last 3 and half years to get to this point and allowing me to be a part of this lab group. I would also like to thank the BBSRC LIDo scheme for providing me with the funding to complete my PhD. Thank you to Mojca and Lulu, for all the support and time you've given me over my PhD. Also, to the team at the BSI for being so welcoming over my PIPS, a special thank you to Erika, Teresa, Jasmine, and Jennie.

To all the people who have supported me on my scientific journey so far: Ann Steward, Jess Dale, Cristina Montiel-Duarte, Pravesh Dhanilal, Hitiksha Maru, Jahanvi Joshi, Sarah Philips, and Lindsay Hall. I wouldn't have made it to this point without the teaching and guidance you have all provided me.

To "The Hub" members, past and present: Amy, Anna, Emilia, Ernest, Gary, Jamille, Joe, Julian, Lin, Mark, Myrela, Pepi, and Woranich. Whether knowingly or not, you've all played a massive part in my PhD. So, thank you for all the intellectual and emotional support you've all given.

A few more specific thank yous: to the help gc, Ashley, Emma, Gabbie, Leen, and Sophie. Thank you for all the laughing and crying sessions. There isn't a group of people I'd rather have gone through this with– you're all the best! Matt, the chilliest there ever was. Dan, m8, keep on slaying it. Thank you for the guidance you've both given over the last 4 years; it's been invaluable.

My girl, bestie Francesca, we started this journey together back in summer 2k18, and I would never have made it without the delulu, tea spilling, and excellent tik tok taste. Can't wait to give you your friendship bracelets!! To Elisabeth, my lab partner forever, and the other half of my brain, I would never have pursued a PhD if it wasn't for your infectious enthusiasm, so thank you.

To Louise, thank you for the support and always taking an interest in what I'm doing - I appreciate every time you ask. To Serah, all the amazing food and drink you've served over the last 4 years have been integral to me getting through, as have your domestics with Michael.

To Michael, who'd have thought? Not me! Thank you for your constant support, Aries energy, and unwavering belief in me. I couldn't have done it without you, here is to all our future pets and I can't wait to celebrate our freedom- we did it to em.

To both my Nanny's (Big and Little), my Grandad, and the rest of my family, on every side, thank you for all your love and support. I promise this is definitely the last degree, and I'll get a proper job now- can't say there won't be more mosquito squashing though.

To my biggest cheerleaders, Mum, and Dad, thank you. You always said to me in school, just try your best that's all we can ask- well I might have taken that slightly further than you expected but here you go! I'm so grateful for the endless love and support you've given me over the last 28 years. To my Mum, your dedication and determination inspire me

every day, thank you for all you do - LYTTMAB. To my Dad, thank you for your unfaltering belief, sense of humour, and always being there. Also, thank you for giving me your excellent taste in football teams – MOT.

And finally, to anyone I haven't mentioned by name, but who have been there along the way: thank you for your support. An especially big thank you to anyone who let me pet sit and to Taylor Swift.

Thesis Publications and Manuscripts

1. Acford-Palmer, H., Phelan, J.E., Tadesse, F.G. *et al.* Identification of two insecticide resistance markers in Ethiopian *Anopheles stephensi* mosquitoes using a multiplex amplicon sequencing assay. *Sci Rep* **13**, 5612 (2023). https://doi.org/10.1038/s41598-023-32336-7
2. Acford-Palmer, H., Manko, E., Phelan, J.E., *et al*. Genome-wide population genetics, and molecular surveillance of insecticide resistance in Ethiopian *Anopheles stephensi*. (To submit)
3. Acford-Palmer, H., Andrade, A.O., Phelan, J.E., *et al.* Application of a targeted amplicon sequencing panel to screen for insecticide resistance mutations in *Anopheles darlingi* populations from Brazil. Sci Rep. (In review) https://doi.org/10.21203/rs.3.rs-3053716/v1
4. Acford-Palmer, H., Andrade, A.O., Manko, E., *et al.* Genome -wide population genetics of *Anopheles darlingi*, a malaria vector from the State of Rondônia, Brazil. (To submit)
5. Acford-Palmer, H., Campos, M., Phelan, J.E., *et al.* Detection of insecticide resistance markers in *Anopheles funestus* from the Democratic Republic of the Congo using a targeted amplicon sequencing panel. Sci Rep **13**, 17363 (2023). https://doi.org/10.1038/s41598-023-44457-0

Additional Publications

1. Collins, E.L., Phelan, J.E, Hubner, M., Spadar, A., Campos, M.O., Ward,D., **Acford-Palmer, H.** 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 neglected tropical diseases* **16**,12 (2022) https://doi.org/10.1371/journal.pntd.0010935
2. Kristan, M., **Acford-Palmer, H**., Campos, M.O. *et al.* Towards environmental detection, quantification, and molecular characterization of *Anopheles stephensi* and *Aedes aegypti* from experimental larval breeding sites. *Sci Rep* **13**, 2729 (2023). https://doi.org/10.1038/s41598-023-29657-y
3. Moss. S., Pretorius, E., Ceesay, S., Hutchins, H., Teixeira Da Silva, E., Ndiath, M.O., Phelan, J.E., **Acford-Palmer, H.,** *et al.* Genomic surveillance of 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

Table of Contents

Abbreviations

Ace-1	Acetylcholinesterase 1
ACT	Artemisinin Combination Therapy
Amp-seq	Amplicon Sequencing
Bp	Base-pair
Cox-1	Cytochrome Oxidase I
CYP	Cytochrome P450
DBP1	Duffy Binding Protein 1
DDT	Dichlorodiphenyltrichloroethane
DNA	Deoxyribonucleic Acid
DRC	The Democratic Republic of the Congo
DVS	Dominant vector species
eDNA	Environmental DNA
Fst	Fixation Index
GABA	Gamma-Aminobutyric Acid
GST	Glutathione S-Transferase
GSTe2	Glutathione S-Transferase Epsilon 2
HOA	Horn of Africa
iHS	Integrated Haplotype Score
INDEL	Insertion/Deletion
IRS	Indoor Residual Spraying
ITN	Insecticide Treated Net
ITS2	Internally Transcribed Spacer 2
Kdr	Knockdown resistance
LD	Linkage Disequilibrium
LLINs	Long-Lasting Insecticidal Nets
Mb	Mega-base pairs
MCT	Monocarboxylate Transporter
MOA	Mode of Action
Mt	mitochondrial
ND5	Nicotinamide Adenine Dinucleotide Hydrogen (NADH) dehydrogenase 5
NGS	Next Generation Sequencing
NS	Non-Synonymous
NJ	Neighbourhood Joining
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
Poc	Plasmodium ovale curtisi
Pow	Plasmodium ovale wallikeri
qPCR	Quantitative PCR
RBC	Red Blood Cell

Rdl	resistant to dieldrin
RDT	Rapid Diagnostic Test
RFLP	Restriction fragment length polymorphism
s.l.	sensu lato
SNP	Single Nucleotide Polymorphism
SP	Sulphadoxine-Pyrimethamine
Spp	Species
s.s	sensu stricto
Td	Tajima's D
TfR1	host transferrin receptor 1
VCF	Variant Call Format
Vgsc	Voltage gated sodium channel
UAE	United Arab Emirates
WGS	Whole Genome Sequencing
WHO	World Health Organisation
XP-EHH	Extended Haplotype Homozygosity

Chapter 1: Introduction

Malaria

Malaria is a significant global health problem and is caused by parasites from the genus *Plasmodium*. Six *Plasmodium* species can cause disease in humans: *P. falciparum, P. vivax, P. ovale curtisi (Poc), P. ovale wallikeri (Pow), P. malariae* and *P. knowlesi*. These parasites can be found across the subtropical and tropical regions in Africa, Asia, and Central and South America. Of these six species, most mortality results from *P. falciparum* infections. Historically concentrated in Sub-Saharan Africa, this species results in an estimated 233 million infections in the World Health Organization (WHO) African Region[1]. In total, malaria causes ~593,000 deaths a year, the majority the result of *P. falciparum* infection[1]. Children under the age of 5 years, pregnant woman and the immunocompromised are the most vulnerable groups.

P. vivax, the second most virulent parasite species, is the most prevalent cause of malaria outside Africa. This parasite has a wider geographic distribution and is found in Asia, South America, Middle East, and Oceania and in some regions in Africa. *P. vivax* was originally thought to rarely occur in African populations due to the absence of the Duffy antigen on red blood cells surface[2]. Caused by an alteration in the Duffy antigen receptor for chemokine gene, this protective mutation was believed to prevent parasite entry into red blood cells (RBCs)[3,4]. However, many studies have detected *P. vivax* infections in Duffy negative individuals[5]. Research has implicated further proteins such as host transferrin receptor 1 (TfR1) in *P. vivax* entry into RBCs[5–7]. It is believed that the severity of *P. falciparum* malaria has prevented the detection of *P. vivax* which can often appear asymptomatically[8].

P. knowlesi, a zoonotic malaria and accounts for up to 95% of malaria cases in certain areas in Southeast Asia, particularly in Malaysia, with some severe and fatal outcomes[9,10]. Infections with the neglected malaria parasites *Poc, Pow* and *P. malariae* are rarely life threatening but

can also cause substantial morbidity, it is commonly detected as a co-infection with *P. falciparum* and *P. vivax.*

The life cycle of these parasites' species is similar (**Figure 1**), involving two hosts: humans and female *Anopheles* mosquitoes. The asexual stages of the *Plasmodium* parasite occur in humans, but the sexual stages occur when both male and female gametocytes are ingested by the *Anopheles* mosquito during a blood meal from an infected host, this is known as the sporogenic cycle[11].

Following ingestion, the malaria gametocytes activate into gametes (gametogenesis) in the midgut lumen. The gametes fuse to form a zygote, meiosis and genetic recombination then occur to transform the zygote into an ookinete and invades the midgut epithelium. Here they develop into oocysts with sporozoites forming inside, which eventually are released into the mosquito haemocoel and transported, via haemolymph, to the mosquito salivary glands[2]. Where, when the mosquito takes its next bloodmeal, the sporozoites will be released from into the blood stream of the mammal to continue the infection cycle.

In humans, these sporozoites move to the liver and infect hepatocytes to begin maturation and asexual replication. It is here the parasite lifecycle varies in *P. vivax* and *P. ovale spp.*, in that dormant liver hypnozoites are established[12,13]. This extra liver stage can remain undetected and reactivate weeks, months or even years after initial infection clearance[14,15]. After 1-2 weeks the matured schizonts move out of liver cells via vesicles, into the bloodstream where they burst releasing merozoites, that invade RBCs. The merozoites then multiply, and mature into trophozoites or gametocytes. Trophozoites evolve into schizonts, which continue to invade and rupture RBCs, which results in clinical malaria symptoms[16]. Whilst the gametocytes are taken up in blood meals by female *Anopheles spp* mosquitoes to continue the reproductive cycle.

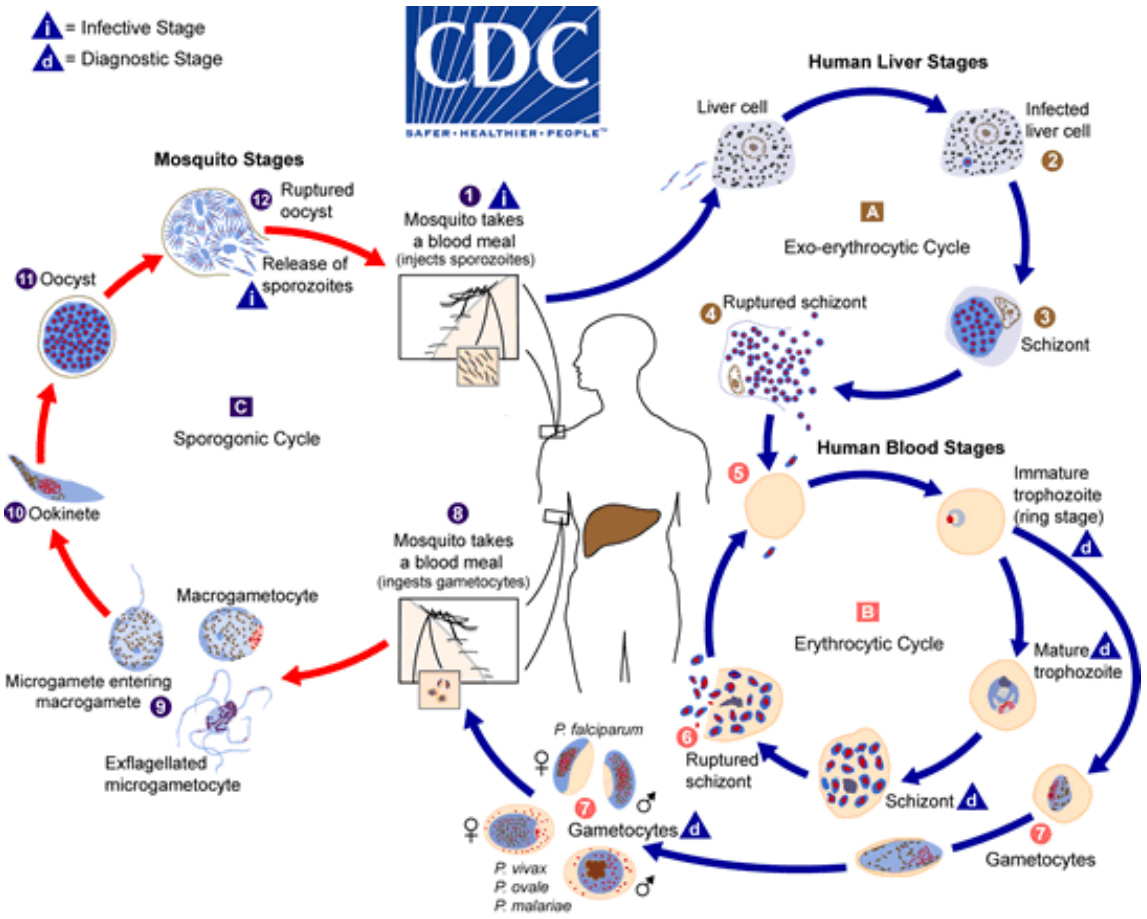


Figure 1. Lifecycle of malaria parasite *P. falciparum* through vector and host[11].

Malaria Control

Malaria is a complex, and multi-faceted disease, with parasite, human, and vector needing to be accounted for when attempting to control it. Control of malaria typically relies on transmission prevention, including early diagnosis using microscopy and rapid diagnostic tests (RDTs), and easy access to antimalarial drugs such as chloroquine, sulphadoxine-pyrimethamine (SP), artemisinin-based compounds and primaquine. Arguably the control of *Anopheles spp* mosquitoes has made the most significant contribution to reducing the burden of malaria since 2000. The use of indoor residual spraying (IRS), larval-breeding site spraying, and long-lasting insecticide treated bed nets (LLINs) are estimated to be responsible for 68% of malaria case aversions between 2000 and 2015[17]. Other more novel vector control

methods currently under development include, release of sterile males, or the release of mosquitoes infected with *Wolbachia* – a bacterium that prevents the ability of *Plasmodium spp* to reproduce in the mosquito midgut preventing onward transmission of malaria[18–23].

The WHO currently has the target of reducing global malaria incidence and mortality rates by at least 90% by 2030[24]. Great progress has been made in the reduction of malaria cases and deaths over the last two decades. However recently, the annual reduction of malaria morbidity and mortality has started to stagnate[1]. Between 2015 and 2020 cases were still decreasing, although at a slower rate, but disruptions caused by the COVID-19 pandemic saw cases rise in the subsequent years. Research has suggested that COVID-19 reduced access to malaria testing, LLIN distribution, and drug administration, resulting in a 6% increase in malaria infections from 232 million in 2019 to 247 million in 2021[1,25]. The impact of the pandemic will be felt within the next decade, but as mentioned above, the decrease in rates of mortality and infection were slowing prior to 2020. This has been attributed to the emergence of resistance to front line drug treatments such as Artemisinin Combination Therapy (ACT) across South-East Asia and Africa, alongside rising rates of insecticide reistsance[26–28]. Insecticides still remain an effective control method, even in areas with reported resistance, but this data demonstrates the need for surveillance of both insecticide and anti-malarial drug resistance, in the race to eliminate malaria[29].

Anopheles spp.

Female *Anopheles* mosquitoes are responsible for the transmission of malaria, with half the *Plasmodium* lifecycle occurring in the mosquito midgut (**Figure 1**). *Anopheles spp.,* have different host feeding preferences and transmit various forms of mammalian malaria.

There are over 400 species of *Anopheles* mosquitoes, all with varying vectoral capabilities, geographical distributions, and morphological variations[30]. Around 30-40 of these species are believed to be vectors for human malaria parasite species[30]. There has also been evidence of *Anopheles* mosquitoes transmitting other pathogens such as the O'Nyong O'Nyong virus[31]. Currently this mosquito genus puts an estimated 3.3 billion people at risk from malaria[32,33]. However, with rising climate temperatures this number is likely to increase due to increase in areas with high enough temperatures to sustain malaria reproduction in vectors[34].

Sub-Saharan Africa represents the region with the highest number of malaria infections and deaths[1]. The high mortality burden in this region is a result of the large proportion of infections caused by *P. falciparum*. The primary vector in Africa is *Anopheles gambiae*, which is responsible for a large proportion of malaria transmission[35]. *An. gambiae sensu lato (s.l.)* is part of a species complex with at least nine species within it, representing a morphologically similar sub-group of *Anopheles* species[36,37]. Of this complex, five are believed capable of transmitting *Plasmodium* to humans: *An. gambiae s.l.*; *An. arabiensis; An. coluzzi*; *An. merus;* and *An. melas*[37,38].

Other primary vectors in Africa include *An. arabiensis* (also part of the *An. gambiae* species complex), and *An. funestus sensu stricto (s.s.).* Species such as *An. melas, An. merus* and *An. nili*, are highly anthropophilic and efficient vectors, and can also contribute to malaria transmission across the continent[30,35].

Recently, *An. stephensi*, has moved from Asia into Africa, and has demonstrated the ability to transmit local strains of both *P. vivax* and *P. falciparum*[39]. Originally *An. stephensi* was a dominant vector species (DVS) across South Asia, acting as a primary malaria vector in India and Pakistan[30,40]. In the Asian-Pacific region there are high amounts of diversity amongst vector species, with many primary and secondary vectors contributing to malaria

transmission. *An. culicifacies* occupies a similar geographical distribution to *An. stephensi* in South Asia. Whilst in South-East Asia, species such as *An. dirus s.l., An. minimus s.l., An. sinensis, An. punctulatus s.l.,* and *An. faurauti s.l.* are largely responsible for malaria transmission[30,41].

There are few vectors for human malaria in South America, with *An. darlingi* being the most dominant and responsible for the majority of malaria transmission across Venezuela, Peru, Colombia, and Brazil[30]. Other vectors such as *An. albimanus* also occupy this region, and *An. albitarsis s.l.,* is found in Brazil and Northern Argentina. The latter two species are less efficient vectors than *An. darlingi,* but nevertheless contribute to the region's malaria burden[42].

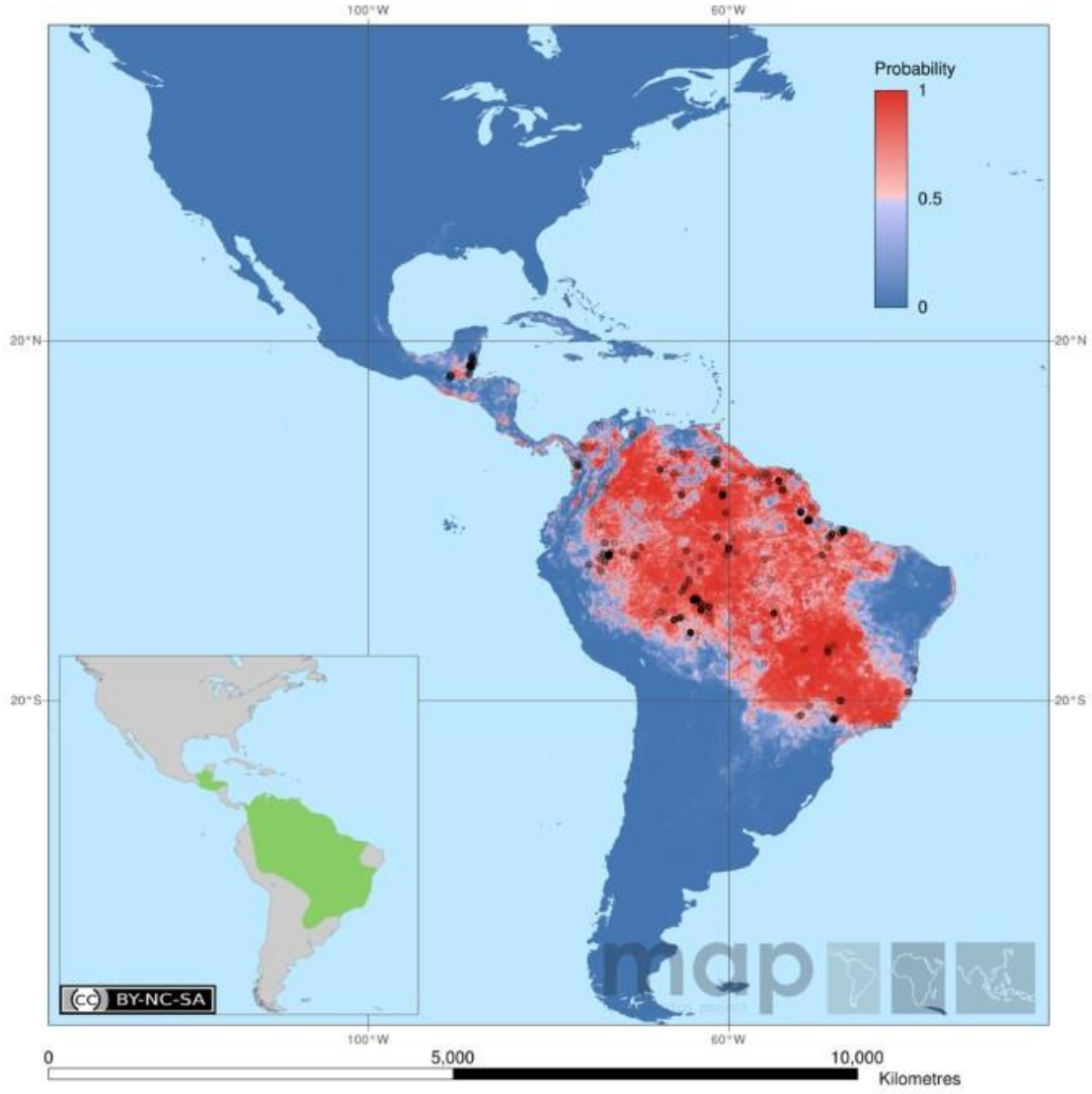
Although many *Anopheles* species can transmit malaria, as most mortality occurs in Africa, most vector control methods are designed around *An. gambiae* behaviours e.g., the use of LLINs as *An. gambiae* predominantly demonstrates night-time biting, or larval spraying of habitats preferred by *An. gambiae*[43]. These methods have been largely successful at reducing the *An. gambiae* population[44,45].

There are large behavioural differences between *Anopheles spp.,* and even within species, regarding their feeding and resting behaviour. *An. gambiae* is a nocturnal feeder, whilst *An. darlingi* often demonstrates peak biting at twilight[46]. *An. darlingi* is a good example of behavioural plasticity in a vector, as various populations have been observed to highly selective anthropophilic vectors, but also displays opportunistic zoophily[47,48]. *An. arabiensis* feeds and rests outdoors (exophagic and exophilic), whilst *An. gambiae* is largely endophagic[49,50]. These examples of behavioural variations in *Anopheles spp*, highlights the need for species specific vector control methods, which requires knowledge of the biology, behaviour, ecology, and susceptibility to insecticides of local *Anopheles spp.*

This thesis will focus on three species of *Anopheles*: *An. darlingi, An. funestus,* and *An. stephensi.*

Anopheles darlingi

An. darlingi, or *Nyssorhynchus darlingi* is distributed throughout Central and South America (**Figure 2**). *An. darlingi* predominantly transmits *P. vivax,* with this species of malaria accounting for the majority of infections in South America[51,52]. This mosquito species is also capable of transmitting *P. falciparum,* however this parasite is less prevalent across the South America continent compared to Africa[1,51]. Found across Colombia, Peru, Venezuela, and Brazil, *An. darlingi* has recently moved more into focus due to the increasing proportion of *P. falciparum* malaria cases across South America[1]. The increased severity of *P. falciparum* malaria, coupled with lower rates of immunity could cause significant morbidity and mortality in the continent, and the need to control this vector is increasingly important.



144 **Figure 2.** Estimated distribution of *An. darlingi* throughout the Central and South Americas[42].

145

146 *An. darlingi* is a proficient vector and can support *Plasmodium* transmission even when
147 parasites are at low densities[53]. It is both endo- and exophagic, it generally rests outside
148 houses or in animal outhouses, and typically feeds on humans in the early morning and
149 evening[54]. These behaviours allow *An. darlingi* to avoid traditional vector control methods
150 such as LLINs. *An. darlingi* is a highly adaptable species and can often fill ecological niches left
151 when other *Anopheles* species are unable to survive environmental changes such as
152 deforestation, which is increasingly common in the Amazon region[53,54]. Studies have

demonstrated that in these "frontier" areas, *An. darlingi* numbers increase along-side malaria transmission[55]. Typically, *An. darlingi* breeds during the rainy season when riverbeds expand and become flooded, and during this time species abundance would peak along with malaria cases. However, *An. darlingi's* adaptability is leading to potential year-round malaria transmission, with human-made environmental changes resulting in permanently available larval habitats which could lead to perennial malaria transmission[46,56]. Published reports of the genetic diversity of *An. darlingi* are inconsistent, with countries such as Brazil reporting a high genetic diversity in the population, whilst others such as Colombia reporting the opposite[46,57]. Genetic diversity between different geographical locations has been demonstrated to be high, whilst within regions is low[58,59]. This difference could be the result of localised geographical isolation, due to geographic barriers such as the Amazon River or mountain ranges resulting in a lack of gene flow.

Limited work has been conducted on this vector, so our understanding of its genetic structure, and the mechanisms involved in insecticide resistance need further investigation.

Anopheles funestus

Anopheles funestus s.s (hereafter known as *An. funestus*), is a highly competent vector for both *P. falciparum* and *P. vivax*. This mosquito species is the dominant vector in the *An. funestus* species complex. The complex comprises of 9 species, of which four others have been implicated in the transmission of human malaria: *An. parensis*, *An. longipalpis, An. vaneedeni* and *An. rivulorum*[43,60]. *An. funestus* is the most abundant and widely distributed of the complex and is found across Sub-Saharan Africa (**Figure 3**). It is a primary vector in Tanzania, where it contributes to almost 90% of the malaria transmission burden even in areas with low abundance[61,62]. Other countries where it acts as a primary vector include

Kenya, Malawi, Burkina Faso, Cameroon, Zambia, and the Democratic Republic of the Congo[35].

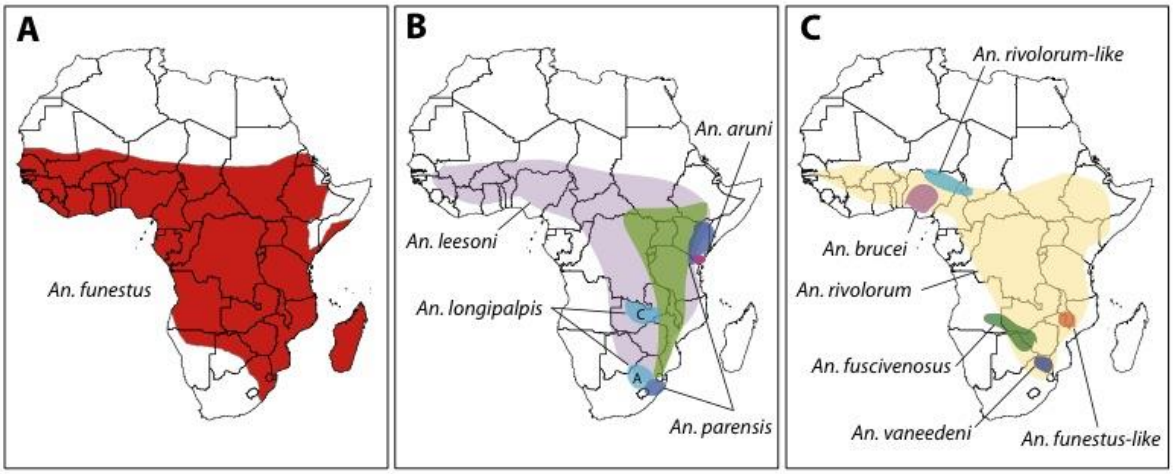


Figure 3. Geographical Distribution of the different members of the *An. funestus* species complex. Window A shows the widespread distribution of *An. funestus s.s.,* while window B and C indicate the distribution of other members of the *An. funestus* complex, of these *An. longipalpis, An. vandeeni, An. rivulorum* and *An. parensis* are the only species that have demonstrated malaria transmission[35,63,64].

This vector is anthropophilic with endophilic resting behaviour, and like *An. gambiae s.l.* predominantly bites during at night[65,66]. However, unlike *An. gambiae,* it prefers more permanent aquatic habitats for breeding, such as rivers, ponds, and swamps. This behaviour allows it to sustain population numbers, and malaria transmission, year-round[43,62]. It has been identified that after successful efforts to reduce *An. gambiae* populations, *An. funestus* and *An. rivulorum* often fill the ecological gap that remains[61,67,68].

An. funestus is predominantly a rural vector but has been found to contribute to malaria transmission in urban areas also[69]. It is the most geographically widespread species of its

complex due to its suitability across a wide range of climate conditions. This climate adaptability is believed to stem from the variety of ecotypes present, potentially arising from the high numbers of chromosomal inversions occurring in this species[60,65,70]. This adaptability is of significant concern as climate change impacts Sub-Saharan Africa, it could lead to a much larger geographical distribution for this vector. Other variations in phenotypic traits linked to these structural variants include mating behaviour, insecticide resistance, resting behaviour, and malaria transmission[70]. Analysis of *An. funestus* mitochondrial genes, *ND5* and *cox-1*, has indicated two distinct genetic lineages (I and II). One lineage was distributed across Sub-Saharan Africa, and the other in Southeast Africa throughout Madagascar, Mozambique and Tanzania, amongst others[60]. There were also genetic groupings visible from East, West and Central Africa indicating a limited gene flow between these populations[71,72]. Analysis of the entire mitochondrial genome also demonstrated two distinct lineages, as well as indicating species introgression within the *An. funestus* species complex[71,73,74]. The analysis suggested a significant genetic exchange (31.6 Mb) between *An. funestus s. s.* and *An. parensis*[74]. Although, the introgression of mitochondrial DNA can occur without the exchange of nuclear material[75,76].

An. funestus is one of the most studied *Anopheles* species, but more work is needed to identify insecticide resistance mechanisms and the vectoral capacity of species within the *An. funestus* complex.

Anopheles stephensi

Historically, *An. stephensi* has been found across South Asia and the Middle East[30]. A primary vector across large parts of India and Pakistan, *An. stephensi* is responsible for large amounts of *P. vivax* transmission in these regions. Three forms of *An. stephensi* have been identified

with varying vectorial capacities and geographical distributions. The forms are distinguishable through their morphology and molecularly with the *odorant binding protein*, intron 1[77]. The three forms are: type, intermediate, and mysorensis[78]. Type and intermediate forms are found in urban environments, predominantly feeding on humans, both are proficient vectors of *P. falciparum* and *P. vivax* [79]. Whereas the mysorensis form of *An. stephensi* is a poor vector for human malaria, mainly zoophagic, and is found in rural areas[80].

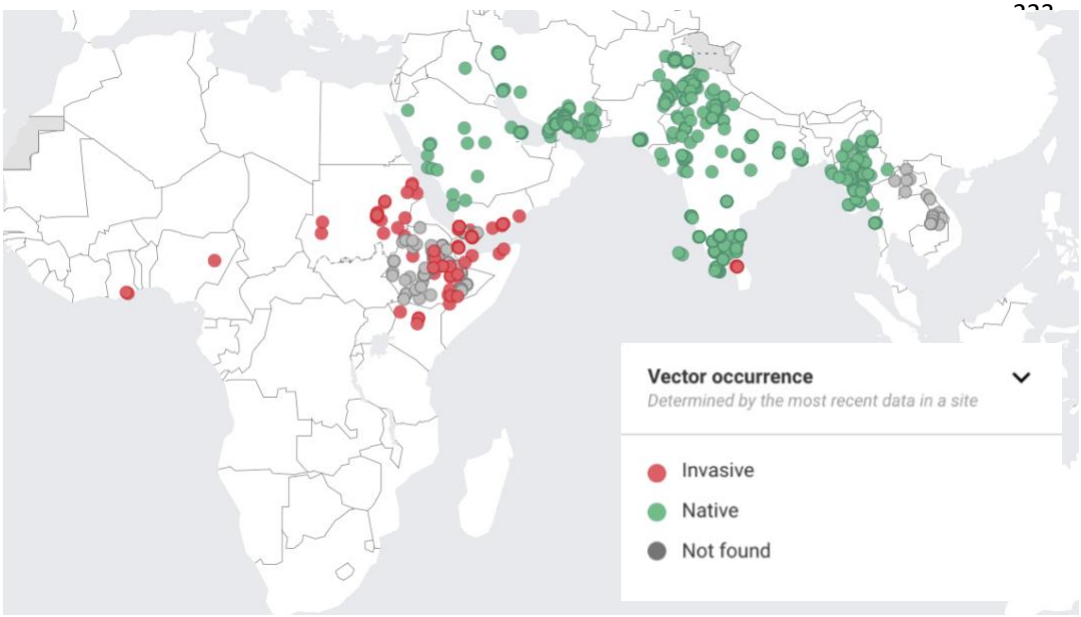


Figure 4. Geographical distribution of *An. stephensi* in native regions, and locations of identified invasions[81]

In 2012, *An. stephensi* was found in the Horn of Africa (HOA), specifically in Djibouti[82]. Between 2012 and 2017 malaria cases in Djibouti city increased five-fold, attributed to the arrival of this vector[39,83]. Since 2012, *An. stephensi* has migrated further within the continent (**Figure 4**) to Ethiopia, Sudan, Somalia, Nigeria and Kenya[81,84–87]. This spread has raised significant concerns for the eradication of malaria in the region, due to *An. stephensi's* ability

to transmit regional strains of *Plasmodium spp* so proficiently. Whilst *An. arabiensis* and *An. funestus* are associated with malaria transmission in rural regions of Sub-Saharan Africa, *An. stephensi*'s behaviour differs, and is regularly found in cities and towns where population density is higher, leaving more people are at risk from malaria infection[40,88,89]. Estimates have suggested *An. stephensi* could easily spread across Sub-Saharan Africa, putting an estimated 126 million new people are risk from infection[40]. *An. stephensi* was named on a WHO vector alert due to its expanding geographical distribution coupled with its high vectoral capacity[90]. Our understanding of the invasive African *An. stephensi* population is limited, and multiple theories have been presented concerning which area in Asia these mosquitoes originated from and how they entered the HOA. Initial theories suggested *An. stephensi* was imported from Asia through shipping lanes and ports, as *Aedes aegypti* and *Aedes albopictus* did[82,91,92]. Dijbouti city is only 5 miles from the coast, and as the site of original discovery implicates this as a probable importation event. Other theories have suggested the geographical proximity of the HOA to the Middle East meant the mosquitoes could have easily moved on their own[93]. Others have hypothesised this invasive population originated from India[94]. Sri Lanka has also recently experienced the introduction of this species - sparking concerns for malaria transmission due to its proximity to India. Genetic studies have demonstrated a high haplotype diversity of Ethiopian *An. stephensi,* which implies either a large initial invasion event or several small invasion events that have led to the long-term establishment of this species in the country[94]. The need to control this highly capable vector is clear and elucidating its origins may help us to better understand the threat it poses in the African continent.

Insecticides

Insecticides are often routinely used in countries at risk from malaria or other arthropod transmitted diseases (e.g., Dengue or Zika). Insecticides offer a low-cost method to reduce vector populations and thereby disease transmission. The use of insecticides varies between countries, and often by region within countries. Insecticides can be split into larvicides and adulticides. Larvicides target mosquito larval stages and are used for spraying of water bodies used as mosquito breeding grounds. Whereas adulticides target adult mosquitoes, and are used for IRS, and the impregnation of bed nets (LLINs).

There are six classes of adulticides used: pyrethroids, organochlorines, organophosphates, carbamates, neonicotinoids, and pyrroles. Each class of adulticides has a different mode of action, but typically target parts of the arthropods nervous system. Pyrethroids work by binding to the *voltage-gated sodium channel* (*vgsc*), preventing inactivation and causing consistent depolarisation of the membrane resulting in insect paralysis[95]. There has been research to suggest pyrethroids can also act on voltage-gated chloride and calcium channels[96,97]. The exact mode of action of organochlorines is unknown, but DDT-like organochlorines are believed to work by preventing sodium and calcium gated channel closure creating repeating neuronal discharges[98]. They also block the gamma-aminobutyric acid (*gaba*) receptor preventing GABA neurotransmission[99].

Organophosphates work by irreversibly phosphorylating residues on acetylcholinesterase's and thereby inactivating them and causing consistent muscle contractions[100,101]. Carbamates also inhibit acetylcholinesterase's but through carbamylation instead and is also reversible unlike organophosphate inhibition[101]. Both pyrroles and neonicotinoids represent new classes of insecticides, with no cross-resistance due to their modes of action[102–104].

Neonicotinoids work as agonists causing an irreversible blockade of nicotinergic acetylcholine receptors[105]. They are also a selective class of insecticide, as they only function in arthropods with specific acetylcholine receptors. Pyrroles are the only class where these chemicals act by disrupting mitochondrial function[106]. However there have been observed variations in efficacy across time of day, as pyrroles are activated by the process they inhibit: oxidative phosphorylation. This process is dependent on insect metabolism, which varies by climate and time of day. LLINs containing pyrroles are thereby particularly effective at night when the mosquito's circadian rhythm means metabolic activity is highest[107]. These newer classes of insecticides, alongside being unlikely to be affected by cross-resistance are also less toxic than the four traditionally used classes: organochlorines, organophosphates, carbamates, and pyrethroids[108,109].

Insecticide Resistance

Resistance to at least one class of insecticide has been reported in 90 countries globally. However, most countries have resistance to multiple insecticide classes, thereby threatening the basis of many malaria control programmes (**Figure 5**)[81].

Monitoring the emergence and frequency of resistance to each insecticide class is key to informing local public health policy to select the best method for vector control.

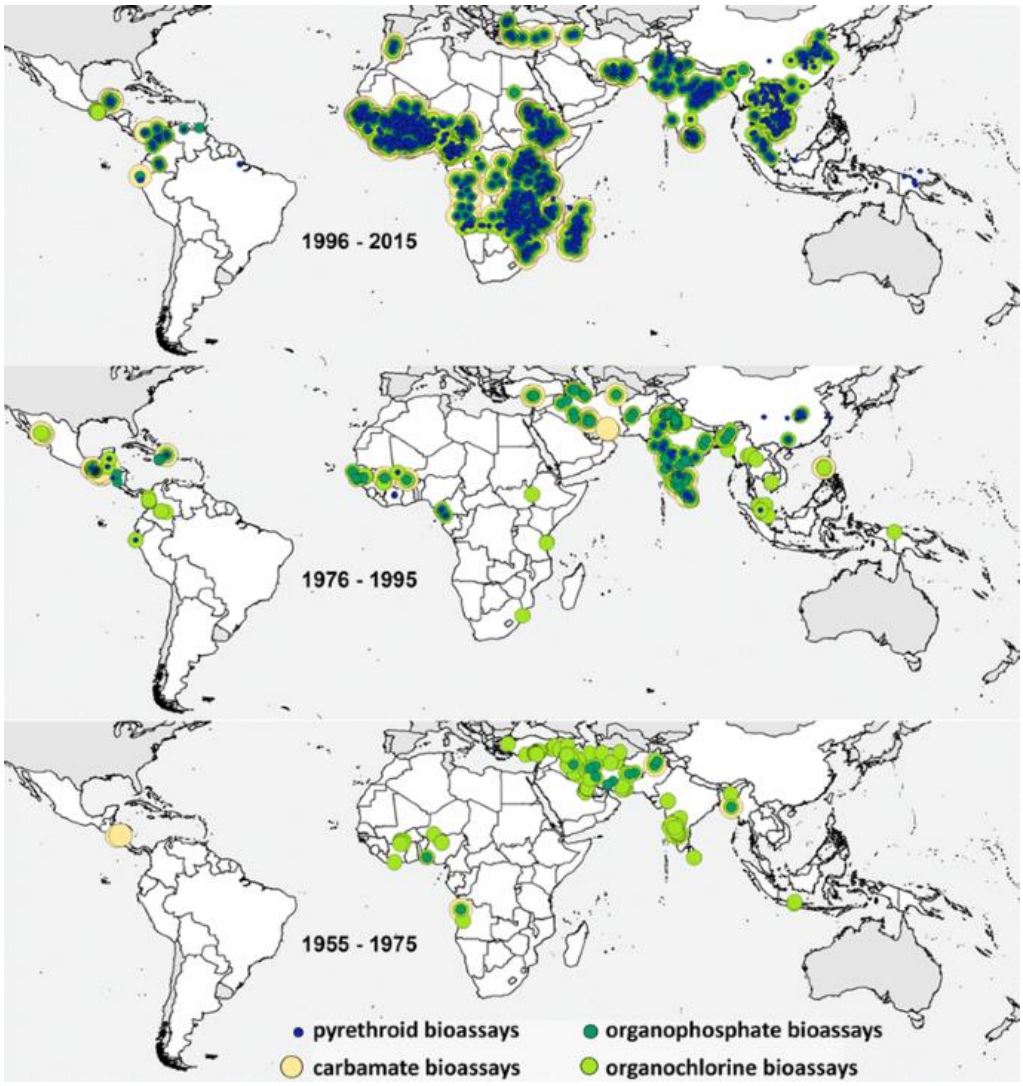


Figure 5. Identification of resistance to four main adulticides classes over time[110]. The increase in resistance across the globe in 60 years is a cause for concern due to the impact insecticides have has on reducing the malaria burden. There has been a clear increase in pyrethroid resistance globally, likely due to the heavy usage of this insecticide, but also increased bioassay testing.

The five main mechanisms that cause insecticide resistance, defined by WHO, are:

1. **Target site insensitivity:** The protein targeted by the insecticide has been structurally altered, so the insecticide is unable to bind to the protein in the intended way reducing its efficacy and resulting toxicity. These structural alterations are typically the result

of non-synonymous (NS) single nucleotide polymorphisms (SNPs) or insertions/deletions (INDELs) that cause changes to amino acid sequences[111,112].

2. **Metabolic resistance:** The upregulation of the activity of a detoxifying enzyme e.g., *cytochrome P450s* and *glutathione-S-transferases*, causing the insecticide to be metabolised faster making it less effective[111,112]. This can be driven by genetic polymorphisms in gene promoter regions, increased gene copy numbers or gene overexpression.
3. **Cuticle modification:** Insecticides are absorbed through the mosquito cuticle - a hard, protective layer surrounding its body. Increased cuticle thickness results in reduced insecticide penetration. Mosquito cuticles naturally become thicker with age but also can be caused by the overexpression of cuticular proteins[111,112].
4. **Avoidance**: Mosquitoes alter their behaviour to avoid insecticide sprayed surfaces, for example, living in cattle dwellings, avoiding landing on inside walls, or changing typical biting times[111,112].
5. **Microbiome Alterations**: Strains of bacteria commonly found in mosquitoes midgut microbiome appear capable of metabolising insecticides, thereby reducing the efficacy of the insecticide[113–115].

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

Target Site Resistance

Several missense SNPs associated with insecticide resistance have been identified in *Anopheles* mosquitoes. Such SNPs result in amino acid changes that can reduce the binding affinity or completely prevent a given insecticide from binding to its target protein.

The *kdr (knock-down resistance)* or L1014F/S alteration in the *vgsc* is the most researched of these mutations. This SNP results in pyrethroid resistance and is one of many missense SNPs observed in the *vgsc*[116]. Two variants, East (F) and West (S) have been identified, their names based on where they were first identified in East and West Africa[117,118]. However recently, the geographical distinction between the two has become muddled with the spread of both alleles across the African continent[119,120]. In fact, both these SNPs have been found in multiple *Anopheles spp.,* globally. For example, across the WHO Eastern Mediterranean and Central Asia regions, in species including *An. stephensi, An. sinensis,* and *An. culicifacies*[81,121–123]. It has also been observed in Central and South America in *An. albimanus*[124–126].

However, *kdr* mutations have not been universally characterised in all *Anopheles* species. This may be due to a lack of research as suspected with *An. darlingi,* or in the case of *An. funestus* there may be alternative molecular mechanisms at work (e.g., metabolic-based resistance).

Other SNPs that alter pyrethroid susceptibility in the *vgsc,* have predominantly been identified in *An. gambiae*. They include the V402L identified in *An. gambiae* and *An. coluzzii* in West Africa[127–129]. The F1529C mutation was first described in *An. gambiae* from Ivory Coast. The I1527C, N1575Y, and A1746S mutations were again found in *An. gambiae* in West and Central Africa.[129–131]. The high number of SNPs detected in *An. gambiae* is likely due to the extensive research performed on this vector, and it is likely these SNPs occur in other *Anopheles spp*.

The high number of SNPs identified in the *vgsc* may be due to the large size of the gene in comparison to the others, but also pyrethroids are the most widely used insecticides globally and so high selective pressure is applied to the gene resulting in these mutations.

Other commonly seen SNPs include the G119S mutation in the *ace-1* gene, resulting in organophosphate and carbamate resistance. This SNP is detected less frequently than the

kdr-L1014 mutation, but has been detected across Africa, East Asia and South America in *An. gambiae, An. sinensis, and An. albimanus* respectively[81,132–134]. The other *ace-1* SNP associated with carbamate resistance is the N485I alteration, which has only been reported in *An. funestus* thus far[135].

The A296S alteration in the *gaba* gene or the *rdl* (resistant to dieldrin) is seen frequently, and has been identified in *An. funestus, An. gambiae, An. stephensi, An. arabiensis,* and *An. sinensis* species[136–140]. Since the banning of dieldrin use due to human health concerns, the A296S mutation has become less relevant, but can still be found in recently collected mosquitoes indicating the fixation of this SNP in some populations[138]. This occurrence may be due to the application of pesticides such as fipronil that also act on *gaba* receptors. Other SNPs found in this gene include the V327I and T345S, both of which were identified in *An. sinensis*, and exist in tight linkage[136,137].

Metabolic Markers of Resistance

In instances where phenotypic resistance is present, but no target site mutations can be detected, metabolic mechanisms are often responsible. Metabolic-based resistance is mediated by different groups of enzymes – cytochrome P450's (CYP), glutathione S-transferases (GST), mixed-function oxidases, carboxy/cholinesterase's, and other non-specific esterase's[112,141–144]. Increased activity, or expression of these enzymes results in increased metabolism of insecticides thus reducing their efficacy.

Gene duplication events often result in the increased expression of these enzymes that cause resistance. Examples of this include the *ace-1* duplication in *An. gambiae s.s.*, which then introgressed into *An. coluzzi,* which causes resistance to pirimiphos-methyl (an organophosphate)[145,146]. Duplication of *CYP6P9* and *CYP6P4* cause pyrethroid resistance in

382 *An. funestus*[147]. In *An. stephensi,* duplications of *GST2* and *GST4* genes have been found and
383 require further investigation to be linked to a DDT resistant phenotype[148].
384 Other metabolically mediated mechanisms of resistance include the occurrence of amino acid
385 substitutions that result in increased insecticide breakdown. The L119F *GSTe2* mutation is an
386 example of this, where the mutation is coupled with over-transcription of the *GSTe2* gene,
387 and allows for increased DDT binding and metabolism, reducing the insecticide
388 effectiveness[149]. Originally identified in *An. funestus,* L119F has since been detected in *An.*
389 *gambiae* and *An. coluzzii* in Africa[129]. Similarly, the I114T and F120L mutations have also been
390 linked to DDT and pyrethroid resistance in *An. gambiae*[150].
391 There are further metabolic markers for resistance including the presence of mutations in the
392 *CYP6P9a* and *CYP6P9b* genes. In the *CYP6P9a* gene, a 2bp insertion was found 359bp from
393 the start codon and was associated with reduced efficacy of pyrethroid treated LLINs[151]. *Cis*-
394 regulatory variants were identified in the *CYP6P9b* gene were tightly linked with pyrethroid
395 resistance. A 3bp deletion 703bp upstream from the gene start codon can be used as a marker
396 for this resistance[152]. There has been substantial work conducted investigating the metabolic
397 basis of resistance in *An. funestus,* primarily due to the lack of *kdr* mutations found in the
398 species.
399 Similarly, in the *CYP6P4* gene in *An. gambiae,* the I236M mutation was strongly predictive of
400 pyrethroid resistance[153]. The L43F amino acid alteration in the *CYP4J5* gene is linked to
401 resistance to pyrethroids in *An. gambiae* and *An. funestus* also[154,155]. The contribution of
402 increased insecticide metabolism on resistance should not be underestimated. Studies have
403 found evidence of metabolic resistance in many species, including *An. darlingi, An. funestus,*
404 and *An. stephensi* amongst many others[156–158].
405

Detection of Insecticide Resistance

The current gold-standard for detecting insecticide resistance is through biochemical susceptibility assays or bioassays[111]. These tests are time consuming, require multiple repeats for validation, and can be insensitive. They often only detect resistance when the frequency of resistant phenotypes is already high. PCR-based assays are used to detect these molecular changes associated with insecticide resistance. However, these assays are often limited to single regions in particular genes or SNPs, so are difficult to implement as a cost-effective method of detecting resistance. Capillary or Sanger sequencing is used as a sequence-based method of insecticide resistance SNP detection but can be expensive for large numbers of samples, and again requires the SNP to be at a high enough allelic frequency for detection. So often these mutations are not identified until they are occurring at high frequencies in the population.

The confirmation of metabolic-based resistance is more difficult as it cannot be easily detected using classic sequencing or PCR methods. The development of qPCR assays are an easy way to measure expression of these enzymes, but this requires knowledge of the genes involved in resistance, and for many species these mechanisms are still unknown[159,160]. The use of WGS and RNA-sequencing (RNA-seq) has helped to identify many more duplications and overexpression events in *Anopheles spp.* that can be linked to insecticide resistance, but these approaches are expensive[161,162].

The monitoring of insecticide resistance is key component in reducing malaria transmission. The development of high-throughput molecular assays to detect resistance with potential in-field applications could help inform insecticide usage in real-time to optimise vector control. These assays used in tandem with phenotypic bioassays, can also support the identification of new genetic variants associated with resistance.

Genomics of *Anopheles spp.*

The development of high-throughput molecular assays for the detection of insecticide resistance markers, requires a robust understanding of the organism's genomic landscape. The first *An. gambiae* genome was published in 2002, and since then 38 other *Anopheles spp* now have reference genomes assembled to at least contig level[163]. The *Anopheles spp* genome is made up of 3 chromosomes, along with a mitochondrial genome. The length of the genome varies dependent on which species, with the most complete assembled *Anopheles* genome is *An. gambiae,* having a total length of 265Mbp. Of the three *Anopheles spp.* examined here, *An. darlingi* has the smallest genome at 181.6Mbp, while *An. funestus s.s* and *An. stephensi* have genome lengths more similar to *An. gambiae* at 250.7Mb, and 243.5Mb respectively[164,165]. The variations in length are predominantly due to differences in intronic variants, intergenic regions, and the number of transposable elements present across the genomes.

All three of these species have chromosomal level assemblies, which gives a more complete view of the organism's genome. The previous *An. darlingi* reference genome was assembled to a contig level and was 60Mbp smaller than the current reference grade assembly available[166].

Low quality reference genomes often contained mis-assembled regions or are missing genes entirely. This poor-quality impacts on the identification of genetic elements that are potentially functionally relevant for malaria transmission or insecticide resistance. Genomic regions with copy number variants or have highly repetitive sequences are often missed or incorrectly assembled. These sequences often play important functional roles or help shape the genetic diversity of organisms and are key to understanding population dynamics and the emergence of resistance. **Figure 6** demonstrates the genetic diversity visible from just the mitochondrial genomes of a few *Anopheles* species.

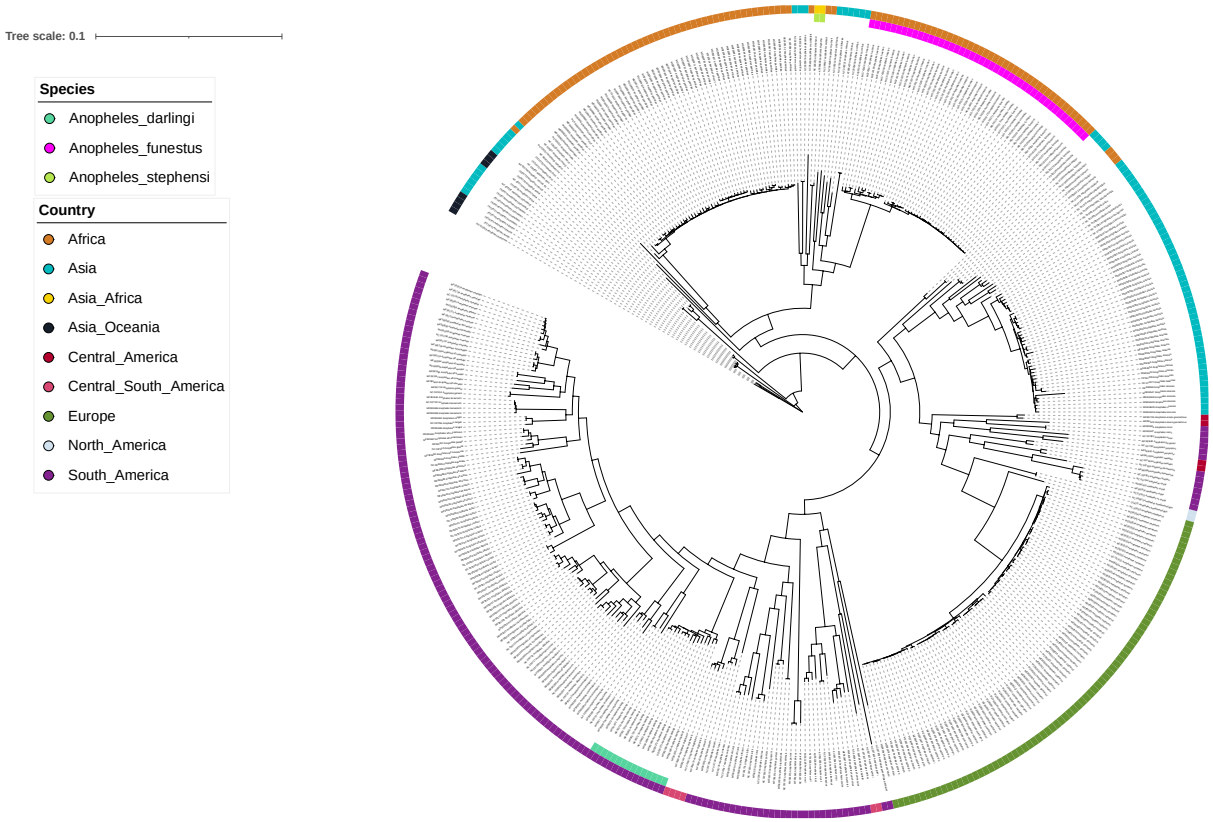


Figure 6. Maximum-likelihood tree generated from all available *Anopheles* species mitochondrial genomes. Sequences were aligned in MAFFT software. The tree was created using RAxML software with 1000 bootstraps, and visualised in iTOL[167–169].

Large scale population genetics studies conducted on *An. gambiae, An. funestus, An. minimus*, *An. sinensis* and *An. coluzzii* have helped to improve our understanding of these mosquitoes,

462 demonstrating the high genetic diversity of these vector populations and identifying genomic
463 regions involved in insecticide resistance under selective pressure[162,170–174].
464 However, there are limited sequences available for other species. For example, for *An.*
465 *stephensi* there are <50 available samples with WGS data, and there is no such data for *An.*
466 *darlingi*. This leads to large knowledge gaps and limits our understanding of these vectors.
467 The limited WGS data available for some *Anopheles spp*., is predominantly due to the high
468 cost of applying next-generation sequencing techniques along with the need for high
469 quantities of DNA from single mosquitoes to enable good coverage over their large genomes.
470 Targeted amplicon sequencing ("amp-seq") provides an alternative to WGS. This technique
471 applies next-generation sequencing to PCR amplicons, coupled with a dual-index barcoding
472 system to provide a high-throughput, low-cost screening method targeted to multiple
473 genomic loci. The cost of amp-seq per amplicon works out around $0.50 per amplicon due to
474 the ability to pool samples, in comparison capillary sequencing costs ~$4 an amplicon. It can
475 be used to screen insecticide resistance associated SNPs in candidate genes, or to gain insight
476 into population structure. This approach has previously been applied to *An. gambiae, An.*
477 *coluzzii,* and *Ae. aegypti*[129,175,176]*.* Amp-seq can be conducted on both short-read Illumina
478 platforms, and long-read Oxford Nanopore Technology platforms (e.g., MinION). Assays
479 based on amplicon sequencing have the potential to identify emerging resistance, by
480 detecting low frequency alleles, and to support population genetic analysis of vectors to
481 understand how resistance genotypes spread throughout *Anopheles* populations.

Hypotheses

Anopheles stephensi

Resistance to pyrethroids, carbamates, organochlorides, and organophosphates was found in *An. stephensi* from Awash Sebat Kilo in the same year (2019) these isolates were collected[81]. So, it is highly likely the samples analysed in this research were phenotypically resistant to a range of insecticides. Given that target site mutations such as *kdr* have previously been identified in resistant Indian and Afghanistan *An. stephensi,* it is expected these SNPs would be found here also[122,177–179]. Ethiopia relies heavily on pyrethroids for vector control via IRS and LLINs, so there would likely be large amounts of selective pressure on this mosquito to evolve resistance[180]. Although, it could be argued *An. stephensi*'s exposure to these chemicals may be limited due to its crepuscular biting behaviour and propensity to rest indoors in cattle dwelling, but this is clearly not the case as phenotypic studies have demonstrated[81,181,182].

Anopheles darlingi

Little is known about the resistance status of *An. darlingi,* particularly in Brazil. Phenotypic studies are either sparsely done, or not reported. Resistance to pyrethroids (cypermethrin and deltamethrin) has been reported in Cruzeiro du Sol, State of Acre Brazil, but this is the only data available for Brazilian *An. darlingi*[81]. Cruzeiro du sol is >1000km from the collection sites of the isolates used in these studies, so gene flow between these populations is likely limited. Only anecdotal information is available for these isolates about the likelihood of insecticide exposure. In Porto Velho insecticides are routinely used, whereas in Candeias dos Jamari there was minimal usage. As discussed earlier, *An. darlingi* displays great behavioural plasticity with variations in resting and feeding behaviour present even in locations within small distances. As a vector, it is highly anthropophilic exhibiting unimodal, bimodal, and even

trimodal biting patterns. Biting times vary based on location tested, and even within Brazil variation was seen between endo/exophagic and endo/exophilic biting[53]. This behavioural variation makes it difficult to speculate on what will be observed on a phenotypic and molecular level in these *An. darlingi* isolates. This is exacerbated by the lack of molecular markers found in *An. darlingi* populations, these studies have been small and sparse so may have missed low frequency SNPs that appear as a heterozygous genotype[183,184]. But the application of WGS and amp-seq to these isolates may yet find known markers of insecticide resistance in this species or find putatively novel mutations.

Anopheles funestus

The phenotypic status of the isolates tested were unknown, but resistance to pyrethroids has been detected in the province these isolates were collected in[81]. *An. funestus* typically exhibits bimodal biting behaviour, at night and in the morning, although recent reports have suggested it preferentially bites in the early morning[185,186]. Such behaviour would mean they would likely be exposed to insecticide treated nets. This mosquito also tends to rest and feed indoors, indicating if IRS is carried out, they would likely be exposed to insecticides. Unlike *An. stephensi,* and *An. darlingi,* this species behaves in way that would indicate they would routinely be exposed to insecticides and as such would have a higher chance of being resistant. As demonstrated by the previously testing carried out on thesis isolates using established methods such as RFLP, 82% of isolates analysed here were found to have the *CYP6P9a* marker for pyrethroid resistance[155]. As such we would expect to find this marker at a similar frequency, alongside other novel SNPs.

Thesis Structure

In this thesis, I investigate the genomics of three *Anopheles spp.*: *An. darlingi, An. funestus,* and *An. stephensi*. I developed a species-specific amplicon sequencing assay for each species. In addition, I applied whole genome sequencing methods to *An. darlingi* and *An. stephensi* and used the resulting data to investigate the genetic diversity and insecticide resistance status of these vectors.

Chapter 2 (Published paper)

In chapter 2, I develop a high-throughput, multiplexable amplicon sequencing assay for *An. stephensi*. This assay targets nine genomic regions, seven within four genes (*ace-1, GSTe2, vgsc,* and *rdl)* associated with insecticide resistance. The remaining two targeted genes, *cox-1* and *ITS2*, were chosen for species identification and/or phylogenetic analysis. The panel was applied to Ethiopian field isolates and SDA500 colony strains of *An. stephensi*. We identified two known molecular markers for insecticide resistance in the invasive Ethiopian mosquitoes, as well as putatively novel missense mutations in genes of interest. Shared haplotypes between *An. stephensi* from Ethiopian and Pakistan were also identified.

Chapter 3 (Manuscript yet to be submitted)

Here, for the first time, I apply WGS to field isolates of *An. stephensi* from Ethiopia and SDA500 colony strains. In the context of publicly available Indian and Pakistani isolates, I utilised population genetics analysis methods such as phylogenetics, principal component analysis, and admixture, to unravel the population dynamics of *An. stephensi*. I identified 3 molecular markers of insecticide resistance and gained greater resolution into the ancestry of the

552 invasive Ethiopian populations of *An. stephensi*, including identifying shared ancestry with
553 west coast Indian field and Pakistani colony isolates.
554
555 ## Chapter 4 (Manuscript submitted, under review)
556 In chapter 4, I design a next-generation sequencing amplicon assay specifically for *An.*
557 *darlingi.* This assay consists of 11 amplicons, covering four key genes (*ace-1, GSTe2, vgsc,* and
558 *rdl*) associated with insecticide resistance, and two loci (*cox-1* and *ITS2*) informative for
559 speciation and phylogenetic analysis. The amplicons are multiplexable and can be sequenced
560 in a high throughput manner using dual-index barcoding. This approach resulted in the
561 identification of 10 putatively novel missense mutations in *An. darlingi* (n=200) collected
562 within the state of Rondônia, Brazil, as well as insights into the genetic diversity in that region.
563
564 ## Chapter 5 (Manuscript yet to be submitted)
565 This chapter contains the first instance of WGS being applied to *An. darlingi.* Isolates collected
566 from the State of Rondônia, were sequenced using Illumina technology, and population
567 genomic analyses were performed on the resulting data. Genome-wide tests of selective
568 pressure were applied to identify regions that could be linked to insecticide resistance. The
569 molecular mechanisms of resistance in *An. darlingi* are poorly understood, and this analysis
570 identified several regions including *CYP* genes with recent or ongoing selection pressure.
571
572 ## Chapter 6 (Manuscript submitted, under review)
573 In this chapter, I design an amplicon sequencing panel for *An. funestus*, suitable for next
574 generation sequencing and dual-tandem barcoding. The panel targets 17 amplicons, across
575 nine genes; six of which have been previously associated with insecticide resistance, and a

576 further three for phylogenetic inference and speciation. The amp-seq assay was used on *An.*
577 *funestus* isolates collected in the DRC. Two markers of resistance were identified, one
578 associated with pyrethroid resistance, and the other with carbamate resistance. We identified
579 a further 18 non-synonymous SNPs in insecticide resistance associated genes and gained
580 insight into the genetic diversity of this population.

Thesis Chapter	**Manuscript Title**	**Authors**	**Status**
Chapter 2	Identification of two insecticide resistance markers in Ethiopian *Anopheles stephensi* mosquitoes using a multiplex amplicon sequencing assay	**Holly Acford-Palmer,** Jody E. Phelan, Fitsum G. Tadesse, Mojca Kristan, Emma Collins, Anton Spadar, Thomas Walker, Teun Bousema, Louisa A. Messenger, Taane G. Clark & Susana Campino	Published: *Sci Rep.* 2023; April 5; 13: 5612
Chapter 3	Genome-wide population genetics, and molecular surveillance of insecticide resistance in Ethiopian *Anopheles stephensi*	**Holly Acford-Palmer**, Emilia Manko, Jody Phelan, Fitsum G. Tadesse, Mojca Kristan, Ashley Osborne, Matthew Higgins, Teun Bousema, Louisa A. Messenger, Taane G. Clark & Susana Campino	To submit
Chapter 4	Application of a targeted amplicon sequencing panel to screen for insecticide resistance mutations in *Anopheles darlingi* populations from Brazil	**Holly Acford-Palmer,** Alice O. Andrade, Jody E. Phelan, Rosa A. Santana, Stefanie C.P. Lopes, Jansen F. Medeiros, Taane G. Clark, Maisa S. Araujo, Susana Campino	Submitted
Chapter 5	Genome -wide population genetics of *Anopheles darlingi*, a malaria vector from the State of Rondônia, Brazil	**Holly Acford-Palmer**, Emilia Manko, Alice O. Andrade, Jody Phelan, Rosa A. Santana, Stefanie C. P. Lopes, Matthew Higgins, Jansen F Medeiros, Taane G. Clark, Maisa S. Araujo, Susana Campino	To submit
Chapter 6	Detection of insecticide resistance markers in *Anopheles funestus* from the Democratic Republic of the Congo using a targeted amplicon sequencing panel	**Holly Acford-Palmer**, Monica Campos, Janvier Bandibabone, Sévérin N'Do, Chimanuka Bantuweko, Bertin Zawadi, Thomas Walker, Jody E. Phelan, Louisa A. Messenger, Taane G. Clark, Susana Campino	Submitted

References

1. Global Malaria Programme: WHO Global. *WHO World Malaria Report 2022*. *WHO*.

2. Howes, R. E. *et al.* The global distribution of the Duffy blood group. *Nat. Commun.* **2**, 266 (2011).

3. Chitnis, C. E. & Sharma, A. Targeting the Plasmodium vivax Duffy-binding protein. *Trends Parasitol.* **24**, 29–34 (2008).

4. Horuk, R. *et al.* A receptor for the malarial parasite Plasmodium vivax: the erythrocyte chemokine receptor. *Science* **261**, 1182–1184 (1993).

5. Ménard, D. *et al.* Plasmodium vivax clinical malaria is commonly observed in Duffy-negative Malagasy people. *Proc. Natl. Acad. Sci.* **107**, 5967–5971 (2010).

6. Gunalan, K. *et al.* Role of Plasmodium vivax Duffy-binding protein 1 in invasion of Duffy-null Africans. *Proc. Natl. Acad. Sci.* **113**, 6271–6276 (2016).

7. Gruszczyk, J. *et al.* Transferrin receptor 1 is a reticulocyte-specific receptor for Plasmodium vivax. *Science* **359**, 48–55 (2018).

8. Battle, K. E. & Baird, J. K. The global burden of Plasmodium vivax malaria is obscure and insidious. *PLOS Med.* **18**, e1003799 (2021).

9. Human Infections and Detection of Plasmodium knowlesi - PMC. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3623376/.

10. William, T. *et al.* Changing epidemiology of malaria in Sabah, Malaysia: increasing incidence of Plasmodium knowlesi. *Malar. J.* **13**, 390 (2014).

11. CDC. Malaria Lifecycle. *https://www.cdc.gov/malaria/about/biology/index.html* (2020).

12. 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).

13. Collins, W. E. & Jeffery, G. M. Plasmodium ovale: Parasite and Disease. *Clin. Microbiol. Rev.* **18**, 570–581 (2005).

14. Howes, R. E. *et al.* Global Epidemiology of Plasmodium vivax. *Am. J. Trop. Med. Hyg.* **95**, 15–34 (2016).

15. Battle, K. E. *et al.* Geographical variation in Plasmodium vivax relapse. *Malar. J.* **13**, 144 (2014).

16. Aly, A. S. I., Vaughan, A. M. & Kappe, S. H. I. Malaria Parasite Development in the Mosquito and Infection of the Mammalian Host. *Annu. Rev. Microbiol.* **63**, 195–221 (2009).

17. Bhatt, S. *et al.* The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. *Nature* **526**, 207–211 (2015).

18. Walker, T. *et al.* Stable high-density and maternally inherited Wolbachia infections in Anopheles moucheti and Anopheles demeilloni mosquitoes. *Curr. Biol.* **31**, 2310-2320.e5 (2021).

19. Gomes, F. M. *et al.* Effect of naturally occurring Wolbachia in Anopheles gambiae s.l. mosquitoes from Mali on Plasmodium falciparum malaria transmission. *Proc. Natl. Acad. Sci.* **114**, 12566–12571 (2017).

20. Gomes, F. M. & Barillas-Mury, C. Infection of anopheline mosquitoes with Wolbachia: Implications for malaria control. *PLoS Pathog.* **14**, e1007333–e1007333 (2018).

21. Klein, T. A., Windbichler, N., Deredec, A., Burt, A. & Benedict, M. Q. Infertility resulting from transgenic I-PpoI male Anopheles gambiae in large cage trials. *Pathog. Glob. Health* **106**, 20–31 (2012).

22. Nolan, T. *et al.* Developing transgenic Anopheles mosquitoes for the sterile insect technique. *Genetica* **139**, 33–39 (2011).

23. Anguelov, R., Dumont, Y. & Lubuma, J. Mathematical modeling of sterile insect technology for control of anopheles mosquito. *Comput. Math. Appl.* **64**, 374–389 (2012).

24. Global technical strategy for malaria 2016-2030, 2021 update. https://www.who.int/publications-detail-redirect/9789240031357.

25. Weiss, D. J. *et al.* Indirect effects of the COVID-19 pandemic on malaria intervention coverage, morbidity, and mortality in Africa: a geospatial modelling analysis. *Lancet Infect. Dis.* **21**, 59–69 (2021).

26. Balikagala, B. *et al.* Evidence of Artemisinin-Resistant Malaria in Africa. *N. Engl. J. Med.* **385**, 1163–1171 (2021).

27. Phyo, A. P., Nosten, F., Phyo, A. P. & Nosten, F. The Artemisinin Resistance in Southeast Asia: An Imminent Global Threat to Malaria Elimination. in *Towards Malaria Elimination - A Leap Forward* (IntechOpen, 2018). doi:10.5772/intechopen.76519.

28. Uwimana, A. *et al.* Emergence and clonal expansion of in vitro artemisinin-resistant Plasmodium falciparum kelch13 R561H mutant parasites in Rwanda. *Nat. Med.* **26**, 1602–1608 (2020).

29. Threats to the effectiveness of insecticide-treated bednets for malaria control: thinking beyond insecticide resistance - The Lancet Global Health. https://www.thelancet.com/journals/langlo/article/PIIS2214-109X(21)00216-3/fulltext.

30. Sinka, M. E. *et al.* A global map of dominant malaria vectors. *Parasit. Vectors* **5**, 69 (2012).

31. Rezza, G., Chen, R. & Weaver, S. C. O'nyong-nyong fever: a neglected mosquito-borne viral disease. *Pathog. Glob. Health* **111**, 271–275 (2017).

32. Hay, S. I. *et al.* A World Malaria Map: Plasmodium falciparum Endemicity in 2007. *PLOS Med.* **6**, e1000048 (2009).

33. Battle, K. E. *et al.* Mapping the global endemicity and clinical burden of Plasmodium vivax, 2000-2017: a spatial and temporal modelling study. *The Lancet* **394**, 332–343 (2019).

34. Impact of climate change on global malaria distribution | PNAS. https://www.pnas.org/doi/full/10.1073/pnas.1302089111.

35. Sinka, M. E. *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* **3**, 117 (2010).

36. Loughlin, S. O. The expanding Anopheles gambiae species complex. *Pathog. Glob. Health* **114**, 1 (2020).

37. 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).

38. Fontaine, M. C. *et al.* Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science* **347**, 1258524 (2015).

39. Seyfarth, M., Khaireh, B. A., Abdi, A. A., Bouh, S. M. & Faulde, M. K. Five years following first detection of Anopheles stephensi (Diptera: Culicidae) in Djibouti, Horn of Africa: populations established—malaria emerging. *Parasitol. Res.* **118**, 725–732 (2019).

40. Sinka, M. E. *et al.* A new malaria vector in Africa: Predicting the expansion range of Anopheles stephensi and identifying the urban populations at risk. *Proc. Natl. Acad. Sci.* **117**, 24900 LP – 24908 (2020).

41. Sinka, M. E. *et al.* The dominant Anopheles vectors of human malaria in the Asia-Pacific region: occurrence data, distribution maps and bionomic précis. *Parasit. Vectors* **4**, 89 (2011).

42. Sinka, M. E. *et al.* The dominant Anopheles vectors of human malaria in the Americas: occurrence data, distribution maps and bionomic précis. *Parasites Vectors 2010 31* **3**, 1–26 (2010).

43. Kahamba, N. F. *et al.* Using ecological observations to improve malaria control in areas where Anopheles funestus is the dominant vector. *Malar. J.* **21**, 158 (2022).

44. O'Loughlin, S. M. *et al.* Genomic signatures of population decline in the malaria mosquito Anopheles gambiae. *Malar. J.* **15**, 182 (2016).

45. Bayoh, M. N. *et al.* Anopheles gambiae: historical population decline associated with regional distribution of insecticide-treated bed nets in western Nyanza Province, Kenya. *Malar. J.* **9**, 62 (2010).

46. Moutinho, P. R., Gil, L. H. S., Cruz, R. B. & Ribolla, P. E. M. Population dynamics, structure and behavior of Anopheles darlingi in a rural settlement in the Amazon rainforest of Acre, Brazil. *Malar. J.* **10**, 1–13 (2011).

47. Moreno, M. *et al.* Intensive trapping of blood-fed Anopheles darlingi in Amazonian Peru reveals unexpectedly high proportions of avian blood-meals. *PLoS Negl. Trop. Dis.* **11**, e0005337 (2017).

48. Barbosa, L. M. C., Souto, R. N. P., dos Anjos Ferreira, R. M. & Scarpassa, V. M. Behavioral patterns, parity rate and natural infection analysis in anopheline species involved in the transmission of malaria in the northeastern Brazilian Amazon region. *Acta Trop.* **164**, 216–225 (2016).

49. Fornadel, C. M., Norris, L. C., Glass, G. E. & Norris, D. E. Analysis of Anopheles arabiensis Blood Feeding Behavior in Southern Zambia during the Two Years after Introduction of Insecticide-Treated Bed Nets. *Am. J. Trop. Med. Hyg.* **83**, 848–853 (2010).

50. Debebe, Y., Hill, S. R., Tekie, H., Ignell, R. & Hopkins, R. J. Shady business: understanding the spatial ecology of exophilic Anopheles mosquitoes. *Malar. J.* **17**, 351 (2018).

51. Carlos, B. C., Rona, L. D. P., Christophides, G. K. & Souza-Neto, J. A. A comprehensive analysis of malaria transmission in Brazil. *Pathog. Glob. Health* **113**, 1–13 (2019).

52. Siqueira, A. M. *et al.* Plasmodium vivax Landscape in Brazil: Scenario and Challenges. *Am. J. Trop. Med. Hyg.* **95**, 87–96 (2016).

53. Hiwat, H. & Bretas, G. Ecology of Anopheles darlingi Root with respect to vector importance: a review. *Parasit. Vectors* **4**, 177 (2011).

54. Charlwood, J. D. Biological variation in Anopheles darlingi root. *Mem. Inst. Oswaldo Cruz* **91**, 391–398 (1996).

55. Da Silva-Nunes, M. *et al.* Malaria on the amazonian frontier: Transmission dynamics, risk factors, spatial distribution, and prospects for control. *Am. J. Trop. Med. Hyg.* **79**, 624–635 (2008).

56. Angêlla, A. F. *et al.* Seasonal genetic partitioning in the neotropical malaria vector, Anopheles darlingi. *Malar. J.* **13**, 1–10 (2014).

57. Mirabello, L. & Conn, J. E. Molecular population genetics of the malaria vector Anopheles darlingi in Central and South America. *Heredity* **96**, 311–321 (2006).

58. Emerson, K. J., Conn, J. E., Bergo, E. S., Randel, M. A. & Sallum, M. A. M. Brazilian Anopheles darlingi Root (Diptera: Culicidae) Clusters by Major Biogeographical Region. *PLOS ONE* **10**, e0130773 (2015).

59. Pedro, P. M. & Sallum, M. A. M. Spatial expansion and population structure of the neotropical malaria vector, Anopheles darlingi (Diptera: Culicidae). *Biol. J. Linn. Soc.* **97**, 854–866 (2009).

60. Odero, J. O. *et al.* Advances in the genetic characterization of the malaria vector, Anopheles funestus, and implications for improved surveillance and control. *Malar. J.* **22**, 230 (2023).

61. Kaindoa, E. W. *et al.* Interventions that effectively target Anopheles funestus mosquitoes could significantly improve control of persistent malaria transmission in south–eastern Tanzania. *PLOS ONE* **12**, e0177807 (2017).

62. Mapua, S. A. *et al.* Persistently high proportions of plasmodium-infected Anopheles funestus mosquitoes in two villages in the Kilombero valley, South-Eastern Tanzania. *Parasite Epidemiol. Control* **18**, e00264 (2022).

63. Manguin, S. *Anopheles Mosquitoes - New Insights into Malaria Vectors*. (2013). doi:10.5772/3392.

64. Ibrahima Dia, Moussa Wamdaogo Guelbeogo, & Diego Ayala. Advances and Perspectives in the Study of the Malaria Mosquito Anopheles funestus | IntechOpen. in.

65. Coetzee, M. & Fontenille, D. Advances in the study of Anopheles funestus, a major vector of malaria in Africa. *Insect Biochem. Mol. Biol.* **34**, 599–605 (2004).

66. Coetzee, M. & Koekemoer, L. L. Molecular Systematics and Insecticide Resistance in the Major African Malaria Vector Anopheles funestus. *Annu. Rev. Entomol.* **58**, 393–412 (2013).

67. Lwetoijera, D. W. *et al.* Increasing role of Anopheles funestus and Anopheles arabiensis in malaria transmission in the Kilombero Valley, Tanzania. *Malar. J.* **13**, 331 (2014).

68. Kawada, H. *et al.* Reconsideration of Anopheles rivulorum as a vector of Plasmodium falciparum in western Kenya: some evidence from biting time, blood preference, sporozoite positive rate, and pyrethroid resistance. *Parasit. Vectors* **5**, 230 (2012).

69. Djamouko-Djonkam, L. *et al.* Implication of Anopheles funestus in malaria transmission in the city of Yaoundé, Cameroon. *Parasite* **27**, 10.

738 70. Ayala, D., Ullastres, A. & González, J. Adaptation through chromosomal inversions in
739 Anopheles. *Front. Genet.* **5**, (2014).

740 71. Michel, A. P. *et al.* Rangewide population genetic structure of the African malaria vector
741 Anopheles funestus. *Mol. Ecol.* **14**, 4235–4248 (2005).

742 72. Choi, K. S., Koekemoer, L. L. & Coetzee, M. Population genetic structure of the major malaria
743 vector Anopheles funestus s.s. and allied species in southern Africa. *Parasit. Vectors* **5**, 283
744 (2012).

745 73. Jones, C. M. *et al.* Complete Anopheles funestus mitogenomes reveal an ancient history of
746 mitochondrial lineages and their distribution in southern and central Africa. *Sci. Rep.* **8**, 9054
747 (2018).

748 74. Small, S. T. *et al.* Radiation with reticulation marks the origin of a major malaria vector. *Proc.*
749 *Natl. Acad. Sci.* **117**, 31583–31590 (2020).

750 75. Bachtrog, D., Thornton, K., Clark, A. & Andolfatto, P. EXTENSIVE INTROGRESSION OF
751 MITOCHONDRIAL DNA RELATIVE TO NUCLEAR GENES IN THE DROSOPHILA YAKUBA SPECIES
752 GROUP. *Evolution* **60**, 292–302 (2006).

753 76. Toews, D. P. L. & Brelsford, A. The biogeography of mitochondrial and nuclear discordance in
754 animals. *Mol. Ecol.* **21**, 3907–3930 (2012).

755 77. Gholizadeh, S. *et al.* The Anopheles stephensi odorant binding protein 1 (AsteObp1) gene: A
756 new molecular marker for biological forms diagnosis. *Acta Trop.* **146**, 101–113 (2015).

757 78. Vipin, Dube, M. & Gakhar, S. K. Genetic differentiation between three ecological variants
758 ('type', 'mysorensis' and 'intermediate') of malaria vector Anopheles stephensi (Diptera:
759 Culicidae). *Insect Sci.* **17**, 335–343 (2010).

79. Alam, M. T., Bora, H., Das, M. K. & Sharma, Y. D. The type and mysorensis forms of the Anopheles stephensi (Diptera: Culicidae) in India exhibit identical ribosomal DNA ITS2 and domain-3 sequences. *Parasitol. Res.* **103**, 75 (2008).

80. Chavshin, A. R. *et al.* Molecular characterization, biological forms and sporozoite rate of Anopheles stephensi in southern Iran. *Asian Pac. J. Trop. Biomed.* **4**, 47–51 (2014).

81. WHO. Global Malaria Programme: Malaria Threats Map. https://www.who.int/teams/global-malaria-programme/surveillance/malaria-threats-map.

82. 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 Trop.* **139**, 39–43 (2014).

83. Tadesse, F. G. *et al.* Anopheles stephensi Mosquitoes as Vectors of Plasmodium vivax and falciparum, Horn of Africa, 2019. *Emerg. Infect. Dis.* **27**, 603–607 (2021).

84. Balkew, M. *et al.* An update on the distribution, bionomics, and insecticide susceptibility of Anopheles stephensi in Ethiopia, 2018–2020. *Malar. J.* **20**, 263 (2021).

85. 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. *Parasit. Vectors* **14**, 511 (2021).

86. Ahmed, A. *et al.* Invasive Malaria Vector Anopheles stephensi Mosquitoes in Sudan, 2016–2018. *Emerg. Infect. Dis.* **27**, 2952–2954 (2021).

87. Adeogun, A. *et al.* Spatial distribution and ecological niche modeling of geographical spread of Anopheles gambiae complex in Nigeria using real time data. *Sci. Rep.* **13**, 13679 (2023).

88. Messenger, L. A. *et al.* Insecticide resistance in Anopheles arabiensis from Ethiopia (2012–2016): a nationwide study for insecticide resistance monitoring. *Malar. J.* **16**, 469 (2017).

89. 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. *Malar. J.* **22**, 187 (2023).

90. Vector alert: Anopheles stephensi invasion and spread.

91. Ahn, J., Sinka, M., Irish, S. & Zohdy, S. Modeling marine cargo traffic to identify countries in Africa with greatest risk of invasion by Anopheles stephensi. *Sci. Rep.* **13**, 876 (2023).

92. Lounibos, L. P. Invasions by Insect Vectors of Human Disease. *Annu. Rev. Entomol.* **47**, 233–266 (2002).

93. Lehmann, T. *et al.* Urban malaria may be spreading via the wind—here's why that's important. *Proc. Natl. Acad. Sci.* **120**, e2301666120 (2023).

94. Carter, T. E. *et al.* Genetic diversity of Anopheles stephensi in Ethiopia provides insight into patterns of spread. *Parasit. Vectors* **14**, 602 (2021).

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

96. Forshaw, P. J., Lister, T. & Ray, D. E. The role of voltage-gated chloride channels in type II pyrethroid insecticide poisoning. *Toxicol. Appl. Pharmacol.* **163**, 1–8 (2000).

97. Cao, Z., Shafer, T. J. & Murray, T. F. Mechanisms of Pyrethroid Insecticide-Induced Stimulation of Calcium Influx in Neocortical Neurons. *J. Pharmacol. Exp. Ther.* **336**, 197–205 (2011).

98. Coats, J. R. Mechanisms of toxic action and structure-activity relationships for organochlorine and synthetic pyrethroid insecticides. *Environ. Health Perspect.* **87**, 255–262 (1990).

99. Islam, R. & Lynch, J. W. Mechanism of action of the insecticides, lindane and fipronil, on glycine receptor chloride channels. *Br. J. Pharmacol.* **165**, 2707–2720 (2012).

100. Robb, E. L. & Baker, M. B. Organophosphate Toxicity. in *StatPearls* (StatPearls Publishing, Treasure Island (FL), 2023).

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

102. Dagg, K. *et al.* Evaluation of toxicity of clothianidin (neonicotinoid) and chlorfenapyr (pyrrole) insecticides and cross-resistance to other public health insecticides in Anopheles arabiensis from Ethiopia. *Malar. J.* **18**, 49 (2019).

103. Raghavendra, K. *et al.* Evaluation of the pyrrole insecticide chlorfenapyr for the control of Culex quinquefasciatus Say. *Acta Trop.* **118**, 50–55 (2011).

104. Evaluation of the pyrrole insecticide chlorfenapyr against pyrethroid resistant and susceptible Anopheles funestus (Diptera: Culicidae) - Oliver - 2010 - Tropical Medicine & International Health - Wiley Online Library. https://onlinelibrary.wiley.com/doi/full/10.1111/j.1365-3156.2009.02416.x.

105. Ihara, M. & Matsuda, K. Neonicotinoids: molecular mechanisms of action, insights into resistance and impact on pollinators. *Curr. Opin. Insect Sci.* **30**, 86–92 (2018).

106. N'Guessan, R. *et al.* Chlorfenapyr: A pyrrole insecticide for the control of pyrethroid or DDT resistant Anopheles gambiae (Diptera: Culicidae) mosquitoes. *Acta Trop.* **102**, 69–78 (2007).

107. Oxborough, R. M. *et al.* The activity of the pyrrole insecticide chlorfenapyr in mosquito bioassay: towards a more rational testing and screening of non-neurotoxic insecticides for malaria vector control. *Malar. J.* **14**, 124 (2015).

108. Oxborough, R. M. *et al.* Susceptibility testing of Anopheles malaria vectors with the neonicotinoid insecticide clothianidin; results from 16 African countries, in preparation for indoor residual spraying with new insecticide formulations. *Malar. J.* **18**, 264 (2019).

109. Zoh, M. G. *et al.* Experimental evolution supports the potential of neonicotinoid-pyrethroid combination for managing insecticide resistance in malaria vectors. *Sci. Rep.* **11**, 19501 (2021).

110. Coleman, M. *et al.* Developing global maps of insecticide resistance risk to improve vector control. *Malar. J.* **16**, 86 (2017).

111. Organisation, W. H. *Global Report on Insecticide Resistance in Malaria Vectors: 2010-2016*. (2017).

112. Liu, N. Insecticide Resistance in Mosquitoes: Impact, Mechanisms, and Research Directions. *Annu. Rev. Entomol.* **60**, 537–559 (2015).

113. Pelloquin, B. *et al.* Overabundance of Asaia and Serratia Bacteria Is Associated with Deltamethrin Insecticide Susceptibility in Anopheles coluzzii from Agboville, Côte d'Ivoire. *Microbiol. Spectr.* **9**, e00157-21 (2021).

114. Barnard, K., Jeanrenaud, A. C. S. N., Brooke, B. D. & Oliver, S. V. The contribution of gut bacteria to insecticide resistance and the life histories of the major malaria vector Anopheles arabiensis (Diptera: Culicidae). *Sci. Rep.* **9**, 9117 (2019).

115. Dada, N., Sheth, M., Liebman, K., Pinto, J. & Lenhart, A. Whole metagenome sequencing reveals links between mosquito microbiota and insecticide resistance in malaria vectors. *Sci. Rep.* **8**, 2084 (2018).

116. Williamson, M. S., Martinez-Torres, D., Hick, C. A. & Devonshire, A. L. Identification of mutations in the houseflypara-type sodium channel gene associated with knockdown resistance (kdr) to pyrethroid insecticides. *Mol. Gen. Genet. MGG* **252**, 51–60 (1996).

117. 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).

118. 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).

119. 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).

120. Mouhamadou, C. S. *et al.* Rapid spread of double East- and West-African kdr mutations in wild Anopheles coluzzi from Côte d'Ivoire. *Wellcome Open Res.* **4**, 31 (2019).

121. Yang, C., Feng, X., Huang, Z., Li, M. & Qiu, X. Diversity and frequency of kdr mutations within Anopheles sinensis populations from Guangxi, China. *Malar. J.* **15**, 411 (2016).

122. Dykes, C. L. *et al.* Knockdown Resistance (kdr) Mutations in Indian Anopheles stephensi (Diptera: Culicidae) Populations. *J. Med. Entomol.* **53**, 315–320 (2016).

123. 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. *Parasit. Vectors* **7**, 450 (2014).

124. Lol, J. C. *et al.* Molecular evidence for historical presence of knock-down resistance in Anopheles albimanus, a key malaria vector in Latin America. *Parasit. Vectors* **6**, 268 (2013).

125. Mackenzie-Impoinvil, L. *et al.* Contrasting patterns of gene expression indicate differing pyrethroid resistance mechanisms across the range of the New World malaria vector Anopheles albimanus. *PLoS ONE* **14**, e0210586 (2019).

126. Fernández, J. *et al.* Evaluation of primers reported for KDR detection in Anopheles species collected in the departments of Caaguazú and Alto Paraná in Paraguay. *Mem. Inst. Investig. En Cienc. Salud* **18**, 27–32 (2020).

127. 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. *Mol. Ecol.* **30**, 5303–5317 (2021).

128. 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).

129. 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).

130. Edi, A. V. C. *et al.* First detection of N1575Y mutation in pyrethroid resistant Anopheles gambiae in Southern Côte d'Ivoire. *Wellcome Open Res.* **2**, 71 (2017).

131. 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.* **109**, 6614–6619 (2012).

132. Park, S. H. *et al.* Monitoring Insecticide Resistance and Target Site Mutations of L1014 Kdr And G119 Ace Alleles in Five Mosquito Populations in Korea. *Korean J. Parasitol.* **58**, 543–550 (2020).

133. Feng, X. *et al.* Distribution and frequency of G119S mutation in ace-1 gene within Anopheles sinensis populations from Guangxi, China. *Malar. J.* **14**, 470 (2015).

134. Liebman, K. A. *et al.* Novel mutations on the ace-1 gene of the malaria vector Anopheles albimanus provide evidence for balancing selection in an area of high insecticide resistance in Peru. *Malar. J.* **14**, 74 (2015).

135. Ibrahim, S. S., Ndula, M., Riveron, J. M., Irving, H. & Wondji, C. S. The P450 CYP6Z1 confers carbamate/pyrethroid cross-resistance in a major African malaria vector beside a novel carbamate-insensitive N485I acetylcholinesterase-1 mutation. *Mol. Ecol.* **25**, 3436–3452 (2016).

136. Liu, N., Feng, X. & Qiu, X. RDL mutations in Guangxi Anopheles sinensis populations along the China–Vietnam border: distribution frequency and evolutionary origin of A296S resistance allele. *Malar. J.* **19**, 23 (2020).

137. Yang, C., Huang, Z., Li, M., Feng, X. & Qiu, X. RDL mutations predict multiple insecticide resistance in Anopheles sinensis in Guangxi, China. *Malar. J.* **16**, 482 (2017).

138. 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).

139. Tchouakui, M. *et al.* A marker of glutathione S-transferase-mediated resistance to insecticides is associated with higher Plasmodium infection in the African malaria vector Anopheles funestus. *Sci. Rep.* **9**, 5772 (2019).

140. 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).

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

142. 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).

143. Scott, J. G. Cytochromes P450 and insecticide resistance. *Insect Biochem. Mol. Biol.* **29**, 757–777 (1999).

144. Bergé, J. B., Feyereisen, R. & Amichot, M. Cytochrome P450 monooxygenases and insecticide resistance in insects. *Philos. Trans. R. Soc. B Biol. Sci.* **353**, 1701–1705 (1998).

145. Djogbénou, L., Labbé, P., Chandre, F., Pasteur, N. & Weill, M. Ace-1 duplication in Anopheles gambiae: a challenge for malaria control. *Malar. J.* **8**, 70 (2009).

146. Grau-Bové, X. *et al.* Resistance to pirimiphos-methyl in West African Anopheles is spreading via duplication and introgression of the Ace1 locus. *PLOS Genet.* **17**, e1009253 (2021).

147. Wondji, C. S. *et al.* Two duplicated P450 genes are associated with pyrethroid resistance in Anopheles funestus, a major malaria vector. *Genome Res.* **19**, 452–459 (2009).

148. Tandem duplication of a genomic region encoding glutathione S-transferase epsilon-2 and -4 genes in DDT-resistant Anopheles stephensi strain from India | Scientific Reports. https://www.nature.com/articles/s41598-022-21522-8.

149. 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 Biol.* **15**, R27 (2014).

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

151. Weedall, G. D. *et al.* A cytochrome P450 allele confers pyrethroid resistance on a major African malaria vector, reducing insecticide-treated bednet efficacy. *Sci. Transl. Med.* **11**, eaat7386 (2019).

152. Mugenzi, L. M. J. *et al.* Cis-regulatory CYP6P9b P450 variants associated with loss of insecticide-treated bed net efficacy against Anopheles funestus. *Nat. Commun.* **10**, 4652 (2019).

153. 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).

154. 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).

155. Bandibabone, J. *et al.* Investigating molecular mechanisms of insecticide resistance in the Eastern Democratic Republic of the Congo. *Malar. J.* **20**, 464 (2021).

156. Safi, N. H. Z. *et al.* Evidence of metabolic mechanisms playing a role in multiple insecticides resistance in Anopheles stephensi populations from Afghanistan. *Malar. J.* **16**, 100 (2017).

157. Fonseca-González, I., Quiñones, M. L., McAllister, J. & Brogdon, W. G. Mixed-function oxidases and esterases associated with cross-r esistance between DDT and lambda-cyhalothrin in Anopheles darlingi Root 1926 populations from Colombia. *Mem. Inst. Oswaldo Cruz* **104**, 18–26 (2009).

158. Mulamba, C. *et al.* Widespread Pyrethroid and DDT Resistance in the Major Malaria Vector Anopheles funestus in East Africa Is Driven by Metabolic Resistance Mechanisms. *PLOS ONE* **9**, e110058 (2014).

159. Gueye, Oumou. K. *et al.* Insecticide Resistance Profiling of Anopheles coluzzii and Anopheles gambiae Populations in the Southern Senegal: Role of Target Sites and Metabolic Resistance Mechanisms. *Genes* **11**, 1403 (2020).

160. Mavridis, K. *et al.* Rapid multiplex gene expression assays for monitoring metabolic resistance in the major malaria vector Anopheles gambiae. *Parasit. Vectors* **12**, 9 (2019).

161. Bonizzoni, M. *et al.* RNA-seq analyses of changes in the Anopheles gambiae transcriptome associated with resistance to pyrethroids in Kenya: identification of candidate-resistance genes and candidate-resistance SNPs. *Parasit. Vectors* **8**, 474 (2015).

162. Weedall, G. D. *et al.* An Africa-wide genomic evolution of insecticide resistance in the malaria vector Anopheles funestus involves selective sweeps, copy number variations, gene conversion and transposons. *PLOS Genet.* **16**, e1008822 (2020).

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

164. Chakraborty, M. *et al.* Hidden genomic features of an invasive malaria vector, Anopheles stephensi, revealed by a chromosome-level genome assembly. *BMC Biol.* **19**, 28 (2021).

165. Ayala, D. *et al.* The genome sequence of the malaria mosquito, Anopheles funestus, Giles, 1900. *Wellcome Open Res.* **7**, (2022).

166. Marinotti, O. *et al.* The genome of Anopheles darlingi, the main neotropical malaria vector. *Nucleic Acids Res.* **41**, 7387–7400 (2013).

167. Katoh, K., Misawa, K., Kuma, K. I. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).

168. Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455 (2019).

169. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).

170. Consortium, A. gambiae 1000 G. *et al.* Genetic diversity of the African malaria vector Anopheles gambiae. *Nature* **552**, 96–100 (2017).

171. St. Laurent, B. *et al.* Population genomics reveal distinct and diverging populations of An. minimus in Cambodia. *Commun. Biol.* **5**, 1–10 (2022).

172. The Anopheles gambiae 1000 Genome 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).

173. Neafsey, D. E. *et al.* Highly evolvable malaria vectors: The genomes of 16 Anopheles mosquitoes. *Science* **347**, 1258522 (2015).

174. Chang, X. *et al.* Genomic Variant Analyses in Pyrethroid Resistant and Susceptible Malaria Vector, Anopheles sinensis. *G3 GenesGenomesGenetics* **10**, 2185–2193 (2020).

175. Makunin, A. *et al.* A targeted amplicon sequencing panel to simultaneously identify mosquito species and Plasmodium presence across the entire Anopheles genus. *Mol. Ecol. Resour.* **22**, 28–44 (2022).

176. Collins ID, 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).

177. Ahmad, M. *et al.* Status of insecticide resistance in high-risk malaria provinces in Afghanistan. *Malar. J.* **15**, 98 (2016).

178. Enayati, A., Hanafi-Bojd, A. A., Sedaghat, M. M., Zaim, M. & Hemingway, J. Evolution of insecticide resistance and its mechanisms in Anopheles stephensi in the WHO Eastern Mediterranean Region. *Malar. J.* **19**, 258 (2020).

179. Singh, O. P., Dykes, C. L., Lather, M., Agrawal, O. P. & Adak, T. Knockdown resistance (kdr)-like mutations in the voltage-gated sodium channel of a malaria vector Anopheles stephensi and PCR assays for their detection. *Malar. J.* **10**, 59 (2011).

180. 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 Biochem. Mol. Biol.* **139**, 103655 (2021).

181. Whittaker, C. *et al.* Seasonal dynamics of Anopheles stephensi and its implications for mosquito detection and emergent malaria control in the Horn of Africa. *Proc. Natl. Acad. Sci.* **120**, e2216142120 (2023).

182. Thomas, S. *et al.* Resting and feeding preferences of Anopheles stephensi in an urban setting, perennial for malaria. *Malar. J.* **16**, 111 (2017).

183. Orjuela, L. I. *et al.* Absence of knockdown mutations in pyrethroid and DDT resistant populations of the main malaria vectors in Colombia. *Malar. J.* **18**, 1–9 (2019).

184. Lol, J. C. *et al.* Development of molecular assays to detect target-site mechanisms associated with insecticide resistance in malaria vectors from Latin America. *Malar. J.* **18**, 202 (2019).

185. Moiroux, N. *et al.* Human Exposure to Early Morning Anopheles funestus Biting Behavior and Personal Protection Provided by Long-Lasting Insecticidal Nets. *PLOS ONE* **9**, e104967 (2014).

186. Doucoure, S. *et al.* Anopheles arabiensis and Anopheles funestus biting patterns in Dielmo, an area of low level exposure to malaria vectors. *Malar. J.* **19**, 230 (2020).

Chapter 2: Identification of two insecticide resistance markers in Ethiopian *Anopheles stephensi* mosquitoes using a multiplex amplicon sequencing assay

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	lsh1704757	**Title**	Miss
First Name(s)	Holly		
Surname/Family Name	Acford-Palmer		
Thesis Title	A genomic investigation into three malaria vectors: An. darlingi, An. funestus, and An. stephensi		
Primary Supervisor	Prof. Susana Campino		

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?	05/04/2023		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
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 designed the amplicon panel, tested the primers, and applied the primer panel to the isolates. I then performed the bioinformatic analysis, interpreted the results under the supervision from my supervisors. I wrote the first draft of the paper, which was commented upon by my supervisors and co-authors.

SECTION E

Student Signature	[redacted]
Date	

Supervisor Signature	
Date	

scientific reports

Check for updates

OPEN

Identification of two insecticide resistance markers in Ethiopian *Anopheles stephensi* mosquitoes using a multiplex amplicon sequencing assay

Holly Acford-Palmer[1], Jody E. Phelan[1], Fitsum G. Tadesse[2], Mojca Kristan[1], Emma Collins[1], Anton Spadar[1], Thomas Walker[1], Teun Bousema[3], Louisa A. Messenger[1,4], Taane G. Clark[1,5] & Susana Campino[1]✉

Since its first detection in 2012 in Djibouti, *Anopheles stephensi* has invaded and established in the Horn of Africa, and more recently Nigeria. The expansion of this vector poses a significant threat to malaria control and elimination efforts. Integrated vector management is the primary strategy used to interrupt disease transmission; however, growing insecticide resistance is threatening to reverse gains in global malaria control. We present a next-generation amplicon-sequencing approach, for high-throughput monitoring of insecticide resistance genes (*ace1*, *GSTe2*, *vgsc* and *rdl*), species identification and characterization of genetic diversity (*its2* and *cox1*) in *An. stephensi*. Ninety-five *An. stephensi* mosquitoes, collected in Ethiopia, were screened, identifying 104 SNPs, including the knock-down mutation L958F (L1014F in *Musca domestica*), and for the first time in this vector species, the A296S substitution (A301S in *Drosophila melanogaster*) in the *rdl* locus. Two other amino acid substitutions (*ace1*-N177D, *GSTe2*-V189L) were also identified but have not been previously implicated in insecticide resistance. Genetic diversity in the mitochondrial *cox1* gene revealed shared haplotypes between Ethiopian *An. stephensi* with samples from Pakistan, Sudan, and Djibouti. Overall, we present a reliable, cost-effective strategy using amplicon-sequencing to monitor known insecticide resistance mutations, with the potential to identify new genetic variants, to assist in the high-throughput surveillance of insecticide resistance in *An. stephensi* populations.

The first confirmed finding of the Asian mosquito, *Anopheles stephensi*, in Africa was reported in Djibouti City in 2012[1], and it has now spread throughout the Horn of Africa into Ethiopia, the Republic of Sudan, Somalia, and recently Nigeria (2020) and Yemen[2–7]. *Anopheles stephensi* is a highly competent vector for both *Plasmodium falciparum* and *P. vivax* malaria parasites. A primary malaria vector in South Asian countries such as India and Pakistan, *An. stephensi* also present in the Middle East[1]. The spread of this vector into Africa has sparked concerns for the control and elimination of malaria, particularly since its first report in Djibouti malaria cases drastically increased from only 1,684 cases in 2013 to 73,535 in 2020[3,8–10]. Whilst not confirmed, this drastic increase is potentially a result of this vector's ability to occupy different ecological niches compared to other mosquito species in the region. The primary malaria vector in the Horn of Africa, *An. arabiensis*, transmits malaria in rural areas or extensive areas of urban agriculture [10–12]. However, *An. stephensi* is a proficient vector in urban areas, being able to use artificial water sources such as water tanks and polluted water sources for larval habitats[13]. As the continent of Africa becomes increasingly urbanised, the spread of *An. stephensi* is predicted to put 126 million Africans, with little to no acquired immunity, at risk of malaria without immediate control[12].

[1]Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, UK. [2]Malaria and NTD Directorate, Armauer Hansen Research Institute, ALERT Hospital Compound, P.O. Box 1005, Addis Ababa, Ethiopia. [3]Department of Medical Microbiology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands. [4]Environmental and Occupational Health, School of Public Health, University of Nevada, Las Vegas, Las Vegas, USA. [5]Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK. ✉email: Susana.campino@lshtm.ac.uk

Over the last two decades, interventions to control mosquito vectors have greatly contributed to the decline in malaria morbidity and mortality. Long lasting insecticidal nets (LLINs) and indoor residual spraying (IRS) have been successfully used to control African malaria vectors such as *An. funestus* and *An. gambiae* sensu lato complex species. However, *An. stephensi* exhibits different feeding (outdoor, evening biting) and resting behaviours (animal shelters are common dwellings) rendering methods such as IRS less effective[12,14]. In addition, the effectiveness of insecticides is threatened by increasing resistance in *Anopheles* vectors, which has been reported in almost all African countries[15,16].

The main mechanisms associated with insecticide resistance are target site insensitivity and metabolic resistance[17]. Several substitutions in the *voltage-gated sodium channel* (*vgsc*) gene, have been associated with resistance to pyrethroids and dichlorodiphenyltrichloroethane (DDT) (knock down resistance; *kdr*)[18,19]. Mutations and duplications in the *acetylcholinesterase 1* (*ace1*) gene have been associated with resistance to both carbamates and organophosphates[20,21], whereas mutations in the *GABA* receptor (dieldrin-*rdl* locus) have been associated with resistance to phenylpyrazoles and organochlorides, particularly dieldrin[22–24]. GABA receptors are also believed to be a target of pyrethroids[25,26]. Metabolic resistance resulting from enhanced detoxification of insecticides has been reported as one of the causes of resistance to insecticides from diverse classes and are usually associated with the over-expression of detoxification enzymes[27]. Other mechanisms of insecticide resistance identified include mosquito microbiome components and cuticle alterations[27–30]. Resistance to several insecticides of all major classes has been reported in *An. stephensi* in India, Sri Lanka, many countries of the WHO Eastern Mediterranean Region and also in Ethiopia[20,31–35]. Besides the use of susceptibility bioassays and biochemical techniques to characterize insecticide resistance, the occurrence of the *vgsc*-V1014 *kdr* mutation has also been investigated in *An. stephensi* populations[20,31–35]. In Ethiopia, the L1014F mutation was reported at low frequencies. In a cohort of permethrin-resistant *An. stephensi* the *kdr* mutation was not detected, this suggests other molecular mechanisms maybe involved with pyrethroid susceptibility in this vector. Therefore, additional candidate genes need to be explored to better understand the wider landscape of insecticide-associated mutations in *An. stephensi* populations.

The use of whole genome sequencing (WGS) gives more insight into the genomic landscape of organisms, however limitations such as high cost and low DNA quantities make it an unsuitable method for high-throughput surveillance. The application of multiplex amplicon sequencing with next generation sequencing (NGS) technologies provides a good, high throughput alternative to screen target regions of interest in large datasets, in a cost-efficient manner. Target amplicon sequencing (Amp-seq) is based on the sequencing of pools of PCR amplicons of interest and has been applied in malaria surveillance to characterise *P. falciparum*[36] and *An. gambiae s.l.* populations[37,38]. Here we designed and validated a targeted Amp-seq assay, combined with dual indexing barcodes and Illumina sequencing, for profiling *An. stephensi* across *vgsc*, *rdl*, *gste2* and *ace1* loci related to insecticide resistance. Further, we included amplicons in the assay for species identification and phylogenetic analysis, specifically the internal transcribed spacer (*its2)* and the mitochondrial gene (*cytochrome c oxidase subunit 1*, *cox1)*. Using this approach, it is possible to pool multiple samples across several loci, distinguishing individual samples based on the unique index barcodes, increasing efficiency, and decreasing costs. Amp-seq can be used to screen for known and novel mutations, which when used in combination with phenotyping assays, can identify genetic variants predictive of resistance to interventions/insecticides. The assay is easy to implement and can be applied to many samples at low cost, achieved through PCR multiplexing and dual barcoding. It is a promising tool to confirm species, support the surveillance of insecticide resistance in *An. stephensi* and inform vector control strategies targeting this invasive species in the African continent.

Results

Detection of SNPs in genes associated with insecticide resistance. The amplicon data generated from 95 *An. stephensi* sourced from Ethiopia, identified a total of 104 high quality SNPs and 20 indels (Table 1, Supplementary Table 1). The average coverage per amplicon was high (range: 437.54 to 14,483.47 reads) (Table 1). Most of the SNPs identified were synonymous, not leading to any changes in the protein sequence.

		Number of SNPs/INDELs		
Amplicon	Average coverage	All	Ethiopia	Colony
ACE1_I	13,624.74	15	12	6
ACE1_II	614.67	14	14	11
CO1	14,483.47	6	6	2
GSTe2	2394.67	16	13	12
ITS2	2231.35	2	2	0
Rdl1	1499.71	20	19	7
Rdl2	1241.53	27	25	14
VGSCI	1242.09	3	2	1
VGSCII	437.54	21	16	12
VGSCIII	750.55	6	4	6
VGSCIV	1841.83	12	11	3

Table 1. Average amplicon coverage, and number of genetic variants detected.

Only four SNPs were found to result in a missense mutation, with two previously described to be associated with insecticide resistance (Supplementary Table 2). The *kdr* mutation L958F (L1014F in *M. domestica*) in the 2nd domain of the *vgsc* locus, was detected in 13 Ethiopian samples (13.8%). The other known SNP was a A296S substitution (A301S in *D. melanogaster*) in the *GABA*-gated chloride channel (*rdl* locus), which has been previously reported in *Anopheles* species but not *An. stephensi*[39]. In our study, A296S was detected in 21 (22.1%) Ethiopian samples. Two putatively novel missense variants were identified in our study. First, a SNP in the *ace1* gene (N177D) found in 19 (57.8%) Ethiopian samples and in the insecticide susceptible colony populations (n = 2). Second, a SNP in the *GSTe2* gene (V189L) was detected at low frequency (n = 3). Sanger sequencing confirmed that these SNPs were true polymorphisms and not sequencing artifacts.

Genetic diversity of An. stephensi populations from Asia and Africa. Amplicon sequences of the ribosomal DNA internal transcribed spacers *its2* and mitochondrial *cox1* gene were analysed alongside publicly available data. Across 81 Ethiopian samples, 38 publicly available sequences and 1 colony (SDA500 colony specimen, descended from a Pakistani isolate), only two SNPs were identified in *its2*, one present only in 6% of Ethiopian samples, and the other present in all Ethiopian samples and a few samples from India. A few singletons were also found. Phylogenetic analyses revealed a clade that includes all Ethiopian samples and three samples from India (Fig. 1), while the samples from other geographic regions clustered together with no specific geographic cluster observed, reflecting the limited genetic diversity at this locus, also supported by the low nucleotide diversity ($\pi = 0.00169$). For the *cox1* gene, 49 Ethiopian samples, 41 publicly available sequences and 1 colony sample were analysed. Two SNPs were identified in Ethiopia (22%; 38%), and some singleton SNPs were also observed. For the phylogenetic analysis, additional *cox1* sequences from other *Anopheles* species, (*An. albimanus*, *An. arabiensis*, *An. coluzzii*, *An. darlingi*, *An. dirus*, *An. funestus*, *An. gambiae*, *An. minimus*), were also included. The

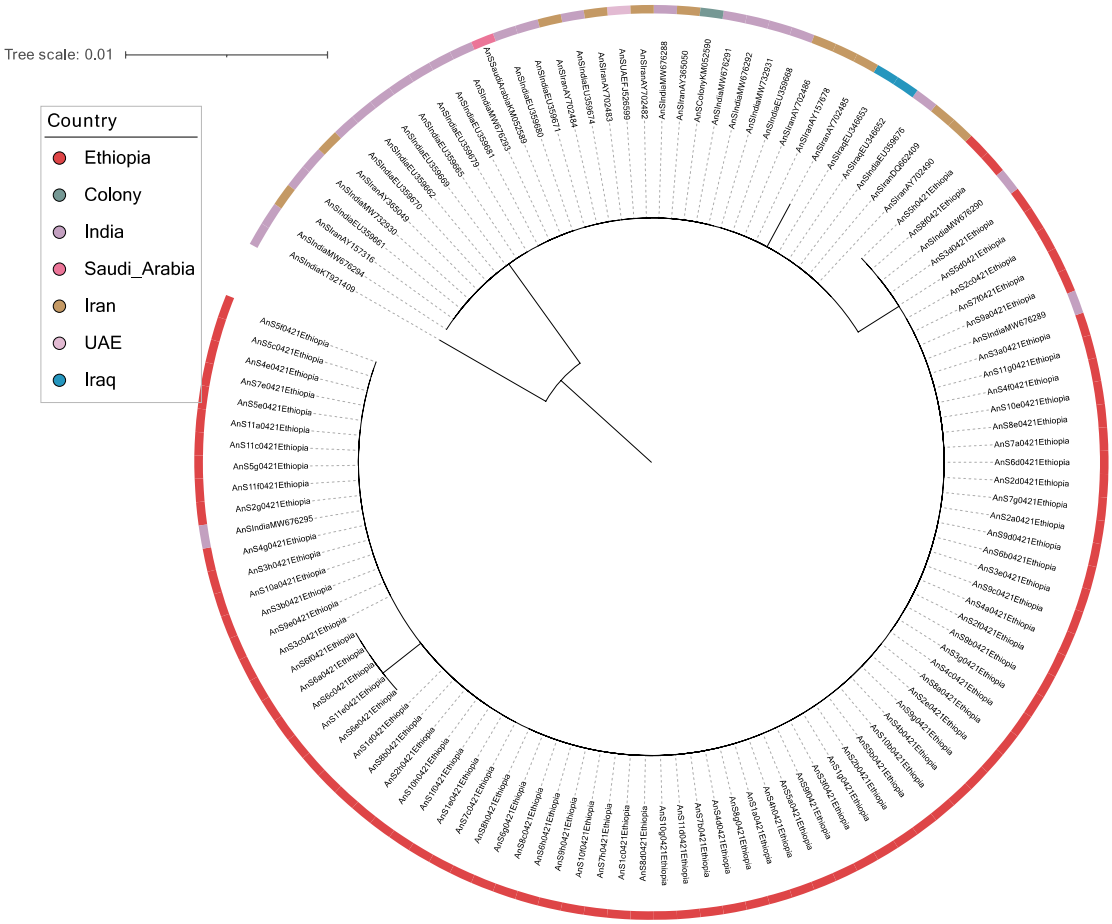


Figure 1. Neighbourhood joining tree constructed using ITS2 gene sequences generated from this study, alongside publicly available sequence data. The tree was built using the maximum-likelihood method assuming GTR model of nucleotide substitution, with the gamma model of heterogeneity rate. This tree was generated using RAxML, and visualised with iTOL[65,66].

phylogenetic analyses revealed a clade that separated the non-*An. stephensi* samples, including a subclade with 3 samples from the *An. gambiae* complex. For the remaining *An. stephensi*, due to the limited genetic diversity at this locus, supported by the low nucleotide diversity ($\pi = 0.00521$), there were no specific geographic clusters observed (Fig. 2). The nucleotide diversity was in accordance with data previously reported for this species at this locus[40].

Haplotype networks were constructed to analyse and visualize the relationships among the DNA sequences (Fig. 3). For *its2*, nine haplotypes were identified, with > 72% of Ethiopian samples sharing a single haplotype. The samples from other geographic regions also shared a core haplotype. More haplotypes were identified for *cox1* (n = 24), with a core haplotype present in almost all samples from the different countries, except Saudi Arabia that did not share any haplotypes with other populations. A single sample from India also shared a less frequent haplotype with a Sri Lankan specimen. For the samples from Ethiopia, besides the core haplotype, other haplotypes were also observed, particularly one shared with samples from Pakistan and Sudan, and a less frequent haplotype shared with samples from Djibouti.

Discussion

The introduction of *An. stephensi* mosquitos into the Horn of Africa is a significant threat to malaria control. Vector control strategies have been successful in reducing malaria transmission[41,42] and surveillance of insecticide resistance is essential to inform malaria control programmes. To assist surveillance activities, we evaluate an Amp-seq assay to screen for known and novel mutations associated with a range of insecticides. In the *An. stephensi* specimens screened, the *kdr* mutation L958F (L1014F in *M. domestica;* L1014F *kdr-west*) was present at a frequency of 10.2%. The L1014F *kdr-west* mutation was first detected in *An. gambiae* in West Africa. It was reported at a low frequency (5.7%) in Ethiopian *An. stephensi* specimens collected in 2018 from other regions (Bati, Degehabur, Dire Dawa, Gewane, and Semera)[34]. The same study also screened samples from Awash Sebat Kilo, the region where this studies samples were collected, but did not detect this mutation. However, a year later, we have detected the L1014F *kdr-west* mutation. This may be a result of this studies larger sample size (n = 95),

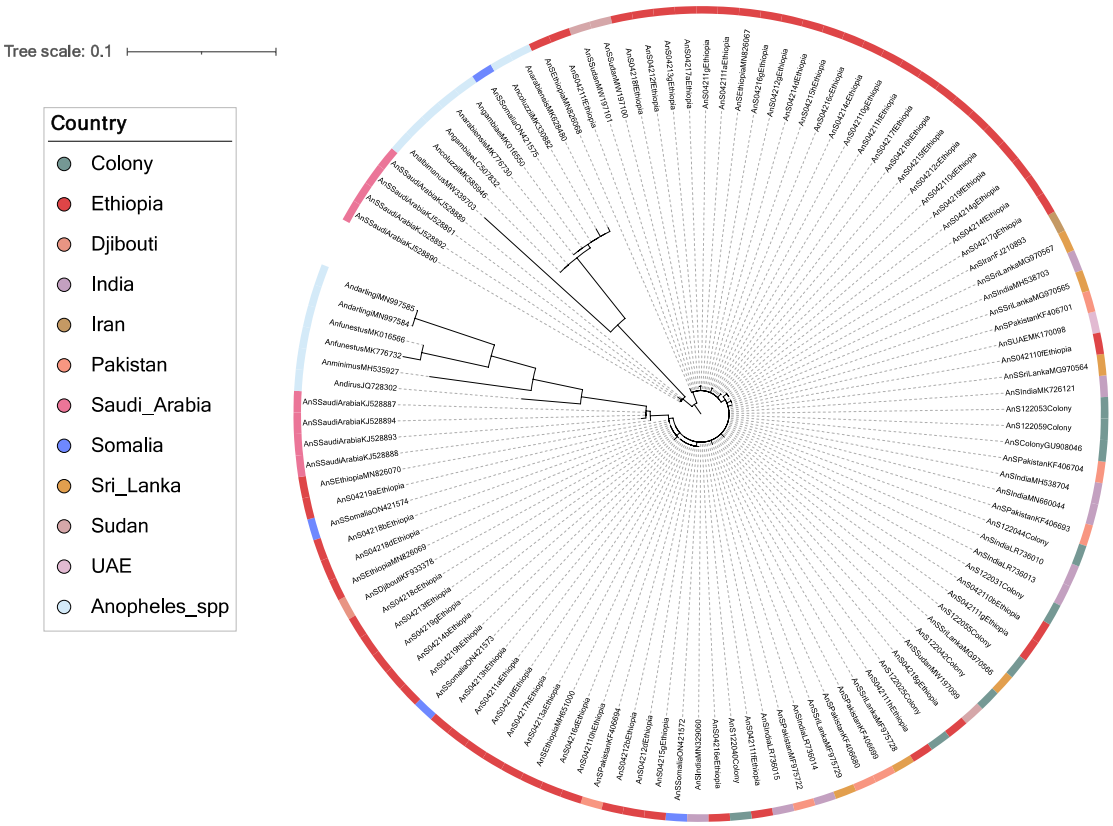


Figure 2. Neighbourhood joining tree constructed using *cox1* gene sequences generated from this study, alongside publicly available sequence data from *An. stephensi* and a variety of other *Anopheles spp*. This includes *An. albimanus, An. arabiensis, An. coluzzi, An. darlingi, An. dirus, An. funestus,* and *An. gambiae*. The tree was built using the maximum-likelihood method assuming GTR model of nucleotide substitution, with the gamma model of heterogeneity rate. This tree was generated using RAxML, and visualised with iTOL[65,66].

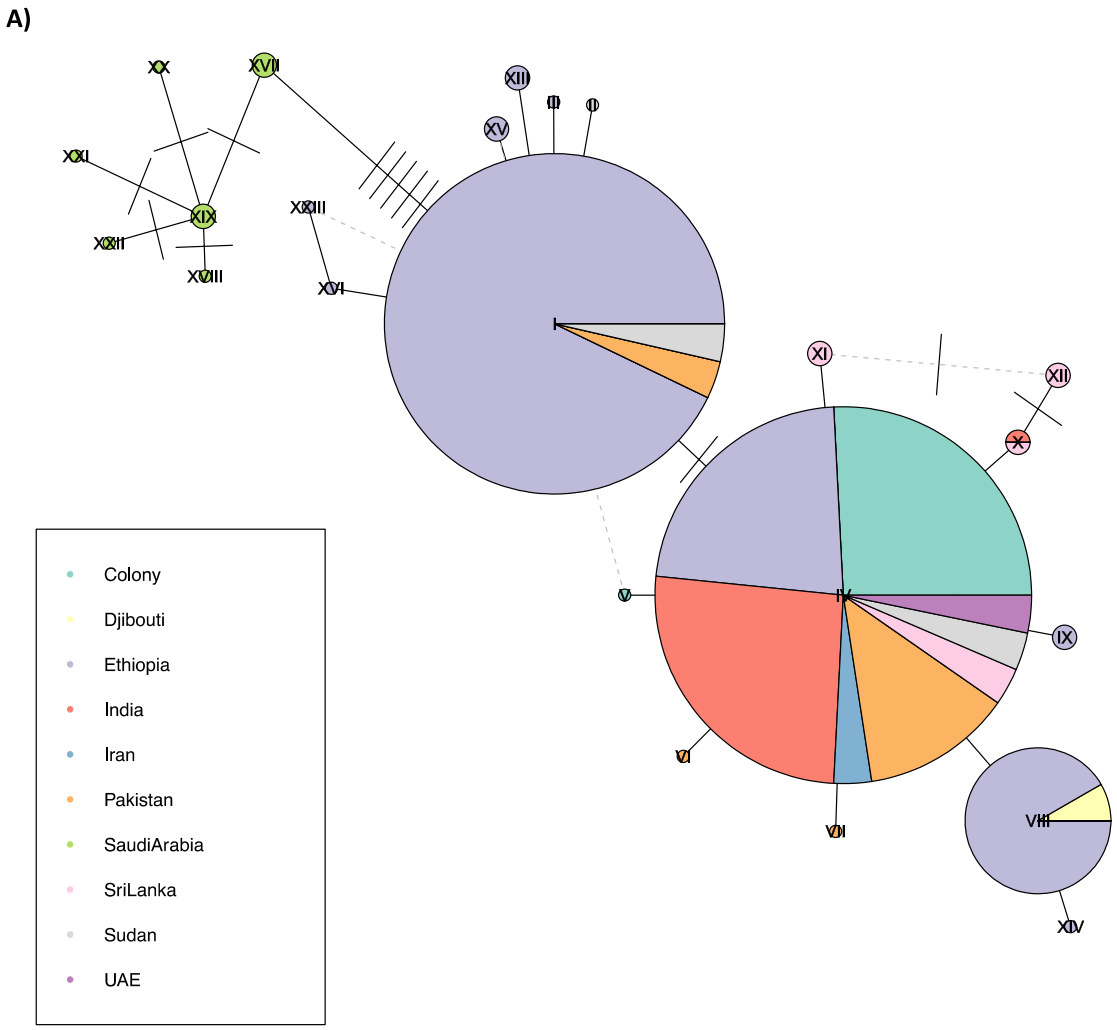


Figure 3. Haplotype or minimal-spanning network constructed using *cox-1* sequences (A) and *its2* sequences (B) generated in this study alongside publicly available data. Each node represents a haplotype, with each segment in the node representing a country, proportionally sized to the number of samples present in the segment and node. The number of nucleotides differences between haplotypes is represented by the number of ticks between nodes.

compared to n = 8. This mutation has also been confirmed in populations of *An. stephensi* in the Middle East and South Asia, with both *kdr-west* (L1014F) and *kdr-east* (L1014S) being described[20,31,32,43–45].

In this study we also identified the A296S mutations in the GABA gene (*rdl*). This mutation has not been reported previously in *An. stephensi*, likely because this locus has not been screened in this mosquito species. Dieldrin, alongside other cyclodienes that target the GABA receptor, are no longer officially used due to environmental toxicity and the potential impact on human health. However, the A296S mutation may confer resistance to other insecticides, such as fipronil; and broflanilide. These are newly discovered GABA-gated chloride channel allosteric modulators under evaluation as an IRS tool[46,47], so it remains important to identify and document its presence. We targeted two other mutations in the *rdl* gene (V327I and T345S), which are seen in tandem with the A296S SNP[39,48,49]. However, we found no evidence of these non-synonymous SNPs in the specimens analysed. These secondary SNPs have been described in both Asian and African *Anopheles* mosquito populations, but always in the presence of the A296S mutation[39,48,49]. No other commonly reported mutations, such as the *ace1* G119S, were identified in the specimens analysed. However, two new mutations were detected in the Ethiopian samples: N177D in the *ace1* gene, and V189L in the *GSTe2* gene. Both were also present in insecticide susceptible colony samples, suggesting these mutations are not involved in insecticide resistance.

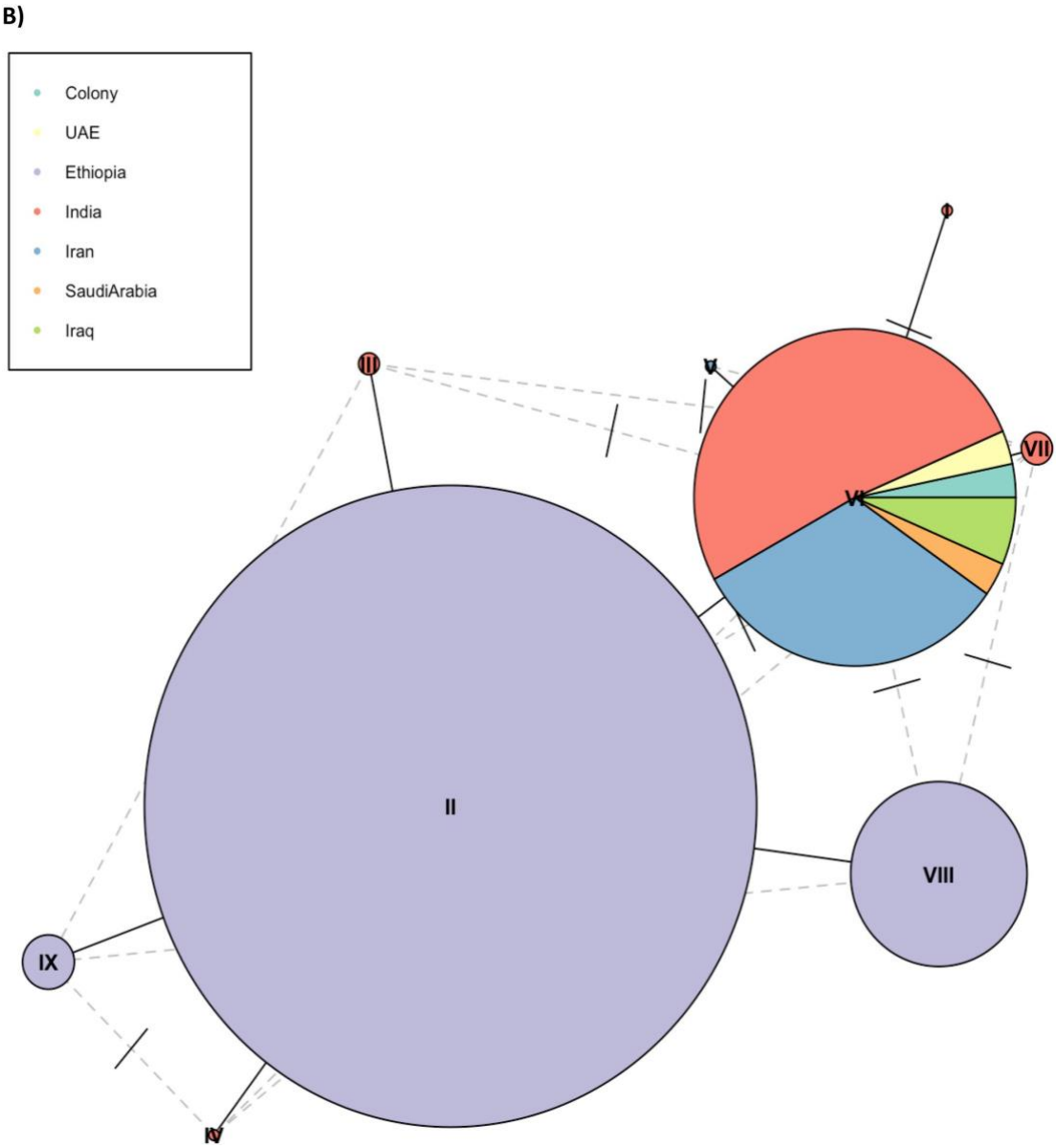


Figure 3. (continued)

The genetic analysis using data from *its2* and *cox1* genes revealed a low level of genetic diversity, consistent with studies of native *An. stephensi* populations in Sri Lanka and India, where few variants and minimal genetic differentiation in these loci was observed[13,50–52]. For *its2*, the Ethiopian samples cluster together and separate from the *An. stephensi* from other countries. The Ethiopian samples were all collected in the same village, likely explaining the clustering pattern observed. More haplotypes were observed for the *cox1* gene, with a core haplotype shared by all samples, except those sourced from Saudi Arabia. Ethiopian samples shared haplotypes with samples of Pakistan origin, as previously reported[53], but also shared haplotypes with samples from neighbouring countries (Sudan, Djibouti), supporting evidence of the spread of this vector across the Horn of Africa. It is possible that this vector is already present in other regions in Africa, as recently reported in Nigeria and Kenya (year 2020)[54,55]. Strengthening the surveillance activities of this species, through large-scale genetic characterisation of this vector, is urgently needed to rapidly implement targeted control strategies.

Overall, the cost-effective and high throughput Amp-seq approach presented here, can be implemented by vector control programs to screen *An. stephensi* mosquitoes for both known and putative novel insecticide resistance mutations. A limitation of our approach is that it only targets loci that are established mechanisms of

insecticide resistance. However, it is easy to add other targets to the Amp-seq method as new loci are detected. This approach can be used to complement diagnostic bioassays, by providing genotyping data alongside phenotypic information, to detect new genetic variants underlying insecticide resistance. Our approach is highly flexible, with the easy addition of genomic regions of interest to the panel, for example, to include loci related to metabolic resistance or that can target pathogens such as malaria parasites, thereby enhancing surveillance activities.

Conclusion

This study validated the use of an amplicon sequencing panel to perform molecular monitoring of *An. stephensi* insecticide resistance and explore genetic diversity. The identification of two known SNPs associated with insecticide resistance, alongside other non-synonymous SNPs not previously described in any mosquitoes highlights the potential for this technology to find existing and novel mutations that could affect insecticide response. Here we confirm the presence of the *vgsc*-L1014F *kdr* mutation and, for the first time in this species, the A296S *rdl* mutation. We were also able to investigate the genetic structure of *An. stephensi* and haplotype sharing between this Ethiopian population and neighbour countries. The extension of the approach to other loci can further assist with supporting control strategies of vector-borne diseases and reducing their global burden.

Materials and methods

Mosquito collection and identification. *An. stephensi* mosquitoes were obtained from the LSHTM colony (Sind Kasur strain, from Pakistan), and field samples were collected in Awash Sebat Kilo, Ethiopia. Samples collected in Ethiopia were caught between April and September 2019, in one of three ways: CDC mini light traps, aspiration from cattle shelters, and human landing collection. The mosquitoes were first identified morphologically as *An. stephensi*, before undergoing a qPCR targeting the *ITS2* and *cox1* genes for molecular confirmation using previously described methods[9].

DNA extraction. Individual mosquitoes were suspended in 1X PBS and lysed mechanically for 30 s or until all body parts were no longer visible using a Tissue Ruptor II (Qiagen, Hilden, Germany). The DNA was then extracted using Qiagen DNAeasy Blood and Tissue kits, according to manufacturer's instructions. DNA for each sample was quantified using Qubit 2.0 fluorimeter HS DNA kit (ThermoFisher, Waltham, MA, USA) and stored at – 20 °C.

Primer design for Amp-seq. Sequences were downloaded from VectorBase (https://vectorbase.org/vectorbase/app). The primers were designed using Primer3[56] and aimed to amplify an approximately 500 bp region, which contained previously described SNPs associated with target site resistance in *Anopheles* and other vectors (Supplementary Table 3). Primers were designed to bind to exonic regions where possible. Primers were designed to target 7 genomic regions in SNPs previously associated with insecticide resistance: 4 different domains of the voltage-gated sodium channel (*vgsc* DI-IV), one partial sequence of the gene *acetylcholinesterase 1* (*ace1*), one sequence in the resistance to dieldrin gamma-amino butyric acid receptor (*rdl*), and another in the *gste2* gene (Supplementary Table 4). Two genomic regions were targeted for species identification and/or phylogenetic analyses: the nuclear ribosomal internal transcribed spacer 2 (*ITS2*) and the mitochondrially encoded cytochrome c oxidase I (*cox1*), (Supplementary Table 3). The primer sequences were concatenated with unique 5' tag barcode (6 bp long) to discriminate individual samples and enable multiplexing, along with sequences complementary to Illumina adapters for sequencing as described previously[38]. Prior to PCR amplification, each sample was assigned one unique barcode in each forward and reverse primer to be used for amplification of all loci. To identify primers suitable for multiplexing, ThermoFisher Scientific Multiple Primer Analyser was used with sensitivity set to one to identify potential primer dimerization events.

Amplicon generation and next generation sequencing. Amplicons were generated using NEB Q5 Hot-start polymerase (NEB, New England Biolabs, UK) in 25ul final reactions. Final primer concentration averaged 0.5um for all assays, and 1 µl (around 2 ng/µl) of sample extract was used. The cycling conditions were as follows: hot start polymerase activation for 3 min at 95 °C, then 30 cycles of 95 °C for 10 s, 58 °C for 30 s and 72 °C for 45 s, followed by a final elongation step of 72 °C for 5 min. After gene-specific multiplex PCR reactions, amplicons were visualized in a 1% agarose gel to check for band size and intensity. Each PCR multiplex from each mosquito sample were pooled together, and then pooled with other mosquitoes that had different 5'tag barcodes to reach a maximum of 200 amplicons per pool. Pools were purified with using Kapa beads (Roche) as per the manufacturer's instructions in a 0.7:1 bead to sample volume ratio to remove excess primers and PCR reagents. The concentration of each pool was then measured using the Qubit 2.0 fluorimeter HS DNA kit (ThermoFisher, Waltham, MA, USA). The final pool containing 200 amplicons was then used as the template in the indexing PCR (second PCR step), which was performed to introduce the Illumina adaptors and barcode. This 2nd PCR step was performed as part of the Illumina-based Amplicon-EZ service (Genewiz, UK), followed by sequenced using Illumina MiSeq, on a configuration of paired end 2 × 250 bp. A minimum of 50,000 reads were obtained per pool (250 reads per amplicon in a pool of 200 amplicons) using the Genewiz service (< US$ 0.5 per amplicon).

Amplicon bioinformatic analysis. Raw sequences were de-multiplexed based on the barcode combination assigned to each sample using an in-house python script (https://github.com/LSHTMPathogenSeqLab/amplicon-seq). The resulting raw sequence data was then analysed using an in-house pipeline, where raw

sequence data for each sample was trimmed using Trimmomatic software, mapped to the reference sequence (UCI_ANSTEP_V1.0) using bwa-mem software, and reads clipped using Samclip package with a maximum clip length of 50[57–59]. Sample fastq files are available at European Nucleotide Archive website (Project ID: PRJEB57331, Accession numbers: ERR10484018—ERR10484147).

Multiple variant callers were utilised to maximise the number of mutations identified, and as an additional quality control procedure. SNPs and small indels were called using freebayes (v1.3.5, –haplotype-length -1) and GATK HaplotypeCaller (v4.1.4.1, default parameters) software tools[60,61]. The variants were then filtered (DP > 30) using bcftools software to ensure only high-quality variants were included. High quality SNPs were identified using filters that included a minimum phred quality of 30 per called base, minimum depth of 30 reads, and minimum alternate allele depth of 10. Finally, only SNPs that were present in > 1 sample, and present across two independent pools were retained. Variant annotation was carried out using the SnpEff tool, combined with a database based on the UCI_V1 reference genome[62,63]. The percentage of alternative allele to total depth coverage was used to classify genotypes to homozygous reference (< 25% alternate allele reads), homozygous alternate (> 75% alternate allele reads) or heterozygous (25–75% alternate allele), as described previously[38].

Phylogenetic analysis. For the *ITS2* and *cox1* amplicons, following SNP calling, all bam files from samples with > 50-fold coverage for these genes were converted to fasta files using an in-house pipeline (https://github.com/LSHTMPathogenSeqLab/fastq2matrix). The FASTA files were then aligned using MAFFT for each gene. The alignments also included 37 *ITS2 An. stephensi* nucleotide sequences public available (accession numbers: India (n = 22, EU359661, EU359662, EU359665, EU359668-EU359671, U359674, EU359676, U359679, EU359680, EU359681, KT921409, MW676288-MW676295, MW732930), Iran (n = 11, AY157316, AY157678, AY365049, AY365050, AY702482, AY702483, AY702484, AY702485, AY702486, AY702490, DQ662409, Iran (n = 2, EU346653, EU346652), Saudi Arabia (KM052589), UAE (FJ526599)). For *cox1*, alignments also included 40 nucleotide sequences public available (accession numbers: Saudi Arabia (n = 8, KJ528887- KJ528894), Djibouti n = 1(MK170098), Ethiopia (n = 5,MH651000, MN826067-MN826070), India (n = 9, LR736010, LR736013-LR736015, MH538703, MH538704, MK726121, MN329060, MN660044) Iran (FJ210893), Pakistan (n = 7 , KF406680, KF406693, KF406694, KF406699, KF406704, KF406704), Sri Lanka (n = 6, MF975728, MF975729, MG970564-MG970567), Sudan (n = 3, MW197099-MW197101), UAE (MK170098)). For COI other Anopheles species were included (*An. albimanus* (MW339703), *An_arabiensis*_MK628480, *An. coluzzii* (MK330882), *An. darlingi* MN997585, *An. dirus* JQ728302, *An. funestus* MK776732, *An. gambiae*_LC507832, *An. minimus* MH535927). The resulting alignment was then viewed and trimmed on Aliview software[64]. Following this, a phylogenetic tree was constructed using RAxML software[65]. RAxML uses a maximum-likelihood approach, and the GTRGAMMA option applied which assumes GTR model of nucleotide substitution, with the gamma model of heterogeneity rate. A bootstrap value of 500 replicates was used to construct the tree. Trees were then visualised in the iTOL tool[66].

Haplotype networks. The aligned FASTA files for ITS2 and *cox1* genes, were used to construct haplotype networks using the Pegas package in R[67]. The same package was also used to calculate nucleotide diversity, haplotype diversity and Tajima's D statistic.

Data availability

All raw sequence data is listed in the European Nucleotide Archive (Project ID: PRJEB57331, Accession numbers: ERR10484018—ERR10484147).

Received: 26 December 2022; Accepted: 26 March 2023
Published online: 05 April 2023

References

1. 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 Trop.* **139**, 39–43 (2014).
2. R, A. *et al.* Confirmation of the presence of Anopheles stephensi among a conflict-affected host community in Aden City, Yemen. (2022) doi:https://doi.org/10.21203/RS.3.RS-1977582/V1.
3. Balkew, M. *et al.* An update on the distribution, bionomics, and insecticide susceptibility of Anopheles stephensi in Ethiopia, 2018–2020. *Malar. J.* **20**, 263 (2021).
4. WHO. Malaria Threat Map: Invasive Vector Species. *WHO* (2021).
5. Malaria Threat Map. https://apps.who.int/malaria/maps/threats/?theme=invasive&mapType=invasive%3A0&bounds=%5B%5B-23.293603560044943%2C-18.128661258089999%5D%2C%5B65.36581083765986%2C32.29279894563096%5D%5D&insecticideClass=PYRETHROIDS&insecticideTypes=&assayTypes=MOLECULAR_ASSAY%2CBIOCHEMICAL_ASSAY%2CSYNERGIST-INSECTICIDE_BIOASSAY&synergistTypes=&species=&vectorSpecies=&surveyTypes=&deletionType=HRP2_PROPORTION_DELETION&plasmodiumSpecies=P._FALCIPARUM&drug=DRUG_AL&mmType=1&excludeLowerPatients=false&excludeLowerSamples=false&endemicity=false&countryMode=false&storyMode=false&storyModeStep=0&filterOpen=true&filtersMode=filters&years=1985%2C2022.
6. 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. *Parasit. Vectors* **14**, (2021).
7. Ahmed, A. *et al.* Invasive malaria vector anopheles stephensi mosquitoes in Sudan, 2016–2018. *Emerg. Infect. Dis.* **27**, 2952–2954 (2021).
8. World Health Orgnization (WHO). *WHO initiative to stop the spread of Anopheles stephensi in Africa*.
9. Tadesse, F. G. *et al.* Anopheles stephensi Mosquitoes as Vectors of Plasmodium vivax and falciparum, Horn of Africa, 2019. *Emerg. Infect. Dis.* **27**, 603–607 (2021).
10. Seyfarth, M., Khaireh, B. A., Abdi, A. A., Bouh, S. M. & Faulde, M. K. Five years following first detection of Anopheles stephensi (Diptera: Culicidae) in Djibouti, Horn of Africa: Populations established—malaria emerging. *Parasitol. Res.* **118**, 725–732 (2019).

11. Messenger, L. A. *et al.* Insecticide resistance in Anopheles arabiensis from Ethiopia (2012–2016): A nationwide study for insecticide resistance monitoring. *Malar. J.* **16**, 469 (2017).
12. Sinka, M. E. *et al.* A new malaria vector in Africa: Predicting the expansion range of "Anopheles stephensi" and identifying the urban populations at risk. *Proc. Natl. Acad. Sci.* **117**, 24900–24908 (2020).
13. Surendran, S. N. *et al. Anthropogenic factors driving recent range expansion of the malaria vector anopheles stephensi. Frontiers in Public Health* vol. 7 (2019).
14. Surendran, S. N. *et al.* Anthropogenic factors driving recent range expansion of the malaria vector anopheles stephensi. *Frontiers in Public Health* (2019).
15. Hancock, P. A. *et al.* Associated patterns of insecticide resistance in field populations of malaria vectors across Africa. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 5938–5943 (2018).
16. Moyes, C. L. *et al.* Evaluating insecticide resistance across African districts to aid malaria control decisions. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 22042–22050 (2020).
17. Organização Mundial da Saúde. *Guidelines for Malaria Vector Control. Guidelines for Malaria Vector Control* (2019).
18. Martinez-Torres, D. *et al.* Molecular characterization of pyrethroid knockdown resistance (kdr) in the major malaria vector Anopheles gambiae ss. *Insect Mol. Biol.* **7**, 179–184 (1998).
19. 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 Biol.* **15**, R27 (2014).
20. Soltani, A. *et al.* Resistance mechanisms of anopheles stephensi (Diptera: Culicidae) to temephos. *J. Arthropod-Borne Dis.* **9**, 71–83 (2014).
21. Ibrahim, S. S., Ndula, M., Riveron, J. M., Irving, H. & Wondji, C. S. The P450 CYP6Z1 confers carbamate/pyrethroid cross-resistance in a major African malaria vector beside a novel carbamate-insensitive N485I acetylcholinesterase-1 mutation. *Mol. Ecol.* **25**, 3436–3452 (2016).
22. 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. *Parasit. Vectors* **7**, (2014).
23. Du, W. *et al.* Independent mutations in the Rdl locus confer dieldrin resistance to Anopheles gambiae and An. arabiensis. *Insect Mol. Biol.* **14**, 179–183 (2005).
24. Ffrench-Constant, R. H., Anthony, N., Aronstein, K., Rocheleau, T. & Stilwell, G. Cyclodiene insecticide resistance: From molecular to population genetics. *Annu. Rev. Entomol.* **45**, 449–466 (2000).
25. 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).
26. Taylor-Wells, J., Brooke, B. D., Bermudez, I. & Jones, A. K. The neonicotinoid imidacloprid, and the pyrethroid deltamethrin, are antagonists of the insect Rdl GABA receptor. *J. Neurochem.* **135**, 705–713 (2015).
27. Balabanidou, V. *et al.* Cytochrome P450 associated with insecticide resistance catalyzes cuticular hydrocarbon production in Anopheles gambiae. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 9268–9273 (2016).
28. Yahouédo, G. A. *et al.* Contributions of cuticle permeability and enzyme detoxification to pyrethroid resistance in the major malaria vector Anopheles gambiae. *Sci. Rep.* **71**(7), 1–10 (2017).
29. Pelloquin, B. *et al.* Overabundance of asaia and serratia bacteria is associated with deltamethrin insecticide susceptibility in anopheles coluzzii from agboville, Côte d'Ivoire. *Microbiol. Spectr.* **9**, (2021).
30. Omoke, D. *et al.* Western Kenyan Anopheles gambiae showing intense permethrin resistance harbour distinct microbiota. *Malar. J.* **20**, (2021).
31. Singh, O. P., Dykes, C. L., Lather, M., Agrawal, O. P. & Adak, T. Knockdown resistance (kdr)-like mutations in the voltage-gated sodium channel of a malaria vector Anopheles stephensi and PCR assays for their detection. *Malar. J.* **10**, 59 (2011).
32. Dykes, C. L. *et al.* Knockdown resistance (kdr) mutations in Indian anopheles stephensi (Diptera: Culicidae) populations. *J. Med. Entomol.* **53**, 315–320 (2016).
33. Yared, S. *et al.* Insecticide resistance in Anopheles stephensi in Somali Region, eastern Ethiopia. *Malar. J.* **19**, (2020).
34. Samake, J. N. *et al.* Detection and population genetic analysis of kdr L1014F variant in eastern Ethiopian Anopheles stephensi. *Infect. Genet. Evol.* **99**, 105235 (2022).
35. Carter, T. E. *et al.* Analysis of the Knockdown Resistance Locus (kdr) in Anopheles stephensi, An. arabiensis, and Culex pipiens s.l. for Insight Into the Evolution of Target-site Pyrethroid Resistance in Eastern Ethiopia. *Am. J. Trop. Med. Hyg.* **106**, 632–638 (2022).
36. Nag, S. *et al.* High throughput resistance profiling of Plasmodium falciparum infections based on custom dual indexing and Illumina next generation sequencing-technology. *Sci. Rep.* **7**, 2398 (2017).
37. Makunin, A. *et al.* A targeted amplicon sequencing panel to simultaneously identify mosquito species and Plasmodium presence across the entire Anopheles genus. *Mol. Ecol. Resour.* **22**, 28–44 (2022).
38. 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**, (2022).
39. Yang, C., Huang, Z., Li, M., Feng, X. & Qiu, X. RDL mutations predict multiple insecticide resistance in Anopheles sinensis in Guangxi, China. *Malar. J.* **16**, (2017).
40. Carter, T. E. *et al.* Genetic diversity of Anopheles stephensi in Ethiopia provides insight into patterns of spread. *Parasit. Vectors* **14**, 1–12 (2021).
41. Nkya, T. E., Akhouayri, I., Kisinza, W. & David, J. P. Impact of environment on mosquito response to pyrethroid insecticides: Facts, evidences and prospects. *Insect Biochem. Mol. Biol.* **43**, 407–416 (2013).
42. Bhatt, S. *et al.* The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. *Nat. 2015 5267572* **526**, 207–211 (2015).
43. Khan, J. *et al.* Identification of a biological form in the Anopheles stephensi laboratory colony using the odorant-binding protein 1 intron I sequence. *PLoS ONE* **17**, e0263836 (2022).
44. Enayati, A. A., Vatandoost, H., Ladonni, H., Townson, H. & Hemingway, J. Molecular evidence for a kdr-like pyrethroid resistance mechanism in the malaria vector mosquito Anopheles stephensi. *Med. Vet. Entomol.* **17**, 138–144 (2003).
45. Tiwari, S., Ghosh, S. K., Ojha, V. P., Dash, A. P. & Raghavendra, K. Reduced susceptibility to selected synthetic pyrethroids in urban malaria vector Anopheles stephensi: A case study in Mangalore city, South India. *Malar. J.* **9**, (2010).
46. Tungu, P. K. *et al.* Large-scale (Phase III) evaluation of broflanilide 50WP (VECTRON™ T500) for indoor residual spraying for malaria vector control in Northeast Tanzania: study protocol for a two-arm, non-inferiority, cluster-randomised community trial. *BMC Infect. Dis.* **22**, (2022).
47. Portwood, N. M. *et al. Multi-centre discriminating concentration determination of broflanilide and potential for cross- resistance to other public health insecticides in Anopheles vector populations.* https://doi.org/10.21203/RS.3.RS-2022718/V1 (2022).
48. Liu, N., Feng, X. & Qiu, X. RDL mutations in Guangxi Anopheles sinensis populations along the China-Vietnam border: distribution frequency and evolutionary origin of A296S resistance allele. *Malar. J.* **19**, 23 (2020).
49. Grau-Bové, X. *et al.* Evolution of the insecticide target Rdl in African Anopheles is driven by interspecific and interkaryotypic introgression. *Mol. Biol. Evol.* **37**, 2900–2917 (2020).
50. Surendran, S. N. *et al.* Genotype and biotype of invasive Anopheles stephensi in Mannar Island of Sri Lanka. *Parasit. Vectors* **11**, 3 (2018).

51. Mishra, S., Sharma, G., Das, M. K., Pande, V. & Singh, O. P. Intragenomic sequence variations in the second internal transcribed spacer (ITS2) ribosomal DNA of the malaria vector Anopheles stephensi. *PLoS ONE* **16**, e0253173 (2021).
52. Alam, M. T., Bora, H., Das, M. K. & Sharma, Y. D. The type and mysorensis forms of the Anopheles stephensi (Diptera: Culicidae) in India exhibit identical ribosomal DNA ITS2 and domain-3 sequences. *Parasitol. Res.* **103**, 75 (2008).
53. 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).
54. WHO initiative to stop the spread of Anopheles stephensi in Africa. https://www.who.int/publications/i/item/WHO-UCN-GMP-2022.06.
55. Ochomo, E. O. *et al. Molecular surveillance leads to the first detection of Anopheles stephensi in Kenya.* https://doi.org/10.21203/RS.3.RS-2498485/V1 (2023).
56. Untergasser, A. *et al.* Primer3--new capabilities and interfaces. *Nucleic Acids Res.* **40**, (2012).
57. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinforma. Oxf. Engl.* **30**, 2114–2120 (2014).
58. H.Li. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. (2013).
59. tseemann/samclip: Filter SAM file for soft and hard clipped alignments. https://github.com/tseemann/samclip.
60. Garrison, E. & Marth, G. *Haplotype-based variant detection from short-read sequencing.* https://doi.org/10.48550/arxiv.1207.3907 (2012).
61. Poplin, R. *et al.* Scaling accurate genetic variant discovery to tens of thousands of samples. 201178 Preprint at https://doi.org/10.1101/201178 (2018).
62. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w(1118); iso-2; iso-3. *Fly (Austin)* **6**, 80–92 (2012).
63. Chakraborty, M. *et al.* Hidden genomic features of an invasive malaria vector, Anopheles stephensi, revealed by a chromosome-level genome assembly. *BMC Biol.* **19**, 28 (2021).
64. Larsson, A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30**, 3276–3278 (2014).
65. Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455 (2019).
66. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).
67. Paradis, E. pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics* **26**, 419–420 (2010).

Acknowledgements

H.A.P. is funded by a lido-dtp PhD studentship. E.C is funded by Medical Research Council LiD PhD studentships. J.P. was funded by a Newton Institutional Links Grant (British Council, no. 261868591). T.G.C. is funded by the MRC UK (Grant no. MR/M01360X/1, MR/N010469/1, MR/R025576/1, and MR/R020973/1). S.C. was funded by MRC UK grants (ref. MR/M01360X/1, MR/R025576/1, and MR/R020973/1). T.W. is funded by a Wellcome Trust/Royal Society Sir Henry Dale Fellowship (101285/Z/13/Z). L.A.M. is jointly funded by the Foreign, Commonwealth & Development Office (FCDO), the Medical Research Council, Wellcome and Department of Health and Social Care (Grant Ref: MR/R006040/1). M.K. was funded by the Foreign, Commonwealth & Development Office (FCDO) RPC and a Wellcome Trust Biomedical Resources grant.

Author contributions

T.G.C and S.C. conceived and directed the project. T.B., L.A.M., and F.T. provided samples and performed species identification. L.A.M, M.K., T.W., H.A.P. and S.C. undertook sample processing, DNA extraction and molecular assays. H.A.P., E.C., A.S. performed the bioinformatic analyses and interpreted results under the supervision of T.G.C. , S.C. J.P. and L.A.M. H.A.P. and S.C. wrote the first draft of the manuscript. All authors commented and edited on various versions of the draft manuscript and approved the final manuscript. T.G.C , L.A.M. and S.C. compiled the final manuscript.

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/s41598-023-32336-7.

Correspondence and requests for materials should be addressed to 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 Table 1. Variants Identified in each population

Sample	Sample Number	Number of SNPs	No. of Missense SNPs	Number of INDELs
All	127	115	4	27
Ethiopia	95	104	4	20
Colony	33	63	1	11

Supplementary Table 2. Locations of the four missense SNPs identified in all samples

Gene	Chromosome	Position	Amino Acid
ace1	NC_050202.1	60913884	N177D
gaba (rdl)	NC_050203.1	8353055	A296S
vgsc	NC_050203.1	42817709	L958F
gste2	NC_050203.1	70580373	V189L

Supplementary Table 3. Locations of insecticide resistance associated SNPs with targeted amplicons

Target Gene	Amplicon	Target SNP	Reference Species	Location in *An. stephensi*	Described in *An. stephensi*	First described in
ace1	ACE1_I	G119S	*T.californica*	G276S	N	*D. melanogaster*[1]
	ACE1_II	N485I		N639I	N	*Anopheles funestus*[2]
gste2	GSTe2	L119F	*An. funestus*	L119F	N	*An. funestus*[3]
vgsc	VGSCI	V410L	*M.domestica*	V371L	N	*Aedes aegypti*[4]
	VGSCII	L1014F		L958F	Y [5,6]	*An.gambiae*[7]
	VGSCIII	F1552C		F1569C	N	*Ae.aegypti*[8]
	VGSCIV	D1794Y		D1783Y	N	*Ae. aegypti*[9],
gaba (rdl)	RDL1	A302S	*D. melanogaster*	A296S	N	*D. melanogaster*[10]
	RDL2	V327I		V327I	N	*An. funestus*[11]
	RDL2	T345S/M		T345S	N	*Anopheles sinensis*[12]

Supplementary Table 4. Amplicon Primer sequences with target SNPs

Associated with	Target Gene	Amplicon	Accession ID	Target SNP	Exon Span	Forward primer	Reverse Primer	Product Size (bp)
Insecticide Resistance	*vgsc*	VGSCI	ASTEI08161	V410L	9	TTTCTCCAGCACCAAACATT	TGCTCCAAAAATGAACAAAAA	499
		VGSCII		L1014F	26	GATTGTGTTCCGTGTGCTGT	GGTTGGTAGCGGTAAGGTGA	498
		VGSCIII		F1152C	36	TCTTCGGATCGTTCTTCACC	AGATTCCACGGCTCGATAAA	496
		VGSCIV		D1794Y	39	AAGAGTGGTCTGGACGATGTG	GTAGTCGTCGTCGGTCAAGC	523
	ace1	ACE1_I	ASTE007197	G119S	3	AGGTTCCGTCATCCTCGAC	GGAAACAGCACCAGCACTCT	483
		ACE1_II		N485I	5	GTGGGCGATTACCATTTCAC	GGTGCTGCCACTTGTAGGTT	452
	gste2	GSTe2	ASTE016034	L119F	3	ATCATTACCGAGAGCCATGC	CACCGTTAGCCTCCTCGTAG	490
	gaba	rdl1	ASTE016089	A296S/V327I	7	AGTTTGTACGTTCGATGGGTTA	AGTGGCAGAAAGTGGTGTCC	498
		rdl2		T345S	8	GTCGTTCAGCGCACCACT	GTGAGCTTTCGGGTCGTG	399
Phylogeny	*cox1*	COI	AF116835.1	n/a	n/a	AATTAGGACACCCAGGAGCA	GCTCCAGCTAATACAGGTAATGA	490
	its2	ITS2	AY157678.1	n/a	n/a	TCGATGAAGACGCAGCTAAA	GCAACTGGATGCGAGGAA	492

Chapter 3: Genome-wide population genetics and molecular surveillance of insecticide resistance in *Anopheles stephensi* mosquitoes from Ethiopia

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	lsh1704757	**Title**	Miss
First Name(s)	Holly		
Surname/Family Name	Acford-Palmer		
Thesis Title	A genomic investigation into three malaria vectors: An. darlingi, An. funestus, and An. stephensi		
Primary Supervisor	Prof. Susana Campino		

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?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
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?	PLOS Genetics
Please list the paper's authors in the intended authorship order:	Holly Acford-Palmer, Emilia Manko, Jody Phelan, Fitsum G. Tadesse, Mojca Kristan, Ashley Osborne, Matthew Higgins, Teun Bousema, Louisa A. Messenger, Taane G. Clark & Susana Campino
Stage of publication	**Not yet 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 extracted and cleaned the DNA for sequencing. I then performed the bioinformatic analyses and interpreted results under the guidance of my supervisors. I wrote the first draft of the manuscript, which was then commented on by my supervisors and co-authors.

SECTION E

Student Signature	[redacted]
Date	

Supervisor Signature	[redacted]
Date	

Genome-wide population genetics and molecular surveillance of insecticide resistance in *Anopheles stephensi* mosquitoes from Ethiopia

Holly Acford-Palmer[1], Fitsum G. Tadesse[1,2], Emilia Manko[1], Jody E. Phelan[1], Matthew Higgins[1], Ashley Osborne[1], Mojca Kristan[1], Tom Walker[3], Teun Bousema[4], Louisa A. Messenger[5], Taane G. Clark[1,6], Susana Campino[1]

[1] Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom

[2] Malaria and NTD Directorate, Armauer Hansen Research Institute, ALERT Hospital Compound, Addis Ababa, Ethiopia

[3] School of Life Sciences, Gibbet Hill Campus, University of Warwick, Coventry, CV4 7AL, United Kingdom

[4] Department of Medical Microbiology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands

[5] Department of Environmental and Occupational Health, School of Public Health, University of Nevada, Las Vegas. Las Vegas, United States

[6] Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom

Corresponding author:

Prof. Susana Campino

Department of Infection Biology, Faculty of Infectious and Tropical Diseases

London School of Hygiene and Tropical Medicine

Susana.campino@lshtm.ac.uk

Abstract

Since the first detection of the Asian mosquito *Anopheles stephensi* in Dijbouti in 2012, this malaria vector has spread throughout the Horn of Africa. This invasive vector has continued to move across the continent, occupying a new niche in urban areas, it is becoming an increasing problem for malaria control programmes. Methods for vector control, e.g., the use of insecticide impregnated bed nets and insecticide spraying, have substantially reduced the global malaria burden. However, the increasing prevalence of mosquitos resistant to these measures, including *An. stephensi* populations, are threatening the reductions made so far. In this study, we generated whole genome sequencing data for *An. stephensi* sourced from a town in Ethiopia (n=27) and compared with populations in South Asia (n=45; India and Pakistan) to assess differences in genomic diversity, population structure, and uncover the presence of insecticide resistance markers. Using genome-wide single nucleotide polymorphisms (n=15,533,476), analysis of population structure revealed that the Ethiopian isolates cluster together forming a distinct ancestral group, separate from South Asian isolates. Three SNPs associated with insecticide resistance (*gaba* gene: A296S and V327I; *vgsc* L1014F) were detected. The *rdl*-A296S SNP was found in all populations analysed. Whilst the *kdr-L1014F* mutation was only identified in Ethiopian isolates, and the *rdl*-V327I SNP occurred only in Indian field samples. Evidence of ongoing selection was found in several loci, including genes previously associated with pesticide use (neonicotinoids), ivermectin resistance, as well as DDT and pyrethroid resistance. Notable genes included acetylcholine receptors (nAChR) subunits, glutamate-gated channel, *cytochrome P450 307a1-like* (*CYP307a1*), and the *gaba* subunit beta. Overall, this study represents the first genome-wide population genetics study of the invasive *An. stephensi* mosquito in the Horn of Africa, and reveals genomic differences from South Asian populations, which can be used for future assessments of vector movement and molecular surveillance of insecticide resistance.

Word count: 300

Introduction

The Asian malaria mosquito vector*, Anopheles stephensi*, was first detected in Africa in Djibouti in 2012[1]. Since then, this mosquito species has spread throughout the Horn of Africa (HOA), including Somalia, Ethiopia, and South Sudan, as well as Kenya, and with recent reports in Nigeria and Ghana[2–6]. Historically, *An. stephensi* was distributed across South Asia and throughout the Arabian Peninsula. It is a primary vector in India and Pakistan, proficiently transmitting both *Plasmodium falciparum* and *P. vivax*[7,8]. Its movement into the HOA and rapid expansion across the region has coincided with a surge in malaria cases and urban outbreaks. in Djibouti, malaria cases increased 40-fold between the years 2012 and 2020, likely due to the presence of *An. stephensi*[9,10]. In Ethiopia, the role of *An. stephensi* in regional urban malaria transmission was recently confirmed in an outbreak of *P. falciparum* in Ethiopia[11]. The species is filling a currently empty ecological niche, namely, the ability to breed and transmit malaria in urban environments by demonstrating the ability to breed in both small artificial water sources, and natural aquatic habitats near human dwellings [12,13]. The current primary vector in the Horn of Africa is *An. arabiensis*, which typically occupies rural areas or areas of high agricultural activity. *An. stephensi* is occupies. *An. arabiensis* takes advantage of the availability of large bodies of unpolluted water, however *An. stephensi* can make use of smaller artificial water sources as larval habitats[14]. Increasing urbanisation occurring across the African continent, coupled with the expanding geographical range *of An. stephensi*, means a further 120 million Africans are estimated to be at risk from malaria[6,13,15]. With the World Health Organization's (WHO's) target to reduce the global malaria burden by 90% by 2030, the need to control this vector has never been more crucial[16].

Vector control currently relies on long-lasting insecticide treated nets (LLINs), indoor residual spraying (IRS), and larval breeding site spraying. There are four main classes of adulticides in use: pyrethroids, organochlorides, organophosphates, and carbamates, each with varying modes of action. However, resistance to all four major adulticide classes have been reported for many *Anopheles* species, including for *An. stephensi* across the HOA, India, Pakistan, Sri Lanka, and the WHO Eastern Mediterranean region[3,17–19]. Resistance to insecticides can arise through multiple mechanisms: target site resistance, metabolically

77 mediated, behavioural changes, microbiome alterations, and thickening of the insect cuticle.
78 Predominantly, resistance results from metabolic or target site alterations. Target-site resistance arises
79 from single nucleotide polymorphisms (SNPs) that alter the targeted protein's amino acid sequence and
80 result in conformational changes that prevent the insecticide binding correctly. The target site mutation
81 A296S in the *GABA* receptor has been reported in *An. stephensi* from Ethiopia, and is associated with
82 resistance to dieldrin (*rdl),* an organochloride banned in the 1990s due to concerns about its impact on
83 human health[20–22]. The L1014 knockdown resistance (*kdr*) mutation in the *voltage-gated sodium channel*
84 *(vgsc*) is linked to pyrethroid resistance and has been reported in *An. stephensi* from Ethiopia, Afghanistan,
85 and India[21,23–26]. However, there have been examples of pyrethroid resistance in *An. stephensi* in the
86 absence of this mutation, implying other mechanisms may contribute to the development of resistance[19,27].
87 These alternative mechanisms could include metabolically mediated resistance, characterized by increased
88 detoxification of insecticides through the overexpression of *glutathione-s transferases* or *cytochrome*
89 *P450s*, thereby reducing insecticide efficacy[28,29]
90
91 To date, the detection of insecticide resistance markers and the population genetics of the invasive *An.*
92 *stephensi* species has been limited to the examination of a few candidate genes. Its emergence in Africa is
93 not fully understood, but one hypothesis is that *An. stephensi* was transported through either human
94 mediated travel or transport ships carrying cattle[30,31]. *An. stephensi* is both anthropophilic and zoophilic as
95 a vector, and the mosquito was first found in Africa in Djibouti city, located 20 kilometres from a trade
96 port[6,8]. Other theories include long-distance wind-borne migration[32].
97
98 The application of whole genome sequencing (WGS) provides a holistic view of the genomic landscape,
99 allowing for insights into mosquito ancestry, ongoing genetic selection, and the identification of insecticide
100 resistance markers. The availability of increasingly affordable and high throughput platforms for WGS, are
101 resulting in a growing number of *Anopheles* species with available genomic data[33–35]. Here, we generate the
102 first whole genome sequence data from *An. stephensi* collected in Africa and conduct population genetic

analysis of these Ethiopian *An. stephensi* WGS data (n=27) alongside publicly available data sourced from South Asia (India and Pakistan, n=45), to gain a deeper understanding of this vectors' genomic architecture, and gain insight into its ancestry, phylogenetics, and uncover loci under selective pressure.

Results

Whole Genome Sequence Data and Nucleotide Diversity

Whole genome data was generated from 28 mosquito isolates sourced from Ethiopia; mapping data identified one isolate as non-*stephensi* which was excluded from further analysis. This non-stephensi isolate was likely an *An. arabiensis* mosquito, with 96.28% of reads mapping to the *An. arabiensis* genome. The number of sequenced paired-reads across the 27 samples, ranged between 26,841,458 and 85,763,451, and after mapping to *An. stephensi*'s three chromosomes, this resulted in an average coverage of 33.7-fold (standard deviation 7.5-fold). A further 45 isolates with publicly available WGS data from India (Bangalore and Mangalore, N= 21; Colonies, N=16) and Pakistan (Colony, N=8) were also mapped, and across the combined dataset (N=72), 15,533,476 high quality SNPs were identified. Sliding window analysis (size 100kbp) revealed low nucleotide diversity with π averaging < 0.03 across *An. stephensi*'s three chromosomes **(Supp Fig. 1)**.

Population Differentiation and Ancestral Analysis Reveals Distinct Geographic Groups

Using the 15,533,476 high quality SNPs, both principal component and phylogenetic analysis revealed that the Ethiopian, Pakistan colony, and Indian field *An. stephensi* isolates formed distinct clusters (**Figure 1a and 1b**). The Indian field isolates also clustered with a subset of Indian colony mosquitoes, whilst the remaining (n=5) Indian colony isolates appear genetically closer to the Pakistani SDA500 colony isolates. This separation is based on their origin, the five clustering with Pakistani colony samples are the Walter Reid colony strain; and those appearing with the Indian field isolates are a different laboratory-reared colony strain. Application of the pairwise F_{ST} population differentiation metric, which quantifies differences in allele frequencies, confirmed that the greatest genetic distinctness was between Ethiopia and Pakistan

Colony samples (70,898 SNPs with $F_{ST} > 0.8$; 3,472 SNPs with $F_{ST} = 1$) **(Table 1)**. The most genetic similarity was between Indian field and colony samples (2,411 SNPs with $F_{ST} > 0.8$, 354 SNPs with Fst =1), and Ethiopia (7539 SNPs with $F_{ST} > 0.8$, 194 SNPs with Fst =1) (Supplementary Table 1).

The admixture analysis was performed to identify any potential ancestral relationships between the isolates, and revealed five ancestral populations: Ethiopian, Indian field, Pakistani colony, and the separation of Indian colony isolates, as observed in the phylogenetic tree. The Ethiopian isolates consisted of one population (K1) **(Figure 2)**. The K1 ancestral population was also observed in a single Pakistan colony and Indian isolates (14/21; mixed with K3). In the Indian isolates the K4 ancestry was dominant, and no clear distinction could be seen between Bangalore and Mangalore isolates. The K3 ancestry appeared as the only ancestry present in a subset of Indian colony samples. In the remaining Indian colony samples, K5 was dominant, although two isolates shared K2 ancestry that was predominant in colony samples from Pakistan. The single Pakistani colony sample sharing K1 ancestry with Ethiopia, also had K3 and K5 ancestral types seen in Indian wild-type and colony samples. This sample had been previously morphologically identified as the intermediate form of *An. stephensi.*

Selection Analysis Reveals Genes Involved with Insecticide Resistance

Genome-wide selection scans were performed across and within populations to identify signals of recent positive selection. In an across-population analysis, 275 SNPs were identified as having ongoing directional selection (XP-EHH > 4.0), with 148 SNPs across 81 genes, and 127 SNPs within introns (Supplementary Table 2). One gene was identified as a candidate potentially involved in insecticide resistance (*acetylcholine receptor (nAChR) subunit beta-like 2*) (two SNPs, XP-EHH values > 5.3) in a comparison of Indian and Ethiopian field samples **(Table 2)**.

In single population (iHS) analysis, a total of 997 loci were identified as having significant iHS scores, 439 were within non-coding regions and the remaining 559 occurred across 325 different genes (Supplementary Table 3). Of these, five were candidate genes that had loci with iHS scores indicative of selection, including

the *nAChR subunit beta-like* 2 in the Indian isolates. The remaining four genes (IHS values > 4.0) were *nAChR alpha-like* (Ethiopia and Indian wild samples), glutamate-gated chloride channel (Pakistani Colony), *cytochrome P450 307a1*-like (*CYP307a1*) (Indian isolates), and *gaba* subunit beta (India field samples and Pakistan colony).

Identification of Three Insecticide Resistance Associated SNPs

A total of 25 missense SNPs were identified across four genes previously associated with insecticide resistance in *Anopheles* mosquitoes (*ace-1*, *gaba* (*rdl*), *GSTe2*, and *vgsc*) **(Table 3)**. Three of these are SNPs known to be associated with target-site mediated resistance (*vgsc* L1014F; *gaba* A296S and V327I). The *vgsc* L1014F SNP mediates resistance to pyrethroids and was identified exclusively in the Ethiopian population. It was only identified as a heterozygous genotype in 2 samples (allelic frequency of 7.4% (2/27 samples)). The *gaba* A296S mutation was found in all 4 populations (allelic frequency of 18.8%; 27/72 samples), with almost all samples heterozygous at this position. Whilst the V327I alteration was only present in the Indian field and colony populations (5 samples). This SNP was also present as the heterozygous genotype, and only in samples that also had the *gaba* A296S mutation. A further 22 missense mutations were identified across the four key insecticide resistance associated genes (*ace-1 (n*=5), *gaba* (*n*=9), *GSTe2 (n*=9), *vgsc* (n=2)), with allelic frequencies ranging from 1% to 83% (**Table 3**).

Detection of Structural Variants: Copy Number Variants and INDELs

For coverage-based analysis, only one significant deviation from the median genome coverage for the Ethiopian population was observed (Supp fig 2). A possible copy number variant (CNV) was detected in a region containing a cluster of *cytochrome P450* genes (CYP). The *CYP6a* cluster on chromosome two (NC_050202.1) was observed to have elevated coverage, in comparison to both the median coverage and other populations.

From analysis with Delly, a total of 57,815 deletions were identified after quality control filtering. Of these, 36 were identified across 18 genes previously identified as involved in insecticide resistance, or as belonging

to a gene family possibly involved in insecticide metabolism. Two of these deletions, in *CYP307a1* and *CYP6a1* genes, were annotated as resulting in significant structural alterations. For *CYP307a1*, it was identified as a gene fusion event resulting from a 494bp deletion that was present in one Ethiopian sample, but upon further inspection was identified as an intronic variant. For *CYP6a1,* a frameshift variant was identified in a single Indian colony sample (SRR1529388). There were 2,901 duplications originally found across 44 genes and 25 intergenic regions. Of these, 351 (12.1%) resulted in a frameshift variant. One 181bp duplication event was identified in the *CYP9f2* candidate gene in one Indian colony sample (SRR1529388, see above), but was annotated as a non-coding transcript. Delly did not detect any Copy Number Variants (CNVs) in the chromosome 2 CYP6a cluster with elevated coverage identified in Supplementary Figure 2.

Discussion

The invasion of *An. stephensi* into the Horn of Africa is a significant threat to malaria control and elimination efforts in Sub-Saharan Africa. Understanding the origins of this invasive species and ongoing gene flow can provide greater insight into its emergence in this region and improve predictions of future spread. Here we explore the genomic landscape of *An. stephensi* and demonstrate, using population structure analysis, that our Ethiopian field isolates are genetically distinct from Indian and Pakistani colony populations. Ancestry analysis indicated that each field population constituted a distinct ancestral population. The Indian field isolates, despite the large geographical distance between the regions (Bengaluru and Mangaluru, around 350km), shared an ancestral population (K4) in the presence of several minor ones. The Indian colony samples were divided in two groups, based on their origin, (Walter Reid or laboratory-reared strains). All samples from Ethiopia inherited the K1 ancestry, which was present at a minor proportion in some wild-caught Indian samples, and a single Pakistan SDA500 colony sample (intermediate form). This observation suggests a possible South Asian origin of the Ethiopian samples, as previously reported[2]. Previous studies using candidate gene analysis have identified high levels of genetic diversity across Ethiopian populations, implying either a one large invasion incident, or multiple smaller colonisation events[31]. The samples

206 analysed here were collected in the same market town, which could have resulted from a population
207 expansion in this area, resulting in the homogenous ancestry observed.

208 The ancestry observed in Ethiopia might have been dominant in other areas in South Asia at the time of the
209 importation event. A similar relationship is observed between the Indian laboratory-reared and field
210 isolates, where the former was collected in 2016 in Chennai, India and have only K5 ancestry. This K5
211 ancestry also appears in 14 Indian field samples with the K1 Ethiopian ancestry, suggestive of shared
212 ancestry between the Indian populations, despite the large geographical distance between Chennai,
213 Bengaluru, and Mangaluru. Other studies have indicated that *Anopheles spp.* population structure remains
214 stable over time, and physical distance is a larger driver of genetic variation[36,37]. To further understand the
215 origins of the *An. stephensi* in the Horn of Africa, further WGS should be performed on more field isolates
216 from *An. stephensi's* native geographical regions, including India, Pakistan, Afghanistan, Iran, as well as
217 elsewhere in the Horn of Africa.
218
219 The *kdr* L1014F insecticide resistance mutation was detected at a low frequency in this Ethiopian
220 population, and absence in the Indian field samples. This *kdr* mutation confers resistance to pyrethroids
221 and has been found previously in Indian populations of *An. stephensi* collected in 2016, and in Afghan
222 populations, collected in 2018[24,25]. The *kdr* L1014F mutation is absent in the Mangalore and Bangalore
223 populations analysed here, despite evidence of extensive pyrethroid resistance in both these cities[3,38]. The
224 low proportion of Ethiopian isolates with this SNP, along with its presence as a heterozygous genotype,
225 implies it has recently arisen in this population. The *kdr* L1014F mutation has been previously identified in
226 Ethiopian samples at a similar frequency to our study[21].
227
228 The other known target site mutation identified in the Ethiopian isolates was *gaba* A296S, which confers
229 resistance to dieldrin[20,22]. This substitution was also identified in the Indian field populations, highlighting
230 its prevalence, despite the insecticide being banned since the 1990s. The presence of this mutation
231 corroborates the historical shared ancestry of the isolates analysed here. The *gaba* V327I mutation was also

identified, and has a strong association to the A296S alteration, with five Indian field and colony samples carrying both SNPs[39,40]. A further 22 putatively novel SNPs were identified in genes associated with insecticide resistance. Six of these SNPs appeared in two Pakistan Colony samples (Ste32 and Ste58) known to be insecticide susceptible. Two of these SNPs occurred in the *ace-1* gene (N177D and V59A), and four in *GSTe2* (F196Y, C146S, H97A, and G66A). The remaining 16 missense SNPs occur in populations where the resistance status is unknown, so it is not possible to infer any impact on insecticide susceptibility.

Of the genes identified as having ongoing directional selection, five could be potentially involved in insecticide resistance. The *nAChR receptor subunit beta* was found under selection by both within (iHS) and between (XP-EHH) population analysis. The *nAChR receptor subunit alpha* was also identified using the iHS metric as being under ongoing selection in Indian and Ethiopian populations. These two subunits of the receptor are the result of splicing of the *nAChR* gene; the presence (alpha) or absence (beta) of two cysteines determines their type[41]. Mutations within *nAChR* have been reported to result in resistance to neonicotinoids, which are pesticides that mediate synaptic transmission via nAChR, resulting in insect mortality[42,43]. This type of pesticide usage has been reported to result in neonicotinoid resistance in other *Anopheles species*[44]. Worryingly, neonicotinoids are considered an alternative to pyrethroids, for vectors with high levels of resistance to pyrethroids[3,45–47]. In Ethiopia, the President's Malaria Initiative have using SumiShield, which contains a neonicotinoid since 2021. In addition, significant directional selection was detected in glutamate-gated channel genes in the Pakistan colony. Mutations in these loci have been associated with ivermectin resistance in *Drosophila*[48–50]. Ivermectin is an anti-parasitic, often used in mass drug administrations, which also kills *Anopheles spp.* mosquitoes when they ingest the blood of a treated host[51]. Ivermectin has been trialled as a vector control method using mass drug administration to help reduce malaria cases[52].

Distinct signals of selection in a *gaba* gene were identified in India field and Pakistan colony samples, fipronil is an insecticide that acts on *gaba* receptor subunit beta to cause neurotoxicity[53]. This phenylpyrazole that has been suggested as a one health approach to vector control[54]. Similarly to ivermectin, fipronil can be

used in mass drug administration, particularly in livestock (e.g., cattle targeting zoophilic vectors (e.g., *An. stephensi*), where this approach has been trialled successfully[55–57]. Resistance to fipronil has been reported in Iranian *An. stephensi* isolates, with both *kdr* and *rdl* mutations hypothesised to be involved[58]. Mutations within GABA receptors have been associated with reduced insecticide efficacy of fipronil, although this has not been observed in *Anopheles spp*[59–61]. Further surveillance of this gene could provide valuable insights into its potential involvement in fipronil resistance.

Another notable gene exhibiting strong directional selection by iHS was *CYP307a1*, identified in the Indian field populations. A 494bp deletion was detected in the intronic region of this locus in one Ethiopian isolate, and may impact on gene expression or result in alternative splicing[62]. *CYP307a1* is a member of the cytochrome P450 gene family, and has previously been linked to resistance to both DDT and pyrethroids in *An. funestus*[63,64]. Similarly, in other insect species (*Cydia pomonella*), the upregulation of *CYP307a1* has been associated with deltamethrin (a pyrethroid) resistance[65]. Typically, *CYP307a1* is involved in ecdysteroid hormone biosynthesis, these hormones control mosquito behaviour, nervous system development, and reproduction[66]. This gene has not been confirmed to directly result in insecticide resistance but warrants further investigation.

Other structural variants were detected that resulted in amino acid alterations, including a 67bp deletion in the *CYP6a1* gene, found in a single Indian colony sample. This deletion was one of two detected in this gene, where the other was a 3' UTR variant found in both Indian colony and field samples. Altered expression levels of *CYP6a1* have previously been associated with deltamethrin resistance in *D. melanogaster* and *C. pipiens*[67,68]. With the absence of *kdr* mutations in these Indian field samples, but phenotypic pyrethroid resistance reported near the collection sites, it is likely other mechanisms contribute to resistance; such as metabolically mediated resistance[3,19,27,69].

This *CYP6a1* gene identified here is not in the *CYP6a* cluster that was observed to have elevated coverage proportional to the genome median in Ethiopian isolates. To further investigate this gene cluster, read

orientations and breakpoints would need to be identified to confirm whether this increased coverage was due to a copy number variation in the population.

Further investigations, using WGS or targeted amplicon sequencing, in tandem with susceptibility bioassays, are needed to investigate the impact of these mutations on insecticide response. The novel missense SNPs potentially linked with resistance should be used as targets in high-throughput molecular assays, to support surveillance and assist functional work to understand and validate underlying mechanisms associated with resistant phenotypes.

In conclusion, this study gives greater insight into the population genetics of cross-continental *An. stephensi*. Applications of WGS analysis to larger *An. stephensi* sample cohorts, across different geographical regions, will be key to understanding gene flow and identifying insecticide resistance markers. Such insights will enable public health authorities to make informed choices about vector surveillance and insecticide usage.

Methods

Mosquito Collection and Identification

An. stephensi mosquitoes were sourced from an LSHTM colony (Sind Kasur strain, originally from Pakistan in 1982). Field samples were collected in Awash Sebat Kilo, Ethiopia, between April and September 2019. They were collected in one of three ways: CDC mini light traps, aspiration from cattle shelters, and human landing collection. All mosquitoes were identified morphologically as *An. stephensi* before a multiplex qPCR with *ITS2* and *cox-1* genes was used for molecular confirmation[7,70].

DNA Extraction

Mosquitoes were individually suspended in 1X PBS, before being mechanically lysed with a Tissue Ruptor II (Qiagen, Hilden, Germany) for 30 seconds, or until all body parts were no longer visible. DNA was extracted

using the Qiagen Blood and Tissue kits, according to manufacturer's instructions. DNA concentrations for each sample were quantified using the Qubit 2.0 fluorimeter HS DNA kit (ThermoFisher). DNA was then stored at -20^{o}C.

Whole Genome Sequencing and Bioinformatic Analysis

The DNA of 30 *An. stephensi* isolates were sequenced on the Illumina MiSeq using 2 x 250bp paired end configuration. Twenty-eight of these isolates were collected in Ethiopia, and the remaining two were colony mosquitoes. Publicly available *An. stephensi* WGS data was included in the population analyses, this includes 21 wild-type Indian samples (11 from Bangalore, and 10 from Mangalore). A total of 24 publicly colony samples were included in analysis, 16 of which were Indian colony samples, and eight Pakistan colony samples. The raw WGS sequence data was first trimmed using trimmomatic software (version 0.39), before being aligned to the UCI_ANSTEP_V1.0 (An. stephensi) reference genome, using bwa-mem software (with default parameters)[71–73]. The isolate identified as non-*stephensi* was then mapped to available genomes from *Anopheles* species known to be present in the region. These included *An. gambiae* (AgamP4), *An. arabiensis* (AaraD3), *An. funestus* (idAnoFuneDA-416_04) and *An. coustani* (idAnoCousDA_361_x.2). and coverage statistics from the resulting bam files were calculated using samtools[74]. Variants (SNPs and INDELs) were called and validated using GATK software with the HaplotypeCaller[75]. Once individual VCF files had been generated, a multi-sample VCF was created using GATK's GenomicsDBImport and GenotypeGVCFs function. The multi-sample VCF was then filtered to contain only chromosomal variant using bcftools, this package was also used to sort and normalise multi-allelic variant sites. Further filtering was conducted using vcftools, removing variants with a depth > 5, more than 50% missingness[76]. A total of 16,580,599 SNPs were initially identified from the remaining 72 samples but reduced to 15,533,476 after filtering.

Identification of Insecticide Resistance Associated SNPs

A bed file containing genes associated with insecticide resistance was created based on a literature search. This search included the *para, gaba,* and *ace-1* genes associated with target site resistance, along with cytochrome P450's, and esterase's linked to metabolic-based resistance. The bed file included the 500bp before the gene start codon, and 500bp after the stop codon to identify any variants within the promoter or terminator regions. The bed file was then applied to the filtered multi-sample VCF using bcftools. The package snpEff was then used to annotate these variants, using a custom-built database from the UCI_ANSTEP_V1.0 gff file[71,77].

Population Genetic Analysis

A pairwise-genetic distance matrix was generated from the filtered multi-sample VCF file, using an in-house script[78]. This distance matrix was used as the basis for the generation of a neighbourhood joining tree, and principal component analyses generated in R using ape and qqman packages[79,80]. The resulting NJ tree was visualised and annotated using iTOL[81]. ADMIXTURE software (v1.3) was used to conduct admixture analysis[82]. PLINK package was first used to convert the VCF file to a bed file for these analyses[83]. The optimum K value (estimated number of ancestral populations) was calculated by cross-validation of 1-10 dimensions of eigenvalue decay (k=5). This value along with the bed file was used by ADMIXTURE software to identify shared ancestral populations. The output was then visualised in R. To investigate the genetic diversity in *An. stephensi* F_{ST} was calculated using the Weir and Cockerham estimator between Ethiopian and Indian field populations, using vcftools and visualised in the R statistical tool. Nucleotide diversity was also calculated across the genome, using 100kb windows. Genomic regions under directional selection, were detected with the R-based package rehh[84]. The Integrated haplotype statistic (iHS) was used to find selection within populations, and extended haplotype homozygosity (XP-EHH) was used to identify selection ongoing between populations.

Identification of Large Structural Variants

Two methods were utilised to detect copy number variants for this date set. First, a coverage-based method was used focussing on clusters of CYP genes identified in the genome (Supplementary Table 4). Sample coverage was averaged by collection location resulting in four populations: Ethiopia, Indian field, Indian Colony, and Pakistani Colony. Coverage for each population was normalised using Kernel-smoothing, and then plotted against to the median genome coverage depth for that population.

Secondly, Delly software was used to identify large structural variants (SVs) >60bp[85]. For individual samples that had insertions and deletions, sample BCF format files were merged and filtered based on sample missingness (<50%), followed by the removal of heterozygous calls. As described above, the population differentiation statistic F_{ST} was calculated using vcftools, to identify SVs unique to populations. To confirm any SVs occurring in genes associated with insecticide resistance, the bcftools software was used for visual inspection.

References

1. 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 Trop.* **139**, 39–43 (2014).
2. 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).
3. WHO. Global Malaria Programme: Malaria Threats Map. https://www.who.int/teams/global-malaria-programme/surveillance/malaria-threats-map.
4. 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. *Parasit. Vectors* **14**, 511 (2021).
5. Ali, S., Samake, J. N., Spear, J. & Carter, T. E. Morphological identification and genetic characterization of Anopheles stephensi in Somaliland. *Parasit. Vectors* **15**, 247 (2022).
6. Sinka, M. E. *et al.* A new malaria vector in Africa: Predicting the expansion range of Anopheles stephensi and identifying the urban populations at risk. *Proc. Natl. Acad. Sci.* **117**, 24900–24908 (2020).
7. Tadesse, F. G. *et al.* Anopheles stephensi Mosquitoes as Vectors of Plasmodium vivax and falciparum, Horn of Africa, 2019. *Emerg. Infect. Dis.* **27**, 603–607 (2021).
8. The dominant Anopheles vectors of human malaria in the Asia-Pacific region: occurrence data, distribution maps and bionomic précis | Parasites & Vectors | Full Text. https://parasitesandvectors.biomedcentral.com/articles/10.1186/1756-3305-4-89.
9. Seyfarth, M., Khaireh, B. A., Abdi, A. A., Bouh, S. M. & Faulde, M. K. Five years following first detection of Anopheles stephensi (Diptera: Culicidae) in Djibouti, Horn of Africa: populations established—malaria emerging. *Parasitol. Res.* **118**, 725–732 (2019).
10. Moussa, R. A. *et al.* Molecular investigation of malaria-infected patients in Djibouti city (2018–2021). *Malar. J.* **22**, 147 (2023).
11. 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).

12. 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. *Malar. J.* **22**, 187 (2023).
13. Takken, W. & Lindsay, S. Increased Threat of Urban Malaria from Anopheles stephensi Mosquitoes, Africa. *Emerg. Infect. Dis.* **25**, 1431–1433 (2019).
14. Surendran, S. N. *et al. Anthropogenic Factors Driving Recent Range Expansion of the Malaria Vector Anopheles Stephensi*. *Frontiers in Public Health* vol. 7 (2019).
15. WHO. *WHO Initiative to Stop the Spread of Anopheles Stephensi in Africa*. https://www.who.int/publications/i/item/WHO-UCN-GMP-2022.06 (2022).
16. Global Malaria Programme: WHO Global. *WHO World Malaria Report 2022*. *WHO*.
17. Balkew, M. *et al.* An update on the distribution, bionomics, and insecticide susceptibility of Anopheles stephensi in Ethiopia, 2018–2020. *Malar. J.* **20**, 263 (2021).
18. Yared, S. *et al.* Insecticide resistance in Anopheles stephensi in Somali Region, eastern Ethiopia. *Malar. J.* **19**, 180 (2020).
19. Enayati, A., Hanafi-Bojd, A. A., Sedaghat, M. M., Zaim, M. & Hemingway, J. Evolution of insecticide resistance and its mechanisms in Anopheles stephensi in the WHO Eastern Mediterranean Region. *Malar. J.* **19**, 258 (2020).
20. ffrench-Constant, R. H., Anthony, N., Aronstein, K., Rocheleau, T. & Stilwell, G. Cyclodiene Insecticide Resistance: From Molecular to Population Genetics. *Annu. Rev. Entomol.* **45**, 449–466 (2000).
21. 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).
22. Independent mutations in the Rdl locus confer dieldrin resistance to Anopheles gambiae and An. arabiensis - Du - 2005 - Insect Molecular Biology - Wiley Online Library. https://resjournals.onlinelibrary.wiley.com/doi/10.1111/j.1365-2583.2005.00544.x.
23. Samake, J. N. *et al.* Detection and population genetic analysis of kdr L1014F variant in eastern Ethiopian Anopheles stephensi. *Infect. Genet. Evol.* **99**, 105235 (2022).

24. Dykes, C. L. *et al.* Knockdown Resistance (kdr) Mutations in Indian Anopheles stephensi (Diptera: Culicidae) Populations. *J. Med. Entomol.* **53**, 315–320 (2016).
25. Safi, N. H. Z. *et al.* Status of insecticide resistance and its biochemical and molecular mechanisms in Anopheles stephensi (Diptera: Culicidae) from Afghanistan. *Malar. J.* **18**, 249 (2019).
26. 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).
27. Safi, N. H. Z. *et al.* Evidence of metabolic mechanisms playing a role in multiple insecticides resistance in Anopheles stephensi populations from Afghanistan. *Malar. J.* **16**, 100 (2017).
28. World Health Organisation. *Global Report on Insecticide Resistance in Malaria Vectors: 2010-2016 Global Malaria Programme*.
29. Liu, N. Insecticide Resistance in Mosquitoes: Impact, Mechanisms, and Research Directions. *Annu. Rev. Entomol.* **60**, 537–559 (2015).
30. Ahn, J., Sinka, M., Irish, S. & Zohdy, S. Modeling marine cargo traffic to identify countries in Africa with greatest risk of invasion by Anopheles stephensi. *Sci. Rep.* **13**, 876 (2023).
31. Carter, T. E. *et al.* Genetic diversity of Anopheles stephensi in Ethiopia provides insight into patterns of spread. *Parasit. Vectors* **14**, 602 (2021).
32. Lehmann, T. *et al.* Urban malaria may be spreading via the wind—here's why that's important. *Proc. Natl. Acad. Sci.* **120**, e2301666120 (2023).
33. Ingham, V. A. *et al.* Integration of whole genome sequencing and transcriptomics reveals a complex picture of the reestablishment of insecticide resistance in the major malaria vector Anopheles coluzzii. *PLoS Genet.* **17**, e1009970 (2021).
34. 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).
35. Consortium, A. gambiae 1000 G. *et al.* Genetic diversity of the African malaria vector Anopheles gambiae. *Nature* **552**, 96–100 (2017).

36. Population genetic structure of the malaria mosquito Anopheles arabiensis across Nigeria suggests range expansion - Onyabe - 2001 - Molecular Ecology - Wiley Online Library. https://onlinelibrary.wiley.com/doi/full/10.1046/j.0962-1083.2001.01387.x?sid=nlm%3Apubmed.

37. Kent, R. J., Mharakurwa, S. & Norris, D. E. Spatial and Temporal Genetic Structure of Anopheles arabiensis in Southern Zambia over Consecutive Wet and Drought Years. *Am. J. Trop. Med. Hyg.* **77**, 316–323 (2007).

38. Tiwari, S., Ghosh, S. K., Ojha, V. P., Dash, A. P. & Raghavendra, K. Reduced susceptibility to selected synthetic pyrethroids in urban malaria vector Anopheles stephensi: a case study in Mangalore city, South India. *Malar. J.* **9**, 179 (2010).

39. Yang, C., Huang, Z., Li, M., Feng, X. & Qiu, X. RDL mutations predict multiple insecticide resistance in Anopheles sinensis in Guangxi, China. *Malar. J.* **16**, 482 (2017).

40. Liu, N., Feng, X. & Qiu, X. RDL mutations in Guangxi Anopheles sinensis populations along the China–Vietnam border: distribution frequency and evolutionary origin of A296S resistance allele. *Malar. J.* **19**, 23 (2020).

41. Jones, A. K., Buckingham, S. D., Brown, L. A. & Sattelle, D. B. Alternative splicing of the Anopheles gambiae nicotinic acetylcholine receptor, Agamαβ9, generates both alpha and beta subunits. *Invert. Neurosci.* **9**, 77 (2009).

42. Shimada, S. *et al.* The mechanism of loop C-neonicotinoid interactions at insect nicotinic acetylcholine receptor α1 subunit predicts resistance emergence in pests. *Sci. Rep.* **10**, 7529 (2020).

43. Bass, C. *et al.* Mutation of a nicotinic acetylcholine receptor β subunit is associated with resistance to neonicotinoid insecticides in the aphid Myzus persicae. *BMC Neurosci.* **12**, 51 (2011).

44. Mouhamadou, C. S. *et al.* Evidence of insecticide resistance selection in wild Anopheles coluzzii mosquitoes due to agricultural pesticide use. *Infect. Dis. Poverty* **8**, 64 (2019).

45. Zoh, M. G. *et al.* Experimental evolution supports the potential of neonicotinoid-pyrethroid combination for managing insecticide resistance in malaria vectors. *Sci. Rep.* **11**, 19501 (2021).

46. Oxborough, R. M. *et al.* Susceptibility testing of Anopheles malaria vectors with the neonicotinoid insecticide clothianidin; results from 16 African countries, in preparation for indoor residual spraying with new insecticide formulations. *Malar. J.* **18**, 264 (2019).
47. Tchouakui, M. *et al.* Comparative study of the effect of solvents on the efficacy of neonicotinoid insecticides against malaria vector populations across Africa. *Infect. Dis. Poverty* **11**, 35 (2022).
48. Kane, N. S. *et al.* Drug-resistant Drosophila indicate glutamate-gated chloride channels are targets for the antiparasitics nodulisporic acid and ivermectin. *Proc. Natl. Acad. Sci.* **97**, 13949–13954 (2000).
49. McCavera, S., Rogers, A. T., Yates, D. M., Woods, D. J. & Wolstenholme, A. J. An Ivermectin-Sensitive Glutamate-Gated Chloride Channel from the Parasitic Nematode Haemonchus contortus. *Mol. Pharmacol.* **75**, 1347–1355 (2009).
50. Atif, M. *et al.* GluClR-mediated inhibitory postsynaptic currents reveal targets for ivermectin and potential mechanisms of ivermectin resistance. *PLOS Pathog.* **15**, e1007570 (2019).
51. Chaccour, C., Lines, J. & Whitty, C. J. M. Effect of Ivermectin on Anopheles gambiae Mosquitoes Fed on Humans: The Potential of Oral Insecticides in Malaria Control. *J. Infect. Dis.* **202**, 113–116 (2010).
52. Marathe, A. *et al.* Potential impact of 5 years of ivermectin mass drug administration on malaria outcomes in high burden countries. *BMJ Glob. Health* **6**, e006424 (2021).
53. Ikeda, T., Nagata, K., Kono, Y., Yeh, J. Z. & Narahashi, T. Fipronil modulation of GABAA receptor single-channel currents. *Pest Manag. Sci.* **60**, 487–492 (2004).
54. Ruiz-Castillo, P., Rist, C., Rabinovich, R. & Chaccour, C. Insecticide-treated livestock: a potential One Health approach to malaria control in Africa. *Trends Parasitol.* **38**, 112–123 (2022).
55. Makhanthisa, T. I., Braack, L. & Lutermann, H. The effect of cattle-administered ivermectin and fipronil on the mortality and fecundity of Anopheles arabiensis Patton. *Parasit. Vectors* **14**, 349 (2021).
56. Dreyer, S. M. *et al.* Fipronil and ivermectin treatment of cattle reduced the survival and ovarian development of field-collected Anopheles albimanus in a pilot trial conducted in northern Belize. *Malar. J.* **18**, 296 (2019).
57. Poché, R. M. *et al.* Preliminary efficacy investigations of oral fipronil against Anopheles arabiensis when administered to Zebu cattle (Bos indicus) under field conditions. *Acta Trop.* **176**, 126–133 (2017).

512 58. Davari, B. *et al.* Selection of Anopheles stephensi with DDT and dieldrin and cross-resistance spectrum to
513 pyrethroids and fipronil. *Pestic. Biochem. Physiol.* **89**, 97–103 (2007).
514 59. Zhang, Y. *et al.* Synergistic and compensatory effects of two point mutations conferring target-site
515 resistance to fipronil in the insect GABA receptor RDL. *Sci. Rep.* **6**, 32335 (2016).
516 60. Nakao, T. Mechanisms of resistance to insecticides targeting RDL GABA receptors in planthoppers.
517 *NeuroToxicology* **60**, 293–298 (2017).
518 61. Tian, Y., Gao, Y., Chen, Y., Liu, G. & Ju, X. Identification of The Fipronil Resistance Associated Mutations in
519 Nilaparvata lugens GABA Receptors by Molecular Modeling. *Molecules* **24**, 4116 (2019).
520 62. How introns enhance gene expression - ScienceDirect.
521 https://www.sciencedirect.com/science/article/pii/S1357272517301541?via%3Dihub.
522 63. Tchigossou, G. *et al.* Molecular basis of permethrin and DDT resistance in an Anopheles funestus
523 population from Benin. *Parasit. Vectors* **11**, 602 (2018).
524 64. Sandeu, M. M., Mulamba, C., Weedall, G. D. & Wondji, C. S. A differential expression of pyrethroid
525 resistance genes in the malaria vector Anopheles funestus across Uganda is associated with patterns of
526 gene flow. *PLoS ONE* **15**, e0240743 (2020).
527 65. Insects | Free Full-Text | Genome-Wide Selective Signature Analysis Revealed Insecticide Resistance
528 Mechanisms in Cydia pomonella. https://www.mdpi.com/2075-4450/13/1/2.
529 66. Pondeville, E. *et al.* Microarray and RNAi Analysis of P450s in Anopheles gambiae Male and Female
530 Steroidogenic Tissues: CYP307A1 Is Required for Ecdysteroid Synthesis. *PLOS ONE* **8**, e79861 (2013).
531 67. Sun, Y. *et al.* Functional characterization of an arrestin gene on insecticide resistance of Culex pipiens
532 pallens. *Parasit. Vectors* **5**, 134 (2012).
533 68. Cariño, F. A., Koener, J. F., Plapp, F. W. & Feyereisen, R. Constitutive overexpression of the cytochrome
534 P450 gene CYP6A1 in a house fly strain with metabolic resistance to insecticides. *Insect Biochem. Mol.*
535 *Biol.* **24**, 411–418 (1994).
536 69. Shetty, V., Sanil, D. & Shetty, N. J. Insecticide susceptibility status in three medically important species of
537 mosquitoes, Anopheles stephensi, Aedes aegypti and Culex quinquefasciatus, from Bruhat Bengaluru
538 Mahanagara Palike, Karnataka, India. *Pest Manag. Sci.* **69**, 257–267 (2013).

70. 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.* **3**, 294–299 (1994).

71. Chakraborty, M. *et al.* Hidden genomic features of an invasive malaria vector, Anopheles stephensi, revealed by a chromosome-level genome assembly. *BMC Biol.* **19**, 28 (2021).

72. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinforma. Oxf. Engl.* **30**, 2114–2120 (2014).

73. H.Li. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arxiv.org* **n/a**, n/a (2013).

74. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10**, 1–4 (2021).

75. Poplin, R. *et al.* Scaling accurate genetic variant discovery to tens of thousands of samples. 201178 Preprint at https://doi.org/10.1101/201178 (2018).

76. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

77. Cingolani, P. *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 (Austin)* **6**, 80–92 (2012).

78. fastq2matrix/scripts at master · LSHTMPathogenSeqLab/fastq2matrix. *GitHub* https://github.com/LSHTMPathogenSeqLab/fastq2matrix/tree/master/scripts.

79. Turner, S. D. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *J. Open Source Softw.* **3**, 731 (2018).

80. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).

81. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).

82. Alexander, D. H. & Lange, K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* **12**, 246 (2011).

83. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
84. rehh 2.0: a reimplementation of the R package rehh to detect positive selection from haplotype structure - Gautier - 2017 - Molecular Ecology Resources - Wiley Online Library. https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.12634.
85. Rausch, T. *et al.* DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**, i333–i339 (2012).

Acknowledgements

H.A.P. is funded by a LIDo-DTP PhD studentship. J.P. was funded by a Newton Institutional Links Grant (British Council, no. 261868591). T.G.C. is funded by the MRC UK (Grant no. MR/M01360X/1, MR/N010469/1, MR/R025576/1, and MR/R020973/1). S.C. was funded by MRC UK grants (ref. MR/M01360X/1, MR/R025576/1, and MR/R020973/1). T.W. is funded by a Wellcome Trust/Royal Society Sir Henry Dale Fellowship (101285/Z/13/Z). L.A.M. is jointly funded by the Foreign, Commonwealth & Development Office (FCDO), the Medical Research Council, Wellcome and Department of Health and Social Care (Grant Ref: MR/R006040/1). M.K. was funded by the Foreign, Commonwealth & Development Office (FCDO) RPC and a Wellcome Trust Biomedical Resources grant.

Data availability

All raw data generated in this work for Ethiopian *An. stephensi* is publicly available (see PRJEB66077 for accession numbers). Accession numbers for publicly available raw data used in this study can be found in supplementary table 5.

Author contributions statement

T.G.C and S.C. conceived and directed the project. T.B., L.A.M., and F.T. provided samples and performed species identification. L.A.M, M.K., T.W. undertook sample processing, DNA extraction and molecular identification. H.A-P., performed the bioinformatic analyses and interpreted results, with the assistance of

E.M, J.P, M.H and A.O, and under the supervision of T.G.C., S.C. J.P. and L.A.M. H.A.P. wrote the first draft of the manuscript. All authors commented and edited the draft manuscript and approved the final manuscript.

ADDITIONAL INFORMATION

Ethics approval and consent

Study protocol was approved by the Institutional Ethical Review Board of the Aklilu Lemma Institute of Pathobiology of Addis Ababa University (ALIPB IRB/025/2011/2019), the Oromia Regional Health Bureau (BEFO/MBTFH/1331), and AHRI/ALERT Ethics Review Committee (AF-10-015.1, PO07/19).

Consent for publication

All authors have consented to the publication of this manuscript.

Availability of data and materials

Analysis scripts are available at https://github.com/LSHTMPathogenSeqLab

Competing interests

The authors declare no conflicts of interest.

Tables and Figures

Table 1. Number of SNPs with significant Pairwise Fst Calculations

Populations	Fst > 0.8	Fst = 1.0
Ethiopia vs India Wildtype	7539	194
Ethiopia vs India Colony	30567	938
Ethiopia vs Pakistan Colony	70898	3472
India Wildtype vs India Colony	2411	355
India Wildtype vs Pakistan Colony	35077	2942
India Colony vs Pakistan Colony	34479	3390

Table 2. Genes identified as undergoing directional selection with either XP-EHH or iHS statistics

XP-EHH				
Chromosome	**Position**	**Score**	**Gene**	**Populations**
NC_050202.1	68720664	5.30439559	acetylcholine receptor subunit beta-like 2	Ethiopia|India
NC_050202.1	68720666	5.30439559	acetylcholine receptor subunit beta-like 2	Ethiopia|India
iHS				
Chromosome	**Position**	**Score**	**Gene**	**Population**
NC_050201.1	13607505	4.157532377	gamma-aminobutyric acid receptor subunit beta-like	Pakistan colony
NC_050202.1	68736715	4.16320592	acetylcholine receptor subunit beta-like 2	India wildtype
NC_050202.1	68736717	4.24447916	acetylcholine receptor subunit beta-like 2	India wildtype
NC_050202.1	68789775	-4.5335888	acetylcholine receptor subunit alpha-like	India wildtype
NC_050202.1	68790027	-4.2056141	acetylcholine receptor subunit alpha-like	India wildtype
NC_050202.1	68790037	-4.5254523	acetylcholine receptor subunit alpha-like	India wildtype
NC_050202.1	68790043	-4.5118037	acetylcholine receptor subunit alpha-like	India wildtype
NC_050202.1	68874338	3.97023238	acetylcholine receptor subunit alpha-like	Ethiopia
NC_050203.1	8347167	4.26288047	gamma-aminobutyric acid receptor subunit beta	India wildtype
NC_050203.1	20668384	3.94395547	glutamate-gated chloride channel	Pakistan colony
NC_050203.1	20668385	3.94395547	glutamate-gated chloride channel	Pakistan colony
NC_050203.1	67922600	4.0921626	cytochrome P450 307a1-like	India wildtype
NC_050203.1	67925612	3.91798343	cytochrome P450 307a1-like	India wildtype
NC_050203.1	67925620	3.92503232	cytochrome P450 307a1-like	India wildtype
NC_050203.1	67925893	4.3089353	cytochrome P450 307a1-like	India wildtype
NC_050203.1	67926103	4.25137268	cytochrome P450 307a1-like	India wildtype
NC_050203.1	67926118	4.03362144	cytochrome P450 307a1-like	India wildtype
NC_050203.1	67926235	4.11747932	cytochrome P450 307a1-like	India wildtype

Table 3. Missense SNPs identified across four main insecticide resistance associated genes

Chromosome	Position	Amino acid alteration	Populations	Allele frequency		Genotype frequency		
				0	1	0/0	0/1	1/1
NC_050202.1	60913808	Val189Ile	E, IC & IW	0.91	0.09	0.83	0.16	0.01
NC_050202.1	60913844	Asn177Asp	All	0.17	0.83	0.07	0.21	0.72
NC_050202.1	60916071	Gly94Ser	All	0.22	0.78	0.07	0.31	0.63
NC_050202.1	60916173	Gly60Ser	PC	0.99	0.01	0.99	0.01	0.00
NC_050202.1	60916175	Val59Ala	All	0.24	0.76	0.13	0.22	0.65
NC_050203.1	8349210	Met349Ile	PC	0.99	0.01	0.99	0.01	0.00
NC_050203.1	8352962	Val327Ile	IC & IW	0.97	0.03	0.93	0.07	0.00
NC_050203.1	8353055	Ala296Ser	All	0.81	0.19	0.63	0.36	0.01
NC_050203.1	8380167	Leu101Ser	IC	0.98	0.02	0.96	0.04	0.00
NC_050203.1	8391526	Pro77Ala	Ethiopia	0.97	0.03	0.93	0.07	0.00
NC_050203.1	8391528	Pro76Gln	IW	0.99	0.01	0.97	0.03	0.00
NC_050203.1	8391529	Pro76Thr	Ethiopia	0.99	0.01	0.99	0.01	0.00
NC_050203.1	8395314	Gly74Glu	IC	0.95	0.05	0.90	0.10	0.00
NC_050203.1	8395317	Tyr73Cys	Ethiopia	0.96	0.04	0.93	0.06	0.01
NC_050203.1	42808541	Ala1955Cys	Ethiopia	0.96	0.04	0.97	0.03	0.00
NC_050203.1	42817709	Leu958F	Ethiopia	0.99	0.01	0.94	0.06	0.00
NC_050203.1	70580273	Thr222Ala	E & PC	0.97	0.03	0.90	0.10	0.00
NC_050203.1	70581404	Phe196Tyr	E & PC	0.95	0.05	0.08	0.06	0.86
NC_050203.1	70581440	Ile184Thr	IC	0.11	0.89	0.98	0.01	0.01
NC_050203.1	70581534	Arg153Cys	All	0.98	0.02	0.57	0.26	0.17
NC_050203.1	70581555	Cys146Ser	All	0.70	0.30	0.08	0.10	0.82
NC_050203.1	70581607	Asp128Glu	E & PC	0.93	0.07	0.90	0.06	0.04
NC_050203.1	70581621	Pro124Thr	Ethiopia	0.98	0.02	0.96	0.04	0.00
NC_050203.1	70581766	His97Asn	All	0.21	0.79	0.14	0.14	0.72
NC_050203.1	70581858	Gly66Ala	All	0.24	0.76	0.18	0.11	0.71

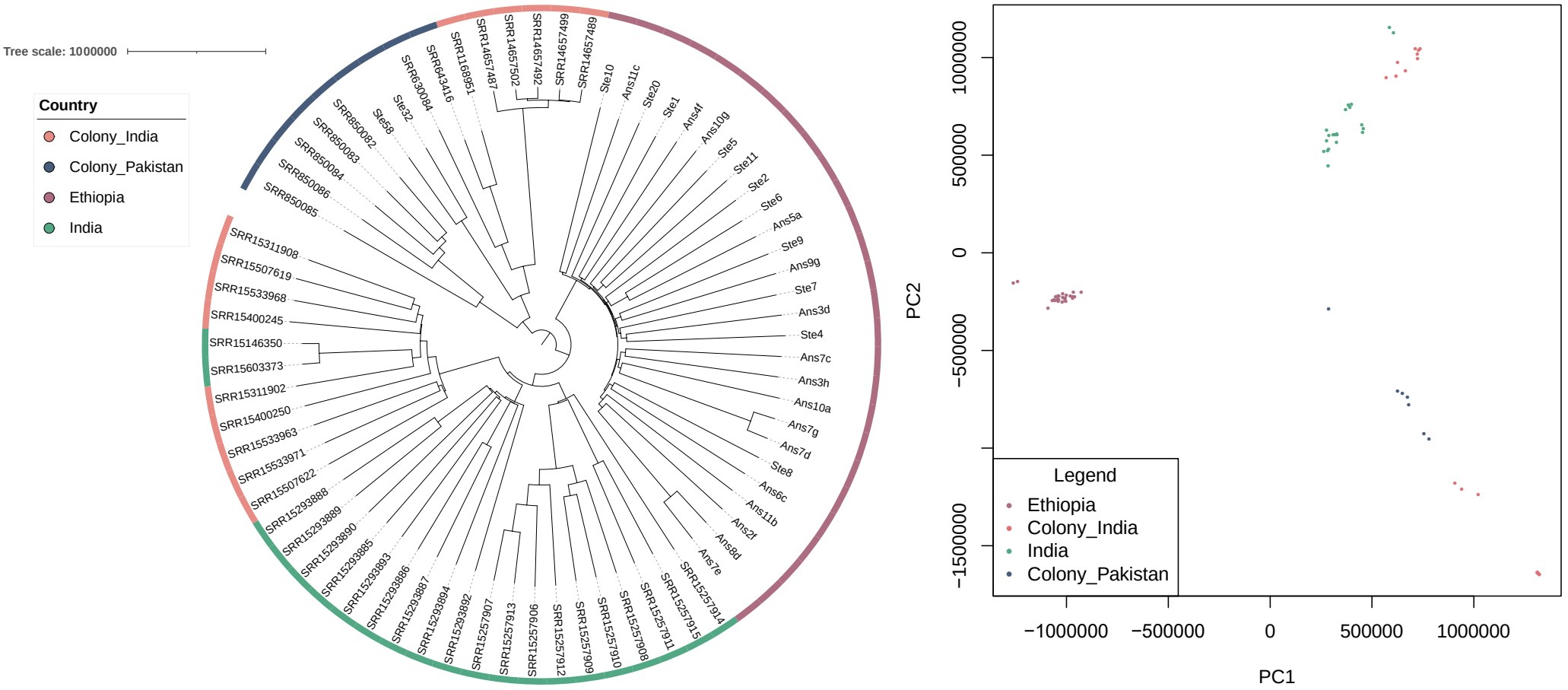


Figure 1. Neighbourhood joining tree, and principal component analysis plot generated from pairwise distance matrix of available *An. stephensi* WGS isolates (Ethiopia (n= 27), Indian colony (n=21), Indian wildtype (n=16), and Pakistan colony (n=8)

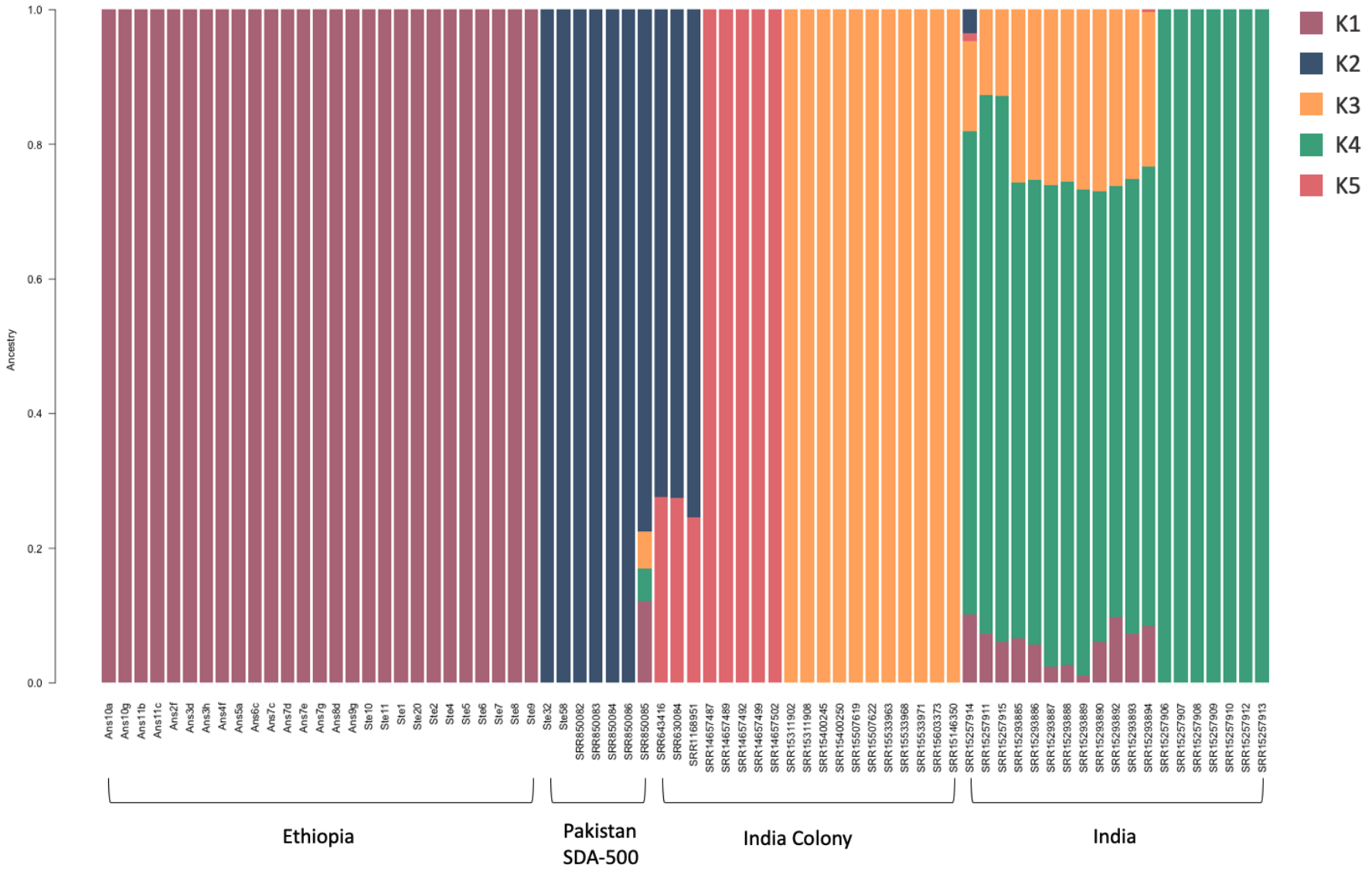
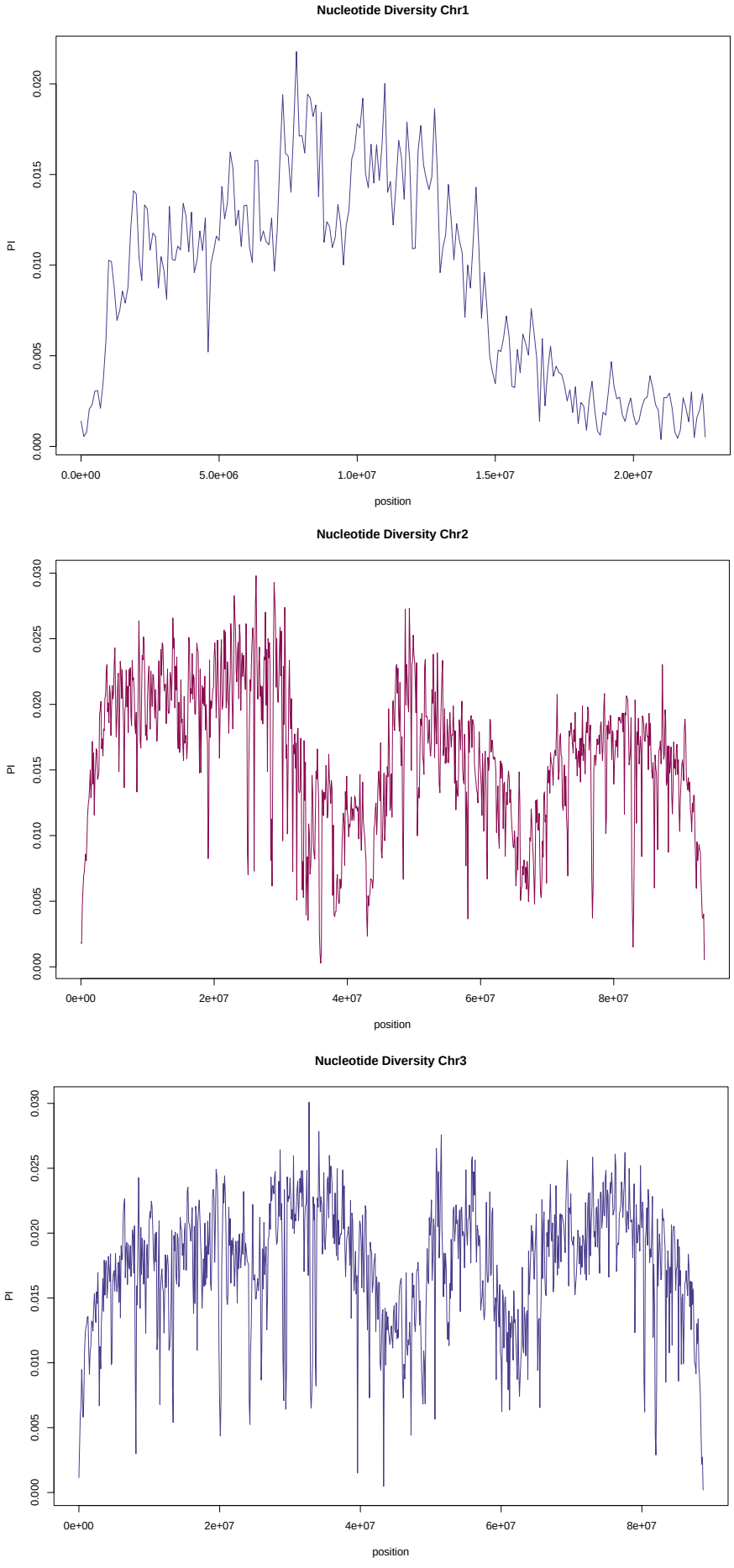


Figure 2. Genome wide Admixture analysis of An. stephensi isolates

Each isolate is a column. Five ancestral populations were identified (k=5), across the four isolate groups (Ethiopia (n= 27), Indian colony (n=21), Indian wildtype (n=16), and Pakistan colony (n=8)) analysed



Supp. Figure 1. Sliding window analysis of nucleotide diversity in *An. stephensi* by chromosome
Calculated across populations in 100kb windows, (pi) represents the average number of nucleotide differences per site. Sliding window analysis generated using vcftools and visualised in R.

Supplementary Table 1. Positions within genes identified as having significant XP-EHH scores.

Chromosome	Position	XPEHH Score	Reference allele	Alternate Allele	Gene_name	category_name
50201	9465180	-5.005665328	A	C	sodium channel protein Nach	Ethiopia|India Wildtype
50201	9465180	5.005665328	A	C	sodium channel protein Nach	Ethiopia|India Wildtype
50201	9465189	-4.444481662	G	A	sodium channel protein Nach	Ethiopia|India Wildtype
50201	9465189	4.444481662	G	A	sodium channel protein Nach	Ethiopia|India Wildtype
50201	9467621	-4.616753578	T	G	sodium channel protein Nach	Ethiopia|India Wildtype
50201	9467621	4.616753578	T	G	sodium channel protein Nach	Ethiopia|India Wildtype
50201	9480436	-4.458442882	A	G	sodium channel protein Nach	Ethiopia|India Wildtype
50201	9480436	4.458442882	A	G	sodium channel protein Nach	Ethiopia|India Wildtype
50201	9480437	-4.458442882	A	C	sodium channel protein Nach	Ethiopia|India Wildtype
50201	9480437	4.458442882	A	C	sodium channel protein Nach	Ethiopia|India Wildtype
50201	13100960	-4.709717919	T	G	AF4/FMR2 family member lilli-like	India Colony|Ethiopia
50201	13100960	4.709717919	T	G	AF4/FMR2 family member lilli-like	India Colony|Ethiopia
50202	1055305	-5.113814473	C	A	POU domain, class 6, transcription factor 2	India Colony|India Wildtype
50202	1055305	5.113814473	C	A	POU domain, class 6, transcription factor 2	India Colony|India Wildtype
50202	2083985	-4.55636738	C	T	endoribonuclease Dicer	Ethiopia|India Wildtype
50202	2083985	4.55636738	C	T	endoribonuclease Dicer	Ethiopia|India Wildtype
50202	4742060	4.599331789	C	A	uncharacterized LOC118508163	Ethiopia|India Wildtype
50202	4742060	-4.599331789	C	A	uncharacterized LOC118508163	Ethiopia|India Wildtype
50202	5592348	-4.530278196	T	A	limbic system-associated membrane protein	India Colony|Ethiopia
50202	5592348	4.530278196	T	A	limbic system-associated membrane protein	India Colony|Ethiopia
50202	12062186	-4.727903879	G	T	speract receptor	India Colony|India Wildtype
50202	12062186	4.727903879	G	T	speract receptor	India Colony|India Wildtype
50202	12064274	-4.521993755	C	T	speract receptor	India Colony|India Wildtype
50202	12064274	4.521993755	C	T	speract receptor	India Colony|India Wildtype

50202	12064324	-5.091355371	C	T	speract receptor	India Colony|India Wildtype
50202	12064324	5.091355371	C	T	speract receptor	India Colony|India Wildtype
50202	17196920	-4.762553453	T	G	meiotic recombination protein SPO11	Ethiopia|India Wildtype
50202	17196920	4.762553453	T	G	meiotic recombination protein SPO11	Ethiopia|India Wildtype
50202	21207678	-4.516864651	C	A	ankyrin repeat domain-containing protein 17	India Colony|India Wildtype
50202	21207678	4.516864651	C	A	ankyrin repeat domain-containing protein 17	India Colony|India Wildtype
50202	21207681	-5.094011005	C	T	ankyrin repeat domain-containing protein 17	India Colony|India Wildtype
50202	21207681	5.094011005	C	T	ankyrin repeat domain-containing protein 17	India Colony|India Wildtype
50202	21356934	-4.510156561	G	A	ankyrin repeat domain-containing protein 17	India Colony|India Wildtype
50202	21356934	4.510156561	G	A	ankyrin repeat domain-containing protein 17	India Colony|India Wildtype
50202	21385853	-5.059016074	A	G	ankyrin repeat domain-containing protein 17	India Colony|India Wildtype
50202	21385853	5.059016074	A	G	ankyrin repeat domain-containing protein 17	India Colony|India Wildtype
50202	25400970	-4.421035661	T	C	protein germ cell-less	India Colony|India Wildtype
50202	25400970	4.421035661	T	C	protein germ cell-less	India Colony|India Wildtype
50202	27087057	-4.537178445	G	A	growth factor receptor-bound protein 2	India Colony|India Wildtype
50202	27087057	4.537178445	G	A	growth factor receptor-bound protein 2	India Colony|India Wildtype
50202	35316698	-4.497163664	G	C	uncharacterized LOC118506318	India Colony|India Wildtype
50202	35316698	4.497163664	G	C	uncharacterized LOC118506318	India Colony|India Wildtype
50202	36554735	-4.670777572	A	T	N-acetylgalactosaminyltransferase 4-like	India Colony|India Wildtype
50202	36554735	4.670777572	A	T	N-acetylgalactosaminyltransferase 4-like	India Colony|India Wildtype
50202	36622449	-4.425557617	A	C	protein gooseberry-neuro	India Colony|India Wildtype
50202	36622449	4.425557617	A	C	protein gooseberry-neuro	India Colony|India Wildtype
50202	36622455	-4.464974184	C	G	protein gooseberry-neuro	India Colony|India Wildtype
50202	36622455	4.464974184	C	G	protein gooseberry-neuro	India Colony|India Wildtype
50202	37475265	-4.453004824	G	T	F-box/SPRY domain-containing protein 1	India Colony|India Wildtype
50202	37475265	4.453004824	G	T	F-box/SPRY domain-containing protein 1	India Colony|India Wildtype
50202	38473284	-4.708242107	C	T	homeobox protein abdominal-B-like	India Colony|India Wildtype

50202	38473284	4.708242107	C	T	homeobox protein abdominal-B-like	India Colony|India Wildtype
50202	39404435	-4.555763217	T	C	uncharacterized LOC118507833	India Colony|India Wildtype
50202	39404435	4.555763217	T	C	uncharacterized LOC118507833	India Colony|India Wildtype
50202	39818059	-4.595908924	C	T	uncharacterized LOC118507877	India Colony|India Wildtype
50202	39818059	4.595908924	C	T	uncharacterized LOC118507877	India Colony|India Wildtype
50202	41527499	-4.41754895	C	T	dopamine D2-like recepto	India Colony|India Wildtype
50202	41527499	4.41754895	C	T	dopamine D2-like recepto	India Colony|India Wildtype
50202	41861815	-4.796144402	T	C	lethal(2) giant larvae protein	India Colony|India Wildtype
50202	41861815	4.796144402	T	C	lethal(2) giant larvae protein	India Colony|India Wildtype
50202	48706614	-4.800396475	C	G	centrosomin	India Colony|India Wildtype
50202	48706614	4.800396475	C	G	centrosomin	India Colony|India Wildtype
50202	52988816	-4.641803851	T	A	cell wall protein AWA1	India Colony|India Wildtype
50202	52988816	4.641803851	T	A	cell wall protein AWA1	India Colony|India Wildtype
50202	52996720	-4.562666071	G	A	cell wall protein AWA1	India Colony|India Wildtype
50202	52996720	4.562666071	G	A	cell wall protein AWA1	India Colony|India Wildtype
50202	53039951	-5.773605556	C	A	cell wall protein AWA1	India Colony|India Wildtype
50202	53039951	5.773605556	C	A	cell wall protein AWA1	India Colony|India Wildtype
50202	53047183	-4.441675247	C	T	cell wall protein AWA1	India Colony|India Wildtype
50202	53047183	4.441675247	C	T	cell wall protein AWA1	India Colony|India Wildtype
50202	59287030	-4.516720667	C	A	nicotinamidase	India Colony|Ethiopia
50202	59287030	4.516720667	C	A	nicotinamidase	India Colony|Ethiopia
50202	59667235	-4.431415707	A	G	cubilin	India Colony|Ethiopia
50202	59667235	4.431415707	A	G	cubilin	India Colony|Ethiopia
50202	60835853	-4.475126562	C	T	pre-mRNA 3' end processing protein WDR33	India Colony|Ethiopia
50202	60835853	4.475126562	C	T	pre-mRNA 3' end processing protein WDR33	India Colony|Ethiopia
50202	61257455	-4.509738748	T	C	TBC1 domain family member whacked	India Colony|India Wildtype
50202	61257455	4.509738748	T	C	TBC1 domain family member whacked	India Colony|India Wildtype

50202	61257491	-5.300927412	G	A	TBC1 domain family member whacked	India Colony|India Wildtype
50202	61257491	5.300927412	G	A	TBC1 domain family member whacked	India Colony|India Wildtype
50202	61259531	-4.517148456	T	G	TBC1 domain family member whacked	India Colony|India Wildtype
50202	61259531	4.517148456	T	G	TBC1 domain family member whacked	India Colony|India Wildtype
50202	68720664	-5.304395585	G	A	acetylcholine receptor subunit beta-like 2	Ethiopia|India Wildtype
50202	68720664	5.304395585	G	A	acetylcholine receptor subunit beta-like 2	Ethiopia|India Wildtype
50202	68720666	-5.304395585	A	G	acetylcholine receptor subunit beta-like 2	Ethiopia|India Wildtype
50202	68720666	5.304395585	A	G	acetylcholine receptor subunit beta-like 2	Ethiopia|India Wildtype
50202	76811040	-4.434506507	T	C	hemicentin-1	India Colony|Ethiopia
50202	76811040	-5.089076007	T	C	hemicentin-1	India Colony|India Wildtype
50202	76811040	4.434506507	T	C	hemicentin-1	India Colony|Ethiopia
50202	76811040	5.089076007	T	C	hemicentin-1	India Colony|India Wildtype
50202	76811043	-5.495219203	C	T	hemicentin-1	India Colony|Ethiopia
50202	76811043	-6.217604889	C	T	hemicentin-1	India Colony|India Wildtype
50202	76811043	5.495219203	C	T	hemicentin-1	India Colony|Ethiopia
50202	76811043	6.217604889	C	T	hemicentin-1	India Colony|India Wildtype
50202	76811044	-5.394962388	G	A	hemicentin-1	India Colony|Ethiopia
50202	76811044	-6.110186512	G	A	hemicentin-1	India Colony|India Wildtype
50202	76811044	5.394962388	G	A	hemicentin-1	India Colony|Ethiopia
50202	76811044	6.110186512	G	A	hemicentin-1	India Colony|India Wildtype
50202	76811048	-5.97789318	C	T	hemicentin-1	India Colony|India Wildtype
50202	76811048	5.97789318	C	T	hemicentin-1	India Colony|India Wildtype
50202	76811059	-4.8361411	A	G	hemicentin-1	India Colony|India Wildtype
50202	76811059	4.8361411	A	G	hemicentin-1	India Colony|India Wildtype
50202	76811061	-4.927425796	T	G	hemicentin-1	India Colony|India Wildtype
50202	76811061	4.927425796	T	G	hemicentin-1	India Colony|India Wildtype
50202	78746627	-4.445383926	G	T	putative transcription factor capicua	India Colony|India Wildtype

50202	78746627	4.445383926	G	T	putative transcription factor capicua	India Colony|India Wildtype
50202	78964825	-4.637697645	G	T	serine/threonine-protein phosphatase PP1-beta catalytic subuni	India Colony|Ethiopia
50202	78964825	4.637697645	G	T	serine/threonine-protein phosphatase PP1-beta catalytic subuni	India Colony|Ethiopia
50202	80145364	-4.551285507	G	T	Kv channel-interacting protein 1	Ethiopia|India Wildtype
50202	80145364	4.551285507	G	T	Kv channel-interacting protein 1	Ethiopia|India Wildtype
50202	80281624	-5.668252926	C	A	angiopoietin-related protein 1-like	India Colony|India Wildtype
50202	80281624	5.668252926	C	A	angiopoietin-related protein 1-like	India Colony|India Wildtype
50202	80281625	-5.755461491	G	A	angiopoietin-related protein 1-like	India Colony|India Wildtype
50202	80281625	5.755461491	G	A	angiopoietin-related protein 1-like	India Colony|India Wildtype
50202	80281630	-5.180678507	G	T	angiopoietin-related protein 1-like	India Colony|India Wildtype
50202	80281630	5.180678507	G	T	angiopoietin-related protein 1-like	India Colony|India Wildtype
50202	80295481	-4.573458762	C	T	connectin-like	India Colony|India Wildtype
50202	80295481	4.573458762	C	T	connectin-like	India Colony|India Wildtype
50202	80374058	-4.830799954	A	C	connectin-like	India Colony|India Wildtype
50202	80374058	4.830799954	A	C	connectin-like	India Colony|India Wildtype
50202	80374072	-4.462623292	A	T	connectin-like	India Colony|India Wildtype
50202	80374072	4.462623292	A	T	connectin-like	India Colony|India Wildtype
50202	80377015	-4.552850054	T	C	connectin-like	India Colony|India Wildtype
50202	80377015	4.552850054	T	C	connectin-like	India Colony|India Wildtype
50202	80377018	-4.55744321	T	G	connectin-like	India Colony|India Wildtype
50202	80377018	4.55744321	T	G	connectin-like	India Colony|India Wildtype
50202	80378945	-4.680565057	T	G	connectin-like	India Colony|India Wildtype
50202	80378945	4.680565057	T	G	connectin-like	India Colony|India Wildtype
50202	80378948	-4.612051821	A	T	connectin-like	India Colony|India Wildtype
50202	80378948	4.612051821	A	T	connectin-like	India Colony|India Wildtype

50202	80385412	-4.441177488	A	G	connectin-like	India Colony|India Wildtype
50202	80385412	4.441177488	A	G	connectin-like	India Colony|India Wildtype
50202	85249950	-4.441874871	A	T	CCA tRNA nucleotidyltransferase 1, mitochondrial	India Colony|Ethiopia
50202	85249950	4.441874871	A	T	CCA tRNA nucleotidyltransferase 1, mitochondrial	India Colony|Ethiopia
50202	86768223	-4.816762081	A	T	coiled-coil domain-containing protein lobo	India Colony|India Wildtype
50202	86768223	4.816762081	A	T	coiled-coil domain-containing protein lobo	India Colony|India Wildtype
50202	86768493	-4.595960863	A	G	coiled-coil domain-containing protein lobo	India Colony|India Wildtype
50202	86768493	4.595960863	A	G	coiled-coil domain-containing protein lobo	India Colony|India Wildtype
50202	86771047	-4.601367228	A	G	coiled-coil domain-containing protein lobo	India Colony|India Wildtype
50202	86771047	4.601367228	A	G	coiled-coil domain-containing protein lobo	India Colony|India Wildtype
50202	86771072	-4.674419435	T	A	coiled-coil domain-containing protein lobo	India Colony|India Wildtype
50202	86771072	4.674419435	T	A	coiled-coil domain-containing protein lobo	India Colony|India Wildtype
50202	86771101	-4.556824774	T	C	coiled-coil domain-containing protein lobo	India Colony|India Wildtype
50202	86771101	4.556824774	T	C	coiled-coil domain-containing protein lobo	India Colony|India Wildtype
50202	86771140	-4.560784184	C	A	coiled-coil domain-containing protein lobo	India Colony|India Wildtype
50202	86771140	4.560784184	C	A	coiled-coil domain-containing protein lobo	India Colony|India Wildtype
50202	90739286	-4.545661488	A	T	E3 ubiquitin-protein ligase TRIM33	India Colony|India Wildtype
50202	90739286	4.545661488	A	T	E3 ubiquitin-protein ligase TRIM33	India Colony|India Wildtype
50203	53153112	-4.817793589	C	A	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase classes I and II	India Colony|Ethiopia
50203	53153113	-4.542803026	T	C	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase classes I and II	India Colony|Ethiopia
50203	53153112	4.817793589	C	A	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase classes I and II	India Colony|Ethiopia
50203	53153113	4.542803026	T	C	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase classes I and II	India Colony|Ethiopia
50203	31749273	-4.569502324	C	T	potassium voltage-gated channel subfamily KQT member 1-like	India Colony|India Wildtype

50203	31749273	4.569502324	C	T	potassium voltage-gated channel subfamily KQT member 1-like	India Colony|India Wildtype
50203	31873994	-4.577619538	A	G	potassium voltage-gated channel subfamily KQT member 1-like	India Colony|India Wildtype
50203	31874076	-4.909721832	G	A	potassium voltage-gated channel subfamily KQT member 1-like	India Colony|India Wildtype
50203	31890671	-4.710033064	G	A	potassium voltage-gated channel subfamily KQT member 1-like	India Colony|India Wildtype
50203	31873994	4.577619538	A	G	potassium voltage-gated channel subfamily KQT member 1-like	India Colony|India Wildtype
50203	31874076	4.909721832	G	A	potassium voltage-gated channel subfamily KQT member 1-like	India Colony|India Wildtype
50203	31890671	4.710033064	G	A	potassium voltage-gated channel subfamily KQT member 1-like	India Colony|India Wildtype
50203	36632733	-4.451344621	A	G	hillarin	India Colony|India Wildtype
50203	36632739	-4.427059512	T	C	hillarin	India Colony|India Wildtype
50203	36632788	-4.709684319	G	A	hillarin	India Colony|India Wildtype
50203	36632821	-4.433101043	G	A	hillarin	India Colony|India Wildtype
50203	36632823	-4.531065587	A	G	hillarin	India Colony|India Wildtype
50203	36632827	-4.446273504	G	A	hillarin	India Colony|India Wildtype
50203	36632733	4.451344621	A	G	hillarin	India Colony|India Wildtype
50203	36632739	4.427059512	T	C	hillarin	India Colony|India Wildtype
50203	36632788	4.709684319	G	A	hillarin	India Colony|India Wildtype
50203	36632821	4.433101043	G	A	hillarin	India Colony|India Wildtype
50203	36632823	4.531065587	A	G	hillarin	India Colony|India Wildtype
50203	36632827	4.446273504	G	A	hillarin	India Colony|India Wildtype
50203	37105747	-4.463267349	A	G	ankyrin repeat domain-containing protein SOWAHB	India Colony|Ethiopia
50203	37127258	-4.53341508	G	A	ankyrin repeat domain-containing protein SOWAHB	India Colony|India Wildtype

50203	37105747	4.463267349	A	G	ankyrin repeat domain-containing protein SOWAHB	India Colony|Ethiopia
50203	37127258	4.53341508	G	A	ankyrin repeat domain-containing protein SOWAHB	India Colony|India Wildtype
50203	36445500	-4.575575657	A	T	uncharacterized LOC118509705 (	India Colony|Ethiopia
50203	36445500	4.575575657	A	T	uncharacterized LOC118509705 (	India Colony|Ethiopia
50203	28008767	-5.283229508	C	T	uncharacterized LOC118510096	India Colony|India Wildtype
50203	28008767	5.283229508	C	T	uncharacterized LOC118510096	India Colony|India Wildtype
50203	5005839	-4.459142803	A	C	DNA N6-methyl adenine demethylase	India Colony|Ethiopia
50203	5005845	-4.460652716	G	T	DNA N6-methyl adenine demethylase	India Colony|Ethiopia
50203	5005839	4.459142803	A	C	DNA N6-methyl adenine demethylase	India Colony|Ethiopia
50203	5005845	4.460652716	G	T	DNA N6-methyl adenine demethylase	India Colony|Ethiopia
50203	7684853	-4.521264918	T	A	neogenin	India Colony|India Wildtype
50203	7684853	4.521264918	T	A	neogenin	India Colony|India Wildtype
50203	7808325	-4.695531329	G	A	uncharacterized LOC118510486	India Colony|India Wildtype
50203	7808325	4.695531329	G	A	uncharacterized LOC118510486	India Colony|India Wildtype
50203	7822198	-4.437732051	C	T	putative mediator of RNA polymerase II transcription subunit 26	India Colony|Ethiopia
50203	7822268	-4.681329214	G	C	putative mediator of RNA polymerase II transcription subunit 26	India Colony|Ethiopia
50203	7822328	-4.458668577	C	A	putative mediator of RNA polymerase II transcription subunit 26	India Colony|Ethiopia
50203	7822268	-4.852902721	G	C	putative mediator of RNA polymerase II transcription subunit 26	India Colony|India Wildtype
50203	7822328	-4.613152098	C	A	putative mediator of RNA polymerase II transcription subunit 26	India Colony|India Wildtype
50203	7832999	-4.542523371	C	A	putative mediator of RNA polymerase II transcription subunit 26	India Colony|India Wildtype
50203	7822198	4.437732051	C	T	putative mediator of RNA polymerase II transcription subunit 26	India Colony|Ethiopia

50203	7822268	4.681329214	G	C	putative mediator of RNA polymerase II transcription subunit 26	India Colony|Ethiopia
50203	7822328	4.458668577	C	A	putative mediator of RNA polymerase II transcription subunit 26	India Colony|Ethiopia
50203	7822268	4.852902721	G	C	putative mediator of RNA polymerase II transcription subunit 26	India Colony|India Wildtype
50203	7822328	4.613152098	C	A	putative mediator of RNA polymerase II transcription subunit 26	India Colony|India Wildtype
50203	7832999	4.542523371	C	A	putative mediator of RNA polymerase II transcription subunit 26	India Colony|India Wildtype
50203	7853974	-4.435484555	T	A	ATPase family AAA domain-containing protein 5	India Colony|India Wildtype
50203	7853974	4.435484555	T	A	ATPase family AAA domain-containing protein 5	India Colony|India Wildtype
50203	7858613	-4.653240286	G	A	cytochrome b5-like	India Colony|India Wildtype
50203	7858613	4.653240286	G	A	cytochrome b5-like	India Colony|India Wildtype
50203	8485166	-5.763886396	T	G	glutamate receptor 1-like	India Colony|India Wildtype
50203	8485168	-5.90966027	A	G	glutamate receptor 1-like	India Colony|India Wildtype
50203	8485172	-5.412882141	T	G	glutamate receptor 1-like	India Colony|India Wildtype
50203	8485166	5.763886396	T	G	glutamate receptor 1-like	India Colony|India Wildtype
50203	8485168	5.90966027	A	G	glutamate receptor 1-like	India Colony|India Wildtype
50203	8485171	5.810627055	G	C	glutamate receptor 1-like	India Colony|India Wildtype
50203	8485172	5.412882141	T	G	glutamate receptor 1-like	India Colony|India Wildtype
50203	77326077	-4.419417499	C	A	P protein	India Colony|India Wildtype
50203	77326077	4.419417499	C	A	P protein	India Colony|India Wildtype
50203	73673505	-4.535296303	A	G	apoptosis-resistant E3 ubiquitin protein ligase 1	India Colony|India Wildtype
50203	73673505	4.535296303	A	G	apoptosis-resistant E3 ubiquitin protein ligase 1	India Colony|India Wildtype
50203	15276293	-4.680159147	T	G	trithorax group protein osa	India Colony|India Wildtype
50203	15276293	4.680159147	T	G	trithorax group protein osa	India Colony|India Wildtype
50203	2895486	-4.590372242	C	T	uncharacterized LOC118511244	India Colony|Ethiopia

50203	2895486	4.590372242	C	T	uncharacterized LOC118511244	India Colony|Ethiopia
50203	46918232	-4.530998326	T	G	serine-rich adhesin for platelets-like	India Colony|India Wildtype
50203	46918232	4.530998326	T	G	serine-rich adhesin for platelets-like	India Colony|India Wildtype
50203	27625615	-4.475598481	G	A	dead ringer-like	India Colony|India Wildtype
50203	27625615	4.475598481	G	A	dead ringer-like	India Colony|India Wildtype
50203	65163599	-4.692754597	C	A	paxillin	India Colony|Ethiopia
50203	65163599	4.692754597	C	A	paxillin	India Colony|Ethiopia
50203	40136361	-4.903095958	A	G	myb-like protein Q	India Colony|India Wildtype
50203	40136361	4.903095958	A	G	myb-like protein Q	India Colony|India Wildtype
50203	40679757	-4.444395417	C	T	NACHT domain- and WD repeat-containing protein 1	India Colony|Ethiopia
50203	40679757	4.444395417	C	T	NACHT domain- and WD repeat-containing protein 1	India Colony|Ethiopia
50203	1538814	4.710636611	C	T	lysophospholipid acyltransferase 1	Ethiopia|India Wildtype
50203	1538814	-4.710636611	C	T	lysophospholipid acyltransferase 1	Ethiopia|India Wildtype
50203	1445974	-4.804446916	A	C	uncharacterized LOC118512080	India Colony|India Wildtype
50203	1445974	4.804446916	A	C	uncharacterized LOC118512080	India Colony|India Wildtype
50203	69463786	-4.521174362	C	T	T-box protein H15-like	India Colony|India Wildtype
50203	69463786	4.521174362	C	T	T-box protein H15-like	India Colony|India Wildtype
50203	15561768	-4.709458944	A	G	platelet binding protein GspB	India Colony|Pakistan Colony
50203	15561771	-4.610603285	A	G	platelet binding protein GspB	India Colony|Pakistan Colony
50203	15561772	-4.565390713	G	A	platelet binding protein GspB	India Colony|Pakistan Colony
50203	15561775	-4.422688562	C	G	platelet binding protein GspB	India Colony|Pakistan Colony
50203	15561776	-4.556652629	A	T	platelet binding protein GspB	India Colony|Pakistan Colony
50203	15561768	-4.478021367	A	G	platelet binding protein GspB	India Colony|India Wildtype
50203	15561768	4.709458944	A	G	platelet binding protein GspB	India Colony|Pakistan Colony
50203	15561771	4.610603285	A	G	platelet binding protein GspB	India Colony|Pakistan Colony

50203	15561772	4.565390713	G	A	platelet binding protein GspB	India Colony|Pakistan Colony
50203	15561775	4.422688562	C	G	platelet binding protein GspB	India Colony|Pakistan Colony
50203	15561776	4.556652629	A	T	platelet binding protein GspB	India Colony|Pakistan Colony
50203	15561768	4.478021367	A	G	platelet binding protein GspB	India Colony|India Wildtype
50203	54844619	-4.474442724	A	G	protein outspread	India Colony|Ethiopia
50203	54844623	-4.608394092	C	T	protein outspread	India Colony|Ethiopia
50203	54844631	-4.634456523	T	C	protein outspread	India Colony|Ethiopia
50203	54844619	4.474442724	A	G	protein outspread	India Colony|Ethiopia
50203	54844623	4.608394092	C	T	protein outspread	India Colony|Ethiopia
50203	54844631	4.634456523	T	C	protein outspread	India Colony|Ethiopia
50203	71907635	-4.421811549	T	G	signal-induced proliferation-associated 1-like protein 2	Ethiopia|India Wildtype
50203	71907635	4.421811549	T	G	signal-induced proliferation-associated 1-like protein 2	Ethiopia|India Wildtype
50203	17396549	-4.467277364	G	C	AF4/FMR2 family member lilli-like	India Colony|Ethiopia
50203	17396549	4.467277364	G	C	AF4/FMR2 family member lilli-like	India Colony|Ethiopia
50203	86803105	-4.443099736	T	G	F-box/LRR-repeat protein 20	India Colony|India Wildtype
50203	86803105	4.443099736	T	G	F-box/LRR-repeat protein 20	India Colony|India Wildtype
50203	12067769	-4.689398922	C	T	polypeptide N-acetylgalactosaminyltransferase	India Colony|India Wildtype
50203	12067783	-4.486559527	T	A	polypeptide N-acetylgalactosaminyltransferase	India Colony|India Wildtype
50203	12067769	4.689398922	C	T	polypeptide N-acetylgalactosaminyltransferase	India Colony|India Wildtype
50203	12067783	4.486559527	T	A	polypeptide N-acetylgalactosaminyltransferase	India Colony|India Wildtype
50203	12052275	-4.529419893	A	G	uncharacterized LOC118514496	India Colony|India Wildtype
50203	12052275	4.529419893	A	G	uncharacterized LOC118514496	India Colony|India Wildtype
50203	67833037	-4.558188554	G	T	atrial natriuretic peptide receptor 1	India Colony|India Wildtype
50203	67833043	-4.688324454	T	C	atrial natriuretic peptide receptor 1	India Colony|India Wildtype
50203	67833116	-4.997379153	G	T	atrial natriuretic peptide receptor 1	India Colony|India Wildtype

50203	67833037	4.558188554	G	T	atrial natriuretic peptide receptor 1	India Colony|India Wildtype
50203	67833043	4.688324454	T	C	atrial natriuretic peptide receptor 1	India Colony|India Wildtype
50203	67833116	4.997379153	G	T	atrial natriuretic peptide receptor 1	India Colony|India Wildtype
50203	50219557	-4.554666695	G	C	lipopolysaccharide-induced tumor necrosis factor-alpha factor-like	India Colony|India Wildtype
50203	50219557	4.554666695	G	C	lipopolysaccharide-induced tumor necrosis factor-alpha factor-like	India Colony|India Wildtype

1

Chapter 4: Application of a targeted amplicon sequencing panel to screen for insecticide resistance mutations in Anopheles darlingi populations from Brazil

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	lsh1704757	**Title**	Miss
First Name(s)	Holly		
Surname/Family Name	Acford-Palmer		
Thesis Title	A genomic investigation into three malaria vectors: An. darlingi, An. funestus, and An. stephensi		
Primary Supervisor	Prof. Susana Campino		

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?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

*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?	Scientfic Reports
Please list the paper's authors in the intended authorship order:	Holly Acford-Palmer, Alice O. Andrade, Jody E. Phelan, Rosa A. Santana, Stefanie C.P. Lopes, Jansen F. Medeiros, Taane G. Clark, Maisa S. Araujo, 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 designed the amplicon panel, tested the primers, and applied the primer panel to the isolates. I then performed the bioinformatic analysis, interpreted the results under the supervision from my supervisors. I wrote the first draft of the paper, which was commented upon by my supervisors and co-authors.

SECTION E

Student Signature	
Date	

Supervisor Signature	
Date	

Application of a targeted amplicon sequencing panel to screen for insecticide resistance mutations in *Anopheles darlingi* populations from Brazil

Short title: Amplicon sequencing to identify insecticide resistance mutations in *Anopheles darlingi*

Holly Acford-Palmer[1], Alice O. Andrade[2], Jody E. Phelan[1], Rosa A Santana[3], Stefanie C.P. Lopes[3,4], Jansen F. Medeiros[2,6], Taane G. Clark [1,5], Maisa S. Araujo[2,7,8 *], Susana Campino[1, *]

[*] joint last authors

[1] Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom

[2] Plataforma de Produção e Infecção de Vetores da Malária- PIVEM, Laboratório de Entomologia, Fiocruz Rondonia, Porto Velho, RO, Brazil

[3] Instituto de Pesquisa Clínica Carlos Borborema, Fundação de Medicina Tropical Dr. Heitor Vieira Dourado, Manaus, AM, Brazil

[4] Instituto Leônidas & Maria Deane, FIOCRUZ, Manaus, AM, Brazil

[5] Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom

[6] Programa de Pós-Graduação em Biologia Experimental – PGBIOEXP, Fundação Universidade Federal de Rondonia, Porto Velho, RO, Brazil

[7] Programa de Pós-Graduação em Conservação e uso de Recursos Naturais – PPGReN, Fundação Universida-de Federal de Rondonia, Porto Velho, RO, Brazil

[8] Laboratório de Pesquisa Translacional e Clínica, Centro de Pesquisa em Medicina Tropical, Porto Velho, RO, Brazil

[**] Corresponding Authors:

Susana Campino, LSHTM, Susana.campino@lshtm.ac.uk

Maisa S. Araujo, PIVEM, maisa.araujo @fiocruz.br

Abstract

Large-scale surveillance and informed vector control approaches are urgently needed to ensure that national malaria programs remain effective in reducing transmission and, ultimately, achieving malaria elimination targets. In South and Central America, *Anopheles darlingi* is the primary malaria vector, responsible for the majority of *Plasmodium* species transmission. However, little is known about their molecular markers associated with insecticide resistance. Here we developed a low-cost, high throughput amplicon sequencing ("amp-seq") panel, consisting of 11 amplicons that target genes linked to mosquito species (*cox-1* and *its2*) and insecticide resistance (*ace-1*, *GSTe2*, *vgsc* and *rdl*). Used in tandem with dual index barcoding of amplicons, our approach permits high numbers of loci and samples to be sequenced in single runs, thereby decreasing costs, and increasing efficiency. By screening 200 *An. darlingi* mosquitoes collected in Brazil, our amp-seq approach identified 10-point mutations leading to amino acid alterations in *ace-1* (V243I, N194H, S673N, S674N/T) and *GSTe2* genes (I114V, D128E, T166I, T179I, and T205A). Overall, our work has demonstrated the utility of amp-seq to provide insights into the genetic diversity of *An*. *darlingi* mosquitoes. The amp-seq approach can be applied as a wide-scale insecticide-resistance surveillance technique to better inform vector-control methods.

Word count: 194

Background

Malaria, caused by *Plasmodium* parasites and transmitted by *Anopheles* mosquitoes, is one of the most severe public health problems in the Americas, where ~597,000 cases were diagnosed in 2021 alone[1]. Approximately 1 in 5 of those malaria cases occurred in Brazil[1], where the number of cases rose by 3% between 2015 and 2020 (137,982 vs. 142,112)[2]. The transmission of *Plasmodium* spp. in Brazil predominantly occurs in the Amazon region[2,3]. *Plasmodium vivax* parasites cause the majority (~83%) of malaria cases, followed by *P. falciparum* (~17%), and then other *Plasmodium* spp. (0.1%). Over the last 7 years, the proportion of cases caused by *P. falciparum*, the species that causes most severe disease, has increased by 7%[2].

The primary malaria vector in Brazil, *Anopheles darlingi* (also called *Nyssorhynchus darlingi*), is highly susceptible to *Plasmodium* infection and can maintain malaria transmission even when parasites are at low densities[4–9]. This mosquito species is highly adaptable to recently anthropized environments and exhibits both exophagic and endophagic behaviour. *An. darlingi* primarily feeds on humans and can occupy ecological niches left empty by other *Anopheles* spp[8,10]. Since 2017, deforestation has increased in the Brazilian Amazon, with previous studies showing these newly deforested areas or "frontiers" have higher *An. darlingi* abundance and increased malaria transmission[11]. The adaptability of *An. darlingi* mosquitoes is believed to be leading to year-round malaria transmission. Previously, forest populations of *An. darlingi* would peak during, and towards the end of the rainy season due to the increased availability of larval habitats in flooded areas near rivers. However, environmental changes produced by humans have created permanently available larval habitats, thereby supporting perennial malaria transmission[12,13]. These challenges represent a clear risk to Brazil to accomplish its National Elimination plan and achieve the World Health Organization (WHO) goal of reducing malaria cases by 90% within the next 7 years [14].

Typical malaria control practice in Brazil uses insecticides, especially indoor residual spraying (IRS) or long lasting insecticide treated nets (LLINs)[15,16]. Pyrethroids have been the insecticide class applied in recent years, but organophosphates (malathion) and carbamates are used on *Aedes aegypti* as part of arbovirus control. Since the reintroduction of dengue in Brazil and the occurrence of outbreaks of other arboviruses (e.g., Zika and chikungunya), *Ae. aegypti* has been part of a national insecticide resistance screening programme to optimise elimination strategies, but no such programme exists for *Anopheles* spp[17,18]. Resistance to pyrethroids has been reported across *Anopheles* spp. globally, and for *An. darlingi* in countries surrounding Brazil, including Bolivia, Peru, Colombia, and French Guiana[19–21]. Resistance to carbamates has also been reported in Peru and Bolivia, and organochloride resistance has been observed in Peru and Colombia[22,23]. No resistance has been reported for organophosphates in Brazil, and no data is available for pyrroles or neonicotinoids. The lack of reporting on insecticide resistance in *An. darlingi* is a cause for concern, due to the high levels of associated *Plasmodium* spp. transmission by this vector.

The main insecticide resistance mechanisms observed by mosquito species are target site, metabolic and cuticular, and behavioural avoidance[24]. Target site resistance is mediated by mutations in insecticide target genes, such as the acetylcholinesterase-1 (*ace-1*), γ-aminobutyric acid (GABA) receptor (*rdl*), and voltage-gated sodium channel (*vgsc*), as well as mutations in glutathione-s-transferase *epsilon* (*GSTe2),* which encodes an insecticide metabolising enzyme. Particularly well-studied are the knockdown resistance (*kdr*) mutations associated with Dichlorodiphenyltrichloroethane (DDT) and pyrethroid-based insecticides, including the L1014 mutation in *vgsc* [25,26]. Point mutations in the *GSTe2* gene are also associated with these insecticides, and the L119F and I114T mutations lead to resistance to permethrin in *An. funestus*[27,28]. Mutations and duplications in the *ace-1* gene result in organophosphate and carbamate resistance, and amino acid alterations in *rdl* have been associated with resistance to organochlorines, particularly dieldrin [29–31]. Molecular surveillance studies of insecticide resistance mutations in *An. darlingi* are scarce, with

only a few investigations exploring *vgsc* and *ace-1* genes[32,33]. None of the single nucleotide polymorphisms (SNPs) previously found to be associated with insecticide resistance in other *Anopheles* species have been observed in *An. darlingi*, including in pyrethroid resistant populations[33]. However, only limited regions of these two genes have been analysed. It is possible that genetic variants in other gene regions are present, or other mechanisms are involved.

Whole genome sequencing (WGS) has been applied to many vectors to better understand their genomic landscapes and identify candidate genes to unravel mechanisms of insecticide resistance. However, *Anopheles* genomes are large (~300 Mbp), and WGS is an expensive method that requires high quantities of DNA to gain suitable genomic insights, meaning it is inappropriate as a high-throughput surveillance method. The application of next-generation sequencing to targeted PCR amplicons, in tandem with dual-index barcoding, has been successfully used in other *Anopheles* spp*.,* *Aedes* spp*.*, and *Plasmodium* spp*.,* as a high-throughput and low-cost screening method for insecticide or drug resistance mutations in target *loci* [34–37]. Targeting several candidate genes in many samples permits the tracking of emerging resistance and spread of known mutations in the population. This approach also allows for an analysis of genotype-phenotype associations to identify novel mutations linked to insecticide phenotypic assays.

Here we have designed an amplicon-sequencing ("amp-seq") assay, consisting of a panel of 11 amplicons (each ~500bp) covering the 4 genes (*vgsc, ace-1, rdl*, and *GSTe2*) commonly associated with insecticide resistance, and a further 2 genes (*Its2* and *cox1*) used for species identification and phylogenetic analysis. The assay was used to screen 200 *An. darlingi* mosquitoes collected in Brazil, revealing new mutations. Our assay represents a cost-effective method to confirm mosquito species and conduct insecticide resistance surveillance, with the potential to inform control strategies for an understudied vector responsible for high levels of malaria transmission in South America.

Methods

Sample Collection, Species Identification and DNA extraction

An. darlingi field mosquitoes were mostly collected from localities in Rondônia state (n= 171; Candeias do Jamari 116, Porto Velho 55) but also include four samples from the Amazonas state (n = 4; Manaus 1, Manacapuru 3) collected to established a *An. darlingi* colony[38,39] . The collections were performed during studies of vector density in malaria endemic areas of Rondônia in 2005-2006 and 2018-2019[38,39]. Mosquitoes from the *An. darlingi* colony of Porto Velho/Rondônia (colony generations : F2-F4, F9-F11, F21, F33-F35, F39, F40-F42) were also included in the study (n=25)[40]. In total, 200 mosquitoes were screened. The field samples were selected based on whether the localities had used insecticides (Porto Velho) or had little/no recent insecticide usage (Candeias do Jamari). The specimens were initially identified by stereoscopic microscopy, using the established dichotomous keys [41]. Genomic DNA was extracted from whole mosquitoes using the Qiagen DNeasy ® Tissue and Blood kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions. Briefly, each mosquito was grounded with 30 µL and the DNA extract was resuspended in 50 µL elution buffer.

Primer Design

Amplicon primers were designed using Primer3 software, against sequences downloaded from VectorBase[42]. The primers aimed to amplify an approximate 500bp region containing SNPs previously described as associated with insecticide resistance in *Anopheles* or *Aedes* mosquito's species. This resulted in a panel of 9 primers targeting 4 genes: (i) *vgsc* (4 amplicons, targeting four domains); (ii) *rdl* (2 amplicons, targeting 3 SNPs); (iii) *ace-1* (2 amplicons, targeting two SNPs, including A280S[32]); and (iv) *GSTe2* (1 amplicon, targeting a single SNP). When possible, these primers were designed to bind to exonic regions. Two other amplicons were designed to target genes commonly used for species identification and phylogenetic investigation: the ribosomally encoded gene internal transcribed spacer 2 (*its2*); and cytochrome c oxidase I (*cox1*), a locus found in the mitochondria. This resulted in a final panel of 11 assays, covering 6 genes (**Table 1**). Each primer sequence was concatenated with a

unique 5’ barcode (8bp), to enable sample pooling during sequencing. Samples were assigned a unique forward and reverse barcode combination used for the generation of each amplicon. To identify amplicons suitable for multiplexing the ThermoFisher Scientific Multiple Primer Analyser was used with sensitivity for dimmer detection set to one.

Amplicon Generation

Multiplex Polymerase chain reactions (PCR) were performed using NEB Q5 hot start polymerase (New England BioLabs, UK) with a total volume of 25ul per reactions. Sample volume of 1µL (~2ng/µL) was used, with an average final primer concentration of 0.5µm in each PCR. The amplification was conducted as follows: hot-start polymerase activation for 3 minutes at 95^{o}C, followed by 30 cycles of 95^{o}C for 10 seconds, 58^{o}C for 30 seconds and 72^{o}C for 45 seconds, followed by a final elongation step of 72^{o}C for 5 minutes. Post-multiplex PCR reaction, amplicons were visualised on a 1% agarose gel to confirm amplification, alongside band size and intensity. The multiplex PCR amplicons consisting of 11 amplicons were first pooled by sample, and subsequently pooled with other samples that had different 5’ barcode tags. Sample pools were purified using Roche Kapa beads following manufacturer’s instructions. A bead to sample ratio of 0.7:1 was used to remove excess primers and PCR reagents. The Qubit 2.0 fluorimeter HS DNA kit was used to quantify the pool concentration. Illumina adaptors and barcodes were added as part of the indexing PCR to the sample pool as a part of the Illumina-based Amplicon-EZ service (Genewiz, UK). Pools contained a maximum of 200 amplicons to maximise coverage. Each indexed pool was sequenced using a 2 x 250bp (paired-end) configuration on an Illumina MiSeq. A minimum of 50,000 paired-end reads were attained per pool, which equates to at least 450 reads per amplicon in a pool of 110 amplicons, at a cost of <US$0.5 per amplicon.

178 **Amplicon Analysis**

179 Raw fastq files were de-multiplexed using the unique barcode combination assigned to each sample,
180 using an in-house python script (https://github.com/LSHTMPathogenSeqLab/amplicon-seq). The
181 resulting sample fastq files were then analysed using another in-house pipeline, where files are first
182 trimmed using Trimmomatic software, then mapped to the reference sequence idAnoDarlMG_H_01
183 (from NCBI) using the bwa-mem package, and reads are then filtered using Samclip software[43–45].
184 GATK HaplotypeCaller (v4.1.4.1, default parameters) and Freebayes (v1.3.5, --haplotype-length -1)
185 software were used to call variants [46,47]. The SNPs and small insertions/deletions (INDELs) identified
186 were then filtered using bcftools[48]. To pass quality control checks, a minimum depth of 30 reads, phred
187 score of > 30 per base, and a minimum allele depth of 10 was required. Variants had to be present in
188 >1 sample, and across >1 of the sample pools sequenced. The SnpEff tool was applied to annotate
189 variants using a database built from the idAnoDarlMG_H_01 reference genome[49]. Variants were then
190 genotyped based on the proportion of alternative allele to total depth coverage, called as homozygous
191 reference (<20% alternate allele reads), heterozygous (20-80% alternate allele), or homozygous
192 alternate (>80% alternate allele reads)[34].

193

194 **Phylogenetic Analysis**

195 For the *Its2* and *cox1* amplicons, SNP calls with >50-fold read depth were converted to fasta files using
196 an in-house pipeline (https://github.com/LSHTMPathogenSeqLab/fastq2matrix). Only sequences with
197 SNP calls that reached this depth were included in phylogenetic analysis. For each gene, sequence
198 data was aligned using the MAFFT tool. Sequences from the NCBI other countries were included in the
199 resulting alignments[50]. For *cox-1*, 62 sequences were added from Brazil, Honduras, Belize, Colombia,
200 Panama, Ecuador, and Peru. For *its2*, an additional 27 sequences were aligned from Brazil, Colombia,
201 Belize, and Bolivia. For tree generation, sequences from both genes were concatenated, and the
202 resulting alignments were viewed and trimmed using Aliview[51]. Phylogenetic trees were constructed
203 using RAxML software[52]. The trees were built using a maximum-likelihood method, with the

GTRGAMMA option. This approach assumes a GTR model of nucleotide substitution, and a gamma model of rate heterogeneity. A bootstrap value of 1000 was used for tree construction, and the resulting tree was visualised using iTOL software[53].

Haplotype networks and Maps

To construct the haplotype network, fasta sequences for each gene were aligned for all samples, and then the Pegas package in R[54] was applied. The same package was used to calculate nucleotide diversity, haplotype diversity, Tajima's D statistic, fixation indel (Fst), heterozygosity and linkage disequilibrium.

Results

Detection of novel SNPs in genes associated with insecticide resistance

A total of 200 *An. darlingi* samples were sequenced, with the resulting average amplicon coverage ranging from 171.17- to 5621.32-fold (**Table 2**). From the alignments, 246 SNPs and 20 INDELs passed all quality control measures, the majority of which were either synonymous (37.8%) or intronic variants (45.9%). Ten SNPs and one INDEL were annotated as non-synonymous and resulted in an amino acid change (**Table 3**). Missense SNPs were only found in *ace-1* and *GSTe2* genes, and all were present in at least two samples and in two or more populations **(Table 3**). These SNPs have not been previously reported. In the *ace-1* gene, 111 SNPs were found across the two amplicons, including five missense SNPs (V243I, N194H, S673N, S674N, and S674T), with S674N occurring at the highest frequency (22.3%). Three of these five mutations (V243I, N194H and S674T) occurred only in field populations, while S673N and S674N appeared in field populations and colony samples (**Table 4).** A further five non-synonymous SNPs were found in the *GSTe2* gene. Three of the missense SNPs (I114V, T166I, and T179I) occurred at frequencies below 23%, and two others (D128E and T205A) appear to be at or approaching fixation as no samples were genotyped as homozygous reference, and over 90% were homozygous alternate for both amino acid alterations. Two of these amino acid alterations in

the *GSTe2* gene were found only in field populations (I114V and T166I), and the remaining three (D128E, T191I, and T205A) were also observed in colony samples.

For v*gsc*, the analysis revealed a INDEL caused by a 3bp deletion, resulting in an isoleucine deletion at position 422 in the first domain. This mutation has not previously been reported and occurred at a frequency of 25% across field and colony samples. Sanger sequencing confirmed these mutations, ruling out sequencing artifacts. No missense polymorphisms were detected in the *gaba* gene (*rdl*).

Genetic diversity of *An. darlingi* populations in Brazil

Sequences for mitochondrial *cox-1* gene and ribosomal *its2* were generated for genetic diversity analysis. In the *its2* gene, six SNPs were identified. Originally a total of 91 SNPs were identified in the *cox-1* amplicon, but upon further inspection, 71 of these SNPs were present in one sample (AnDar600), which was subsequently identified as an *An. peryassui* isolate (Blast score: 99.2% identity) and excluded from further analysis. Twenty SNPs were identified in the *cox-1* gene, 19 of which appeared in the Candeias do Jamari population, 13 in colony samples, 10 in the Porto Velho populations, and six in the State of Amazonas samples. A fixation index analysis for each SNP revealed no significant population differentiation across these genes (Fst < 0.032).

Phylogenetic analysis with the *its2* gene reflected the small number of SNPs shown in these populations, and very little differentiation was observed between Brazilian isolates and those from other countries (**Figure 1**). The tree separated into two main clades, the first of which contains three publicly available samples from Brazil, Belize, and Bolivia. The second clade contained several subclades that included all sequences generated in this study (n=198) along with the remaining publicly available Brazilian and Colombian samples (n=26). No differentiation between the different Brazilian populations was observed, which was supported by low nucleotide diversity (π=0.00536) (**Table 6**).

For the *cox-1* gene, a total of 128 sequences from this study alongside 67 publicly available samples were analysed**.** (**Figure 2)**. Using this gene, it was possible to see clusters separating most Colombian, Honduran and Brazilian samples. Although, sequences from other countries can be seen interspersed particularly within the Brazilian *cox-1* sequences. There was both low genetic and nucleotide diversity (0.00697) in *Cox-1* (**Table 5**), similar to the *its2* results.

A phylogenetic tree based on concatenated *its2* and *cox-1* sequences revealed little differentiation between samples across the geographical regions (**Supplementary figure 1**). Both loci demonstrated high haplotype diversity (**Table 5, Table 6**), with 44 haplotypes identified for the c*ox-1* gene, and 31 for *its2*. Most of the haplotypes, 77.3% for *cox-1* (n=34), and 54.8% for *Its2* (n=17), were present in only one sample (singletons). The high number of singleton haplotypes reflects the high proportion of SNPs occurring at low frequency in the populations. Haplotype 34, representing the *cox-1* gene, was the most frequent, and present in samples from the colony, and states of Amazonas and Rondônia (**Supplementary Figure 2)**. A higher number of *its2* haplotypes (n=14, 45.2%) were present in more than one sample, compared to *cox1* (n=10, 22.7%). Haplotypes 1 and 9 were the most frequent, present in both colony and state of Rondônia populations (**Supplementary Figure 3**). The samples from the state of Amazonas shared haplotypes with both colony and state of Rondônia samples. The networks revealed shared haplotypes for both genes across the three populations and included several other samples available from other Brazilian states.

Discussion

The application of our amp-seq panel to Brazilian field and colony *An. darlingi* samples has demonstrated its potential utility for species identification, and the discovery of SNPs in genes associated with insecticide resistance. Whilst no previously reported SNPs associated with insecticide resistance in other *Anopheles* species were found in this study, ten other non-synonymous SNPs were detected. Of the five SNPs found in the *GSTe2* gene, all except one (D128E) are either in amino acid

positions that are highly variable across *Anopheles* spp. or the mutation results in a change to an amino acid that is present in the reference of another *Anopheles* spp. The I114V substitution is in the same location as a previously reported amino acid alteration in *An. gambiae*[56]. In *An. gambiae,* the mutation results in a I114T alteration, which is hypothesised to cause resistance through the introduction of hydroxyl (-OH) group on the substituted threonine. The hydroxyl group decreases product affinity in the hydrophobic DDT binding site, thereby increasing metabolic turnover of the insecticide. In this instance the valine substitution does not introduce this same hydroxyl group, and valine is present as a reference amino acid in *An. atroparvus.* The D128E mutation occurs at a highly conserved site across *Anopheles* spp., with aspartic acid (D) present as the reference for all. The alteration to glutamic acid results in a similar amino acid structure with the addition of an extra carbon. This mutation appears to be near fixation as 92% of samples were genotyped as homozygous alternate, and the remaining 8% as heterozygous.

Pyrethroids are the predominant insecticide class being applied for malaria-focussed vector control in Brazil[16]. This implies that if target site alterations were to arise, they would predominantly occur in *vgsc* and *GSTe2* genes, as these are the proteins pyrethroids interact with. However, in Brazilian dengue control programmes, *Aedes* spp. are targeted with both pyrethroids and organophosphates, and this usage could impact *Anopheles* spp also present. This could explain the three non-synonymous SNPs found in the *ace-1* gene, a target for organophosphates. The previous reported target-site mutations G119S and N485I (positions in *Torpedo califonica,* G305S and S642I) alter susceptibility to the organophosphate and carbamate classes of insecticide[29]. Resistance to carbamates has been reported in Bolivia, close to the Brazilian border[22]. Of the five amino acid alterations detected here in *Ace-1* (V243I, N294H, S673N, and S674N/T), none have been previously reported. The S673N and S674N/T *ace-1* mutations occur near one of the three catalytic sites (H440; H625 in *An. darlingi*), so may impact the binding of insecticides. The N294H alteration occurs 9 amino acids upstream of the G119S (G305S) mutation, and results in a change from an amino acid with a polar uncharged side chain

(asparagine) to one with a charged side chain with an aromatic imidazole ring. For all five of these amino acid alterations, the reference amino acid is conserved across other *Anopheles* species. Futher studies, which include bioassays, are needed to confirm if these mutations result in organophosphate or carbamate resistance. It is not possible at present to exclude potential candidate SNPs on the basis of their presence in the colony samples, as these have demonstrated resistance to both carbamates and pyrethroids, data that is currently being confirmed with bioassays.

The four study locations had varying insecticide usage, with Porto Velho (State of Rondônia), and Manaus (State of Amazonia) sites having a history of intensive use of pyrethroids and carbamates previously through IRS. Whilst Candeias do Jamari (State of Rondônia) and Manacapuru (State of Amazonas), had little to no insecticide usage. All missense SNPs identified in this study were found in locations with intensive insecticide usage, and locations with little to no insecticide usage. Additional studies, combining phenotypic and molecular surveillance data, are needed to understand the impact of the SNPs detected here on insecticide resistance in *An. darlingi.*

The lack of SNPs found in the *rdl* gene is unsurprising. Mutations in *gaba* result in dieldrin resistance, an organochloride that has been prohibited from use for at least the past decade, due to its adverse effects on human health. In relation to the *vgsc* gene, no *kdr* mutations were detected in the populations surveyed here. This observation was also reported in pyrethroid resistant populations of *An. darlingi* from Colombia[33]. The lack of *kdr* mutations suggests that SNPs in v*gsc* may play a reduced role in *An. darlingi* insecticide resistance. An INDEL resulting in isoleucine deletion at position 422 was detected in 25% of sample, across both field and colony sources. Further studies are needed to understand the involvement of this deletion in insecticide resistance.

The lack of known resistance associated SNPs in *An. darlingi* could be due to the amount of gene flow between geographically close populations. The mixing of possible resistant urban populations with

insecticide-sensitive forest populations on the borders of urban areas may minimise the frequency of resistant alleles[19]. An alternative hypothesis is other genetic variants are important for this species, or the involvement of other mechanisms, such as differential expression of genes.

In relation to the genetic diversity across populations, there was little differentiation between *An. darlingi* from Brazil and other regions using *its2* gene data. However, the phylogenetic tree constructed using *cox-1* gene data revealed that samples generally grouped by country. It has been hypothesised that physical barriers like the Atlantic forest mountain range or the Amazon River, prevent the mixing of these populations and so they appear genetically distinct[58,59]. Data from the *cox-1* gene revealed an outlying sample, which coincided with another species (*An. peryassui*).

Within the Brazilian population, low nucleotide diversity was seen in tandem with high haplotype diversity for both *cox-1* and *its2* genes, indicating many low frequency variants. This observation is consistent with previous studies conducted on *cox-1* across Central and South American, and within Brazilian and Colombian *An. darlingi* populations [13,57,60]. The geographical proximity of the collection sites of these mosquitoes may also contribute to the low genetic diversity observed here. The inclusion of study sites that are more geographically distant or from other countries may give greater resolution to the population dynamics of this species.

Overall, our amp-seq panel provides a tool for the better understanding of the genomic landscapes of this understudied *An. darlingi* vector. It is a high-throughput, low-cost assay for species identification and the detection of novel SNPs in insecticide resistance associated genes. Further investigation is required to identify whether these SNPs contribute to insecticide resistance in *An. darlingi*. A clear limitation of this method is that it currently only targets known loci, however the panel is easily adaptable to new targets, including metabolic markers. The panel can be used in tandem with phenotypic assays to identify SNPs that result in functional changes. Large-scale surveillance methods

are urgently needed to inform malaria vector control methods in Brazil, particularly to assist initiatives to reduce malaria transmission. Our panel represents the first steps towards a molecular surveillance method to track known and identify potencial markers of resistance.

Data Availability

All raw sequence data is listed in the European Nucleotide Archive (Project ID: PRJEB61194, Accession numbers: ERR11204754 to ERR11204953).

Acknowledgements

We thank for the members of the PIVEM/Fiocruz Rondônia for *An. darlingi* culture. JFM also gratefully acknowledge their CNPq productivity grant (process number: 304830/2022-4).

Funding

This work was supported, in part, by the Brazilian Ministry of Health/DECIT/CNPq N° 23/2019 (grant number 442653/2019-0) and the Bill & Melinda Gates Foundation (INV-003970). Under the grant conditions of the foundation, a Creative Commons Attribution 4.0 Generic License has already been assigned to the author accepted manuscript version that might arise from this submission. The study also received funding from the International Centres of Excellence for Malaria Research (ICEMR) program GR109237 (CON-80002357). H.A.P. is funded by a BBSRC LIDo-DTP PhD studentship. J.P. was funded by a Newton Institutional Links Grant (British Council, no. 261868591). T.G.C. and S.C. were funded by UKRI grants (ref. BB/X018156/1, MR/X005895/1).

Conflicts of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Global Malaria Programme: WHO Global. *WHO World Malaria Report 2022*. *WHO*.

2. Laporta, G. Z., Grillet, M. E., Rodovalho, S. R., Massad, E. & Sallum, M. A. M. Reaching the malaria elimination goal in Brazil: a spatial analysis and time-series study. *Infect. Dis. Poverty* **11**, 39 (2022).

3. Carlos, B. C., Rona, L. D. P., Christophides, G. K. & Souza-Neto, J. A. A comprehensive analysis of malaria transmission in Brazil. *Pathog. Glob. Health* **113**, 1–13 (2019).

4. Martins-Campos, K. M. *et al.* Integrated vector management targeting Anopheles darlingi populations decreases malaria incidence in an unstable transmission area, in the rural Brazilian Amazon. *Malar. J.* **11**, 351 (2012).

5. Girod, R., Gaborit, P., Carinci, R., Issaly, J. & Fouque, F. Anopheles darlingi bionomics and transmission of Plasmodium falciparum, Plasmodium vivax and Plasmodium malariae in Amerindian villages of the Upper-Maroni Amazonian forest, French Guiana. *Mem. Inst. Oswaldo Cruz* **103**, 702–710 (2008).

6. Prado, C. C. *et al.* Behavior and abundance of Anopheles darlingi in communities living in the Colombian Amazon riverside. *PLoS ONE* **14**, (2019).

7. Oliveira-Ferreira, J. *et al.* Malaria in Brazil: An overview. *Malar. J.* **9**, 1–15 (2010).

8. Hiwat, H. & Bretas, G. Ecology of Anopheles darlingi Root with respect to vector importance: a review. *Parasit. Vectors* **4**, 177 (2011).

9. Foster, P. G. *et al.* Phylogeny of Anophelinae using mitochondrial protein coding genes. *R. Soc. Open Sci.* **4**, 170758 (2017).

10. Charlwood, J. D. Biological variation in Anopheles darlingi root. *Mem. Inst. Oswaldo Cruz* **91**, 391–398 (1996).

11. Da Silva-Nunes, M. *et al.* Malaria on the amazonian frontier: Transmission dynamics, risk factors, spatial distribution, and prospects for control. *Am. J. Trop. Med. Hyg.* **79**, 624–635 (2008).

12. Angêlla, A. F. *et al.* Seasonal genetic partitioning in the neotropical malaria vector, Anopheles darlingi. *Malar. J.* **13**, 1–10 (2014).

13. Moutinho, P. R., Gil, L. H. S., Cruz, R. B. & Ribolla, P. E. M. Population dynamics, structure and behavior of Anopheles darlingi in a rural settlement in the Amazon rainforest of Acre, Brazil. *Malar. J.* **10**, 1–13 (2011).

14. Torres, K. *et al.* Malaria Resilience in South America: Epidemiology, Vector Biology, and Immunology Insights from the Amazonian International Center of Excellence in Malaria Research Network in Peru and Brazil. *Am. J. Trop. Med. Hyg.* **107**, 168–181 (2022).

15. Vector-Focused Approaches to Curb Malaria Transmission in the Brazilian Amazon: An Overview of Current and Future Challenges and Strategies - PMC. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7709627/.

16. Baia-da-Silva, D. C. *et al.* Current vector control challenges in the fight against malaria in Brazil. *Rev. Soc. Bras. Med. Trop.* **52**, e20180542 (2019).

17. Campos, K. B. *et al.* Assessment of the susceptibility status of Aedes aegypti (Diptera: Culicidae) populations to pyriproxyfen and malathion in a nation-wide monitoring of insecticide resistance performed in Brazil from 2017 to 2018. *Parasit. Vectors* **13**, 531 (2020).

18. Leandro, A. de S. *et al.* Malathion insecticide resistance in Aedes aegypti: laboratory conditions and in situ experimental approach through adult entomological surveillance. *Trop. Med. Int. Health* **25**, 1271–1282 (2020).

19. Vezenegho, S. *et al.* Variation in Pyrethroid Resistance Phenotypes in Anopheles darlingi in an Area with Residual Malaria Transmission: Warning of Suspected Resistance in French Guiana. *Am. J. Trop. Med. Hyg.* **1**, (2022).

20. Fonseca-González, I., Quiñones, M. L., McAllister, J. & Brogdon, W. G. Mixed-function oxidases and esterases associated with cross-r esistance between DDT and lambda-cyhalothrin in Anopheles darlingi Root 1926 populations from Colombia. *Mem. Inst. Oswaldo Cruz* **104**, 18–26 (2009).

21. Organisation, W. H. *Global report on insecticide resistance in malaria vectors: 2010-2016*. (2017).

22. World Health Organization. *Global report on insecticide resistance in malaria vectors: 2010-2016*. *WHO* (2016).

23. WHO. Global Malaria Programme: Malaria Threats Map. https://www.who.int/teams/global-malaria-programme/surveillance/malaria-threats-map.

24. Namias, A., Jobe, N. B., Paaijmans, K. P. & Huijben, S. The need for practical insecticide-resistance guidelines to effectively inform mosquito-borne disease control programs. *eLife* **10**, e65655.

25. 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).

26. 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).

27. 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 Biol.* **15**, R27 (2014).

454 28. Helvecio, E. *et al.* Polymorphisms in GSTE2 is associated with temephos resistance in Aedes
455 aegypti. *Pestic. Biochem. Physiol.* **165**, 104464 (2020).

456 29. Ibrahim, S. S., Ndula, M., Riveron, J. M., Irving, H. & Wondji, C. S. The P450 CYP6Z1 confers
457 carbamate/pyrethroid cross-resistance in a major African malaria vector beside a novel
458 carbamate-insensitive N485I acetylcholinesterase-1 mutation. *Mol. Ecol.* **25**, 3436–3452
459 (2016).

460 30. Ffrench-Constant, R. H., Rocheleau, T. A., Steichen, J. C. & Chalmers, A. E. A point mutation in a
461 Drosophila GABA receptor confers insecticide resistance. *Nature* **363**, 449–451 (1993).

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

464 32. Lol, J. C. *et al.* Development of molecular assays to detect target-site mechanisms associated
465 with insecticide resistance in malaria vectors from Latin America. *Malar. J.* **18**, 202 (2019).

466 33. Orjuela, L. I. *et al.* Absence of knockdown mutations in pyrethroid and DDT resistant
467 populations of the main malaria vectors in Colombia. *Malar. J.* **18**, 1–9 (2019).

468 34. Campos, M. *et al.* High-throughput barcoding method for the genetic surveillance of insecticide
469 resistance and species identification in Anopheles gambiae complex malaria vectors. *Sci. Rep.*
470 **12**, (2022).

471 35. Collins, E. L. *et al.* A next generation targeted amplicon sequencing method to rapidly screen
472 for insecticide resistance mutations in Aedes aegypti populations reveals a rdl mutation in
473 mosquitoes from Cabo Verde. 1–28.

474 36. Makunin, A. *et al.* A targeted amplicon sequencing panel to simultaneously identify mosquito
475 species and Plasmodium presence across the entire Anopheles genus. *Mol. Ecol. Resour.* **n/a**,
476 (2021).

37. Nag, S. *et al.* High throughput resistance profiling of Plasmodium falciparum infections based on custom dual indexing and Illumina next generation sequencing-technology. *Sci. Rep.* **7**, 2398 (2017).

38. Andrade, A. O. *et al.* Description of malaria vectors (Diptera: Culicidae) in two agricultural settlements in the Western Brazilian Amazon. *Rev. Inst. Med. Trop. São Paulo* **63**, e60 (2021).

39. Katsuragawa, T. H. *et al.* The Dynamics of Transmission and Spatial Distribution of Malaria in Riverside Areas of Porto Velho, Rondônia, in the Amazon Region of Brazil. *PLOS ONE* **5**, e9245 (2010).

40. Araujo, M. da S. *et al.* Brazil's first free-mating laboratory colony of *Nyssorhynchus darlingi*. *Rev. Soc. Bras. Med. Trop.* **52**, e20190159 (2019).

41. Consoli, R. A. G. B. & Oliveira, R. L. de. *Principais mosquitos de importância sanitária no Brasil*. (Editora FIOCRUZ, 1994).

42. Untergasser, A. *et al.* Primer3—new capabilities and interfaces. *Nucleic Acids Res.* **40**, e115–e115 (2012).

43. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinforma. Oxf. Engl.* **30**, 2114–2120 (2014).

44. H.Li. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arxiv.org* **n/a**, n/a (2013).

45. Anopheles darlingi genome assembly idAnoDarlMG_H_01. *NCBI* https://www.ncbi.nlm.nih.gov/data-hub/assembly/GCF_943734745.1/.

46. McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

47. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. (2012).

48. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10**, 1–4 (2021).

49. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w(1118); iso-2; iso-3. *Fly (Austin)* **6**, 80–92 (2012).

50. Katoh, K., Misawa, K., Kuma, K. I. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).

51. Larsson, A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30**, 3276–3278 (2014).

52. Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455 (2019).

53. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).

54. Paradis, E. pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics* **26**, 419–420 (2010).

55. Wickham, H. ggplot2. *ggplot2* (2009) doi:10.1007/978-0-387-98141-3.

56. Mitchell, S. N. *et al.* Metabolic and Target-Site Mechanisms Combine to Confer Strong DDT Resistance in Anopheles gambiae. *PLoS ONE* **9**, 92662 (2014).

57. Mirabello, L. & Conn, J. E. Molecular population genetics of the malaria vector Anopheles darlingi in Central and South America. *Heredity* **96**, 311–321 (2006).

58. Emerson, K. J., Conn, J. E., Bergo, E. S., Randel, M. A. & Sallum, M. A. M. Brazilian Anopheles darlingi Root (Diptera: Culicidae) Clusters by Major Biogeographical Region. *PLOS ONE* **10**, e0130773 (2015).

59. Pedro, P. M. & Sallum, M. A. M. Spatial expansion and population structure of the neotropical malaria vector, Anopheles darlingi (Diptera: Culicidae). *Biol. J. Linn. Soc.* **97**, 854–866 (2009).

60. Angêlla, A. F., Gil, L. H. S., Da Silva, L. H. P. & Ribolla, P. E. M. Population structure of the malaria vector Anopheles darlingi in Rondônia, Brazilian Amazon, based on mitochondrial DNA. *Mem. Inst. Oswaldo Cruz* **102**, 953–958 (2007).

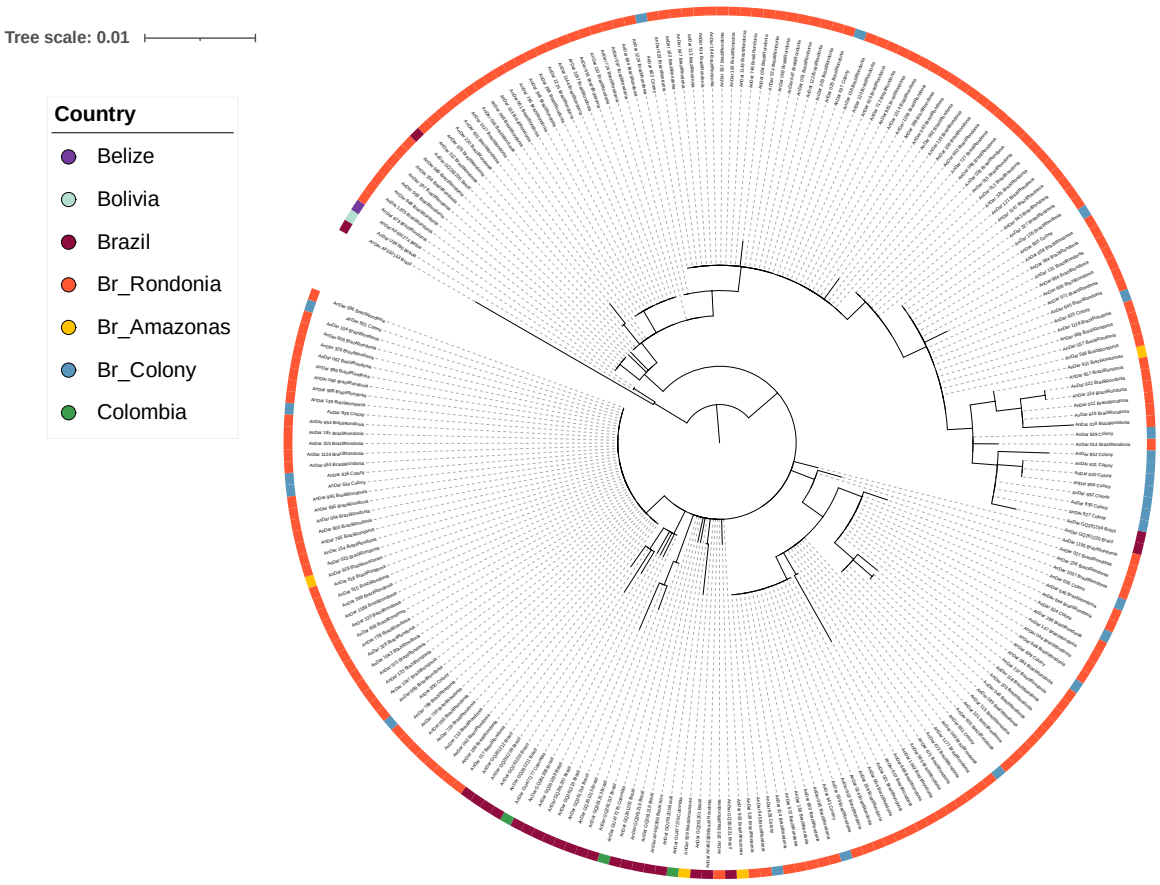


Figure 1. Maximum-likelihood tree constructed using Its2 gene sequences generated in this study (total=198, Br_Rondonia=171, Br_Amazonas=4, Br_Colony=23), alongside other publicly available *An. darlingi* ITS2 sequences (n=26, Brazil, Colombia, Belize, and Bolivia). The tree was built using the maximum-likelihood method assuming GTR model of nucleotide substitution, with the gamma model of heterogeneity rate.

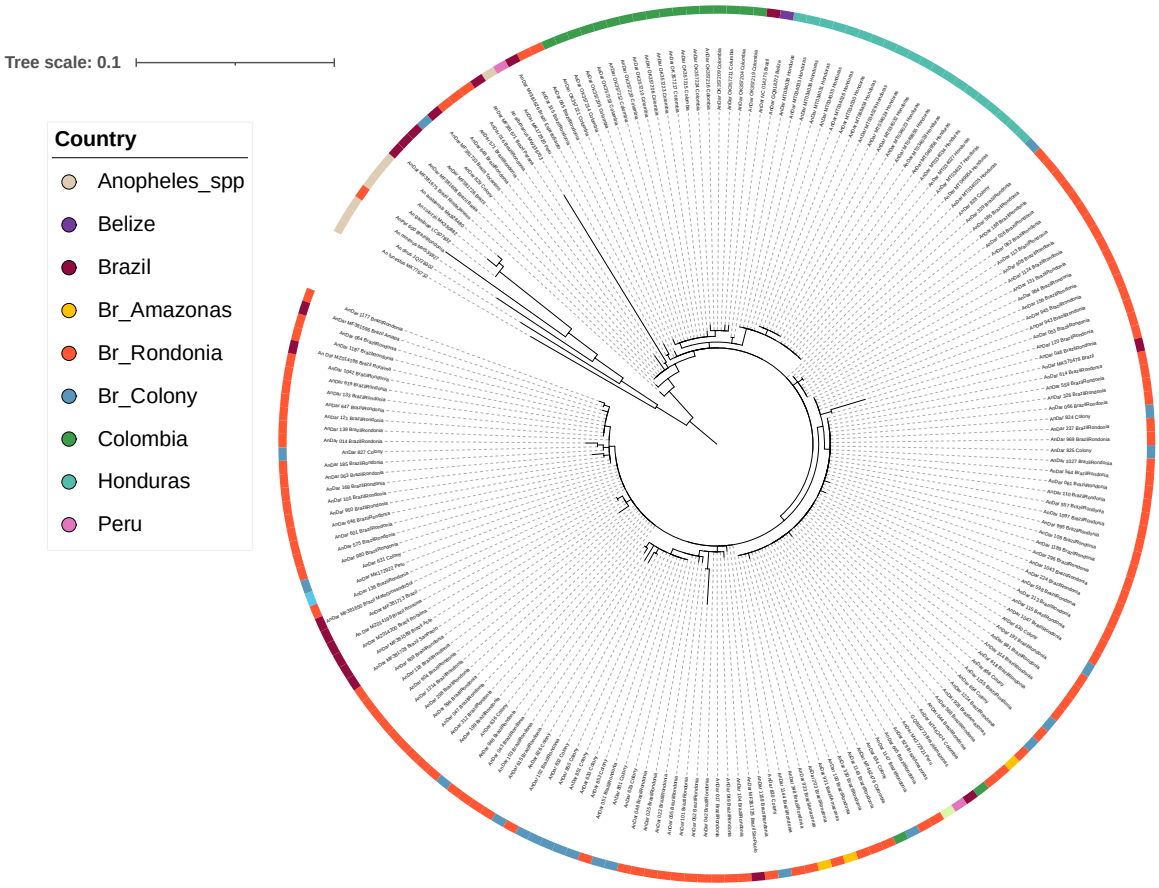


Figure 2. Maximum-likelihood tree constructed using *Cox-1* gene sequences generated in this study (total=129, Br_Rondonia=106, Br_Amazonas=3, Br_Colony=20), alongside other publicly available *An. darlingi cox-1* sequences (n=67, Brazil, Honduras, Belize, Colombia, Panama, Ecuador, and Peru). This tree also has a group of *Anopheles spp.* including *An.albimanus, An.arabiensis, An.coluzzi, An.dirus, An.funestus,* and *An.gambiae.* The tree was built using the maximum-likelihood method assuming GTR model of nucleotide substitution, with the gamma model of heterogeneity rate.

Table 1. Primers and targets of *An. darlingi* amp-seq panel

***Anopheles darlingi* Primers**								
Aim	**Target Gene**	**Amplicon**	**Accession ID**	**Target SNP***	**Exon Span**	**Forward primer**	**Reverse Primer**	**Product Size (bp)**
Insecticide Resistance	vgsc	VGSCI	ADAC011160	V416L	9 - 10	GCCTTTCGTCTAATGACTCAAGA	GCCAAGATTAAATTTACAAGGTAAAAC	500
		VGSCII		L1014F	20 - 21	ACCGTTTCCCCGATAAAGAC	ACGGACGCAATTTGACTTGT	450
		VGSCIII		F1511C/ N1552Y	30	TTTTCCAGGTTGCCACTTTC	ATTGCTTGTGGCCTCCACT	475
		VGSCIV		D1739Y	32 - 33	AAAATATTTCGTTTCCCCAACA	TCCCAGGATAACCTTTGTCG	447
	ace-1	ACE1_1	ADAC000377	G305S	2	TAAGAAGGTGGACGTGTGGC	AGAGCAAGGTTCTGATCGAA	450
		ACE1_II		N642I	4-5	GACGGGGTACGTCGACAA	AAGGCGCTACTTTCACACG	500
	GSTe2	GSTe2	ADAC008205	L119F	3	TTCGAATCCGGTGTGATCTT	TGGTCACGATCATCTTTATTGG	471
	rdl	RDL1	ADAC005672	A296S/ V371I	7	CACCAACACCAGTCTGATCG	TGGCAAATACCATGACGAAG	490
		RDL2		T345S	8	TGGTTTTTCCCAATCGTTTT	CTGCCCATCTGCTGCTTC	492
Phylogeny	*cox*-1	COX-1	HM022406.1	n/a	n/a	TCTCCAGGGATTACTTTAGATCG	GCTGGGCTGTATGTTAATTGAG	494
	ITS2	ITS2	KF436940.1	n/a	n/a	GACTCAGTGCGAGGTACACA	GAGGCCCACTTGAGATCCTA	455

*Target SNP loci in *An. darlingi*

Table 2. Average amplicon coverage, and number of genetic variants identified

Amplicon	Average coverage	Number of SNPs	Number of Non-synonymous SNPs	Number of INDELs	Number of Non-synonymous INDELs
ACE1_I	3190.40	43	2	0	0
ACE1_II	1415.81	68	3	12	0
COI	1996.52	20	0	0	0
GSTe2	3196.05	40	5	1	0
ITS2	5621.32	6	0	0	0
VGSCI	272.17	9	0	3	1
VGSCII	216.81	4	0	0	0
VGSCIII	780.34	4	0	0	0
VGSCIV	1620.89	1	0	2	0
Rdl1	171.17	16	0	1	0
Rdl2	895.84	35	0	1	0

Table 3. Locations and allelic frequencies of detected non-synonymous variants

Chromosome	Amplicon	Position	Sample Number	Annotation	Genotype Frequencies			Allele Frequencies	
					0/0	0/1	1/1	0	1
SNPs									
NC_064874.1	ACE1_I	15679573	191	Val243Ile	92.67	6.81	0.52	96.07	3.93
		15679726	191	Asn294His	97.38	2.62	0	98.70	1.30
	ACE1_II	15681121	149	Ser673Asn	95.30	4.03	0.67	97.32	2.68
		15681124	148	Ser674Asn	77.70	19.60	2.70	87.50	12.50
		15681124	145	Ser674Thr	98.64	0.68	0.68	98.97	1.03
	GSTe2	89825807	129	Ile114Val	95.35	4.65	0	97.67	2.33
		89825922	129	Asp128Glu	0	7.75	92.25	3.88	96.12
		89826035	128	Thr166Ile	87.50	12.50	0	93.75	6.25
		89826074	128	Thr179Ile	78.91	19.53	1.56	88.67	11.33
		89826151	129	Thr205Ala	0	6.20	93.80	3.10	96.90
INDELs									
NC_064875.1	VGSCI	35317107	40	Ile422del	67.5	30.0	2.5	82.5	17.5

Table 4. Number of genetic variants identified per population

Population	Sample Number	Number of SNPs	Number of Non-synonymous SNPs	Number of INDELs	Number of Non-synonymous INDELS
All	200	246	10	20	1
Amazonas State	4	91	7	14	0
Rondônia State	171	241	10	19	1
Colony	25	133	6	16	1

Table 5. Nucleotide and haplotype diversities of *cox-1* per population

Population	Nucleotide diversity	Haplotype diversity
All (n=129)	0.00653	0.856
Amazonas (n=4)	N/A	N/A
Rondônia (n=105)	0.00688	0.867
Colony (n=20)	0.00611	0.837

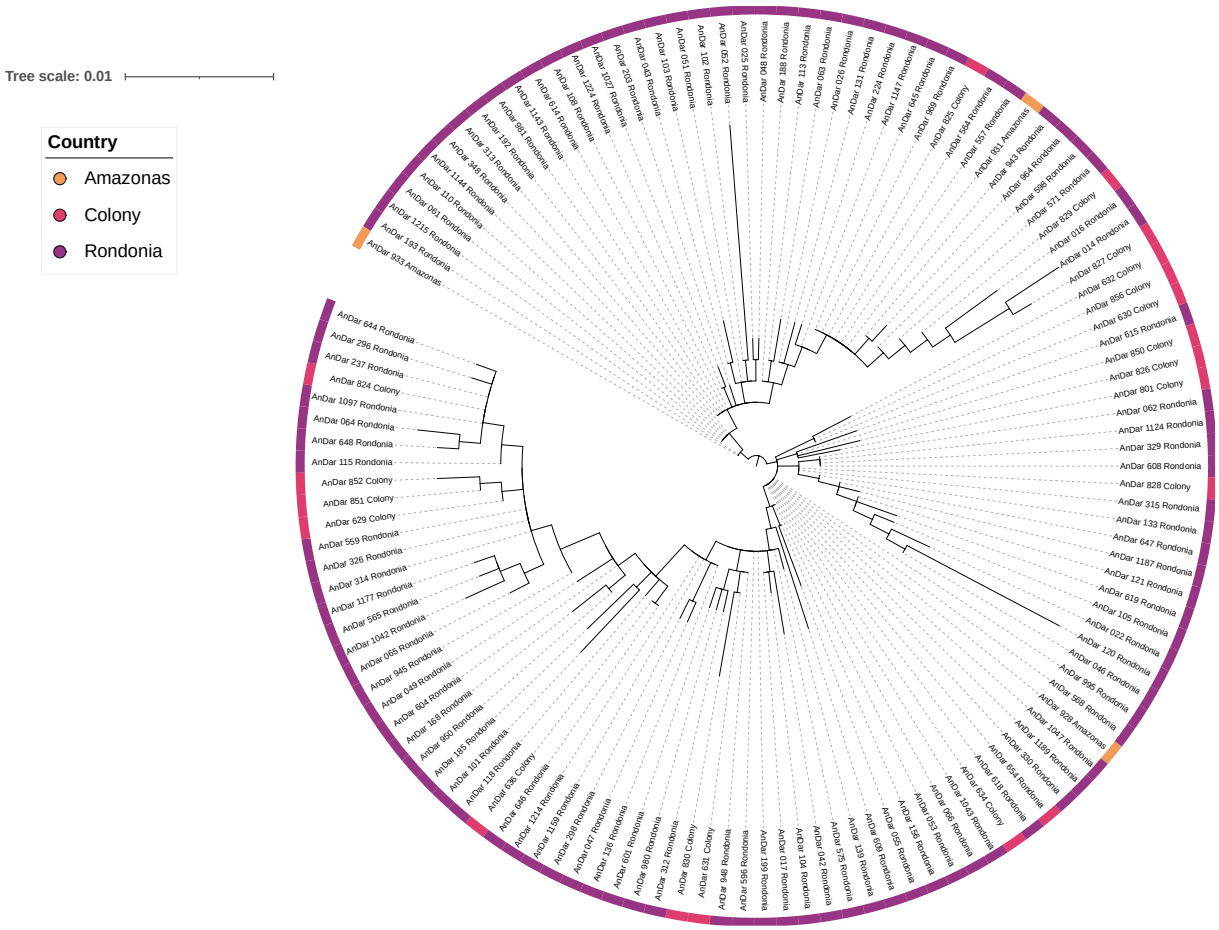
Table 6. Nucleotide and haplotype diversities of ITS2 per population

Population	Nucleotide diversity	Haplotype diversity
All (n=198)	0.00536	0.893
Amazonas (n=4)	N/A	N/A
Rondônia (n=172)	0.00607	0.883
Colony (n=23)	0.00639	0.913

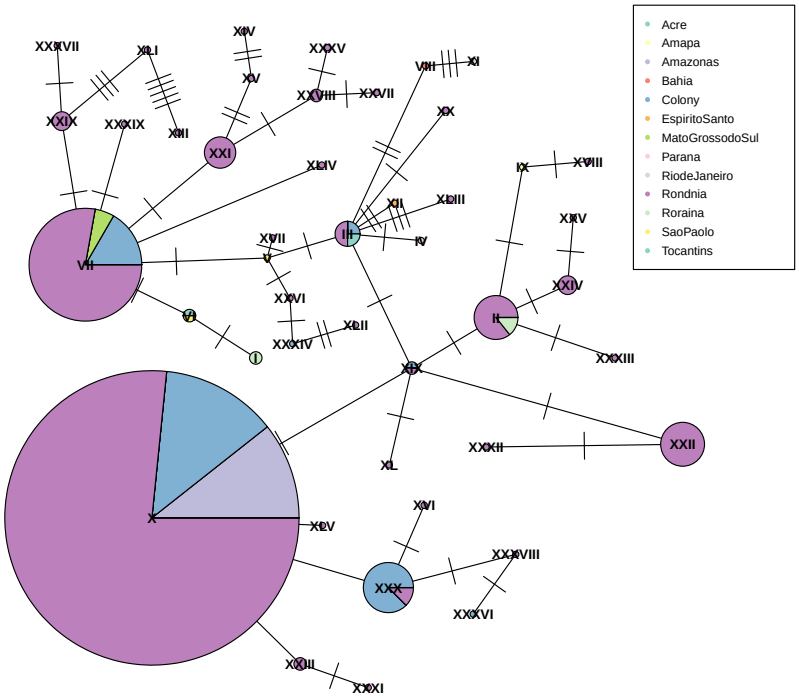
Table 7. Genotype frequencies of NS SNPs in the four collection sites

Amplicon	SNP Position	Amino Acid Change	State of Rondonia						State of Amazonas					
			Porto Velho* (50)			Candeias do Jamari (n=113)			Manacapuru (n=3)			Manaus* (n=1)		
			0/0	0/1	1/1	0/0	0/1	1/1	0/0	0/1	1/1	0/0	0/1	1/1
ACE1_I	15679573	Val243Ile	90.0%	8.0%	2.0%	93.8%	6.2%	0.0%	66.7%	33.3%	0.0%	0.0%	100%	0.0%
	15679726	Asn294His	96.0%	4.0%	0.0%	97.3%	2.7%	0.0%	100%	0.0%	0.0%	100%	0.0%	0.0%
ACE1_II	15681121	Ser673Asn	100%	0.0%	0.0%	96.4%	2.4%	1.20%	66.7%	33.3%	0.0%	0.0%	100%	0.0%
	15681124	Ser674Asn	78.3%	17.4%	4.3%	77.4%	20.2%	2.40%	100%	0.0%	0.0%	100%	0.0%	0.0%
	15681124	Ser674Thr	97.7%	0.0%	2.3%	98.8%	1.2%	0.0%	100%	0.0%	0.0%	100%	0.0%	0.0%
GSTe2	89825807	Ile114Val	97.6%	2.4%	0.0%	93.3%	6.7%	0.0%	100%	0.0%	0.0%	100%	0.0%	0.0%
	89825922	Asp128Glu	0.00%	4.9%	95.1%	0.0%	10.7%	89.3%	0.0%	0.0%	100%	0.0%	0.0%	100.0%
	89826035	Thr166Ile	90.2%	9.8%	0.0%	86.5%	13.5%	0.0%	66.7%	33.3%	0.0%	100%	0.0%	0.0%
	89826074	Thr179Ile	80.5%	17.1%	2.4%	77.0%	23.0%	0.0%	66.7%	0.0%	33.7%	100%	0.0%	0.0%
	89826151	Thr205Ala	0.00%	4.8%	95.2%	0.0%	8.0%	92.0%	0.0%	0.0%	100%	0.0%	0.0%	100%

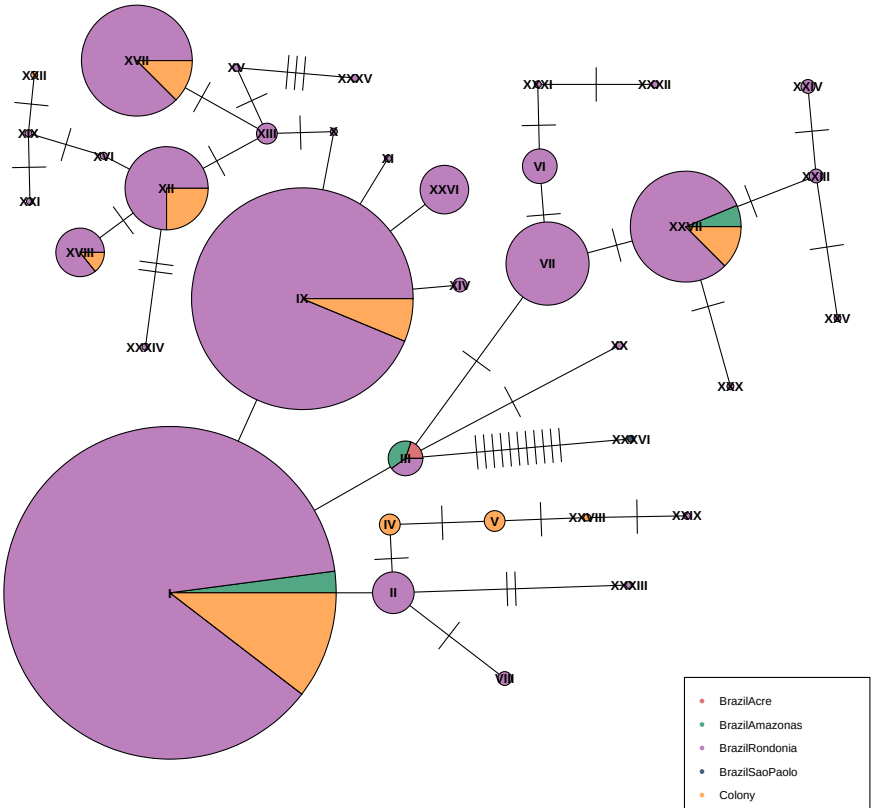
*Indicates insecticides are regularly used in this location



Supplementary figure 1. Maximum-likelihood tree constructed using concatenated *ITS2* and *cox-1* gene sequences generated in this study (n=127). This included samples from Rondônia (n = 106), Colony (n=18), Amazonas (n=3). The tree was built using the maximum-likelihood method assuming GTR model of nucleotide substitution, with the gamma model of heterogeneity rate.



Supplementary Figure 2. Haplotype or minimal-spanning network constructed using *cox-1* sequences generated in this study. Each node represents a haplotype, each segment within the node represents a region, and is proportionally sized to the number of sequences present in the segment and node. The number of number of ticks between nodes represents the number of genetic differences between nodes.



Supplementary Figure 3. Haplotype or minimal-spanning network constructed using ITS2 sequences generated in this study. Each node represents a haplotype, each segment within the node represents a region, and is proportionally sized to the number of sequences present in the segment and node. The number of number of ticks between nodes represents the number of genetic differences between nodes.

Chapter 5: Genome-wide population genetics of *Anopheles darlingi*, a malaria vector from the State of Rondônia, Brazil

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	lsh1704757	**Title**	Miss
First Name(s)	Holly		
Surname/Family Name	Acford-Palmer		
Thesis Title	A genomic investigation into three malaria vectors: An. darlingi, An. funestus, and An. stephensi		
Primary Supervisor	Prof. Susana Campino		

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?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
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?	Parasites & Vectors
Please list the paper's authors in the intended authorship order:	Holly Acford-Palmer, Emilia Manko, Alice O. Andrade, Jody Phelan, Rosa A. Santana, Stefanie C. P. Lopes, Matthew Higgins, Jansen F Medeiros, Taane G. Clark, Maisa S. Araujo, Susana Campino
Stage of publication	**Not yet 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 extracted and cleaned the DNA for sequencing. I then performed the bioinformatic analyses and interpreted results under the guidance of my supervisors. I wrote the first draft of the manuscript, which was then commented on by my supervisors and co-authors.

SECTION E

Student Signature	[redacted]
Date	

Supervisor Signature	
Date	

Title: Genome-wide population genetics of *Anopheles darlingi*, a malaria vector from the State of Rondônia, Brazil

Authors: Holly Acford-Palmer[1], Emilia Manko[1], Alice O. Andrade[2], Jody Phelan[1], Rosa A. Santana[3], Stefanie C. P. Lopes[3,4], Matthew Higgins[1], Jansen F Medeiros[2,6], Taane G. Clark[1,5*] Maisa S. Araujo[2,7,8*], Susana Campino [1*]

[1] Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom

[2] Plataforma de Produção e Infecção de Vetores da Malária- PIVEM, Laboratório de Entomologia, Fiocruz Rondonia, Porto Velho, RO, Brazil

[3] Instituto de Pesquisa Clínica Carlos Borborema, Fundação de Medicina Tropical Dr. Heitor Vieira Dourado, Manaus, AM, Brazil

[4] Instituto Leônidas & Maria Deane, FIOCRUZ, Manaus, AM, Brazil

[5] Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom

[6] Programa de Pós-Graduação em Biologia Experimental – PGBIOEXP, Fundação Universidade Federal de Rondonia, Porto Velho, RO, Brazil

[7] Programa de Pós-Graduação em Conservação e uso de Recursos Naturais – PPGReN, Fundação Universida-de Federal de Rondonia, Porto Velho, RO, Brazil

[8] Laboratório de Pesquisa Translacional e Clínica, Centro de Pesquisa em Medicina Tropical, Porto Velho, RO, Brazil

[*]Joint last authors

[**]Corresponding Author:

Professor Susana Campino, LSHTM, Susana.campino@lshtm.ac.uk

Maisa S. Araujo, PIVEM, maisa.araujo @fiocruz.br

Abstract

Anopheles (*Nyssorhynchus*) *darlingi* is a primary malaria vector in Central and South America and contributes to the majority of the area's malaria transmission, particularly in the Amazon region. The importance of *An. darlingi*'s contribution to the region's malaria burden, incentivises efforts for the genomic investigation of this mosquito to support new opportunities for vector control and identify markers of insecticide resistance. Here, for the first time, we applied whole genome sequencing (WGS) to field (n=23) and colony populations (n=8) of *An. darlingi,* sourced from different locations in the State of Rondônia, Brazil. A total of 16.7 million high-quality single nucleotide polymorphisms (SNPs) were identified across 31 *An. darlingi* isolates. Population genomic analysis revealed the colony and field isolates are genetically distinct from each other, consistent with their geographical differentiation, and homogenous within each isolate type. No SNPs previously associated with insecticide resistance were detected, but several non-synonymous mutations were observed in the four genes linked to insecticide resistance (*ace-1, gaba receptor (rdl), GSTe2,* and *vgsc*). Signals of selection were detected in genes previously associated with resistance in *Anopheles* mosquitoes and other insects (e.g., *CYP4c1, CYP4c3, and CYP307a1),* which require follow-up investigation to assess their contribution to insecticide resistant phenotypes. Overall, this work provides the first catalogue of genome-wide polymorphisms in *An. darlingi*, thereby guiding the identification of resistance drivers in this vector species and support new efforts for vector control.

Word count: 230

Introduction

The *Anopheles darlingi* (or *Nyssorhynchus darlingi)* mosquito is a key malaria vector throughout South and Central Americas[1]. In Brazil, one of the countries with highest malaria burden in the region, this vector species is responsible for a substantial proportion of the country's malaria transmission. *An. darlingi* is a very efficient malaria vector due to its high susceptibility to *Plasmodium spp* infection. It can maintain *Plasmodium* transmission even with low parasite densities, and is highly anthropophilic in behaviour[2,3]. This vector is both exophagic and endophagic, primarily feeding on humans[1]. As a species, it's highly adaptable, often filling ecological niches left empty by other *Anopheles spp*[1,4]. These niches are often recently deforested areas or "frontiers", which tend to have higher abundances of *An. darlingi* comparative to other species[5]. The ongoing anthropization of these areas is believed to be causing perennial transmission of malaria, due to the increased availability of larval habitats[6,7]. Typically, forest populations of *An. darlingi* would be highest during or at the end of the rainy season, due to increased flooding. However, human-led changes have made these once temporary larval habitats more permanent, thereby promoting longer-term malaria transmission, leading to a greater public health risk[6,7].

In 2021, one-fifth of the 600,000 malaria cases diagnosed in South America, occurred in Brazil[8]. Since 2015, the number of malaria cases in Brazil have risen 3% from ~138,000 to in excess of 142,000, with the vast majority occurring within the country's Amazon region[8–10]. Almost all (99%) of malaria cases in Brazil are caused by *Plasmodium vivax* (83%), followed by *P. falciparum* (16%)[8]. Any increase in malaria cases is a public health concern, but particularly in Brazil as the proportion of cases caused by *P. falciparum* has increased by 7%[10]. *P. falciparum* causes the most severe disease and results in the highest mortality, particularly in children. Malaria control in Brazil has never been more crucial to prevent further increases in transmission, and deliver on the country's own National Elimination Plan alongside the World Health Organization's goal of reducing malaria infection by 90% by 2030[11].

Current malaria control efforts in Brazil rely on indoor residual spraying (IRS) of insecticides, and the use of long-lasting insecticide treated nets (LLINs)[12,13]. Pyrethroids are the most used insecticide class for *Anopheles spp.* in Brazil, but carbamates and organophosphates are also applied for *Aedes spp.* mosquitoes in arbovirus control programmes[14,15]. Screening for insecticide resistance is key to optimising insecticide usage to reduce malaria transmission. However, there is currently no national screening in place for insecticide resistance in *Anopheles*, despite the existence of such a programme for *Ae. Aegypti*[14]. Resistance to some classes of adulticides (carbamates, organochlorides, and pyrethroids) in *An. darlingi* has been reported in countries neighbouring Brazil, such as Bolivia, Colombia, French Guiana, and Peru[16,17]. However, no resistance to insecticides in *An. darlingi* has been reported in Brazil, likely due to a lack of surveillance and reporting[16].

Insecticide resistance can arise through multiple mechanisms including, target site resistance, metabolically mediated resistance, behavioural avoidance, cuticle thickening, and microbiome alterations[18,19]. The most observed mechanism is target site resistance, where selective pressure results in single nucleotide polymorphisms (SNPs), causing amino acid changes in the proteins targeted by the insecticide. Of the few studies screening for molecular mechanisms of resistance in *An. darlingi* mosquitos, none have identified known SNPs associated with resistance in other *Anopheles* vectors[20–24]. Markers such as the knockdown resistance (*kdr*) mutation or L1014F/S, in the *voltage-gated sodium channel* (*vgsc*) linked to pyrethroid and Dichlorodiphenyltrichloroethane (DDT) resistance; or the G119S alteration in the *acetylcholinesterase* (*ace-1)* gene causing carbamate resistance, have been absent in *An. darlingi*[25–27]*.* Other commonly observed mutations include the L119F in the *GSTe2* gene, which is a metabolic marker for DDT and pyrethroid resistance[28]. The A296, V327I, T345S markers are found in the *gaba* gene, otherwise known as *rdl* or resistant to dieldrin[29]. Dieldrin belongs to a subgroup of organochlorides known as cyclodienes, which were now banned due to human health concerns. None of the *rdl* or *GSTe2* mutations have been observed in *An. darlingi* thus far.

The other frequently observed mechanism of resistance in insects is metabolic, where gene expression changes result in increased activity of detoxification enzymes, reducing insecticide efficacy[28,30–32]. Due to the lack of target site resistance mutations identified in *An. darlingi*, it has been hypothesised that metabolic mechanisms may play a prominent role in resistance for this species than others[33].

To date, molecular studies of *An. darlingi* have focused on screening a few regions in candidate genes (e.g., *vgsc, ace-1*), but genome-wide approaches are required to explore the entire genomic landscape and support the identification of new candidate loci involved in resistance. The resulting insights can inform the development of molecular tools to rapidly screen for the emergence of resistant alleles, which is key to optimising vector control methods. Whole genome sequencing (WGS) has been applied to many vectors, and specifically *Anopheles spp.,* to better understand population dynamics, and insecticide resistance mechanisms. The development of next generation sequencing techniques means the cost of WGS has drastically decreased, and sequence quality has improved. This increase in throughput allows for greater insight into the genomic architecture of vector populations, including the identification of genomic regions under recent positive selection. However, WGS has never been conducted on *An. darlingi,* until now.

Here we apply WGS to two Brazilian populations of *An. darlingi*, a wildtype population from Candeias dos Jamari, State of Rondônia, and a colony population. Through examining ancestry, population structure, and genetic regions under selection, we sought to gain insights into genomic diversity and molecular markers of insecticide resistance[34].

Results

Whole Genome Sequence Data for *An. darlingi*

Forty *An. darlingi* isolates underwent WGS on an Illumina platform, of which 31 had average genome-wide coverage of more than 5-fold, which were taken forward for further analyses. Of these 31, 23 were wild isolates sourced from Candeias dos Jamari (State of Rondônia) and eight were colony samples originally

collected from Porto Velho (State of Rondônia) in 2016. The number of sequencing reads generated ranged from 3,553,670 to 23,369,013, resulting in an average genome-wide coverage across of >8.5-fold (standard deviation of 2.7-fold). For these 31 samples, 22,481,832 SNPs were called across three chromosomes, reducing to 16,749,254 after quality control filtering.

An. darlingi colony samples are genetically distinct from State of Rondônia field populations

Using the 16.75 million high quality SNPs, a population and ancestral structure analysis demonstrated limited genetic diversity within the colony (n=8) and wildtype (n=23) populations, but clear differentiation between them. An admixture analysis indicated only two ancestral groups (K=2), separating clearly the two populations analysed **(Figure 1**). The colony samples have only K1 ancestry, and State of Rondônia field isolates have K2 ancestry except for one sample (AnDar989) that had both ancestries present. This sample also appeared to cluster closely to colony samples in a neighbour-joining (NJ) tree. The principal component analysis plot indicates greater genetic diversity within the State of Rondônia samples, than within the colony population (**Figure 1**). The colony isolates cluster closely together, except for sample AnDar1382, which on the NJ tree appears as its own clade, closer to AnDar989 (**Figure 1**).

Low nucleotide diversity ($\pi < 0.012$) was observed across the chromosomes, especially in chromosome 1, further demonstrating the limited genetic diversity across the *An. darlingi* genome (**Figure 2**). Similarly, a population differential F_{ST} analysis revealed evidence of allele frequency differences between the two populations (**Supp Figure 1**), with a total of 55,586 (<0.3%) SNPs with an F_{ST} of > 0.8, and 7,829 SNPs with an F_{ST} of 1 (**Supp Table 1, Supp Figure 1**). In a window based F_{ST} analysis (sizes: 1kb), two regions on chromosome two appear to have high population differentiation (window average $F_{ST} > 0.2$). One spans an intronic region (chr 2: 1897001-1898000) and the other encodes the *monocarboxylate transporter 7* (*MCT7*) (chr 2: 2070001-2071000).

No known insecticide resistance associated SNPs occur within these populations

A total of 133 genes of interest were initially identified as potentially involved in insecticide resistance. These included genes where previously reported SNPs had resulted in reduced insecticide efficacy, including *vgsc, ace-1,* all *gaba*, cytochrome P450s, carboxylesterases, glutathione transferases, and glucornyltransferases. A total of 35,418 SNPs were identified in these genes, but no known SNPs resulting in insecticide resistance in other *Anopheles* species were found. Within the four main genes associated with resistance (*ace-1, gaba, GSTe2,* and *vgsc*), 9,379 SNPs were identified, of which 58 were missense mutations (**Table 1, Supp Table 2**).

Several candidate genes are under selection

To infer any loci under recent positive selection, both the single population integrated haplotype (iHS) and across-population extended haplotype homozygosity (XP-EHH) statistics were calculated. Evidence of selection (absolute iHS values > 4) was found in the field (1,976 SNPs) and colony samples (207 SNPs) (**Supp Figure 2a**). Of the 723 genes and 913 intergenic regions with significant iHS values, four were included in genes potentially involved in insecticide resistance (glutamate gated chloride channel (colony), *CYP307a1* (field), *CYP4c1* (field), and *CYP4c3* (colony)). Substantially fewer positions were identified in the across population analysis, with only 26 positions identified over 8 genes and 12 intergenic regions (abs. XP-EHH >4) (**Supp Figure 2b**). None of the genes had been previously identified as involved in insecticide resistance in *An. darlingi* or other insects. Of the 26 positions, 14 had negative values suggesting they were under positive selection in the State of Rondônia field isolates, and 12 had positive XP-EHH scores suggesting they were under positive selection in the Colony population.

Linkage disequilibrium (LD; pairwise r^2) plots by physical distance were generated to investigate long-range haplotypes across chromosomes. LD decayed rapidly within 10kb as expected across the chromosomes for both populations, indicating a lack of long-range haplotypes or selection (**Figure 2**). The plots were also generated for the four genes originally known to be involved in insecticide resistance (*ace-1, GSTe2, gaba* and *vgsc*), and the four genes further identified with significant iHS values. No significant long-term

association was observed for any of these eight genes. To further investigate signatures of selection, Tajima's D (T_D) statistic was calculated in 100kb windows across the chromosomes for each population. The average T_D for the colony and field isolates were 1.00 and -0.42, respectively (**Supp Figure 3**). For the colony samples, more positive values were indicative of balancing selection or a recent population bottleneck. The latter is more likely as these mosquitoes were collected from the field and then reared in the laboratory after several generations. For the field isolates, it appears to be weak ongoing balancing selection and an excess of low frequency polymorphisms, indicating population size expansion. No loci were identified as significant (T_D of > 2.5 or < -2.5) for the State of Rondônia isolates. For the colony isolates, a total of 35 loci, covering 220 genes, had a T_D value > 2.5. Seven of these occurred on chromosome 2 and included candidate gene *glutathione-S transferase 1* (*GST1*). A further 28 windows had a T_D value > 2.5 on chromosome three, including one acetylcholinesterase-like (*ace-like*) gene, distinct from the *ace-1* loci associated with carbamate and organophosphate resistance. For these regions, sliding window analysis was repeated in 1kb windows and resulted in high T_D scores for *GST1* (2.73) and *ace-like (*2.60).

Discussion

This study presents whole genome sequence data from two *An. darlingi* populations sourced from the State of Rondônia (field and colony), Brazil. We have identified putatively novel missense SNPs, distinct population structure between field and colony samples, and signatures of directional selection in genes associated with insecticide resistance. We identified limited genetic diversity within these populations, which was to be expected as the collection sites for each population was in a limited geographic radius. The population structure analysis demonstrated a lack of gene flow between these populations, with them appearing as two distinct ancestries. This was to be expected as one population were colony isolates that likely had extensive inbreeding. However, this does demonstrate admixture's ability to accurately distinguish these ancestries, and thus its utility as part of this methodology to examine the ancestral groups present in larger more diverse analysis of *An. darlingi* WGS data. The PCA plot and phylogenetic tree also reflected this. Previous studies examining the genetic diversity of *An. darlingi* have used individual genes such as *cox-1* and found geographical distance is a great contributor of genetic distance, with physical

barriers such as the Andes Mountain ranges or the Amazon River have also been hypothesised to prevent population mixing[35–38]. Expanding from singular genes to whole genome will help further illuminate the population dynamics of this behaviourally complex species. The population differentiation (F_{ST}) analysis supported this with >55,000 SNPs having differences in allele frequency (F_{ST} > 0.8) between the populations[39]. An *MCT* gene with high Fst was identified in the windowed analysis, *MCT's* are involved in cellular metabolism, but have not been linked to insecticide metabolism[40]. Similarly, calculation of Tajima's D statistic with sliding window analysis, indicated two candidate genes were undergoing balancing selection, including *ace*-like and *GST1* genes. This *ace* gene varies significantly from *ace-1,* which has been associated with carbamate and organophosphate resistance[41,42]. Glutathione S-transferases (e.g., *GST1*) are known to confer resistance to organophosphates and DDT. Class 1 GSTs confer this resistance in *M. domestica* and are believed to do the same in *Anopheles spp*, although further supporting evidence is required[43–45].

The lack of known insecticide resistance associated missense SNPs identified in this study is not unexpected, as no previous studies have identified any such markers[21,23,24]. It is hypothesised that *An. darlingi* relies more on metabolic mechanisms of detoxification to reduce the efficacy of insecticides[33]. Three of the four candidate genes identified as under selection with the iHS metric, were part of the cytochrome P450 superfamily of genes involved in xenobiotic detoxification (e.g., insecticides). These genes, *CYP307a1, CYP4c1*, and *CYP4c3,* have not been formally linked to insecticide resistance. Previously *CYP307a1* has been found to be overexpressed in pyrethroid and DDT resistant *An. funestus,* and deltamethrin (a pyrethroid) resistant *Cydia pomonella*[46–48]. Functionally, *CYP307a1* is involved in ecdysteroid hormone synthesis, and these hormones regulate insect behaviour, nervous system development, and reproduction[49]. The remaining two genes *CYP4c1* and *CYP4c3*, are both members of the *CYP4* clan, which have been linked to insecticide resistance[50,51]. They were both found to be upregulated in DDT, pyrethroid and carbamate resistant *Ae. aegypti* and highly expressed in deltamethrin resistant *C. pomonella* larvae[48,52]. *CYP4c1* is believed to be involved in metabolism compensation in periods of starvation in cockroaches and is hormonally regulated[53,54]. It was also found to be up regulated in

240 *Leptinotarsa decemlineata* (a crop pest), after exposure to neonicotinoids[55]. Even less is known about the
241 function of *CYP4c3*, but it was reported to be upregulated in honeybees after amitraz treatment[56]. Amitraz
242 is a formamidine, a pesticide and insecticide class used to eliminate ticks and mites. To further understand
243 the potential roles of these genes in insecticide resistance, it is necessary to have phenotypic data, which
244 was not available for the samples analysed here. Comparative RNA-seq analysis on insecticide resistant
245 and susceptible populations can also support to identify genes highly expressed in resistant populations,
246 and thereby involved in metabolic-based insecticide resistance.
247 There may be other reasons for known molecular markers of resistance not being identified. It's not
248 implausible that the mosquitoes analysed here are completely susceptible to insecticides and this why no
249 insecticide resistance markers were found. As mentioned above, resistance data for *An. darlingi* is
250 extremely limited, with no reports in a >1000km radius of the collection sites for these isolates[16]. Making
251 it difficult to hypothesise about the phenotype of the isolates analysed here. Typical vector control in Brazil
252 focusses on *Aedes spp.,* so fogging and space spraying with organophosphates and pyrethroids are
253 commonly conducted in urban areas[57–59]. *An. darlingi* has been reported in peri-urban environments and
254 has a preference to rest indoors so could theoretically have been exposed to insecticides as part of
255 arbovirus control methods[1]. However, Candeias dos Jamari, has little to no insecticide use, so the isolates
256 collected there were likely susceptible. The colony isolates originated from Porto Velho in 2016, where
257 insecticides were in use, so the chances of those isolates being phenotypically resistant are higher. But
258 exposure to insecticides cannot be assumed to result in a resistant phenotype. Another reason known
259 target site mutations could be missing from this data set is *An. darlingi* mosquitos maybe similar to *An.*
260 *funestus,* which lacks the classic *kdr* mutation[60–64]. Extensive investigation of *An. funestus* has led to the
261 identification of other molecular markers of resistance, and the same is needed for *An. darlingi*[31,65,66].
262 Further WGS work with higher read depth could allow for the identification of copy number variants that
263 also play a role in the development of resistance[41,67]. The 58 missense SNPs identified across the four genes
264 (*ace-1, gaba, GSTe2,* and *vgsc*) commonly associated with decreased insecticide susceptibility should also
265 be examined further. The lack of phenotypic resistance data available for these isolates limits any
266 conclusions that can be drawn from this data, but these SNPs offer opportunities for further studies to

identify markers of resistance for *An. darlingi.* These mutations could be included as part of a high-throughput molecular assays to screen larger size vector populations studies and support the identification of the molecular mechanism underpinning resistance in this species

In summary, more work is needed to fully understand the population dynamics of *An. darlingi* and uncover markers associated with resistance*,* but this study represents the first steps towards this. Our study is the first application of WGS to *An. darlingi and* provides the blueprint for future studies with more expansive sample sizes and greater field population representation. Large scale surveillance of *An. darlingi* is needed to understand its molecular ecology and create high-throughput assays to inform control programmes and work towards malaria elimination.

Methods

Mosquito Collection, Species Identification, and DNA extraction

Two populations of *An. darlingi* from the State of Rondônia were used in this study, one field population from Candeias do Jamari, (n=28), and a cohort of colony isolates from Porto Velho (n=12)[68]. The field isolates were collected during vector density studies in malaria endemic regions in Rondônia in 2018-19, while the colony was initiated in 2018[68–70]. The geographical distance between the colony isolates initial collection point and the field isolate collection point is < 25km. Specimens were morphologically identified via microscopy. Genomic DNA from the isolates were extracted from the whole mosquito using Qiagen DNaeasy Blood and Tissue kit (Qiagen, Hilden, Germany), according to manufacturer's instructions. DNA was quantified on the Qubit 2.0 fluorimeter HS DNA kit (Thermofisher), and stored at -20^{o}C

Whole Genome Sequencing, and Bioinformatics Analysis

A total of 40 *An. darlingi* isolates were sequenced using the Illumina MiSeq on 2x250bp paired end configuration. There was no WGS data publicly available for *An. darling*i to incorporate into analysis. The raw paired fastq files were trimmed using trimmomatic software (version 0.39), and then aligned using bwa-mem software to the AnoDar_H01 (*An. darlingi)* reference genome using default parameters[71,72].

Genome coverage from the mapped bam files was calculated using samtools, and variants called and validated using HaplotypeCaller and VailidateVariants from GATK software respectively[73,74]. Once VCFs had been generated for each sample, GATK was used to create a multi-sample VCF with GenomicsDBImport and GenotypeGVCF functions. This multi-sample VCF was then filtered to contain only chromosomal variants, sorted, and normalised to remove multi-allelic sites using bcftools. Further filtering was conducted to remove variants with low depth (<5), and high sample and site missingness with vcftools[75].

Population Genetic Analysis

From the filtered multi-sample VCF, a pairwise-genetic distance matrix was generated using an in-house script[76]. This matrix was used for the generation of a neighbour joining (NJ) tree, and principal component analysis using R packages ape and qqman[77,78]. The NJ tree generated was annotated and visualised in iTOL[79]. Ancestry admixture analysis was conducted using ADMIXTURE software (version 1.3)[80]. The multi-sample VCF was converted to a bed file using PLINK, and the optimum K value (estimated number of ancestral populations) was calculated by cross-validation of 1-10 eigenvalue decay dimensions[81]. In this instance, K of 2 was estimated, and this value along with the bed file was used by the ADMIXTURE software to analyse the shared ancestral populations in these samples. The output was then visualised in R using ggplot2 package[82].

Genomic regions under selection were identified with the R package rehh[83]. The Integrated Haplotype Statistic (iHS) was used to find directional selection within populations, and extended haplotype homozygosity (XP-EHH) between the two populations[84,85]. The Tajima's D (T_D) statistic was also calculated using vcftools in 100kb windows across chromosomes to identify any balancing selection that may be occurring[75,86]. Similarly, nucleotide diversity (π) was calculated in 100kb windows across chromosomes using vcftools. These three statistics were then plotted in R. The Weir and Cockerham's F statistic (F_{ST}) was estimated per SNP between the field and colony samples using vcftools[75,87]. F_{ST} was plotted by SNP and averaged over 100kb windows using python. The outputs were then visualised in R. Linkage disequilibrium (LD) was estimated using the pairwise r^2 statistic, and visualised in two ways: (i) LD decay across

chromosomes, and in candidate genes, using PopLDdecay software[88]; (ii) heatmaps for candidate genes visualised in the R package Gaston[89].

Examination of Insecticide Resistance Associated Genes

A bed file of genes commonly associated with target-site insecticide resistance SNPs (e.g., *para, gaba,* and *ace-1*) was created. All annotated cytochrome P450's and esterase's linked to metabolically mediated insecticide resistance were included. The bed file was then applied to the multi-sample VCF using bcftools view function. The snpEff software was then used to annotate the effect each SNP would have on the protein's amino acid sequence[90]. A custom-built database was created for this, using the AnoDar_H01 GFF file.

Data Availability

All raw data used in this work is publicly available (see PRJEB66076 for accession numbers).

Acknowledgements

We thank for the members of the PIVEM/Fiocruz Rondônia for *An. darlingi* culture. JFM also gratefully acknowledge their CNPq productivity grant (process number: 304830/2022-4).

Funding

This work was supported, in part, by the Brazilian Ministry of Health/DECIT/CNPq N° 23/2019 (grant number 442653/2019-0) and the Bill & Melinda Gates Foundation (INV-003970). Under the grant conditions of the foundation, a Creative Commons Attribution 4.0 Generic License has already been assigned to the author accepted manuscript version that might arise from this submission. The study also received funding from the International Centres of Excellence for Malaria Research (ICEMR) program GR109237 (CON-80002357). H.A.P. is funded by a BBSRC LIDo-DTP PhD studentship. J.P. was funded by a Newton Institutional Links Grant (British Council, no. 261868591). T.G.C. and S.C. were funded by UKRI grants (ref. BB/X018156/1, MR/X005895/1). The authors declare no conflicts of interest. The funders had

no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript should be declared.

Authors contributions

TGC and SC conceived and directed the project. HA-P performed bioinformatic and statistical analyses under the supervision of TGC and SC. XX provided resources. HA-P, TGC and SC interpreted results. HA-P wrote the first draft of the manuscript with inputs from XXX. All authors commented and edited on various versions of the draft manuscript and approved the final version. HA-P compiled the final manuscript.

Additional Information

Ethics approval and consent

No ethics approvals were required as all data is publicly available.

Consent for publication

All authors have consented to the publication of this manuscript.

Availability of data and materials

Analysis scripts are available at https://github.com/LSHTMPathogenSeqLab

Competing interests

Authors declare no competing interests.

References

1. Hiwat, H. & Bretas, G. Ecology of Anopheles darlingi Root with respect to vector importance: a review. *Parasit. Vectors* **4**, 177 (2011).

2. Klein, T. A., Lima, J. B. P. & Tada, M. S. Comparative Susceptibility of Anopheline Mosquitoes to Plasmodium Falciparum in Rondonia, Brazil. *Am. J. Trop. Med. Hyg.* **44**, 598–603 (1991).

3. Gil, L. H. S., Tada, M. S., Katsuragawa, T. H., Ribolla, P. E. M. & Silva, L. H. P. da. Urban and suburban malaria in Rondônia (Brazilian Western Amazon) II: perennial transmissions with high anopheline densities are associated with human environmental changes. *Mem. Inst. Oswaldo Cruz* **102**, 271–276 (2007).

4. Charlwood, J. D. Biological variation in Anopheles darlingi root. *Mem. Inst. Oswaldo Cruz* **91**, 391–398 (1996).

5. Da Silva-Nunes, M. *et al.* Malaria on the amazonian frontier: Transmission dynamics, risk factors, spatial distribution, and prospects for control. *Am. J. Trop. Med. Hyg.* **79**, 624–635 (2008).

6. Angêlla, A. F. *et al.* Seasonal genetic partitioning in the neotropical malaria vector, Anopheles darlingi. *Malar. J.* **13**, 1–10 (2014).

7. Moutinho, P. R., Gil, L. H. S., Cruz, R. B. & Ribolla, P. E. M. Population dynamics, structure and behavior of Anopheles darlingi in a rural settlement in the Amazon rainforest of Acre, Brazil. *Malar. J.* **10**, 1–13 (2011).

8. Global Malaria Programme: WHO Global. *WHO World Malaria Report 2022*. *WHO*.

9. Carlos, B. C., Rona, L. D. P., Christophides, G. K. & Souza-Neto, J. A. A comprehensive analysis of malaria transmission in Brazil. *Pathog. Glob. Health* **113**, 1–13 (2019).

10. Laporta, G. Z., Grillet, M. E., Rodovalho, S. R., Massad, E. & Sallum, M. A. M. Reaching the malaria elimination goal in Brazil: a spatial analysis and time-series study. *Infect. Dis. Poverty* **11**, 39 (2022).

11. Torres, K. *et al.* Malaria Resilience in South America: Epidemiology, Vector Biology, and Immunology Insights from the Amazonian International Center of Excellence in Malaria Research Network in Peru and Brazil. *Am. J. Trop. Med. Hyg.* **107**, 168–181 (2022).

12. Rocha, E. M. *et al.* Vector-Focused Approaches to Curb Malaria Transmission in the Brazilian Amazon: An Overview of Current and Future Challenges and Strategies. *Trop. Med. Infect. Dis.* **5**, 161 (2020).

13. Baia-da-Silva, D. C. *et al.* Current vector control challenges in the fight against malaria in Brazil. *Rev. Soc. Bras. Med. Trop.* **52**, e20180542 (2019).

14. Campos, K. B. *et al.* Assessment of the susceptibility status of Aedes aegypti (Diptera: Culicidae) populations to pyriproxyfen and malathion in a nation-wide monitoring of insecticide resistance performed in Brazil from 2017 to 2018. *Parasit. Vectors* **13**, 531 (2020).

15. Leandro, A. de S. *et al.* Malathion insecticide resistance in Aedes aegypti: laboratory conditions and in situ experimental approach through adult entomological surveillance. *Trop. Med. Int. Health* **25**, 1271–1282 (2020).

16. WHO. Global Malaria Programme: Malaria Threats Map. https://www.who.int/teams/global-malaria-programme/surveillance/malaria-threats-map.

17. World Health Organisation. *Global Report on Insecticide Resistance in Malaria Vectors: 2010-2016 Global Malaria Programme*.

18. Namias, A., Jobe, N. B., Paaijmans, K. P. & Huijben, S. The need for practical insecticide-resistance guidelines to effectively inform mosquito-borne disease control programs. *eLife* **10**, e65655.

19. Organização Mundial da Saúde. *Guidelines for Malaria Vector Control*. *Guidelines for Malaria Vector Control* (2019).

20. Lol, J. C. *et al.* Development of molecular assays to detect target-site mechanisms associated with insecticide resistance in malaria vectors from Latin America. *Malar. J.* **18**, 202 (2019).

21. Orjuela, L. I. *et al.* Absence of knockdown mutations in pyrethroid and DDT resistant populations of the main malaria vectors in Colombia. *Malar. J.* **18**, 1–9 (2019).

22. Application of a targeted amplicon sequencing panel to screen for insecticide resistance mutations in Anopheles darlingi populations from Brazil. https://www.researchsquare.com (2023) doi:10.21203/rs.3.rs-3053716/v1.

23. Loureiro, A. C. *et al.* Molecular diversity of genes related to biological rhythms (period and timeless) and insecticide resistance NaV and ace-1in Anopheles darlingi. *Mem. Inst. Oswaldo Cruz* **118**, e220159 (2023).

24. Bickersmith, S. A. *et al.* Mutations Linked to Insecticide Resistance Not Detected in the Ace-1 or VGSC Genes in Nyssorhynchus darlingi from Multiple Localities in Amazonian Brazil and Peru. *Genes* **14**, (2023).

25. 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).

26. 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).

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

28. 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 Biol.* **15**, R27 (2014).

29. Yang, C., Huang, Z., Li, M., Feng, X. & Qiu, X. RDL mutations predict multiple insecticide resistance in Anopheles sinensis in Guangxi, China. *Malar. J.* **16**, 482 (2017).

30. Cariño, F. A., Koener, J. F., Plapp, F. W. & Feyereisen, R. Constitutive overexpression of the cytochrome P450 gene CYP6A1 in a house fly strain with metabolic resistance to insecticides. *Insect Biochem. Mol. Biol.* **24**, 411–418 (1994).

31. Ibrahim, S. S., Ndula, M., Riveron, J. M., Irving, H. & Wondji, C. S. The P450 CYP6Z1 confers carbamate/pyrethroid cross-resistance in a major African malaria vector beside a novel carbamate-insensitive N485I acetylcholinesterase-1 mutation. *Mol. Ecol.* **25**, 3436–3452 (2016).

32. 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).

33. Fonseca-González, I., Quiñones, M. L., McAllister, J. & Brogdon, W. G. Mixed-function oxidases and esterases associated with cross-r esistance between DDT and lambda-cyhalothrin in Anopheles darlingi Root 1926 populations from Colombia. *Mem. Inst. Oswaldo Cruz* **104**, 18–26 (2009).

34. Riveron, J. M. *et al.* Directionally selected cytochrome P450 alleles are driving the spread of pyrethroid resistance in the major malaria vector Anopheles funestus. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 252–257 (2013).

35. Angêlla, A. F., Gil, L. H. S., Da Silva, L. H. P. & Ribolla, P. E. M. Population structure of the malaria vector Anopheles darlingi in Rondônia, Brazilian Amazon, based on mitochondrial DNA. *Mem. Inst. Oswaldo Cruz* **102**, 953–958 (2007).

36. Mirabello, L. & Conn, J. E. Molecular population genetics of the malaria vector Anopheles darlingi in Central and South America. *Heredity* **96**, 311–321 (2006).

37. Emerson, K. J., Conn, J. E., Bergo, E. S., Randel, M. A. & Sallum, M. A. M. Brazilian Anopheles darlingi Root (Diptera: Culicidae) Clusters by Major Biogeographical Region. *PLOS ONE* **10**, e0130773 (2015).

38. Pedro, P. M. & Sallum, M. A. M. Spatial expansion and population structure of the neotropical malaria vector, Anopheles darlingi (Diptera: Culicidae). *Biol. J. Linn. Soc.* **97**, 854–866 (2009).

39. Wright, S. *Evolution and the Genetics of Populations, Volume 4: Variability Within and Among Natural Populations*. (University of Chicago Press, 1984).

40. Liu, Q., Dou, S., Wang, G., Li, Z. & Feng, Y. Evolution and functional divergence of monocarboxylate transporter genes in vertebrates. *Gene* **423**, 14–22 (2008).

41. Djogbénou, L., Labbé, P., Chandre, F., Pasteur, N. & Weill, M. Ace-1 duplication in Anopheles gambiae: a challenge for malaria control. *Malar. J.* **8**, 70 (2009).

42. Assogba, B. S. *et al.* An ace-1 gene duplication resorbs the fitness cost associated with resistance in Anopheles gambiae, the main malaria mosquito. *Sci. Rep.* **5**, 14529 (2015).

43. Ranson, H. *et al.* Cloning and Localization of a Glutathione S-transferase Class I Gene from Anopheles gambiae*. *J. Biol. Chem.* **272**, 5464–5468 (1997).

44. Fournier, D., Bride, J. M., Poirie, M., Bergé, J. B. & Plapp, F. W. Insect glutathione S-transferases. Biochemical characteristics of the major forms from houseflies susceptible and resistant to insecticides. *J. Biol. Chem.* **267**, 1840–1845 (1992).

45. Pasteur, N. & Raymond, M. Insecticide Resistance Genes in Mosquitoes: Their Mutations, Migration, and Selection in Field Populations. *J. Hered.* **87**, 444–449 (1996).

46. Sandeu, M. M., Mulamba, C., Weedall, G. D. & Wondji, C. S. A differential expression of pyrethroid resistance genes in the malaria vector Anopheles funestus across Uganda is associated with patterns of gene flow. *PLoS ONE* **15**, e0240743 (2020).

47. Riveron, J. M. *et al.* Genome-Wide Transcription and Functional Analyses Reveal Heterogeneous Molecular Mechanisms Driving Pyrethroids Resistance in the Major Malaria Vector Anopheles funestus Across Africa. *G3 GenesGenomesGenetics* **7**, 1819–1832 (2017).

48. Insects | Free Full-Text | Genome-Wide Selective Signature Analysis Revealed Insecticide Resistance Mechanisms in Cydia pomonella. https://www.mdpi.com/2075-4450/13/1/2.

49. Pondeville, E. *et al.* Microarray and RNAi Analysis of P450s in Anopheles gambiae Male and Female Steroidogenic Tissues: CYP307A1 Is Required for Ecdysteroid Synthesis. *PLOS ONE* **8**, e79861 (2013).

50. Dulbecco, A. B., Calderón-Fernández, G. M. & Pedrini, N. Cytochrome P450 Genes of the CYP4 Clan and Pyrethroid Resistance in Chagas Disease Vectors. *Front. Trop. Dis.* **3**, (2022).

492 51. Killiny, N., Hajeri, S., Tiwari, S., Gowda, S. & Stelinski, L. L. Double-Stranded RNA Uptake through Topical
493 Application, Mediates Silencing of Five CYP4 Genes and Suppresses Insecticide Resistance in Diaphorina
494 citri. *PLOS ONE* **9**, e110536 (2014).

495 52. Lien, N. T. K. *et al.* Transcriptome Sequencing and Analysis of Changes Associated with Insecticide
496 Resistance in the Dengue Mosquito (Aedes aegypti) in Vietnam. *Am. J. Trop. Med. Hyg.* **100**, 1240–1248
497 (2019).

498 53. Lu, K.-H., Bradfield, J. Y. & Keeley, L. L. Juvenile hormone inhibition of gene expression for cytochrome
499 P4504C1 in adult females of the cockroach, Blaberus discoidalis. *Insect Biochem. Mol. Biol.* **29**, 667–673
500 (1999).

501 54. Cytochrome P450 family 4 in a cockroach: molecular cloning and regulation by regulation by
502 hypertrehalosemic hormone. https://www.pnas.org/doi/10.1073/pnas.88.10.4558
503 doi:10.1073/pnas.88.10.4558.

504 55. Bouafoura, R. *et al.* Characterization of Insecticide Response-Associated Transcripts in the Colorado
505 Potato Beetle: Relevance of Selected Cytochrome P450s and Clothianidin. *Insects* **13**, 505 (2022).

506 56. Ye, L. *et al.* Transcriptomic analysis to elucidate the response of honeybees (Hymenoptera: Apidae) to
507 amitraz treatment. *PLOS ONE* **15**, e0228933 (2020).

508 57. Garcia, G. de A. *et al.* The impact of insecticide applications on the dynamics of resistance: The case of
509 four Aedes aegypti populations from different Brazilian regions. *PLoS Negl. Trop. Dis.* **12**, e0006227
510 (2018).

511 58. Melo Costa, M. *et al.* Kdr genotyping in Aedes aegypti from Brazil on a nation-wide scale from 2017 to
512 2018. *Sci. Rep.* **10**, 13267 (2020).

513 59. Lima, E. P. *et al.* Insecticide resistance in Aedes aegypti populations from Ceará, Brazil. *Parasit. Vectors*
514 **4**, 5 (2011).

515 60. Coetzee, M. & Koekemoer, L. L. Molecular Systematics and Insecticide Resistance in the Major African
516 Malaria Vector Anopheles funestus. *Annu. Rev. Entomol.* **58**, 393–412 (2013).

61. Kamgang, B. *et al.* Exploring insecticide resistance mechanisms in three major malaria vectors from Bangui in Central African Republic. *Pathog. Glob. Health* **112**, 349–359 (2018).

62. Tchigossou, G. *et al.* Molecular basis of permethrin and DDT resistance in an Anopheles funestus population from Benin. *Parasit. Vectors* **11**, 602 (2018).

63. Irving, H. & Wondji, C. S. Investigating knockdown resistance (kdr) mechanism against pyrethroids/DDT in the malaria vector Anopheles funestus across Africa. *BMC Genet.* **18**, 76 (2017).

64. Morgan, J. C., Irving, H., Okedi, L. M., Steven, A. & Wondji, C. S. Pyrethroid Resistance in an Anopheles funestus Population from Uganda. *PLOS ONE* **5**, e11872 (2010).

65. Mugenzi, L. M. J. *et al.* Cis-regulatory CYP6P9b P450 variants associated with loss of insecticide-treated bed net efficacy against Anopheles funestus. *Nat. Commun.* **10**, 4652 (2019).

66. Weedall, G. D. *et al.* A cytochrome P450 allele confers pyrethroid resistance on a major African malaria vector, reducing insecticide-treated bednet efficacy. *Sci. Transl. Med.* **11**, eaat7386 (2019).

67. Mugenzi, L. M. J. *et al.* The duplicated P450s CYP6P9a/b drive carbamates and pyrethroids cross-resistance in the major African malaria vector Anopheles funestus. *PLOS Genet.* **19**, e1010678 (2023).

68. Araujo, M. da S. *et al.* Brazil's first free-mating laboratory colony of *Nyssorhynchus darlingi*. *Rev. Soc. Bras. Med. Trop.* **52**, e20190159 (2019).

69. Andrade, A. O. *et al.* Description of malaria vectors (Diptera: Culicidae) in two agricultural settlements in the Western Brazilian Amazon. *Rev. Inst. Med. Trop. São Paulo* **63**, e60 (2021).

70. Katsuragawa, T. H. *et al.* The Dynamics of Transmission and Spatial Distribution of Malaria in Riverside Areas of Porto Velho, Rondônia, in the Amazon Region of Brazil. *PLOS ONE* **5**, e9245 (2010).

71. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinforma. Oxf. Engl.* **30**, 2114–2120 (2014).

72. H.Li. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arxiv.org* **n/a**, n/a (2013).

73. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10**, 1–4 (2021).

74. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. (2012) doi:10.48550/arxiv.1207.3907.

75. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

76. fastq2matrix/scripts at master · LSHTMPathogenSeqLab/fastq2matrix. *GitHub* https://github.com/LSHTMPathogenSeqLab/fastq2matrix/tree/master/scripts.

77. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).

78. Turner, S. D. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *J. Open Source Softw.* **3**, 731 (2018).

79. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).

80. Alexander, D. H. & Lange, K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* **12**, 246 (2011).

81. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

82. Wickham, H. ggplot2. *ggplot2* (2009) doi:10.1007/978-0-387-98141-3.

83. rehh 2.0: a reimplementation of the R package rehh to detect positive selection from haplotype structure - Gautier - 2017 - Molecular Ecology Resources - Wiley Online Library. https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.12634.

84. Sabeti, P. C. *et al.* Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913–918 (2007).

85. Voight, B. F., Kudaravalli, S., Wen, X. & Pritchard, J. K. A Map of Recent Positive Selection in the Human Genome. *PLoS Biol.* **4**, e72 (2006).

565 86. Tajima, F. Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism.
566 *Genetics* **123**, 585–595 (1989).

567 87. Weir, B. S. & Cockerham, C. C. Estimating F-Statistics for the Analysis of Population Structure. *Evolution*
568 **38**, 1358–1370 (1984).

569 88. Zhang, C., Dong, S.-S., Xu, J.-Y., He, W.-M. & Yang, T.-L. PopLDdecay: a fast and effective tool for linkage
570 disequilibrium decay analysis based on variant call format files. *Bioinformatics* **35**, 1786–1788 (2019).

571 89. genostats/gaston. genostats (2022).

572 90. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide
573 polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly*
574 *(Austin)* **6**, 80–92 (2012).

575

576

Table 1. Number of missense SNPs identified in each key gene associated with insecticide resistance.

Gene	No. of missense SNPs identified
ace-1	14
GSTe2	26
gaba/rdl	2
vgsc	16

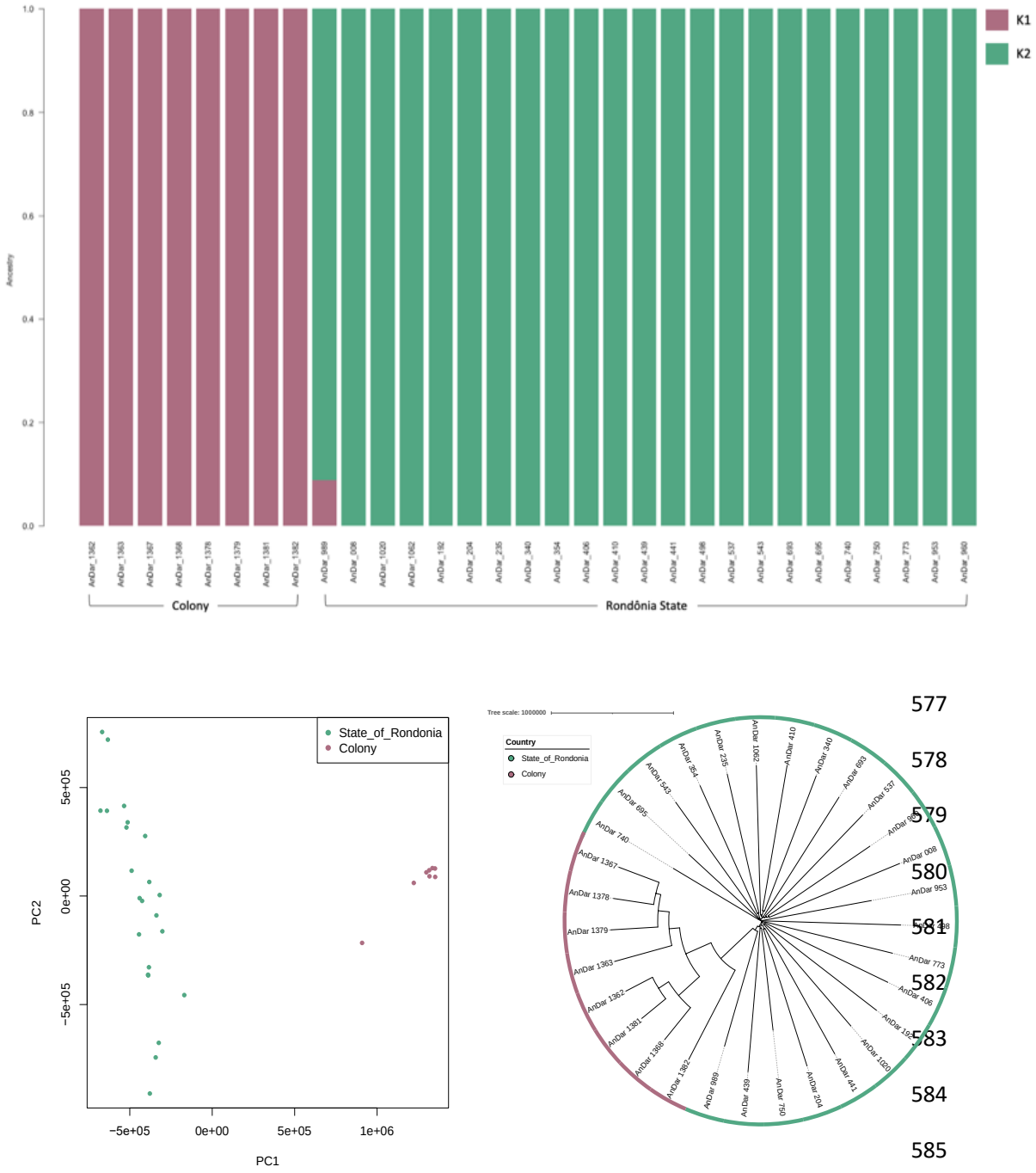


Figure 1. (TOP) Ancestry analysis of *An. darlingi* colony or wildtype (State of Rondônia) isolates. The number of ancestral populations (K) was estimated to be 2. (BOTTOM left) Principal Component Analysis plot (variation explained: PC1 17.5%, PC2 5.1%); (BOTTOM right) neighbourhood joining tree generated from a pairwise genetic distance matrix with 16,749,254 SNPs, demonstrating two clear population clusters for the Colony (n=8), and State of Rondônia field (n=23) isolates.

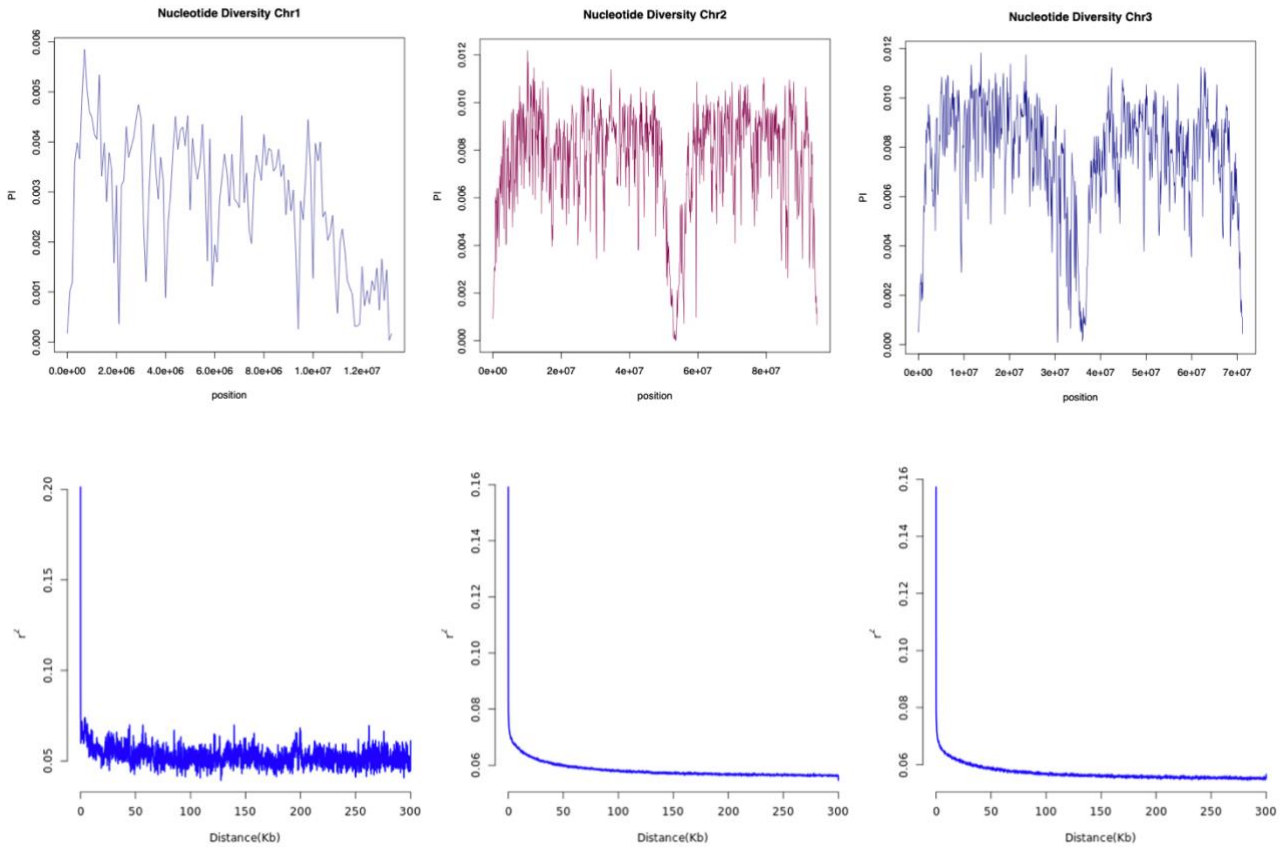
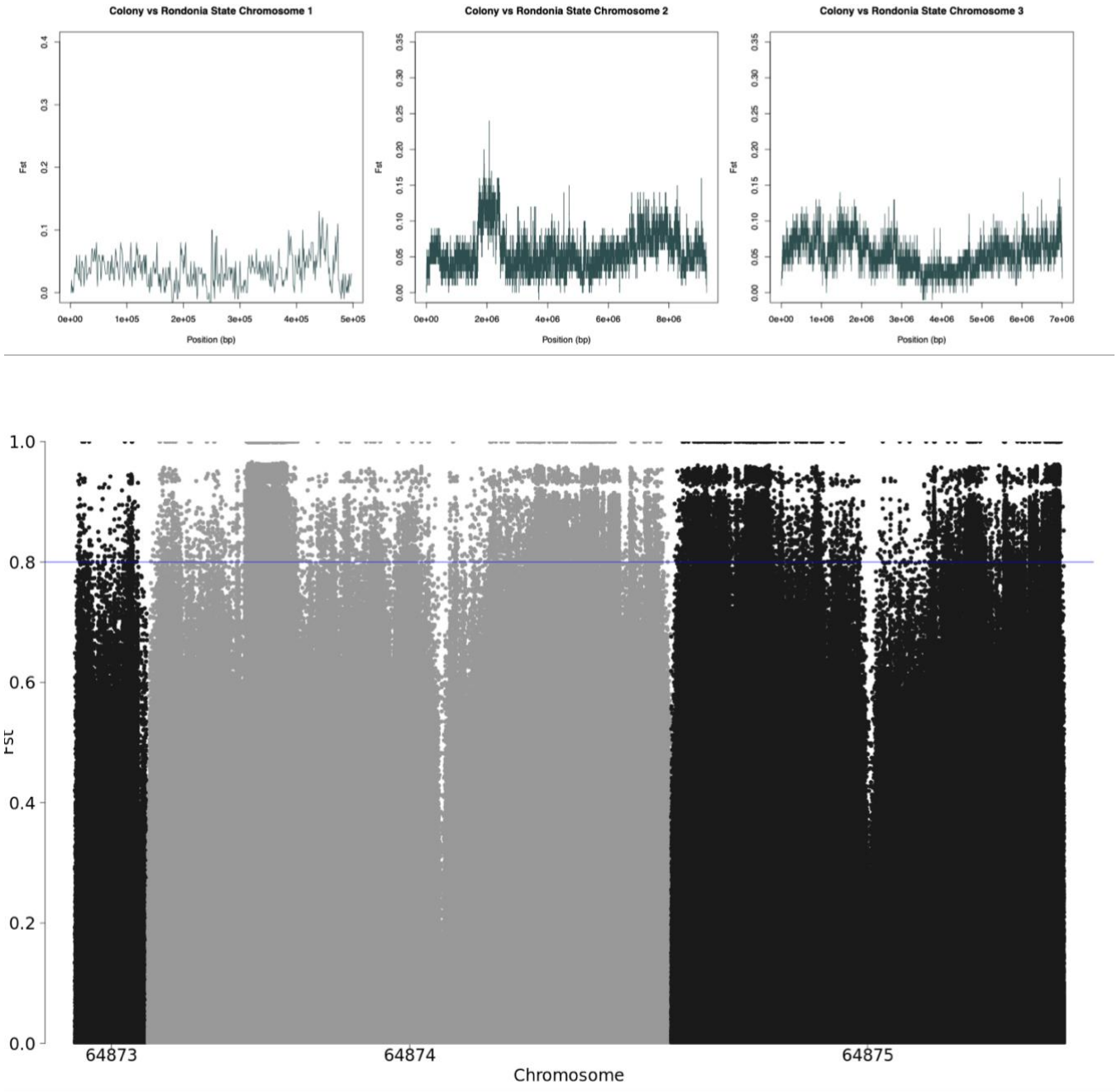
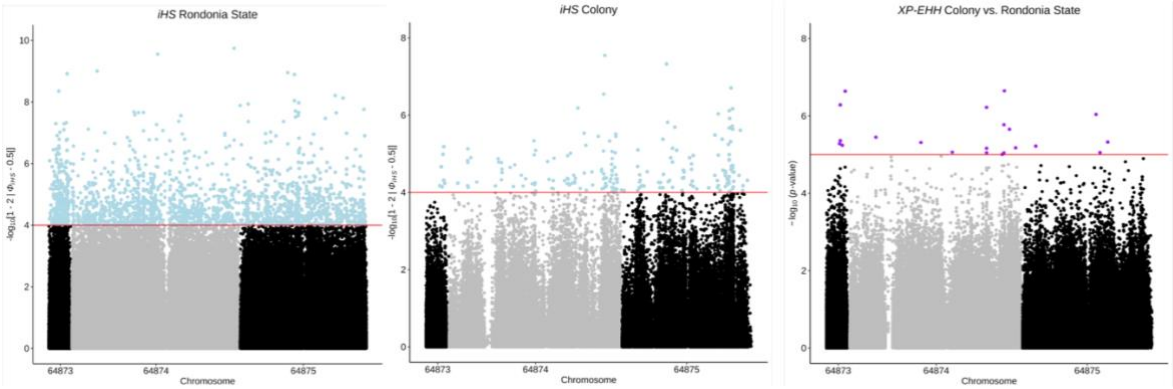


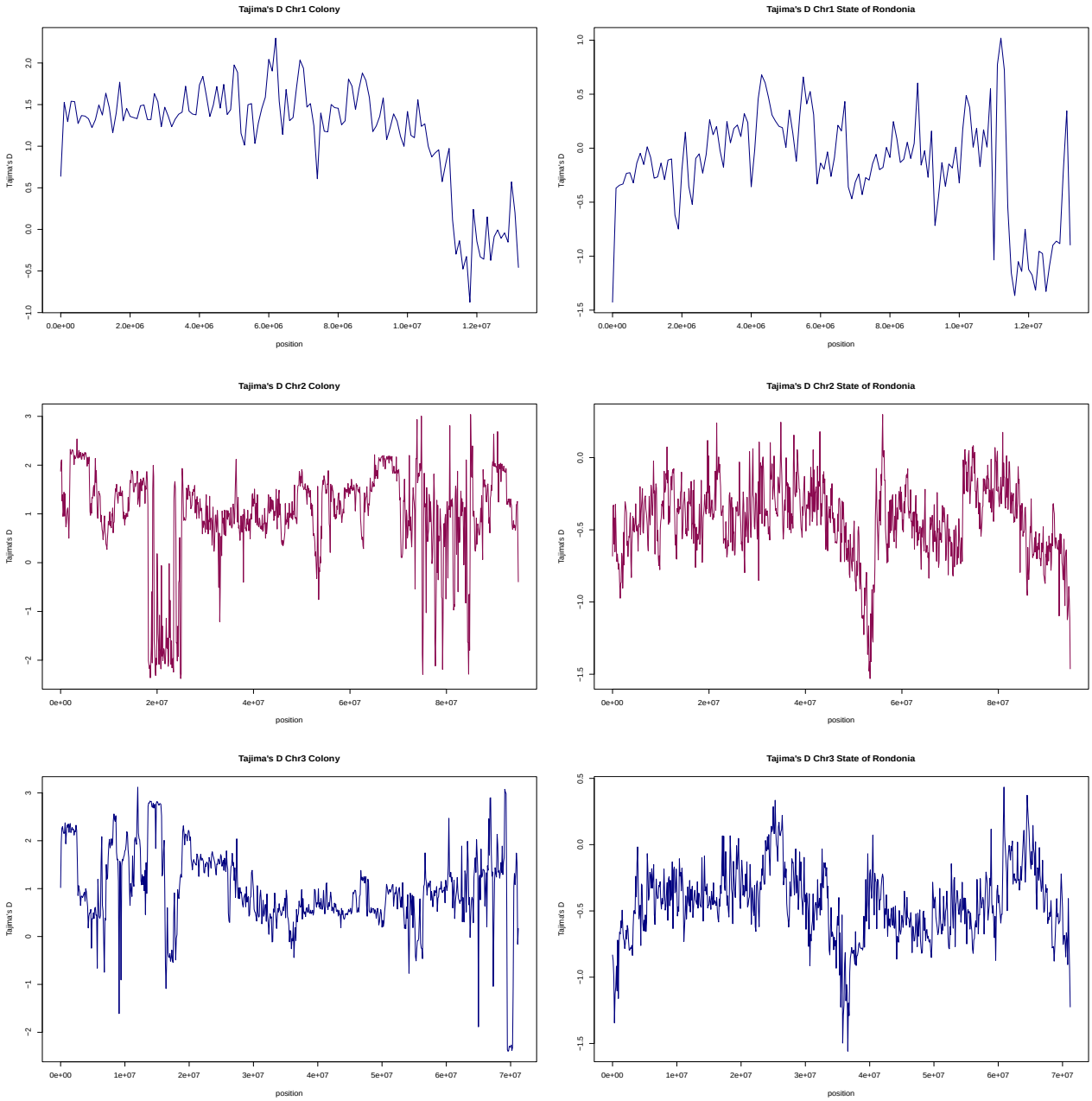
Figure 2. (TOP) Nucleotide diversity (π) for each chromosome with both populations (colony and field). 100kb windows were used and demonstrated low π (<0.012); (BOTTOM) Linkage disequilibrium (pairwise r^2) decay by distance for each chromosome, across both populations. Window sizes of 300kb are used.



Supplementary Figure A. (TOP) Population differentiation F_{ST} between the colony and field isolates. 100kb windows were used for each chromosome. (BOTTOM) Manhattan plot of population differentiation F_{ST} scores between colony and field samples for each filtered SNP identified across the three chromosomes.



Supplementary Figure B. Manhattan plots of SNPs identified as significant during scans for recent directional selection: A) Integrated Haplotype Score (iHS) for SNPs in State of Rondônia isolates; B) iHS for SNPs identified in Colony isolates; C) Extended Haplotype Heterozygosity (XP-EHH) cross population test for SNPs at fixation in one population but remaining polymorphic in the other.



Supplementary Figure C. Average Tajima's D statistic calculated in 100kb windows, across each chromosome and population.

Supplementary Table A. Number and percentage of SNPs with population differentiation F_{ST} values between colony and field samples at different thresholds.

F_{ST} value	No. SNPs	% SNPs
>0.15	2,670,704	15.94
>0.8	55,586	0.33
1.0	7,829	0.05

Supplementary Table B. Missense SNPs identified in four key genes associated with insecticide resistance.

Chromosome	Position	Gene	*An. darlingi* Amino acid change	*An. gambiae* Amino acid position	Reference species amino acid position*
NC_064874.1	15676189	*Ace-1*	L6H	n/a	n/a
NC_064874.1	15676201	*Ace-1*	G10A	n/a	n/a
NC_064874.1	15676222	*Ace-1*	D17G	n/a	n/a
NC_064874.1	15676267	*Ace-1*	V32A	n/a	n/a
NC_064874.1	15676390	*Ace-1*	G73V	n/a	n/a
NC_064874.1	15679129	*Ace-1*	T123S	Q8	n/a
NC_064874.1	15679222	*Ace-1*	G154S	n/a	n/a
NC_064874.1	15679232	*Ace-1*	V157G	n/a	n/a
NC_064874.1	15679234	*Ace-1*	G158S	n/a	n/a
NC_064874.1	15679235	*Ace-1*	G158V	n/a	n/a
NC_064874.1	15679465	*Ace-1*	A207T	n/a	V43
NC_064874.1	15679466	*Ace-1*	A207V	n/a	V43
NC_064874.1	15680528	*Ace-1*	T501S	I382	T338
NC_064874.1	15681124	*Ace-1*	S674N	n/a	S504
NC_064874.1	89825504	*GSTe2*	E33A	T33	-
NC_064874.1	89825711	*GSTe2*	N82D	D82	-
NC_064874.1	89825712	*GSTe2*	N82S	D82	-
NC_064874.1	89825747	*GSTe2*	A94T	A94	-
NC_064874.1	89825922	*GSTe2*	D128Q	D128	-
NC_064874.1	89825962	*GSTe2*	E143K	E143	-
NC_064874.1	89826035	*GSTe2*	T166I	I166	-
NC_064874.1	89826074	*GSTe2*	T179I	H179	-
NC_064874.1	89826151	*GSTe2*	T205A	n/a	-
NC_064874.1	89827133	*GSTe2*	L217S	n/a	-
NC_064874.1	89827263	*GSTe2*	Q260H	n/a	-
NC_064874.1	89827280	*GSTe2*	R266L	n/a	-
NC_064874.1	89827309	*GSTe2*	V276L	n/a	-
NC_064874.1	89827367	*GSTe2*	A295V	n/a	-
NC_064874.1	89827414	*GSTe2*	T311A	n/a	-
NC_064874.1	89827415	*GSTe2*	T311I	n/a	-
NC_064874.1	89827426	*GSTe2*	V315I	n/a	-
NC_064874.1	89827429	*GSTe2*	N316Y	n/a	-
NC_064874.1	89827430	*GSTe2*	N316S	n/a	-
NC_064874.1	89827471	*GSTe2*	F330V	n/a	-
NC_064874.1	89827472	*GSTe2*	F330S	n/a	-
NC_064874.1	89827483	*GSTe2*	L334V	n/a	-
NC_064874.1	89827604	*GSTe2*	L353F	n/a	-

NC_064874.1	89827755	*GSTe2*	A404T	n/a	-
NC_064874.1	89827816	*GSTe2*	A424V	n/a	-
NC_064874.1	89827869	*GSTe2*	A442T	n/a	-
NC_064875.1	35309547	*vgsc*	T2I	T2	T2
NC_064875.1	35309652	*vgsc*	K37M	K37	Q37
NC_064875.1	35313439	*vgsc*	V177F	V166	V167
NC_064875.1	35314187	*vgsc*	W215L	W204	W205
NC_064875.1	35317106	*vgsc*	A421V	A413	A415
NC_064875.1	35318429	*vgsc*	P523L	P515	A528
NC_064875.1	35328791	*vgsc*	G1090E	n/a	n/a
NC_064875.1	35331900	*vgsc*	G1254R	G1217	G1276
NC_064875.1	35335129	*vgsc*	W1407R	W1370	W1429
NC_064875.1	35335134	*vgsc*	L1408F	W1371	W1430
NC_064875.1	35335624	*vgsc*	V1524M	V1487	V1546
NC_064875.1	35335757	*vgsc*	P1568L	P1531	P1590
NC_064875.1	35335762	*vgsc*	P1570S	P1533	P1592
NC_064875.1	35337347	*vgsc*	G2027E	G1992	G2030
NC_064875.1	35337353	*vgsc*	G2029E	G1994	S2031
NC_064875.1	35337437	*vgsc*	G2057E	S2022	V2066
NC_064875.1	35337458	*vgsc*	R2064H	R2029	R2073
NC_064875.1	53427108	*gaba/rdl*	S90N	S90	-
NC_064875.1	53427135	*gaba/rdl*	G81E	G81	-

*reference species for *ace-1* is the pacific electric ray (*Tetronarce californica*), and for *vgsc* is the common house fly (*Musca domestica).* There is no reference species for *gaba* and *GSTe2.*

**n/a indicates position was not present in this species

Chapter 6: Detection of insecticide resistance markers in *Anopheles funestus* from the Democratic Republic of the Congo using a targeted amplicon sequencing panel

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	lsh1704757	**Title**	Miss
First Name(s)	Holly		
Surname/Family Name	Acford-Palmer		
Thesis Title	A genomic investigation into three malaria vectors: An. darlingi, An. funestus, and An. stephensi		
Primary Supervisor	Prof. Susana Campino		

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?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

*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?	Scientfic Reports
Please list the paper's authors in the intended authorship order:	Holly Acford-Palmer, Monica Campos, Janvier Bandibabone, Sévérin N'Do, Chimanuka Bantuweko, Bertin Zawadi, Thomas Walker, Jody E. Phelan, Louisa A. Messenger, Taane G. Clark, 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 designed the amplicon panel, and tested the primers. I then performed the bioinformatic analysis, interpreted the results under the supervision from my supervisors. I wrote the first draft of the paper, which was commented upon by my supervisors and co-authors.

SECTION E

Student Signature	
Date	

Supervisor Signature	
Date	

Detection of insecticide resistance markers in *Anopheles funestus* from the Democratic Republic of the Congo using a targeted amplicon sequencing panel

Holly Acford-Palmer[1], Monica Campos[1], Janvier Bandibabone[2], Sévérin N'Do[3,4], Chimanuka Bantuweko[2,4], Bertin Zawadi[2], Thomas Walker[6], Jody E. Phelan[1] , Louisa A. Messenger[1,7], Taane G. Clark[1,8], Susana Campino[1]

[1] Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom

[2] Laboratoire d'Entomologie Médicale et Parasitologie, Centre de Recherche en Sciences Naturelles (CRSN/Lwiro), Sud-Kivu, Democratic Republic of the Congo.

[3] Médecins Sans Frontières (MSF) OCBA, Barcelona, Spain.

[4] Institut de Recherche en Sciences de la Santé (IRSS), Bobo-Dioulasso, Burkina Faso

[5] Université Officielle de Bukavu, Bukavu, Democratic Republic of the Congo

[6] School of Life Sciences, Gibbet Hill Campus, University of Warwick, Coventry, CV4 7AL, UK

[7] Department of Environmental and Occupational Health, School of Public Health, University of Nevada, Las Vegas. Las Vegas, United States

[8] Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom

Corresponding author:

Prof. Susana Campino

Department of Infection Biology, London School of Hygiene and Tropical Medicine

susana.campino@lshtm.ac.uk

Abstract

Vector control strategies have been successful in reducing the number of malaria cases and deaths globally, but the spread of insecticide resistance represents a significant threat to disease control. Insecticide resistance has been reported across *Anopheles (An.)* vector populations, including species within the *An. funestus* group. These mosquitoes are responsible for intense malaria transmission across sub-Saharan Africa, including in the Democratic Republic of the Congo (DRC), a country contributing >12% of global malaria infections and mortality events. To support the continuous efficacy of vector control strategies, it is essential to monitor insecticide resistance using molecular surveillance tools. In this study, we developed an amplicon sequencing ("Amp-seq") approach targeting *An. funestus*, and using multiplex PCR, dual index barcoding, and next-generation sequencing for high throughput and low-cost applications. Using our Amp-seq approach, we screened 80 *An. funestus* field isolates from the DRC across a panel of nine genes with mutations linked to insecticide resistance (*ace-1, CYP6P4, CYP6P9a, GSTe2, vgsc,* and *rdl*) and mosquito speciation (*cox-1, mtND5*, and *ITS2*). Amongst the 18 non-synonymous mutations detected, was N485I, in the *ace-1* gene associated with carbamate resistance. Overall, our panel represents an extendable and much-needed method for the molecular surveillance of insecticide resistance in *An. funestus* populations.

Word count: 202

Background

Malaria, caused by *Plasmodium* parasites and transmitted by *Anopheles spp.* mosquitoes, is a major public health problem contributing to substantial global morbidity and mortality[1]. The prevention of malaria relies on vector control measures, particularly the distribution of long-lasting insecticidal nets (LLINs)[1,2]. Whilst LLINs have contributed to significant drops in malaria burden since year 2000, there has been a plateauing in improvements in case reductions, coinciding with the spread of insecticide resistance across many *Anopheles* spp[3,4].

Species of the *Anopheles gambiae* (*An. gambiae)* group are the dominant malaria vectors across most of sub-Saharan Africa, but other species from the *An. funestus* group (*An. funest*us *sensu stricto*, *An. parensis*, *An. vandeeni*, and *An. rivulorum*) are also vectors and contribute to malaria transmission[5–7]. *An. funestus s.s.* mosquitos make up the largest population in the complex and have the widest geographical distribution[8]. This vector can thrive in varying climate conditions, is highly anthropophilic, and has night-time biting and endophilic resting behaviour[3,9]. These behaviours make *An. funestus* highly susceptible to traditional vector control methods, but resistance to insecticides has emerged[3,10–12].

The Democratic Republic of Congo (DRC) is a malaria hotspot, with >25 million cases (12% of global total), and transmission caused by both *An. gambiae* and *An. funestus* complex species[1]. Since 2015, there has been a 15% increase in malaria cases in DRC[1,13], with in-country vector control relying on mass distribution of LLINs, complemented by smaller-scale indoor and outdoor (ORS) residual spraying (IRS) in focal areas, by private mining enterprises. However, resistance to the four major classes of insecticides (carbamates, cyclodienes, organophosphates, pyrethroids) has emerged in *An. gambiae*, and *An. funestus s.s.*[13,14]

The underlying mechanisms of insecticide resistance across several mosquito species include target site mutations and metabolic-based, but alterations in microbiome composition and cuticles, as well as behavioural modifications, have been found to alter vector susceptibility[15–18]. Target site resistance results from single nucleotide polymorphisms (SNPs) that cause changes to the amino acid sequence in proteins involved in insecticide binding. The most well-known are the *kdr* (knock-down resistance) mutations in the *voltage-gated sodium channel* (*vgsc*), including *kdr* L1014F/S, V410L, F1508C, N1549Y, and D1763Y[19–22], which result in resistance to pyrethroids and DDT[23,24]. None of these mutations have been described in *An. funestus* populations despite extensive studies[9,25–28]. Other commonly observed mutations in *Anopheles spp*., including the G119S mutation in the *acetylcholinesterase-1* (*ace-1*) gene, leading to resistance to organophosphates and carbamates, but has not been observed in *An. funestus* populations[29–31]. However, the *ace-1* N485I mutation was identified in *An. funestus* samples from Malawi and linked to bendiocarb (carbamate) resistance[32]. Similarly, the A296S mutation in the *gaba* receptor, also known as the *rdl* (resistant to dieldrin) mutation, has been observed in *An. funestus* populationsand linked to several insecticides, including cyclodienes, a subgroup of organochlorides[31,33,34].

Metabolically mediated resistance mutations include the L119F mutation in the glutathione-S-transferase epsilon 2 (*GSTe2*) gene, which is linked to DDT resistance, and has been found in *An. funestus*[35,36]. Other work in this vector has sought to identify resistance associated alleles in cytochrome P450 genes (e.g., *CYP6P9a* and C*YP6P4*), with pyrethroid resistance linked to overexpression of the *CYP6P9a* gene in isolates from Southern Africa, driven by cis-regulatory polymorphisms[32,37–39].

The increasing resistance to insecticides in *An. funestus* highlights the need for rapid molecular surveillance techniques to identify underlying mutations, and thereby inform National Malaria Control Program for appropriate decisions about insecticide usage. Whole genome sequencing is limited by a

need for high DNA concentrations and the large size of mosquito genomes(~350Mbp) results in a high cost per sample. Amplicon sequencing ("amp-seq"), which can simultaneously target many genomic regions (each ~500bp) across candidate genes, has previously been applied to other vectors such as *An. gambiae, An. stephensi* and *Aedes aegypti*[40–43]. Amplicon primers designed for *An. gambiae* were tested *in silico* to check whether they were suitable for use on *An. funestus*, however the number of mismatches (n=>4) per primer, when compared to the reference sequence, meant they were unlikely to work efficiently on field specimens. Here we developed a targeted *An. funestus* amp-seq assay and applied it to 80 wild caught mosquitoes from the DRC to screen for molecular markers of insecticide resistance. The 17-amplicon panel covers regions in *vgsc, ace-1, CY9P6a, CYP9P4*, *GSTe2* and *rdl* loci for insecticide resistance profiling, as well as mitochondrial genes (cytochrome oxidase 1 (*cox-1)*, NADH dehydrogenase 5 (*mt-ND5*)) and the ribosomal locus *ITS2* (internally transcribed spacer 2) for speciation and phylogenetic analysis. The amp-seq assay uses a dual index barcoding system to facilitate the pooling of amplicons across many samples, thereby increasing throughput and decreasing costs. Our assay represents a promising strategy to support *An. funestus* vector control surveillance.

Methods

Sample Collection

Adult *Anopheles* were collected from households in two sites in Sud-Kivu province around 145km apart (Tchonka; 2° 19' 18" S, 27° 32' 24" E and Tushunguti; 1° 48' 19" S, 28° 45' 00.5" E) using Centers for Disease Control (CDC) light traps during the rainy seasons (Tchonka: April-June 2018; Tushunguti December 2017-February 2018). Mosquitoes were identified morphologically as members of the *An. funestus s.l.* group[44]. A total of 80 isolates were used for this study (Tchonka 70; Tushunguti 10). Individual mosquitoes were homogenized in a Qiagen TissueLyser II with sterilized 5 mm stainless steel

beads for 5 min at 30 Hz and incubated overnight at 56°C. DNA was extracted using a Qiagen DNeasy 96 blood and tissue kit (Qiagen, UK), according to the manufacturer's protocol.

Primer Design

Amplicon primers were designed with Primer3 software, using sequences from the *An. funestus* FUMOZ reference strain downloaded from VectorBase[45,46]. The primers were designed to amplify a region of around 500bp, typically around a SNP previously reported as associated with insecticide resistance in *Anopheles, Aedes*, or *Culex* mosquitoes. Where possible, these primers were designed to bind to exons. This panel comprised of 17 primer pairs (amplicons) targeting nine genes, including *vgsc* (6 amplicons), g*aba* (2), *ace-1* (3), *GSTe2* (1), *CYP6P4* (1), *CYP6P9a* (1) for insecticide resistance, and the *cox1* (1 amplicon), *ITS2* (1), and *mt-ND5* (1) for species identification or phylogenetic analysis (**Supplementary table 1**). Primers for the CYP6P9a amplicon were taken from Weedall *et al*.,2019[39]. Each primer was concatenated with one of ten unique 8bp barcodes at the 5' end. Each sample was assigned a barcode combination to be used throughout amplicon generation. This allowed for amplicons from samples with different barcodes to be pooled. To assess their suitability for multiplexing, samples were checked for potential dimer formation using ThermoFisher Scientific Multiple Primer Analyser software with sensitivity set to one.

Amplicon Generation

Using NEB Q5 hot start polymerase (New England BioLabs, UK), amplicons (500bp) were generated in 25ul reactions. Sample volume of 1ul (~2ng/ul) was used, with an average final primer concentration of 0.5uM in each PCR. The amplification was conducted as follows: hot-start polymerase activation for 3 minutes at 95°C, followed by 30 cycles of 95°C for 10 seconds, 60°C for 30 seconds and 72°C for 45 seconds, followed by a final elongation step of 72°C for 2 minutes. Post-multiplex PCR reaction, amplicons were visualised on a 1% agarose gel to confirm band size. Multiplexed PCR amplicons were first pooled by sample, and then with other samples with different 5' barcode combinations. Sample pools were purified using Roche Kapa beads following manufacturer's instructions. A bead to sample

ratio of 0.7:1 was used to remove excess primers and PCR reagents. The Qubit 2.0 fluorimeter HS DNA kit was used to quantify the pool concentration. Illumina adaptors and barcodes were ligated to the sample pool as a part of the Illumina-based Amplicon-EZ service (Genewiz, UK). The indexed pool was then sequenced using a 2 x 250bp paired end configuration on an Illumina MiSeq. A minimum of 50,000 reads were attained per pool, equating to at least 290 reads per amplicon in a pool of 170 amplicons (at a low cost of <US$0.5 per amplicon).

Amplicon Analysis

The multi-sample fastq files were first demultiplexed using an in-house python script (https://github.com/LSHTMPathogenSeqLab/amplicon-seq) into individual sample fastq files, through the unique barcode combination previously assigned. The reads were trimmed using the Trimmomatic package, then mapped to the reference sequence with bwa-mem and mapped reads clipped using the Samclip package[47–49]. Using the alignments, GATK HaplotypeCaller (v4.1.4.1, default parameters) and Freebayes (v1.3.5, --haplotype-length -1) software were applied for variant calling[50,51]. Any identified SNPs or insertions/deletions (INDELS) were filtered using bcftools for a minimum allele depth of 20 reads. The Phred score was also used for filtering, where a score of >30 per base was required to pass quality control checks. To determine the consequence of variants at an amino acid level, the SnpEff tool was applied with a database built from the FUMOZ reference genome[52]. The available reference genomes, at the time of analysis, either had no information about insecticide susceptibility, or were the pyrethroid resistant FUMOZ strain. Variants were genotyped using the proportion of alternate allele reads to total position reads for each sample. Samples were genotyped as homozygous reference (<20% alternate allele), heterozygous (20-80% alternate allele) or homozygous alternate (>80% alternate allele)[40,41,43].

Phylogenetic Analysis

For the *ITS2, cox1* and *mt-ND5* amplicons, each sample bam file was converted into fasta format using an in-house script (https://github.com/LSHTMPathogenSeqLab/fastq2matrix). This analysis required a depth of at least 20-fold in each position, and if samples had a large proportion (>90%) of uncalled bases, they were excluded from this analysis. For each gene, sequences were aligned using MAFFT software, along with publicly available sequences of *An. funestus* specimens from other countries[53]. For *ITS2*, this included 35 samples from Cameroon, DRC, Ethiopia, Kenya, Madagascar, Malawi, Mozambique, and Zambia. The *cox1* alignment included 111 sequences from Cameroon, Central African Republic, DRC, Gabon, Ghana, Kenya, Madagascar, Malawi, Mozambique, Tanzania, Uganda, and Zambia. For mt-ND5, 66 sequences from DRC, Ghana, Kenya, Malawi, Mozambique, Tanzania, Uganda, and Zambia were used. The alignments were then viewed and trimmed in Aliview[54]. RAxML software was used to construct maximum likelihood phylogenetic trees, with a bootstrap value of 1000, and gamma model of heterogeneity and GTR model of nucleotide substitution assumed[55]. The resulting tree model was visualised using iTOL software[56].

Haplotype Analysis

Specimen sequences were aligned using MAFFT software, and haplotype networks were constructed using the R package Pegas[57]. Amplicon nucleotide diversity and haplotype diversity were also calculated using the Pegas package. The vcftools package was used to calculate nucleotide diversity per SNP, fixation index, and linkage disequilibrium metrics. Linkage disequilibrium output was visualised with the Gaston R package[58,59].

Results

Detection of SNPs associated with insecticide resistance

Eighty *An. funestus* specimens were sequenced resulting in the identification of 377 variants (351 SNPs and 26 INDELs not previously described) across the 17 amplicons (**Supplementary Table 2**). The average coverage of amplicons varied from 193- to 3684-fold. Of the 351 SNPs identified, 92% were

either intronic variants or synonymous variants. A total of 18 missense SNPs were found, but no INDELs resulted in amino acid changes (**Table 1**). Of these 18 non-synonymous SNPs, only one had been previously reported – the N643I mutation in the *ace-1* gene (N485I in *Torpedo californica* otherwise known as Pacific electric ray). The *ace-1* N643I SNP occurred in samples from both Tchonka (7/70) and Tushunguti (1/10). The remaining 17 novel missense SNPs appeared in either the *CYP6P4* (I288N, G289R, N291S/T, L294V, K295E, E297K, D404N, and I414L), *GSTe2* (G80A, V134M, and K146T) or *vgsc* gene (domain II) (F763L, I768L/M, L788F, and G793C). All SNPs were detected at low frequencies, with allelic frequencies varying from 0.7 to 13.6%. Also identified was a 2bp insertion in the *CYP6P9a* amplicon, which occurs in a non-coding region, thereby not resulting in an amino acid change, but has been identified previously as a marker for pyrethroid resistance[39]. This insertion occurred in 91.3% of samples, with 73.9% of specimen's genotyped as homozygous alternate (R/R), 17.4% as heterozygous (R/S), and 8.7% as homozygous reference (S/S).

Linkage disequilibrium was calculated for each *CYP6P4* and *vgsc* (domain II) amplicons due to the high number of non-synonymous SNPs present. For the VGSCIIa amplicon, perfect linkage disequilibrium (LD r^2=1) was present between the I768L and G793C mutations. High LD was observed in the VGSCIIa amplicons between other sets of SNP pairs (F763L, I768M; F763L, L788F; I768M, L788F; I768L, G793C; all r^2>0.75), suggesting a strong association between these mutations. In the *CYP6P4* amplicons, perfect linkage (r^2=1) was observed between E297K and K295E, and L294V and N291T SNPs (**Supplementary Figure 1**).

Genetic diversity of *An. funestus* in Eastern DRC

Low genetic diversity was observed at all three loci (**Table 2**). The *cox-1* gene had the highest nucleotide diversity (0.011) and number of SNPs identified (64/351). This high number of SNPs resulted in high haplotype diversity, with 18 haplotypes identified and 50% being singletons. When 111 *cox-1* sequences from 11 other countries were included, 54 haplotypes were identified, 32 (58%)

of which were singletons. The *mt-ND5* gene also exhibited high haplotype diversity, but from a smaller number of SNPs (n=13), with 17 haplotypes identified (53% singletons). The number of haplotypes identified expanded to 53 (66% singleton) when including in the analysis 66 publicly available *mt-ND5* sequences covering seven countries. For *ITS2* sequences from DRC, four SNPs were identified, resulting in three haplotypes none of which had fewer than seven isolates present. When expanding these networks to include publicly available *ITS2* sequences (n=35; 8 countries), the number of haplotypes remained the same (n=3), with >20 isolates per haplotype. The haplotype networks for each gene (**Figure 1 a-c**), showed that most samples from the different countries shared a core haplotype, including the DRC samples. For the *ND5* sequences many DRC samples had haplotypes that were not present in the other countries (**Figure 1b**).

The data from the three genes demonstrated little population differentiation within the phylogenetic tree constructed **(Figure 2; Figure 3; Supplementary Figure 2).** Both mitochondrial genes demonstrated an ability to speciate, with separate clades for each *Anopheles spp* (**Figure 2; Figure 3**). The *cox-1* gene was able to distinguish at a greater resolution, with samples likely to be incorrectly identified through morphology as *An. funestus* (Anfun01, Anfun13, Anfun27 and Anfun71) appearing within the other *Anopheles spp.* clade. Anfun71 was in a cluster with the *An. arabiensis* and *An. gambiae s.s.* sequences, which was supported by a NCBI BLAST analysis that revealed it shares a 99.0% identity with *An. gambiae cox-1* isolates, and a 98.5% identity with *An. arabiensis cox-1* sequences. NCBI BLAST identified the remaining three samples (Anfun01, Anfun13, Anfun27) as *An. coustani* (identity >97%). In comparison, the *mt-ND5* gene did not speciate these samples as non-*An. funestus* but did reveal the clearest population differentiation between the DRC isolates and the publicly available sequences from other countries (**Figure 3**).

Discussion

The application of our amp-seq panel to *An. funestus* collected in Eastern DRC demonstrates its utility as a surveillance technique for genotypic-based insecticide resistance and species identification. Across the 80 DRC samples, we identified the *ace-1* N643I SNP, alongside 17 other putatively novel non-synonymous SNPs in genes associated with insecticide susceptibility. The *ace-1* N643I resistant allele, also known as N485I in *Torpedo californica*, was originally found in Southern African countries such as Mozambique and Malawi, but appears to have spread or emerged independently in DRC[32]. This SNP has been associated with increased resistance to the carbamate, bendiocarb[32]. Resistance to bendiocarb has also been reported in other *An. funestus* isolates from DRC collected in Tchonka[14]. The original study in Southern Africa only found heterozygous (R/S) genotypes, but in Malawi homozygous (R/R) were also detected, demonstrating higher resistance to bendiocarb than R/S genotypes[60]. In this study, the resistant allele appeared in 14.7% of samples, with only one sample classified as R/R. Since there are no reports of the use of carbamates or organophosphates in the DRC, it is possible that the N643I mutation is playing another role in resistance/cross-resistance to other insecticides or imparts a fitness advantage.

In the *CYP6P4* gene, the I288N, G289R, N291S/T, L294V, K295E, and E297K SNPs all occur within the variable region of the protein, close to I236M, which is a mutation linked to pyrethroid resistance in *An. gambiae* s.s. and *An. coluzzii*[37]. Some of these SNPs appear in a block of high linkage, probably due to close proximity. Other genetic variants were detected in this gene, including in codon 414, where an isoleucine changes to a leucine. This is unlikely to result in resistance to insecticides in *An. funestus*, as leucine is the reference amino acid for both *An. gambiae* s.s. and *An. arabiensis*. The other detected substitution D404N occurs in the conserved amino acid region, but is not proximal to the catalytic sites, so probably not involved in resistance. Our *CYP6P4* amplicon was designed based on the identification of deltamethrin binding site described previously, and believed to bind around the

Pro[376,] Leu[380,], and Ser[381 61]. However, additional, or modified amplicons can be included for *CYP6P4* if other positions are found to be important.

Of the three missense SNPs (G80A, K146T, V134M) found in the *GSTe2* amplicon, G80A is unlikely to have an impact on resistance, as other *Anopheles spp.* such as *An. sinensis* and *An. atroparvus* have alanine as the reference amino acid at this position. Similarly, the K146T alteration is present in other *Anopheles spp.* that have threonine as the reference amino acid at this position. For the V134M mutation, this position appears highly conserved across *Anopheles spp*. Codon position 134 exists within the H5 helix in the *GSTe2* protein, however H5 does not appear to play a part in DDT binding to *GSTe2*[35]. Mutations at nearby codon positions 131 and 139 have been previously reported and are not believed to alter insecticide susceptibility[35,36]. V134 was identified as a highly replaceable site across the GST family.

The five non-synonymous mutations found in the *vgsc* gene (F763L, I768L/M, L788F, and G793C) occur in the IIS1 domain of the VGSC protein. A T791M mutation has been previously reported in this region in *An. gambiae*, but no association with insecticide resistance was established[62]. In our work, the F763L, I768L, and G793C mutations all result in changes to amino acids found in other species at that position. The I768M and L788F mutations have not been observed in other *Anopheles spp.* For the I768M mutation, there was variation observed between species at codon 768, but methionine was not present. Whilst the leucine at codon 788 was highly conserved across species with no phenylalanine being reported previously. Future studies involving the analysis of genotype-phenotype associations in *An. funestus* populations could identify the possible involvement of these SNPs in insecticide resistance.

The absence of previously reported *vgsc-kdr* mutations and the *ace-1* G119S SNP [27,30] is not unexpected, as these have not yet been observed in *An. funestus* populations. Other molecular

mechanisms are involved in resistance to pyrethroids in this vector species. However, the continued attempts to detect the classic *kdr* mutations in *An. funestus* are necessary due the speed this highly favourable polymorphism can spread through the population, as seen with *An.gambiae*[27,63]. In DRC, *An. funestus* populations resistant to pyrethroid have been reported, likely due to the use of pyrethroid-only LLINs in the country[64]. It is therefore essential to investigate the genetic variants involved in insecticide resistance in *An. funestus*, due to the speed with which some of these highly favourable polymorphisms can spread through a population, as observed previously with *An. gambiae*[27,63].

The cytochrome P450 genes were included in our panel because of their association with metabolic-based insecticide resistance. The previously described 2bp insertion within the *CYP6P9a* promoter region was detected in >90% of our samples. This frequency is consistent with estimates based on applying a restriction fragment length polymorphism (RFLP) *CYP6P9a* diagnostic assay to a cohort of Tchonka and Tushunguti samples (82-98%)[10,39]. The 2bp insertion has been identified as a potential marker for pyrethroid resistance, being tightly linked with a resistant phenotype[39]. Other cytochrome P450 genes, such as *CYP6P9b*, have also been involved in insecticide resistant phenotype, particularly associated with elevated gene expression[65]. These loci can be integrated in the amplicon assay, particularly when genetic markers in these genes are uncovered to be involved in the resistance phenotype.

Of the phylogenetic markers included in the amplicon panel, those in *ITS2* showed the least utility for investigations into genetic diversity or relatedness. In contrast, the mitochondrial genes, *cox-1* and *mt-ND5*, showed more promise for speciation and population delineation. The *cox-1* gene was able to identify four samples that had been misclassified as *An. funestus*. Visual identification of mosquito species requires skilled, and experienced individuals, but such identification can often be of limited use due to sample degradation. The non-*An. funestus* isolates were found to be *An. gambiae* and *An.*

coustani, both known vectors in the region. *An. gambiae* is the focus of many vector control strategies across Sub-Saharan Africa, due to its large contribution to malaria transmission. *Anopheles coustani* is considered a secondary malaria vector across Central and Southern Africa. It is highly zoophilic and endophilic, differs sufficiently enough in behaviour to avoid many traditional vector control methods, and therefore its capacity for transmitting malaria is beginning to be taken more seriously[66,67]. Also such ability to escape IRS and LLINs through its behaviour, might be the origin of a future epidemic resurgence of malaria after the main malaria vector An. *gambiae* has been controlled. Whilst the occurrence of these species in this study could be the result of incorrect morphological identification, it may also be an example of species introgression[68]. BLAST analysis of the *cox-1* sequences for these misclassified samples revealed a 91-94% identity to *An. funestus,* compared to the 97-98% identity to the other species. Introgression of genes in *An. funestus* has previously been reported, but to confirm it is occurring here would require whole genome sequencing combined with a comparative genomic analysis[69].

For both mitochondrial genes, high haplotype diversity was observed in the context of the very low nucleotide diversity. This suggests a high number of low frequency variants, which has been observed for other *Anopheles spp.*[43]. The smaller sample size tested here may contribute to the low frequencies observed, so increasing the number of specimens screened with this amplicon panel would provide greater insights into the population dynamics.

Our study has demonstrated the utility of an amp-seq panel as a viable screening technique for SNPs associated with insecticide resistance. The detection of previously unreported missense SNPs also demonstrates its potential usage for the identification of new SNPs that may be involved in insecticide resistance, if used in tandem with phenotypic studies. Currently the use of this panel in a field setting may be limited by access to sequencing platforms, and a lack of bioinformatics expertise, and as such could be of more use in a research setting. However the use of a portable sequencer such as the long-

read MinION could help to overcome this, along with a graphical web interface platform for data analysis. Which has been done, and successfully implemented with both malaria and Tuberculosis[70,71]. Importantly, informed vector control methods are needed to meet the World Health Organization goals of reducing malaria mortality by 90% within the next seven years. Whilst gains have been made since this target was established, in recent years the number of cases has stabilised. New impetus is needed for large-scale surveillance studies with high throughput molecular tools to rapidly inform policy choices and reduce malaria cases. Our assay, which can be easily extended to other loci, represents a tool and opportunity to perform molecular surveillance in a vector heavily involved in malaria transmission across Africa.

DATA AVAILABILITY

All raw sequence data is listed in the European Nucleotide Archive (Project ID: PRJEB61194,

Accession numbers: ERR11507573 – ERR11507628

FUNDING & ACKNOWLEDGEMENTS

H.A.P. is funded by a BBSRC LIDo-DTP PhD studentship. T.G.C. and S.C. are funded by UKRI grants (ref. BB/X018156/1, MR/X005895/1). T.W. was funded by a Wellcome Trust/Royal Society Sir Henry Dale Fellowship (101285/Z/13/Z). L.A.M. is jointly funded by the Foreign, Commonwealth & Development Office (FCDO), the Medical Research Council, Wellcome and Department of Health and Social Care (Grant Ref: MR/R006040/1).

REFERENCES

1. Global Malaria Programme: WHO Global. *WHO World Malaria Report 2022*. *WHO*.

2. Wat'senga, F. *et al.* Nationwide insecticide resistance status and biting behaviour of malaria vector species in the Democratic Republic of Congo. *Malar. J.* **17**, 129 (2018).

3. Coetzee, M. & Fontenille, D. Advances in the study of Anopheles funestus, a major vector of malaria in Africa. *Insect Biochem. Mol. Biol.* **34**, 599–605 (2004).

4. Alout, H., Roche, B., Dabiré, R. K. & Cohuet, A. Consequences of insecticide resistance on malaria transmission. *PLoS Pathog.* **13**, e1006499 (2017).

5. Sinka, M. E. *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* **3**, 117 (2010).

6. Kahamba, N. F. *et al.* Using ecological observations to improve malaria control in areas where Anopheles funestus is the dominant vector. *Malar. J.* **21**, 158 (2022).

7. Odero, J. O. *et al.* Advances in the genetic characterization of the malaria vector, Anopheles funestus, and implications for improved surveillance and control. *Malar. J.* **22**, 230 (2023).

8. Ibrahima Dia, Moussa Wamdaogo Guelbeogo, & Diego Ayala. Advances and Perspectives in the Study of the Malaria Mosquito Anopheles funestus | IntechOpen. in.

9. Coetzee, M. & Koekemoer, L. L. Molecular Systematics and Insecticide Resistance in the Major African Malaria Vector Anopheles funestus. *Annu. Rev. Entomol.* **58**, 393–412 (2013).

10. Bandibabone, J. *et al.* Investigating molecular mechanisms of insecticide resistance in the Eastern Democratic Republic of the Congo. *Malar. J.* **20**, 464 (2021).

11. Weedall, G. D. *et al.* An Africa-wide genomic evolution of insecticide resistance in the malaria vector Anopheles funestus involves selective sweeps, copy number variations, gene conversion and transposons. *PLOS Genet.* **16**, e1008822 (2020).

12. Riveron, J. M., Watsenga, F., Irving, H., Irish, S. R. & Wondji, C. S. High Plasmodium Infection Rate and Reduced Bed Net Efficacy in Multiple Insecticide-Resistant Malaria Vectors in Kinshasa, Democratic Republic of Congo. *J. Infect. Dis.* **217**, 320–328 (2018).

13. World Health Organisation. *Global report on insecticide resistance in malaria vectors: 2010-2016 Global Malaria Programme*.

14. WHO. Global Malaria Programme: Malaria Threats Map. https://www.who.int/teams/global-malaria-programme/surveillance/malaria-threats-map.

15. Barnard, K., Jeanrenaud, A. C. S. N., Brooke, B. D. & Oliver, S. V. The contribution of gut bacteria to insecticide resistance and the life histories of the major malaria vector Anopheles arabiensis (Diptera: Culicidae). *Sci. Rep.* **9**, 9117 (2019).

16. Balabanidou, V., Grigoraki, L. & Vontas, J. Insect cuticle: a critical determinant of insecticide resistance. *Curr. Opin. Insect Sci.* **27**, 68–74 (2018).

17. Chareonviriyaphap, T. *et al.* Review of insecticide resistance and behavioral avoidance of vectors of human diseases in Thailand. *Parasit. Vectors* **6**, 280 (2013).

18. Pelloquin, B. *et al.* Overabundance of Asaia and Serratia Bacteria Is Associated with Deltamethrin Insecticide Susceptibility in Anopheles coluzzii from Agboville, Côte d'Ivoire. *Microbiol. Spectr.* **9**, e00157-21 (2021).

19. Chang, C. *et al.* A novel amino acid substitution in a voltage-gated sodium channel is associated with knockdown resistance to permethrin in Aedes aegypti. *Insect Biochem. Mol. Biol.* **39**, 272–278 (2009).

20. Haddi, K. *et al.* Detection of a new pyrethroid resistance mutation (V410L) in the sodium channel of Aedes aegypti: a potential challenge for mosquito control. *Sci. Rep.* **7**, 46549 (2017).

21. 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.* **109**, 6614–6619 (2012).

22. Du, Y. *et al.* Molecular evidence for dual pyrethroid-receptor sites on a mosquito sodium channel. *Proc. Natl. Acad. Sci.* **110**, 11785–11790 (2013).

23. 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).

24. 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).

25. Kamgang, B. *et al.* Exploring insecticide resistance mechanisms in three major malaria vectors from Bangui in Central African Republic. *Pathog. Glob. Health* **112**, 349–359 (2018).

26. Tchigossou, G. *et al.* Molecular basis of permethrin and DDT resistance in an Anopheles funestus population from Benin. *Parasit. Vectors* **11**, 602 (2018).

27. Irving, H. & Wondji, C. S. Investigating knockdown resistance (kdr) mechanism against pyrethroids/DDT in the malaria vector Anopheles funestus across Africa. *BMC Genet.* **18**, 76 (2017).

28. Morgan, J. C., Irving, H., Okedi, L. M., Steven, A. & Wondji, C. S. Pyrethroid Resistance in an Anopheles funestus Population from Uganda. *PLOS ONE* **5**, e11872 (2010).

29. Elanga-Ndille, E. *et al.* The G119S Acetylcholinesterase (Ace-1) Target Site Mutation Confers Carbamate Resistance in the Major Malaria Vector Anopheles gambiae from Cameroon: A Challenge for the Coming IRS Implementation. *Genes* **10**, 790 (2019).

30. Cuamba, N., Morgan, J. C., Irving, H., Steven, A. & Wondji, C. S. High Level of Pyrethroid Resistance in an Anopheles funestus Population of the Chokwe District in Mozambique. *PLOS ONE* **5**, e11010 (2010).

31. Mugenzi, L. M. J. *et al.* Escalating pyrethroid resistance in two major malaria vectors Anopheles funestus and Anopheles gambiae (s.l.) in Atatam, Southern Ghana. *BMC Infect. Dis.* **22**, 799 (2022).

32. Ibrahim, S. S., Ndula, M., Riveron, J. M., Irving, H. & Wondji, C. S. The P450 CYP6Z1 confers carbamate/pyrethroid cross-resistance in a major African malaria vector beside a novel carbamate-insensitive N485I acetylcholinesterase-1 mutation. *Mol. Ecol.* **25**, 3436–3452 (2016).

33. ffrench-Constant, R. H., Rocheleau, T. A., Steichen, J. C. & Chalmers, A. E. A point mutation in a Drosophila GABA receptor confers insecticide resistance. *Nature* **363**, 449–451 (1993).

34. Yang, C., Huang, Z., Li, M., Feng, X. & Qiu, X. RDL mutations predict multiple insecticide resistance in Anopheles sinensis in Guangxi, China. *Malar. J.* **16**, 482 (2017).

35. 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 Biol.* **15**, R27 (2014).

36. Helvecio, E. *et al.* Polymorphisms in GSTE2 is associated with temephos resistance in Aedes aegypti. *Pestic. Biochem. Physiol.* **165**, 104464 (2020).

37. Fotso-Toguem, Y. *et al.* Genetic Diversity of Cytochrome P450s CYP6M2 and CYP6P4 Associated with Pyrethroid Resistance in the Major Malaria Vectors Anopheles coluzzii and Anopheles gambiae from Yaoundé, Cameroon. *Genes* **14**, 52 (2022).

38. Mugenzi, L. M. J. *et al.* Cis-regulatory CYP6P9b P450 variants associated with loss of insecticide-treated bed net efficacy against Anopheles funestus. *Nat. Commun.* **10**, 4652 (2019).

39. Weedall, G. D. *et al.* A cytochrome P450 allele confers pyrethroid resistance on a major African malaria vector, reducing insecticide-treated bednet efficacy. *Sci. Transl. Med.* **11**, eaat7386 (2019).

40. 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).

41. Collins ID, 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).

42. Makunin, A. *et al.* A targeted amplicon sequencing panel to simultaneously identify mosquito species and Plasmodium presence across the entire Anopheles genus. *Mol. Ecol. Resour.* **22**, 28–44 (2022).

43. 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).

44. Gillies, M. T. & De Meillon, B. The Anophelinae of Africa south of the Sahara (Ethiopian Zoogeographical Region). *Anophelinae Afr. South Sahara Ethiop. Zoogeographical Reg.* (1968).

45. Untergasser, A. *et al.* Primer3—new capabilities and interfaces. *Nucleic Acids Res.* **40**, e115–e115 (2012).

46. VectorBase. https://vectorbase.org/vectorbase/app/.

47. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinforma. Oxf. Engl.* **30**, 2114–2120 (2014).

48. H.Li. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arxiv.org* **n/a**, n/a (2013).

49. tseemann/samclip: Filter SAM file for soft and hard clipped alignments. https://github.com/tseemann/samclip.

50. Poplin, R. *et al.* Scaling accurate genetic variant discovery to tens of thousands of samples. 201178 Preprint at https://doi.org/10.1101/201178 (2018).

51. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. (2012) doi:10.48550/arxiv.1207.3907.

52. Cingolani, P. *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 (Austin)* **6**, 80–92 (2012).

53. Katoh, K., Misawa, K., Kuma, K. I. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).

54. Larsson, A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30**, 3276–3278 (2014).

55. Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455 (2019).

56. Alam, M. T., Bora, H., Das, M. K. & Sharma, Y. D. The type and mysorensis forms of the Anopheles stephensi (Diptera: Culicidae) in India exhibit identical ribosomal DNA ITS2 and domain-3 sequences. *Parasitol. Res.* **103**, 75 (2008).

57. Paradis, E. pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics* **26**, 419–420 (2010).

58. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

59. Perdry, H., Dandine-Roulland, C., class), D. B. (C++ gzstream & class), L. K. (C++ gzstream. gaston: Genetic Data Handling (QC, GRM, LD, PCA) & Linear Mixed Models. (2022).

60. Menze, B. D. *et al.* Marked aggravation of pyrethroid resistance in major malaria vectors in Malawi between 2014 and 2021 is partly linked with increased expression of P450 alleles. *BMC Infect. Dis.* **22**, 660 (2022).

517 61. Ibrahim, S. S., Riveron, J. M., Stott, R., Irving, H. & Wondji, C. S. The cytochrome P450 CYP6P4 is
518 responsible for the high pyrethroid resistance in knockdown resistance-free Anopheles arabiensis.
519 *Insect Biochem. Mol. Biol.* **68**, 23 (2016).

520 62. Clarkson, C. S. *et al.* The genetic architecture of target-site resistance to pyrethroid insecticides in
521 the African malaria vectors Anopheles gambiae and Anopheles coluzzii. *Mol. Ecol.* **30**, 5303–5317
522 (2021).

523 63. Ranson, H. *et al.* Pyrethroid resistance in African anopheline mosquitoes: what are the implications
524 for malaria control? *Trends Parasitol.* **27**, 91–98 (2011).

525 64. N'Do, S. *et al.* Insecticide resistance profiles in malaria vector populations from Sud-Kivu in the
526 Democratic Republic of the Congo. *Trans. R. Soc. Trop. Med. Hyg.* **115**, 1339–1344 (2021).

527 65. Mugenzi, L. M. J. *et al.* The duplicated P450s CYP6P9a/b drive carbamates and pyrethroids cross-
528 resistance in the major African malaria vector Anopheles funestus. *PLOS Genet.* **19**, e1010678
529 (2023).

530 66. Ciubotariu, I. I. *et al.* Genetic Diversity of Anopheles coustani (Diptera: Culicidae) in Malaria
531 Transmission Foci in Southern and Central Africa. *J. Med. Entomol.* **57**, 1782–1792 (2020).

532 67. Goupeyou-Youmsi, J. *et al.* Differential contribution of Anopheles coustani and Anopheles
533 arabiensis to the transmission of Plasmodium falciparum and Plasmodium vivax in two
534 neighbouring villages of Madagascar. *Parasit. Vectors* **13**, 430 (2020).

535 68. Fontaine, M. C. *et al.* Extensive introgression in a malaria vector species complex revealed by
536 phylogenomics. *Science* **347**, 1258524 (2015).

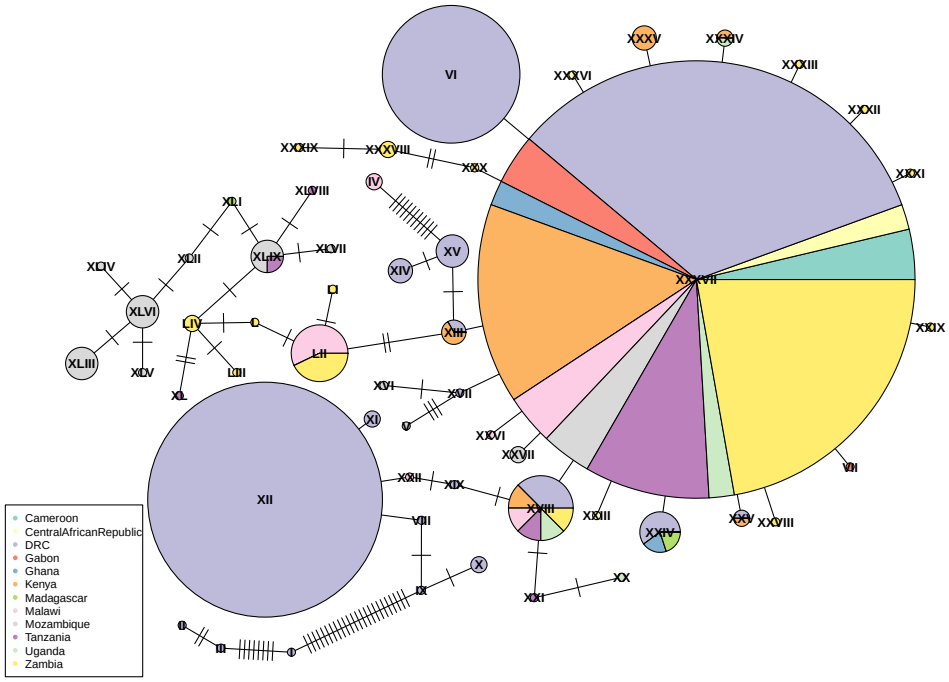
537 69. Riveron, J. M. *et al.* Rise of multiple insecticide resistance in Anopheles funestus in Malawi: a major
538 concern for malaria vector control. *Malar. J.* **14**, 344 (2015).

539 70. Phelan, J. E. *et al.* Integrating informatics tools and portable sequencing technology for rapid
540 detection of resistance to anti-tuberculous drugs. *Genome Med.* **11**, 41 (2019).

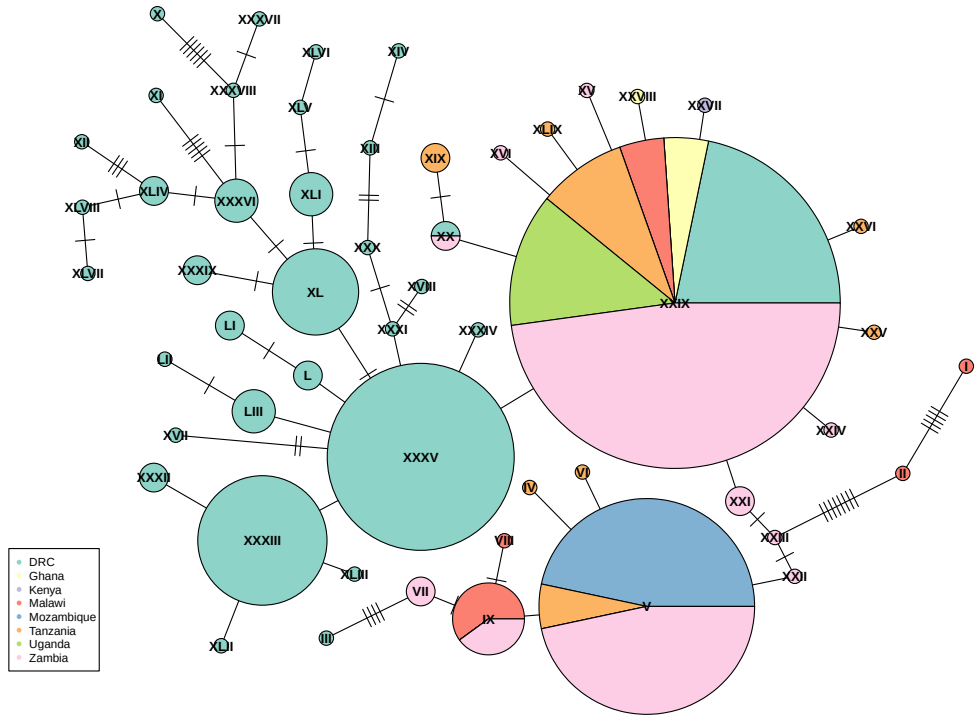
541 71. PathogenSeq lab. https://bioinformatics.lshtm.ac.uk/malaria-profiler/.

542

a)


Cameroon
CentralAfricanRepublic
DRC
Gabon
Ghana
Kenya
Madagascar
Malawi
Mozambique
Tanzania
Uganda
Zambia

b)


DRC
Ghana
Kenya
Malawi
Mozambique
Tanzania
Uganda
Zambia

c)

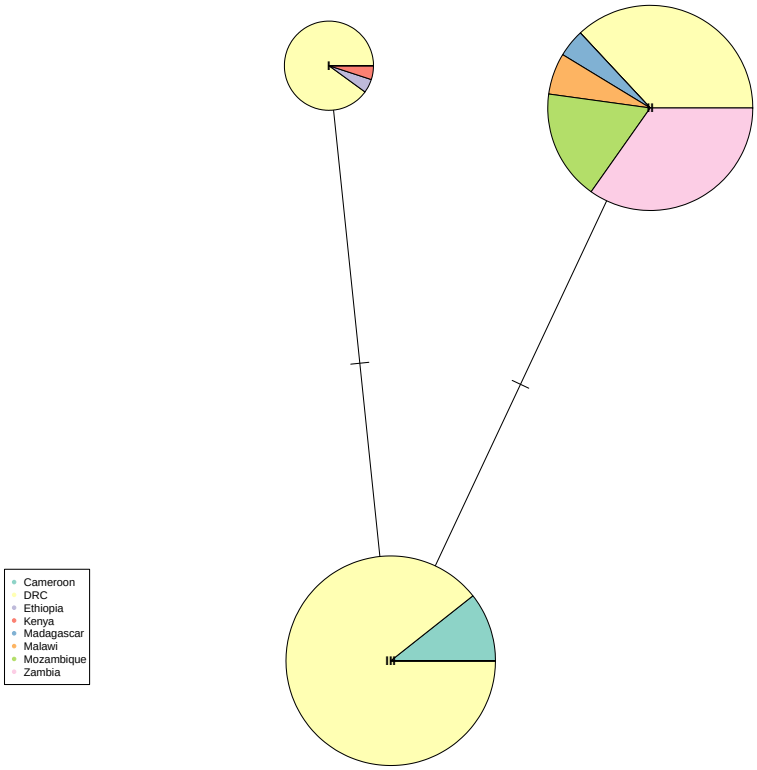


Figure 1. Haplotype or minimal-spanning network constructed using a) *cox-1*, b) *mt-ND5,* and c) ITS2 sequences generated in this study and publicly available samples. Each node represents a haplotype, each segment within the node represents a country, and is proportionally sized to the number of sequences present in the segment and node. The number of number of ticks between nodes represents the number of genetic differences between nodes.

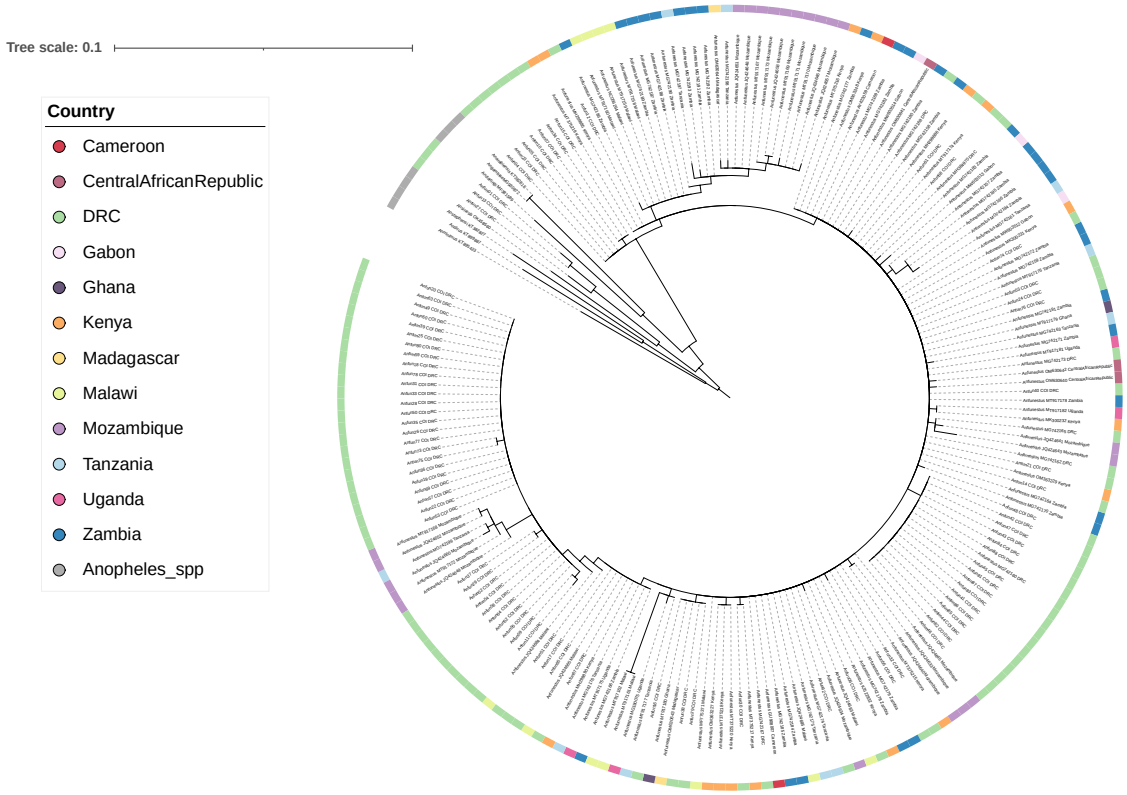


Figure 2. Maximum-likelihood tree constructed using *cox-1* gene sequences generated in this study (n=84), alongside other publicly available *An. funestus cox-1* sequences (n=111), (Cameroon = 2, Central African Republic = 3, DRC = 7, Gabon = 3, Ghana = 2, Kenya = 16, Madagascar = 2, Malawi = 11, Mozambique = 21, Tanzania = 10, Uganda = 4, Zambia = 30). This tree also has a group of *Anopheles spp.* (n = 7), including *An. arabiensis, An. darlingi, An. dirus, An. gambiae s.s, An. minimus, An. sinensis* and *An. stephensi.* The tree was built using the maximum-likelihood method assuming GTR model of nucleotide substitution, with the gamma model of heterogeneity rate.

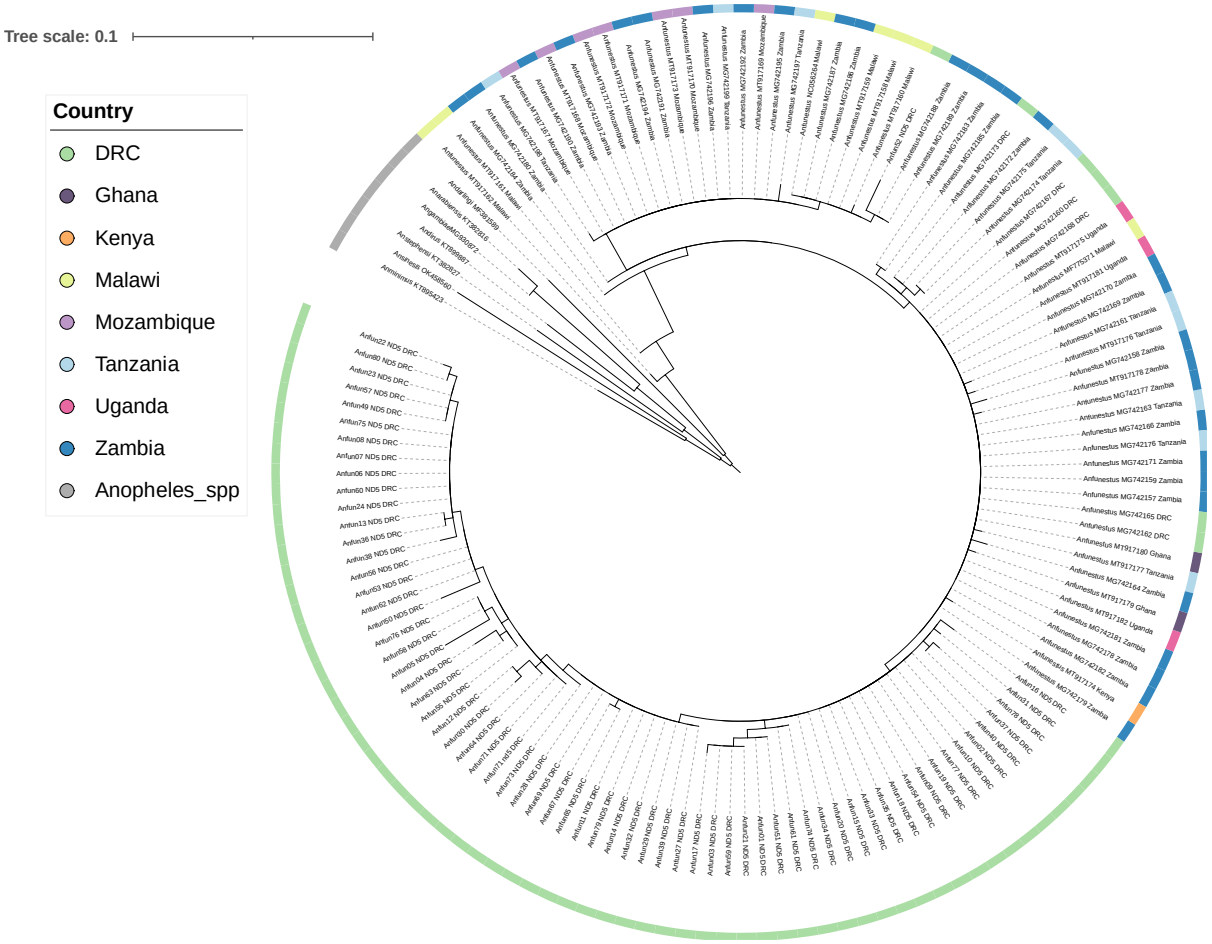


Figure 3. Maximum-likelihood tree constructed using *mt-ND5* gene sequences generated in this study (n=67), alongside other publicly available *An. funestus mt-ND5* sequences (n=66), (DRC = 6, Ghana = 2, Kenya = 1, Malawi = 7, Mozambique = 7, Tanzania = 10, Uganda = 3, Zambia = 30). This tree also has a group of *Anopheles spp.* (n = 7), including *An. arabiensis, An. darlingi, An. dirus, An. gambiae s.s, An. minimus, An. sinensis* and *An. stephensi.* The tree was built using the maximum-likelihood method assuming GTR model of nucleotide substitution, with the gamma model of heterogeneity rate.

Table 1. Location and frequencies of non-synonymous variants detected

Amplicon	Position	Sample Number*	Annotation	Allele frequencies		Nucleotide Diversity
				Reference	Non-reference	
ACE1III	19555221	68	Asn643Ile	91.9	8.1	0.25
CYP6P4	8560733	66	Ile414Leu	97.7	2.3	0.07
	8560763	66	Asp404Asn	93.2	6.8	0.16
	8561152	66	Glu297Lys	92.4	7.6	0.19
	8561158	66	Lys295Glu	92.4	7.6	0.19
	8561161	66	Leu294Val	96.2	3.8	0.14
	8561169	66	Asn291Ser	97.8	2.2	0.07
	8561169	66	Asn291Thr	94.7	5.3	0.14
	8561176	66	Gly289Arg	98.5	1.5	0.06
	8561178	66	Ile288Asn	91.7	8.3	0.18
GSTe2	75252570	70	Lys146Thr	93.6	6.4	0.18
	75252607	70	Val134Met	99.3	0.7	0.03
	75252839	70	Gly80Ala	86.4	13.6	0.27
VGSCIIa	42339660	71	Phe763Leu	99.3	0.7	0.14
	42339675	71	Ile768Leu	97.9	2.1	0.11
	42339677	71	Ile768Met	98.9	1.4	0.14
	42339735	71	Leu788Phe	97.2	2.8	0.15
	42339750	71	Gly793Cys	97.9	2.1	0.11

Table 2. Haplotype and nucleotide diversity of genes in DRC

Gene	No. of haplotypes	Haplotype diversity	Nucleotide diversity
ITS2	3	0.60	0.002
Cox1	18	0.82	0.011
ND5	16	0.93	0.006

Supplementary table 1. Amplicon targets, primers sequences, and size (bp)

***Anopheles funestus* Primers information**							
	Target Gene	Amplicon	Accession ID	Target SNP	Forward primer	Reverse Primer	Product Size (bp)
Insecticide Resistance	*vgsc*	VGSCIa	AFUN000494	n/a	TGCTTGGTTAAACCGAATAAAA	GCTTCGAACAATCGAAATCA	479
		VGSCIb		V410L[22]	TTTTCTTCGATGGTGTGCAA	ATGTGGTTGGCATTGTTTCA	439
		VGSCIIa		n/a	AATGCTCGCCTACGATGTTC	CGTTGCTGTGAAGAACTGGA	463
		VGSCIIb		L1014F/S[19,20]	AGTGCTGGTGAACGCCTAAT	CATTTGCTATGTTCGCCTTG	484
		VGSCIII		F1534C[24]/ N1575Y[23]	GCGACATTCAAAGGATGGAT	GAAGCCGATGAACAACATGA	513
		VGSCIV		D1763Y[21]	TGTACTGCGTTTGGTGAAGG	CTTCCGTTGCTTGGGAATAG	528
	ace-1	ACE1_Ia	AFUN011616	n/a	CATCACTACCACCACCACCA	GCCAAGCCACACATCTACCT	500
		ACE1_Ib	AFUN011616	G1195[25]	GAGCGGCAAAAAGGTAGATG	GCTGATAGCACACCTCCACA	495
		ACE1_II	AFUN011616	N485I[32]	GCTACACCGAGGACGAGAAA	CTAGCAGCACAACCAGATCA	480
	GSTe2	GSTe2	AFUN015809	L119F[30,31]	CCTTCTGGCTGGTGATCATT	ACCCAGCAACAAAGTCATCC	496
	rdl	RDL1	AFUN016339	A296S/ V371I[33,34]	AACTCGCCCACAATGAAAAA	TTTAATGGGCAAACGGAAAG	488
		RDL2	AFUN016339	T345S[34]	AACACTTGTGATGCTTTTCGT	GAACCGTACCTCCTGGAACA	482
	P450 (*CYP6P9a*)	CYP6P9a*	AFUN015792	aa insertion[39]	AGGTTATTTGGACGAACGTGA	CAAAACAAGTTAAGTGCCTGTAAGA	496
	P450 (*CYP6P4*)	CYP6P4	EU852651.1	P376, L380, & S381[62]	AGGTTATTTGGACGAACGTGA	CTGGGTAAAACTCGGGATCA	496
Phylogeny	Mitochondria (*cox-1 & ND5*)	COX-1	AF15KAPO-006	n/a	GAGCCCCTGATATAGCTTTCC	AAACTTCTGGGTGTCCAAAAA	468
		Mt-ND5		n/a	TCTTTTGTTTTATTGATTTCATCG	TCCTGCTGTAACCAAAGTTGAA	496
	ITS2	ITS2	MK129243.1	n/a	TCTAGTGTCGTGGGGGAAAC	AGCACGTTGTCCGAATATCA	498
*These primers were designed by Weedall et al., 2019[38]							

Supplementary Table 2. Average coverage, and number of variants identified for each amplicon

Amplicon	Coverage	SNPs	NS SNPs	INDELs
All	-	351	18	26
ACE1_I	193.19	15	0	2
ACE1_II	722.50	11	0	0
ACE1_III	671.56	23	1	0
COI	2194.28	64	0	1
CYP9P4	574.96	35	9	0
CYP9P6a	938.42	55	0	6
GSTe2	846.42	18	3	2
ITS2	3684.57	6	0	0
mt-ND5	1059.01	13	0	0
VGSCIa	1205.03	5	0	2
VGSCIb	1012.69	5	0	2
VGSCIIa	2521.66	39	5	2
VGSCIIb	839.73	7	0	0
VGSCIII	870.34	12	0	0
VGSCIV	512.47	8	0	0
Rdl1	1326.72	7	0	6
Rdl2	843.04	18	0	3

Chapter 7: Discussion & Conclusions

Discussion

To combat the global malaria burden, a multifaceted approach, underpinned by effective vector control is essential. Insecticides are the most broadly used form of vector control, and monitoring of mosquito genomes through next-generation sequencing is essential in identifying the emergence and spread of DNA alterations that could result in insecticide resistance. Current molecular screening methods in *Anopheles spp.,* heavily relies on capillary sequencing of a few genes of interest (e.g., *vgsc* or *ace-1*). These studies reveal important information about the genetic diversity of these genes and help inform vector control efforts. However, this method is low-throughput, and screening known loci could mean important changes occurring elsewhere in the genome are missed.

The need for multi-locus high-throughput molecular screening of *Anopheles* for insecticide resistance has been highlighted many times, and this thesis demonstrates the development of several such panels. However, also demonstrated in this thesis, is that a one assay fits all approach could be difficult, due to the genetic variation between species. As observed with *An. funestus,* in **Chapter 6,** different molecular mechanisms are at play that result in decreased insecticide susceptibility, including increased occurrence of *CYP*-based resistance with an absence of commonly observed target-site mutations such as *kdr*[1,2]. Wide scale WGS studies are needed for a variety of different *Anopheles* species, to identify SNPs, copy number variations and INDELs that can be used for molecular markers for resistance.

WGS has been successfully applied to several *Anopheles spp.* and has increased our understanding of their genomic landscapes[3–5]. In *An. gambiae* and *An. coluzzii,* WGS illuminated CNVs associated with metabolically mediated resistance to pyrethroids, the same data set was also used to identify mutations in the *vgsc* with a role in pyrethroid resistance[3,6]. Studies examining Asian malaria vector *An. minimus* identified diverging populations within Cambodia, this could impact on the effectiveness of malaria control methods if this genetic diversity reflects behavioural diversity as it does in *An. darlingi*[7]. Simultaneously understanding gene flow within mosquito species can indicate how resistance associated mutations can move through a population[8]**.** However, knowledge of many other of *Anopheles*

spp., is severely lacking, and with the high amount of genetic diversity observed within this genus, it is clear there is more to learn.

At the time of writing this thesis there were fewer than 50 samples with WGS data available for *An. stephensi,* and <5 for *An. darlingi*. There were over 100 for *An. funestus,* which is an important vector for *P. falciparum* in Sub-Saharan Africa, which has substantially more research has focussed on it due to its impact on the mortality burden. Whilst for *An. gambiae s.l.* there are >1,000 available due to genome sequencing projects targeting this vector[4].

An. darlingi and *An. stephensi* have historically been *P. vivax* vectors, due to their geographical distribution (South America and South Asia respectively), and *P. vivax* is typically not a research priority due to the lower mortality burden. However, the spread of *An. stephensi* into the HOA, occupying urban areas and with its proficiency as a vector for both *P. falciparum* and *P. vivax,* has meant this vector has become a higher research priority[9–11].

Access to WGS is limited due to high costs, particularly when applied to large genomes, and the need for high DNA concentrations, often not possible from vectors. The use of amp-seq panels provide an appealing alternative to WGS, being both low cost and requiring smaller DNA quantities. Amp-seq also has the benefit of being multiplexable, meaning it can be a high-throughput method of screening for insecticide resistance mutations[12,13]. It can be easily adapted to include new loci to the panel relevant to the targeted species, while also targeting highly variable genomic regions to gain insight into population dynamics.

In **Chapter 2**, I describe the development of an amp-seq panel for *An. stephensi*, and its subsequent application to both Ethiopian field and SDA-500 colony isolates. The nine-amplicon panel covers four genes associated with insecticide resistance and a further two for phylogenetic and speciation analysis. This work resulted in the identification of two known target site mutations: *kdr*-L1014F and *rdl*-A296S*,* alongside a further two putatively novel missense SNPs in the *ace-1* and *GSTe2* amplicons. The L1014F SNP or *kdr* west SNP, had previously been identified by Samake et al, in five locations throughout Eastern

Ethiopia[14]. However, they did not identify the SNP in the Awash Sebat Kilo location where these isolates were collected from. This may be a result of the higher sensitivity of next-generation sequencing in comparison to capillary sequencing. The L1014F mutation was only detected as a heterozygous genotype, which using capillary sequencing may be difficult to detect. The appearance of this SNP may be due to the temporal differences in the two sample sets. Samake et al, collected mosquitoes August to November 2018, whilst my study isolates were collected from April to September 2019. The temporal difference may mean the *kdr* had risen to a higher allele frequency so was more easily detectable, or gene flow had occurred with *An. stephensi* from nearby villages that had the L1014F mutation in the population. My work had a larger sample size (n=8 versus n=95), and due to the low frequency this SNP was present at (13.8%), it could have been missed previously[14]. The detection of this *kdr* mutation is in line with my hypotheses of the detected pyrethroid resistance in this area would likely indicate the presence of this target site mutation. Although the identification of only this resistance marker despite widespread resistance to the four main insecticide classes, may indicate other mechanisms are behind this phenotypic resistance.

The *rdl* A296S mutation was also detected in this population, whilst this SNP is not typically looked for in traditional molecular screening assays due to the removal of dieldrin as an insecticide. The occurrence here may be a result of pesticide use by local farmers, or from allele fixation in the source population.

From the phylogeny and speciation markers, *cox-1* and *ITS2,* the former demonstrated higher levels of variation and therefore greater utility for phylogenetic inference. The work demonstrated that there were multiple haplotypes present in the Ethiopian populations, suggesting multiple introduction events, as seen previously[10]. One of these haplotypes was shared with populations from Pakistan. Another haplotype (denoted as IV) was shared with Pakistan, India, Iran, UAE, and Sri Lanka, which does not assist the refining of the origins of this invasive species. Previous work has demonstrated Ethiopian isolates share more similarity with South Asian *An. stephensi*, than those from the Arabian Peninsula.[10] However further sampling is needed from more locations in the HOA, along with more countries where *An. stephensi* is native, to elucidate the origins of this species and migration patterns.

This amp-seq panel demonstrated its utility as a high-throughput panel for screening insecticide resistance and investigating population dynamics. Since publication, the amp-seq panel has subsequently been used in a study examining potential applications of amplicon sequencing to environmental DNA (eDNA), where it detected the same two putatively novel missense SNPs in colony samples, also identified in this study in colony and Ethiopian isolates[15].

Further investigation of Ethiopian *An. stephensi* occurred in **Chapter 3**. Here I applied WGS to 29 Ethiopian *An. stephensi*, and two SDA-500 colony samples. The resulting data was analysed in the context of publicly available sequence data from three collection sites in India (Bangalore, Chennai, and Mangalore), two colony strains from India, as well as further SDA-500 colony samples. I identified the two known SNPs (*kdr*-L1014F and *rdl*-A296S) in the Ethiopian samples. The A296S mutation was also detected in the India-wildtype isolates along with V327I, both in the *gaba* gene. The two SNPs identified in the Ethiopian isolates are in line with the amp-seq results from **Chapter 2**. The two putatively novel SNPs identified in that chapter by amp-seq were also found in this population. The identification of the known *gaba* gene mutations in the Indian populations, represents the first time these SNPs have been identified in Indian *An. stephensi*, and the V327I SNP has only been observed in *An. sinensis* and *An. funestus* thus far[16,17]. As discussed above, the presence of insecticide resistance associated SNPs in the *gaba* gene are likely the result of exposure to pesticides used in farming[18]. As discussed above the identification of only two target site mutations in the Ethiopian samples despite the likely phenotypically resistant samples. Further target site mutations e.g., *ace-1 G119S* or GSTe2 L119F, would be expected, indicating other mechanisms may be at play.

Several copy number variants (CNV) were identified in genes of interest, including *CYP307a1*, *CYP6a1*, and *CYP9f2*. The first two being a deletion events, and latter two a duplication. CNVs have been reported to result in resistant phenotypes, so these alterations require further investigation[19].

The ancestry analysis conducted on this data set identified shared ancestry between the Indian isolates (Mangalore) and Ethiopian isolates. This ancestry was also observed in a single SDA-500 sample from Pakistan. The shared ancestry between the Mangalore and Ethiopian samples would suggest a shared

origin, and the similarity between HOA *An. stephensi* and South Asian *An. stephensi* has been noted previously[10]. The Ethiopian samples in this study appear genetically homogenous, with only one ancestry present, whereas the Mangalore isolates all had at least three ancestries present. The Ethiopian isolates were collected in one location, so expanding the sampling area may increase the genetic diversity and number of ancestries observed in HOA populations. The WGS of *An. stephensi* collected across Central and South Asia will help us to better understand the population dynamics of this species in its native environment. Understanding these dynamics will help demonstrate how insecticide resistance is likely to emerge and spread throughout the vector population and where this species is likely to appear next. Understanding gene flow within a population can give insight into how alleles spread through a species, e.g., resistance associated mutations; and identifying where *An. stephensi* came from will help us understand how it moved and how it could move further.

Overall, this chapter represents the first report of results from the application of WGS to *An. stephensi*. The sample size of this data set is limited but demonstrates the utility of such techniques even when applied to a small number of isolates. The application of population genetics techniques has further illuminated the ancestry of this invasive species and identified markers of insecticide resistance in field isolates. It also resulted in the detection of many SNPs in candidate genes, and the resulting genomic dataset is available for other researchers to utilise in the future.

Chapter 4 outlines the development of an amp-seq assay for *An. darlingi*, and its application to field isolates collected from four sites in the State of Rondônia, Brazil. The 11-amplicon panel covers nine insecticide resistance associated SNPs across four genes (*ace-1, gaba/rdl, GSTe2,* and *vgsc),* as well as two amplicons for speciation and phylogenetic analysis *(cox-1* and *ITS2).* No known insecticide resistance mutations were identified in the 200 samples this panel was applied to. Ten putatively non-synonymous SNPs were found, five of these were in the *ace-1* gene, and the remaining five in the *GSTe2* amplicon. The lack of phenotype data meant that it is not currently possible to determine if these SNPs are associated with insecticide resistance. The absence of any known SNPs is in line with previous studies looking for molecular markers of resistance in *An. darling*[20]*.*

Chapter 5 also focused on *An. darlingi* and had similar results to the previous chapter, in that no known insecticide resistance associated SNPs were identified. The lack of known mutations could be a result of the great genetic distinction observed between South American *Anopheles spp.*, and African/Asian *Anopheles*[21]. There may be different molecular markers in *An. darlingi* that indicate resistance in this species, similar to what is observed in *An. funestus*.

An. funestus and *An. gambiae* are relatively genetically similar in the global context of *Anopheles* species, and yet the *kdr* mutations which play such a large role in *An. gambiae* are absent in *An. funestus*[22,23]. This may be the case with *An. darlingi* and *An. albimanus,* where several known SNPs have been identified in *An. albimanus*, but remain absent in *An. darlingi* despite their similar geographical ranges and close genetic relationship[24].

Both *An. darlingi* and *An. funestus* had no signals of selection ongoing in the *vgsc,* but while *An. funestus* demonstrated a high diversity in the *vgsc* the reverse was true for *An. darlingi*[23,25]. Typically, the *vgsc* gene is highly conserved due to its critical function, so *An. darlingi* reflects what is expected for *Anopheles* spp mosquitoes. The existence of frontier regions was believed to be contributing to the lack of resistance markers in *An. darlingi,* the high gene flow in these regions was hypothesised to be keeping resistance associated SNPs at low enough frequencies they were undetected[26,27]. Conversely, in *An. funestus* the lack of gene flow between central and southern African populations was thought to be restricting resistance mutations from moving between populations[23].

It was previously believed that the lack of insecticide resistance associated mutations identified in *An. darlingi* was a result of the limited studies investigating this. However, recent large-scale studies have also failed to find any of these known SNPs, these studies also lacked phenotypic data so it may be the mosquitoes tested were completely susceptible to insecticides hence the lack of molecular markers[25,28]. But it is important to consider that these mutations just may not play a role in resistance in *An. darlingi.* To identify molecular markers of insecticide resistance in *An. darlingi*, further studies are needed with

phenotyped isolates to identify mutations, INDELs, and copy number variants exclusively present in with susceptible or resistant isolates.

The phylogenetic markers in *cox-1* and *ITS2* showed varying utility to explore genetic diversity, but both loci revealed low nucleotide diversity. *ITS2* showed less genetic distinction between the Brazilian isolates, and *An. darlingi* sequences from other South and Central American countries. Whereas c*ox-1* showed more distinction between *An. darlingi* isolates from other countries on a phylogenetic tree, but still demonstrated low genetic diversity. However, the *cox-1* gene was successfully used to speciate a sample as an *An. peryassui* isolate, which had been incorrectly identified as *An. darlingi* during morphological examination.

The use of WGS on *An. darlingi* in **Chapter 5** focussed more on the identification of genomic regions under selection, to potentially identify genes of interest that may be involved in insecticide resistance. A total of 31 *An. darlingi* were successfully sequenced, including eight colony samples, and 23 field isolates from Candeias dos Jamari, State of Rondônia in Brazil. Again, the lack of phenotype data available for these isolates meant associating any positions under selection with insecticide resistance is not possible. However, it is possible to identify new genes or other genes that have previously been associated with resistance in other vector species, to gain insight and provide targets for future studies. Five candidate genes were identified as under some form of selection (directional or balancing) from integrated haplotype statistic (iHS) or Tajima's D analysis. One of these genes was the cytochrome P450 *CYP307a1*. This gene has previously been associated with pyrethroid and DDT resistance in *An. funestus,* and was also found to be under selection in the Indian *An. stephensi* field isolates in **Chapter 3**[29]. This gene requires further investigation as to its potential role in insecticide resistance in both these species. The lack of any selection ongoing in the four key genes known to be associated with insecticide resistance (*ace-1, gaba/rdl, GSTe2,* and *vgsc*) could suggest that these loci are not involved in insecticide resistance in *An. darlingi*. However, a similar lack of selection was observed in these genes in **Chapter 3**, where the *kdr*-L1014F SNP was identified, so this is unlikely.

Of the 10 missense SNPs identified in **Chapter 4**, five were identified in these isolates (S674N in *ace-1,* and D128Q, T166I, T179I, and T205A in *GSTe2)*. These variants all had high genotype frequencies, varying between 13% and 100%, in the amp-seq data related to the samples from Candeias dos Jamari. Whereas the other five non-synonymous SNPs not found in the WGS data were all present at frequencies ~7% in the amp-seq samples. The number of isolates undergoing WGS may be too small to detect these lower frequency variants.

It was difficult to hypothesis about the molecular landscape of these isolates due to the minimal data available for these samples specifically, but also the lack of research on *An. darlingi* in general. We still did not find any known insecticide resistance mutations in this sample cohort using amp-seq or WGS, and neither have other capillary sequencing studies conducted since. The lack of selection ongoing around any of the key genes associated with insecticide resistance could indicate they aren't involved in resistance in this species. But without the phenotypic data there is a strong chance the isolates tested here were completely susceptible, so no assumptions can be made around this lack of selection.

I also applied other population genetics techniques, such as phylogenetic methods, principal component analysis, and admixture ancestry analysis. Limited genetic diversity was observed within the colony and field isolates, with only two ancestral populations identified. But the populations were genetically distinct from each other, separating clearly on the phylogenetic tree, PCA, and admixture plot. Given the geographical proximity of the field isolate collection site to where the colony isolates were originally collected (<24km), 3 years prior, more genetic similarity between the two populations was expected. The extensive inbreeding between colony isolates could contribute to the distinction. Additional WGS of field isolates from Porto Velho, State of Rondônia would be needed to truly understand the underlying population dynamics at play. The lack of WGS data available for *An. darlingi* limits any conclusions that can be drawn from this data. My results represent two small sample sets from State of Rondônia, and sampling *An. darlingi* from geographically distant regions, would give greater insight into the population dynamics of this species. However, this data represents the first steps towards fully understanding the

genomic landscape of *An. darlingi* and is the first instance of WGS being applied to this important South American vector.

My final chapter, **Chapter 6,** involves the development of the final *Anopheles* species specific amplicon panel for *An. funestus*. The amp-seq assay was then applied to isolates from two villages in eastern DRC. The 17-amplicon panel covers six genes (*ace-1, CYP6P4, CYP6P9a, gaba/rdl, GSTe2, and vgsc)* associated with insecticide resistance in *An. funestus,* as well as three further loci (*cox-1, ITS2,* and *mt-ND5)* for speciation and phylogenetic analysis. Two known markers for insecticide resistance were identified, the N485I SNP in the *ace-1* gene that results in bendiocarb resistance, and a 2bp deletion in the *CYP6P9a* gene, a molecular marker for pyrethroid resistance. The original detection assay designed for the deletion was a restriction-fragment length polymorphism (RFLP) assay, whose implementation can be a complicated and laborious process[1]. My results demonstrate that amp-seq can be used as a high-throughput method to screen many samples simultaneously and is an alternative to RFLP for large molecular surveillance of vector populations. Data on some of these samples was previously analysed using qPCR and RFLP, and the results published[30]. The two bp *CYP6P9a* INDEL was identified at a slightly higher frequency in the amp-seq data (91.3% vs 84.0%). This may be due to a higher proportion of Tushunguti isolates being sequenced, where the resistant genotype was higher than in Tchonka (98% vs 82%). It could also be the result of increased sensitivity of the amp-seq assay compared to capillary sequencing.

There were a further 17 missense SNPs identified in the amplicons, of which none appear to have been reported previously. The identification of the two-insecticide resistance associated mutations in such high frequencies, implies the mosquitoes tested here are at least pyrethroid resistant, but further investigation of the new non-synonymous SNPs identified here is needed to confirm their involvement with phenotypic resistance. The expansion of this amp-seq panel to include *CYP* genes demonstrates its potential for adaptation in the future as more markers of insecticide resistance become known.

Both the mitochondrial genes, *cox-1* and *mt-ND5,* gave the greatest insight into *An. funestus* population structure and species identification. Mitochondrial genes do generally display the highest amount of

variation and genetic diversity, in comparison to nuclear or ribosomal genes such as *ITS2.* The mt-ND5 amplicon gave the greatest differentiation between the DRC isolates and those from other countries. Whilst *cox-1* was able to identify three isolates as *An. coustani*, an understudied vector in the region, one further sample closely identified with *An. gambiae* (99%) and *An. arabiensis* (98.5%). This sample could be an example of species introgression occurring between the vectors. They have overlapping geographical distributions, so the likelihood of interspecies breeding and interchange of genetic elements is high[31,32]. Comparative genomics with WGS data would be needed to confirm this. The clear distinction of the Anfun71 isolate on the *cox-1* phylogenetic tree suggests that this isolate was genetically distinct from the other *An. funestus* sequences. The similarity between *An. gambiae* and *An. funestus cox-1* sequences is ~99% on average. But NCBI BLAST analysis for this sample demonstrated a 91.9% similarity with *An. funestus* compared with to a 99% identity with *An. gambiae*, suggesting it was likely just misidentified morphologically. Overall, the panel demonstrated increased sensitivity compared to capillary sequencing and provided an alternative diagnostic assay to detect known insecticide resistance markers. The main aim of this study was to determine whether it was possible to identify more complex markers such as the 2bp *CYP6P9a* insertion that typically requires more complicated detection methods than standard PCR. Amp-seq appears more sensitive than the standard RFLP assay and requires less laborious laboratory techniques. The results of this study are in line with my hypotheses regarding the presence of these markers, and it also detected the N485I *ace-1* alteration, alongside other novel SNPS.

Limitations

Overall, my thesis demonstrates the utility of next generation sequencing in detecting known molecular markers of insecticide resistance, and its potential to identify novel markers also. However, there are still extensive limitations regarding the implementation of such techniques in the field, the accessibility of WGS in malaria-endemic countries is often limited due to reasons discussed above. But also, the bioinformatics skills and high-performance computing clusters required for data analysis can be difficult to access in these settings. Amp-seq avoids the latter of these issues as the data generated is

substantially smaller, and PCR-based methods of detecting SNPs associated with insecticide resistance are routinely conducted in reference laboratories making it a more feasible alternative. But the issue of sequencer accessibility remains an issue.

Despite the PCR-based methods of insecticide resistance associated SNPs being in use for decades, phenotypic bioassays remain the gold standard for resistance detection. There are numerous reasons for this, including a lack of species-specific knowledge of molecular markers of resistance, as seen with *An. darlingi*. Simultaneously even if SNPs previously linked to resistance are found, this does not always correlate with mosquito phenotype. Mutations such as *kdr*-L1014 and *ace-1* G119S have been found in phenotypically susceptible mosquitoes, and conversely, they can be completely absent in resistant mosquitoes[20,33]. So, whilst these SNPs may be useful in surveillance, they are not diagnostic markers of phenotypic resistance.

The main limitation of this thesis is the lack of phenotypic data available for all isolates, without this information it is impossible to draw conclusions about how any of the genomic alterations identified affect the samples resistance status. For *An. funestus,* this data was less crucial due to the pre-existing information available about the insecticide resistance markers present in the sample cohort. In *An. stephensi* the wide-spread resistance to carbamates, organophosphates, organochlorines, and pyrethroids identified in isolates taken from the same village the isolates examined here would indicate the samples were likely also resistant[34]. However, we cannot be certain, so the signals of selection and putatively novel SNPs identified in WGS and amp-seq analysis cannot be assumed to be related to phenotype. The presence of insecticide susceptible colony isolates for *An. stephensi* were a useful control group to help counteract this uncertainty.

However, for *An. darlingi* the susceptibility of even the laboratory-reared colony isolates was unknown. Metadata indicated the isolates collected from the Porto Velho sites had been exposed to insecticides, but exposure does not automatically indicate resistance. As such the importance of the information gleaned from the *An. darlingi* analysis remains unknown.

Conclusions

This thesis presents the application of genomics to three vector species, two of which have been historically understudied. It includes the development of three species specific amp-seq assays for the high-throughput detection of insecticide resistance, and the analysis of population dynamics. This thesis also demonstrates the utility of applying WGS to characterise the genomic landscape of two species of *Anopheles* mosquitoes for the first time, creating a blueprint for the future expansions of such studies.

WGS and amp-seq have the potential to have major impacts on our understanding of vectors and contribute to the fight against malaria. Such techniques have been used to generate large amounts of data about malaria parasites, and *An. gambiae* complex mosquitoes. This thesis has substantially increased the amount of genomic data available for both *An. darlingi* and *An. stephensi*, enabling future investigations into these species. The successful application of amp-seq to both *An. stephensi* and *An. funestus* demonstrates the utility of this method for high-throughput detection of insecticide resistance. The panel still resulted in an increase in genomic data for *An. darlingi* and insights into the species' population dynamics. The lack of known markers found in this sample set, or any others tested so far, indicates more studies are needed in tandem with phenotypic typing, to truly understand the genomic underpinnings of insecticide resistance in this species.

Overall, the WGS studies here have given insight into the population dynamics of two historically understudied malaria vectors. Whilst without key phenotypic data, the methodology used here to analyse WGS data, provides a framework for future work that could identify important genetic markers of resistance. These novel markers could also be incorporated into amp-seq assays used for insecticide resistance surveillance by reference laboratories.

Despite the limitation discussed above, regarding the sometimes-tenuous link between mosquito genotype and phenotype, the identification of such markers could be crucial in optimising insecticide usage in malaria control policy.

To summarise, vector genomics is a rapidly expanding field, and further wide-scale WGS studies are needed for more *Anopheles* species, to understand each species unique genomic landscape and molecular evolution. Such insights can be used to inform high-throughput assays, such as amp-seq, for the detection of insecticide resistance markers as demonstrated here.

Future Perspectives

Genomic data has vastly improved our understanding of many infectious diseases, including malaria. The application of WGS to *Plasmodium* parasites has helped identify drug resistance markers and informed control methods to reduce disease burden. Malaria is a multi-faceted disease, that requires transmission interventions for various parts of its lifecycle. Vector control is a key component of breaking malaria transmission, and the need for efficient control methods has never been more crucial.

However, one concern for the future of insecticide-based vector control is the global emergence of resistance. Other control methods such as sterile male release, *Wolbachia spp.* infection, or mass administration of ivermectin, can help share the burden[35–37]. However, none are as established or low cost as IRS or the use of LLINs. The development of new insecticide classes such as pyrroles and neonicotinoids will also reduce this pressure on traditionally used chemicals[38,39]. Unlike typically used insecticide classes like pyrethroids and organophosphates, the chances of mosquitoes developing cross-resistance to these new insecticides is unlikely. But even with these novel compounds, resistance can still emerge. This worrying outcome is why large scale high-throughput surveillance methods are so important moving forward, to rapidly identify the emergence of new mutations associated with resistance.

As discussed above, the need for WGS studies with large sample numbers from diverse locations will help understand the molecular drivers of resistance, when paired with phenotypic data.
It will be crucial to conduct these studies in a timely manner, so such drivers can be quickly identified with the resulting phenotypic alteration to inform malaria control programmes.

It will also help us understand the population dynamics of the *Anopheles* species and illuminate how resistance is likely to spread throughout the population. **Chapters 3 and 5** outlined how such data could be successfully utilised. The former presented a first of its kind study to look for resistance markers in *An. darlingi.* Whilst the latter, presented an ancestry analysis of invasive *An. stephensi* to probe what introductory invasion events could have taken place, as well as understand where this vector could spread in the future.

The further investigation of putatively novel SNPs identified in this thesis in phenotyped isolates, would help indicate whether these mutations result in resistance. The pairing of bioassay and genomic data is key to moving forward with the methodologies developed in this thesis. The bioassay data would create two sample cohorts – resistant and susceptible, which could allow for the identification of SNPs in genes of interest associated with either phenotype, identified using amp-seq or WGS. The application of WGS to a sample set like this would mean analysis could be done to identify genomic regions under selective pressure associated with insecticide resistance. A further step could also be the incorporation of RNA-seq technology to identify gene expression changes associated with metabolically mediated resistance. For *An. darlingi* and *An. stephensi* no such studies exist, and so are an opportunity to expand our knowledge of the genomic landscapes of these species. Particularly for *An. darlingi,* this could help elucidate the mechanisms by which resistance occurs in this mosquito e.g., if target site mutations play a reduced role in how this species develops resistance.

The amp-seq assay I developed here for *An. darlingi,* is now being used by collaborators in Brazil, in combination with bioassays and other entomological data. This implementation will support the understanding of the molecular basis of insecticide resistance in this species and local malaria control programmes. The application to field samples by scientists in malaria endemic countries is the desired outcome of the amp-seq assays designed in this thesis.

Moving forward, the creation of a *pan-Anopheles* amp-seq assay, incorporating well characterised resistance markers commonly observed across *Anopheles* species, would be suitable for more well-studied species. The development of a such an assay would remove the need to morphologically identify the *Anopheles* species first, before selecting the correct assay to implement. In turn improving the

accessibility and cost-effectiveness of such surveillance. The inclusion of molecular speciation markers such as *cox-1* is an alternative approach to morphological assessments, and amplicons for known molecular markers should be included in the amplicon assays designed in this study. This would also allow for applications from environmental DNA (eDNA), where water collected from *Anopheles* breeding grounds could be tested for insecticide resistance markers and species identification. This has been previously successfully implemented with specific *Anopheles spp.,* using Illumina sequencing platforms (Appendix 1)[15].

The variations in intron lengths in *Anopheles spp.* would mean the creation of such a panel would likely need longer read sequencing (e.g., MinION). I have begun the development of such a panel, for the four key genes: *ace-1, gaba/rdl, GSTe2,* and *vgsc,* following the markers in the amp-seq panels described (**Chapter 2).** Testing of this multi-species target panel has begun on a range of species such as *An. funestus, An. darlingi, An. stephensi,* and *An. gambiae.*

A major strength of MinION technology is its portability and potential for in-field sequencing of vectors. Another use is the possibility of real-time analysis of DNA sequences generated on the platform, which could be used to identify insecticide resistance SNPs within the 48-hour sequencing run[40,41]. Paired with this could be the use of adaptive sampling, where only genes of interest are sequenced[42]. An input FASTA file containing the candidate genes instructs the nanopore pores to reject any sequences that are not included in the file, and results in increased coverage of genomic regions of interest.

One of the more interesting questions remaining is, would it be possible to make these amp-seq assays field applicable? These assays have been demonstrated to be run on the MinION sequencing platform, as well as Illumina. So as such, with a crude DNA extraction and access to a PCR machine this assay could be run in field setting. The next key question is, would it be possible for someone with no bioinformatics knowledge to analyse the data? Currently the script to analyse the data requires minimal programming experience, however, the next step would be the development of a web-based graphic interface that analyses the data with no bioinformatics knowledge necessary. The same thing has been developed for malaria and requires only fastq files as input and will output the report with detected variants.

These methods, particularly the multispecies amp-seq panel, in combination with the MinION sequencing platform, can facilitate the molecular surveillance of insecticide resistance, and could be implemented in malaria endemic regions.

Whilst there is still plenty of work to do, this thesis represents some of the first steps into a brave new world of the application of genomics to malaria vectors, helping to unravel the complexities of the world's deadliest animal.

The further development of novel detection assays, and in-depth investigations of the genomic landscape of insecticide resistant mosquitoes will help usher in a new age of genomics-based vector surveillance, to inform targeted vector control strategies that move us out of the dark ages of this devasting parasitic disease, and into a world free of malaria.

References

1. Weedall, G. D. *et al.* A cytochrome P450 allele confers pyrethroid resistance on a major African malaria vector, reducing insecticide-treated bednet efficacy. *Sci. Transl. Med.* **11**, eaat7386 (2019).

2. 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).

3. 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. *Mol. Ecol.* **30**, 5303–5317 (2021).

4. Consortium, A. gambiae 1000 G. *et al.* Genetic diversity of the African malaria vector Anopheles gambiae. *Nature* **552**, 96–100 (2017).

5. Weedall, G. D. *et al.* An Africa-wide genomic evolution of insecticide resistance in the malaria vector Anopheles funestus involves selective sweeps, copy number variations, gene conversion and transposons. *PLOS Genet.* **16**, e1008822 (2020).

6. The Anopheles gambiae 1000 Genome 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).

7. St. Laurent, B. *et al.* Population genomics reveal distinct and diverging populations of An. minimus in Cambodia. *Commun. Biol.* **5**, 1–10 (2022).

8. Das, S. *et al.* Fine-scale spatial distribution of deltamethrin resistance and population structure of Anopheles funestus and Anopheles arabiensis populations in Southern Mozambique. *Malar. J.* **22**, 94 (2023).

9. 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 Trop.* **139**, 39–43 (2014).

10. Carter, T. E. *et al.* Genetic diversity of Anopheles stephensi in Ethiopia provides insight into patterns of spread. *Parasit. Vectors* **14**, 602 (2021).

11. Tadesse, F. G. *et al.* Anopheles stephensi Mosquitoes as Vectors of Plasmodium vivax and falciparum, Horn of Africa, 2019. *Emerg. Infect. Dis.* **27**, 603–607 (2021).

12. Collins ID, 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).

13. 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).

14. Samake, J. N. *et al.* Detection and population genetic analysis of kdr L1014F variant in eastern Ethiopian Anopheles stephensi. *Infect. Genet. Evol.* **99**, 105235 (2022).

15. Kristan, M. *et al.* Towards environmental detection, quantification, and molecular characterization of Anopheles stephensi and Aedes aegypti from experimental larval breeding sites. *Sci. Rep.* **13**, 2729 (2023).

16. 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).

17. Liu, N., Feng, X. & Qiu, X. RDL mutations in Guangxi Anopheles sinensis populations along the China–Vietnam border: distribution frequency and evolutionary origin of A296S resistance allele. *Malar. J.* **19**, 23 (2020).

18. Bloomquist, J. R. GABA and Glutamate Receptors as Biochemical Sites for Insecticide Action. in *Biochemical Sites of Insecticide Action and Resistance* (ed. Ishaaya, I.) 17–41 (Springer, Berlin, Heidelberg, 2001). doi:10.1007/978-3-642-59549-3_2.

19. Djogbénou, L., Labbé, P., Chandre, F., Pasteur, N. & Weill, M. Ace-1 duplication in Anopheles gambiae: a challenge for malaria control. *Malar. J.* **8**, 70 (2009).

20. Orjuela, L. I. *et al.* Absence of knockdown mutations in pyrethroid and DDT resistant populations of the main malaria vectors in Colombia. *Malar. J.* **18**, 1–9 (2019).

21. Neafsey, D. E. *et al.* Highly evolvable malaria vectors: The genomes of 16 Anopheles mosquitoes. *Science* **347**, 1258522 (2015).

22. Bass, C. *et al.* Detection of knockdown resistance (kdr) mutations in Anopheles gambiae: a comparison of two new high-throughput assays with existing methods. *Malar. J.* **6**, 111 (2007).

23. Irving, H. & Wondji, C. S. Investigating knockdown resistance (kdr) mechanism against pyrethroids/DDT in the malaria vector Anopheles funestus across Africa. *BMC Genet.* **18**, 76 (2017).

24. Lol, J. C. *et al.* Molecular evidence for historical presence of knock-down resistance in Anopheles albimanus, a key malaria vector in Latin America. *Parasit. Vectors* **6**, 268 (2013).

25. Loureiro, A. C. *et al.* Molecular diversity of genes related to biological rhythms (period and timeless) and insecticide resistance NaV and ace-1in Anopheles darlingi. *Mem. Inst. Oswaldo Cruz* **118**, e220159 (2023).

26. Hiwat, H. & Bretas, G. Ecology of Anopheles darlingi Root with respect to vector importance: a review. *Parasit. Vectors* **4**, 177 (2011).

27. Vezenegho, S. *et al.* Variation in Pyrethroid Resistance Phenotypes in Anopheles darlingi in an Area with Residual Malaria Transmission: Warning of Suspected Resistance in French Guiana. *Am. J. Trop. Med. Hyg.* **1**, (2022).

494 28. Bickersmith, S. A. *et al.* Mutations Linked to Insecticide Resistance Not Detected in the Ace-1 or
495 VGSC Genes in Nyssorhynchus darlingi from Multiple Localities in Amazonian Brazil and Peru. *Genes*
496 **14**, (2023).

497 29. Sandeu, M. M., Mulamba, C., Weedall, G. D. & Wondji, C. S. A differential expression of pyrethroid
498 resistance genes in the malaria vector Anopheles funestus across Uganda is associated with
499 patterns of gene flow. *PLoS ONE* **15**, e0240743 (2020).

500 30. Bandibabone, J. *et al.* Investigating molecular mechanisms of insecticide resistance in the Eastern
501 Democratic Republic of the Congo. *Malar. J.* **20**, 464 (2021).

502 31. Sinka, M. E. *et al.* The dominant Anopheles vectors of human malaria in Africa, Europe and the
503 Middle East: occurrence data, distribution maps and bionomic précis. *Parasit. Vectors* **3**, 117
504 (2010).

505 32. Fontaine, M. C. *et al.* Extensive introgression in a malaria vector species complex revealed by
506 phylogenomics. *Science* **347**, 1258524 (2015).

507 33. Ahmad, M. *et al.* Status of insecticide resistance in high-risk malaria provinces in Afghanistan.
508 *Malar. J.* **15**, 98 (2016).

509 34. WHO. Global Malaria Programme: Malaria Threats Map. https://www.who.int/teams/global-
510 malaria-programme/surveillance/malaria-threats-map.

511 35. Hughes, G. L., Koga, R., Xue, P., Fukatsu, T. & Rasgon, J. L. Wolbachia Infections Are Virulent and
512 Inhibit the Human Malaria Parasite Plasmodium Falciparum in Anopheles Gambiae. *PLOS Pathog.* **7**,
513 e1002043 (2011).

514 36. Alphey, L. *et al.* Sterile-Insect Methods for Control of Mosquito-Borne Diseases: An Analysis.
515 *Vector Borne Zoonotic Dis.* **10**, 295–311 (2010).

516 37. Foy, B. D. *et al.* Efficacy and risk of harms of repeat ivermectin mass drug administrations for
517 control of malaria (RIMDAMAL): a cluster-randomised trial. *The Lancet* **393**, 1517–1526 (2019).

518 38. Zoh, M. G. *et al.* Experimental evolution supports the potential of neonicotinoid-pyrethroid
519 combination for managing insecticide resistance in malaria vectors. *Sci. Rep.* **11**, 19501 (2021).

520 39. N'Guessan, R. *et al.* Chlorfenapyr: A pyrrole insecticide for the control of pyrethroid or DDT
521 resistant Anopheles gambiae (Diptera: Culicidae) mosquitoes. *Acta Trop.* **102**, 69–78 (2007).

522 40. Charalampous, T. *et al.* Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower
523 respiratory infection. *Nat. Biotechnol.* **37**, 783–792 (2019).

524 41. Leggett, R. M. *et al.* Rapid MinION profiling of preterm microbiota and antimicrobial-resistant
525 pathogens. *Nat. Microbiol.* **5**, 430–442 (2020).

526 42. Martin, S. *et al.* Nanopore adaptive sampling: a tool for enrichment of low abundance species in
527 metagenomic samples. *Genome Biol.* **23**, 11 (2022).

528