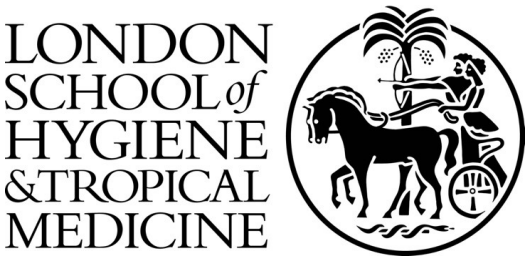


Utilising genomic approaches to explore genetic diversity and insecticide resistance in *Aedes aegypti* populations

Emma Collins

Thesis submitted in accordance with the requirements for the degree of

Doctor of Philosophy of the University of London

2025

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

Funded by The Medical Research Council

Research group affiliation(s): Prof. Susana Campino
Prof. Taane Clark

Declaration

I, Emma Collins, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I can confirm that this has been indicated.

Signed: Emma Collins Date: 06/01/2025

COVID-19 Impact statement

My PhD was significantly disrupted by the COVID-19 pandemic, government lockdowns and related laboratory equipment supply chain issues. I started my PhD in September 2019 and had little time between then and the government-mandated lockdown from the 23rd of March 2020. This interrupted my laboratory work as it prevented primary data collection which forms a critical part of my PhD. The LSHTM laboratory remained closed until June 2020, however, even after reopening there were restrictions on the laboratory access and use, which presented ongoing challenges to performing the work required. Additionally, delays in receiving laboratory consumables (as COVID-related supplies became a priority for the relevant providers) persisted through 2020 and into 2021. I had planned to fulfil a travel grant during 2020 for sample collection for the project. Travel restrictions meant I was unable to travel to the proposed research location (The Philippines). As a result, I had to change the location of the project to Puerto Rico and was only able to complete this sample collection in May 2022. This was one of several ways that I tried to mitigate the effects of COVID-19 on my PhD project, along with taking a placement working at WHO on COVID-19 and an interruption of study between May 2021 and April 2022 to work at UKHSA on COVID-19 related activities. This extended the time available to finish the PhD with less disruption, however inevitably this interruption made it challenging to pick it up again the following year. Finally, intangible disruption occurred through lost contact hours with both colleagues and mentors, especially in the formative first year of the PhD. This is difficult to quantify but is undoubtedly fundamental to learning and PhD progress. Despite the disruption, I have completed four chapters, adapted to mitigate delays and self-taught where possible.

Publications and manuscripts

Thesis publications and manuscripts:

Collins, E. L., Phelan, J. E., Hubner, M., Spadar, A., Campos, M., Ward, D., Acford-Palmer, H., Gomes, A. R., Silva, K., Gomez, L. F., Clark, T. G., & Campino, S. (2022). 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), e0010935. https://doi.org/10.1371/JOURNAL.PNTD.0010935

Collins, E., Medina, J., Puerto, Q., Vector, R., Morales, R., Vector, P. R., Moss, S., Acford-Palmer, H., Higgins, M., Phelan, J., Brown, G., & Campino, S. (2024). Profiling insecticide resistance phenotypes and genotypes in Aedes aegypti populations across four regions in Puerto Rico. https://doi.org/10.21203/RS.3.RS-3943783/V1

Spadar, A., **Collins, E.,** Messenger, L. A., Clark, T. G., & Campino, S. (2024). Uncovering the genetic diversity in Aedes aegypti insecticide resistance genes through global comparative genomics. Scientific Reports 2024 14:1, 14(1), 1–12. https://doi.org/10.1038/s41598-024-64007-6

Other manuscripts with contributions:

Campos, M; Phelan, J; Spadar, A; **Collins, E**; Gonçalves, A; Pelloquin, B; Vaselli, N. M.; Meiwald, A; Clark, E; Stica, C; Orsborne, J; Sylla, M; Edi, C; Camara, D; Mohammed, A. R.; Afrane, Y A; Kristan, M; Walker, T; Gomez, L. F.; Messenger, L. A; Clark, T. G.; Campino, S; (2022) High-throughput barcoding method for the genetic surveillance of insecticide resistance and species identification in Anopheles gambiae complex malaria vectors. Scientific Reports, 12 (1). 13893-. ISSN 2045-2322 DOI: https://doi.org/10.1038/s41598-022-17822-8

Kristan, M.; Acford-Palmer, H; Campos, M. O.; **Collins, E. L.;** Phelan, J; Portwood, N. M.; Pelloquin, B; Clarke, S; Lines, J; Clark, T. G.; Walker, T; Campino, S; Messenger, L. A; (2023) Towards environmental detection, quantification, and molecular characterization of Anopheles stephensi and Aedes aegypti from experimental larval breeding sites. Scientific Reports, 13 (1). 2729-. ISSN 2045-2322 DOI: https://doi.org/10.1038/s41598-023-29657-y

Acford-Palmer H, Phelan JE, Tadesse FG, Kristan M, **Collins E**, Spadar A, Walker T, Bousema T, Messenger LA, Clark TG, Campino S; (2023) Identification of two insecticide resistance markers in Ethiopian *Anopheles stephensi* mosquitoes using a multiplex amplicon sequencing assay. Scientific Reports, 13(1):5612, DOI: 10.1038/S41598-023-32336-7

Moss S, Pretorius E, Ceesay S, Hutchins H, da Silva ET, Ndiath MO, Jones RT, Vasileva H, Phelan J, Acford-Palmer H, **Collins E**, Rodrigues A, Krishna S, Clark TG, Last A, Campino S; (2024) Genomic surveillance of Anopheles mosquitoes on the Bijagós Archipelago using custom targeted amplicon sequencing identifies mutations associated with insecticide resistance. Parasites & Vectors, 17(1):10, DOI: 10.1186/S13071-023-06085-5

ABSTRACT

Aedes aegypti is a mosquito species responsible for considerable global mortality and morbidity, through its role as the vector of many arboviruses, including dengue, Zika and Chikungunya virus. For *Ae. aegypti* and many other mosquito species, the predominant method of vector control is insecticide use. However, their widespread deployment over the last hundred years has led to the inevitable rise of insecticide resistance. Insecticide resistance in *Ae. aegypti* has been documented in most countries where the species is endemic. This thesis uses genomic approaches to develop tools that provide insight into the diversity of global *Ae. aegypti* populations and analyse the genes and mutations associated with insecticide resistance.

Detection of resistance has historically relied on time-consuming phenotypic bioassays, however, in recent years focus has shifted to molecular assays to objectively identify resistance markers. Genomic approaches can inform on mutations that confer resistance and the population structure and diversity within the species. I have developed a barcoded multi-target amplicon sequencing panel for high throughput detection of single nucleotide polymorphisms (SNPs) in gene regions linked to insecticide resistance in *Ae. aegypti* (*voltage-gated sodium channel (vgsc), resistance to dieldrin (rdl), acetylcholinesterase-1 (ace-1)* and later *glutathione-S transferase 2 (GSTe2*)). This panel can be used for the surveillance of resistance alongside traditional bioassays. This methodology has been implemented on multiple populations, including *Ae. aegypti* sourced from Cabo Verde and Puerto Rico and has identified previously reported insecticide resistance SNPs as well as additional putatively novel missense SNPs.

Utilising whole genome sequencing data can provide further insights into ongoing selective pressures due to insecticide use and uncover population dynamics. I carried out a study employing comparative genomics to assess differences in the main insecticide resistance associated genes (*vgsc, rdl, ace-1* and *GSTe2)* in 729 *Ae. aegypti* sequences from 15 countries. This led to the identification of 747 missense mutations, of which five have previously been associated with insecticide resistance. Combining this genomic data with available phenotype data indicates these profiles are variable, and further investigation into the functional link between mutations and phenotype is required. Creating this large catalogue of genotype data along with the geographic distribution will help to identify resistance drivers and aid monitoring and surveillance efforts in *Ae. aegypti.*

Analysis was expanded by whole genome sequencing 33 *Ae. aegypti* from a Puerto Rican population and comparing the results to 215 other publicly available global sequences. This analysis highlighted similarities and differences between the Puerto Rican and other global populations with respect to population structure, and genome-wide nucleotide diversity and selection markers. I identified over 281,000 missense SNPs across all populations including four insecticide resistance SNPs (*vgsc* V410L, V1016I, F1534C; *rdl* A301S). Signals of selection were found in genes associated with insecticide resistance, including gamma-aminobutyric acid receptor subunit alpha, glutathione S-transferases and cytochrome P450s.

This thesis underscores the focal role of genomic techniques and their analysis in enhancing our understanding of insecticide resistance, which can subsequently aid and inform vector control programmes. The identification of mutations known to be associated with resistance is important for assessing vector profiles, and the reporting of candidate novel putative mutations can launch follow-up validation work, including functional studies. Through the implementation of these techniques, surveillance and control can be improved to disrupt transmission and subsequently alleviate the huge global burden of vector-borne disease.

Acknowledgements

I would like to thank my supervisors Professor Susana Campino and Professor Taane Clark for their insight, advice and mentoring throughout my PhD. During my studies, I was also mentored by Louisa (Lulu) Messenger and Mojca Kristan as well as Jody Phelan and Dan Ward.

I am grateful for the funding I received from the Medical Research Council who supported my scholarship through the LID PhD scheme.

I am grateful for the support and friendship provided by other PhD students and postdocs at LSHTM, namely, the 'hub' group in 380. Ashley, Holly, Gabbie, Leen and Sophie thank you for always listening and believing there is no such thing as a stupid question.

Friends outside 380 – you know who you are, I couldn't have done it without all the dinners, walks and debriefs, thank you for always listening to my rants and the never-ending support.

I thank my family and especially my partner for their endless support and patience while I worked to complete my PhD. Will, I'm forever grateful that you've been there every step of the way. Mum, Dad, Soph, and Grandma this might be the only bit you understand but thank you for everything, couldn't have done it without you.

Table of Contents

List of Abbreviations

Abbreviation	Description
ACE	Acetylcholinesterase
ADP	Adenosine Diphosphate
Arbovirus	Arthropod-borne Virus
ATP	Adenosine Triphosphate
CE	Carboxylesterases
CHIKV	Chikungunya Virus
CNV	Copy Number Variation
CRISPR	Clustered Regularly Spaced Short Palindromic Repeats
DDT	Dichlorodiphenyltrichloroethane
DENV	Dengue Virus
DNA	Deoxyribonucleic Acid
GSTE	Glutathione S-transferase
IRS	Indoor Residual Spraying
ISS	Indoor Space Spraying
ITN	Insecticide Treated Net
JEV	Japanese Encephalitis Virus
KDR	Knock Down Resistance
LLIN	Long Lasting Insecticide Net
MFO	Multifunction Oxidases
NCBI	National Centre for Biotechnology Information
NGS	Next Generation Sequencing
PCA	Principal Components Analysis
PCR	Polymerase Chain Reaction
RDL	Resistance to Dieldrin
SLE	St Louis Encephalitis
SNP	Single Nucleotide Polymorphism
TGS	Third Generation Sequencing
ULVS	Ultra-Low Volume Spraying
VGSC	Voltage Gated Sodium Channel
VBD(s)	Vector Borne Disease(s)
WGS	Whole Genome Sequencing
WHO	World Health Organization
WNV	West Nile Virus
ZIKV	Zika Virus

INTRODUCTION

This thesis highlights the pivotal role of genomic techniques and their analysis in advancing our comprehension of insecticide resistance in *Aedes aegypti*, thereby supporting and guiding vector control efforts. By employing these methods, surveillance and control strategies can be refined to interrupt transmission, ultimately mitigating the significant global impact of vector-borne diseases.

Vector-borne Diseases

Adult female mosquitos are crucial vectors for a diverse array of pathogens due to their blood-feeding behaviour. Their contribution to the global transmission of vector-borne diseases (VBDs) presents a formidable risk to public health. Data from the World Health Organization (WHO) in 2020 reported more than 700,000 deaths annually attributable to VBDs [1]. Of the over 112 genera and ~3,500 species of mosquito, those implicated in disease transmission are primarily from three genera: *Anopheles, Culex,* and *Aedes* [2]. The seven major VBDs include malaria, lymphatic filariasis, dengue, Japanese encephalitis, yellow fever, and Chagas disease. They are widely distributed, primarily among the tropics and sub-tropics as shown in **Figure 1** [3].

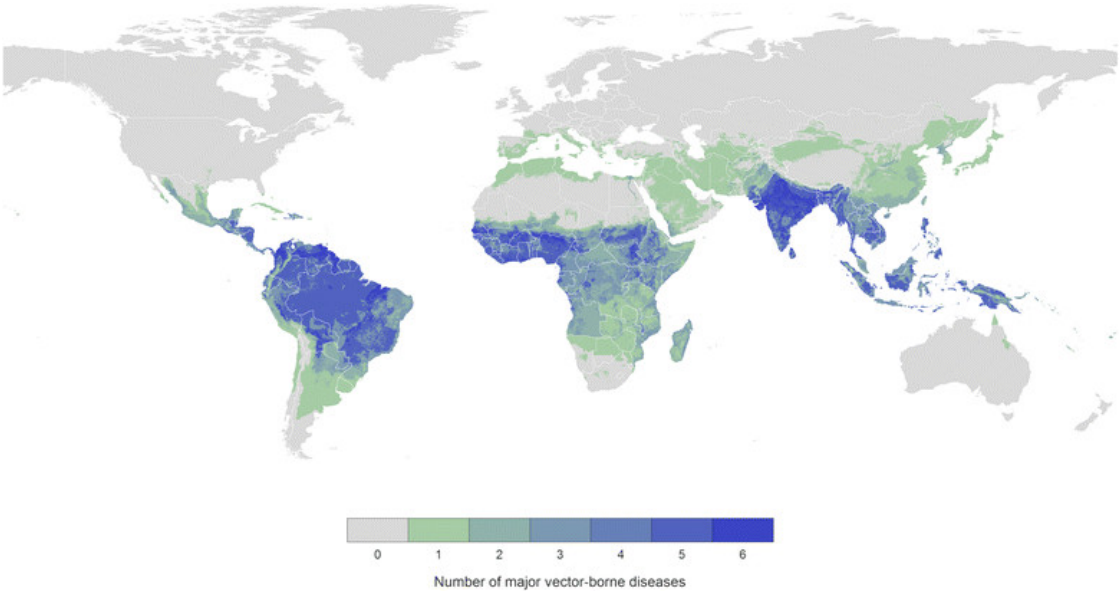


Figure 1. *Global distribution of seven major VBDs: malaria, lymphatic filariasis, leishmaniasis, dengue, Japanese encephalitis, yellow fever, and Chagas disease. Colours indicate the number of VBDs that pose a risk at each 5 × 5 km grid cell. Figure from Golding et al, 2015* [3].

Anopheline mosquitoes (genus *Anopheles*) are distributed across the globe, predominantly in the tropics and subtropics, but also in more northern climes including across Europe. They are the primary vector for several of the seven main VBDs, including malaria, which causes the highest global burden of mortality [4]. In 2022 alone, an estimated malaria prevalence of 249 million cases were recorded

globally [5]. *Culex* mosquitoes have the most widespread geographical distribution of any genus and are found across both tropical and temperate regions worldwide, likely due to their adaptation to human-made environments [6]. *Culex* mosquitos transmit West Nile virus and Japanese encephalitis. Finally, *Aedes* mosquitoes are widely distributed across the tropics and subtropics and transmit a wide array of pathogens, including several arthropod viruses (Arboviruses) including dengue (DENV), Zika (ZIKV) and Chikungunya (CHIKV) viruses, as well as other diseases including yellow fever (YF), Rift valley fever (RVF) and lymphatic filariasis (LF) which contribute substantially to the global burden of disease.

Aedes aegypti is one of the most competent *Aedes* vectors, being highly anthropophilic and susceptible to many pathogens, notably DENV, CHIKV, and ZIKV [7]. The *Ae. aegypti* mosquito is particularly adapted for urban environments, and consequently, its distribution has increased dramatically alongside the spread of urbanisation [8,9]. *Aedes albopictus,* also known as the tiger mosquito, is another highly competent and widely distributed vector, responsible for the transmission of pathogens, principally West Nile Virus (WNV), St Louis Encephalitis (SLE), Japanese Encephalitis Virus (JEV), and lymphatic filariasis.

Close contact with humans paired with their ability to transmit several pathogens make mosquitoes a major threat to human health. Surveillance and control of these vectors is instrumental to improving global public health, especially in tropical and sub-tropical regions.

Mosquito Vector Control Methods

Globally, numerous vector control methods are employed to reduce the transmission of VBDs. These methodologies target different genera behaviours or life stages of vectors, including larval stages, adults, and blood-feeding females [10]. Insecticidal compounds, used for thousands of years to control pests, remain universally employed and can target any life stage to interrupt the transmission of vector-borne diseases [11]. Non-chemical methods include habitat destruction and modification, such as introducing larval predators and preventing the formation of water bodies that could become breeding sites [12]. When water bodies are too large for removal, adding insecticides to these breeding sites is another viable option [13–15]. While larval control strategies are effective against all disease-transmitting mosquito genera, they are limited to situations where larval habitats can be clearly identified which can be challenging in urban areas [16].

Other strategies targeting the adult vector are primarily chemical, such as indoor residual spraying (IRS), indoor space spraying (ISS) and outdoor ultra-low volume spraying (ULVS) [17–20]. IRS involves applying insecticide to the walls of homes, exposing resting vectors to a lethal dose, thereby halting the transmission of any pathogens they carry. ISS involves spraying insecticides inside to target flying mosquitoes that contact the insecticide [21]. Ultra-low volume spraying uses insecticides applied in affected outdoor areas to kill adult mosquitoes [19]. Targeting blood-feeding female mosquitoes before they can infect a human host is another crucial control mechanism. Insecticide-treated bed nets (ITNs) are widely used, particularly in malaria endemic regions [22]. Other non-chemical methods include house screening and traps [23,24]. Household screening prevents mosquitoes from entering homes, while traps attract and kill female mosquitoes searching for places to lay their eggs.

When control programmes are selecting which interventions to implement, it is vital to account for the characteristics and behaviours of the mosquito being targeted. For example, a malaria control programme might use ITNs to protect people sleeping, due to the anthropophilic preference of Anopheline mosquitoes to feed on humans at night [25], but they are thought to be less effective against *Aedes* or *Culex* mosquitoes due to their daytime feeding behaviour. However, there is some evidence they may still have an effect due to their indoor resting behaviour [19,26]. ITNs have been shown to be highly effective in controlling malaria in Africa, with estimates that ITNs have contributed to 68% of averted cases[12]. *Ae. aegypti* mosquitos, typically found in urban settings, are often endophilic, resting inside homes and anthropophilic seeking out humans as their blood meal [13]. Many control programs for *Aedes* utilise outdoor ultra-low volume spraying (ULVS); however, some are sceptical of its efficacy, given *Aedes* mosquitoes' behaviour, and instead favour indoor insecticide application [14,15]. A campaign spraying dichlorodiphenyltrichloroethane (DDT) successfully eradicated *Ae. aegypti* from 19 countries by the 1970s, however emerging and widespread resistance and environmental impacts stopped its use[16]. A mixture of methods can be used in combination where *Anopheles* and *Aedes* are sympatric, however costs of implementing multiple interventions is a major limiting factor in their implementation, especially in low resource regions. All methods have been proven effective in reducing the burden of VBD, dependent on the context, with chemical methods being the most widely employed [10,17,18].

Insecticide Resistance

Insecticides have been a mainstay of VBD control for decades, however there are limitations with their use, firstly from insecticide resistance and secondly from the limited number of insecticides available for vector control. There are currently nine classes of insecticide prequalified by the World Health

Organisation (WHO) for use in vector control: pyrethroids, carbamates, organophosphates, organochlorines, neonicotinoids, pyrroles, butenolides, juvenile hormone mimics and spinosyns (**Table 1**) [27,28]. Pyrethroids, carbamates, organophosphate and organochlorines are most widely used [29]. These insecticides have different mechanisms of action within the mosquito [30]. Except for pyrroles and juvenile growth hormones, the majority of these insecticides act by disrupting the insect nervous system. Neonicotinoids and butenolides are insecticide classes that have more recently been implemented for vector control since their prequalification status by WHO in 2017 [27]. Spinosyns differ from other insecticide as they are derived from the fermentation of two species of *Saccharopolyspora* [31].

Pyrethroids bind to voltage gated sodium channels (sodium channel modulators) in insect nerve axons to prevent the nerve depolarizing after it has fired, causing paralysis and death at sufficient doses [32]. Organophosphates and carbamates are acetylcholinesterase inhibitors, preventing the enzyme from breaking down acetylcholine, a common neurotransmitter that causes a nerve to fire. A neurotransmitter that is not broken down will cause the nerve to continuously to fire, causing death [33]. For most organochlorines, the mode of action is via binding to the gamma-aminobutyric acid (GABA) chlorine complex which, prevents chloride flow into the nerve causing hyperexcitation [34]. However, DDT is an organochlorine with a different mechanism of action similar to pyrethroids targeting the voltage-gated sodium channels [35]. Pyrroles are a novel insecticide class targeting the metabolic respiratory pathways (oxidative phosphorylation) in the mitochondria to disrupt adenosine triphosphate (ATP) production from adenosine diphosphate (ADP) (respiratory pathway modulators) [36–38]. Neonicotinoids and butenolides are used in combination with pyrethroids for both IRS and space spraying [27]. These insecticides work by acting as a competitive modulators of nicotinic acetylcholine receptors, fatally interfering with neural transmission [39]. Finally, spinosyns also disrupt nicotinic acetylcholine receptors, there was also mixed evidence about cross resistance and suggestions it was related to a secondary effect on the GABA receptors [40,41], however alternative evidence has emerged that there is no effect on RDL GABA receptors [42]. Example compounds and their uses are outlined in **Table 1.**

The prequalification of additional insecticide classes, such as pyrroles and neonicotinoids, for vector control has aided options for regions with resistance, however the number of available insecticides is still limited, and resistance has already emerged to neonicotinoids in multiple mosquito species [43–45], decreasing the efficacy of their use against vectors [46]. **Figure 2** indicates how insecticide resistance to the four commonly used classes now has an almost global distribution for both *Aedes* and *Anopheles*

mosquitoes. Given the wide-scale use of insecticides, insecticide resistance has emerged across the globe [47]. To combat this, alternative insecticides are required, and numerous alternative vector control methods are under development. However, these methods, such as genetic manipulation (e.g., CRISPR), bacterial infection of vectors (e.g., Wolbachia) and a sterile insect technique are not yet widely used, although it has been trialled with varying results in a few locations including Australia, Brazil and the Singapore [48–52]. With development being unable to keep up with the increasing insecticide resistance we are heading towards a crisis with a lack of effective control methods.

Table 1. *Insecticide classes and corresponding mechanism of action, example compound and vector control use* [19,21,31]

Insecticide	Primary Mechanism of Action	Example	Use
Pyrethroids	Sodium channel modulators	Deltamethrin, Permethrin	ITNs, IRS
Carbamates	Acetylcholinesterase inhibitors	Bendiocarb, Pirimiphos-methyl	IRS
Organophosphates	Acetylcholinesterase inhibitors	Temephos, Malathion	IRS, larvicide
Organochlorines	GABA-gated channel inhibitors	Dieldrin	IRS
Pyrroles	Respiratory pathway modulators	Chlorfenapyr	ITN (in combination)
Neonicitinoids	Nicotinic acetylcholine receptor competitive modulators	Clothianidin	IRS, Space spraying (in combination)
Butenolides	Nicotinic acetylcholine receptor competitive modulators	Flupyradifurone	Space spraying (in combination)
Juvenile hormone mimics	Developmental growth regulator	Pyriproxyfen	ITN (in combination), larvicide
Spinosyns	Nicotinic acetylcholine receptor allosteric modulators	Spinosad	Larvicide

Mechanisms and Genetics of Insecticide Resistance

The selective pressure from insecticide use on mosquito populations has led to resistance to all commonly used insecticide classes (pyrethroids, carbamates, organochlorines, and organophosphates) [47]. Prevalence and distribution are increasing with ever more countries reporting insecticide resistance (**Figure 2**) [47]. However, the picture is incomplete as many countries do not carry out routine monitoring of resistance or do not report this in a timely manner or are slow to react once

resistance is identified [29]. Moreover, most available data is shared in an open-source manner which can impact the ability for comparison between studies [47,54]. Furthermore, much of the focus has been on *Anopheles*, while far fewer studies have been carried out on *Aedes* or *Culex* vectors; there is therefore a research gap on insecticide research in *Aedes* mosquitoes globally.

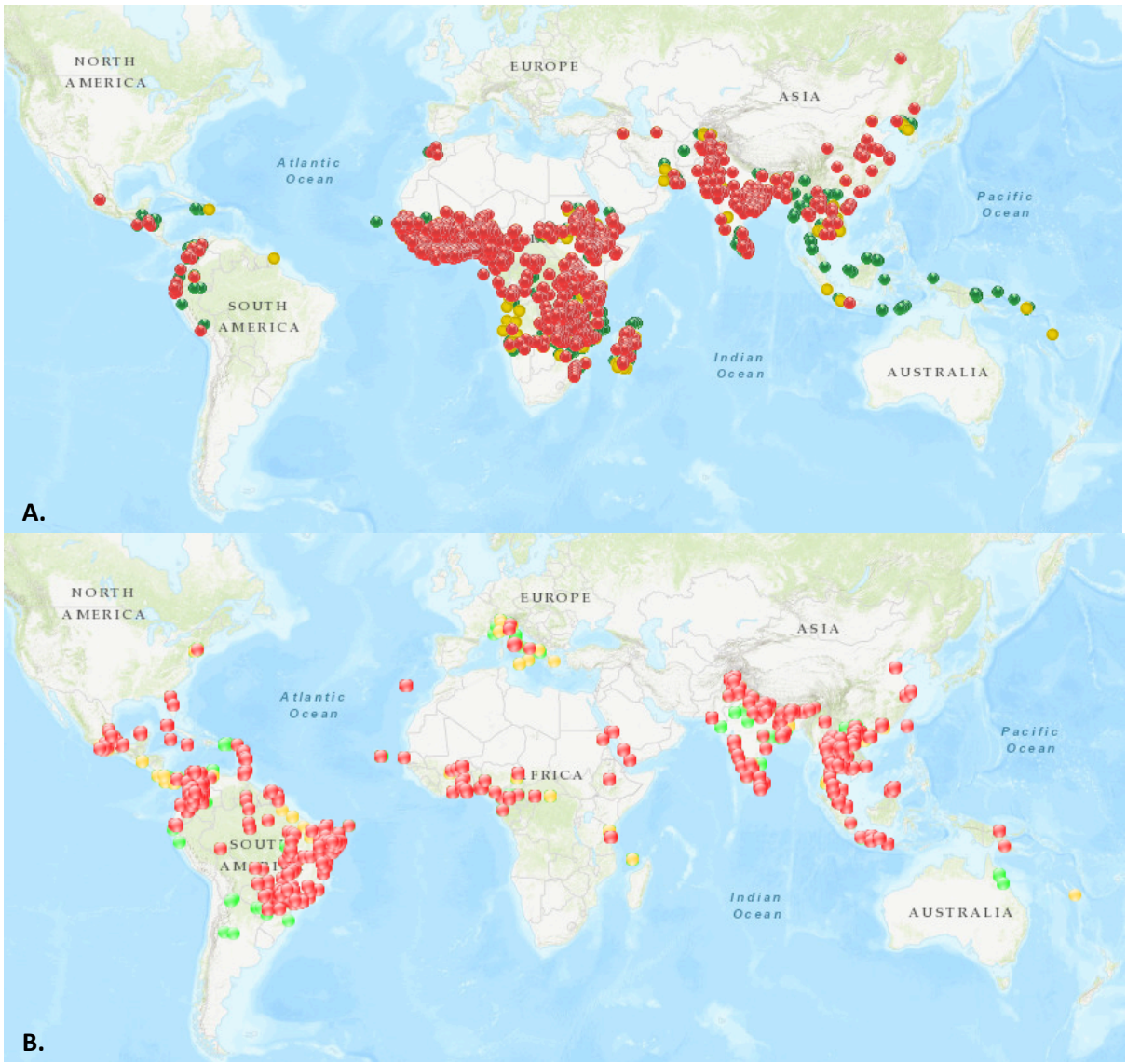


Figure 2. A. *Distribution and results of bioassay data for insecticide resistance in Anopheles spp.* ***B.*** *Distribution and results of bioassay data for insecticide resistance in Ae. aegypti and Ae. albopictus. Red points indicate resistance observed, yellow indicates possible resistance and green points indicate susceptibility. Data covering 2005-2024 from IR Mapper (https://aedes.irmapper.com/, accessed 03/01/24).*

As there are a limited number of safe and cost-effective insecticides available for public health purposes, the reduction in efficacy due to resistance is a major public health threat. Resistance to pyrethroids is a particular concern as they were for many decades the only class of insecticide

recommended for ITNs, however in recent years this has been adjusted with insecticide and synergist combination nets endorsed by the WHO [27,55].

The complexity of insecticide resistance can be seen in the variety of mechanisms, the additive combinations of the mutations, and the interaction between regulatory and coding genes. The main mechanisms of insecticide resistance are target site resistance, metabolic resistance, cuticular resistance and behavioural avoidance (**Figure 3)** [27,56,57]. Target site and metabolic resistance have been found to have a genetic underpinning, which have been explored in several insect species. Target site and metabolic resistance are the most well studied mechanism, and investigations of *Musca domestica* have been key to elucidating the molecular basis of insecticide resistance [58]. A multitude of genomic alterations have been linked to insecticide resistance in agricultural and medical vectors, however, there remains gaps in knowledge and understanding of the interplay between these mechanisms.

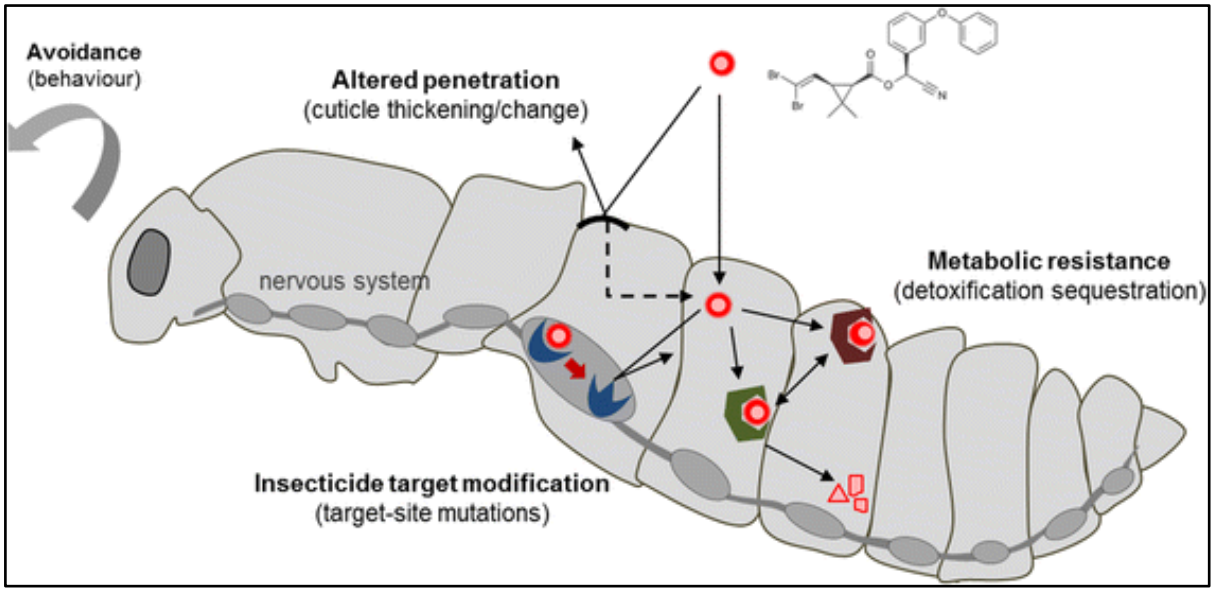


Figure 3. *Illustration of the four mechanisms of insecticide resistance adapted from Corbel et al, (2017)* [37]

Target site resistance is caused by point mutations in specific genes which encode proteins targeted by the insecticide; this has the effect of reducing or eliminating the normal response to the insecticide. These resistance-conferring mutations have been described most widely in the genes encoding the voltage gated sodium channels (*vgsc*) [60–62], acetylcholinesterase (*ace-1*) [63,64] and resistance to dieldrin or gamma-aminobutyric acid (GABA) (*rdl*) [65–68]. Target site resistance SNPs have been identified in

many species including medically important vectors such as *Ae. aegypti.* The role of these SNPs has been confirmed to different degrees [54].

Mutations in the *vgsc* gene encoding a sodium channel key to nerve function have been widely described associated with resistance. Some common and well described SNPs that have been experimentally confirmed to confer resistance to pyrethroids in the *vgsc* include F1534C/L, I1011M/V and V1016G/I (*Musca domestica* numbering) in *Aedes aegypti* [54]. A mutation in the *ace-1* gene encoding acetylcholinesterase (AchE1), which is involved in terminating neural signals, G119S (*Torpedo californica* numbering) has been described in many *Anopheles* and *Culex spp.* and is associated with resistance to organophosphates and carbamates [69,70]. This mutation has only been described once in *Aedes spp.* [71]; likely due to two mutations being required for glycine to serine conversion in *Aedes* compared to a single mutation in *Anopheles* and *Culex*. The A301S/G mutation (*Drosophila melanogaster* numbering) occurring in the *rdl* gene has been described in multiple vector species and confers resistance to dieldrin and organochlorines. This mutation has widely been described in *Anopheles spp.* but less so in *Aedes spp* [39–42]. The mutation has only been identified in a few wild populations, including *Ae. albopictus* from Malaysia [43] and in *Ae. aegypti* populations in this thesis [72,73].

Metabolic resistance is caused by the upregulation of detoxification enzymes. Production of additional or more efficient detoxification enzymes by the insect causes metabolism of insecticide molecules before they can influence their target site [54]. There are three major detoxification systems: cytochrome P450 monooxygenases (P405s), carboxyl/cholinesterases (CEs), UDP-glucuronosyltransferases (UGTs), ABC transporters and glutathione S-transferase (GSTs) [74–77]. Multiple genes have been associated with this type of resistance in *Aedes* mosquitoes, however they vary dramatically between geographical regions. Examples include GSTe2, CCEAE3A, CYP6P12, CYP4D24 which are involved in resistance to pyrethroids [38].

Glutathione S-transferase is a detoxifying enzyme which has been associated with metabolic resistance, however, point mutations in the *GSTe2* gene have been found to also be associated with resistance. The L119V described in *Anopheles funestus* was shown to be associated with DDT and pyrethroid resistance, while the I114T described in *Anopheles gambiae,* have been implicated in resistance to DDT [78,79]. In *Ae. aegypti* modelling studies have associated two mutations; L111S and I150V, with resistance to temephos [80].

Behavioural resistance results from changes in mosquito behaviour resulting in reduced contact with insecticides. This can be stimulus dependent or independent. Stimulus independent is natural avoidance of an environment with insecticide, for example an increase in exophilic behaviour in a location with IRS or ITNs [81]. Stimulus dependent behaviour refers to an increased ability to detect insecticide and avoid it accordingly [81]. Behavioural resistance has been observed in *Anopheles* species, such as, *An. punctulatus* and *An. farauti* s.s., which, following an ITN distribution campaign, modified their behaviour to bite earlier in the day [82,83]. Another study indicated that in the four years following mass long lasting insecticidal net distribution in Tanzania, vectors *An. funestus* and *An. arabiensis* both showed an increase in exophilic and in the case of *An. arabiensis* an alteration in host preference was observed with a decrease in anthropophilic behaviour [84]. However, there is limited information on behavioural resistance for a number of reasons; it is complex to observe and measure, bioassays can't detect behavioural changes, and fewer people are researching this area of resistance [81].

Other insecticide resistance mechanisms include changes to cuticle structure and microbiome effects. Cuticular resistance occurs when a thickened or altered cuticle prevents or slows the mosquito's absorption of insecticide when it rests on surfaces [85,86]. This mechanism may allow more time for detoxifying enzymes to act, and the mosquito may thereby receive a sub-lethal dose [85]. Cuticular resistance has not been well studied in *Aedes* vectors. In most cases cuticle modifications have been found in combination with other mechanisms (target site and metabolic), this might suggest that it may have more effect more synergistically with other mechanisms than alone [85]. Additionally, increased hydrocarbon and polysaccharide content in the cuticles may aid resistance and could be increased due to overexpression of cytochrome P450 genes also involved in metabolic resistance [86,87].

There is emerging evidence that an insect's microbiota may have some effect on insecticide resistance, however the exact mechanisms remain unclear [88]. One study found that gut bacteria may play a role in insecticide resistance in diamondback moths (*Plutella xylostella)* [89]. Studies in *An. albimanus* and *Ae. albopictus* mosquitoes found that the gut microbiota was significantly different in susceptible and resistant adults, indicating that bacteria may play a role in microbiota-mediated insecticide resistance [90,91]. Further work is required to elucidate the link and potential mechanisms between the insect microbiota and insecticide resistance.

Finally other mechanisms have been identified utilising alternative systems including chemosensory proteins, transcription factors, alpha-crystallin's chaperon proteins and hexamerin storage and

transport proteins [92–95]. For example, chemosensory proteins such as the sensory appendage protein (SAP2) have been implicated in resistance to pyrethroids in *An. gambiae*. Overexpression of SAP2 conferred resistance, while silencing the gene restored susceptibility [92]. Similarly, a negative correlation between the expression of odorant-binding protein 28 (OBP28) and bioassay mortality was observed in *Culex quinquefasciatus*, suggesting its involvement in resistance[94].

Importantly, all these mechanisms are not mutually exclusive and there is some evidence that insecticide resistance can be additive [96]. The total number of genes involved in insecticide resistance remains unknown, however, many genes involved in insecticide resistance are highly conserved across multiple vector species. Promisingly, this means that studies that focus on a single vector may benefit the whole field of insecticide resistance [97–99], however this is variable between gene families and one to many homologues are also common. There is still much to understand about the interplay and interactions between mechanisms which adds further complexity.

Vector Monitoring and Surveillance

Mosquito surveillance and monitoring activities are essential to maintain effective local and national vector control. These activities involve the regular, systematic collection of data at representative local sites, which informs stratification of risk for prioritization of resources, as well as identifying outbreaks and threats to control such as insecticide resistance [100]. It is becoming increasingly recognised that, as well as monitoring factors such as case burden and vector abundance, insecticide resistance monitoring should be a foundation of vector control management plans.

Current gold standard methods of identifying insecticide resistance are susceptibility biological assays [101,102]and biochemical assays [103]. Biological assays involve exposing larvae or adult mosquitoes in bottles to a diagnostic dose of insecticide, then regularly monitoring larvae/mosquito knockdown. These experiments are essential to provide an understanding of the insecticide resistance profile of a population of mosquitoes, however they are time consuming with multiple repeats needed for reliable estimation and may be subjective due to the difficulty in judging knockdown. Alternatively, biochemical assays can be used to measure the activity of detoxification enzymes within the mosquito. However, this is technically complex and requires a cold chain to avoid loss of enzyme activity. Moreover, due to the large enzyme families involved, over-expression of one gene may not affect the expression of the whole enzyme family. Consequently, surveillance programmes may benefit from complementing phenotypic bioassays with genomic surveillance.

Molecular assays can be used to monitor resistance markers and can provide key information on the underpinning mechanism of resistance therefore aiding in choosing the best control strategy available given the profile of the population. These molecular methods are usually polymerase chain reaction (PCR), or quantitative PCR based, and therefore only provide information on validated DNA markers [104–106]. Molecular methods are beneficial because they are objective and can provide high throughput screening of a population. Moreover, they may identify novel mechanisms of resistance that may not be detected via phenotypic assays. Unfortunately, most vector control programmes are not established to carry out molecular surveillance for several reasons, primarily resources and capability. There is a need for more rapid and economically viable tests of molecular markers of insecticide resistance. Molecular tools may provide a substantial benefit and would allow public health organizations to monitor the emergence and spread of known resistance mutations and identify new putative resistant markers.

Sequencing technologies for epidemiological surveillance

The scientific community has come a long way since the human genome sequence was completed in 2003, chiefly due to developments in sequencing technologies that have allowed quicker, more accurate, and more affordable sequencing. We are in an era of next-generation sequencing (NGS, second generation) as well as the ongoing development of third generation sequencing (TGS), which means an entire human genome can now be sequenced in less than a day [107]. The development of NGS and TGS has led to increased data generating capacity, along with a reduction in costs. This has increased the scope of opportunities to study and control of infectious diseases, with hundreds of genes crucial to disease being discovered in this manner [108,109].

Sequencing can be broadly split into short and long read systems, each with their advantages. Short read methods are quicker and more accurate with error rates of ~0.1-0.5% per base, however these methodologies are limited by the maximum read length which is ~600 nucleotides [110]. Whereas long read methods are less accurate with error rates of between 2-20% per base however much longer read lengths are possible (>10,000 nucleotides) [111]. Short read platforms include Illumina MiSeq and SOliD, while long read platforms include Pacific Biosciences Sequel and Oxford Nanopore Technology (ONT). The accuracy of short read sequencing is beneficial when aiming to detect variants in small target regions with low coverage. Long read sequencing is advantageous for resolving long repetitive regions of DNA and therefore essential to construct complex regions of genome assemblies [109]. The

additional benefits of TGS include sequencing directly from the DNA molecule removing the requirement of PCR, ultra-long read production, and real time sequencing [112].

Modern sequencing technologies have improved in various ways including the reduction in the size of the template sample required for sequencing, accuracy, and information yield, these developments are highly beneficial for epidemiological disease surveillance. Using NGS, a sequence can be obtained from a single strand of DNA which is effective for surveillance of pathogens which may be detected at low abundance such as arboviruses in mosquito vectors. Parallel processing of NGS has improved both speed and accuracy by accommodating the production of more sequence copies; this can aid sequencing of more individuals improving understanding of genetic variation, which may be important for epidemiological characteristics such as resistance [113]. As well as the earlier detection of emerging detrimental mutations at low frequency, comparative genomics is also useful for examining outbreaks and transmission to gain further understanding and aid future control strategies. Finally, reduced labour, time and expense for reagents mean that costs have been significantly reduced, increasing the range of settings where it can be implemented; this is particularly important given infectious diseases disproportionately affect low-income countries [114]. However, it is important to note that barriers to use remain in low-resource settings, including cost and access to reagents. Finally, there are emerging technologies such as Oxford Nanopore technology which is promising for both clinical and field settings given its portable nature as the smallest sequencing platform available [115].

With the improvements in these technologies, sequencing is now a standard part of scientific research, with many envisaging its ubiquitous use in clinical settings and its expansion into low resource settings. Projects such as MalariaGEN and Ag1000G have illustrated the benefits of a concerted effort to sequence a species and centralise available data and analysis [116,117]. Leading to benefits such as the discovery of multiple drug resistance SNPs in *Plasmodium* parasites [118] as well as identifying genes associated with insecticide resistance in *An. gambiae* [119]. The successes of sequencing partnered with bioinformatics to improve understanding of disease, diagnosis, treatment, and control; it is promising that they can be usefully applied to *Ae. aegypti* mosquitoes [120].

Amplicon Sequencing

As previously outlined, molecular techniques are becoming increasingly affordable and attractive methodologies for a variety of applications in public health [121,122]. This includes sequencing technologies ranging from sequencing short fragments of DNA (amplicons) all the way up to the entire

genome of an organism. There are some situations where sequencing larger sections of the genome (from entire genes of interest, up to the whole genome) may be preferable. Whole genome sequencing (WGS) can elucidate large structural variants such as insertions, deletions and inversions which may be missed by amplicon sequencing [123]. Additionally, less initial information is required about the organism for WGS to be effective which can be particularly important in less well studied organisms. However, this must be balanced with disadvantages such as the increased cost of WGS compared to other methodologies including amplicon sequencing [124].

Amplicon sequencing, by employing small fragments of DNA, allows numerous amplicons from multiple samples to be sequenced in a single run, using a next generation or third generation sequencing platform, permitting high throughput at a lower cost. This methodology has been applied to many diverse diseases and pathogens, including screening malaria parasites for antimalarial resistance, and insecticide resistance in *Anopheles* and examining the microbiome of children [119,125,126].

Amplicon sequencing involves (multiplex) PCR to create amplicons, combined with sequencing to obtain deep coverage of targeted sequenced fragments. The three basic steps used for amplicon sequencing in the Illumina platform, are outlined in **Figure 4**. First, a PCR is performed to produce short specific amplicon sequences, with their length depending on the sequencing platform used (usually 300-1000bp). For multiplexing of multiple samples, barcodes can be combined with primers to permit discrimination of each sample for analysis. Additionally, if using TGS platforms such as Oxford Nanopore, samples can be barcoded again for even higher throughput. For Illumina platforms, adapters can be ligated to amplicon fragments to allow the amplicons to adhere to the flow cell for sequencing. Finally, deep sequencing of fragments is performed [127].

Amplicon sequencing can quickly and easily provide high coverage sequence data which can be employed to identify low frequency variants in specific genomic regions of interest. It is especially useful in cases where the organism of interest has a large genome, such as *Ae. aegypti* (1.3Gb). The barcoding and multiplexing capabilities allow high throughput, so amplicon sequencing can be used to explore many genomic regions and efficiently process large numbers of samples. Moreover, the selective nature of the sequencing means that data can be analysed more efficiently with less computational power required than in whole genome sequencing.

Amplicon sequencing is a powerful tool that may rapidly and economically identify insecticide resistance mutations making it a useful approach in geographical regions that rely heavily on

insecticide vector control methods to monitor insecticide resistance. Utilizing genomic information can offer insight on changes in the genetic diversity of both mosquitos and pathogens to increase our knowledge of their biology and evolution, which could inform future VBD control efforts.

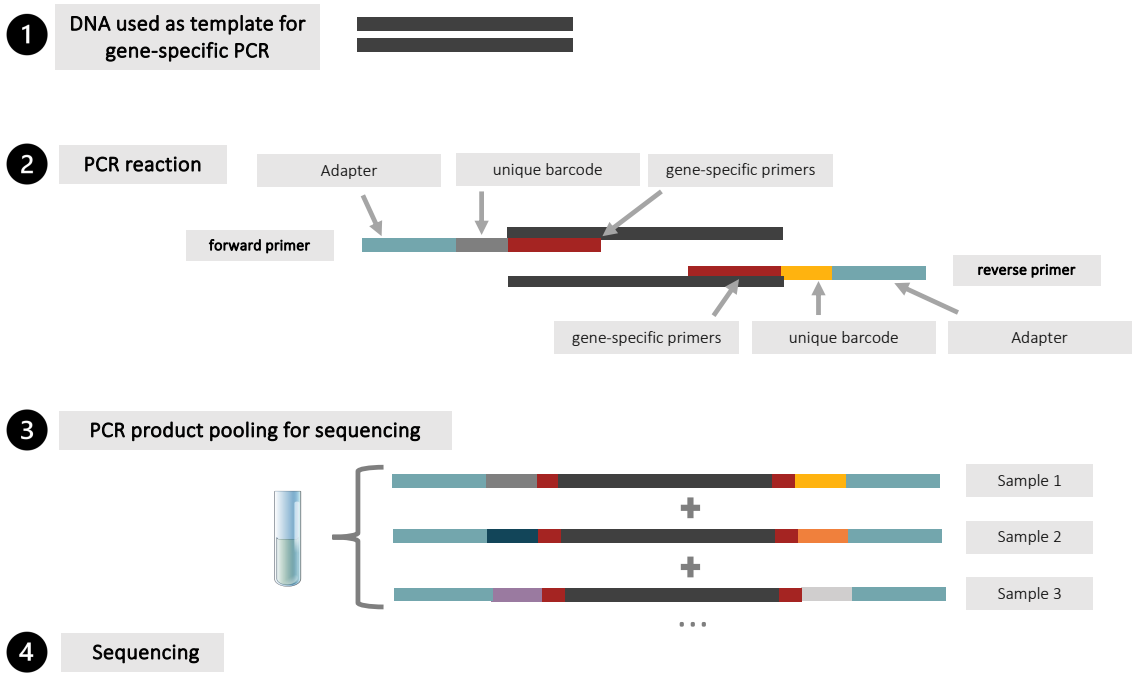


Figure 1. *Steps for amplicon sequencing. The amplicon PCR outlines the use of different barcodes per sample to allow amplification of different regions of interest in multiple samples to identify each sample within a multiplex PCR. Adapter ligation to product is required for sequencing platform compatibility. Finally sequencing can be carried out.*

Genomics of Mosquitoes

As a comprehensively studied model organism, *Drosophila melanogaster* was the first insect genome to be sequenced in the year 2000 [128]. Two years later, *Anopheles gambiae* became the first mosquito to be sequenced and assembled [129]. Differences in *An. gambiae* and *D. melanogaster* indicated that there were many orthologues, but for about half the genes there were substantial differences [130]; this observation indicated the need for more mosquito genomes to be sequenced. However, of over 3,500 extant Culicidae species just 129 have a reference genome sequence available in the NCBI's Taxonomy Database [131,132].

The Culicidae family includes more than 129 genera, with three primary genera known for transmitting vector-borne diseases (VBDs): *Anopheles, Aedes,* and *Culex*. Among all the available genome sequences for Culicidae, 74% represent Anopheles, whereas *Aedes* and *Culex* genomes account for only 14% and 5%, respectively [131]. Examining the sequencing record archive of NCBI shows there has been a focus on sequencing *Anopheles gambiae* mosquito genomes, assisted in part, by the Ag1000G project which aimed to sequence more than 1000 field *An. gambiae* genomes [117,133]. The paucity of genome records from *Aedes* and *Culex* vectors indicates the need for more research into non-*Anopheles* mosquitos, which are also prominent vectors of disease.

Sequencing and working with mosquito genomes present several challenges. One such challenge is the genome size; while *Plasmodium* genomes are around ~30 Mb, Anopheles genomes are around ten times larger at ~300 Mb and *Aedes* genomes are even larger still at ~1,300 Mb [134–136]. This makes sequencing more difficult to get the required coverage across the genome and analysis more computationally intensive due to the amount of data produced. Secondly, their genome complexity can cause problems. Mosquito species have a multitude of genes and a high degree of repetitive elements and transposable elements. *Aedes* mosquitoes have a highly repetitive genome along with 19,790 genes [136,137]. Mosquitos' diploid nature and high levels of genetic diversity, probably due to their large population size and sexual reproduction, means robust analytical methods are required to understand population structure and dynamics, which can be more computationally intense.

However, molecular techniques and analysis have been used to compare mosquito species and can reveal significant differences between them [138]. For example, there are huge variations in the size of mosquito genomes, ranging from 225 to 1,862 Mb. Culicinae have large genomes and have the most inter- and intra-species genomic variation [139]. Interestingly there have also been reports of up to three times intraspecies differences in genome size within *A. albopictus* [140]. The interspecies variation in genome size is probably due to the number of repetitive elements (e.g., microsatellites, transposable elements and ribosomal RNA), which may be driven by selection [141]. Transposable elements account for ~50% of the *Ae. aegypti* genome [142]. Both Anopheline and Culicine sequences have around ~46% GC content. Orthology of genes is seen across insect species. An analysis of single copy orthologs found ~74% amino acid identity between *Ae. aegypti* and *An. gambiae* compared with ~58% identity between mosquito and *D. melanogaster* (**Figure 5**) [142]. This analysis illustrates the power of comparative genomics to elucidate evolutionary gene dynamics in mosquitoes.

Applying genomic techniques and population analysis within a species has led to a greater understanding of the species' evolutionary history, gene flow and observed phenotypes. For example, a study investigating the genetic differences between differing *Ae. aegypti* populations found that the genetic diversity of African populations is much higher than non-African populations [143,144]. This supports an out-of-Africa migration theory whereby it is understood that *Aedes* were distributed from Africa to the Americas initially, and then secondly distributed from the Americas to Asia [145]. A similar study considering invasive populations of *Ae. aegypti* in California, USA, found evidence for the recent introduction of *Ae. aegypti* from multiple genetically diverse populations [146]. An analysis of *Ae. aegypti* discovered a genetic basis for changes in host preference, including that adaptions in odorant receptor 4 (AAEL015147) play a significant role in *Ae. aegypti's* 'domestication' to human feeding preference [147]. Moreover, genomic analysis can highlight reductions in variation in genes indicating advantageous mutations are undergoing a selective sweep. For instance, a study of *Culex* mosquitos found a high number of pathways under positive selection, including in the genes in the P450 family and salivary proteins associated with insecticide resistance [148].

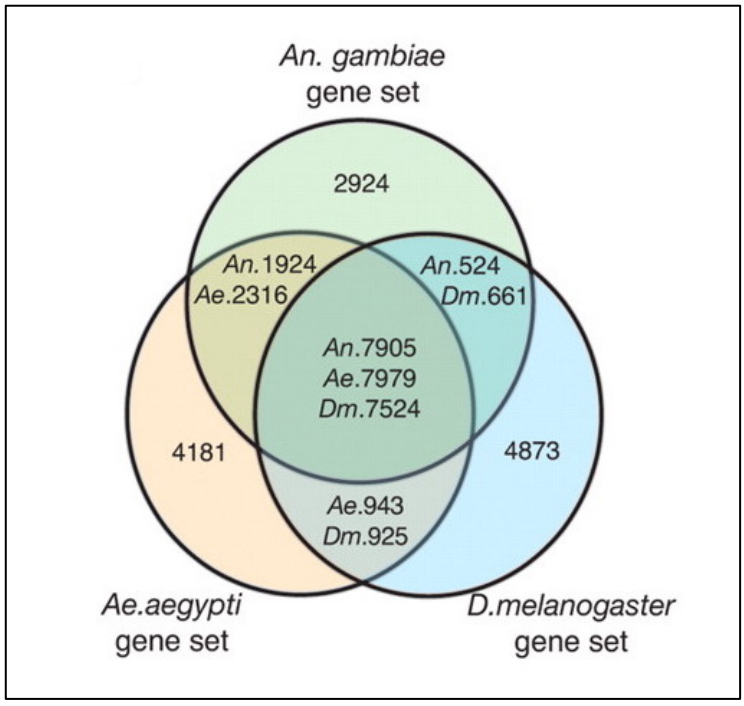


Figure 5. *Orthology between Ae. aegypti, An. gambiae and D. melanogaster genes. Adapted from Nene et al. (2007)* [91].

Insecticide resistance poses a formidable challenge to global initiatives aiming to control mosquitoes that transmit vector-borne diseases (VBDs). Utilizing molecular methodologies to elucidate the markers linked to insecticide resistance is helpful to try and ensure the best interventions are in place

and developing effective strategies to mitigate its impact. This thesis investigates and delineates these markers, aiming to provide insights that may reshape our strategies in vector control and public health management. Furthermore, it seeks to enhance our understanding of the global diversity of *Aedes* populations, thereby informing comprehensive approaches to combatting insecticide resistance.

Objectives

Overall, this thesis aimed to utilise genomic approaches to investigate genetic diversity and insecticide resistance in *Ae. aegypti* mosquitoes. This aim is explored in four objectives, outlined below, corresponding to each chapter of the thesis.

Objective 1: Create an amplicon sequencing insecticide resistance surveillance panel to capture SNPs associated with insecticide resistance and apply it to a population of field *Ae. aegypti* samples from Cabo Verde.

Objective 2: Pair an expanded amplicon sequencing insecticide resistance screening panel (created in **Objective 1**), with phenotypic testing on a field population of *Ae. aegypti* from Puerto Rico.

Objective 3: Use public *Ae. aegypti* WGS data to investigate genetic diversity of key insecticide resistance associated genes (*vgsc, rdl, ace-1* and *GSTe2*) globally.

Objective 4: Generate and analyse WGS data from a population of *Ae. aegypti sourced* from Puerto Rico and examine their signals of selection and relatedness to those from other geographical regions.

Outline of Thesis

In this thesis I demonstrate how both wet and dry lab techniques can be applied to investigate insecticide resistance in *Ae. aegypti* mosquitoes. In **Chapter 1**, I show the development of an amplicon sequencing panel and methodology to examine well-described target site resistance SNPs in a population of mosquitoes from Cabo Verde (Published Manuscript [73]). **Chapter 2** illustrates further development of this methodology to include more sites in the panel as well as importantly pair the molecular data with phenotypic data from Puerto Rico (Submitted Manuscript, under revision [72]). **Chapters 1** and **2** utilise Illumina NGS technology. Both **Chapters 3** and **4** focus on the utilisation of WGS to provide further insights into the population genomics of *Ae. aegypti*. **Chapter 3** further examines the field samples collected in Puerto Rico, presenting population genomics of the Caribbean region compared to the Americas, which had not previously been examined (Published Manuscript [149]). Finally, **Chapter 4** utilises publicly available WGS data for *Ae. aegypti* and performs comparative genomics with a focus on insecticide resistance associated genes (*vgsc, rdl, ace-1* and *GSTe2*) (Submitted Manuscript).

References

1. World Health Organisation. Vector-borne diseases. https://www.who.int/news-room/fact-sheets/detail/vector-borne-diseases (2020).
2. Elbers, A. R. W., Koenraadt, C. J. M. & Meiswinkel, R. Mosquitoes and Culicoides biting midges: vector range and the influence of climate change. *Rev. Sci. Tech. Off. Int. Epiz* 34, 123–137 (2015).
3. Golding, N. *et al.* Integrating vector control across diseases. *BMC Med* 13, 1–6 (2015).
4. Sinka, M. E. *et al.* A global map of dominant malaria vectors. *Parasit Vectors* 5, 1–11 (2012).
5. World malaria report 2023. https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2023.
6. Bhattacharya, S., Basu, P. & Sajal Bhattacharya, C. The Southern House Mosquito, Culex quinquefasciatus: profile of a smart vector. *~ 73 ~ Journal of Entomology and Zoology Studies* 4, 73–81 (2016).
7. Souza-Neto, J. A., Powell, J. R. & Bonizzoni, M. Aedes aegypti vector competence studies: A review. *Infection, Genetics and Evolution* 67, 191–209 (2019).
8. Brown, J. E. *et al.* Human impacts have shaped historical and recent evolution in Aedes aegypti, the dengue and yellow fever mosquito. *Evolution* 68, 514–525 (2014).
9. Kraemer, M. U. G. *et al.* Past and future spread of the arbovirus vectors Aedes aegypti and Aedes albopictus. *Nature Microbiology 2019 4:5* 4, 854–863 (2019).
10. Wilson, A. L. *et al.* The importance of vector control for the control and elimination of vector-borne diseases. *PLoS Negl Trop Dis* 14, 1–31 (2020).
11. Araújo, M. F., Castanheira, E. M. S. & Sousa, S. F. The Buzz on Insecticides: A Review of Uses, Molecular Structures, Targets, Adverse Effects, and Alternatives. *Molecules* 28, (2023).
12. Gubler, D. J. & Clark, G. G. Community involvement in the control of Aedes aegypti. *Acta Trop* 61, 169–179 (1996).
13. Antonio-Nkondjio, C., Sandjo, N. N., Awono-Ambene, P. & Wondji, C. S. Implementing a larviciding efficacy or effectiveness control intervention against malaria vectors: Key parameters for success. *Parasit Vectors* 11, 1–12 (2018).
14. Controlling Mosquitoes at the Larval Stage | US EPA. https://www.epa.gov/mosquitocontrol/controlling-mosquitoes-larval-stage.
15. Lamaningao, P. *et al.* Aedes mosquito surveillance and the use of a larvicide for vector control in a rural area of the Lao People's Democratic Republic. *Trop Med Health* 48, 1–9 (2020).
16. Wilke, A. B. B. *et al.* Proliferation of Aedes aegypti in urban environments mediated by the availability of key aquatic habitats. *Scientific Reports 2020 10:1* 10, 1–10 (2020).
17. Weeratunga, P., Rodrigo, C., Fernando, S. D. & Rajapakse, S. Control methods for Aedes albopictus and Aedes aegypti. *Cochrane Database Syst Rev* 2017, (2017).
18. Gratz, N. G. Space Sprays for the Control of Aedes aegypti in South-East Asia and the Western Pacific. *Dengue Bull* 23, 80 (1999).
19. Gunning, C. E. *et al.* Efficacy of Aedes aegypti control by indoor Ultra Low Volume (ULV) insecticide spraying in Iquitos, Peru. *PLoS Negl Trop Dis* 12, 3230 (2018).
20. Lorenz, C. & Chiaravalloti-Neto, F. Control methods for Aedes aegypti: Have we lost the battle? *Travel Med Infect Dis* 49, 102428 (2022).
21. Samuel, M. *et al.* Community effectiveness of indoor spraying as a dengue vector control method: A systematic review. *PLoS Negl Trop Dis* 11, (2017).
22. Insecticide-Treated Nets | Malaria | CDC. https://www.cdc.gov/malaria/php/public-health-strategy/insecticide-treated-nets.html.
23. Saili, K. *et al.* House Screening Reduces Exposure to Indoor Host-Seeking and Biting Malaria Vectors: Evidence from Rural South-East Zambia. *Trop Med Infect Dis* 9, (2024).
24. Manrique-Saide, P. *et al.* Protective effect of house screening against indoor Aedes aegypti in Mérida, Mexico: A cluster randomised controlled trial. *Tropical Medicine & International Health* 26, 1677 (2021).
25. Ortega-Morales, A. I. & Nava, M. R. Mosquito species of neighboring states of Mexico. *Mosquitoes, Communities, and Public Health in Texas* 279–306 (2019) doi:10.1016/B978-0-12-814545-6.00009-2.
26. Lenhart, A. *et al.* Insecticide-treated bednets to control dengue vectors: preliminary evidence from a controlled trial in Haiti. *Tropical Medicine & International Health* 13, 56–67 (2008).

27. World Health Organization. Manual for monitoring insecticide resistance in mosquito vectors and selecting appropriate interventions.
28. WHO. Test procedures for insecticide resistance monitoring in malaria vector mosquitoes (Second edition) (Updated June 2018). *Who* 48 (2018).
29. van den Berg, H. *et al.* Recent trends in global insecticide use for disease vector control and potential implications for resistance management. *Scientific Reports 2021 11:1* 11, 1–12 (2021).
30. Sparks, T. C. & Nauen, R. IRAC: Mode of action classification and insecticide resistance management. *Pestic Biochem Physiol* 121, 122–128 (2015).
31. Kirst, H. A. The spinosyn family of insecticides: realizing the potential of natural products research. *The Journal of Antibiotics 2010 63:3* 63, 101–111 (2010).
32. Narahashi, T. Mode of action of pyrethroids. *Bull World Health Organ* 44, 337 (1971).
33. Colovic, M. B., Krstic, D. Z., Lazarevic-Pasti, T. D., Bondzic, A. M. & Vasic, V. M. Acetylcholinesterase Inhibitors: Pharmacology and Toxicology. *Curr Neuropharmacol* 11, 315 (2013).
34. Cole, L. M., Nicholson, R. A. & Casida, J. E. Action of Phenylpyrazole Insecticides at the GABA-Gated Chloride Channel. *Pestic Biochem Physiol* 46, 47–54 (1993).
35. Zhorov, B. S. & Dong, K. Elucidation of pyrethroid and DDT receptor sites in the voltage-gated sodium channel. *Neurotoxicology* 60, 171 (2017).
36. EPA. Pesticide Fact Sheet Chlorfenapyr. (2001).
37. N'Guessan, R., Corbel, V., Akogbéto, M. & Rowland, M. Reduced efficacy of insecticide-treated nets and indoor residual spraying for malaria control in pyrethroid resistance area, Benin. *Emerg Infect Dis* 13, 199–206 (2007).
38. 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).
39. Insecticide Resistance Action Committee. *MODE OF ACTION CLASSIFICATION SCHEME*. https://irac-online.org/documents/moa-classification/ (2024).
40. Bacci, L., Lupi, D., Savoldelli, S. & Rossaro, B. A review of Spinosyns, a derivative of biological acting substances as a class of insecticides with a broad range of action against many insect pests. *Journal of Entomological and Acarological Research* 48, 40–52 (2016).
41. Sparks, T. C., Dripps, J. E., Watson, G. B. & Paroonagian, D. Resistance and cross-resistance to the spinosyns – A review and analysis. *Pestic Biochem Physiol* 102, 1–10 (2012).
42. 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).
43. Ambadiang, M. *et al.* Anopheles gambiae larvae's ability to grow and emerge in water containing lethal concentrations of clothianidin, acetamiprid, or imidacloprid is consistent with cross-resistance to neonicotinoids. *Parasit Vectors* 17, 1–12 (2024).
44. Zoh, M. G. *et al.* Molecular bases of P450-mediated resistance to the neonicotinoid insecticide imidacloprid in the mosquito Ae. aegypti. *Aquatic Toxicology* 236, 105860 (2021).
45. Tchigossou, G. *et al.* Resistance to neonicotinoids is associated with metabolic detoxification mechanisms in Anopheles coluzzii from agricultural and urban sites in southern Benin. *Frontiers in Tropical Diseases* 5, 1339811 (2024).
46. WHOPES. *Guidelines on Public Health Pesticide Management Policy for the WHO African Region*. (2011).
47. IR MAPPER. https://www.irmapper.com/ (2024).
48. Alphey, L. *et al.* Genetic control of Aedes mosquitoes. *Pathogens and Global Health* vol. 107 170–179 Preprint at https://doi.org/10.1179/2047773213Y.0000000095 (2013).
49. Ogunlade, S. T., Adekunle, A. I., Meehan, M. T. & McBryde, E. S. Quantifying the impact of Wolbachia releases on dengue infection in Townsville, Australia. *Scientific Reports 2023 13:1* 13, 1–12 (2023).
50. Lim, J. T. *et al.* Efficacy of Wolbachia-mediated sterility to reduce the incidence of dengue: a synthetic control study in Singapore. *Lancet Microbe* 5, e422–e432 (2024).
51. Pavan, M. G., Garcia, G. A., David, M. R. & Maciel-de-Freitas, R. The double-edged sword effect of expanding Wolbachia deployment in dengue endemic settings. *The Lancet Regional Health - Americas* 27, 100610 (2023).
52. Anders, K. L. *et al.* The impact of large-scale deployment of Wolbachia mosquitoes on dengue and other Aedes-borne diseases in Rio de Janeiro and Niterói, Brazil: study protocol for a controlled interrupted time series analysis using routine disease surveillance data. *F1000Res* 8, (2019).

53. IRAC. *IRAC Mode of Action Classification*. www.irac-online.org (2017).
54. Moyes, C. L. *et al.* Contemporary status of insecticide resistance in the major Aedes vectors of arboviruses infecting humans. *PLoS Negl Trop Dis* 11, (2017).
55. Zaim, M. & Guillet, P. Alternative insecticides: An urgent need. *Trends in Parasitology* vol. 18 161–163 Preprint at https://doi.org/10.1016/S1471-4922(01)02220-6 (2002).
56. Itokawa, K. *et al.* High-throughput genotyping of a full voltage-gated sodium channel gene via genomic DNA using target capture sequencing and analytical pipeline MoNaS to discover novel insecticide resistance mutations. *PLoS Negl Trop Dis* 13, e0007818 (2019).
57. World Health Organization. *Global Report on Insecticide Resistance in Malaria Vectors: 2010–2016*. *World Health Organization* https://www.who.int/malaria/publications/atoz/9789241514057/en/ (2018).
58. Crow, J. F. Genetics of Insect Resistance to Chemicals. *Annu Rev Entomol* 2, 227–246 (1957).
59. Corbel, V. *et al.* International workshop on insecticide resistance in vectors of arboviruses, December 2016, Rio de Janeiro, Brazil. *Parasit Vectors* 10, 278 (2017).
60. Vais, H. *et al.* Activation of Drosophila sodium channels promotes modification by deltamethrin. Reductions in affinity caused by knock-down resistance mutations. *J Gen Physiol* 115, 305–318 (2000).
61. Soderlund, D. M. & Knipple, D. C. The molecular biology of knockdown resistance to pyrethroid insecticides. *Insect Biochem Mol Biol* 33, 563–577 (2003).
62. Dong, K. Insect sodium channels and insecticide resistance. *Invert Neurosci* 7, 17 (2007).
63. Menozzi, P., Shi, M. A., Lougarre, A., Tang, Z. H. & Fournier, D. Mutations of acetylcholinesterase which confer insecticide resistance in Drosophila melanogaster populations. *BMC Evol Biol* 4, 4 (2004).
64. Mutero, A., Pralavorio, M., Bride, J. M. & Fournier, D. Resistance-associated point mutations in insecticide-insensitive acetylcholinesterase. *Proc Natl Acad Sci U S A* 91, 5922 (1994).
65. Remnant, E. J. *et al.* The role of Rdl in resistance to phenylpyrazoles in Drosophila melanogaster. *Insect Biochem Mol Biol* 54, 11–21 (2014).
66. Anthony, N., Unruh, T., Ganser, D. & Ffrench-Constant, R. Duplication of the Rdl GABA receptor subunit gene in an insecticide-resistant aphid, myzus persicae. *Molecular and General Genetics* 260, 165–175 (1998).
67. 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).
68. Asih, P. B. *et al.* Existence of the Rdl mutant alleles among the Anopheles malaria vector in Indonesia. *Malar J* 11, 57 (2012).
69. Weill, M. *et al.* Insecticide resistance in mosquito vectors. *Nature* 423, 136–137 (2003).
70. Feyereisen, R., Dermauw, W. & Van Leeuwen, T. Genotype to phenotype, the molecular and physiological dimensions of resistance in arthropods. *Pestic Biochem Physiol* 121, 61–77 (2015).
71. Engdahl, C. *et al.* Acetylcholinesterases from the Disease Vectors Aedes aegypti and Anopheles gambiae: Functional Characterization and Comparisons with Vertebrate Orthologues. *PLoS One* 10, (2015).
72. Collins, E. *et al.* Profiling insecticide resistance phenotypes and genotypes in Aedes aegypti populations across four regions in Puerto Rico. (2024) doi:10.21203/RS.3.RS-3943783/V1.
73. Collins, E. L. *et al.* A next generation targeted amplicon sequencing method to screen for insecticide resistance mutations in Aedes aegypti populations reveals a rdl mutation in mosquitoes from Cabo Verde. *PLoS Negl Trop Dis* 16, e0010935 (2022).
74. Hemingway, J. & Ranson, H. Insecticide Resistance in Insect Vectors of Human Disease. *Annu Rev Entomol* 45, 371–391 (2000).
75. David, J. P., Ismail, H. M., Chandor-Proust, A. & Paine, M. J. I. Role of cytochrome P450s in insecticide resistance: Impact on the control of mosquito-borne diseases and use of insecticides on earth. *Philosophical Transactions of the Royal Society B: Biological Sciences* 368, (2013).
76. Perry, T., Batterham, P. & Daborn, P. J. The biology of insecticidal activity and resistance. *Insect Biochem Mol Biol* 41, 411–422 (2011).
77. Amezian, D., Nauen, R. & Van Leeuwen, T. The role of ATP-binding cassette transporters in arthropod pesticide toxicity and resistance. *Curr Opin Insect Sci* 63, 101200 (2024).
78. 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).
79. Mitchell, S. N. *et al.* Metabolic and target-site mechanisms combine to confer strong DDT resistance in Anopheles gambiae. *PLoS One* 9, e92662 (2014).

80. Helvecio, E. *et al.* Polymorphisms in GSTE2 is associated with temephos resistance in Aedes aegypti. *Pestic Biochem Physiol* 165, (2020).
81. Hubbard, C. B. & Murillo, A. C. Behavioral resistance to insecticides: current understanding, challenges, and future directions. *Curr Opin Insect Sci* 63, 101177 (2024).
82. Reimer, L. J. *et al.* Malaria transmission dynamics surrounding the first nationwide long-lasting insecticidal net distribution in Papua New Guinea. *Malar J* 15, 1–11 (2016).
83. Thomsen, E. K. *et al.* Mosquito Behavior Change After Distribution of Bednets Results in Decreased Protection Against Malaria Exposure. *J Infect Dis* 215, 790–797 (2017).
84. Kreppel, K. S. *et al.* Emergence of behavioural avoidance strategies of malaria vectors in areas of high LLIN coverage in Tanzania. *Scientific Reports 2020 10:1* 10, 1–11 (2020).
85. Balabanidou, V., Grigoraki, L. & Vontas, J. Insect cuticle: a critical determinant of insecticide resistance. *Curr Opin Insect Sci* 27, 68–74 (2018).
86. Jacobs, E. *et al.* Cuticular profiling of insecticide resistant Aedes aegypti. *Scientific Reports 2023 13:1* 13, 1–15 (2023).
87. Adams, K. L. *et al.* Cuticular hydrocarbons are associated with mating success and insecticide resistance in malaria vectors. *Communications Biology 2021 4:1* 4, 1–8 (2021).
88. Zhang, Y. *et al.* Microbiome variation correlates with the insecticide susceptibility in different geographic strains of a significant agricultural pest, Nilaparvata lugens. *npj Biofilms and Microbiomes 2023 9:1* 9, 1–8 (2023).
89. Xia, X. *et al.* Gut Microbiota Mediate Insecticide Resistance in the Diamondback Moth, Plutella xylostella (L.). *Front Microbiol* 9, 25 (2018).
90. Wang, H. *et al.* Differences in the intestinal microbiota between insecticide-resistant and -sensitive Aedes albopictus based on full-length 16S rRNA sequencing. *Microbiologyopen* 10, (2021).
91. 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, (2018).
92. Ingham, V. A. *et al.* A sensory appendage protein protects malaria vectors from pyrethroids. *Nature* 577, 376–380 (2020).
93. Ingham, V. A., Wagstaff, S. & Ranson, H. Transcriptomic meta-signatures identified in Anopheles gambiae populations reveal previously undetected insecticide resistance mechanisms. *Nature Communications 2018 9:1* 9, 1–11 (2018).
94. Shen, R. xin *et al.* Deltamethrin interacts with Culex quinquefasciatus odorant-binding protein: a novel potential resistance mechanism. *Parasit Vectors* 15, 1–9 (2022).
95. Ingham, V. A., Pignatelli, P., Moore, J. D., Wagstaff, S. & Ranson, H. The transcription factor Maf-S regulates metabolic resistance to insecticides in the malaria vector Anopheles gambiae. *BMC Genomics* 18, (2017).
96. Liu, N. Insecticide resistance in mosquitoes: impact, mechanisms, and research directions. *Annu Rev Entomol* 60, 537–559 (2015).
97. Davies, T. G. E., Field, L. M., Usherwood, P. N. R. & Williamson, M. S. DDT, pyrethrins, pyrethroids and insect sodium channels. *IUBMB Life* 59, 151–162 (2007).
98. Douris, V. *et al.* Resistance mutation conserved between insects and mites unravels the benzoylurea insecticide mode of action on chitin biosynthesis. *Proc Natl Acad Sci U S A* 113, 14692–14697 (2016).
99. 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).
100. WHO. *Global Vector Control Response 2017-2030*. https://apps.who.int/iris/bitstream/handle/10665/259205/9789241512978-eng.pdf?sequence=1 (2017).
101. WHO. *Monitoring and Managing Insecticide Resistance in Aedes Mosquito Populations Interim Guidance for Entomologists*. *WHO* www.who.int (2016).
102. WHO. *Test Procedures for Insecticide Resistance Monitoring in Malaria Vector Mosquitoes Global Malaria Programme*. http://apps.who.int/iris/bitstream/10665/250677/1/9789241511575-eng.pdf (2016).
103. Dusfour, I. *et al.* Management of insecticide resistance in the major Aedes vectors of arboviruses: Advances and challenges. *PLoS Negl Trop Dis* 13, e0007615 (2019).
104. Saingamsook, J. *et al.* A multiplex PCR for detection of knockdown resistance mutations, V1016G and F1534C, in pyrethroid-resistant Aedes aegypti. *Parasit Vectors* 10, 1–9 (2017).
105. 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).

106. Bass, C., Nikou, D., Vontas, J., Williamson, M. S. & Field, L. M. Development of high-throughput real-time PCR assays for the identification of insensitive acetylcholinesterase (ace-1R) in Anopheles gambiae. *Pestic Biochem Physiol* 96, 80–85 (2010).
107. Behjati, S. & Tarpey, P. S. What is next generation sequencing? *Arch Dis Child Educ Pract Ed* 98, 236–238 (2013).
108. Rabbani, B., Tekin, M. & Mahdieh, N. The promise of whole-exome sequencing in medical genetics. *Journal of Human Genetics 2014 59:1* 59, 5–15 (2013).
109. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics 2016 17:6* 17, 333–351 (2016).
110. Stoler, N. & Nekrutenko, A. Sequencing error profiles of Illumina sequencing instruments. *NAR Genom Bioinform* 3, (2021).
111. Sahlin, K., Sipos, B., James, P. L. & Medvedev, P. Error correction enables use of Oxford Nanopore technology for reference-free transcriptome analysis. *Nature Communications 2021 12:1* 12, 1–13 (2021).
112. Xiao, T. & Zhou, W. The third generation sequencing: the advanced approach to genetic diseases. *Transl Pediatr* 9, 16373–16173 (2020).
113. Satam, H. *et al.* Next-Generation Sequencing Technology: Current Trends and Advancements. *Biology (Basel)* 12, (2023).
114. Wetterstrand KA, N. H. G. R. I. DNA Sequencing Costs: Data. https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data.
115. Lu, H., Giordano, F. & Ning, Z. Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics Proteomics Bioinformatics* 14, 265–279 (2016).
116. MalariaGEN: The Malaria Genomic Epidemiology Network. https://www.malariagen.net/.
117. Miles, A. *et al.* Genetic diversity of the African malaria vector Anopheles gambiae. *Nature 2017 552:7683* 552, 96–100 (2017).
118. Miotto, O. *et al.* Genetic architecture of artemisinin-resistant Plasmodium falciparum. *Nature Genetics 2015 47:3* 47, 226–234 (2015).
119. Lucas, E. R. *et al.* Genome-wide association studies reveal novel loci associated with pyrethroid and organophosphate resistance in Anopheles gambiae s.l. *bioRxiv* (2023) doi:10.1101/2023.01.13.523889.
120. Bah, S. Y., Morang'a, C. M., Kengne-Ouafo, J. A., Amenga–Etego, L. & Awandare, G. A. Highlights on the Application of Genomics and Bioinformatics in the Fight Against Infectious Diseases: Challenges and Opportunities in Africa. *Front Genet* 9, 408772 (2018).
121. Gilbert, G. L. Molecular diagnostics in infectious diseases and public health microbiology: Cottage industry to postgenomics. *Trends Mol Med* 8, 280–287 (2002).
122. Liu, Q. *et al.* Advances in the application of molecular diagnostic techniques for the detection of infectious disease pathogens (Review). *Mol Med Rep* 27, (2023).
123. Rausch, T. *et al.* DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* 28, i333–i339 (2012).
124. Whitford, W. *et al.* Proof of concept for multiplex amplicon sequencing for mutation identification using the MinION nanopore sequencer. *Scientific Reports 2022 12:1* 12, 1–9 (2022).
125. 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, (2017).
126. Gupta, S. *et al.* Amplicon sequencing provides more accurate microbiome information in healthy children compared to culturing. *Commun Biol* 2, 1–7 (2019).
127. Integrated DNA Technologies. Targeted NGS Amplicon Sequencing | IDT. https://www.idtdna.com/pages/technology/next-generation-sequencing/dna-sequencing/targeted-sequencing/amplicon-sequencing.
128. Adams, M. D. *et al.* The genome sequence of Drosophila melanogaster. *Science* 287, 2185–2195 (2000).
129. Holt, R. A. *et al.* The genome sequence of the malaria mosquito Anopheles gambiae. *Science* 298, 129–149 (2002).
130. Zdobnov, E. M. *et al.* Comparative genome and proteome analysis of Anopheles gambiae and Drosophila melanogaster. *Science* 298, 149–159 (2002).
131. NCBI. Genome - NCBI - NLM. https://www.ncbi.nlm.nih.gov/datasets/genome/?taxon=7157.
132. Hawkes, F. M. & Hopkins, R. J. The mosquito. *Mosquitopia* 16–31 (2022) doi:10.4324/9781003056034-3.
133. Ag1000g. https://www.sanger.ac.uk/tool/ag1000g/.

134. Su, X. Z., Lane, K. D., Xia, L., Sá, J. M. & Wellems, T. E. Plasmodium Genomics and Genetics: New Insights into Malaria Pathogenesis, Drug Resistance, Epidemiology, and Evolution. *Clin Microbiol Rev* 32, (2019).
135. Ayala, D. *et al.* The genome sequence of the malaria mosquito, Anopheles funestus, Giles, 1900. *Wellcome Open Res* 7, (2022).
136. Matthews, B. J. *et al.* Improved reference genome of Aedes aegypti informs arbovirus vector control. *Nature 2018 563:7732* 563, 501–507 (2018).
137. Aedes aegypti genome assembly AaegL5.0 - NCBI - NLM. https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_002204515.2/.
138. Timoshevskiy, V. A. *et al.* Genomic composition and evolution of Aedes aegypti chromosomes revealed by the analysis of physically mapped supercontigs. *BMC Biol* 12, 1–13 (2014).
139. Rao, P. N. & Rai, K. S. Genome evolution in the mosquitoes and other closely related members of superfamily Culicoidea. *Hereditas* 113, 139–144 (2008).
140. Kumar, A. & Rai, K. S. Intraspecific variation in nuclear DNA content among world populations of a mosquito, Aedes albopictus (Skuse). *Theoretical and Applied Genetics* 79, 748–752 (1990).
141. Severson, D. W. & Behura, S. K. Mosquito genomics: Progress and challenges. *Annu Rev Entomol* 57, 143–166 (2012).
142. Nene, V. *et al.* Genome sequence of Aedes aegypti, a major arbovirus vector. *Science (1979)* 316, 1718–1723 (2007).
143. Love, R. R., Sikder, J. R., Vivero, R. J., Matute, D. R. & Schrider, D. R. Strong Positive Selection in Aedes aegypti and the Rapid Evolution of Insecticide Resistance. *Mol Biol Evol* 40, (2023).
144. Gloria-Soria, A. *et al.* Global Genetic Diversity of Aedes aegypti. *Mol Ecol* 25, 5377 (2016).
145. Love, R. R., Sikder, J. R., Vivero, R. J., Matute, D. R. & Schrider, D. R. Strong Positive Selection in Aedes aegypti and the Rapid Evolution of Insecticide Resistance. *Mol Biol Evol* 40, (2023).
146. Lee, Y. *et al.* Genome-wide divergence among invasive populations of Aedes aegypti in California. *BMC Genomics* 20, 1–10 (2019).
147. McBride, C. S. *et al.* Evolution of mosquito preference for humans linked to an odorant receptor. *Nature* 515, 222–227 (2014).
148. Asgharian, H. *et al.* Evolutionary genomics of Culex pipiens: global and local adaptations associated with climate, life-history traits and anthropogenic factors. *Proceedings of the Royal Society B: Biological Sciences* 282, (2015).
149. Spadar, A., Collins, E., Messenger, L. A., Clark, T. G. & Campino, S. Uncovering the genetic diversity in Aedes aegypti insecticide resistance genes through global comparative genomics. *Scientific Reports 2024 14:1* 14, 1–12 (2024).

CHAPTER ONE

A next generation targeted amplicon sequencing technique to screen for insecticide resistance mutations in *Aedes aegypti* populations reveals an *rdl* mutation in mosquitoes from Cabo Verde

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	1513859	**Title**	Ms
First Name(s)	Emma Louise		
Surname/Family Name	Collins		
Thesis Title	Utilising genomic approaches to explore genetic diversity and insecticide resistance in Aedes aegypti populations		
Primary Supervisor	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?	PLOS Neglected Tropical Diseases		
When was the work published?	December 2022		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	NA		
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	

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)	For this work I was involved in data generation, analysis, validation. Analysis, visualisation, writing up and reviewing of the results. Conceptualization: Ana Rita Gomes, Lara Ferrero Gomez, Taane G. Clark, Susana Campino. Data curation: Emma L. Collins, Magdalena Hubner, Anton Spadar, Lara Ferrero Gomez, Taane G. Clark, Susana Campino. Formal analysis: Emma L. Collins, Jody E. Phelan, Magdalena Hubner, Susana Campino. Funding acquisition: Lara Ferrero Gomez, Taane G. Clark, Susana Campino. Methodology: Emma L. Collins, Magdalena Hubner, Monica Campos, Daniel Ward, Holly Acford-Palmer, Ana Rita Gomes, Keily Silva. Project administration: Susana Campino. Resources: Ana Rita Gomes, Susana Campino. Supervision: Monica Campos, Lara Ferrero Gomez, Taane G. Clark, Susana Campino. Validation: Emma L. Collins, Monica Campos, Lara Ferrero Gomez, Susana Campino. Visualization: Emma L. Collins Writing – original draft: Emma L. Collins, Susana Campino. Writing – review & editing: Emma L. Collins, Jody E. Phelan, Monica Campos, Holly Acford-Palmer, Ana Rita Gomes, Keily Silva, Lara Ferrero Gomez, Taane G. Clark, Susana Campino

SECTION E

Student Signature	Emma Collins
Date	18/06/2024

Supervisor Signature	Susana Campino
Date	18/06/2024

PLOS NEGLECTED TROPICAL DISEASES

RESEARCH ARTICLE

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

Emma L. Collins[1], Jody E. Phelan[1], Magdalena Hubner[1], Anton Spadar[1], Monica Campos[1], Daniel Ward[1], Holly Acford-Palmer[1], Ana Rita Gomes[2], Keily Silva[3], Lara Ferrero Gomez[3‡], Taane G. Clark[1,4‡], Susana Campino[1‡*]

1 Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom, **2** Laboratory of Pathogen-Host Interactions (LPHI), CNRS, Montpellier University, Montpellier, France, **3** Universidade Jean Piaget (UniPiaget), Praia, Cabo Verde, **4** Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom

‡ These authors are joint authors on this work.
* Susana.campino@lshtm.ac.uk

Citation: Collins EL, Phelan JE, Hubner M, Spadar A, Campos M, Ward D, et al. (2022) 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(12): e0010935. https://doi.org/10.1371/journal.pntd.0010935

Editor: Gordana Rasic, QIMR Berghofer Medical Research Institute, AUSTRALIA

Received: March 30, 2022

Accepted: November 6, 2022

Published: December 13, 2022

Data Availability Statement: All files are available from the ENA database (accession numbers ERS12467832 to ERS12467857).

Funding: E.L.C. is funded by Medical Research Council LiD PhD studentships. J.E.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).

Abstract

Aedes mosquito vectors transmit many viruses of global health concern, including dengue, chikungunya and Zika. These vector-borne viral diseases have a limited number of treatment options, and vaccines vary in their effectiveness. Consequently, integrated vector management is a primary strategy for disease control. However, the increasing emergence and spread of insecticide resistance is threatening the efficacy of vector control methods. Identifying mutations associated with resistance in vector populations is important to monitor the occurrence and evolution of insecticide resistance and inform control strategies. Rapid and cost-effective genome sequencing approaches are urgently needed. Here we present an adaptable targeted amplicon approach for cost-effective implementation within next generation sequencing platforms. This approach can identify single nucleotide polymorphisms (SNPs) and small insertions and deletions (indels) in genes involved in insecticide resistance in *Aedes aegypti* mosquitoes. We designed and tested eleven amplicons, which included segments of the *ace-1* (carbamate target), the *Voltage-Gated Sodium Channel* (*vgsc*; pyrethroids, DDT and organochlorines), and *rdl* (dieldrin) genes; thereby covering established knockdown resistance (kdr) mutations (e.g., S989P, I1011M/V, V1016G/I and F1534C), with the potential to identify novel ones. The amplicon assays were designed with internal barcodes, to facilitate multiplexing of large numbers of mosquitoes at low cost, and were sequenced using an Illumina platform. Our approach was evaluated on 152 *Ae. aegypti* mosquitoes collected in Cabo Verde, an archipelago with a history of arbovirus outbreaks. The amplicon sequence data revealed 146 SNPs, including four non-synonymous polymorphisms in the *vgsc* gene, one in *ace-1* and the 296S *rdl* mutation previously associated with resistance to organochlorines. The 296S *rdl* mutation was identified in 98% of

H.A.P and M.H are funded by a BBSRC LIDo PhD studentship. S.C. was funded by MRC UK grants (ref. MR/M01360X/1, MR/R025576/1, and MR/R020973/1). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

mosquitoes screened, consistent with the past use of an organochlorine compound [redacted]. Overall, our work shows that targeted amplicon sequencing is a rapid, robust, and cost-effective tool that can be used to perform high throughput monitoring of insecticide resistance.

Author summary

Many viruses, such as dengue and Zika, are transmitted by *Aedes aegypti* mosquitoes. These vector-borne diseases are a major public health problem in the tropics and subtropics worldwide. The primary strategy to reduce their burden is vector control, including through the application of insecticides. However, many mosquito populations have developed resistance to insecticides. Here, we present a rapid, robust, and cost-effective tool that allows the screening of genes associated with insecticide resistance across many *Aedes aegypti* mosquitoes, identifying known resistance and novel mutations. This assay will support vector control strategies by informing on the emergence and spread of insecticide resistance mutations across *Aedes aegypti* populations.

Introduction

Vector-borne diseases pose a major risk to public health, causing ~700k deaths every year [1]. Arthropod-borne viruses (arboviruses), causing dengue, Zika and chikungunya infections, contribute substantially to the global burden of disease. Mosquitoes of the genus *Aedes* are responsible for the transmission of many arboviruses, with *Ae*. *aegypti* being one of the most competent vectors. The distribution of *Ae*. *aegypti* has increased dramatically in recent years, predominantly due to their adaptation to urban environments and the globalization of human activities [2,3]. Vector control strategies are essential to prevent arboviral spread, largely due to the lack of effective vaccines and available antiviral drugs. Vector control predominantly involves the use of insecticides, either in the form of spraying or treated bed nets [4,5]. However, the intensive use of insecticides worldwide has led to the emergence of resistance to pyrethroids, organochlorines, carbamates, neonicotinoids and organophosphates [6,7], which is threatening the effectiveness of vector control campaigns for important vector-borne diseases.

The main mechanisms of insecticide resistance are target site, metabolic and cuticular, and behavioural avoidance [6,8]. Target site resistance is caused by point mutations in genes that encode the protein targeted by the insecticide, including voltage gated sodium channels (*vgsc* gene), acetylcholinesterase (*ace-1*) and the γ-aminobutyric acid (GABA) receptor (resistance to dieldrin locus *rdl*) [9]. VGSC proteins are present in the nervous system and are a target for DDT and pyrethroids. Knockdown resistance (kdr) to these two insecticides has been linked to multiple target site mutations in the *vgsc* gene in many insects [9]. Acetylcholinesterase (AChE) enzymes hydrolyse the neurotransmitter acetylcholine at the synaptic cleft and hence terminate nerve signals. Organophosphates and carbamate insecticides bind to AChE thus disrupting nerve impulses and ultimately causing death. A single target site mutation in the *ace-1* gene (G119S), encoding AChE, has been shown to inhibit the insecticidal action in many mosquito vectors [10,11] including *Ae*. *aegypi* [12]. Finally, mutations in the *rdl* gene (e.g. A301S *Drosophila melanogaster*, A296 in many mosquito species including *Ae*. *aegypti*, *Ae*. *albopictus and Anopheles arabiensis* [13] *and* V327I *An*. *funestus* [14]), have been associated with resistance to organochlorine insecticides in *Anopheles*, *Aedes* and *Culex* vectors [15–17].

Current methods for the identification of insecticide resistance involve biological and biochemical assays [18–20] which are time-consuming, require multiple repeats, and involve subjective judgement of mosquito knockdown. Additionally, bioassays can often only detect resistance when frequencies are already high, and molecular methods may be required if resistant alleles are at lower frequencies [21,22]. Molecular methods have been developed for the detection of mutations associated with insecticide resistance and can be an effective approach to monitor resistant alleles when diagnostic markers predictive of vector control intervention failure are known [10,23–25]. Given the recent innovations and cost reductions in molecular techniques, testing based on the molecular underpinning of insecticide resistance is likely to be an effective approach to support monitoring. This innovation would allow public health organizations to monitor the emergence and spread of known resistance mutations and detect the appearance of new genetic polymorphisms. In addition, alongside biological and biochemical assays of susceptibility, molecular surveillance can identify novel resistance markers and provide insights into the mechanisms of action.

Amplicon sequencing is a targeted next-generation sequencing method that allows for the high throughput detection of low frequency variants in specific genomic regions of interest. Here we describe a multiplexed amplicon sequencing approach targeting the *vgsc*, *ace-1* and *rdl* loci of *Ae*. *aegypti* mosquitoes. The assays target eleven genomic regions across these three genes where mutations associated with insecticide resistance have been reported in *Aedes* and other vectors. A dual index approach was used with an individual barcoding system that allows for the pooling and simultaneous sequencing of multiple PCR products. Sequence data are later demultiplexed to individual mosquitoes and genes from raw sequence data, providing a fast and cost-effective surveillance method to detect mutations involved in insecticide resistance. To demonstrate the utility of our approach, it was applied to *Ae*. *aegypti* mosquitoes from Cabo Verde, an archipelago located 500 kilometres off the coast of West Africa. Dengue and Zika outbreaks have been reported in Cabo Verde [26–28]. In 2009, more than 21,000 cases of dengue fever were diagnosed and in 2015 an epidemic of Zika caused at least 7,580 reported cases. To prevent vector-borne disease, Cabo Verde has a history of applying several strategies to combat *Anopheles*, *Aedes* and *Culex* vectors, including the past use of DDT (organochlorine) and recent spraying of temephos (organophosphate) and deltamethrin (pyrethroid) insecticides [29]. Compared to *Anopheles* mosquitoes, little is known about *Aedes* insecticide resistance and associated mutations, in both Cabo Verde, and in Africa as a whole [30]. VGSC mutations (e.g., V1016I, F1534C) that confer resistance to pyrethroids have been reported at low frequency in Cabo Verde [26] but no other mutations were investigated. Using the dual index amplicon-based approach on an Illumina sequencing platform, we screen for known mutations associated with insecticide resistance in *Ae*. *aegypti* sourced from Cabo Verde. Through this work, we demonstrate the utility of our approach for detecting insecticide resistance mutations as well as novel polymorphism, to inform vector-borne disease control efforts.

METHODS

Amplicon primer design

A list of target site insecticide resistance mutations in *Aedes* vectors was extracted from an OVID search and recent reviews [9,31–41]. Overall, twelve mutations linked to insecticide resistance were found, nine in the *vgsc* (V410L, G923V, L982W, S989P, I1011V/M, V1016I/G, T1520I, F1534C/L, D1763Y), one in *rdl* (A301S), and one in *ace-1* (G119S). Sequences for *vgsc* (AAEL023266-RL), *ace-1* (AAEL000511-RJ) and *rdl* (AAEL008354-RA) were extracted from publicly available assemblies for *Ae*. *aegypti* (LVP AGWG). The mutations of interest were

Table 1. Amplicon regions, mutations previously described and primer sequences.

Gene	Primer Name	Mutations	Primer Sequence (5'-3')*	Amplicon Length	Multiplex
vgsc	DomainI	V410L	F: [TTTCGTCTAATGACCCAAGA] R: [ARAGAWTTCGCTCACCCG]	464	1
vgsc	DomainIIIExon35	T1520I, F1534C/L	F: [GGATCCAGATCATGAACGAY] R: [GATGATCATGTCGAACTTCT]	480	1
Rdl	Rdl_Aeg	A301S	F: [CCAACCGATGTATCTTCTTC] R: [CTGGTTATTTGTACAAGTAGCA]	498	1
Ace-1	Ace1	G119S	F: [TCGCYTRGCCGAAGCCGT] R: [CASGTGAARTGATAATCTCCSAC]	468	1
vgsc	DomainIIS4	G923V	F: [TCTAGATTTAGYGACTCCAR] R: [TACCGATGTAGTTCTTGCC]	444	2
vgsc	DomainII	L982W, S989P, I1011V/M, V1016I/G	F: [ACTCRTTCATGATCGTGTTC] R: [GACTTGATCCAGTTGGAGA]	498	2
vgsc	DomainIIIExon36	NA	F: [GTGTCATCATCGACAACTTC] R: [CACACCTAAAATGGACAGGA]	489	2
vgsc	DomainIV	D1763Y	F: [GCGATCTSATCGAGAAGTA] R: [ATGCTAGCAARTACGTGATG]	495	2
vgsc	DomainIIExon26	982W, S989P, I1011V/M	F: [TCACCTTATGCTAAGACTTCA] R: [GGGAAACAATTTGTCGGTTA]	494	3
vgsc	DomainIIIExon33_34	T1520I, F1534C/L	F: [AACTCTCTATTCCCGCTTG] R: [GCAGATCATTCGTAACAAGT]	469	3
vgsc	Domain IVS6	NA	F: [TGTTGGACGGTATCATCAA] R: [CCTCGATCGGRTTACCTTT]	456	3

*Primer sequences underlined show incompatible nucleotides with *Aedes albopictus* reference sequence (FOSHAN).

https://doi.org/10.1371/journal.pntd.0010935.t001

identified by performing a BLAST sequence alignment of the *Ae. aegypti* reference genome against the sequences of species in which the mutation had been described (see S1 Table).

A summary of the amplicon approach is provided (S1 Fig). Forward and reverse primers were designed using PrimerBLAST software to amplify regions of 450-500bp that contained known SNP loci or regions of interest. The primers ranged from 18-25bp in length and were designed to have similar annealing conditions to allow for multiplexing (Table 1). Primers targeting eleven regions were selected across three genes. Primers were checked for cross-hybridization in each multiplex combination using ThermoFisher Multiplex Primer Analyser. Each multiplex PCR used a combination of at least 3 targets (Table 1). To allow pooling of individual mosquito PCR products, a 6bp barcode was added to the 5' end of each primer (forward and reverse unique barcodes) to distinguish individual mosquito products after sequencing (S2 Table). To each mosquito DNA sample, forward and reverse 6bp barcodes were assigned across all loci, allowing the pooling of many different mosquito PCR products. Partial Illumina tails of ~30bp were also added to the 5' end of each primer, just after the 6bp barcode primer-tag, for compatibility with commercial sequencing. This feature allows the pools to be sequenced using any Illumina sequencer, in-house or by commercial providers. A second PCR carried out by the commercial sequencing company enabled the addition of Illumina adaptors and indexes if necessary to pool experiments (S1 Fig).

Mosquito collection, DNA extraction, PCR, and purification

A pooled sample of *Ae. aegypti* was used as control to test the amplicon sequencing primers. *Ae. aegypti* mosquitoes were collected in the city of Praia, on Santiago island in Cabo Verde [42]. The mosquitoes were all morphologically identified as *Ae. aegypti*. Mosquito DNA was

extracted according to the manufacturer's instructions for Qiagen DNeasy Blood and Tissue kits. DNA concentration was determined using Qubit. Multiplex PCR was carried out under the conditions: Initial denaturation (98.0˚C, 30 seconds) followed by 35 cycles of denaturation (98.0˚C, 10 seconds), annealing (60.4˚C, 70 seconds), and extension (72.0˚C, 90 seconds). Each reaction comprised reagents from Q5 High-Fidelity PCR kit (New England Biolabs, UK), 4μl of DNA template and 0.5 μl of each forward and reverse primer at 10pmol/μl. Three multiplex PCRs were carried out (Table 1). PCR products were visualised on a SYBR safe (Cambridge Bioscience,UK) 1% agarose gel alongside a 100bp ladder. PCR products were purified with AMPure XP magnetic beads (Beckman Coulter), using a ratio of 0.8:1 (μl of beads to DNA).

Sequencing and bioinformatics analysis

The DNA concentration of purified PCR products was tested with the Qubit 2.0 fluorimeter HS DNA kit (ThermoFisher, Waltham, MA, USA). DNA concentrations varied between 7.9 and 47.7 ng/μl. All PCRs were diluted and grouped in equal concentrations to create an overall pool of 20 ng/μl in 25 μl total volume, containing around 220 amplicons (11 amplicons across 20 mosquitoes per pool = 220 amplicons). A second PCR to insert IIlumina adaptors and indexes to pool experiments (no further library preparation is required), which allows many pools to be sequenced in the same Illumina run. This second PCR step, followed by amplicon sequencing, was performed by Genewiz (from Azenta Life Sciences) at a cost of ~US$ 60 per pool (~220 amplicons, US$ 0.30 per amplicon). A minimum of 50,000 reads (250bp read pairs) were obtained per pool, equivalent to an average of ~220 reads per amplicon. From the sequenced pool, individual mosquito data were demultiplexed based on the 6bp barcode primer-tag in each forward and reverse primer using an inhouse pipeline (https://github.com/LSHTMPathogenSeqLab/amplicon-seq), which removed any mis-tagging across barcodes. Sequences were trimmed and aligned to the *Ae. aegypti* reference (LVP AGWG). Sequence data were checked for quality using FastQC (v 0.11.5). Paired end reads were mapped against the reference sequence using the BWA-MEM algorithm (v0.7.17, default parameters). SNPs and small indels were called using freebayes (v1.3.5,—haplotype-length -1) and GATK HaplotypeCaller (v 4.1.4.1, default parameters) software tools. Variants detected across either software caller were used as an initial set for characterisation across amplicons. High quality SNPs were identified using filters that included a minimum phred quality score of 30 per called base, a minimum depth of 50 reads, and a minimum allele depth of 10-fold. Only SNPs that were present in more than one mosquito, and present across two independent pools were retained. The bioinformatics pipeline is summarised (S2 Fig). The distribution of allele depth and frequency for each SNP was analysed to assign threshold cut-offs for genotyping calls as described for diploid organisms [43]. The annotation of the SNP identified was called using bcftools csq (v1.1.0, default parameters). Sanger sequencing using individual forward or reverse primers, was performed for 38 mosquitoes to confirm the findings of the amplicon sequencing for the *ace* and *rdl* amplicon regions. Chi squared tests were performed to assess possible deviations from the Hardy-Weinberg Equilibrium. Tajima's D test was applied to distinguish between sequences evolving randomly ("neutrally") and evolving under a non-random process (e.g., selection, demographic expansion/contraction). This test was implemented using MEGA11 software [44].

Results

Target amplicon representation and variant calling

A total of eleven ~500bp amplicon assays were designed across *vgsc* (nine amplicons spanning the positions of nine known insecticide mutations), *ace-1* (1 amplicon; 1 known mutation)

and *rdl* (1 amplicon; 1 known mutation) loci (see Fig 1). Each amplicon assay was validated individually and across multiplex PCRs, and it was possible to amplify the eleven loci multiplexed in three different PCR reactions (Table 1).

A total of 152 *Ae*. *aegypti* mosquitoes sourced from Cabo Verde were processed individually using the multi-locus amplicon assays. For each mosquito, 11 amplicons were obtained containing the same combination of barcodes (forward and reverse 6bp barcode primer-tag) (S2 Table). A unique combination was used for each mosquito across the 11 amplicons in each pool. Each pool consisted of 220 amplicons (11 loci across 20 mosquitoes) and was sequenced on an Illumina platform. This number was selected to obtain a high coverage per amplicon, with a minimum of 50,000 reads per pool sequenced (average 220 reads per amplicon) being obtained (See Methods). A bioinformatics pipeline was developed to demultiplex from the pools each mosquito data (using 6bp barcode primer-tag) from raw sequencing data, remove mistagging sequences, perform alignment to reference strain, call variants and genotypes, whilst removing low-quality data and variants (see Methods). The pipeline revealed some minor differences in genomic coverage, reflecting differences in the amplification of regions, due to the expected differences in the efficiency of primer binding. Overall, coverage varied between amplicons with DomainIIExon26Aeg (*vgsc* gene) having the lowest mean coverage of 120-fold while DomainIIIExon36 (*vgsc* gene) had nearly 20 times higher-fold coverage. The average read count over all amplicons was 936-fold (Table 2).

A total of 146 SNP variants were identified. To validate our pipeline, 87 *Ae*. *aegypti* individual mosquitoes were re-sequenced and added to different pools, leading to all SNPs being confirmed and a concordance of 84.2% obtained for genotype calls. The genotype differences were observed at only five positions, between homozygous and heterozygous calls, and only in the mosquitoes with allele ratios of 0.8–1.0, where one allele in all tested mosquitoes and replicates is present in the majority of total reads.

Further, Sanger sequencing was performed for 38 mosquitoes to confirm the findings of the amplicon sequencing for the *ace* and *rdl* amplicon regions. A 90% concordance in genotyping calls between Sanger sequencing and amplicon sequencing was observed for the *ace* amplicon. For the only SNP detected in the *rdl* gene a 67% genotype concordance was observed between the two methods. Again, discordant genotype calls were observed in heterozygous mosquitoes using amplicon sequencing that were homozygous with Sanger sequencing, and these heterozygous mosquitoes had an allele ratio close to 0.8, showing an increase of one allele in the total reads. It is possible that mistagging rearrangement, as previously highlighted [45,46], could lead to an unexpected distribution of allele frequencies across mosquitoes and differences in genotype calls, particularly leading to excess heterozygous genotypes. By assigning threshold cut-offs based on allele ratios as described for other diploid organisms [43] we can identify and reassign the most likely genotype. Therefore, by recalling heterozygous genotypes with an allele ratio from 0.8–1 into homozygous, a 100% genotype concordance was obtained. The allele frequency spectrum across all genes reveals an excess of low frequency alleles (~40% of SNPS with minor allele frequency (MAF) < 0.1) close to neutrality (Tajima D' = -0.56; close to zero). There are some distortions from HWE (37% of exonic and 43% of intronic SNPs with $P < 0.001$) (Table 3), which could be the result of the amplicon method leading to an excess of heterozygous. These results need to be further investigated in larger studies, and by including more mosquito generations.

SNPs and insecticide resistance variants

The analysis pipeline detected 146 SNP variants of which 45 were exonic. The number of SNPs identified in each region was highly variable, with the majority in the *vgsc* Exon 35 amplicon

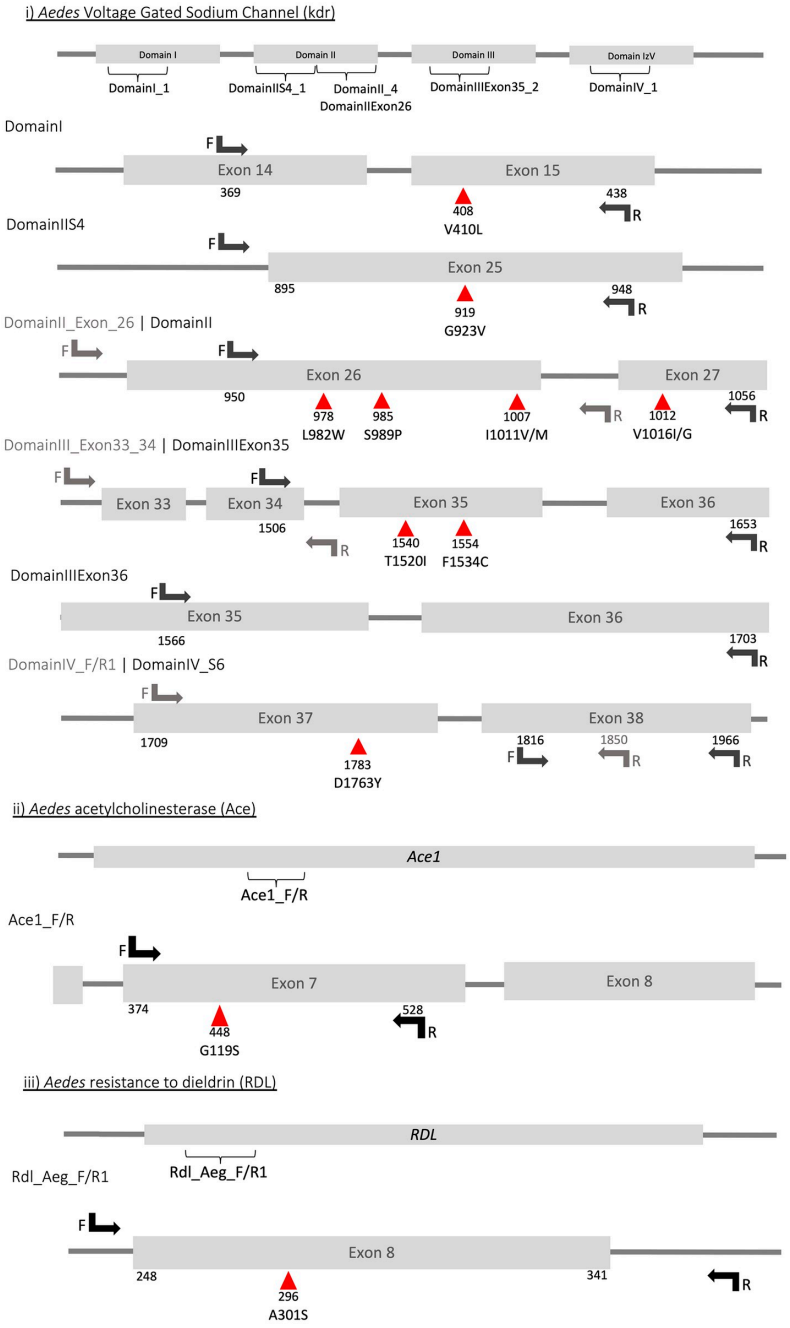


Fig 1. Aedes mutations and primer positions for *ace-1*, *vgsc* and *rdl* genes. Previously described mutations associated with insecticide resistance are represented with red triangles and primers are shown with arrows. The first and last codon that are amplified are under each amplicon diagram.

https://doi.org/10.1371/journal.pntd.0010935.g001

Table 2. The eleven amplicons and their sequencing performance in 152 Cabo Verde *Ae. aegypti*.

Locus	Domain	Exon	known mutations [observed]	Amplicon Length	Mean read depth before filtering	No. SNPs * (Exonic)
ace-1	-	-	1 [0]	468	669	3 (3)
vgsc	I	-	1 [0]	464	328	24 (0)
vgsc	II	-	4 [0]	498	137	25 (2)
vgsc	II	26	3 [0]	494	120	27(3)
vgsc	III	33_34	0 [0]	469	1389	26 (6)
vgsc	III	35	2 [0]	480	2289	38 (15)
vgsc	III	36	0 [0]	489	2352	32 (13)
vgsc	II	S4	1 [0]	444	191	7 (1)
vgsc	IV	-	1 [0]	495	516	6 (3)
vgsc	IV	S6	0 [0]	456	613	10 (10)
rdl	-	-	1 [1]	498	418	1 (1)
Total	-	-	14 (9 unique)	-	-	240 (72) (146 unique)

*Amplicon regions overlap therefore SNPs are found in multiple amplicons

https://doi.org/10.1371/journal.pntd.0010935.t002

(38 SNPs, 11 exonic) and the least in the *rdl* amplicon (1 exonic SNP) (Table 3 and Fig 2). Almost all exonic SNPs led to synonymous changes (39 SNPs), and six led to missense genetic polymorphisms. Thirteen SNPs occurred in predicted splice regions, determined using the bcftools csq tool. Polymorphisms in splice regions may have little effect on gene function, but can alter the splicing pattern, such as skipping an exon or keeping of large segments of an intron in the mRNA, which can affect the protein function [47]. Further functional studies will be needed to confirm the *in silico* predictions of SNPs in splice regions.

The highest number of polymorphisms was found in the *vgsc* gene, largely due to the higher number of primers (9 of 11 amplicons) targeting these loci, with 142 SNPs identified. Mutations in the *vgsc* gene or *ace* gene previously associated with insecticide resistance (S1 Table) were not identified. Four amino-acid substitutions (V977L, K1577T, N1595T, P1612H) were found in *vgsc* gene. Of interest is the substitution V977L as it is next to a known mutation, L978W, (homologous to position L982W in *Drosophila melanogaster*), which is reported to confer resistance to pyrethroids [48,49].

In the *rdl* amplicon, we identified the amino acid substitution A296S, known to be associated with resistance to organochlorines [33]. The A296S mutation (analogous to position 301 in *D. melanogaster*) was identified in 47 mosquitoes (47/48; 70.1% heterozygous; 27.1% mutant homozygous). In the *ace-1* amplicon, three SNPs were found, two being synonymous (T506T, D444D) and one missense translation (L466V). For the T506T mutation (161500076 T>A; n = 108), which has been previously reported in Indonesian mosquitoes [50], all mosquitoes were homozygous for the non-reference allele. The L466V amino acid change was found at a low frequency (1/105, 0.95%; 100% heterozygous). The known G119S insecticide resistance mutation (position G448S in *Ae. aegypti*) was not detected in any of the mosquitoes investigated here.

Discussion

Insecticide resistance is a threat to vector control programs worldwide. Traditional methods to identify resistance can be subjective and time-consuming, therefore molecular surveillance is becoming an attractive option to determine the widespread distribution of insecticide resistance and to complement diagnostic bioassays. Our study successfully demonstrates that multi-locus target amplicon sequencing can be used to identify insecticide resistance

Table 3. Position and frequency of synonymous and non-synonymous SNPs in each gene for *Ae. aegypti*.

Chromosome	Exon	Position	Reference allele	Alternative allele	Consequence	Number	Alternative Allele Frequency (%)
3	Exon 7	161500198	A	C	466L > 466V	105	1
3	Exon 26	315984096	C	T	977V > 977L	80	51.4*
3	Exon 35	315939050	G	T	1612P > 1612H	147	12.2
3	Exon 35	315939101	T	G	1595N > 1595T	139	9
3	Exon 35	315939155	T	G	1577K > 1577T	106	0.9
2	Exon 8	41847790	G	T	296A > 296S	68	62.5*
3	Exon 7	161500076	T	A	Synonymous	108	100
3	Exon 7	161500262	G	A	Synonymous	104	29.3*
3	Exon 26	315984075	G	A	Synonymous	88	10.3
3	Exon 26	315984159	C	T	Synonymous	38	40.8*
3	Exon 34	315939469	G	A	Synonymous	108	2.8
3	Exon 34	315939517	G	T	Synonymous	107	3.3
3	Exon 34	315939547	G	A	Synonymous	112	7.6
3	Exon 33	315939648	C	T	Synonymous	105	27.1*
3	Exon 35	315939229	G	A	Synonymous	140	10.4
3	Exon 35	315939241	A	G	Synonymous	137	71.9*
3	Exon 35	315939244	G	A	Synonymous	138	1.8
3	Exon 35	315939274	A	G	Synonymous	139	1.8
3	Exon 35	315939283	C	T	Synonymous	131	11.5
3	Exon 34	315939367	C	A	Synonymous	148	10.5*
3	Exon 34	315939373	G	A	Synonymous	149	0.3
3	Exon 36	315938745	G	A	Synonymous	83	53.4*
3	Exon 36	315938760	A	G	Synonymous	78	72.8*
3	Exon 36	315938772	C	T	Synonymous	80	3.6
3	Exon 36	315938775	C	T	Synonymous	69	28.3*
3	Exon 36	315938778	C	T	Synonymous	66	43.2*
3	Exon 36	315938832	C	T	Synonymous	75	2.7
3	Exon 36	315938946	A	C	Synonymous	141	0.4
3	Exon 35	315939088	C	T	Synonymous	130	1.5
3	Exon 35	315939112	T	C	Synonymous	110	73.6*
3	Exon 35	315939166	C	T	Synonymous	108	55.1*
3	Exon 25	315998391	C	T	Synonymous	60	3
3	Exon 37	315932072	G	A	Synonymous	95	16.7
3	Exon 37	315932142	A	G	Synonymous	96	42.1*
3	Exon 37	315932184	G	A	Synonymous	92	72.9*
3	Exon 38	315931422	G	A	Synonymous	96	12
3	Exon 38	315931428	C	T	Synonymous	97	14.4
3	Exon 38	315931440	T	C	Synonymous	95	5.3
3	Exon 38	315931470	T	C	Synonymous	94	14.9*
3	Exon 38	315931479	C	A	Synonymous	97	3.6
3	Exon 38	315931485	G	A	Synonymous	95	5.3
3	Exon 38	315931557	T	C	Synonymous	97	57.7*
3	Exon 38	315931563	T	C	Synonymous	95	16.3*
3	Exon 38	315931575	G	A	Synonymous	94	4.8
3	Exon 38	315931578	G	A	Synonymous	93	11.3

* Significant deviation from HWE ($p<0.001$)

https://doi.org/10.1371/journal.pntd.0010935.t003

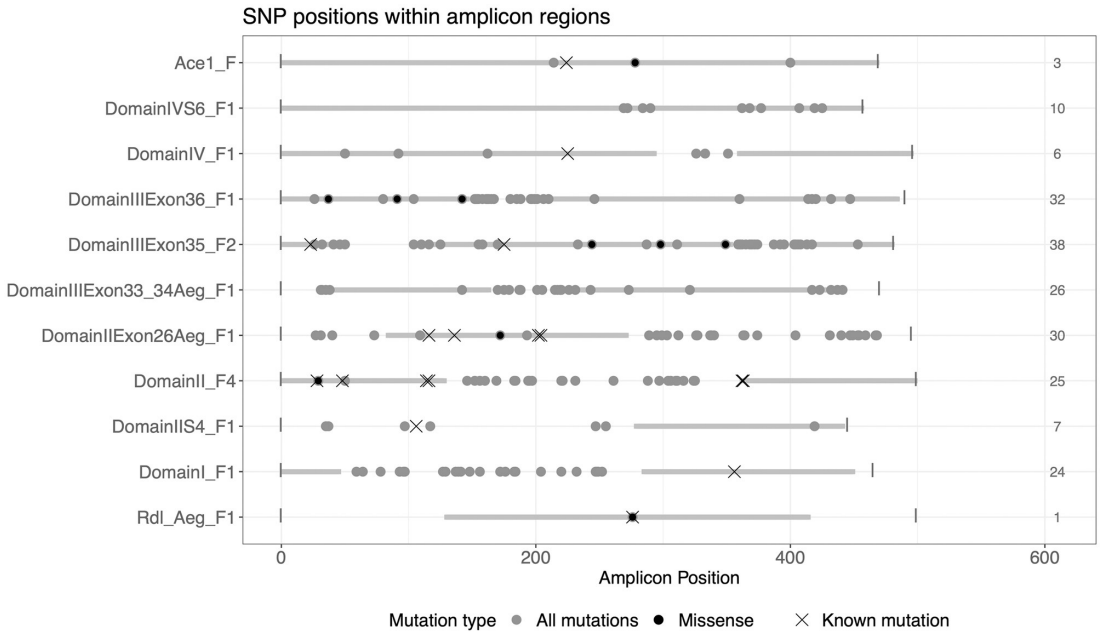


Fig 2. Distribution of SNPs detected in the eleven amplicons. Grey shaded bars illustrate exonic regions. Black points show the position of missense mutations identified, grey points are other SNPs identified, and crosses mark where known mutations associated with insecticide resistance are positioned. The value on the right is the number of SNPs identified in each amplicon.

https://doi.org/10.1371/journal.pntd.0010935.g002

polymorphisms in *Ae. aegypti* mosquitoes in a robust, cost-effective, and high-throughput manner. By testing the approach on *Ae. aegypti* mosquitoes from the city of Praia in Santiago, Cabo Verde, it was possible to identify 146 SNPs across *vgsc*, *rdl*, and *ace-1* loci, including the *rdl* A301S mutation linked to organochlorine insecticide resistance. This observation is probably due to the past use of an organochlorine [redacted] in Cabo Verde. Resistance to this insecticide has been reported to have the least effect on fitness compared to other pyrethroid resistant populations [51]. No other known polymorphisms associated with insecticide resistance were detected. A previous study in Santiago, with samples collected between 2017 and 2018, detected the kdr polymorphism 1016I in two heterozygous individuals from São Lourenço dos Órgãos, and the 1534C mutation at low frequency ($\leq$ 2.0%) in mosquitoes from Praia, but these variants were not observed in previous years (2007 to 2016) [26,42,52]. These results suggest a recent origin of kdr mutations in this island, that could be an independent event or an introduction from neighbouring countries of mainland West Africa. Cabo Verde is located ~500 kilometres off the coast of Senegal, where none of these mutations have been identified in a survey performed in 2017 [53]. These mutations were also not identified in Cameroon, Congo and Central African Republic, but were reported in Ghana and Burkina Faso in high frequency [54–57].

We also have identified further polymorphisms, the majority in intronic regions, some in predicted splice regions or leading to synonymous changes, and only five non-synonymous amino acid substitutions in the *ace-1* and *vgsc* genes. For example, the substitution V977L in the *vgsc* gene which is next to the previously reported L978W (position L982W in *D. melanogaster*) reported to confer resistance to pyrethroids [48,49]. There is also the synonymous

variant in the *ace-1* gene (T506T in *Ae*. *aegypti* reference), which has been described in an Indonesian *Ae*. *aegypti* temephos resistant population [50], and could be in high linkage disequilibrium with a functional polymorphism, but these results have not been confirmed in other studies.

Identification and exploration of these additional polymorphisms can be useful to understand the evolution of these loci and their possible involvement in mechanisms of insecticide resistance, including through linkage with other polymorphisms or *cis*-regulatory elements and the generation of alternative transcripts. Further genotype-phenotype studies will be fundamental to explore the possible involvement of these mutations in insecticide resistance.

More generally, investigations of insecticide resistance in the *Aedes* population have largely focused on *kdr* variations in the *vgsc* gene, with little focus on the other loci investigated here. Molecular testing for insecticide resistance has been limited across West Africa and the wider continent, with only a few studies focusing on a limited number of polymorphisms, typically using a genotyping approach. For instance, the recent studies performed in Senegal, Cameroon, Ghana, Cabo Verde and Ivory Coast [42,53–55,58] focused only on the study of the kdr mutations F1534C, V410L, V1016G/I and S989P in *Ae*. *aegypti*. More surveys are necessary across the *vgsc* gene and other loci to understand the frequency, emergence and spread of genetic variants and their association with insecticide resistance.

As demonstrated here, the multi-locus amplicon approach gives the possibility to inform on both known and discover novel genetic variants in many loci simultaneously and across large numbers of samples. Examining both known and novel SNPs is highly valuable due to the large unexplained variance observed in insecticide resistance. The assays are adaptable and can be extended to include new loci linked to resistance. For example, the current panel does not account for metabolic mechanisms of resistance, which involve the upregulation of detoxifying enzymes in the mosquito, such as cytochrome P450. Only a few SNPs have been associated with this type of resistance and the underlying genes involved remain unclear, but our assays can be extended with further understanding. Sequence capture followed by deep sequencing has been applied to *Ae*. *aegypti* to investigate copy number variations (CNVs) and polymorphisms of detoxification enzymes [59]. However, this system requires the production of capture libraries with overlapping RNA probes, becoming a more expensive and complicated approach than using multiplex PCR assays. Quantitative PCR, like the one developed by Cattel *et al* [60], is still a main approach for the rapid detection and copy number quantification in *Ae*. *Aegypti*.

The multi-locus PCR amplicon approach can be applied across many loci and it is possible to pool large numbers of samples which can be differentiated by unique barcode combinations. Further, there is no need to prepare libraries for Illumina sequencing, as the PCR stages already include Illumina-compatible flow-cell adaptor sequences, leading to multi-locus amplicon pools that can be sequenced by commercial providers at relatively low cost (<$0.50 USD per amplicon). The same amplicons can also be sequenced using portable sequencers (e.g., Oxford Nanopore MinIon), leading to more rapid and informative surveillance of insecticide resistance at field sites, and an improved response to emerging resistance. Relatedly, there is the potential to pool mosquitoes before the PCR step, for rapid overall population surveillance, as opposed to individual mosquito amplification as performed here [21]. This facilitates application to lower income settings, where *Aedes* borne diseases are endemic, and the benefits of informed vector control will be greater. Specifically, the identification of important insecticide resistance mutations would provide early warning and evidence for the need to change the insecticide class prior to total inefficacy.

The main limitation of our amplicon approach is the prior need of information concerning which loci are associated with insecticide resistance. There is no substitute for insecticide

bioassays to determine phenotype response. The multi-locus amplicon approach can be used to complement bioassays, and obtain genotyping data alongside phenotypic information, to inform functional work seeking to understand the molecular mechanisms underlying insecticide resistance. Our assay does not currently detect the genomic contributors of metabolic resistance; however, our approach is highly flexible, and further regions of interest can be added to the panel where required. The design of the methodology to target regions where mutations are likely to occur is also beneficial as it means that some mutations that are newly discovered are already captured in the panel. For example, a new mutation (A1007G) was recently described in *Ae*. *aegypti* to be putatively associated with DDT resistance, and is already captured by the DomainII_F4 amplicon, although it was not detected in our mosquitoes [61]. Moreover, there is the possibility to include pathogen screening in parallel with insecticide resistance characterisation to allow for monitoring of arboviruses and other pathogens in an integrated surveillance programme. It has already been demonstrated that it is possible to detect the malaria parasites from human blood using amplicon sequencing [62,63].

Overall, our work outlines a cost-effective, robust, and high-throughput methodology to screen *Ae*. *aegypti* mosquitoes for both known and putative novel insecticide resistance mutations. The integration of 5' tag barcodes allow the pooling of many mosquitoes and loci within applications of next-generation sequencing, and their subsequent separation during analysis. These molecular and bioinformatic approaches can be implemented by vector control programs to monitor insecticide resistance and improve the efficacy of other approaches. The extension of the approach to other genomics regions, pathogens and other vectors will further assist with supporting control strategies of vector-borne diseases and reducing their global burden.

Supporting information

S1 Fig. Steps for multiplex amplicon sequencing. In the first PCR, target genes are amplified and partial Illumina tails and 6bp barcodes included in primers to differentiate individual samples. In a second step the amplicons are pooled across samples. After, a second PCR is performed in each pool, the Illumina adapters and indexes are added, and pools are ready to be sequenced using an Illumina platform.
(TIF)

S2 Fig. Flow chart of the bioinformatics pipeline.
(TIF)

S1 Table. Amino acid mutation positions for the reference organism and corresponding *Ae. aegypti*.
(DOCX)

S2 Table. The 6bp barcodes and partial Illumina tails added to the 5' end of the forward and reverse primers.
(DOCX)

S3 Table. Details of the SNPs detected in splice regions.
(DOCX)

Author Contributions

Conceptualization: Ana Rita Gomes, Lara Ferrero Gomez, Taane G. Clark, Susana Campino.

Data curation: Jody E. Phelan, Magdalena Hubner, Anton Spadar, Lara Ferrero Gomez, Taane G. Clark, Susana Campino.

Formal analysis: Emma L. Collins, Jody E. Phelan, Magdalena Hubner, Susana Campino.

Funding acquisition: Lara Ferrero Gomez, Taane G. Clark, Susana Campino.

Investigation: Susana Campino.

Methodology: Emma L. Collins, Magdalena Hubner, Monica Campos, Daniel Ward, Holly Acford-Palmer, Ana Rita Gomes, Keily Silva.

Project administration: Susana Campino.

Resources: Ana Rita Gomes, Susana Campino.

Supervision: Monica Campos, Lara Ferrero Gomez, Taane G. Clark, Susana Campino.

Validation: Emma L. Collins, Monica Campos, Lara Ferrero Gomez, Susana Campino.

Visualization: Jody E. Phelan.

Writing – original draft: Emma L. Collins, Susana Campino.

Writing – review & editing: Emma L. Collins, Jody E. Phelan, Monica Campos, Holly Acford-Palmer, Ana Rita Gomes, Keily Silva, Lara Ferrero Gomez, Taane G. Clark, Susana Campino.

References

1. World Health Organisation. Vector-borne diseases. 2020.
2. Kraemer MUG, Sinka ME, Duda KA, Mylne AQN, Shearer FM, Barker CM, et al. The global distribution of the arbovirus vectors Aedes aegypti and Ae. Albopictus. Elife. 2015; 4. https://doi.org/10.7554/eLife.08347 PMID: 26126267
3. Brown JE, Evans BR, Zheng W, Obas V, Barrera-Martinez L, Egizi A, et al. Human impacts have shaped historical and recent evolution in Aedes aegypti, the dengue and yellow fever mosquito. Evolution. 2014; 68: 514–25. https://doi.org/10.1111/evo.12281 PMID: 24111703
4. World Health Organization. Guidelines for Malaria Vector Control. 2019.
5. Achee NL, Gould F, Perkins TA, Reiner RC, Morrison AC, Ritchie SA, et al. A Critical Assessment of Vector Control for Dengue Prevention. PLoS Negl Trop Dis. 2015; 9: 1–19. https://doi.org/10.1371/journal.pntd.0003655 PMID: 25951103
6. World Health Organization. Global report on insecticide resistance in malaria vectors: 2010–2016. World Heal Organ. World Health Organization; 2018.
7. Rivero A, Vézilier J, Weill M, Read AF, Gandon S. Insecticide control of vector-borne diseases: when is insecticide resistance a problem? PLoS Pathog. 2010; 6: e1001000. https://doi.org/10.1371/journal.ppat.1001000 PMID: 20700451
8. Itokawa K, Sekizuka T, Maekawa Y, Yatsu K, Komagata O, Sugiura M, et al. High-throughput genotyping of a full voltagegated sodium channel gene via genomic DNA using target capture sequencing and analytical pipeline MoNaS to discover novel insecticide resistance mutations. PLoS Negl Trop Dis. 2019;13. https://doi.org/10.1371/journal.pntd.0007818 PMID: 31738756
9. Moyes CL, Vontas J, Martins AJ, Ng LC, Koou SY, Dusfour I, et al. Contemporary status of insecticide resistance in the major Aedes vectors of arboviruses infecting humans. PLoS Negl Trop Dis. 2017; 11: 1–20. https://doi.org/10.1371/journal.pntd.0005625 PMID: 28727779
10. Weill M, Malcolm C, Chandre F, Mogensen K, Berthomieu A, Marquine M, et al. The unique mutation in ace-1 giving high insecticide resistance is easily detectable in mosquito vectors. 2004; 13: 1–7. Available: https://pubmed.ncbi.nlm.nih.gov/14728661/
11. Weill M, Luffalla G, Mogensen K, Chandre F, Berthomieu A, Berticat C, et al. Insecticide resistance in mosquito vectors. Nature. 2003; 423: 136–137. https://doi.org/10.1038/423136b PMID: 12736674
12. Muthusamy R, Shivakumar MS. Susceptibility status of Aedes aegypti (L.) (Diptera: Culicidae) to temephos from three districts of Tamil Nadu, India. J Vector Borne Dis. 2015; 52: 159–165. Available: https://www.researchgate.net/publication/279631792_Susceptibility_status_of_Aedes_aegypti_L_Diptera_Culicidae_to_temephos_from_three_districts_of_Tamil_Nadu_India PMID: 26119549
13. Tan WL, Li CX, Wang ZM, Liu MD, Dong YD, Feng XY, et al. First detection of multiple knockdown resistance (kdr)-like mutations in voltage-gated sodium channel using three new genotyping methods in

Anopheles sinensis from Guangxi Province, China. J Med Entomol. 2012; 49: 1012–1020. https://doi.org/10.1603/me11266 PMID: 23025181

14. Wondji CS, Dabire RK, Tukur Z, Irving H, Djouaka R, Morgan JC. Identification and distribution of a GABA receptor mutation conferring dieldrin resistance in the malaria vector Anopheles funestus in Africa. Insect Biochem Mol Biol. 2011; 41: 484–491. https://doi.org/10.1016/j.ibmb.2011.03.012 PMID: 21501685
15. Remnant EJ, Good RT, Schmidt JM, Lumb C, Robin C, Daborn PJ, et al. Gene duplication in the major insecticide target site, Rdl, in Drosophila melanogaster. Proc Natl Acad Sci U S A. 2013; 110: 14706–14710. https://doi.org/10.1073/pnas.1311341110 PMID: 23959864
16. Asih PB, Syahrani L, Rozi IE, Pratama NR, Marantina SS, Arsyad DS, et al. Existence of the Rdl mutant alleles among the Anopheles malaria vector in Indonesia. Malar J. 2012; 11: 57. https://doi.org/10.1186/1475-2875-11-57 PMID: 22364613
17. Shotkoski F, Lee H -J, Zhang H -G, Jackson MB, Ffrench-Constant RH. Functional expression of insecticide-resistant GABA receptors from the mosquito Aedes aegypti. Insect Mol Biol. 1994; 3: 283–287. https://doi.org/10.1111/j.1365-2583.1994.tb00179.x PMID: 7704314
18. WHO. Test procedures for insecticide resistance monitoring in malaria vector mosquitoes Global Malaria Programme. 2016.
19. WHO. Monitoring and managing insecticide resistance in Aedes mosquito populations. Who. 2016;16: 7.
20. Dusfour I, Vontas J, David J-P, Weetman D, Fonseca DM, Corbel V, et al. Management of insecticide resistance in the major Aedes vectors of arboviruses: Advances and challenges. Fuehrer H-P, editor. PLoS Negl Trop Dis. 2019; 13: e0007615. https://doi.org/10.1371/journal.pntd.0007615 PMID: 31600206
21. Mavridis K, Wipf N, Müller P, Traoré MM, Muller G, Vontas J. Detection and monitoring of insecticide resistance mutations in anopheles gambiae: Individual vs pooled specimens. Genes (Basel). 2018; 9: 1.–10. https://doi.org/10.3390/genes9100479 PMID: 30282959
22. Lucas ER, Rockett KA, Lynd A, Essandoh J, Grisales N, Kemei B, et al. A high throughput multi-locus insecticide resistance marker panel for tracking resistance emergence and spread in Anopheles gambiae. Sci Reports 2019 91. 2019; 9: 1–10. https://doi.org/10.1038/s41598-019-49892-6 PMID: 31527637
23. Donnelly MJ, Isaacs AT, Weetman D. Identification, validation and application of molecular diagnostics for insecticide resistance in malaria vectors. Trends Parasitol. 2016; 32: 197. https://doi.org/10.1016/j.pt.2015.12.001 PMID: 26750864
24. Jones CM, Liyanapathirana M, Agossa FR, Weetman D, Ranson H, Donnelly MJ, et al. Footprints of positive selection associated with a mutation (N1575Y) in the voltage-gated sodium channel of Anopheles gambiae. Proc Natl Acad Sci U S A. 2012; 109: 6614–6619. https://doi.org/10.1073/pnas.1201475109 PMID: 22493253
25. Martinez-Torres D, Chandre F, Williamson MS, Darriet F, Bergé JB, Devonshire AL, et al. Molecular characterization of pyrethroid knockdown resistance (kdr) in the major malaria vector Anopheles gambiae s.s. Insect Mol Biol. 1998; 7: 179–184. https://doi.org/10.1046/j.1365-2583.1998.72062.x PMID: 9535162
26. Ayres CFJ, Seixas G, Borrego S, Marques C, Monteiro I, Marques CS, et al. The V410L knockdown resistance mutation occurs in island and continental populations of Aedes aegypti in West and Central Africa. Bonizzoni M, editor. PLoS Negl Trop Dis. 2020; 14: e0008216. https://doi.org/10.1371/journal.pntd.0008216 PMID: 32384079
27. Lourenço J, Monteiro M, Tomás T, Monteiro Rodrigues J, Pybus O, Rodrigues Faria N. Epidemiology of the Zika Virus Outbreak in the Cabo Verde Islands, West Africa. PLoS Curr. 2018; 10. https://doi.org/10.1371/currents.outbreaks.19433b1e4d007451c691f138e1e67e8c PMID: 29637009
28. Salgueiro P, Serrano C, Gomes B, Alves J, Sousa CA, Abecasis A, et al. Phylogeography and invasion history of Aedes aegypti, the Dengue and Zika mosquito vector in Cape Verde islands (West Africa). Evol Appl. 2019; 12: 1797–1811. https://doi.org/10.1111/eva.12834 PMID: 31548858
29. Pires S, Alves J, Dia I, Gómez LF. Susceptibility of mosquito vectors of the city of Praia, Cabo Verde, to Temephos and Bacillus thuringiensis var israelensis. PLoS One. 2020; 15: e0234242. https://doi.org/10.1371/journal.pone.0234242 PMID: 32520941
30. Weetman D, Kamgang B, Badolo A, Moyes C, Shearer F, Coulibaly M, et al. Aedes Mosquitoes and Aedes-Borne Arboviruses in Africa: Current and Future Threats. Int J Environ Res Public Health. 2018; 15: 220. https://doi.org/10.3390/ijerph15020220 PMID: 29382107
31. Haddi K, Tomé HV V., Du Y, Valbon WR, Nomura Y, Martins GF, 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. 2017; 7: 46549. https://doi.org/10.1038/srep46549 PMID: 28422157

32. Low V. L., Vinne-Siow W. Y., Lim A. L., Tan T. K., Leong C. S., Chen C. D., et al. First molecular genotyping of A302S mutation in the gamma aminobutyric acid (GABA) receptor in Aedes albopictus from Malaysia—PubMed. Trop Biomed. 2015;32.

33. Thompson M, Shotkoski F, Ffrench-Constant R. Cloning and sequencing of the cyclodiene insecticide resistance gene from the yellow fever mosquito Aedes aegypti. FEBS Lett. 1993; 325: 187–190. https://doi.org/10.1016/0014-5793(93)81070-G

34. Saavedra-Rodriguez K, Urdaneta-Marquez L, Rajatileka S, Moulton M, Flores AE, Fernandez-Salas I, et al. A mutation in the voltage-gated sodium channel gene associated with pyrethroid resistance in Latin American Aedes aegypti. Insect Mol Biol. 2007; 16: 785–798. https://doi.org/10.1111/j.1365-2583.2007.00774.x PMID: 18093007

35. Brengues C, Hawkes NJ, Chandre F, McCarroll L, Duchon S, Guillet P, et al. Pyrethroid and DDT cross-resistance in Aedes aegypti is correlated with novel mutations in the voltage-gated sodium channel gene. Med Vet Entomol. 2003; 17: 87–94. https://doi.org/10.1046/j.1365-2915.2003.00412.x PMID: 12680930

36. Srisawat R, Komalamisra N, Eshita Y, Zheng M, Ono K, Itoh TQ, et al. Point mutations in domain II of the voltage-gated sodium channel gene in deltamethrin-resistant Aedes aegypti (Diptera: Culicidae). Appl Entomol Zool. 2010; 45: 275–282. https://doi.org/10.1303/AEZ.2010.275

37. Yanola J, Somboon P, Walton C, Nachaiwieng W, Somwang P, Prapanthadara L aied. High-throughput assays for detection of the F1534C mutation in the voltage-gated sodium channel gene in permethrin-resistant Aedes aegypti and the distribution of this mutation throughout Thailand. Trop Med Int Health. 2011; 16: 501–509. https://doi.org/10.1111/j.1365-3156.2011.02725.x PMID: 21342372

38. Chang C, Shen WK, Wang TT, Lin YH, Hsu EL, Dai SM. 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. 2009; 39: 272–278. https://doi.org/10.1016/j.ibmb.2009.01.001 PMID: 19171193

39. Kushwah RBS, Dykes CL, Kapoor N, Adak T, Singh OP. Pyrethroid-Resistance and Presence of Two Knockdown Resistance (kdr) Mutations, F1534C and a Novel Mutation T1520I, in Indian Aedes aegypti. PLoS Negl Trop Dis. 2015; 9: e3332. https://doi.org/10.1371/journal.pntd.0003332 PMID: 25569164

40. Granada Y, Mejía-Jaramillo A, Strode C, Triana-Chavez O. A Point Mutation V419L in the Sodium Channel Gene from Natural Populations of Aedes aegypti Is Involved in Resistance to λ-Cyhalothrin in Colombia. Insects. 2018; 9: 23. https://doi.org/10.3390/insects9010023 PMID: 29443870

41. Kushwah RBS, Kaur T, Dykes CL, Ravi Kumar H, Kapoor N, Singh OP. A new knockdown resistance (kdr) mutation, F1534L, in the voltage-gated sodium channel of Aedes aegypti, co-occurring with F1534C, S989P and V1016G. Parasit Vectors. 2020; 13. https://doi.org/10.1186/s13071-020-04201-3 PMID: 32600469

42. Campos M, Ward D, Morales RF, Gomes AR, Silva K, Sepúlveda N, et al. Surveillance of Aedes aegypti populations in the city of Praia, Cape Verde: Zika virus infection, insecticide resistance and genetic diversity. Parasites and Vectors. 2020; 13. https://doi.org/10.1186/s13071-020-04356-z PMID: 32958043

43. Corrêa RA, Santos D, Goldman GH, Riaño-Pachón DM. ploidyNGS: Visually exploring ploidy with Next Generation Sequencing data. 2016. https://doi.org/10.1101/086488

44. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol Biol Evol. 2021; 38: 3022–3027. https://doi.org/10.1093/molbev/msab120 PMID: 33892491

45. Esling P, Lejzerowicz F, Pawlowski J. Accurate multiplexing and filtering for high-throughput amplicon-sequencing. Nucleic Acids Res. 2015; 43. Available: /pmc/articles/PMC4357712/ https://doi.org/10.1093/nar/gkv107 PMID: 25690897

46. Fiévet A, Bernard V, Tenreiro H, Dehainault C, Girard E, Deshaies V, et al. ART-DeCo: easy tool for detection and characterization of cross-contamination of DNA samples in diagnostic next-generation sequencing analysis. Eur J Hum Genet. 2019; 27: 792–800. https://doi.org/10.1038/s41431-018-0317-x PMID: 30683922

47. Djihinto O, Saizonou HDM, Djogbenou LS, Nolan T, Lee Y. Single nucleotide polymorphism (SNP) in the doublesex (dsx) gene splice sites and relevance for its alternative splicing in the malaria vector Anopheles gambiae [version 1; peer review: 1 approved with reservations, 1 not approved]. 2022 [cited 6 Oct 2022]. https://doi.org/10.12688/wellcomeopenres.17572.1

48. Gan SJ, Leong YQ, bin Barhanuddin MFH, Wong ST, Wong SF, Mak JW, et al. Dengue fever and insecticide resistance in Aedes mosquitoes in Southeast Asia: a review. Parasites Vectors 2021 141. 2021; 14: 1–19. https://doi.org/10.1186/s13071-021-04785-4 PMID: 34112220

49. Kasai S, Itokawa K, Uemura N, Takaoka A, Furutani S, Maekawa Y, et al. Emergence of hyper insecticide-resistant dengue vectors in Indochina Peninsula: threats of concomitant knockdown resistance mutations. bioRxiv. 2022; 2022.03.05.483084. https://doi.org/10.1101/2022.03.05.483084

50. Hasmiwati Rusjdi SR, Nofita E. Detection of ace-1 gene with insecticides resistance in aedes aegypti populations from DHF-endemic areas in Padang, Indonesia. Biodiversitas. 2018; 19: 31–36. https://doi.org/10.13057/biodiv/d190105

51. Saingamsook J, Yanola J, Lumjuan N, Walton C, Somboon P. Investigation of Relative Development and Reproductivity Fitness Cost in Three Insecticide-Resistant Strains of Aedes aegypti from Thailand. Insects. 2019; 10. https://doi.org/10.3390/INSECTS10090265 PMID: 31443487

52. Rocha HDR, Paiva MHS, Silva NM, de Araújo AP, Camacho DDR da R de A, da Moura AJF, et al. Susceptibility profile of Aedes aegypti from Santiago Island, Cabo Verde, to insecticides. Acta Trop. 2015; 152: 66–73. https://doi.org/10.1016/j.actatropica.2015.08.013 PMID: 26307496

53. Sene NM, Mavridis K, Ndiaye EH, Diagne CT, Gaye A, Ngom EHM, et al. Insecticide resistance status and mechanisms in Aedes aegypti populations from Senegal. PLoS Negl Trop Dis. 2021; 15. https://doi.org/10.1371/journal.pntd.0009393 PMID: 33970904

54. Kamgang B, Yougang AP, Tchoupo M, Riveron JM, Wondji C. Temporal distribution and insecticide resistance profile of two major arbovirus vectors Aedes aegypti and Aedes albopictus in Yaoundé, the capital city of Cameroon. Parasit Vectors. 2017; 10. https://doi.org/10.1186/S13071-017-2408-X PMID: 29017606

55. Kamgang B, Wilson-Bahun TA, Yougang AP, Lenga A, Wondji CS. Contrasting resistance patterns to type I and II pyrethroids in two major arbovirus vectors Aedes aegypti and Aedes albopictus in the Republic of the Congo, Central Africa. Infect Dis poverty. 2020; 9. https://doi.org/10.1186/s40249-020-0637-2 PMID: 32114983

56. Djiappi-Tchamen B, Nana-Ndjangwo MS, Mavridis K, Talipouo A, Nchoutpouen E, Makoudjou I, et al. Analyses of insecticide resistance genes in aedes aegypti and aedes albopictus mosquito populations from cameroon. Genes (Basel). 2021; 12. https://doi.org/10.3390/genes12060828 PMID: 34071214

57. Sombié A, Saiki E, Yaméogo F, Sakurai T, Shirozu T, Fukumoto S, et al. High frequencies of F1534C and V1016I kdr mutations and association with pyrethroid resistance in Aedes aegypti from Somgandé (Ouagadougou), Burkina Faso. Trop Med Health. 2019; 47: 2. https://doi.org/10.1186/s41182-018-0134-5 PMID: 30787670

58. Kudom AA. Entomological surveillance to assess potential outbreak of Aedes-borne arboviruses and insecticide resistance status of Aedes aegypti from Cape Coast, Ghana. Acta Trop. 2020;202. https://doi.org/10.1016/j.actatropica.2019.105257 PMID: 31682813

59. Faucon F, Dusfour I, Gaude T, Navratil V, Boyer F, Chandre F, et al. Identifying genomic changes associated with insecticide resistance in the dengue mosquito Aedes aegypti by deep targeted sequencing. Genome Res. 2015; 25: 1347–1359. https://doi.org/10.1101/gr.189225.115 PMID: 26206155

60. Cattel J, Haberkorn C, Laporte F, Gaude T, Cumer T, Renaud J, et al. A genomic amplification affecting a carboxylesterase gene cluster confers organophosphate resistance in the mosquito Aedes aegypti: From genomic characterization to high-throughput field detection. Evol Appl. 2021; 14: 1009–1022. https://doi.org/10.1111/eva.13177 PMID: 33897817

61. Akhir MAM, Wajidi MFF, Lavoué S, Azzam G, Jaafar IS, Awang Besar NAU, et al. Knockdown resistance (kdr) gene of Aedes aegypti in Malaysia with the discovery of a novel regional specific point mutation A1007G. Parasites and Vectors. 2022; 15: 1–15. https://doi.org/10.1186/S13071-022-05192-Z/TABLES/5

62. Nag S, Dalgaard MD, Kofoed PE, Ursing J, Crespo M, Andersen LOB, et al. High throughput resistance profiling of Plasmodium falciparum infections based on custom dual indexing and Illumina next generation sequencing-technology. Sci Rep. 2017; 7. https://doi.org/10.1038/s41598-017-02724-x PMID: 28546554

63. Makunin A, Korlević P, Park N, Goodwin S, Waterhouse RM, Wyschetzki K, et al. A targeted amplicon sequencing panel to simultaneously identify mosquito species and Plasmodium presence across the entire Anopheles genus. Mol Ecol Resour. 2022; 22: 28–44. https://doi.org/10.1111/1755-0998.13436 PMID: 34053186

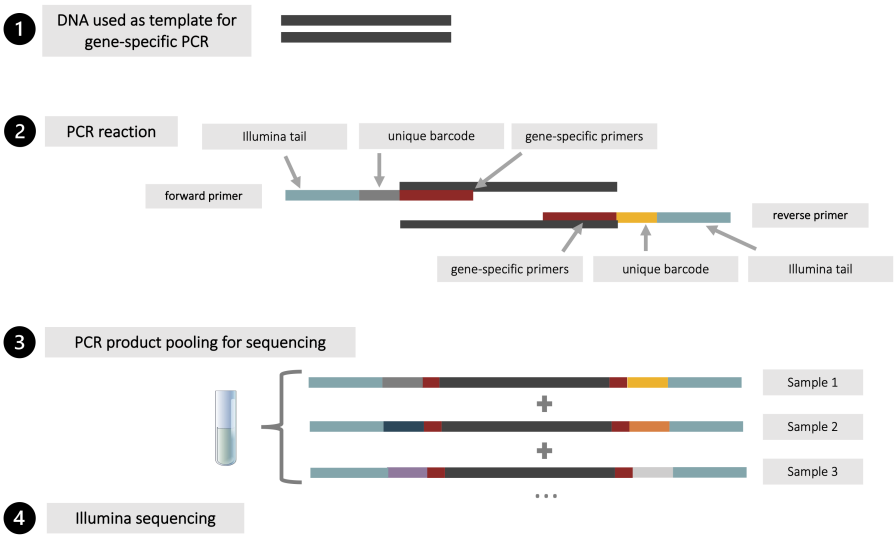
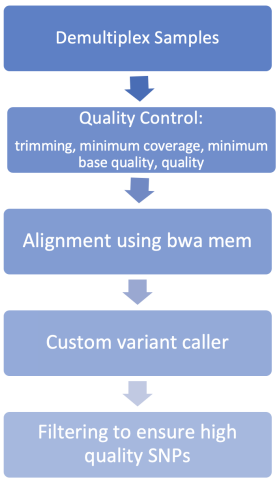


Figure Supplementary 1. *Steps for multiplex amplicon sequencing. In the first PCR, target genes are amplified and partial Illumina tails and 6bp barcodes included in primers to differentiate individual samples. In a second step the amplicons are pooled across samples. After, a second PCR is performed in each pool, the Illumina adapters and indexes are added, and pools are ready to be sequenced using an Illumina platform.*



Filter	Number of SNPs
Trimming 20 + Minimum coverage 50 + Minimum base quality 30 + Overall depth >10 + Overall quality >30	391
Depth >50	388
Allele Depth at least 10 for both alleles	388
SNP occurred in more than 1 amplicon batch	204
Remove indels	200
Remove SNPs within 25bp of barcode	160
Removal of homozygous reference SNPs	146

Figure Supplementary 2. *Flow chart of the bioinformatics pipeline.*

Supplementary Table 1. *Amino acid mutation positions for the reference organism and corresponding Ae. aegypti.*

Gene	Mutation Position Reference	Reference Species	Mutation Position *Ae. aegypti*
vgsc	V410L	*Musca domestica*	419
vgsc	G923V	*Musca domestica*	919
vgsc	L982W	*Musca domestica*	978
vgsc	S989P	*Musca domestica*	985
vgsc	I1011V/M	*Musca domestica*	1007
vgsc	V1016I/G	*Musca domestica*	1012
vgsc	T1520I	*Musca domestica*	1540
vgsc	F1534C/L	*Musca domestica*	1554
vgsc	D1763Y	*Musca domestica*	1783
rdl	A301S	*Drosophila melanogaster*	296
Ace-1	G119S	*Torpedo californica*	448

Supplementary Table 2. *DNA in-line barcodes and Illumina platform tails added to the 5' end of the forward and reverse primers before PCR amplification*

Forward Barcode Name	Illumina tail	Barcode
B1	ACACTCTTTCCCTACACGACFCTCTTCCGATCT	ATCACG
B2	ACACTCTTTCCCTACACGACFCTCTTCCGATCT	CGATGT
B3	ACACTCTTTCCCTACACGACFCTCTTCCGATCT	TTAGGC
B4	ACACTCTTTCCCTACACGACFCTCTTCCGATCT	TGACCA
B5	ACACTCTTTCCCTACACGACFCTCTTCCGATCT	ACATGT
B6	ACACTCTTTCCCTACACGACFCTCTTCCGATCT	TGCCAA
B7	ACACTCTTTCCCTACACGACFCTCTTCCGATCT	AGCTCG
B8	ACACTCTTTCCCTACACGACFCTCTTCCGATCT	ACGTCA
B9	ACACTCTTTCCCTACACGACFCTCTTCCGATCT	GCAGAT
B10	ACACTCTTTCCCTACACGACFCTCTTCCGATCT	GATCAC
Reverse Barcode Name	**Illumina tail**	**Barcode**
BR1	ACACTCTTTCCCTACACGACFCTCTTCCGATCT	CAGATC
BR2	ACACTCTTTCCCTACACGACFCTCTTCCGATCT	ACTTGA
BR3	ACACTCTTTCCCTACACGACFCTCTTCCGATCT	GATCAG
BR4	ACACTCTTTCCCTACACGACFCTCTTCCGATCT	TAGCTT
BR5	ACACTCTTTCCCTACACGACFCTCTTCCGATCT	GGCTAG
BR6	ACACTCTTTCCCTACACGACFCTCTTCCGATCT	CTTGTA
BR7	ACACTCTTTCCCTACACGACFCTCTTCCGATCT	TGAGAT
BR8	ACACTCTTTCCCTACACGACFCTCTTCCGATCT	ATGTGC
BR9	ACACTCTTTCCCTACACGACFCTCTTCCGATCT	GTATCA
BR10	ACACTCTTTCCCTACACGACFCTCTTCCGATCT	CTACTG

Supplementary Table 3. *Details of the SNPs detected in putative splice regions.*

Position	Reference	Alleles	Total samples
315938982	A	G	146
315938986	T	G	144
315939038	T	C	148
315939039	G	T	146
315939040	C	T	147
315939289	C	A	138
315939295	G	A	134
315939353	T	A	151
315939358	T	G	149
315939559	A	C	110
315939620	C	T	107
315939752	A	G	106
315939755	T	G	108

CHAPTER TWO

Profiling insecticide resistance phenotypes and genotypes in *Aedes aegypti* populations across four regions in Puerto Rico

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	1513859	**Title**	Ms
First Name(s)	Emma Louise		
Surname/Family Name	Collins		
Thesis Title	Utilising genomic approaches to explore genetic diversity and insecticide resistance in Aedes aegypti populations		
Primary Supervisor	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?*		Was the work subject to academic peer review?	

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

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	Scientific Reports
Please list the paper's authors in the intended authorship order:	Emma L. Collins, Joanelis Medina Quintana, Reynaldo Morales, Sophie Moss, Holly Acford-Palmer, Matthew Higgins, Jody Phelan, Taane G. Clark, Grayson Brown, 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)	EC designed study. EC collected samples and performed field experiments, with assistance from JMQ and RM. EC did genomic data generation. EC analysed the data, with assistance from SM, HAP, MH and JP. EC wrote the first draft of the manuscript. All authors have edited and approved the final version of the manuscript.

SECTION E

Student Signature	Emma Collins
Date	18/06/2024

Supervisor Signature	Susana Campino
Date	18/06/2024

Profiling insecticide resistance phenotypes and genotypes in *Aedes aegypti* populations across four regions in Puerto Rico

Authors: Emma L. Collins[1], Joanelis Medina Quintana [2], Reynaldo Morales[2], Sophie Moss[1], Holly Acford-Palmer[1], Matthew Higgins[1], Jody Phelan[1], Taane G. Clark[1,3], Grayson Brown[2*], Susana Campino[1*]

[1] Department of Infection Biology, London School of Hygiene & Tropical Medicine, London, UK

[2] Puerto Rico Vector Control Unit, Puerto Rico, USA

[3] Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, UK

[*]Contributed equally

Corresponding author:

Susana.campino@lshtm.ac.uk

Department of Infection Biology

Faculty of Infectious and Tropical Diseases

London School of Hygiene and Tropical Medicine

Keppel Street, London, UK

Abstract

Vector-borne diseases exert a considerable toll on global health. The efficacy of vector control strategies is being threatened by the emergence and spread of insecticide resistance worldwide. In this study, we investigated the insecticide resistance phenotypes and genotypes of *Aedes aegypti* populations in four regions of Puerto Rico. Insecticide resistance intensity CDC bioassays were employed to determine the response to deltamethrin and malathion. In parallel, next-generation targeted amplicon sequencing was used to investigate the presence of insecticide resistance-conferring mutations in nine targets across four genes: the voltage-gated sodium channel (*vgsc*); GABA (*rdl*); acetylcholinesterase (*ace-1*); and glutathione-S-transferase epsilon 2 (*GSTe2*). We observed high resistance levels to deltamethrin and malathion in *Ae. aegypti*, supported by molecular evidence revealing five mutations (V410L (*vgsc*), V1016I/G (*vgsc*), F1534C (*vgsc*), A296S (*rdl*)), previously linked to insecticide resistance. A previously undocumented mutation, L944I (L921I in *Ae. aegypti*, *vgsc*), was identified. While not yet reported in *Aedes* spp. vectors, this mutation has been associated with pyrethroid resistance in other medically important vectors and agricultural pests. Our research highlights the presence of insecticide resistance and associated mutations in Puerto Rico, which is valuable for vector control programs, providing information to guide decisions regarding the implementation of effective control interventions.

Word Count: 202

Keywords: Insecticide Resistance, Arbovirus Vector Control, Molecular surveillance, *Aedes aegypti*

Introduction

Vector-borne diseases (VBDs) cause vast morbidity and at least 700,000 deaths annually worldwide [1]. The majority of VBDs are transmitted by mosquitoes from three genera (*Anopheles, Culex,* and *Aedes)*. Following *Anopheles* mosquitoes, which are primary vectors for malaria parasites, *Aedes* mosquitoes, notably *Ae. aegypti* (L.), stand as a significant contributor to the global disease burden. *Ae. aegypti* is the dominant vector of many arboviruses including Zika virus, dengue virus, yellow fever virus, and Chikungunya virus. Millions of cases of arboviral diseases occur annually, nearly 400 million from dengue alone [2]. These diseases impose significant social and economic burden across the tropics, with the Americas being particularly affected with 3,126,573 cases reported in 2022 [3]. In Puerto Rico, arboviruses such as dengue and Zika have been responsible for substantial outbreaks [4,5]. Chikungunya and Zika were introduced to the island in 2014 and 2015, respectively [4], while dengue maintains endemic status, with an annual average of 5,000 to 7,000 cases [6]. Despite this baseline prevalence, dengue outbreaks occur regularly, notably in 2007, 2010, and 2013 where approximately 20,000 cases were reported in each year [5].

Insecticides have been used to effectively control vector populations and reduce the associated disease burden, notably in the case of insecticide treated nets to combat malaria. There are currently nine classes of adulticides used globally against mosquitoes, including pyrethroids, carbamates, organophosphates, organochlorines, neonicotinoids, pyrroles, butenolides, juvenile hormone mimics and spinosyns [7,8]. Unfortunately, the use of insecticides for both vector control and agriculture has led to the rise of insecticide resistance globally, threatening control programs. In Puerto Rico, vector control measures are applied inconsistently, targeting both adult mosquitoes and larvae with insecticides. Pyrethroid resistance in *Ae. aegypti* has already been documented on the island [9–11], and mutations linked to this resistance—such as V1016I and F1534C in the voltage-gated sodium channel (*vgsc*) gene—have been detected at high frequencies [11]. Additionally, evidence of metabolic resistance mechanisms has been found using synergist assays to isolate the action of detoxifying enzymes, in this case piperonyl butoxide, as well as the use of RNA sequencing (RNA-seq) to identify the upregulation of cytochrome P450 genes [9,12].

Assessing phenotypic resistance is essential to inform vector control programmes and support the implementation of the most effective methods. Conducting bioassays including well-established methods like the WHO tube tests[13], and WHO or CDC bottle bioassays [13,14], to evaluate mosquito mortality following insecticide exposure, can be a time-intensive process, judgement of knockdown

can be difficult, and comparisons across diverse studies or between WHO and CDC methodologies can be challenging [7]. Consequently, there is growing emphasis on monitoring molecular markers of resistance, as advocated by the WHO, highlighting the importance of understanding molecular mechanisms for designing effective vector control strategies [7]. These methodologies may act an early warning system to show emergence of resistance before control methodologies lose complete efficacy, although currently phenotypic testing is still most commonly used.

Molecular methodologies have become a cost-effective approach for the monitoring of insecticide resistance, particularly when using multiplex assays that target many loci in parallel. Targeted amplicon next generation sequencing (Amp-seq) offers the possibility to analyse large number of candidate genetic regions across many samples using next generation sequencing platforms [15–18]. It offers increased sensitivity compared to PCR-RFLP and real-time PCR, which in general only target a few markers, and decreased costs in comparison to whole genome sequencing (WGS). This approach will not only offer insights into the status of insecticide resistance but also aid in the identification of new resistance markers, however new markers should be confirmed with phenotypic testing and functional studies.

In this study we use bioassays to investigate the response of *Ae. aegypti* populations in Puerto Rico to deltamethrin, which targets the voltage gated sodium channels, and malathion, which inhibits acetylcholinesterase when activated to malaoxon [19]. We complement the study with molecular surveillance using a multi target Amp-seq approach, previously validated by us [20], to identify molecular marker associated with insecticide resistance in *Ae. aegypti* collected across five regions in Puerto Rico [16]. We targeted regions in *ace-1, vgsc, rdl* and *GSTe2* genes, with mutations associated with insecticide resistance, and use dual-indexing barcodes to allow for high-throughput processing and enable sequencing of multiple samples simultaneously to decrease costs.

Results

Mosquito collection and insecticide resistance assessment

Over 5,000 eggs were collected from Bayamon, Dorado, Guánica, Ponce and San Juan between 5th April 2022 and 1st June 2022 (**Figure 1**) and reared to adults. As insufficient eggs were collected from Guánica, bioassays were not performed for mosquitoes in this region. The mosquito populations in the other 4 regions were tested for susceptibility to deltamethrin and malathion using CDC bottle bioassays with a diagnostic dose and time of 0.75 µg/bottle, 30 minutes and 400 µg/bottle, and 15 minutes, respectively (**Table 1**) [21]. A total of 1,003 mosquitoes (765 field mosquitoes) were exposed to deltamethrin and a further 925 (686 field mosquitoes) were exposed to malathion, this includes control strain and unexposed field caught controls.

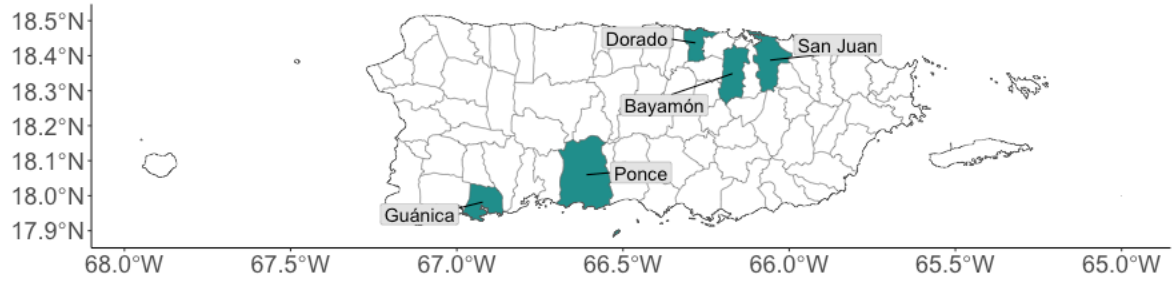


Figure 1. Sampling locations for ovitraps.

The implementation of insecticide bioassays to deltamethrin displayed high levels of resistance (**Figure 2**). The lowest level of mortality was observed in Dorado with only 2.2% mortality against the diagnostic dose, followed by San Juan and Bayamon with 21.4% and 22.7% mortality, respectively. Finally, the *Ae. aegypti* population in Ponce showed the highest mortality rate of 40.4%, however, this rate falls well below the 98% mortality threshold commonly considered indicative of susceptibility [22]. Mortality in all four regions increased when mosquitoes were subjected to five times the diagnostic dose of insecticide; however, these rates remain far below susceptible levels of mortality.

Exposure to malathion showed higher mortality levels than exposure to deltamethrin, however, resistance was still observed (**Figure 2**). Like the results for deltamethrin, the Ponce population showed the highest mortality against the diagnostic dose of malathion (87.2%), followed by San Juan (40.0%), Bayamon (36.4%), and Dorado (17.6%). Every location's mortality rate increased with the exposure to a higher dose of insecticide, however only the population from Ponce displayed

susceptibility after exposure to three times the diagnostic dose (1200 µg/ml). As the number of mosquitos varied between sites, a multivariate generalised log-linear model (Poisson regression) was fitted with (log) mortality count as the outcome, the (log) number of mosquitoes as an offset, with location, concentration and insecticide included as covariates. This approach indicated that as expected concentration ($p = 2x10^{-16}$) and insecticide ($p = 1.46x10^{-7}$) affect the rate of mortality, with malathion having a greater effect than deltamethrin (coefficient estimate = 0.430, deltamethrin as reference). It revealed that the mortality rate in Ponce was higher than the reference location Bayamon (coefficient estimate = 0.343, $p = 1.8x10^{-5}$), while comparisons for San Juan and Dorado showed no differences ($p > 0.07$).

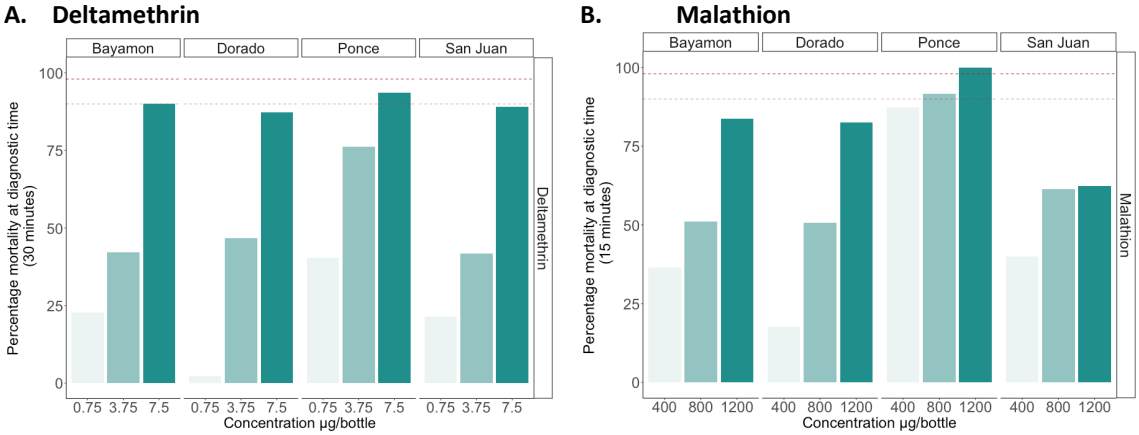


Figure 2. Mortality for the CDC bioassays at each concentration to (**A**) deltamethrin and (**B**) malathion for each location. Dotted red line shows 98% mortality above which indicates susceptibility and light red dotted line shows 90% mortality, between 90-98% indicates possible resistance.

Table 1. Percentage mortality and number of mosquitoes (n) included in each bioassay with differing times the diagnostic dose to deltamethrin and malathion in mosquitoes collected in 2022 across the 4 regions. Diagnostic dose for deltamethrin was 0.75 µg/ml and 400 µg/ml for malathion.

	Deltamethrin Percentage Mortality (n)		
Location	x1 (0.75 µg/bottle)	x5 (3.75 µg/bottle)	X10 (7.5 µg/bottle)
Bayamon (n=211)	22.7 (97)	42.2 (64)	90 (50)
Dorado (n=129)	2.2 (45)	46.7 (45)	87.2 (39)
Ponce (n=136)	40.4 (47)	76.2 (42)	93.6 (47)
San Juan (n=289)	21.4 (98)	41.8 (91)	89 (100)
Total	22.0 (287)	48.8 (242)	89.8 (236)
	Malathion Percentage Mortality (n)		
Location	x1 (400 µg/bottle)	x2 (800 µg/bottle)	x3 (1200 µg/bottle)
Bayamon (n=164)	36.4 (66)	51 (49)	83.7 (49)
Dorado (n=183)	17.6 (91)	50.7 (69)	82.6 (23)
Ponce (n=122)	87.2 (47)	91.7 (48)	100 (27)
San Juan (n=217)	40 (70)	61.4 (70)	62.3 (77)
Total	39.8 (274)	62.3 (236)	76.7 (176)

Amplicon Sequencing to identify molecular markers associated with resistance

A total of 178 samples were sequenced across 10 amplicons, covering *cytochrome oxidase* (COI) for speciation, and 4 genes (*vgsc, rdl, ace-1* and *GSTe2*) associated with insecticide resistance in *Ae. aegypti* [23]. The average sequencing coverage observed across the amplicons was 190.5-fold but varied across the ten amplicons (range: 6.5 to 550.8-fold). The lowest coverage was observed in the longest amplicon which targets *cytochrome oxidase* gene for mosquito speciation (LCO1490 and HCO1298) [24] (**S3 Table**). The average amplicon length was 453 bp (range: 321 to 709bp). All 178 samples passed quality control filters and included 51 from Bayamon, 42 from Dorado, 7 from Guánica, 33 from Ponce, and 45 from San Juan.

Amongst the 178 samples screened, 57 SNPs were identified, of which 14 were missense, one was in a splice region and 14 were synonymous, while the remainder were detected in intronic regions (**Table 2, S4 Table**). Previously identified insecticide resistance SNPs will be referenced according to the organism in which they were first reported. However, *Ae. aegypti* specific nomenclature is provided in **S5 Table** for ease of reference. The missense mutations (n=14) were identified across three genes (*ace-1, vgsc,* and *rdl),* the majority of these occurred in the *vgsc* gene (n=12), while there was a single missense in both the *ace-1* (n=1) and *rdl* (n=1) genes (**Table 2**). No missense mutations were identified in *GSTe2.* Most of these synonymous mutations (n=14) were found in the *vgsc* gene (n=9) while the remainder were in the *ace-1* gene (n=5) (**S4 Table**). The synonymous 506T mutation detected in the *ace-1* gene has previously been documented in resistant *Ae. aegypti* mosquitoes in the Philippines (not in combination with G119S) and Saudi Arabia (in combination with G119S); [25,26], however functional work has not confirmed its association with resistance.

Five missense mutations were detected which have previously been associated with insecticide resistance (*rdl* A296S; *vgsc* F1534C, V1016I, V1016G, V410L) (**Table 2**). The *rdl* A296S mutation was found in 52 samples, of which 24 had heterozygous genotypes. This mutation has been found in multiple insects including *Drosophila*, *Anopheles* and *Aedes* species, and is associated with resistance to dieldrin [16,27–31].

Table 2. Summary of the 14 missense mutations identified in the 178 samples screened. Nomenclature as per *Ae. aegypti. Rdl* mutations were based on the AAEL008354-RF transcript, *vgsc* mutations utilised the AAEL023266-RLtranscript and *ace-1* mutations refer to AAEL034366-RD transcript. This ^ symbol indicated previously described mutation associated with insecticide resistance. A * indicates a deletion.

Chrom	Gene	Position	Nucleic acid change	Annotation	Codon position in ref.	Genotype (n)			Alternative Allele Frequency
						Homo. ref	Hetero.	Homo. alt	
2	*rdl*	41847790	G > T	A296S^	301	7	24	28	67.7%
3	*ace-1*	161500150	C > T	A482T	482	105	1	0	0.47%
3	*vgsc*	315931756	A > C	I1845S	1854	115	1	0	0.43%
3	*vgsc*		G > A	Q1805*			6	0	3.61%
3	*vgsc*	315931943	G > C	Q1805E	1814	73	2	0	1.20%
3	*vgsc*		G > T	Q1805K			2	0	1.20%
3	*vgsc*	315932144	C > T	G1738S	1747	89	1	0	0.56%
3	*vgsc*	315932210	C > A	V1716L	1725	88	2	0	1.11%
3	*vgsc*	315939224	A > C	F1554C^	1534	0	18	109	92.91%
3	*vgsc*	315983762	A > C	V1012G^	1016	50	5	0	4.56%
3	*vgsc*	315983763	G > T	V1012I^	1016	0	5	50	95.45%
3	*vgsc*	315984130	A > C	F967C	979	50	8	0	6.90%
3	*vgsc*	315998386	A > T	F943Y	932	137	1	0	0.36%
3	*vgsc*	315998453	A > T	L921I^	910	95	30	2	22.08%
3	*vgsc*	315998530	A > C	L895R	1008	118	6	0	2.42%
3	*vgsc*	316080722	C > A	V408L^	410	0	15	101	93.53%

The five *vgsc* mutations observed are associated with resistance to the pyrethroid insecticide class. The *vgsc* F1534C mutation was present in 127 samples of which 18 had heterozygous genotypes (n = 127). The V1016I *vgsc* mutation presented as a homozygous alternative for the mutation in 50 samples while 5 were heterozygous (n=55). The V1016G mutation was less frequent, with only 5 samples being identified as a homozygous alternative (n=55). One hundred and one samples were homozygous alternative for the *vgsc* V410L, as well as 15 with heterozygous genotypes (n=116).

The *vgsc* L944I (L921I *Ae. aegypti* AAEL023266-RL transcript numbering) mutation was identified in our samples. This mutation is of particular interest because, although it has not been previously reported in *Ae. aegypti*, the equivalent amino acid change (L925I) has been documented in several other arthropod species, including *Triatoma infestans* and *Bemisia tabaci* [32,33]. The mutation allele frequency was 22.1% with 30 heterozygotes identified and 2 homozygous alternatives (n=127).

All six missense SNPs associated with insecticide resistance were detected in the 5 locations (**Figure 3**), apart from the V1016G mutation, which was absent from Guánica and San Juan. The *vgsc* V410L mutation appears fixed in both San Juan and Guánica and is approaching fixation in Bayamon and Ponce (**Table 3**). Dorado had the lowest proportion of homozygous alternative mutations for *vgsc* V410L, though the allele was still observed at a frequency of 84.6%. The F1534C and V1016I

had the next highest overall allele frequencies of 92.9% and 95.5%, respectively. Overall, there were minimal differences in allele frequencies between the insecticide resistance SNPs and the sampling locations (χ^2 test, $p > 0.05$), suggesting that the observed phenotypic differences may not be primarily driven by the detected SNPs. Spearman's-rank analysis between mortality rate and allele frequency for each insecticide resistance SNP identified by location indicated no correlation ($p > 0.30$).

Table 3. Allele frequency (%) of each of the six detected insecticide resistance associated mutations in each of the locations.

Gene	Mutation (position)	Nucleotide position	Overall	Bayamon	Dorado	Guánica	Ponce	San Juan
rdl	A301S	41847790	66.8	92.3	40.6	83.3	55.9	100
vgsc	F1534C	315939224	92.9	95.6	82.8	100	90.9	100
vgsc	V1016G	315983762	4.5	1.0	2.4	0.0	3.0	0.0
vgsc	V1016I	315983763	95.5	95.0	88.9	100	90.0	100
vgsc	L944I	315998453	13.4	13.9	8.1	14.3	14.0	17.9
vgsc	V408L	316080722	93.5	95.5	84.6	100	91.3	100

A.

	A301S (41847790)	F1534C (315939224)	V1016G (315983762)	V1016I (315983763)	L944I (315998453)	V410L (316080722)
0/0	G/G	A/A	T/T	G/G	A/A	C/C
0/1	G/T	A/C	T/G	G/A	A/T	C/A
1/1	T/T	C/C	G/G	A/A	T/T	A/A

B.

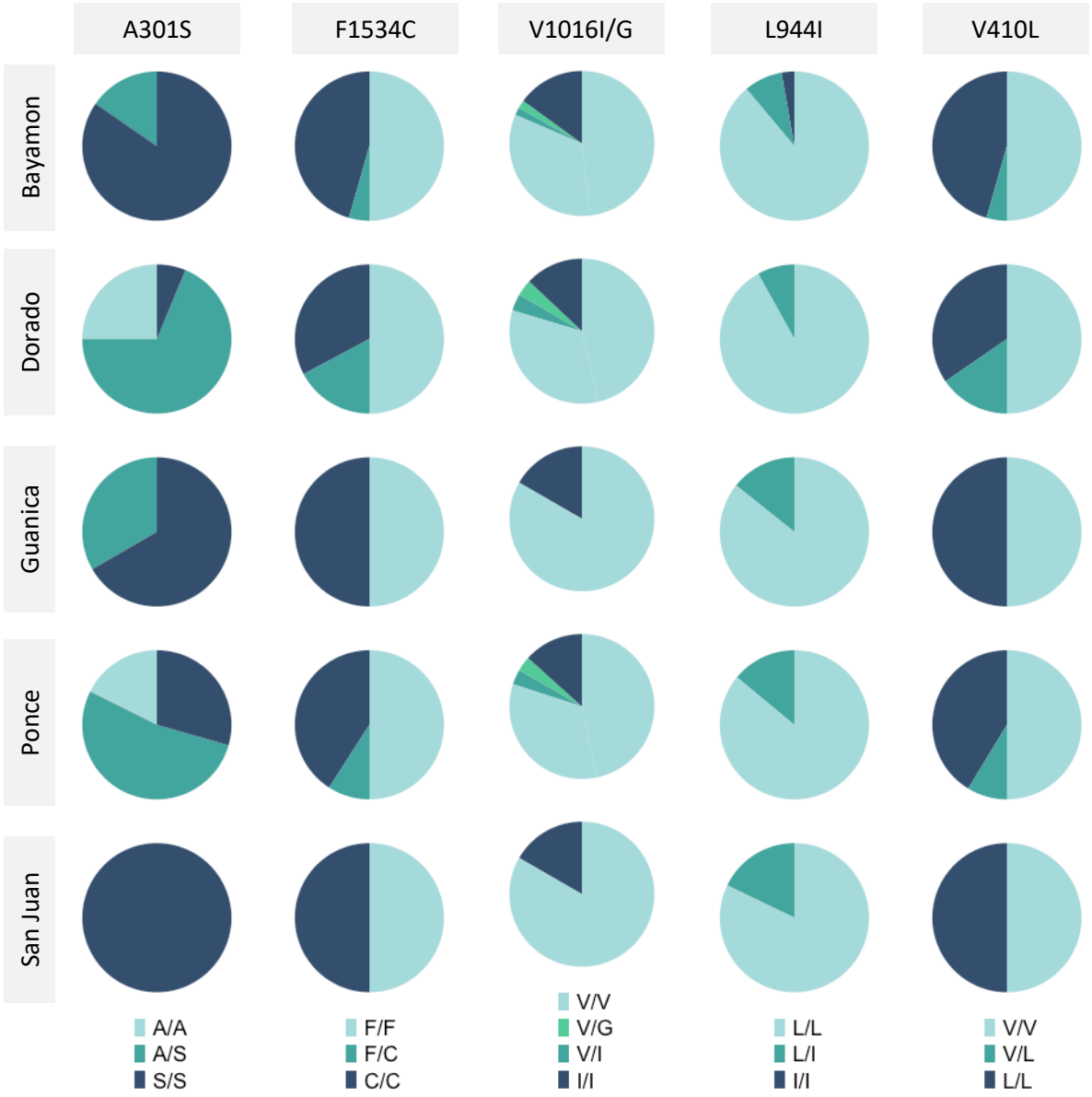


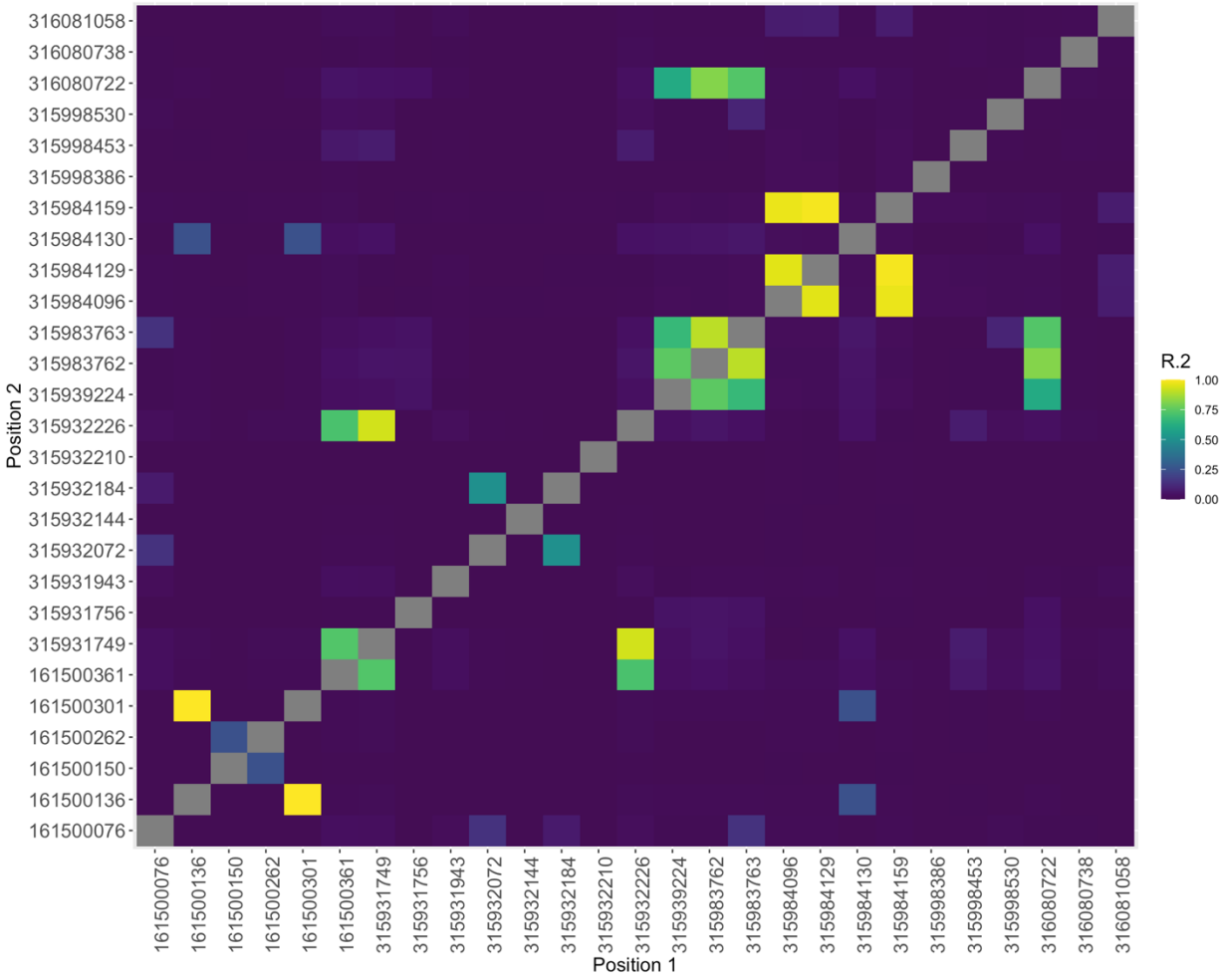
Figure 3. A. Genotype and corresponding alleles for each of the insecticide resistance mutations **B.** The allele proportions for each of the give main insecticide resistance associated mutations overall across all locations. For the V1016 codon a combination of the SNPs at positions 315983762 and 315983763 were used to identify the consequence.

Linked mutations

Linkage disequilibrium analysis was only carried out on chromosome 3 due to few SNPs being identified on chromosome 2. Overall, 236 pairwise SNP combinations of the 3,136 possible, have an R^2 of more than 0.8 [34], which includes 33 unique SNP locations (**S6 Table**). **Figure 4** shows the R^2 values

for synonymous and missense SNPs. Most of these SNPs were in the *vgsc* gene (n = 34), while two were identified in *ace-1.* Strong linkage was identified between four missense mutations located in *vgsc* (V410L, V1016I, V1016G, and F1534C) (R^2 range: 0.61 – 0.90).

Figure 4. *Linkage R^2 values between synonymous and non-synonymous positions*



Discussion

This study provides strong evidence of both phenotypic and genotypic insecticide resistance in Puerto Rico, using a combination of bioassays and a targeted Amp-seq assay. Bioassays were conducted on *Ae. aegypti* samples from four regions, and Amp-seq assays on samples from five regions across Puerto Rico. We observed elevated levels of phenotypic resistance to both deltamethrin and malathion, along with the detection of five genetic markers linked to resistance against organochlorines (cyclodienes and phenylpyrazoles) and pyrethroids.

A 2016 study in Puerto Rico identified resistance in *Aedes aegypti* populations to permethrin, phenothrin, etofenprox, and tetramethrin through phenotypic bioassays [10]. Hemme *et al.* (2019) reported widespread resistance, with deltamethrin achieving effective control in only five out of 38 regions, and no populations showing susceptibility to malathion. Naled was the most effective insecticide, killing 100% of mosquitoes in all locations tested. This study updates those findings and further examines the intensity of resistance to deltamethrin and malathion. High-intensity resistance was observed at up to three times and ten times the diagnostic doses for malathion (1200 µg/bottle) and deltamethrin (7.5 µg/bottle), respectively.

Pyrethroid-associated mutations have also been described previously (V1016I and F1534C) at high frequency in Puerto Rico (60%-100% and 80%-100% respectively) [11]. This study supports previous findings by identifying the V1016I and F1534C SNPs, with allele frequencies ranging from 88.9% to 100% and 82.8% to 100%, respectively, similar to those reported by Ponce-García et al. (2016). The mutation at position 315983762, leading to V1016G, and the mutation at 315983763, resulting in V1016I, have been observed globally, including in Thailand, Indonesia, Ghana, and the USA[35–38]. However, V1016I is primarily found in the Americas, while V1016G is more prevalent in Asia. We identified a combination of these mutations in the V1016 codon (positions 315983762 and 315983763); however, each position is mutually exclusively heterozygous, and does not occur together in the same chromosome; this mutation combination was found at low frequency (4.5%, n = 5) with the resulting heterozygous alleles V1016V/V1016G. The multiple mutations observed at this amino acid position suggest it is under selective pressure [39]. The V1016G mutation, previously described only in Asian *Ae. aegypti* populations, is noteworthy due to its recent detection in the Americas [40,41]. This finding may indicate a breakdown in the geographic separation of these SNPs. This theory is further supported by the recent identification of the V1016G mutation in Benin, West Africa, and Panama[42,43].

This study identified additional insecticide resistance mutations in the *vgsc* gene within this population that had not previously been reported in Puerto Rico, including the V410L mutation and a novel L944I mutation. The V410L mutation was nearing fixation in nearly all populations tested, with an overall allele frequency of 0.908. In contrast, the L944I mutation appears to have recently emerged, as its allele frequencies remain low, ranging from 8.1% in Dorado to 17.9% in San Juan. This is the first documentation of the L944I mutation in *Ae. aegypti* (L921I in *Ae. aegypti* numbering), although it has previously been linked to pyrethroid resistance in other insect vectors and pests [32,33]. The role of the

L944I mutation in pyrethroid resistance has been confirmed in *Drosophila melanogaster* by expressing the mutation in Xenopus oocytes. The conserved nature of the voltage-gated sodium channel and the mutation's location in the critical region for pyrethroid binding in the *vgsc* domain II S4-S5 suggest that it may confer similar resistance phenotypes in *Aedes aegypti* (see alignment **SI Figure 2**) [44]. The V410L and L944I mutations, not described in previous studies, highlight the advantages of the broader amplicon approach over traditional PCR methods. Additionally, we provide the first report of the *rdl* A296S mutation in the Puerto Rican *Ae. aegypti* population. This mutation, which confers resistance to organochlorines (cyclodienes and phenylpyrazoles), was found at an allele frequency ranging from 49.6% to 100%.

Linkage disequilibrium was observed among four missense mutations (V410L, V1016I, V1016G, and F1534C), with R^2 correlations ranging from 0.61 to 0.90. These mutations have demonstrated linkage in multiple *Ae. aegypti* populations. [45–47]. Given the lack of regular or intensive vector control programs in Puerto Rico, the intensity of phenotypic resistance and the presence of genotypic markers are somewhat surprising. However, several previous studies have identified both phenotypic and genotypic resistance on the island. Additionally, *ad hoc* ultra-low volume (ULV) spraying of permethrin is conducted based on population demands or in response to Zika and dengue outbreaks. The recent outbreaks may explain the observed resistance and mutations, as increased implementation of control measures has likely exerted selective pressure on the mosquito population [48,49]. Alongside occasional spraying, the use of insecticides and pesticides in households and agriculture may have further contributed to the resistance profile in *Ae. aegypti* on the island. The combination of various active ingredients in these products, along with the anthropophilic nature of *Ae. aegypti*, likely leads to high exposure levels, promoting the development of the observed broad resistance profiles within the population.

The A296S mutation in the *rdl* gene, which encodes a GABA receptor chloride channel, was also detected in this population. This mutation is well-documented for its association with resistance to dieldrin, which was banned in 1970 due to concerns about its environmental impact and potential carcinogenic properties [50]. Despite this ban, the ongoing use of alternative insecticides or pesticides targeting the GABA chloride channel—such as cyclodiene organochlorines, phenylpyrazoles and pyrethroids (GABA is a secondary target) —may contribute to the persistence of this mutation [51,52]. The A296S mutation has been found in various mosquito populations, including *Aedes aegypti* from Burkina Faso, Cape Verde, and Cameroon [16,31,53], suggesting that it may not impose a significant fitness cost on these populations. However, conflicting evidence exists regarding this hypothesis [54,55].

Our study revealed variations in resistance across different island locations when analysing phenotypic bioassay data. However, no differences were observed in the allele frequencies of insecticide resistance SNPs among these locations. This may suggest the limitations of bioassays in quantifying the intensity of insecticide resistance, or it could indicate a need for larger sample sizes to achieve more accurate quantification. We also highlight that, particularly in highly resistant populations, individual level phenotype and genotype data is preferential for understanding phenotype-genotype associations and correlations. Alternatively, it may imply that other mechanisms contribute to the observed differences in insecticide resistance. This may be more likely given that resistance to malathion was observed however no mutations known to confer resistance to malathion were observed (e.g. G119S), indicating that other mechanisms are mediating this phenotype. Potential mechanisms include metabolic resistance or cuticular modifications, although these aspects were beyond the scope of our current study. Metabolic resistance has previously been documented in Puerto Rico using piperonyl butoxide to isolate the effects of detoxifying enzymes [9] and via RNA-seq [12], which identified the overexpression of cytochrome P450 genes. Future research involving synergistic bioassay testing would be valuable for further investigating metabolic resistance in the identified locations. Additionally, future genotypic studies should focus on genes associated with metabolic resistance, such as P450 monooxygenases, esterases, and glutathione S-transferases [23]. It is also crucial to examine copy number variants and genes linked to cuticular thickening, another recognized resistance mechanism, as highlighted by Faucon *et al.* (2017) [56].

The WHO recognises the importance of molecular markers in understanding the evolving landscape of insecticide resistance mechanisms among medically significant vectors [7]. In our study, we employed a cost-effective and easily implementable assay that can screen numerous samples. By leveraging PCR multiplexing and dual barcoding, this approach enhances scalability and affordability, allowing for the pooling of amplicons across samples and facilitating discrimination during analysis. Additionally, these amplicons can be sequenced on various platforms, including portable sequencers like the Oxford Nanopore Technology MinION, making the method more applicable and accessible in low-resource settings. However, we acknowledge that sequencing technologies are not yet feasible for many vector control programs. Nevertheless, the Ebola and Zika outbreaks, along with the COVID-19 pandemic, have demonstrated the value of sequencing data for monitoring disease transmission. As a result, there has been increased investment in sequencing capabilities, leading to enhanced capacity in many countries.

Our work underscores the value of this methodology as a tool for identifying potential markers by revealing the presence of the novel L944I (L921I in *Aedes aegypti*) mutation. This mutation is associated with pyrethroid resistance in various species, including *Triatoma infestans* (the vector of *Trypanosoma cruzi*, the causative agent of Chagas disease in the Americas) [32] and *Bemisia tabaci* (the silverleaf whitefly, a globally significant agricultural pest) [33], highlighting the potential cross-species relevance of the identified genetic variation. While further studies are necessary to confirm the functional role of this mutation in *Aedes aegypti*, its homology across species emphasizes its potential utility in enhancing our understanding of insecticide resistance in both medically and agriculturally important arthropods [57].

The complex landscape of phenotypic insecticide resistance in mosquitoes encompasses numerous contributing mechanisms and interactions. A deeper understanding of these mechanisms, particularly those related to metabolic resistance, could expand the genomic targets within the Amp-seq panel proposed in our study. This adaptable methodology can be utilised for surveillance of mosquito populations in conjunction with phenotypic testing, providing valuable insights to inform vector control programs that are essential for reducing disease burden.

Methods

Sampling Sites and Bioassays

Mosquito eggs were collected from gravid ovitraps placed in six locations in Puerto Rico: Bayamón, Culebra, Dorado, Guánica, Ponce, and San Juan (**Figure 1**). Between April and May 2022, black cups containing seed germination paper were used as ovitraps, pre-prepared with hay infusion and deployed for one week at a time. The traps were placed within 50 meters of residences. In compliance with the US Health Insurance Portability and Accountability Act of 1996 (HIPAA), we cannot disclose the exact locations of the traps. A total of 14 traps were deployed in Bayamón, 28 in Guánica, 27 in Ponce, and 36 in San Juan. For logistical reasons, only San Juan and Bayamón had multiple trap deployments, with three traps in San Juan and two in Bayamón. Between collections, the traps were washed. After one week of deployment, the traps were collected, and the oviposition papers were dried.

Eggs were reared in the Puerto Rico Vector Control Unit insectary according to standard laboratory protocols until they developed into adults. Adult mosquitoes, aged 3 to 5 days, were then tested for insecticide resistance using the CDC bottle bioassay with 250 mL Wheaton glass bottles. Technical

deltamethrin was tested at 1× (0.75 µg/bottle), 5×, and 10× doses, while technical malathion was tested at 1× (400 µg/bottle), 2×, and 3× doses, following CONUS CDC recommendations [21]. An acetone control was included for both insecticides, and the ROCKEFELLER MR734 strain was used as a reference. The following reagent was obtained from BEI Resources, NIAID, NIH: *Ae. aegypti*, Strain ROCK, MRA-734, contributed by David W. Severson. After treatment, the uncapped bottles were placed on bottle rollers until completely dry.

Mosquitoes were morphologically identified as *Ae. aegypti* prior to testing by a trained entomologist using a key [58]. To test resistance, 18 to 25 F0 mosquitoes were placed in each bottle (3 treated with insecticide, 1 treated with acetone). Each experiment had a bottle for the ROCK susceptible control strain at diagnostic dose and a bottle of field caught mosquitoes exposed to acetone control. Knockdown of the mosquitoes was recorded every 15 minutes for 2 hours as per CDC recommendations [21]. Mosquitoes were recorded as knocked down if they could no longer stand or fly. An assessment of resistance was made using the percentage knocked down at diagnostic time (30 minutes for deltamethrin and 15 minutes for malathion) for the diagnostic dose (0.75 µg/bottle for deltamethrin and 400 µg/bottle for malathion) as per CDC protocol [14,21]. Each bottle was used a maximum of two times before washing and recoating, control mortality showed this was acceptable for efficacy. Testing at each concentration and insecticide was repeated for as many times as number of mosquitoes collection allowed. Mosquitoes were preserved after phenotype testing in RNAlater® and frozen at -80˚C.

Molecular testing

Sample DNA was extracted using the Qiagen DNAeasy blood and tissue extraction kit following the manufacturer's instructions. DNA concentration was tested using the Qubit 2.0 fluorophotometer. Subsequently, five-plex multiplex PCRs were carried out using Q5 High-fidelity PCR kits (New England Biolabs, UK), under the following conditions: initial denaturation (98.0 °C, 30 seconds) followed by 35 cycles of denaturation (98.0 °C, 10 seconds), annealing (57.3 °C, 35 seconds), and extension (72.0°C, 45 seconds). 1 µL of DNA, 0.5 µL of each forward and reverse primer at 10 pmol/µL were combined with 19 µL of master to mix to make up to a 25 µL reaction (**S1 Table; S2 Table**).

For Amp-seq, ten amplicons were designed (9 for insecticide resistance (4 genes), 1 for species identification). Amplicon primers were adapted from Collins et al., (2022), and changes were made to improve efficiency in multiplex combinations and target the *GSTe2* gene. Species primers were taken from Folmer *et al.* (1994) [24], which target the cytochrome c oxidase subunit 1 (*cox-1*) gene of

the mitochondria. Primer regions targeted single nucleotide polymorphisms (SNPs) within regions of ~500 base pairs (amplicon size). Index barcodes were eight base pairs in length (**S1 Table; S2 Table**). A total of 17 SNPs across 4 genes; *vgsc, rdl, ace-1* and *GSTe2* were targeted with this panel (**S3 Table**). PCR assays were carried out in the combinations outlined (**S2 Table**), and the primers had 3' barcodes attached to allow discrimination of individual samples; the barcodes used are outlined (**S1 Table**). PCR products were visualised on 1% agarose gel with SYBR safe (Cambridge Biosciences, UK) alongside a 100bp ladder. The products were purified with AMPure XP magnetic beads (Beckman Coulter), using a ratio of 0.8:1 (µL of beads to DNA). PCR assays were normalised to equal concentrations to create an overall pool of 20 ng/µL in 25 µL total volume (maximum of 10 amplicons across 50 barcoded mosquitoes = 500 amplicons). Sequencing was performed by Genewiz (Azenta Life Sciences), at a cost of ~£60 per pool or >US $0.15 per amplicon.

Data and Statistical Analysis

Mosquito mortality levels were interpreted as per WHO/CDC criteria (98-100% - susceptibility, 90-97% - possible resistance, <90% - resistance). Mortality is taken at the diagnostic time for the insecticide as per WHO and CDC guidelines [22,59]. The mortality rate per site, insecticide and its concentration were modelled using a log-linear (Poisson) regression model with the (log) number of deaths as the outcome and an offset reflecting the (log) group sample size. Likelihood ratio tests were applied to determine the statistical significance of the covariates site, insecticide and concentration. Differences in allele frequencies across populations were estimated using Chi-squared tests and correlations between mortality and allele frequency assessed with Spearman's rank test. All statistical analyses were performed using R (v4.3) software, with a significance level of 0.05.

Bioinformatic Analysis

A minimum of 50,000 reads were obtained per sample pool, and raw pooled FASTQ sequences were demultiplexed based on the 8bp barcode primer-tag in each forward and reverse primer using an inhouse pipeline (available at https://github.com/LSHTMPathogenSeqLab/amplicon-seq). This pipeline removes and mis-tagging caused by errors in sequencing. Sequences were trimmed using *trimmomatic* software (v0.39) using the parameters LEADING:3, TRAILING:3, SLIDING WINDOW:4:20, MINLEN:36 to remove low quality ends of sequences [60]. Sequences were aligned to the *Ae. aegypti* reference (Vectorbase Aag2, GCA_021653915.1) using *bwa-mem* software and default parameters (v0.7.17-r1188) [61]. A small region of the intron in *vgsc* domain II amplicon did not map to the reference due to divergence in the sequences from Puerto Rico, therefore mapping was done to the *vgsc* domain II sequence for this amplicon (MK977835.1) **(S3 Figure).** Following mapping assessments of quality

and mapping using *FastQC* (v0.12.1) and *samtools flagstat* (v1.17) [62]. Mapped reads were clipped using Samclip package. Variants were called using both *GATK* haplotype caller (v4.4.0.0, default parameters) [63]and *freebayes* software (v1.3.6, default parameters) [64], and filtered by *bcftools* to maximise confidence in the called SNPs [65,66]. Further filtering was carried out to ensure there was coverage across a minimum of 6 of the 10 amplicons.

Linkage disequilibrium was calculated using *vcftools*[67] on phased vcf files created with *Beagle* (v 22Jul22.46e) [68,69] software to provide a R^2 value for pairwise combinations of non-synonymous mutations by sample country. Filtering was carried out based on the distance between mutations (minimum 20, maximum 10 Kbp). The related plots were generated using the *gaston* (v1.5.9) package in R.

Alignment of sequences to demonstrate the conservation of the L944I mutation across various species was performed in Aliview (v1.28) [70] using the default MUSCLE alignment settings [71]. The L944I mutation numbering is based on *Musca domestica* accession NW_026712250.

References

1. World Health Organization. Vector-borne diseases. https://www.who.int/news-room/fact-sheets/detail/vector-borne-diseases (2024).
2. Byaruhanga, T. *et al.* Arbovirus circulation, epidemiology and spatiotemporal distribution in Uganda. *IJID Regions* **6**, 171 (2023).
3. PAHO/WHO Data - Annual Arbovirus Bulletin 2022. https://www3.paho.org/data/index.php/en/mnu-topics/indicadores-dengue-en/annual-arbovirus-bulletin-2022.html.
4. PAHO. Health in the Americas - Puerto Rico. https://www.paho.org/salud-en-las-americas-2017/?p=4295.
5. Ministerio de Salud Puetro Rico. *Informe de La Salud En Puerto Rico 2015*. http://www.salud.gov.pr/Estadisticas-Registros-y-Publicaciones/Publicaciones/Informe de la Salud en Puerto Rico 2015_FINAL.pdf (2015).
6. Centres for Disease Control and Prevention. Dengue Historic Data (2010-2023). *https://www.cdc.gov/dengue/statistics-maps/historic-data.html* (2024).
7. World Health Organization. Manual for monitoring insecticide resistance in mosquito vectors and selecting appropriate interventions.
8. World Health Organisation. Vector Control Product List | WHO - Prequalification of Medical Products (IVDs, Medicines, Vaccines and Immunization Devices, Vector Control). https://extranet.who.int/prequal/vector-control-products/prequalified-product-list.
9. Estep, A. S., Sanscrainte, N. D., Waits, C. M., Louton, J. E. & Becnel, J. J. Resistance Status and Resistance Mechanisms in a Strain of Aedes aegypti (Diptera: Culicidae) From Puerto Rico. *J Med Entomol* **54**, 1643–1648 (2017).
10. Hemme, R. R. *et al.* Rapid Screening of Aedes aegypti Mosquitoes for Susceptibility to Insecticides as Part of Zika Emergency Response, Puerto Rico - Volume 25, Number 10—October 2019 - Emerging Infectious Diseases journal - CDC. *Emerg Infect Dis* **25**, 1959–1961 (2019).
11. Ponce-García, G. *et al.* Knockdown resistance mutations in aedes aegypti (Diptera: Culicidae) From puerto rico. *J Med Entomol* **53**, 1410–1414 (2016).
12. Derilus, D. *et al.* Comparative Transcriptomic Analysis of Insecticide-Resistant Aedes aegypti from Puerto Rico Reveals Insecticide-Specific Patterns of Gene Expression. *Genes (Basel)* **14**, 1626 (2023).
13. WHO. Test procedures for insecticide resistance monitoring in malaria vector mosquitoes (Second edition) (Updated June 2018). *Who* 48 (2018).
14. Brogden, W. G. & Chan, A. Guideline for evaluating insecticide resistance in vectors using the CDC bottle bioassay. Preprint at https://stacks.cdc.gov/view/cdc/21777 (2012).
15. 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).
16. Collins, E. L. *et al.* A next generation targeted amplicon sequencing method to screen for insecticide resistance mutations in Aedes aegypti populations reveals a rdl mutation in mosquitoes from Cabo Verde. *PLoS Negl Trop Dis* **16**, e0010935 (2022).
17. Acford-Palmer, H. *et al.* Identification of two insecticide resistance markers in Ethiopian Anopheles stephensi mosquitoes using a multiplex amplicon sequencing assay. *Scientific Reports 2023 13:1* **13**, 1–10 (2023).
18. 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).
19. Malathion Technical Fact Sheet. http://npic.orst.edu/factsheets/archive/malatech.html.
20. Campos, M. *et al.* Surveillance of Aedes aegypti populations in the city of Praia, Cape Verde: Zika virus infection, insecticide resistance and genetic diversity. *Parasit Vectors* **13**, 1–11 (2020).
21. Centers for Disease Control and Prevention. CONUS Manual for Evaluating Insecticide Resistance in Mosquitoes Using the CDC Bottle Bioassay Kit.
22. CDC. *Guideline for Evaluating Insecticide Resistance in Vectors Using the CDC Bottle Bioassay*. https://www.cdc.gov/malaria/resources/pdf/fsp/ir_manual/ir_cdc_bioassay_en.pdf.
23. Moyes, C. L. *et al.* Contemporary status of insecticide resistance in the major Aedes vectors of arboviruses infecting humans. *PLoS Negl Trop Dis* **11**, (2017).

24. 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**, 294–9 (1994).
25. Fang, Y. *et al.* Molecular Analysis of Targeted Insecticide Resistance Gene Mutations in Field-Caught Mosquitos of Medical Importance From Saudi Arabia. *J Med Entomol* **58**, 1839–1848 (2021).
26. HASMIWATI, H., RUSJDI, S. R. & NOFITA, E. Detection of Ace-1 gene with insecticides resistance in Aedes aegypti populations from DHF-endemic areas in Padang, Indonesia. *Biodiversitas* **19**, 31–36 (2018).
27. 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).
28. 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**, 449–51 (1993).
29. Tantely, M. L. *et al.* Insecticide resistance in Culex pipiens quinquefasciatus and Aedes albopictus mosquitoes from La Réunion Island. *Insect Biochem Mol Biol* **40**, 317–324 (2010).
30. 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).
31. 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).
32. Capriotti, N., Mougabure-Cueto, G., Rivera-Pomar, R. & Ons, S. L925I Mutation in the Para-Type Sodium Channel Is Associated with Pyrethroid Resistance in Triatoma infestans from the Gran Chaco Region. *PLoS Negl Trop Dis* **8**, e2659 (2014).
33. Wei, Y., Guan, F., Wang, R., Qu, C. & Luo, C. Amplicon sequencing detects mutations associated with pyrethroid resistance in Bemisia tabaci (Hemiptera: Aleyrodidae). *Pest Manag Sci* **77**, 2914–2923 (2021).
34. Linkage disequilibrium - PLINK 1.9. https://www.cog-genomics.org/plink/1.9/ld.
35. Hernandez, J. R., Longnecker, M., Fredregill, C. L., Debboun, M. & Pietrantonio, P. V. Kdr genotyping (V1016I, F1534C) of the Nav channel of Aedes aegypti (L.) mosquito populations in Harris County (Houston), Texas, USA, after Permanone 31–66 field tests and its influence on probability of survival. *PLoS Negl Trop Dis* **15**, e0009833 (2021).
36. Ayettey, J., Ablorde, A., Amlalo, G. K., Mensah, B. A. & Kudom, A. A. Entomological surveillance on Aedes aegypti during covid 19 period in Cape Coast, Ghana: Risk of arboviral outbreaks, multiple insecticide resistance and distribution of F1534C, V410L and V1016I kdr mutations. *PLoS Negl Trop Dis* **17**, e0011397 (2023).
37. Stenhouse, S. A. *et al.* Detection of the V1016G mutation in the voltage-gated sodium channel gene of Aedes aegypti (Diptera: Culicidae) by allele-specific PCR assay, and its distribution and effect on deltamethrin resistance in Thailand. *Parasit Vectors* **6**, 253 (2013).
38. Hamid, P. H., Ninditya, V. I., Ghiffari, A., Taubert, A. & Hermosilla, C. The V1016G mutation of the voltage-gated sodium channel (VGSC) gene contributes to the insecticide resistance of Aedes aegypti from Makassar, Indonesia. *Parasitol Res* **119**, 2075–2083 (2020).
39. Kushwah, R. B. S. *et al.* A new knockdown resistance (kdr) mutation, F1534L, in the voltage-gated sodium channel of Aedes aegypti, co-occurring with F1534C, S989P and V1016G. *Parasit Vectors* **13**, (2020).
40. Kasai, S. *et al.* Discovery of super–insecticide-resistant dengue mosquitoes in Asia: Threats of concomitant knockdown resistance mutations. *Sci Adv* **8**, eabq7345 (2022).
41. Uemura, N., Itokawa, K., Komagata, O. & Kasai, S. Recent advances in the study of knockdown resistance mutations in Aedes mosquitoes with a focus on several remarkable mutations. *Curr Opin Insect Sci* **63**, 101178 (2024).
42. Murcia, O. *et al.* Presence of the point mutations Val1016Gly in the voltage-gated sodium channel detected in a single mosquito from Panama. *Parasit Vectors* **12**, 1–7 (2019).
43. Tokponnon, T. F. *et al.* Insecticide Resistance in Aedes aegypti Mosquitoes: Possible Detection of kdr F1534C, S989P, and V1016G Triple Mutation in Benin, West Africa. *Insects 2024, Vol. 15, Page 295* **15**, 295 (2024).
44. Silver, K. S. *et al.* Voltage-Gated Sodium Channels as Insecticide Targets. *Adv In Insect Phys* **46**, 389 (2014).

45. Fan, Y. *et al.* Evidence for both sequential mutations and recombination in the evolution of kdr alleles in Aedes aegypti. *PLoS Negl Trop Dis* **14**, e0008154 (2020).
46. 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**, (2017).
47. Saavedra-Rodriguez, K. *et al.* Local evolution of pyrethroid resistance offsets gene flow among Aedes aegypti collections in Yucatan State, Mexico. *Am J Trop Med Hyg* **92**, 201–209 (2015).
48. Historic Data (2010-2023) | Dengue | CDC. https://www.cdc.gov/dengue/data-research/facts-stats/historic-data.html.
49. Centers for Disease Control and Prevention. Zika Cases in the United States | Zika Virus | CDC. https://www.cdc.gov/zika/zika-cases-us/index.html.
50. Department of Health, U., Services, H., for Toxic Substances, A. & Registry, D. Toxicological Profile for Aldrin and Dieldrin. (2022).
51. Ramadan, A. A., Bakry, N. M., Marei, A. S. M., Eldefrawi, A. T. & Eldefrawi, M. E. Action of pyrethroids on GABAA receptor function. *Pestic Biochem Physiol* **32**, 97–105 (1988).
52. Fankhauser, B. *et al.* Repellent and insecticidal efficacy of a new combination of fipronil and permethrin against Three mosquito species (Aedes albopictus, Aedes aegypti and Culex pipiens) on dogs. *Parasit Vectors* **8**, 1–8 (2015).
53. Dabire, K. R. *et al.* Distribution of the members of Anopheles gambiae and pyrethroid knock-down resistance gene (kdr) in Guinea-Bissau, West Africa. *Bull Soc Pathol Exot* **101**, 119–123 (2008).
54. 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 (2020).
55. Platt, N. *et al.* Target-site resistance mutations (kdr and RDL), but not metabolic resistance, negatively impact male mating competiveness in the malaria vector Anopheles gambiae. *Heredity (Edinb)* **115**, 243–252 (2015).
56. Faucon, F. *et al.* In the hunt for genomic markers of metabolic resistance to pyrethroids in the mosquito Aedes aegypti: An integrated next-generation sequencing approach. (2017) doi:10.1371/journal.pntd.0005526.
57. Rinkevich, F. D., Du, Y. & Dong, K. Diversity and Convergence of Sodium Channel Mutations Involved in Resistance to Pyrethroids. *Pestic Biochem Physiol* **106**, 93–100 (2013).
58. Darsie, R. F. *Keys to the Adult Females and Fourth Instar Larvae of the Mosquitoes of Florida (Diptera, Culicidae)*. *Florida Mosquito Control Association* https://books.google.co.uk/books/about/Keys_to_the_Adult_Females_and_Fourth_Ins.html?id=8IYDuQAACAAJ&redir_esc=y (2003).
59. WHO. *Test Procedures for Insecticide Resistance Monitoring in Malaria Vector Mosquitoes Global Malaria Programme*. http://apps.who.int/iris/bitstream/10665/250677/1/9789241511575-eng.pdf (2016).
60. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
61. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
62. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM – ScienceOpen. https://www.scienceopen.com/document?vid=e623e045-f570-42c5-80c8-ef0aea06629c.
63. McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297–1303 (2010).
64. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. (2012).
65. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. (2012).
66. Poplin, R. *et al.* Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv* (2017) doi:10.1101/201178.
67. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
68. Browning, B. L., Tian, X., Zhou, Y. & Browning, S. R. Fast two-stage phasing of large-scale sequence data. *The American Journal of Human Genetics* **108**, 1880–1890 (2021).
69. Browning, B. L., Zhou, Y. & Browning, S. R. A One-Penny Imputed Genome from Next-Generation Reference Panels. *The American Journal of Human Genetics* **103**, 338–348 (2018).
70. Larsson, A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30**, 3276–3278 (2014).
71. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797 (2004).

Acknowledgments

E.L.C. and S.M are funded by an MRC LID PhD studentship (Grant No. MR/N013638/1). H.A-P is funded by BBSRC LIDo PhD studentship. 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. We would like to thank the extended team at Puerto Rico Vector Control unit for assisting with collections and aiding where possible.

Author Contributions

E.L.C completed collections, with assistance from J.M.Q, R.M and G.B. E.L.C completed bioassays, DNA extraction and PCR amplification. E.L.C did bioinformatic analysis and interpretation with assistance from J.P, M.H, S.M, H.A-P, TGC and SC. All authors commented and edited on various versions of the draft manuscript and approved the final version. E.L.C wrote first draft and compiled the final manuscript.

Data Availability

Sequence data that support the findings of this study have been deposited in the European Nucleotide Archive with the primary accession code PRJEB72548.

Additional Information

There were no competing interests for any authors in the production of this paper.

Supplementary Information

Supplementary 1 Table. *Barcodes used with primers sequences to distinguish samples*

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

Supplementary 2 Table. *Amp-seq assay primers used in multiplex 1 and 2 combinations*

Multiplex 1			
Forward Name	**Forward**	**Reverse Name**	**Reverse**
DomainIV_F1	GCGATCTSATCGAGAAGTA	DomainIV_R1	ATGCTAGCAARTACGTGATG
DomainII_F5	ACAATGTGGATCGCTTCCCG	DomainII_NR1	TGAACCGAAATTGGACAAAAG
Ace_NF1	TGGGGAACGCTRGGAATCTGCG	ACE_NR1	GCATATCGCTGGGCAAACTC
LCO1490	GGTCAACAAATCATAAAGATATTGG	HCO2198	TAAACTTCAGGGTGACCAAAAAATCA*
GSTE_111_F	ATGACGAAGCTCATTTTGTACACG	GSTE_111_R	ACAAAAATAAGCCACACACTACTG
Multiplex 2			
Forward Name	**Forward**	**Reverse Name**	**Reverse**
DomainIIS4_F1	TCTAGATTTAGYGACTCCAR	DomainIIS4_R1	TACCGATGTAGTTCTTGCC
DomainI_F1	TTTCGTCTAATGACCCAAGA	DomainI_R1_N	ARAGAWTTCGCTCACCCG
DomainIIIExon35_F2	GGATCCAGATCATGAACGA	DomainIIIExon35_R1	CATGTCGAACTTCTTATTGGT
RDL_NF1	CTTCTAATTTCTCTCATCAC	Primer-Aeg-R1	CTGGTTATTTGTACAAGTAGCA
GSTE_150_F	GAACCAATCCTTTTCGCC	GSTE_150_R	TGCCTTTTGAGCATTCTTCT

*From Folmer et al, 1994 (16)

Supplementary 3 Table. *Average coverage per amplicon, amplicon length, and described mutations found in that amplicon.*

Forward Amplicon Name	Reverse Amplicon Name	Mean Coverage	Amplicon Size	Insecticide Resistance mutations
Ace_NF1	Ace_NR3	550.8	468	G119S
DomainIIIExon35_F2	DomainIIIExon35_R1	304.3	474	T1520I, F1534C
DomainIIS4_F1	DomainIIS4_R1	284.9	444	G923V
DomainII_F5	II_NR1	61.2	466	L982W, S989P, I1011V/M, V1016I/G
DomainIV_F1	DomainIV_R1	149.3	495	D1763Y
DomainI_F1	Domain1_R1_N	244.5	464	V410L
GSTE2_111	GSTE2_111_R1	7.3	462	L111S *(Ae. aegypti)*, C155F
GSTE2_150	GSTE2_150_R	256.0	321	L119F, I150V *(Ae. aegypti)*
RDL_NF1	Primer-Aeg-R1	40.4	482	A301S
LCO1490	HCO1298	6.5	709	-

Supplementary 4 Table. *Summary of synonymous mutations detected in the 178 samples screened.*

Chrom.	Gene	Position	Nucleic acid change	Annotation	Genotype			Alt. Allele Frequency (n)
					Homo. ref	Hetero	Homo. alt	
3	*ace-1*	161500076	T > A	506T[†]	0	7	97	96.6% (201)
3	*ace-1*	161500136	C > T	486L[†]	102	4	0	1.9% (4)
3	*ace-1*	161500262	G > A	444D[†]	100	4	0	1.9% (4)
3	*ace-1*	161500301	G > A	431P[†]	100	4	0	1.9% (4)
3	*ace-1*	161500361	T > C	411I[†]	45	100	0	34.5% (100)
3	*vgsc*	315931749	T > C	1847Y[‡]	16	91	0	42.5% (91)
3	*vgsc*	315932072	G > A	1762L[‡]	179	1	0	0.3% (1)
3	*vgsc*	315932184	G > A	1724V[‡]	88	2	0	1.1% (2)
3	*vgsc*	315932226	G > C	1710L[‡]	80	80	0	25.0% (80)
3	*vgsc*	315984096	C > T	978L[‡]	60	0	60	50.0% (120)
3	*vgsc*	315984129	A > G	967F[‡]	0	0	62	100.0% (124)
3	*vgsc*	315984159	C > T	957K[‡]	0	0	61	100.0% (122)
3	*vgsc*	316080738	C > T	402L[‡]	225	10	1	2.5% (12)
3	*vgsc*	316081058	G > A	375D[‡]	0	1	118	99.6% (237)

[†]As per transcript AAEL000511-RG;
[‡]As per transcript AAEL023266-RL

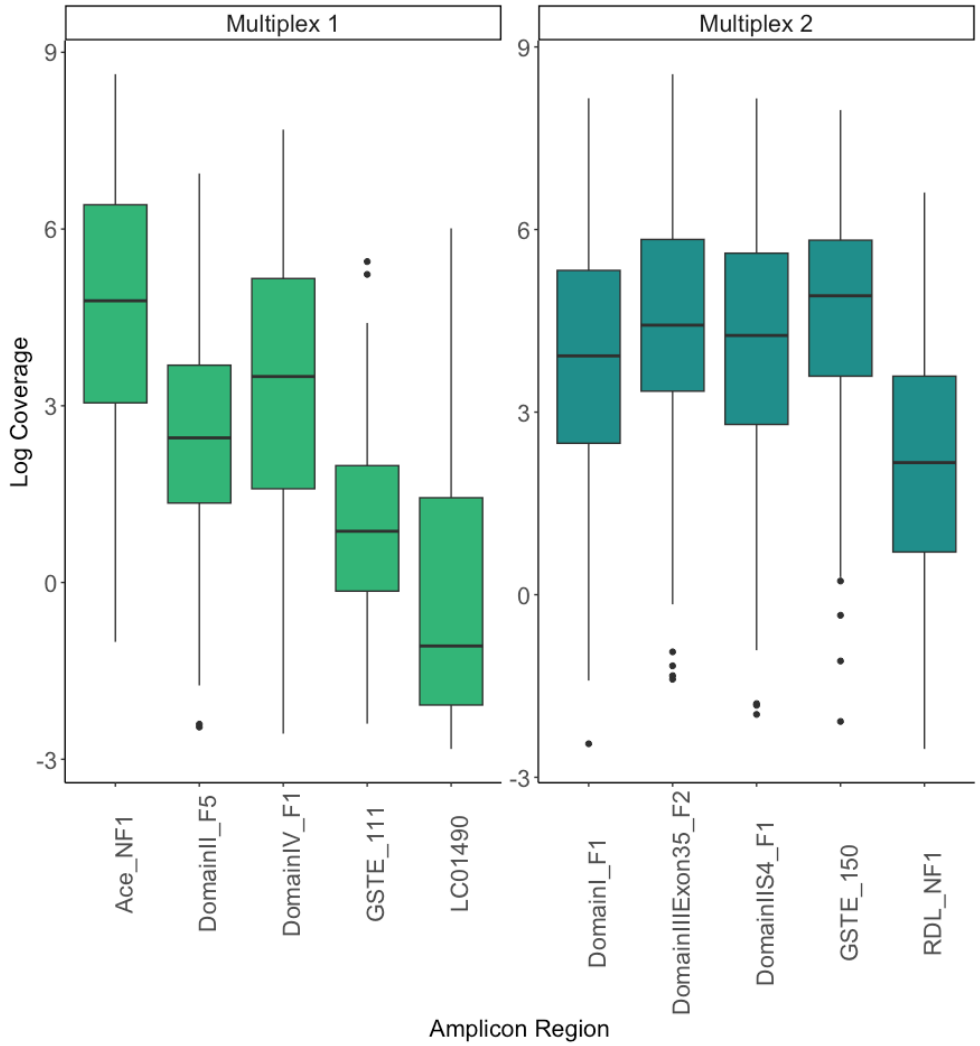
S5 Table. SNPs nomenclature based on reference organism and *Ae. aegypti* from this study.

Gene	Reference organism	Mutation reference	Mutation *Ae. aegypti*	*Ae. aegypti* transcript
vgsc	*Musca domestica*	V410L	V408L	AAEL023266-RL
		L994I	L921I	
		V1016I/G	V1012I/G	
		F1534C	F1554C	
rdl	*Drosophila melanogaster*	A296S	A301S	AAEL008354-RF
Ace-1	Torpedo californica	NA	NA	AAEL034366-RD

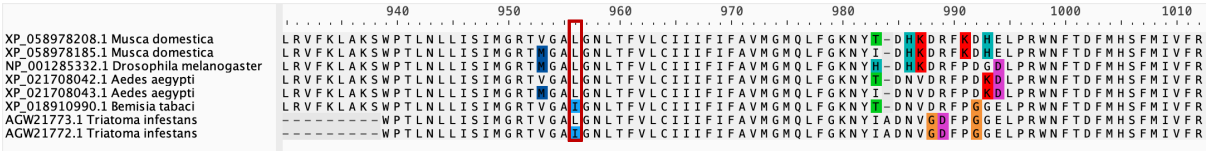
S6 Table. Filtered linked SNP position with $R^2 > 0.8$

Position 1	Position 2	R^2	Position 1	Position 2	R^2
161500301	161500136	1	316080942	316080895	0.8559
315984129	315983809	1	316080950	316080895	0.8559
315984130	315983956	1	316080942	316080896	0.8559
315998782	315998609	1	316080950	316080896	0.8559
316080916	316080895	1	316080942	316080903	0.8559
315983956	315983809	0.9855	316080950	316080903	0.8559
315983973	315983809	0.9855	316080942	316080916	0.8559
315984159	315983809	0.9855	316080950	316080916	0.8559
315984129	315983956	0.9855	316080895	316080847	0.8445
315984129	315983973	0.9855	316080896	316080847	0.8445
315984159	315984129	0.9855	316080903	316080847	0.8445
315984159	315984096	0.9715	316080916	316080847	0.8445
315984159	315983956	0.9711	316080895	316080722	0.8163
315984159	315983973	0.9711	316080896	316080722	0.8163
315984096	315983809	0.9574	316080903	316080722	0.8163
315984129	315984096	0.9574	316080916	316080722	0.8163
316080942	316080722	0.9537	316080983	316080895	0.8163
316080950	316080722	0.9537	316080983	316080896	0.8163
316080983	316080942	0.9537	316080983	316080903	0.8163
316080983	316080950	0.9537	316080983	316080916	0.8163
316080895	316080830	0.9446	316080942	316080830	0.8085
316080896	316080830	0.9446	316080950	316080830	0.8085
316080903	316080830	0.9446	316080942	316080845	0.8085
316080916	316080830	0.9446	316080950	316080845	0.8085
316080895	316080845	0.9446			
316080896	316080845	0.9446			
316080903	316080845	0.9446			
316080916	316080845	0.9446			
315984096	315983956	0.9432			
315984096	315983973	0.9432			
315932226	315931749	0.9337			
316080983	316080722	0.9092			

316081058	316080952	0.8568			



S1 Figure. The log coverage of each amplicon sequenced separated by the multiplex reaction in which they were combined (see **Table 1**).



SI Figure 2. Alignment of protein for voltage gated sodium channel for *Triatoma infestans* (sensitive AGW21773.1 and resistant AWG21772.1 forms), *Bemisia tabaci* and two isoforms of *Aedes aegypti* and *Drosophia melanogaster* showing position of L921I mutation described to be associated with pyrethroid insecticide resistance [32,33].

CHAPTER THREE

Uncovering the genetic diversity in *Aedes aegypti* insecticide resistance genes through global comparative genomics

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	1513859	**Title**	Ms
First Name(s)	Emma Louise		
Surname/Family Name	Collins		
Thesis Title	Utilising genomic approaches to explore genetic diversity and insecticide resistance in Aedes aegypti populations		
Primary Supervisor	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?	June 2024		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	NA		
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	

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)	EC designed the study and collected samples with help from JMQ and RM. EC analysed the data with assistance from SM, JT, HAP. EC wrote the first draft of the manuscript. All authors have edited and approved the final version of the manuscript.

SECTION E

Student Signature	Emma Collins
Date	18/06/2024

Supervisor Signature	Susana Campino
Date	18/06/2024

scientific reports

Check for updates

OPEN

Uncovering the genetic diversity in *Aedes aegypti* insecticide resistance genes through global comparative genomics

Anton Spadar[1,5], Emma Collins[1,5], Louisa A. Messenger[2,3], Taane G. Clark[1,4] & Susana Campino[1]✉

Aedes aegypti **is vector of many arboviruses including Zika, dengue, yellow fever, West Nile, and Chikungunya. Its control efforts are hampered by widespread insecticide resistance reported in the Americas and Asia, while data from Africa is more limited. Here we use publicly available 729** ***Ae. aegypti*** **whole-genome sequencing samples from 15 countries, including nine in Africa, to investigate the genetic diversity in four insecticide resistance linked genes:** ***ace-1***, ***GSTe2***, ***rdl*** **and** ***vgsc*****. Apart from** ***vgsc*****, the other genes have been less investigated in** ***Ae. aegypti*****, and almost no genetic diversity information is available. Among the four genes, we identified 1,829 genetic variants including 474 non-synonymous substitutions, some of which have been previously documented, as well as putative copy number variations in** ***GSTe2*** **and** ***vgsc*****. Global insecticide resistance phenotypic data demonstrated variable resistance in geographic areas with resistant genotypes. Overall, our work provides the first global catalogue and geographic distribution of known and new amino-acid mutations and duplications that can be used to guide the identification of resistance drivers in** ***Ae. aegypti*** **and thereby support monitoring efforts and strategies for vector control.**

Keywords Aedes aegypti, Insecticide resistance, Vector-borne disease, Genomics

Mosquitoes of the genus *Aedes*, particularly *Aedes* (*Ae.*) *aegypti*, are responsible for the transmission of many arboviral diseases, including dengue, Zika, yellow fever, West Nile and Chikungunya, resulting in millions of infections globally per year with limited treatment and vaccination options[1]. The geographical distribution of *Ae. aegypti* has expanded considerably in recent years, predominantly due to adaptation of this vector to urban environments, climate change and the globalization of human activities, thereby increasing the risk of resurgence and spread of arbovirus infections[2–4]. Compounding the problem is the global emergence of insecticide resistance among *Ae. aegypti* and other mosquito species, which is threatening to jeopardise the operational effectiveness of vector control campaigns.

Resistance to the four most common classes of insecticides used against adult mosquitoes (carbamates, organochlorines, organophosphates, and pyrethroids) has now been documented worldwide. Resistance in many mosquito species has been associated with target site mutations, metabolic detoxification, cuticular alterations and behavioural avoidance[5,6] with a suite of alternative resistance mechanisms being revealed[7–10]. Target site resistance is related to mutations in genes that code for insecticide target molecules, such as the voltage-gated sodium channel (*vgsc* also known as knockdown resistance; *kdr*), acetylcholinesterase-1 (*ace-1* also known as *AChE1*) and γ-aminobutyric acid (GABA) receptor (resistance to dieldrin; *rdl*). Mutations in glutathione-s-transferase epsilon two (*GSTe2*), which encodes an insecticide metabolising enzyme, have also been associated with resistance[5,11–13]. The *vgsc* is a large protein that is an integral part of the insect nervous system. DDT (dichloro-diphenyl-trichloroethane) and pyrethroid insecticides interfere with the *vgsc* by prolonging the pore open state leading to insect paralysis and death[14]. In the reference insect for this gene, *Musca domestica*, the most frequent

[1]Faculty of Infectious and Tropical Diseases, Department of Infection Biology, London School of Hygiene and Tropical Medicine, London, UK. [2]Department of Environmental and Occupational Health, School of Public Health, University of Nevada, Las Vegas, Las Vegas, NV, USA. [3]Parasitology and Vector Biology Laboratory (UNLV PARAVEC Lab), School of Public Health, University of Nevada, Las Vegas, NV, USA. [4]Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK. [5]These authors contributed equally: Anton Spadar and Emma Collins. ✉email: Susana.campino@lshtm.ac.uk

kdr resistance mutations are S989 and L1014[15]. In *Ae. aegypti,* the 1014 codon requires at least two mutations to change to a *M. domestica* amino acid known to cause resistance; thus, the substitution L1014F, seen pervasively in *Anopheles* mosquitoes, has not been observed in this species[11]. Instead, F1534C/L, V1016I/G, I1011V/M and V410L mutations have been associated with pyrethroid resistance in *Ae. aegypti* and confirmed experimentally[6]. Other amino acid substitutions reported previously in *Ae. aegypti* include G923V, L982W, S989P, T1520I and D1763Y[11,16–18]. Many of these mutations are often found in combination and appear only on specific continents. For example, V1016G and S989P appear limited to Asia, while V1016I has only been identified in the Americas and Africa and 723T only in the Americas[19].

The *ace-1* gene encodes acetylcholinesterase (AchE1), which is responsible for hydrolysis of acetylcholine terminating the transmission of neural signals. Organophosphates and carbamates bind to the acetylcholinesterase active site which inhibits hydrolysis and consequently neural signal termination, leading to insect death. Unlike mammals and some insects (including *Drosophila melanogaster*), mosquitoes usually have two copies of the *ace-1* gene. In *Anopheles* mosquitoes, the G119S amino acid substitution in *ace-1* is generally associated with resistance (all coordinates are based on *Torpedo californica*)[20,21]. As with the *vgsc*, in *Ae. aegypti* such an amino acid change requires two mutations and has only been observed in one study in India[22]. Despite the lack of described mutations in *ace-1*, resistance to organophosphates in *Aedes* is widespread in the Americas and Asia, while data from Africa is limited[6].

The *rdl* mutation is found in the γ-aminobutyric acid (GABA) receptor gene that controls neural signal inhibition through opening and closing of the transmembrane chloride channel on the cells of the mosquito nervous system. Cyclodienes (e.g., dieldrin) prevent interaction of GABA with its receptor, leading to neuron hyperexcitation and eventual insect death[23–26]. The most common resistance mutation in this gene is A301S/G (*D. melanogaster* numbering) and is observed in multiple insects including mosquitoes of the *Anopheles* and *Aedes* genera[21,27]. Despite a ban on the use of cyclodienes in 2001[28] due to their slow degradation and environmental persistence, *rdl* mutations have persisted for decades later in vector populations, suggesting that they impart limited fitness costs[29,30].

Unlike *rdl*, *ace-1* and *vgsc,* which are targets of insecticides, the homodimer glutathione S-transferase (GST) is a detoxifying enzyme. Most organisms, including *Ae. aegypti*, have multiple GST enzymes of which epsilon two (GSTe2) has been associated with resistance to both DDT and pyrethroids[6,12,31,32]. The *GSTe2* gene contributes to insecticide resistance through both enzyme overexpression and point mutations. Increased expression of this gene was linked to DDT resistance in *An. gambiae*[5,25,26,33]. The L119F substitution in *GSTe2* was observed to enhance resistance to both DDT and pyrethroids in *An. funestus,* and I114T exacerbated resistance to DDT in *An. gambiae*[5,33–35]. In *Ae. aegypti,* L111S and I150V mutations have been linked to temephos resistance in silico[36].

Despite observed phenotypic resistance of *Ae. aegypti* to all main insecticide classes across many countries in Africa, Americas, and Asia[6], the distribution of genetic variants in underlying candidate genes is less studied across *Aedes* populations compared to *Anopheles* species. Here, we examined a large (n = 729), globally diverse dataset of publicly available *Ae. aegypti* whole genome sequencing (WGS) data to uncover the genetic diversity present in *vgsc, ace-1, rdl* and *GSTe2.* The diversity in insecticide resistance loci was interpreted alongside current global trends in phenotypic insecticide resistance in *Ae. aegypti*. This data provides a catalogue of genetic variants that could be involved in insecticide resistance and supports further studies on the molecular surveillance of emerging and spreading insecticide resistance mechanisms amongst *Ae. aegypti* populations.

Material and methods

Aedes aegypti genomic data

We searched the NCBI SRA database for "*Aedes aegypti*" sample data and restricted results to WGS libraries where the number of bases contained implied at least fivefold coverage when mapped to the reference genome AaegL5 (GCF_002204515.2)[32]. We obtained a total of 703 WGS *Ae. aegypti* (non-AaegL5) libraries from 15 countries, across Africa (n = 476, 8 countries), the Americas (n = 191, 3 countries), Oceania (n = 16, 1 country) and Asia (n = 20, 1 country), and 26 colony samples of which 20 had known country of collection. Additionally, we included 7 *Ae. mascarensis* samples from Madagascar (n = 4) and Mauritius (n = 3) as outgroup[37–41] (Table S1).

Insecticide resistance phenotypic data

Insecticide response data was only available for the Bora-Bora susceptible reference strain, which has been maintained in the insectary for 134 generations without any exposure to insecticides[42] and the Nakon Sawan reference strain, which is resistant to deltamethrin and temephos[41,43], (Table S2). Global insecticide resistance phenotype data was retrieved from the IR Mapper tool[44] (sourced on 19/04/2023), which covered 73 countries of which 8 overlap with samples in this study. No data was available for 5 countries (Kenya, Madagascar, Mauritius, South Africa, and Uganda); an additional literature search in PubMed failed to retrieve additional publicly available phenotypic data for *Ae. aegypti* in these countries. We included the data where the phenotype was tested with World Health Organization (WHO) tube or bottle bioassay or Centers for Disease Control and Prevention (CDC) bottle bioassay. Phenotypic data based solely on PCR or RT-PCR methods were excluded. Overall, we analysed 3172 data points for 19 different insecticides across four insecticide classes (Pyrethroids, Organophosphates, Organochlorines and Carbamates) (Table S3). Data points from IR mapper were reported as susceptible, possible resistance or resistant based on mortality as per WHO and CDC guidelines.

Bioinformatic analysis

We aligned the WGS libraries using bowtie2 (v2.4.1) software (with a setting *--fast-local)*[45]. We processed the alignment files using samtools (v1.7) software and SNPs were called using the GATK HaplotypeCaller tool (v4.1.9) with default settings[46,47]. A minimum coverage of 5-fold was used to accept SNPs. We merged the

individual VCF files into a multi-sample file using BCFtools (v1.9)[48]. The impact of SNPs in the multi-sample VCF was predicted using snpEff software (v5.0) with AaegL5 genome annotation (GCF_002204515.2)[49]. The alignment process was performed against the mRNA sequences of twenty *Ae. aegypti* genes (Table 1). Four were loci linked to insecticide resistance [*vgsc* (XM_021852340.1), *rdl* (XM_021840622.1), *ace-1* (XM_021851332.1) and *GSTe2* (XM_021846286.1)] and the remaining sixteen genes were used to establish population structure. One of these was mitochondrial *cox1* (YP_009389261.1) and the remaining fifteen genes were evenly spread across all three *Ae. aegypti* chromosomes (Table 1). These 15 genes were determined to have unique genome-wide exon sequences (using NCBI BLASTn v2.9.0 with—*word-size* 28 and—*evalue* 0.01) which minimised potential mis-mapping of WGS reads to the *Ae. aegypti* genome known to contain many duplications[50]. Read coverage per nucleotide per gene was calculated using the samtools "depth" function and was used to identify possible gene duplications in samples[48]. We merged the coverage data into a single data matrix and removed all regions except gene exons, because intronic regions contained high numbers of repeats. For each sample, we divided each per base coverage value by that sample's overall median coverage across all genes, except *vgsc* and *GSTe2*, which may have copy number variants. We applied UMAP (v0.5.1) software (with a *Euclidean* distance metric) on this scaled matrix to identify gene clusters based purely on the coverage[51].

Population genetics analysis

To determine population structure, we used UMAP software (with *Russell-Rao* distance metric) on the multi-sample VCF, followed by application of HDBSCAN (v0.8.28)[51,52] to determine sample clustering (see[53–55] for recent applications). This work was performed in python (v3.7.6), with scripts available from https://github.com/AntonS-bio/resistance-AedesAegypti. Linkage disequilibrium was calculated using vcftools on phased vcf files created with beagle (v 22Jul22.46e) software to provide a R^2 value for each combination of non-synonymous mutations by sample country. Plots of these values were visualised using the gaston (v1.5.9) package in R.

Protein structure modelling

Protein structure modelling was performed using AlphaFold Multimer software with full protein databases[56,57]. When referring to substitutions and their effects on proteins, we have followed the established nomenclature based on reference resistance linked proteins and structures in the protein databank: ACE1 (2C4H; *Tetronacre californica*), GABA receptor (NP_729462.2; *Drosophila melanogaster*), GSTe2 (XP_319968.3; *An. gambiae*) and VGSC (NP_001273814.1; *Musca domestica*)[58,59]. Unless otherwise specified, all substitution coordinates are with respect to these reference sequences.

Gene	Product	Chr	CDS Len	Resistance gene	Unique missence SNPs	Unique synonymous SNP
XM_021851049.1	TATAmodulator	NC_035107.1	3178		557	1040
XM_001648700.2	Ydcl	NC_035107.1	2902		380	602
XM_021851750.1	LOC110678629	NC_035107.1	1095		35	13
XM_021857384.1	LOC5580295	NC_035107.1	2479		657	437
XM_001652683.2	PotassiumChannel	NC_035107.1	1017		39	262
XM_021840622.1	GABA	NC_035108.1	1653	Yes	64	180
XM_021841341.1	AngiogenicFactor	NC_035108.1	1811		334	439
XM_001664194.2	TIFIID2	NC_035108.1	3755		217	738
XM_001662595.2	Mcm6	NC_035108.1	2429		99	167
XM_001657120.2	Cytochromeb-c1	NC_035108.1	797		48	122
XM_021846286.1	GSTE2	NC_035108.1	666	Yes	109	158
XM_021847043.1	Carbohydratesulfotrans-ferase	NC_035108.1	1330		194	291
XM_001657462.3	LOC5567548	NC_035109.1	1745		334	396
XM_021850261.1	ZincFinger	NC_035109.1	1497		235	267
XM_021851332.1	ACE1	NC_035109.1	2102	Yes	99	144
XM_001649087.2	grpE	NC_035109.1	676		63	59
XM_001649790.2	LOC5565494	NC_035109.1	2445		527	535
XM_021852340.1	VGSC	NC_035109.1	6379	Yes	202	873
XM_021853012.1	LOC5579101	NC_035109.1	1836		378	208
YP_009389261.1	COX1	NC_035159.1	1536		1230*	0*

Table 1. The genes analysed. *The annotation in GCF_002204515.2 assembly has missing start codon for mitochondrial *cox1* and as a result snpEff did not distinguish between synonymous and non-synonymous SNPs. CDS = coding sequence.

Results

Genetic variation and population structure

Across the 729 *Aedes* samples from 15 countries, a total of 1829 SNPs (474 non-synonymous (NS)) were detected across the CDS of four insecticide resistance associated genes (*vgsc*, *rdl*, *ace-1* and *GSTe2*), and 9673 SNPs were identified across the CDS of 15 non-resistance associated genome-wide gene (Tables 1, 2, Table S4).

Using the SNPs from the CDS of 15 genes not associated with insecticide resistance, a UMAP clustering analysis revealed five distinct clusters (Fig. 1A), broadly linked to: (i) eastern Kenya and South Africa (n = 112); (ii) west, central Africa and west Kenya (n = 350); (iii) the Americas, Thailand, and other (n = 258); (d) the Bora-Bora mosquito line from French Polynesia (n = 9); (e) *Ae. mascarensis* from Madagascar and Mauritius (n = 7). Similar results were obtained when analysing only the 1829 SNPs in genes that are associated with resistance (Fig. 1B). These results are broadly consistent with previous reported population structure of *Ae. aegypti* using SNPs and microsatellite data, where African samples formed one cluster and samples from Asia, America and the Caribbean comprised another cluster[60]; however, more focused studies provide better understanding of population structure[38,60–63]. As we observed a separation of most eastern Kenyan samples (n = 121) from west Kenya (n = 37), we investigated the genotype data in these groups independently. Some eastern Kenyan samples (n = 14/121) from a human- biting colony of domestic *Ae. aegypti,* originally collected indoors in Rabai[60,64], clustered with non-African samples (Americas and Thailand and other cluster), as previously observed. When including only non-African samples, the UMAP clustering analysis revealed modest separation of the samples from Brazil, Mexico, French Polynesia, American Samoa and Thailand. For the samples from Africa, clustering separated east Kenyan samples from the rest (Fig. S1). The same patterns were detected across both resistance and non-resistance genes (Fig. S1). Clustering using mitochondrial *cox1* gene was different from the results based on chromosomal loci (Fig. 1C–F). In multiple samples, SNPs had heterozygous *cox1* genotypes possibly multiploidy due to the presence of previously described copies of nuclear mitochondrial (NUMT) DNA which could confound clustering[65,66].

Genetic variation across insecticide resistance associated genes

Vgsc

In the *vgsc* gene, a total of 1075 SNPs (202 non-synonymous; NS) were identified, of which 36 NS SNPs were present in > 1 sample, including eight mutations previously linked to insecticide resistance (V410L, G923V, S989P, I1011M, V1016I/G, T1520I and F1534C) (Table 2, Table S4). We did not observe any other pyrethroid resistance associated substitutions such as L982W, detected previously in Vietnam and Cambodia, and D1763Y reported in Taiwan. However, the D1763G mutation was present in a single USA sample[11,16–18]. The most frequent mutations were F1534C (39%), S723T (23%), V410L (22%) and V1016I (22%) (Fig. 2). The most prevalent F1534C mutations occurred in nearly all samples from the Americas (186/191) and Thailand (20/20). The frequency of F1534C was lower in African samples, appearing only in Burkina Faso (n = 20/34), Ghana (n = 33/58), Nigeria (n = 1/19) and East Kenya (n = 8/107). The F1534C mutation was accompanied by V1016I, S723T and V410L substitutions in most samples from USA, Burkina Faso, and Mexico, as well as in a single Nigerian sample. In Thailand, F1534C co-occurred in many samples with V1016G, T1520I and S989P (Table 2).

*3 samples (SRR11006697, SRR11006705, SRR11006885) are geographically from west Kenya and 1 from Uganda (SRR11006909), but were assigned to East Kenya cluster.		Population Groups	West and Central Africa							East Kenya*		Americas, East Kenya* and others						Indian Ocean		LAB		
			Senegal	Burkina Faso	Ghana	Nigeria	West Kenya*	Gabon	Uganda	S. Africa	East Kenya*	Brazil	USA	American Samoa	Mexico	East Kenya Outlier	Thailand	Madagascar	Mauritius	LAB - NakhR	LAB - Bora-Bora	LAB - Other
	Reference Org Position	Region Total (ex. Lab)	110	34	58	19	37	56	39	2	107	16	160	16	15	14	20	4	3	8	9	9
VSSC_p.Phe1554Cys	p.Phe1534Cys	278	0	20	33	1	0	0	0	0	8	16	155	10	15	0	20	0	0	8	0	4
VSSC_p.Ser711Thr	p.Ser723Thr	163	0	18	0	1	0	0	0	0	0	0	136	0	8	0	0	0	0	0	0	3
VSSC_p.Val408Leu	p.Val410Leu	163	0	18	0	1	0	0	0	0	0	0	136	0	8	0	0	0	0	0	0	2
VSSC_p.Val1012Ile	p.Val1016Ile	160	0	18	0	1	0	0	0	0	0	0	133	0	8	0	0	0	0	0	0	1
VSSC_p.Met940Val	p.Met944Val	42	0	0	0	0	0	0	0	0	42	0	0	0	0	0	0	0	0	0	0	0
VSSC_p.Asp1653Glu	p.Asp1633Glu	18	11	0	0	1	0	0	0	0	0	0	0	0	0	0	0	3	3	0	0	0
VSSC_p.Leu942Gln	p.Leu946Gln	16	0	0	0	1	0	0	0	0	0	15	0	0	0	0	0	0	0	0	0	0
VSSC_p.Ser65Phe	p.Ser66Phe	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	1	0	0
VSSC_p.Thr1540Ile	p.Thr1520Ile	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	1	0	3
GABA_p.Ser110Thr	p.Ser115Thr	707	110	34	58	19	37	56	39	2	107	16	157	16	15	14	20	4	3	8	9	9
GABA_p.Ala296Ser	p.Ala301Ser	130	3	4	2	0	0	7	0	1	3	2	97	0	0	0	11	0	0	1	0	0
GABA_p.Glu79Asp	p.Glu84Asp	18	3	2	2	0	0	4	0	0	2	0	0	0	0	0	5	0	0	0	0	0
ACE1_p.Cys699Ser		42	2	13	9	1	0	4	0	0	0	5	7	0	1	0	0	0	0	0	0	2
ACE1_p.His35Leu		37	9	11	8	2	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ACE1_p.Gly12Ser		35	1	1	0	0	0	0	0	0	0	4	11	0	11	0	7	0	0	0	0	0
ACE1_p.Ser693Leu		20	0	0	0	0	0	2	0	0	17	0	0	0	0	1	0	0	0	0	0	0
ACE1_p.Asp131Glu	p.Asp4Glu	16	0	1	1	1	3	4	0	0	6	0	0	0	0	0	0	0	0	0	4	0
ACE1_p.Leu687Phe		14	1	0	1	0	0	2	2	0	7	0	1	0	0	0	0	0	0	0	0	0
GSTE2_p.Ile150Val	p.Ile150Val	670	110	34	58	19	35	56	39	2	105	16	147	16	15	11	0	4	3	0	8	8
GSTE2_p.Ala198Glu	p.Ala198Glu	670	108	33	58	19	37	55	39	2	104	16	150	16	15	11	0	4	3	0	8	9
GSTE2_p.Cys115Phe	p.Cys115Phe	542	82	19	39	13	37	39	30	2	105	16	114	16	15	14	0	1	0	0	9	8
GSTE2_p.Leu111Ser	p.Leu111Ser	288	25	7	24	4	3	15	3	2	48	16	111	16	14	0	0	0	0	0	9	4
GSTE2_p.Ile169Ser	p.Ile169Ser	172	45	17	13	6	8	20	12	0	8	0	42	0	0	1	0	0	0	0	0	0
GSTE2_p.Leu9Ile	p.Leu9Ile	151	30	17	14	5	10	18	8	0	9	0	39	0	0	1	0	0	0	0	0	0
GSTE2_p.Cys115Ser	p.Cys115Ser	108	35	0	13	3	16	14	20	0	4	0	3	0	0	0	0	0	0	0	1	0
GSTE2_p.Met47Ile	p.Met47Ile	81	14	3	11	4	20	6	18	0	5	0	0	0	0	0	0	0	0	0	0	0
GSTE2_p.Val34Ile	p.Val34Ile	47	0	0	0	0	1	3	0	2	31	0	10	0	0	0	0	0	0	0	0	0
GSTE2_p.Lys23Arg	p.Lys23Arg	38	0	0	0	0	0	0	0	1	31	0	6	0	0	0	0	0	0	0	0	0
GSTE2_p.Gly123Glu	p.Gly123Glu	30	1	0	1	0	14	1	11	0	2	0	0	0	0	0	0	0	0	0	0	0
GSTE2_p.Val11Glu	p.Val11Glu	12	0	0	0	0	0	0	0	0	12	0	0	0	0	0	0	0	0	0	0	0
GSTE2_p.Ser124Tyr	p.Ser124Tyr	12	0	0	0	0	0	0	0	0	2	0	0	0	0	10	0	0	0	0	0	0
GSTE2_p.Pro174Ser	p.Pro174Ser	11	2	1	1	0	0	4	3	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 2. Missense mutations identified in samples and occurring in more than 10 non-lab sample. The full list of mutations is available in Supplementary Table B.

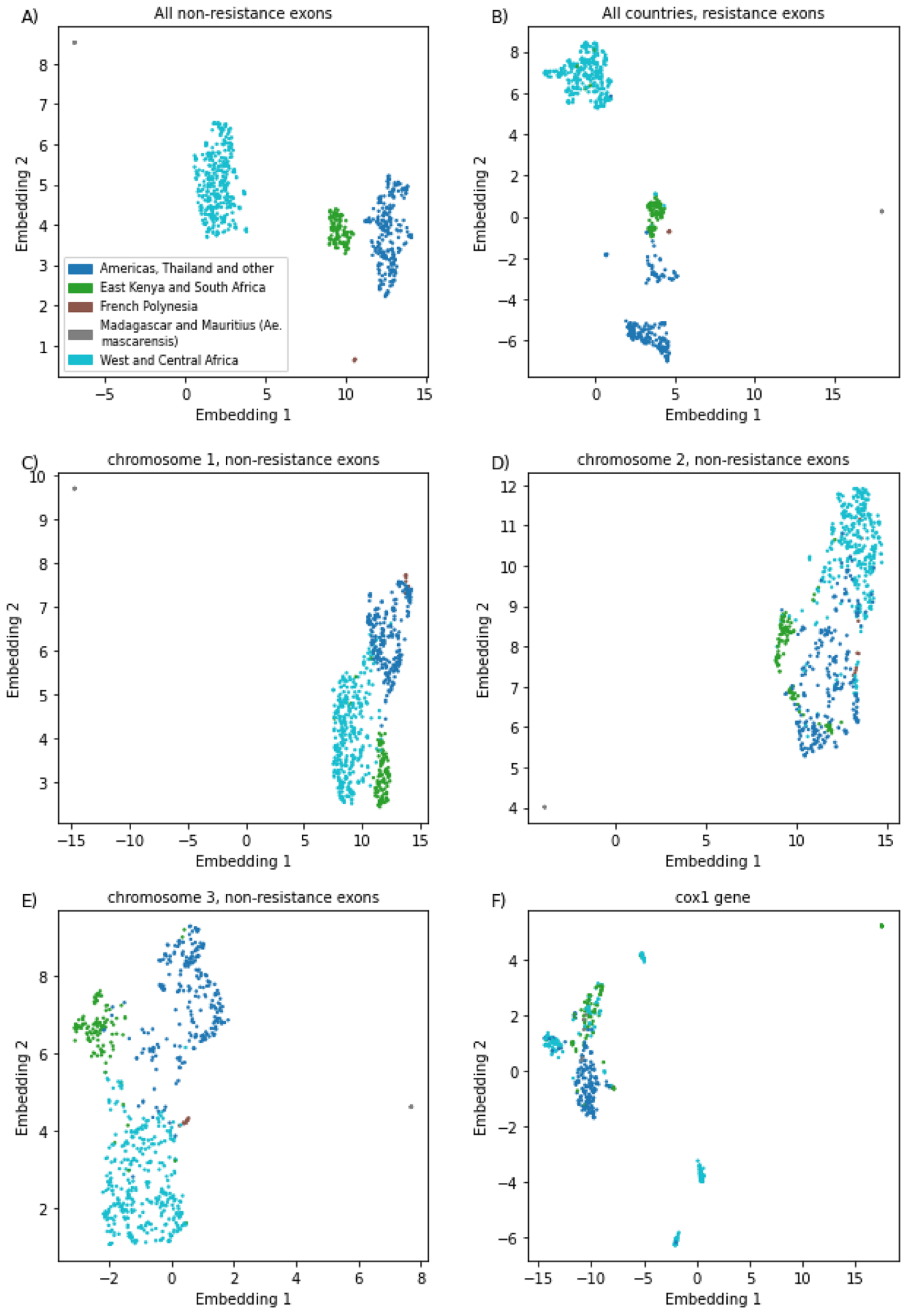


Figure 1. Population structure using UMAP embedding of SNPs from non-resistance linked genes (**A**,**C**,**D**,**E**), and resistance linked genes (**B**) and *cox1* (**F**).

Several mutations were found to be regionally specific. The V1016G mutation was found only in Asia (Thailand) while V1016I was detected in USA, Mexico, and a few countries in Africa[19]. The M944V substitution was unique to East Kenya (n = 42/107), L946G was almost exclusive to Brazil (n = 15/16) except for one Nigerian sample. The V1016G, T1520I (n = 10/20), S989P (n = 7/20), and S66F (n = 11/20) were also almost exclusive to Thailand, apart from a single Nigerian and a Brazilian sample (Table 2). Two conservative in-frame insertions occurred in ~ 20% of west and central African samples, which included an addition of amino acid Glycine (Gly)

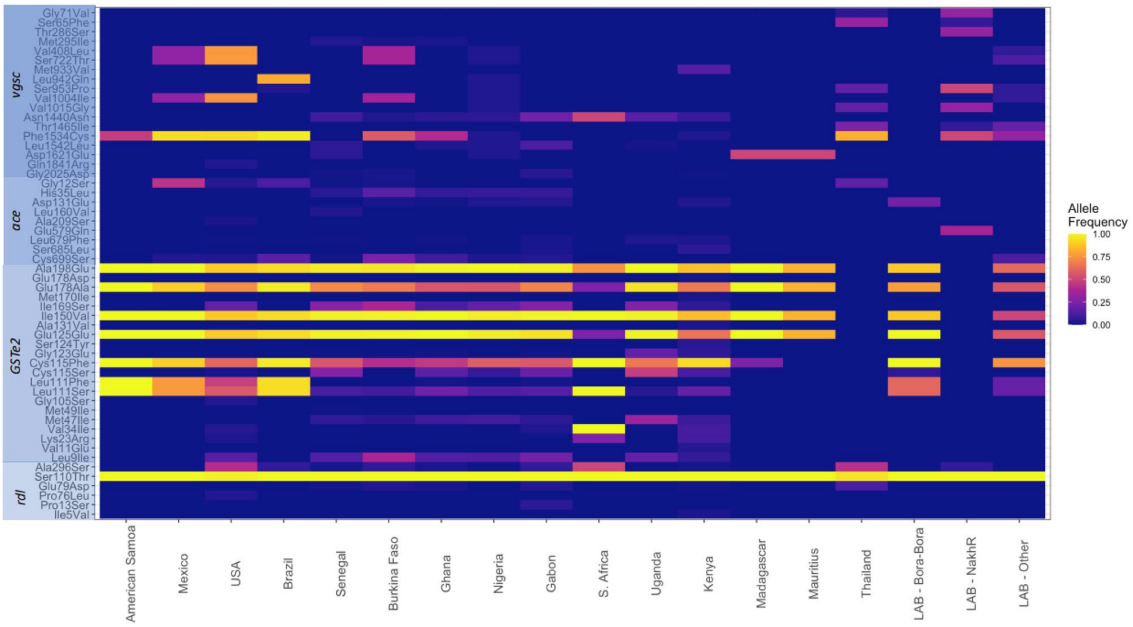


Figure 2. Allele frequency of each missense SNP across the insecticide resistance associated genes; *vgsc, ace-1, rdl, and GSTe2*, by country. Only SNPs with at least 10 samples with a non-reference allele are shown mutation.

into a sequence of four consecutive Gly (positions 2047–2050), and an addition of Serine-Glycine (positions 2016 and 2017).

Rdl (GABA receptor)

In the *rdl* gene, we identified a total of 244 SNPs (64 NS), of which only 17 NS SNPs occurred in > 1 sample and the most frequent were G84A, S115T and A301S. The S115T substitution was present in almost all samples (n = 733/736) including all *Ae. mascarensis* (Fig. 2, Table 2). The T115 is the dominant allele in *An. gambiae* suggesting that the common ancestor of both *An. gambiae* and *Ae. aegypti* had the 115T allele, and a mutation in the *Ae. aegypti* reference strain changed T to S[67].

The previously described A301S substitution, associated with resistance to organochlorines, was frequent in the USA (n = 97/160) and Thailand (n = 11/20), and infrequent in a few countries in Africa (Table 2)[21,27]. This substitution is located on the a-helix forming the protein pore (Fig. S2). The only other notable mutation was E84D present in 18 samples (Africa n = 13, Thailand n = 5), and located on the outward facing section of the protein but could not be robustly modelled by the AlphaFold software.

Ace-1

A total of 243 SNPs were identified in the *ace-1* gene, of which 99 led to amino-acid substitutions, with 30 present in > 1 sample (Table 2). Only 6 amino-acid substitutions (G12S, H35L, D131Q, L687F, S693A, C699S) occurred in > 10 samples (Fig. 2). The most frequent mutation was C699S (n = 42/736), which was present in samples from west and central Africa (n = 29) and the Americas (n = 13). The second most frequent substitution was H35L (5.0%) observed only in west and central African samples. The third most frequent substitution was G12S (4.8%) found mostly in the Americas (n = 26/37) and Thailand (n = 7/37) (Table 2). All three substitutions are defined in *Ae. aegypti* coordinates because these amino acids are outside the range of the *T. californica* reference ACE1 (PDB: 2C4H). In fact, only 20 substitutions had a corresponding coordinate in the *T. californica* protein (Table 2). The only substitution in *Ae. mascarensis* was T55P (*T. californica* coordinates) present in all samples of this species. We modelled the ACE1 protein structure in AlphaFold, and in line with results of crystallographic experiments, the residues 1–131 and 660–702 were disordered, likely reflecting their role in anchoring the protein to the cellular membrane and receptor proteins[68]. The G119S resistance substitution commonly reported in ACE1 in other insect species was not detected in this dataset. This absence is likely because G119S would require two nucleotide substitutions in *Ae. aegypti*. Further, instead of two *ace* genes commonly found in insects, the *Ae. aegypti* reference genome has four *ace* genes including one analysed here (LOC5578456) and three others (LOC5574466, LOC5575867, LOC5570776). The mRNA encoding the cognate proteins had < 5% pair-wise coverage which rules out recent duplication as the origin of these genes. One of these loci (LOC5570776) had the 119S amino acid. We found that despite the very high prevalence of transposable elements in *Ae. aegypti,* this gene remains uninterrupted by them suggesting this locus might be functional[32].

GSTe2

The *GSTe2* gene has a variable copy number in *Ae. aegypti,* and the reference genome contains four copies of this gene[32]. The variable copy number was also evident in our analysis. Because we used short read data, we could not robustly assign each mutation to individual *GSTe2* loci. A total of 267 SNPs were detected in *GSTe2* genes, with 109 leading to amino-acid substitutions, of which 42 were present in >1 sample (Table 2). Seven substitutions were highly frequent: I150V (n = 670), A198E (n = 670), C115F (n = 542), L111S (n = 288), I169S (n = 172), L9I (n = 151) and C115S (n = 108) (Fig. 2). The samples from Thailand had neither synonymous nor missense mutations in *GSTe2,* which we confirmed by visual examination of the read alignments. The C115F substitution was present in almost all countries (except Thailand and Mauritius). The C115S substitution was most common in Africa (n = 101/353). In addition to C115F/S, we observed two other common substitutions (L111S, L9I) at the DDT binding site[69]. The L111S substitution (n = 288/736) appears globally distributed, and L9I was found mainly in Africa and USA, but not observed in *Ae. mascarensis*. The I169S mutation was common in the presence of L9I. Based on a high confidence AlphaFold protein structure model for GSTe2, the I169S mutation is not part of either glutathione or DDT binding site; however, it interacts with both F115 and M111, which are part of the glutathione binding pocket (Fig. S3).

Gene duplications

Gene variable copy numbers were identified based on excess median-scaled read coverage. For the *vgsc* gene, a group of 26 samples had potential duplications, with a median-scaled coverage of 1.4-fold compared to 1.0-fold for the rest of the samples. The samples in this set came from a disparate group of countries: Senegal (n = 13), American Samoa (n = 4), and USA (n = 3), Mexico (n = 2), Mauritius (n = 2), Kenya (n = 1) and Thailand (n = 1) (Table S1).

For *GSTe2*, two groups of samples had likely copy number events. First, a group of samples with median 4.2-fold median-scaled coverage consisting of samples from Thailand (n = 27/28) including samples from the Nakh lab strain, USA (n = 38/160), Mexico (n = 5/15), Brazil (n = 1/16) and two from the Vienna F4 colony[70]. A second group consisted of samples from USA (n = 15/160) and Mexico (n = 9/16) with median-scaled coverage of 9.3-fold compared to 0.9-fold for the rest of the samples (Table S1, Fig. S4). In our search of the literature, we did not identity previous reports of such high duplication rate; this finding requires further validation. However, this result also shows that majority of *Ae. aegypti* reference sequence have single copy of *GSTe2*, in contrast to the reference strain which has four[32].

Linkage disequilibrium between missense mutations

We examined the geographical distribution of the non-synonymous SNPs across the four resistance genes and observed that many mutations co-occur together in certain populations (Fig. 2). For each locus, per population, we assessed the pairwise linkage disequilibrium (LD) of non-synonimous SNPs. We found twenty-seven pairwise SNPs that had, without adjusting for multiple testing, an R^2 value above 0.5 (*GSTe2* n = *15*, *vgsc* n = *9*, *ace-1* n = *2,* and *rdl* n = *1*) (Table S5). The *GSTe2* mutations L9I/I169S (Burkina Faso, Kenya, Gabon, Ghana, Uganda) and I150V/A198E (Kenya, French Polynesia, Mauritius) were detected with a $R^2 > 0.5$ in several countries. In the *vgsc* gene, several SNPs that have been associated with insecticide resistance also had $R^2 > 0.5$, particularly V410L, V1016I, V1016G and F1534C.

Geographical distribution of insecticide resistance mutations and phenotypes

The IR mapper was used to obtain phenotypic data for 8 of the 15 countries examined in this study. These phenotypes show disparity between the availability of phenotypic and genomic data, for example, Brazil and Thailand have the highest number of bioassay records while only having 16 and 20 genomic sequences available, respectively. However, in some countries there was genomic data available with limited phenotypic data, such as Uganda and Kenya. Phenotypic data available for each country from IR Mapper was mapped to the co-occurrence of nine mutations previously associated with insecticide resistance (A301S (RDL) associated with organochlorine resistance, and F1534C, T1520I, V1016I/G, I1011V/M, S989P, G923V, V410L (VGSC) all associated with pyrethroid resistance). Thailand, Burkina Faso, and the USA had the highest proportion of samples with several known insecticide resistance mutations (Fig. 3). This is supported by the Thailand phenotypic data from IR Mapper, which shows reports of resistance to all four main insecticide classes in this country (Fig. 4), particularly to organochlorines, carbamates and pyrethroids. Elevated levels of resistance have also been reported in southeast Asian regions, such as Indonesia, Malaysia, and Thailand; however, there are gaps in the genomic data from these countries[71–74]. For the USA there is no information on phenotype data on IR Mapper, but resistance to pyrethroids has been reported in several states[75–77].

In Africa, 53% of samples from Burkina Faso had more than two insecticide resistance mutations, all in the *vgsc* gene. Burkina Faso also had the highest reported resistance to pyrethroids when compared to the other African samples in this data set (Nigeria, Senegal, Ghana, and Gabon). Levels of resistance to pyrethroids varied between the 8 countries analysed here. The highest levels of resistance were also observed in Brazil, Mexico, and Thailand, coinciding with samples with the most mutations in the *vgsc* gene (excluding the USA, where limited phenotypic data is available) (Figs. 3, 4).

The data from IR mapper showed that the largest number of reports of resistance involved insecticides of the organochlorine class. Mutations associated with this resistance include SNPs in the *vgsc* and *rdl* genes. However, countries with high resistance to organochlorines, such as Senegal and Nigeria have no or very low frequency of mutations in these loci. As the genomic data presented here do not have matching phenotypic information, it is possible that these samples were from a susceptible background or that there are other mechanism of resistance causing the observed phenotype. The least resistance was reported against organophosphates, although resistance

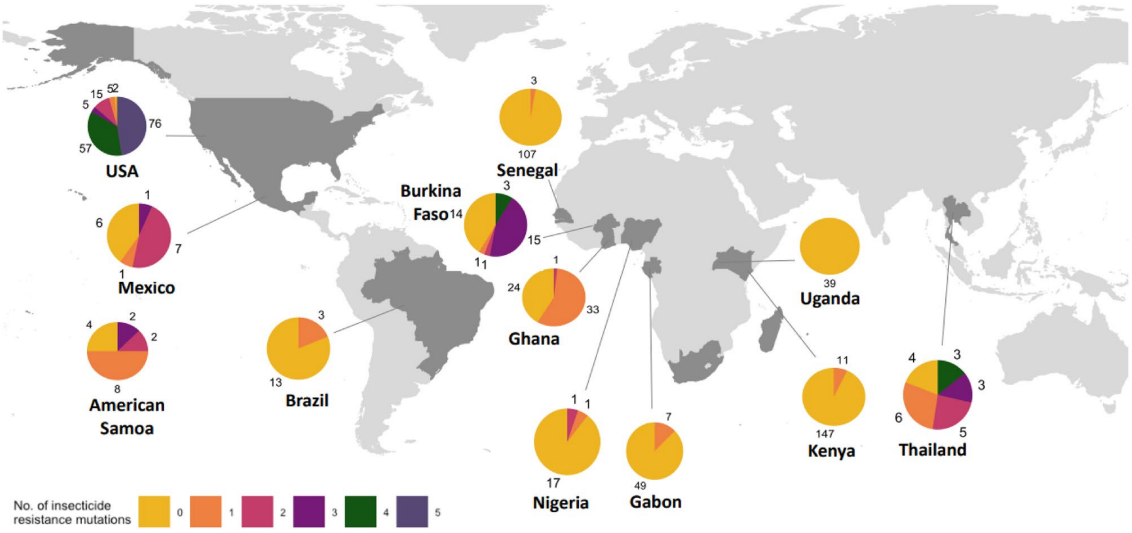


Figure 3. Proportion of samples with 1 or more mutations associated with insecticide resistance in each geographical population. Insecticide resistance SNPs included are: A301S *(rdl)*, F1534L/C, T1520I, V1016I/G, I1011V/M, S989P, G923V, V410L (*vgsc*). Only populations with more than 10 samples were included.

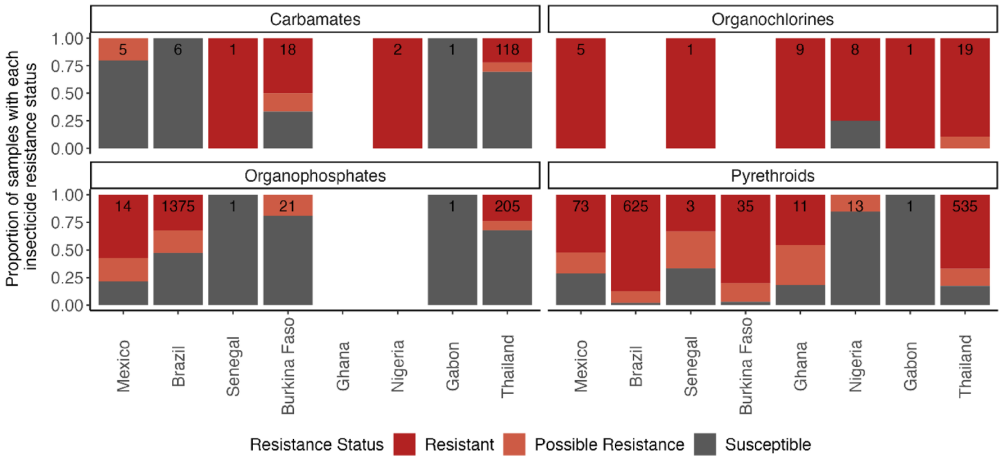


Figure 4. Publicly available phenotype data for *Ae. aegypti* showing the proportion of records that report resistance, possible resistance and susceptibility. Numbers denote total number of records for the insecticide class for that country region[44]. Only data collected on *Aedes aegypti* after 2000 were included for countries that were present in the WGS data set.

is still high in Mexico, followed by Brazil and Thailand (Table 2). These countries only have 1 mutation, G12S, in the *ace* gene common across all of them.

Discussion

We explored the genetic diversity present in four genes (*vgsc, ace-1, rdl* and *GSTe2)* involved in insecticide response across 729 *Ae. aegypti* and 7 *Ae. mascarensis* samples from 15 countries. We identified many known and unreported amino-acid substitutions which may be involved in insecticide resistance. This catalogue of genetic variants is a valuable resource that can be explored to investigate molecular mechanism associated with insecticide resistance together with phenotypic information and used to design diagnostics genetic markers for molecular surveillance.

The populations with greater numbers of amino acid substitutions linked to insecticide resistance were Thailand (RDL: A301S; VGSC: V410L, S989P, V1016G and F1534C) and the USA (RDL A301S; VGSC: V410L, Gly923V, I1011M and F1534C). In Africa, the substitutions most frequently observed were RDL A301S and

VGSC V410L and F1534C, but many countries had none of the reported mutations. We have also observed that VGSC V410L and S723T co-occur in all but one sample. None of the Thai samples had any mutations in the *GSTe2* gene, despite having adequate read coverage. In other countries, we detected two common mutations in GSTe2 (C115F/S and L111S/F) in the DDT binding site. The C115F and C115S mutations were most frequent in Kenya (n = 142, n = 20), the USA (n = 114, n = 20) and Senegal (n = 82, n = 35). Previous work involving DDT docking with *An. gambiae* GSTe2 has suggested that one of the DDT's planar p-chlorophenyl rings can fit into a sub-pocket, but the other ring faces spatial hindrance from M111 and F115 in the side chains[69]. In *An. gambiae*, the M111S substitution would require two nucleotide changes in contrast to one required for L111S/F in *Ae. aegypti*. To our knowledge, there are no reports of *An. gambiae* M111S or F115C/S; although the latter substitution requires a single amino acid change. These two substitutions were detected in almost all countries in this *Aedes* dataset.

We found only two mutations on the surface of the ACE1 pocket directly involved in hydrolysis (A81S, n = 5; D85H, n = 2)[13]. Since we did not have phenotype data, we cannot determine if these mutations are associated with resistance, but their low prevalence would appear at odds with much higher rate and multiple instances of emergence of G119S in *An. gambiae*[20]. Nevertheless, further functional work can contribute to elucidating the involvement of these mutations in resistance phenotypes.

We have also explored the possibility of gene duplications, and detected such variants in *GSTe2* in USA, Mexico, Brazil, and Thailand, which are of interest due to the high rates of permethrin resistance reported in the Americas and Asia[78,79]. We found no duplications in west and central Africa or Eastern Kenya and South Africa regions[6], but bioassay data in these regions is lacking. The possible duplication of the gene encoding VGSC is more puzzling. Previous research in *D. melanogaster* found that individuals lacking VGSC are not viable, but in contrast those with a single functioning gene copy are healthy apart from increased temperature sensitivity[80]. However, DDT and pyrethroids both prolong the open state of VGSC, so the extra gene copy is unlikely to induce resistance through increased number of pores[14]. Experimental work is required to explain the functional role of the extra copy and determine if it is associated with increased insecticide resistance. Long-read sequencing can help to validate the duplications detected and the differences between the *vgsc* sequences.

The inferred population structure was broadly consistent with previous research based on chromosomal loci. We even identified the two previously described distinct subpopulations of *Ae. aegypti* in Rabai District of Kenya[60], but we also observed inconsistency between the structure we inferred from 15 non-resistance genes and 4 resistance genes (Fig. 1A,B). This inconsistency is very clear in case of VGSC where the same four mutations were present in 18/34 samples from Burkina Faso, 133/160 from USA and 8/15 from Mexico. While these could have arisen independently, single emergence and introduction elsewhere appears more parsimonious especially since these samples also share synonymous mutations. Such separate introductions of Ae. aegypti have been examined in the past[38,62]. However, the result may also be artefact of our methodology. The clustering methods we used have two shortcomings. First, they don't have a measure of confidence; second, the relative distances between clusters and spread of points in cluster are usually not meaningful[53]. As a consequence, it's impossible to infer diversity of population within a cluster, nor to determine relatedness between clusters.

An important observation for future research is that the *cox1* gene and other mitochondrial loci may be problematic for population studies in *Ae. aegypti* because of the unknown number of *cox1* copies per genome[65,66]. This is the result of unknown numbers of mitochondria per cell, unknown number of mitochondrial DNA copies on chromosomes, and unknown allelic diversity of all these *cox1* sequences.

While we focused on exploring the genetic diversity in four genes associated with target site insecticide resistance, there are many loci that could have an important role, particularly in metabolic resistance. Multiple P450 genes, particularly members of the CYP6 and CYP9 subfamilies, have been associated with resistance by overexpression when comparing insecticide-resistant to susceptible strains[81–83].

Having both phenotypic and genotypic data is fundamental for the full understanding of the link between phenotypic resistance and genetic mutations, as well as cross resistance mechanisms. Unfortunately, we did not have phenotypic data for all the countries with genotypic data in this study. We strongly advocate that where possible, phenotypic data be generated for samples with genomic sequences.

Further work on exploring genetic diversity in these gene families, particularly using long-read sequencing to support assembly and correct assignment of copy numbers to each individual gene, may reveal important molecular markers that can be involved in insecticide resistance. Genomic studies, like ours, can provide guidance to functional studies and inform the design of genotyping assays for large scale surveillance of insecticide resistance.

Data availability

All data in publicly available. Analysis scripts are available at https://github.com/AntonS-bio/resistance-Aedes Aegypti.

Received: 5 March 2024; Accepted: 4 June 2024
Published online: 11 June 2024

References

1. Lwande, O. W. *et al. Globe-Trotting Aedes aegypti and Aedes albopictus: Risk Factors for Arbovirus Pandemics* 71–81 (Mary Ann Liebert Inc., 2020).
2. Kraemer, M. U. G. *et al.* The global distribution of the arbovirus vectors Aedes aegypti and Ae. Albopictus. *Elife* **1**, 1 (2015).
3. Brown, J. E. *et al.* Human impacts have shaped historical and recent evolution in Aedes aegypti, the dengue and yellow fever mosquito. *Evolution.* **68**(2), 514–525 (2014).
4. Rocklöv, J. & Dubrow, R. Climate change: An enduring challenge for vector-borne disease prevention and control. *Nat. Immunol.* **21**(5), 479–483 (2020).

5. Liu, N. Insecticide resistance in mosquitoes: Impact. *Mech. Res. Direct.* **60**, 537–559. https://doi.org/10.1146/annurev-ento-010814-020828 (2015).
6. Moyes, C. L. *et al.* Contemporary status of insecticide resistance in the major Aedes vectors of arboviruses infecting humans. *PLOS Neglect. Trop. Dis.* **11**(7), e0005625 (2017).
7. Ingham, V. A. *et al.* A sensory appendage protein protects malaria vectors from pyrethroids. *Nature.* **577**(7790), 376–380 (2020).
8. Ingham, V. A., Wagstaff, S. & Ranson, H. Transcriptomic meta-signatures identified in Anopheles gambiae populations reveal previously undetected insecticide resistance mechanisms. *Nat. Commun.* **9**(1), 1 (2018).
9. 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**, 1 (2021).
10. Balabanidou, V. *et al.* Cytochrome P450 associated with insecticide resistance catalyzes cuticular hydrocarbon production in Anopheles gambiae. *Proc. Natl. Acad. Sci. USA.* **113**(33), 9268–9273 (2016).
11. Du, Y., Nomura, Y., Zhorov, B. S. & Dong, K. Sodium channel mutations and pyrethroid resistance in Aedes aegypti. *Insects.* **7**(4), 1 (2016).
12. Pavlidi, N., Vontas, J. & Van Leeuwen, T. The role of glutathione S-transferases (GSTs) in insecticide resistance in crop pests and disease vectors. *Curr. Opin. Insect Sci.* **27**, 97–102 (2018).
13. Cheung, J., Mahmood, A., Kalathur, R., Liu, L. & Carlier, P. R. Structure of the G119S mutant acetylcholinesterase of the malaria vector Anopheles gambiae reveals basis of insecticide resistance. *Structure.* **26**(1), 130 (2018).
14. Field, L. M., Emyr Davies, T. G., O'Reilly, A. O., Williamson, M. S. & Wallace, B. A. Voltage-gated sodium channels as targets for pyrethroid insecticides. *Eur. Biophys. J.* **46**(7), 675 (2017).
15. Scott, J. G. *Life and death at the voltage-sensitive sodium channel: Evolution in response to insecticide use.* **64**, 243–257. https://doi.org/10.1146/annurev-ento-011118-112420 (2019).
16. Kasai, S. *et al.* Discovery of super–insecticide-resistant dengue mosquitoes in Asia: Threats of concomitant knockdown resistance mutations. *Sci. Adv.* **8**(51), 1 (2022).
17. Brengues, C. *et al.* Pyrethroid and DDT cross-resistance in Aedes aegypti is correlated with novel mutations in the voltage-gated sodium channel gene. *Med. Vet. Entomol.* **17**(1), 87–94 (2003).
18. Chung, H. H. *et al.* Voltage-gated sodium channel intron polymorphism and four mutations comprise six haplotypes in an Aedes aegypti population in Taiwan. *PLoS Neglect. Trop. Dis.* **13**(3), 1 (2019).
19. Fan, Y. *et al.* Evidence for both sequential mutations and recombination in the evolution of kdr alleles in Aedes aegypti. *PLoS Neglect. Trop. Dis.* **14**(4), 1–22 (2020).
20. Weill, M. *et al.* Insecticide resistance in mosquito vectors. *Nature* **423**(6936), 136–137 (2003).
21. Feyereisen, R., Dermauw, W. & Van Leeuwen, T. Genotype to phenotype, the molecular and physiological dimensions of resistance in arthropods. *Pesticide Biochem. Physiol.* **121**, 61–77 (2015).
22. Engdahl, C. *et al.* Acetylcholinesterases from the disease vectors Aedes aegypti and Anopheles gambiae: Functional characterization and comparisons with vertebrate orthologues. *PLOS ONE.* **10**(10), e0138598 (2015).
23. Bloomquist, J. R. Toxicology, mode of action and target site-mediated resistance to insecticides acting on chloride channels. *Compar. Biochem. Physiol. Part C Pharmacol. Toxicol. Endocrinol.* **106**(2), 301–314 (1993).
24. Ffrench-Constant, R. H., Williamson, M. S., Davies, T. G. E. & Bass, C. Ion channels as insecticide targets. *J. Neurogenet.* **30**(3–4), 163–177. https://doi.org/10.1080/01677063201612297 81 (2016).
25. Fonseca-González, I., Quiñones, M. L., Lenhart, A. & Brogdon, W. G. Insecticide resistance status of Aedes aegypti (L.) from Colombia. *Pest Manag. Sci.* **67**(4), 430–437 (2011).
26. Goindin, D. *et al.* Levels of insecticide resistance to deltamethrin, malathion, and temephos, and associated mechanisms in Aedes aegypti mosquitoes from the Guadeloupe and Saint Martin islands (French West Indies). *Infect. Dis. Poverty.* **6**(1), 1 (2017).
27. 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**(10), 2900–2917 (2020).
28. Mintz, J. A. Two cheers for global POPs: A summary and assessment of the Stockholm convention on persistent organic pollutants. *Georgetown Int. Environ. Law Rev.* **14**, 1 (2001).
29. 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**(7), 484–491 (2011).
30. 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**(1), 1 (2017).
31. Lumjuan, N. *et al.* The role of the Aedes aegypti Epsilon glutathione transferases in conferring resistance to DDT and pyrethroid insecticides. *Insect Biochem. Mol. Biol.* **41**(3), 203–209 (2011).
32. Matthews, B. J. *et al.* Improved reference genome of Aedes aegypti informs arbovirus vector control. *Nature.* **563**(7732), 501–507 (2018).
33. Ortelli, F., Rossiter, L. C., Vontas, J., Ranson, H. & Hemingway, J. Heterologous expression of four glutathione transferase genes genetically linked to a major insecticide-resistance locus from the malaria vector Anopheles gambiae. *Biochem. J.* **373**(Pt 3), 957–963 (2003).
34. 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**(2), R27 (2014).
35. Mitchell, S. N. *et al.* Metabolic and target-site mechanisms combine to confer strong DDT resistance in Anopheles gambiae. *PloS One.* **9**(3), 1 (2014).
36. Helvecio, E. *et al.* Polymorphisms in GSTE2 is associated with temephos resistance in Aedes aegypti. *Pesticide Biochem. Physiol.* **165**, 1 (2020).
37. Crava, C. M. *et al.* Population genomics in the arboviral vector Aedes aegypti reveals the genomic architecture and evolution of endogenous viral elements. *Mol. Ecol.* **1**, 1 (2021).
38. Rose, N. H. *et al.* Climate and urbanization drive mosquito preference for humans. *Curr. Biol.* **1**, 1 (2020).
39. Lee, Y. *et al.* Genome-wide divergence among invasive populations of Aedes aegypti in California. *BMC Genom.* **20**(1), 1–10 (2019).
40. Kelly, E. T. *et al.* Evidence of local extinction and reintroduction of Aedes aegypti in exeter California. *Front. Trop. Dis.* **2**, 703873 (2021).
41. Faucon, F. *et al.* Identifying genomic changes associated with insecticide resistance in the dengue mosquito Aedes aegypti by deep targeted sequencing. *Genome Res.* **25**(9), 1347–1359 (2015).
42. Leong, C. S. *et al.* Enzymatic and molecular characterization of insecticide resistance mechanisms in field populations of Aedes aegypti from Selangor Malaysia. *Parasites Vectors.* **12**(1), 1–17 (2019).
43. Poupardin, R., Srisukontarat, W., Yunta, C. & Ranson, H. Identification of carboxylesterase genes implicated in temephos resistance in the dengue vector Aedes aegypti. *PLoS Neglect. Trop. Dis.* **8**(3), 1 (2014).
44. Knox, T. B. *et al.* An online tool for mapping insecticide resistance in major Anopheles vectors of human malaria parasites and review of resistance status for the Afrotropical region. *Parasites Vectors.* **7**(1), 1–14 (2014).
45. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods.* **9**(4), 357–359 (2012).
46. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics.* **25**(16), 2078–2079 (2009).
47. McKenna, A. *et al.* The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**(9), 1297–1303 (2010).

48. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience.* **10**(2), 1 (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 w1118; iso-2; iso-3. *Fly.* **6**(2), 80–92 (2012).
50. Boratyn, G. M. *et al.* Domain enhanced lookup time accelerated BLAST. *Biol. Direct.* **1**, 1 (2012).
51. McInnes, L., Healy, J., & Melville, J. UMAP: Uniform manifold approximation and projection for dimension reduction. arXiv (2018).
52. Campello, R. J. G. B., Moulavi, D. & Sander, J. Density-based clustering based on hierarchical density estimates. *Lecture Notes Comput. Sci.* **7819**(2), 160–172 (2013).
53. Diaz-Papkovich, A., Anderson-Trocmé, L. & Gravel, S. A review of UMAP in population genetics. *J. Hum. Genet.* **66**(1), 85–91 (2021).
54. Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **37**(1), 38–47 (2018).
55. Bellin, N. *et al.* Unsupervised machine learning and geometric morphometrics as tools for the identification of inter and intraspecific variations in the Anopheles Maculipennis complex. *Acta Tropica.* **233**, 106585 (2022).
56. Evans, R. *et al.* Protein complex prediction with AlphaFold-Multimer. *bioRxiv.* **1**, 1 (2021).
57. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature.* **596**(7873), 583–589 (2021).
58. Berman, H. M. *et al.* The protein data bank. *Nucleic Acids Res.* **28**(1), 235–242 (2000).
59. Tatusova, T. *et al.* NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* **44**(14), 6614–6624 (2016).
60. Gloria-Soria, A. *et al.* Global genetic diversity of Aedes aegypti. *Mol. Ecol.* **25**(21), 5377–5395 (2016).
61. Kotsakiozi, P. *et al.* Population structure of a vector of human diseases: Aedes aegypti in its ancestral range Africa. *Ecol. Evol.* **8**(16), 7835–7848 (2018).
62. Pless, E. *et al.* Multiple introductions of the dengue vector, Aedes aegypti, into California. *PLoS Negl. Trop. Dis.* **11**(8), e0005718 (2017).
63. Soghigian, J. *et al.* Genetic evidence for the origin of Aedes aegypti, the yellow fever mosquito, in the southwestern Indian Ocean. *Mol. Ecol.* **29**(19), 3593–3606 (2020).
64. McBride, C. S. *et al.* Evolution of mosquito preference for humans linked to an odorant receptor. *Nature* **515**(7526), 222–227 (2014).
65. Richly, E. & Leister, D. NUMTs in sequenced eukaryotic genomes. *Mol. Biol. Evol.* **21**(6), 1081–1084 (2004).
66. Black Iv, W. C. & Bernhardt, S. A. Abundant nuclear copies of mitochondrial origin (NUMTs) in the Aedes aegypti genome. *Insect Mol. Biol.* **18**(6), 705–713 (2009).
67. Clarkson, C. S. *et al.* Genome variation and population structure among 1142 mosquitoes of the African malaria vector species Anopheles gambiae and Anopheles coluzzii. *Genome Res.* **30**(10), 1533–1546 (2020).
68. Colletier, J. P. *et al.* Structural insights into substrate traffic and inhibition in acetylcholinesterase. *EMBO J.* **25**(12), 2746–2756 (2006).
69. Wang, Y. *et al.* Structure of an insect epsilon class glutathione S-transferase from the malaria vector Anopheles gambiae provides an explanation for the high DDT-detoxifying activity. *J. Struct. Biol.* **164**(2), 228–235 (2008).
70. Chen, C. *et al.* Marker-assisted mapping enables forward genetic analysis in Aedes aegypti, an arboviral vector with vast recombination deserts. *Genetics.* **222**(3), 1 (2022).
71. Hamid, P. H., Prastowo, J., Widyasari, A., Taubert, A. & Hermosilla, C. Knockdown resistance (kdr) of the voltage-gated sodium channel gene of Aedes aegypti population in Denpasar, Bali, Indonesia. *Parasites Vectors.* **10**(1), 1 (2017).
72. Rasli, R. *et al.* Susceptibility status and resistance mechanisms in permethrin-selected, laboratory susceptible and field-collected Aedes aegypti from Malaysia. *Insects* **9**(2), 43 (2018).
73. Saha, P. *et al.* Prevalence of kdr mutations and insecticide susceptibility among natural population of Aedes aegypti in West Bengal. *PLOS ONE.* **14**(4), e0215541 (2019).
74. Mano, C. *et al.* Protein expression in female salivary glands of pyrethroid-susceptible and resistant strains of Aedes aegypti mosquitoes. *Parasites Vectors.* **12**(1), 1–19 (2019).
75. Yang, F. *et al.* Insecticide resistance status of Aedes aegypti (Diptera: Culicidae) in California by biochemical assays. *J. Med. Entomol.* **57**(4), 1176 (2020).
76. Kandel, Y. *et al.* Widespread insecticide resistance in Aedes aegypti L. from New Mexico, USA. *PLOS ONE.* **14**(2), e0212693 (2019).
77. Hernandez, H. M., Martinez, F. A. & Vitek, C. J. Insecticide resistance in Aedes aegypti Varies seasonally and geographically in Texas/Mexico border cities. *J. Am. Mosq. Control Assoc.* **38**(1), 59–69 (2022).
78. Chuaycharoensuk, T. *et al.* Frequency of pyrethroid resistance in Aedes aegypti and Aedes albopictus (Diptera: Culicidae) in Thailand. *J. Vector Ecol.* **36**(1), 204–212 (2011).
79. Solis-Santoyo, F. *et al.* Insecticide resistance in Aedes aegypti from Tapachula, Mexico: Spatial variation and response to historical insecticide use. *PLoS Neglect. Trop. Dis.* **15**(9), 1 (2021).
80. Ravenscroft, T. A. *et al.* Drosophila voltage-gated sodium channels are only expressed in active neurons and are localized to distal axonal initial segment-like domains. *J. Neurosci.* **40**(42), 7999–8024 (2020).
81. Vontas, J., Katsavou, E. & Mavridis, K. Cytochrome P450-based metabolic insecticide resistance in Anopheles and Aedes mosquito vectors: Muddying the waters. *Pesticide Biochem. Physiol.* **170**, 104666 (2020).
82. 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**(1), 1–11 (2019).
83. 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**(484), 7386 (2019).

Acknowledgements

E.L.C. is funded by an MRC LID PhD studentship TGC and SC are funded by UKRI grants (BBSRC BB/X018156/1; MRC MR/X005895/1; EPSRC EP/Y018842/1).

Author contributions

AS, SC and TC designed the study. AS and EC analysed the data under the supervision of TC and SC. All authors interpreted the results. AS and EC wrote the first draft of the manuscript. All authors have edited and approved the final version of the 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/s41598-024-64007-6.

Supplementary Information

Supplementary 1 Table. *Sample metadata and median coverage for VSSC, Ace1, GSTe2 and rdl* and coordinates for **Figure 1**.

Please find this table at link:
https://www.nature.com/articles/s41598-024-64007-6#Sec19

Supplementary 2 Table. *Identified sources of resistance phenotype data for some samples in the study*

Samples	Country of Origin	Insecticide Status	Reference
Bora-Bora	French Polynesia	Susceptible	https://parasitesandvectors.biomedcentral.com/articles/10.1186/s13071-019-3472-1
Nakhon Sawan (Nakh-R)	Thailand	Resistant to deltamethrin and temephos	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3961196/ https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4561493/
Thai-Mex (transgenic line)	Thailand x Mexico	Unknown	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9630976/
BRA-PAK (transgenic line)	Brazil x Pakistan	Unknown	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9630976/ https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9009520/
Liverpool-RED recombinant	West Africa	Unknown	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9630976/
Liverpool	West Africa	Reference (Susceptible)	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9630976/ https://www.tandfonline.com/doi/abs/10.1080/00034983.1962.11686134 https://www.science.org/doi/10.1126/sciadv.abq7345

Supplementary 3 Table. *Phenotype data on insecticide resistance based on IR Mapper*

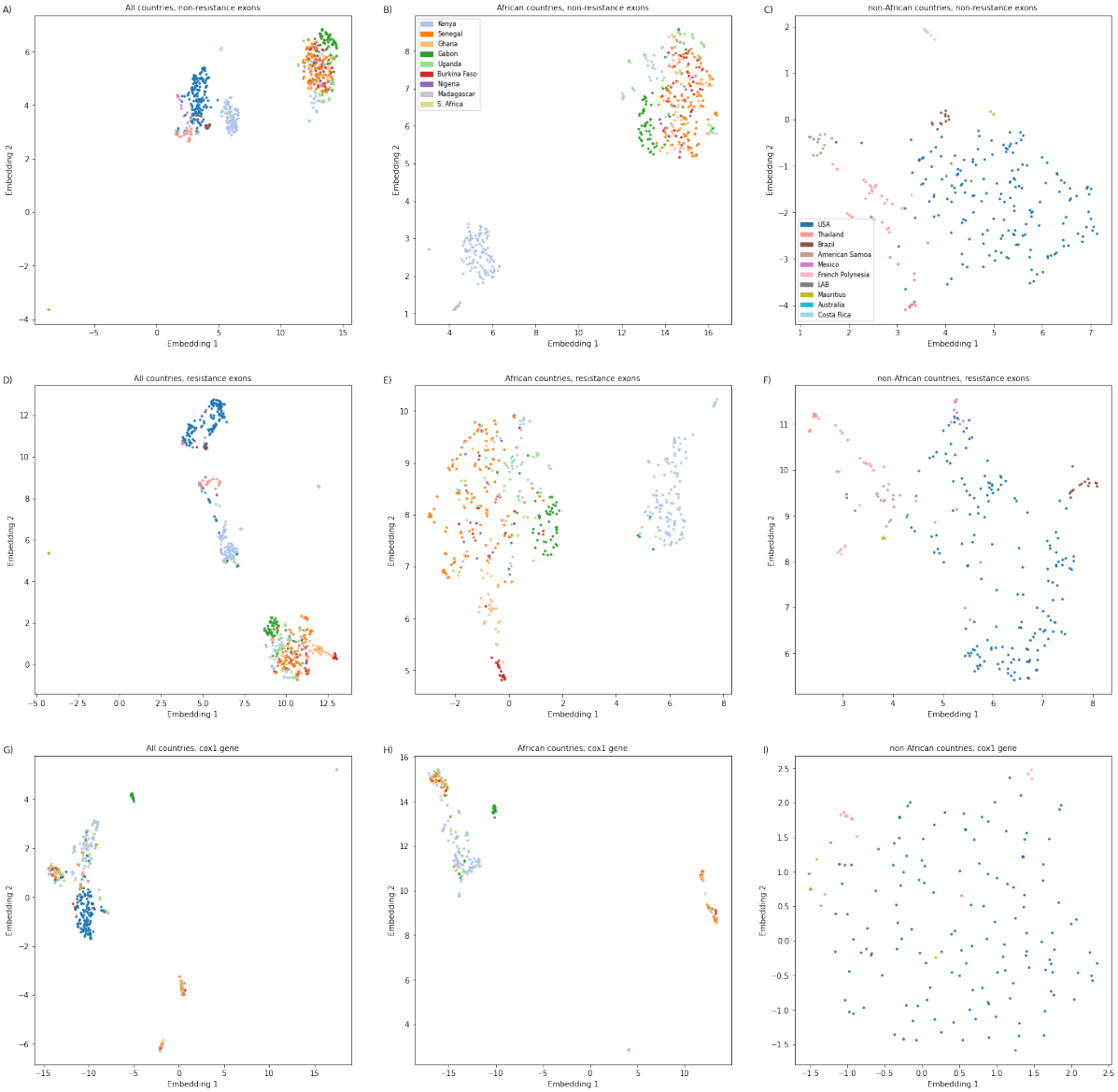
Please find this table at this link:
https://www.nature.com/articles/s41598-024-64007-6#Sec19

Supplementary 4 Table. *All missense mutations identified in samples*

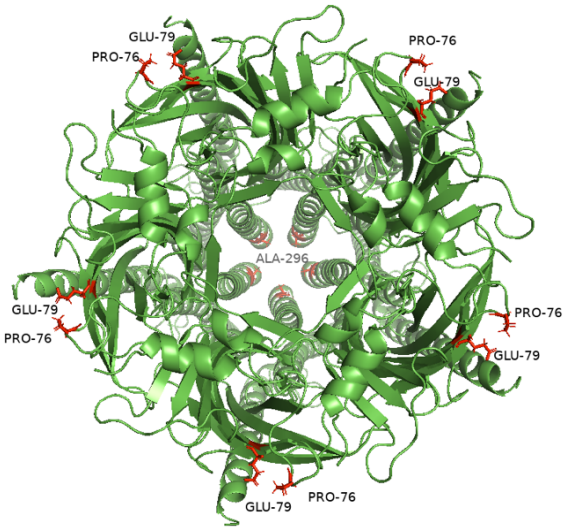
Please find this table at this link:
https://www.nature.com/articles/s41598-024-64007-6#Sec19

Supplementary 5 Table. *Pairwise linkage disequilibrium (LD) of non-synonymous SNPs in vgsc, ace-1, GSTe2 and rdl with an R2 value above 0.5*

Chrom	Pos1	Pos2	R2	D	D prime	Country	Csq1	Csq2	Gene1	Gene2
3	315984009	315998458	1	0.11	1	American Samoa	Ile1011Met	Gly923Val	*vgsc*	*vgsc*
3	316014588	316080722	1	0.23	1	Burkina Faso	Ser723Thr	Val410Leu	*vgsc*	*vgsc*
2	351633680	351634725	0.77	0.21	0.88	Burkina Faso	Leu9Ile	Ile169Ser	*GSTe2*	*GSTe2*
2	351633680	351634725	0.7	0.04	0.88	East Kenya	Leu9Ile	Ile169Ser	*GSTe2*	*GSTe2*
2	351634667	351634812	0.67	0.09	0.83	East Kenya	Ile150Val	Ala198Glu	*GSTe2*	*GSTe2*
2	351633680	351634725	1	0.03	1	East Kenya Outlier	Leu9Ile	Ile169Ser	*GSTe2*	*GSTe2*
2	351634667	351634812	1	0.24	1	East Kenya Outlier	Ile150Val	Ala198Glu	*GSTe2*	*GSTe2*
2	351634667	351634812	1	0.1	1	French Polynesia	Ile150Val	Ala198Glu	*GSTe2*	*GSTe2*
2	41755106	41847790	0.56	0.03	1	Gabon	Glu84Asp	Ala301Ser	*GABA*	*GABA*
2	351633680	351634725	0.79	0.16	1	Gabon	Leu9Ile	Ile169Ser	*GSTe2*	*GSTe2*
2	351633796	351634739	0.65	0.03	1	Gabon	Met47Ile	Pro174Ser	*GSTe2*	*GSTe2*
3	161500299	161504294	1	0.01	1	Ghana	Gln306Leu	Asp4Glu	*ace-1*	*ace-1*
2	351633680	351634725	0.72	0.1	0.92	Ghana	Leu9Ile	Ile169Ser	*GSTe2*	*GSTe2*
3	315984077	316014588	1	0.08	1	LAB	Ser989Pro	Ser723Thr	*vgsc*	*vgsc*
3	315938859	316000838	1	0.25	1	Madagascar	Asp1633Glu	Ala797Ser	*vgsc*	*vgsc*
2	351634602	351634667	1	-0.14	-1	Mauritius	Ala128Val	Ile150Val	*GSTe2*	*GSTe2*
2	351634602	351634812	1	-0.14	-1	Mauritius	Ala128Val	Ala198Glu	*GSTe2*	*GSTe2*
2	351634667	351634812	1	0.14	1	Mauritius	Ile150Val	Ala198Glu	*GSTe2*	*GSTe2*
3	316014588	316080722	1	0.21	1	Mexico	Ser723Thr	Val410Leu	*vgsc*	*vgsc*
3	315984077	315998389	1	0.03	1	Nigeria	Ser989Pro	Leu946Gln	*vgsc*	*vgsc*
3	316014588	316080722	1	0.03	1	Nigeria	Ser723Thr	Val410Leu	*vgsc*	*vgsc*
3	161499835	161500277	0.64	0.06	1	Thailand	Glu460Gln	Arg313Ser	*ace-1*	*ace-1*
3	315983762	315984077	0.75	0.16	1	Thailand	Val1016Gly	Ser989Pro	*vgsc*	*vgsc*
2	351633680	351634725	0.63	0.13	1	Uganda	Leu9Ile	Ile169Ser	*GSTe2*	*GSTe2*
3	316014588	316080722	0.87	0.17	0.95	USA	Ser723Thr	Val410Leu	*vgsc*	*vgsc*
2	41630424	351634748	0.66	0.03	1	West Kenya	Ser20Thr	Lys177Glu	*GABA*	*GSTe2*
2	351633680	351634725	0.57	0.08	0.86	West Kenya	Leu9Ile	Ile169Ser	*GSTe2*	*GSTe2*
2	351633796	351634587	0.59	0.14	1	West Kenya	Met47Ile	Gly123Glu	*GSTe2*	*GSTe2*

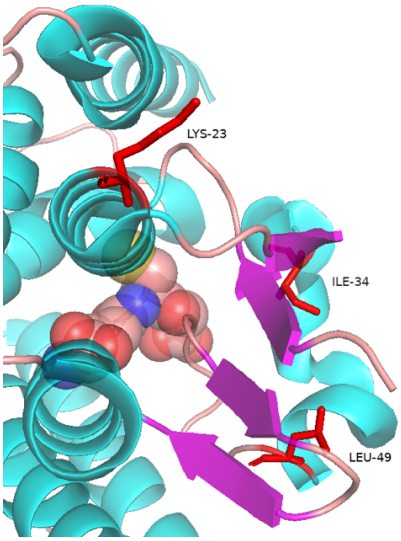


Supplementary 1 Figure. *Population structure using UMAP embedding of SNPs for different geographical regions.*

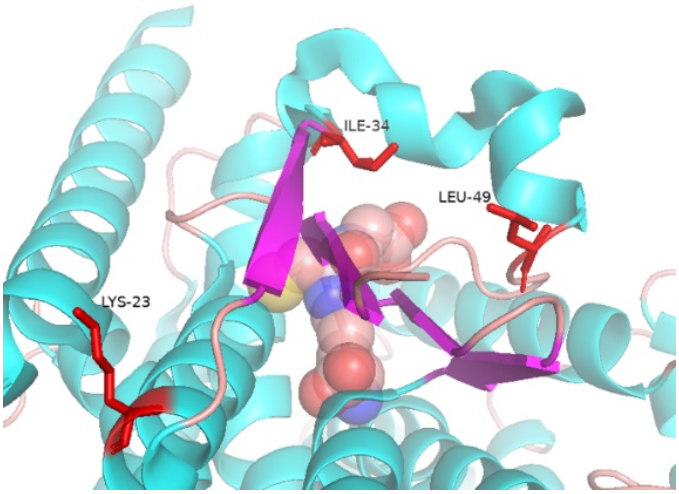


Supplementary 2 Figure. *GABA receptor protein structure including mutations found in >10 isolates.*

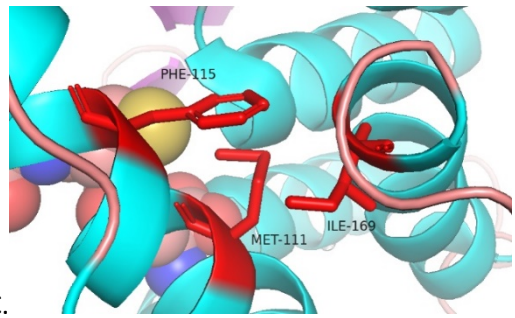
A.



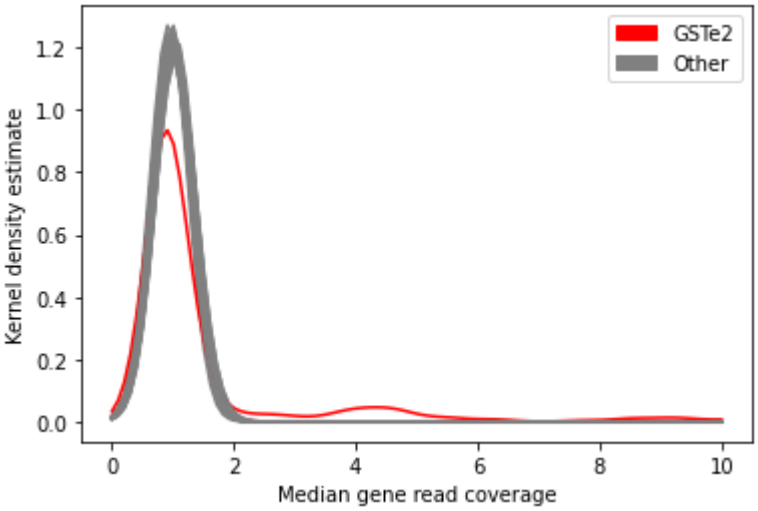
B.



C.



Supplementary 3 Figure. *GSTe2 mutations specific to East Kenya and South Africa (**A,B**) and common substitutions (Cys115Phe/Ser and Leu111Ser) together with west and central Africa specific Ile169Ser substitution. The residue at position 111 is methionine because we used PDB 2IMI structure of An. gambiae to show accurate ligand docking* [136].



Supplementary 4 Figure. *Median per-base read coverage across samples for GSTe2 and other genes. The coverage was normalised for each sample using median coverage across the genes for that sample. Two peaks are visible in GSTe2 at 4 and 9 median gene read coverage.*

Addendum

Duplications of genes can be interesting in the field of insecticide resistance as they may act as a compensatory mechanism to aid resistance to insecticides, usually by increasing the amount of gene product produced. This has been described in *rdl* in *Drosophila melanogaster,* as well as ace-1 duplication in *Culex quinquefasciatus* and *Anopheles gambiae.* Subsequently, it was noted that there are four genes within the Aedes aegypti genome (AaegL5.0, GCF_002204515.2) are annotated as 'acetylcholinesterase'; LOC5578456, LOC5575867, LOC5574466, LOC5570776 **(Table 1)**. This was investigated further to assess if any of these other genes were functional copies.

It is described that LOC5578456 is the main functional *ace-1* gene. LOC5575867 has a 90.5% nucleotide identity and 38.4% protein identity (80% query coverage) with LOC5578456. While LOC5574466 has 73.4% nucleotide identity and 31.9% protein identity (68% query coverage) with LOC5578456. Finally, LOC5570776 has no significant nucleotide similarity with 26.3% protein identity (73% query coverage).

Table 1. *Genes in the Aedes aegypti genome (AaegL5.0, GCF_002204515.2) annotated as acetylcholinesterase*

Chromosome: Position	Gene	Alias	Length
3:161486025-161871375	LOC5578456	AAEL000511	385,350
1:142163173-42283642	LOC5575867	AAEL012141, AAEL8000045	120,469
2:214207432-214234809	LOC5574466	AAEL002376	27,377
3:29506807-29509153	LOC5570776	AAEL008532	2,346

Given the limited protein identity between these genes, it is unlikely that they produce functional acetylcholinesterase enzyme. Given the differences in identity, they may be incorrectly annotated, or they could represent historic duplications that have since diverged significantly from the original sequence. Phylogenetic analysis reveals that the LOC5578456 gene is most similar to the AChE in *Anopheles gambiae,* where the mutation G119S we are investigating has been described, while the other genes are more distinct (**Figure 1**). The LOC5575867 gene is most like that of the AChE in *Drosophila melanogaster.*

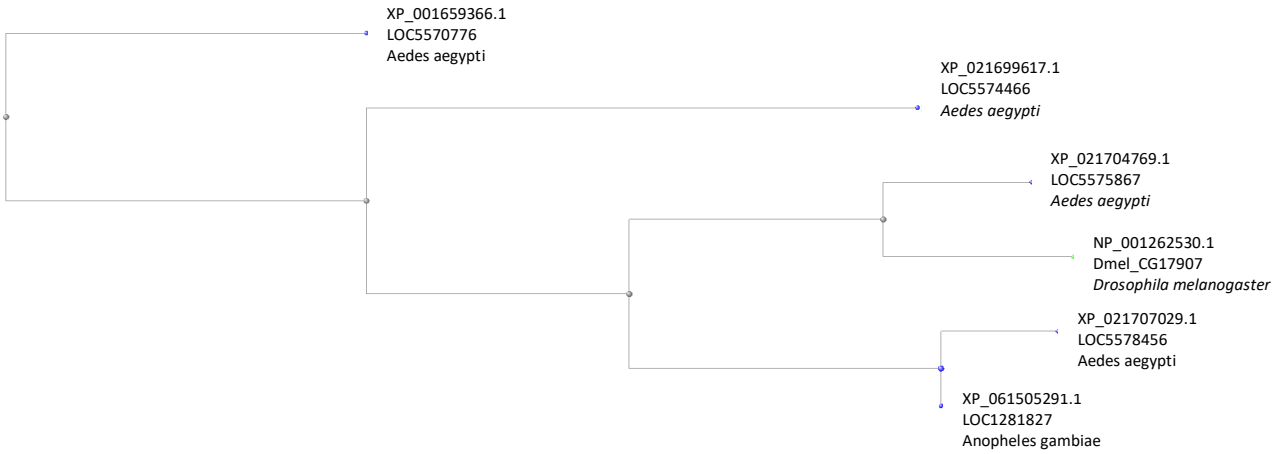


Figure 1. *Phylogenetic tree comparing the protein sequences of the four acetylcholinesterase (AChE) annotated genes in Ae. aegypti, the AChE from An. gambiae and AChE from Drosophila melanogaster.*

CHAPTER FOUR

Genome-wide population genetics and molecular surveillance of insecticide resistance in *Aedes aegypti* mosquitoes from Puerto Rico

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	1513859	**Title**	Ms
First Name(s)	Emma Louise		
Surname/Family Name	Collins		
Thesis Title	Utilising genomic approaches to explore genetic diversity and insecticide resistance in Aedes aegypti populations		
Primary Supervisor	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?*		Was the work subject to academic peer review?	

*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 Neglected Tropical Disease
Please list the paper's authors in the intended authorship order:	Emma L. Collins, Sophie Moss, Joanelis Medina Quintana, Reynaldo Morales, Joseph Thorpe, Holly Acford-Palmer, Taane G. Clark, Grayson Brown, 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)	EC designed the study and collected samples with help from JMQ and RM. EC analysed the data with assistance from SM, JT, HAP. EC wrote the first draft of the manuscript. All authors have edited and approved the final version of the manuscript.

SECTION E

Student Signature	Emma Collins
Date	18/06/2024

Supervisor Signature	Susana Campino
Date	18/06/2024

Genome-wide population genetics and molecular surveillance of insecticide resistance in *Aedes aegypti* mosquitoes from Puerto Rico

Emma L. Collins[1], Sophie Moss[1], Joanelis Medina Quintana[2], Reynaldo Morales[2], Joseph Thorpe [1], Holly Acford-Palmer[1], Taane G. Clark[1,3], Grayson Brown[2], Susana Campino[1]

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

[2] Puerto Rico Vector Control Unit, Puerto Rico, USA

[3] 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

Aedes aegypti is the predominant vector of many arboviruses, including dengue, Zika and Chikungunya. Effective control of these mosquitoes is paramount for human health in endemic regions and to prevent outbreaks; however, there has been a significant increase in resistance to commonly employed insecticides, which poses a considerable challenge to vector control efforts. Genomic analysis can provide valuable insights into the mechanisms underlying insecticide resistance, prevalence of mutations of concern, and population dynamics. In this study, we generated whole genome sequence data for 33 *Ae. aegypti* samples collected from six regions across Puerto Rico, where high levels of insecticide resistance have been reported. The data was analysed alongside WGS data from 215 *Ae. aegypti* samples from seven other countries: Burkina Faso, Uganda, Kenya, Thailand, USA, Mexico, and Brazil. Genome-wide population analyses revealed that the Puerto Rican samples clustered closely with *Ae. aegypti* from Mexico and, together with other American and Thai samples, they were genetically distinct from the African populations. Across all 248 samples, we identified 281,889 missense single nucleotide polymorphisms. Five known insecticide resistance-associated mutations were detected (*vgsc:* V410L, V1016I, V1016G, F1534C; *rdl*: A301S) with particularly high frequencies in Puerto Rico (0.909, 0.909, 0.06, 0.924, 0.773, respectively). Notably, the V1016G mutation, previously only reported in Asian populations, was detected in Puerto Rico, marking the first report of this variant in the Americas. In contrast, these mutations were absent in Ugandan or Kenyan samples, likely due to differing vector control strategies. Additionally, several other non-synonymous SNPs were identified in candidate genes potentially associated with insecticide resistance. Signatures of selection were evident in genes crucial for nerve action, such as the gamma-aminobutyric acid (GABA) receptor subunit alpha. Furthermore, selection signals were evident in genes from various enzyme families linked to resistance, such as glutathione S-transferases and cytochrome P450s. In summary, our research highlights a diverse array of genetic polymorphisms, identifies regions under selection pressure, and offers new genomic insights into the *Ae. aegypti* population in Puerto Rico.

Word count: 326

Introduction

The *Aedes aegypti* mosquito is a proficient vector of numerous arboviral diseases, including dengue, Chikungunya and Zika viruses, with its competence exacerbated by its anthropophilic nature and adaptability to live and breed in urban areas [1]. It is believed that urbanised *Ae. aegypti* originated from a forest-breeding population in sub-Saharan Africa, known as *Ae. aegypti formosus* [2]. The global spread of this species likely occurred via ships traveling to the New World, with genetic data indicating a subsequent westward expansion to Asia and Australia[3]. The introduction of this vector in Asia may have occurred as late as the 19th century, coinciding with the first recorded cases of dengue in the region [4]. Currently, *Ae. aegypti* is widespread in tropical, subtropical areas and some temperate regions, and its distribution continues to expand, driven by factors such as climate change and globalization [5].

With the increasing interactions between *Ae. aegypti* and humans, the threat of disease transmission by these vectors has significantly risen [5–7]. Insecticides have been crucial for vector control, substantially contributing to the reduction of mosquito-borne diseases [8]. For the *Aedes* genus, primary control methods include habitat modification, space spraying and larviciding. However, the extensive use of insecticides has applied significant selective pressure on this vector, leading to the widespread emergence of insecticide resistance in *Ae. aegypti* populations worldwide. According to records from the Insecticide Resistance Mapper (IR Mapper; https://www.irmapper.com/), nearly 3,000 reports of resistance in *Ae. aegypti* have been documented across 40 countries since 2000 [9]. These accounts, based on larval or adult bioassays, indicate resistance to multiple classes of insecticides, including pyrethroids, carbamates, organophosphates, and organochlorines. [10].

The four main mechanisms of insecticide resistance in mosquitoes are target site resistance, metabolic resistance, cuticular alterations, and behavioural changes. Target site resistance results from single nucleotide polymorphisms (SNPs) that reduce insecticide binding efficacy to the target protein [11]. This mechanism of resistance is the most extensively studied, and numerous SNPs have been found across four key genes: voltage-gated sodium channel (*vgsc*), resistance to dieldrin (*rdl*), acetylcholinesterase 1 (*ace-1)* and glutathione-S-transferase 2 (*GSTe2*) [12–14]**.** Mutations in these genes have different effects on different insecticides. Specifically, mutations in *vgsc* are linked to pyrethroid resistance, mutations in *rdl* confer resistance to organochlorines (cyclodienes), and mutations in *ace-1* are associated with resistance to organophosphates and carbamates [14]. *GSTe2* is a metabolic enzyme, and although it is

typically involved in metabolic resistance, a mutation (L119F) in this gene has been identified that alters the DDT binding site, reducing the insecticide's effectiveness[15].

Many knockdown resistance (*kdr*) mutations in the *vgsc* gene have been identified in *Ae. aegypti* to contribute to pyrethroid resistance [14,16]. Metabolic resistance has been associated with increased expression of genes involved in insecticide detoxification, including cytochrome P450 monooxygenases (P450s/CYPs), carboxyl/cholinesterase's (CCEs), glutathione S-transferases (GSTs) and UDP-glycosyl-transferases (UDPGTs) [17,18]. In *Ae. aegypti*, metabolic resistance plays a significant role in insecticide resistance, although the genes identified so far vary between populations [19,20]. Structural variants (SVs) have been identified to contribute to metabolic resistance in *Anopheles species* [21,22]. However, in *Ae. aegypti, t*he characterization of SV has been limited, particularly due to this species' large genome (size: ~1.3Gb) and the presence of numerous repetitive regions [23].

Advances in sequencing technologies, coupled with reduced implementation costs, have made WGS an increasingly valuable tool for investigating genetic diversity, including mutations involved in insecticide resistance, across multiple vector populations worldwide [16,24]. There are likely many uncharacterized genes involved in the development of resistance, and a genome-wide approach can aid in the detection of new, previously unreported variants.

In this study, we generated and analysed WGS data for Ae*. aegypti* mosquitoes *(*n = 33) from Puerto Rico, an island in the Caribbean where high levels of insecticide resistance have been reported [25–27]. *Ae. aegypti* mosquitoes are endemic to Puerto Rico, where they transmit dengue, Zika, and Chikungunya viruses, which have caused several outbreaks in recent years [25,28,29]. Control measures in the region included larval management, larviciding, and ultra-low volume spaying [30].

We explored WGS data of Puerto Rican *Aedes aegypti* samples alongside WGS data available from seven other global *Ae. aegypti* populations (n = 215). We investigated the presence of resistance-associated mutations, high-frequency circulating mutations, gene selection, and population dynamics in Puerto Rico while contextualised this population within the broader genomic landscape of global *Ae. aegypti* [28].

Results

Whole Genome Sequence Data and Nucleotide Diversity

The genomes of 33 Puerto Rican samples collected across six sites **(Figure 1A)** were analysed alongside 215 samples (total n = 248) from seven other countries across the globe (Uganda, Kenya, Burkina Faso, USA, Mexico, Brazil, Thailand) (**Figure 1B**). Among the 33 newly sequenced Puerto Rican samples, the average genome-wide coverage was 15.8-fold, with a mean mapping to the reference genome of 97.0%. The average coverage for the 215 publicly available samples was 9.6-fold, varying between countries from 6.5-fold (USA) to 17.5-fold (Mexico).

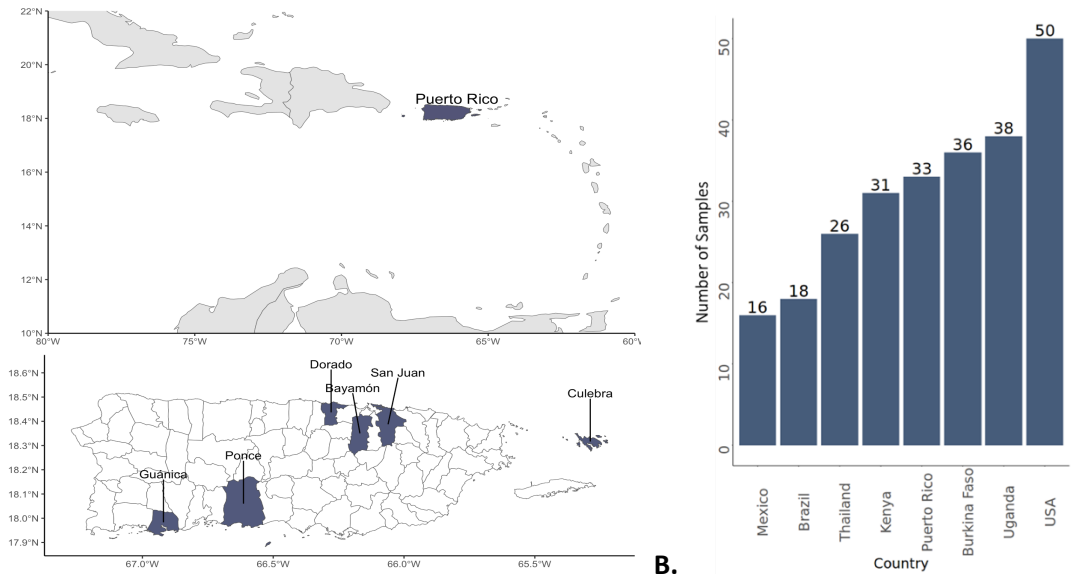
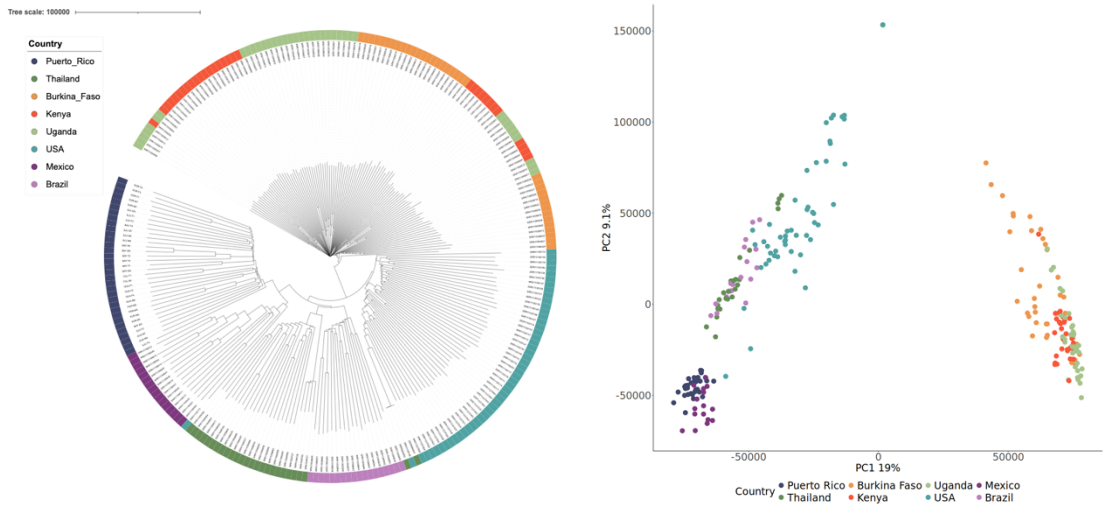


Figure 1. A. *Map of Puerto Rico's location within the Caribbean and the sampling regions across the island.* ***B.*** *Number of sample sequences from each population.*

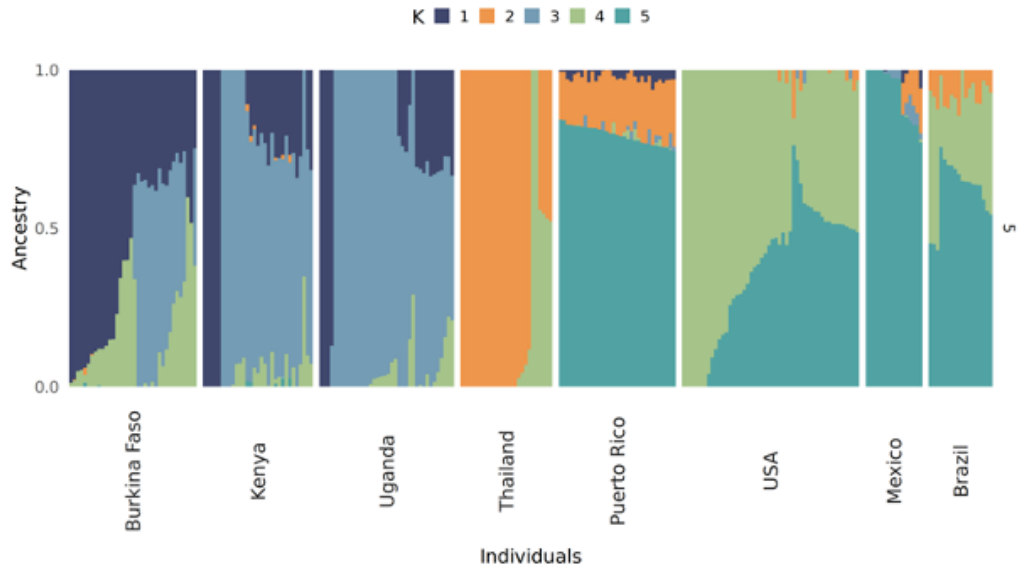
Mean nucleotide diversity (π), calculated in 100kb windows, was 0.000223 across all chromosomes in the Puerto Rican population, with chromosome 1 being the most diverse (π = 0.00024). The average nucleotide diversity (π) of exon regions in the Puerto Rican population was greater than that of all other populations, followed by isolates from Mexico (π = 0.000216), USA (π = 0.000204), Thailand (π = 0.000193), and Brazil (π = 0.000191). African isolates exhibited the lowest levels of genome-wide nucleotide diversity (Burkina Faso; π = 0.000175, Kenya; π = 0.000173, Uganda; π = 0.000172). Overall, patterns in nucleotide diversity across chromosomes was consistent across all populations, with centromeric regions exhibiting the lowest diversity (**Figure S1**).

Population Differentiation and Ancestral Analysis

Using 2,011,616 high-quality SNPs identified across exon regions (total samples = 248), a neighbour joining (NJ) tree was constructed (**Figure 2A**). Three distinct population clusters were observed: the Puerto Rican samples clustered with those from Mexico, forming a group closely related to populations from Brazil, Thailand, and USA, while the African samples (Burkina Faso, Uganda, and Kenya) formed a distinct cluster. To further investigate the African populations, an additional phylogenetic tree was constructed, revealing minimal structure between these countries (**Figure S2).** The corresponding principal component analysis (PCA) confirmed observations from the NJ tree; PC1 separated the larger grouping of the Americas (including Puerto Rico and Mexico) and Thailand from the African cluster (19% variation explained). Within each cluster, evidence of separation was observed at the country level (PC2: 9% variation explained) (**Figure 2B, Figure S3).**



A. **B.**



C.

Figure 2. A*. Phylogenetic tree showing genetic relatedness of different populations.* ***B****. Principal Component Analysis (PCA) plot showing distinct populations* ***C.*** *Admixture analysis showing ancestral relationships.*

Ancestral relationships were inferred using an admixture analysis, which identified five ancestral populations (K1-K5) in line with the geographical origin of samples **(Figure 2C, Figure S4)**. The African samples show high similarity with K1, K3, and K4 ancestries present. K3 predominated in Uganda and Kenya, while K1 was more common in Burkina Faso. The K4 ancestral population was dominant in the USA and was also present in Brazil and Thailand. All American samples shared K5 ancestry, which was most prevalent in Puerto Rico, Mexico, and Brazil. Thai isolates were the most distinct, with a dominance of the K2 ancestry. Small proportions of K2 ancestry were observed in the Puerto Rican, USA, Mexican and Brazilian populations.

Pairwise F_{ST} analysis was performed to investigate population differentiation by comparing variance in allele frequencies between populations. The F_{ST} results corresponded with the PCA, admixture, and NJ tree analysis, showing that Puerto Rico was most genetically similar to Mexico, USA, and Brazilian populations, as indicated by the low number of SNP sites with $F_{ST} > 0.8$ (**Figure S5)**. This finding is in line with the principles of isolation by distance. Kenyan and Ugandan populations were most similar, with no sites with F_{ST} values > 0.8, whereas Ugandan and Mexican populations were most distinct, with 13,858 sites with values > 0.8 and 338 sites with a perfect differentiation (F_{ST} =1).

As anticipated, countries within the same region had a lower proportion of sites with an F_{ST} above 0.8 compared to countries from different regions, aligning with PCA results **(Figure S6).** When examining insecticide resistance genes, the top 20 positions within the *vgsc* gene had an F_{ST} ranging between 0.932 and 0.971. One of the highest F_{ST} values was observed for position 315939224 (LOC5567355) in chromosome 3, where a SNP associated with resistance to pyrethroids (F1534C) was observed in several population comparisons. This SNP was particularly significant in comparisons between Kenya vs Mexico (F_{ST} = 0.932), Mexico vs Uganda (F_{ST} = 0.940), Puerto Rico vs Uganda (F_{ST} = 0.932), and Puerto Rico vs. Kenya (F_{ST} = 0.920). This mutation was not identified in the Kenyan or Ugandan populations.

Population Differentiation within Puerto Rico

A neighbour-joining tree of SNPs showed minimal genomic differences across the six sites where samples were collected in Puerto Rico **(Figure 3).** While most samples are grouped closely together, some from Ponce and Bayamon are separated from the main grouping. PC1 indicated around 8.3% of the variation between the main group and the Bayamon outgroup, and PC2 showed 6.4% of variation between the main group and the Ponce outgroup. Additional component comparisons are available in **Figure S6**. As Culebra is an island off mainland Puerto Rico, we anticipated these samples would form a distinct outgroup. However, the Culebra samples cluster closely within the NJ tree, and the PCA revealed minimal genetic diversity compared to the mainland samples.

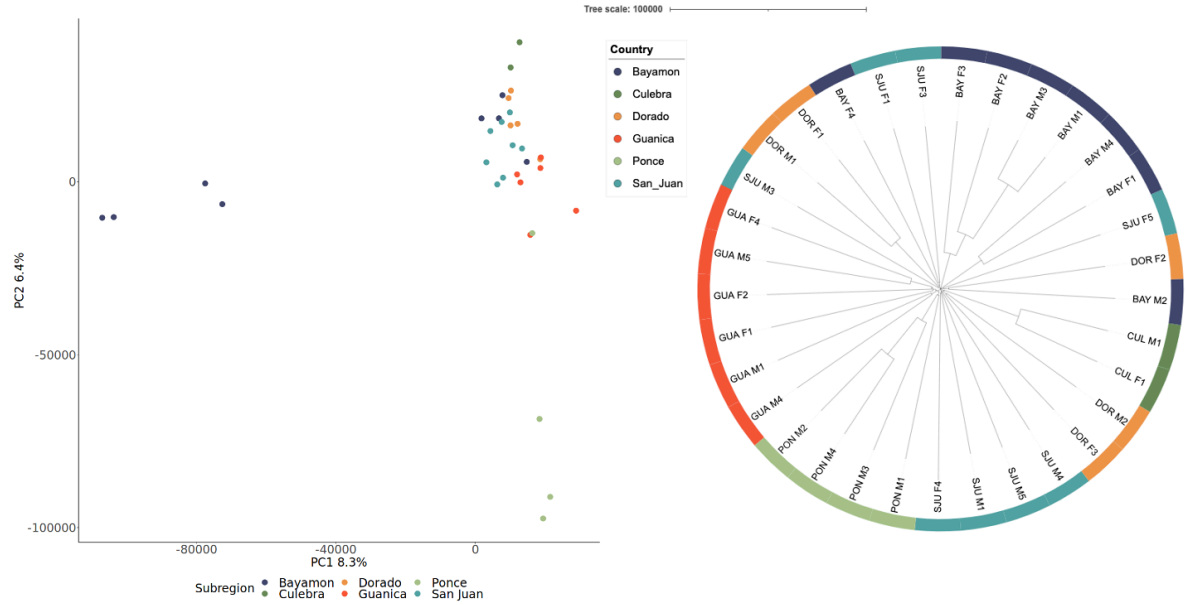


Figure 3. A. *PCA illustrating the samples extracted from Puerto Rico regions,* ***B.*** *A neighbour joining tree based on SNPs for the Puerto Rican samples*

Structural variants

High-quality structural variants (SVs) were identified in the WGS of the Puerto Rican samples using DELLY software, excluding intra-chromosomal variations [31]. Using a literature search and annotation files (see **Methods**), we identified 756 genes with a potential role in insecticide resistance, and SV were detected in 610 genes. After filtering, a total of 5,664 were identified: 1,213 on chromosome 1, 2,861 on chromosome 2, and 1,614 on chromosome 3. Most SVs were deletions (n = 4,841), followed by insertions (n = 413), duplications (n = 221), and inversions (n = 213), as illustrated in **Figure 4A**.

Most SVs occurred in intronic regions (n = 4,676), with smaller proportions in upstream regions (n = 340), downstream regions (n = 133), and inversions (n = 101) (**Figure 4B**). Annotation using SnpEff classified the variants by predicted impact levels: high, moderate, low and modifier [32]. High-impact SVs are defined as having a disruptive impact on the protein function, moderate variants may affect protein effectiveness, low-impact SVs are likely harmless, and modifier variants occur in non-coding regions with minimal or no functional consequences. A total of 296 SVs were classified as high impact, 225 as moderate impact, 93 as low impact, and 5,512 as modifiers.

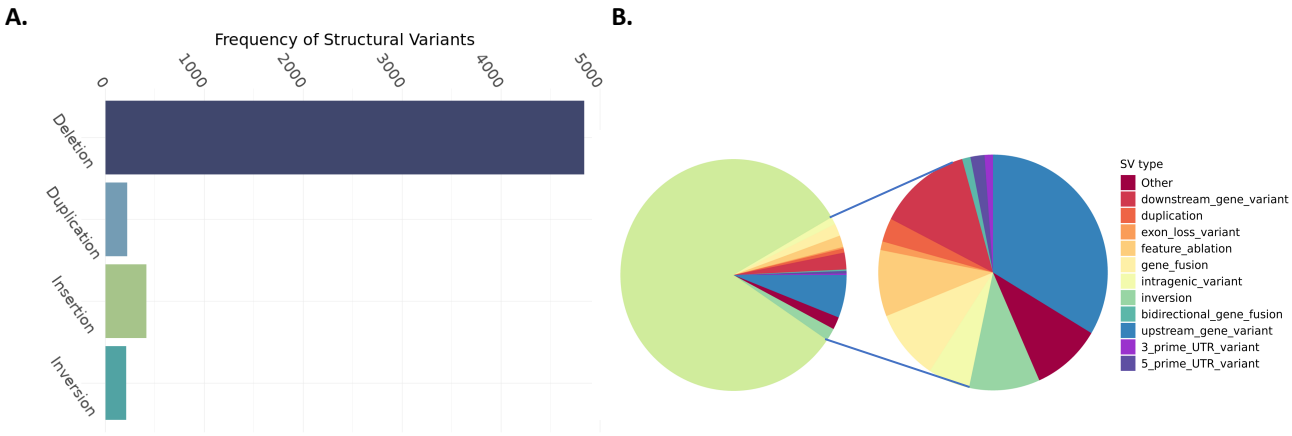


Figure 4. Distribution and Function of SV.

A. *Frequency of each type of SVs within 756 genes of interest,* ***B****. breakdown of the impact of the SVs within 756 genes of interest, the second pie chart shows a breakdown excluding intronic variants.*

In the list of 756 genes of interest, some genes had the highest frequency of SVs across the samples, including five genes associated with G-protein coupled receptors **(Table 1)**. These receptors have been associated with insecticide resistance due to their link with regulating metabolic enzymes such as cytochrome P450s [33,34]. Additionally, four genes from the phosphodiesterase family were highlighted as having a high frequency of SVs **(Table 1)**.

Table 1. Top 10 Structural Variants and Associated Gene Functions

Chrom.	Gene*	Freq.	Function
3	LOC5572215	207	high affinity cGMP-specific 3',5'-cyclic phosphodiesterase 9A
2	LOC23687794	205	probable G-protein coupled receptor Mth-like 1
2	LOC5565389	173	leucine-rich repeat-containing G-protein coupled receptor 5
2	LOC5577718	159	high affinity cAMP-specific and IBMX-insensitive 3',5'-cyclic phosphodiesterase 8
2	LOC23687733	136	G-protein coupled receptor daf-37
1	LOC5571510	127	cAMP-specific 3',5'-cyclic phosphodiesterase
3	LOC23687755	112	G-protein coupled receptor 39
2	LOC5573499	99	calcium/calmodulin-dependent 3',5'-cyclic nucleotide phosphodiesterase 1
3	LOC5571895	94	no receptor potential A
1	LOC5575945	93	G-protein coupled receptor moody

* *Within 756 genes of interest*

Within the *rdl* gene (LOC5570466), 70 SVs were identified, all classified as intronic modifiers. Several variants were found in the ace-1 gene (LOC5578456) (n=70), including one predicted to have high impact effects and 69 categorized as modifiers. The high-impact SV was a duplication spanning positions 161624050 – 161628997 (length = 4,948bp) with an estimated copy number of 1.98. This duplication has various potential effects, including splice acceptor, splice donor, splice region variant,

5' UTR variant, and intron variant. This SV was heterozygous in five samples (n = 32). This duplication is located 123,798 bp away from the described G119S mutation (*XP_021707029.1*) and does not occur in the coding region for this gene's binding site. No SVs were identified in the *vgsc* gene (LOC5567355) or *GSTe2* gene (LOC110676855). The repetitive nature of the *Ae. aegypti* genome limits this analysis, therefore confirming SVs with longer read sequence data would be beneficial for more accurate characterization.

Insecticide Resistance Associated Variant Detection

Of the 2,011,616 high-quality SNPs identified, 4,711 (0.23%) were annotated by snpEff software as having a high impact, potentially resulting in significant functional alterations due to changes such as insertions, deletions, or inversions[32]. Across the data, a total of 281,889 SNPs were identified as missense variants. Within the 756 genes of interest, there were 64,579 SNPs, including 10,799 non-synonymous SNPs. Among these, 136 were annotated as high impact variants. The number of missense SNPs varied between countries, with Puerto Rico having one of the fewest (n=5,528), similar to Mexico (n = 4,409), Brazil (n = 4,124), and Thailand (n = 4,336). In contrast, the USA had slightly more missense SNPs (n = 6,147). The African countries had the highest number of non-synonymous SNPs; Burkina Faso had 8,705, Kenya had 7,943, and Uganda had 8,263 missense SNPs**. Table 2** highlights the genes with the highest frequencies of missense SNPs.

Table 2. *Genes with the highest frequency of missense SNPs*

Gene	Frequency of SNPs	Gene Function
LOC5577718	221	high affinity cAMP-specific and IBMX-insensitive 3',5'-cyclic phosphodiesterase 8
LOC5575945	199	G-protein coupled receptor moody
LOC5571525	123	Uncharacterised
LOC5579144	115	Probable cytochrome P450 4ac1
LOC5573499	108	calcium/calmodulin-dependent 3',5'-cyclic nucleotide phosphodiesterase 1
LOC5578160	105	Esterase B1

Across four insecticide resistance associated genes (*vgsc, rdl, ace-1* and *gste2*) a total of 632 SNPs were identified in all 248 *Ae. aegypti* samples. Most mutations were identified in the *ace*-1 gene (n=219 SNPs), followed by *vgsc* (n = 203), *rdl* (n = 151), and *GSTe2* (n = 59). Among these, 27 SNPs were missense variants, five of which have been previously described to be associated with insecticide resistance (A296S, F1534C, V1016I, V1016G, V410L; using reference species nomenclature) (**Table 3,** bold**)**. Puerto Rico exhibited high allele frequencies for nearly all SNPs located in the *rdl,* vgsc, and GSTe2 genes (AF = A296S; 0.77, F1534C; 0.92, V1016I; 0.91, V1016G; 0.06, V410L; 0.91) compared to the other regions (AF = A296S; 0-0.33, F1534C; 0-0.91, V1016I; 0-0.61, V1016G; 0-0.13, V410L; 0-0.60)

(**Table 3**). The S110T mutation in *rdl* appeared at high frequencies across all populations and was fixed in the Puerto Rican, Brazilian, Mexican, and Kenyan populations. Puerto Rico demonstrated a high frequency of SNPs associated with insecticide resistance, aligning with observed elevated levels of resistance in this population (**Figure 5**) [25–27,35]. A further 152 synonymous mutations were identified across all populations in the main insecticide resistance genes (*vgsc, rdl, ace-1* and *gste2*), the majority of these were in the *vgsc* (n = 72), followed by *ace-1* (n = 36), *GSTe2* (n = 26) and *rdl* (n = 18) genes.

Table 3. *Missense mutations identified in the four genes associated with resistance. Bold mutation indicated described mutation. Allele frequency is provided for each geographical region.*

Gene	Chrom.	Position	Translation	Allele Frequency							
				Thailand	Brazil	Puerto Rico	Mexico	USA	Burkina Faso	Uganda	Kenya
rdl	2	41786063	S110T	0.85	1.00	1.00	1.00	0.88	0.97	0.95	1.00
rdl	2	41847790	**A296S**	0.33	0.06	0.77	-	0.33	0.07	0.00	0.00
ace-1	3	161487441	C699S	0.00	0.14	0.02	0.03	0.06	0.14	0.00	0.00
ace-1	3	161510815	E71K	-	-	0.08	-	0.01	-	0.00	0.00
ace-1	3	161510992	G12S	0.10	0.08	0.05	0.44	-	0.01	-	-
ace-1	3	161679988	L3I	0.08	0.17	0.09	0.06	0.25	0.85	0.88	0.94
ace-1	3	161680347	R83H	0.00	0.06	0.02	0.03	0.18	0.11	0.16	0.10
ace-1	3	161680365	E89A	0.00	0.06	0.02	0.03	0.18	0.07	0.16	0.10
ace-1	3	161694832	S21N	0.00	0.00	-	-	0.00	0.03	0.12	0.05
ace-1	3	161694834	I22L	0.06	0.00	0.03	0.09	0.07	0.33	0.08	0.10
ace-1	3	161694840	R24Y	0.00	0.00	-	-	0.00	0.01	0.00	0.03
ace-1	3	161695278	F149V	0.00	0.00	-	-	0.00	0.06	0.03	0.00
vgsc	3	161695564	F224L	0.00	0.00	-	-	0.00	0.03	0.11	0.19
vgsc	3	315931672	Q1873R	-	-	0.61	-	0.01	-	0.00	0.00
vgsc	3	315939224	**F1554C**	0.62	0.58	0.92	0.91	0.48	0.42	0.00	0.00
vgsc	3	315983762	**V1012G**	0.13	-	0.06	-	-	-	-	-
vgsc	3	315983763	**V1012I**	0.00	0.00	0.91	0.28	0.61	0.36	0.00	0.00
vgsc	3	316014588	S711T	0.00	0.00	0.86	0.25	0.62	0.29	0.00	0.00
vgsc	3	316080722	**V408L**	0.00	0.00	0.91	0.22	0.60	0.33	0.00	0.00
GSTe2	2	351633680	L9I	0.00	0.00	-	-	0.02	0.21	0.13	0.11
GSTe2	2	351633753	T33M	-	-	0.02	-	0.00	-	0.03	0.00
GSTe2	2	351634048	L111S	0.02	0.67	0.85	0.91	0.25	0.01	0.03	0.03
GSTe2	2	351634049	L111S†	0.02	0.67	0.88	0.91	0.23	0.00	0.00	0.00
GSTe2	2	351634667	I150V	0.00	0.56	0.85	0.94	0.49	0.39	0.39	0.32
GSTe2	2	351634725	I169S	0.00	0.00	-	-	0.03	0.08	0.05	0.00
GSTe2	2	351634752	E178A	0.00	0.50	0.86	0.88	0.35	0.08	0.18	0.23
GSTe2	2	351634753	E178A‡	0.00	0.50	0.85	0.88	0.35	0.08	0.18	0.23
GSTe2	2	351634812	A198E	0.00	0.72	0.85	0.94	0.47	0.31	0.32	0.35

† This mutation occurs in the same codon as Leu111Ser, and in this study was present with mutation in 351634048, the resulting change is leucine to serine. ‡ This mutation occurs in the same codon as Glu178Ala, and in this study was present with mutation in 351634752, the resulting change is from glutamine to Alanine.

The GABA A296S mutation (*rdl,* 41847790) linked to resistance to organochlorine insecticides, was found in five populations (54/232, 23.3%), with the highest allele frequency in Puerto Rico (77.3%), USA (33.0%) and Thailand (32.7%), and no mutations identified in Kenya or Uganda (no SNPs were called in this position in the samples from Mexico due to low coverage). The F1534C mutation (*vgsc,* 315939224) associated with pyrethroid resistance was identified in 116 samples across all populations except Kenya and Uganda. This mutation was found in the highest proportion in Puerto Rico (92.4%), Mexico (90.6%) Brazil (84.6%) and Thailand (52.8%). Another *vgsc* mutation (V1016I, position 315983763; n=91) was identified in these same four countries (Puerto Rico, USA, Mexico, and Burkina Faso), while the V1016G mutation (315983762) was detected in four samples from Puerto Rico and seven from Thailand. The V410L mutation (*vgsc,* 316080722), associated with pyrethroid resistance, was less frequent and only identified in populations from Puerto Rico, USA, Mexico and Burkina Faso (between 16.2-23.0% frequency). **Figure 5** highlights the distribution of insecticide resistance mutations across populations, revealing that Puerto Rico, the USA, and Mexico have the highest proportion of samples harbouring multiple resistance mutations.

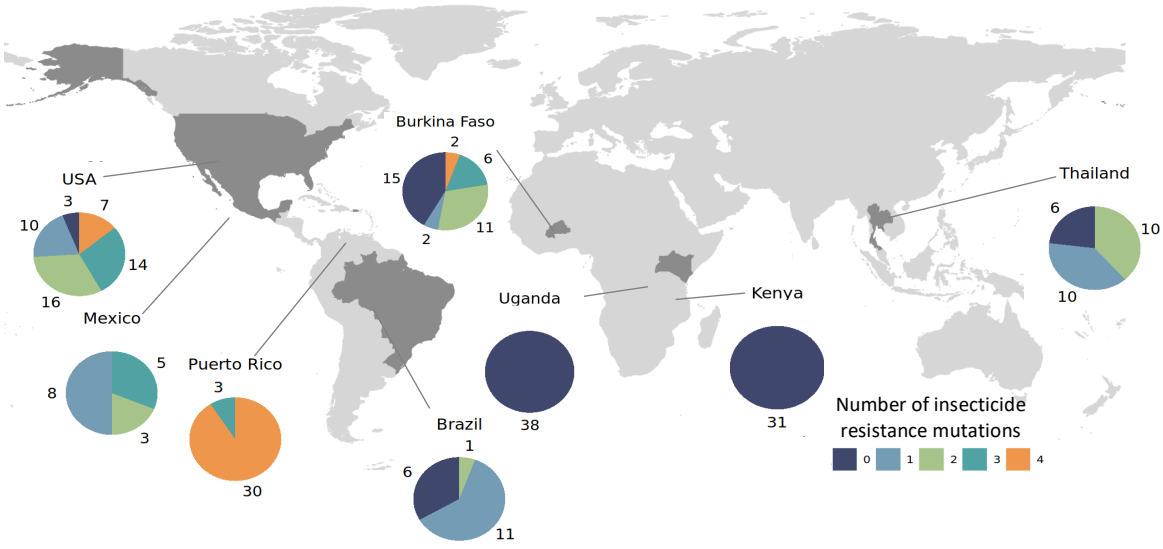


Figure 5. Distribution of samples by number of insecticide resistance mutations *(V401L, V1016I, F1534C in vgsc; A302S in rdl)*

Linkage Disequilibrium

Within the Puerto Rico samples linkage disequilibrium (LD) analysis revealed linkage between a number of SNPs in the insecticide resistance genes (*vgsc, ace-1, rdl,* and *GSTe2*) **(Figure 6)**. Most linked SNPS were identified on chromosome 2, where the *rdl* and *GSTe2* genes are located. The strongest linkage was observed between 315634753 (*GSTe2* - E178D) and 351634812 (*GSTe2* - A198E). There are six positions in that gene region that had relatively high levels of linkage. There were no SNPs in significant linkage in the *rdl* gene. Within the *ace-1* gene (chromosome 3) seven positions exhibited some level of association. Particularly notable were positions 161680347 (R38H) and 161680365 (E89A), 161680365 (E89A) and 161487441 (C699S), and 161487441 (C699S) and 161680347 (R38H), all of which had an LD value of 1, indicating complete linkage. Additionally, on chromosome 3 within the *vgsc* gene, absolute LD was observed between positions 316080722 (*vgsc* - V408L (V410L in *Musca domestica*)) and 315983763 (*vgsc* - V1012I, (V1016I in *Musca domestica*)).

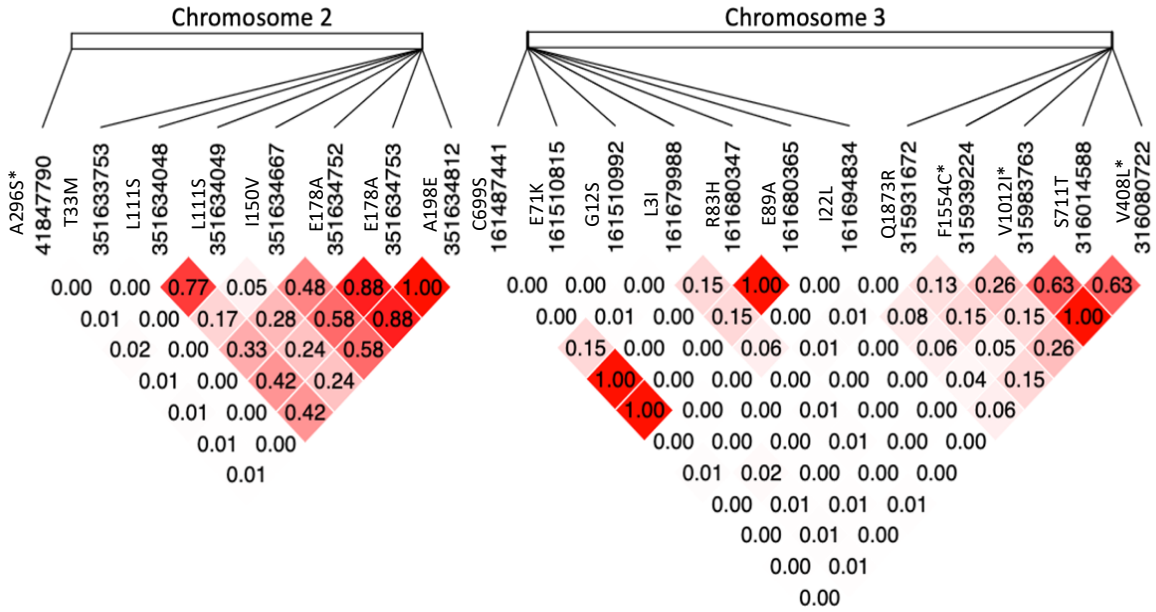


Figure 6. *Linkage disequilibrium R^2 values for the Puerto Rician samples for positions within the rdl and GSTe2 gene on chromosome 2 and the ace-1 and vgsc genes on chromosome 3. Mutations with * indicate the mutation is associated with insecticide resistance.*

Identification of Regions Under Selection

Environmental pressures, such as exposure to insecticides, can drive selective pressures that result in the frequency of certain alleles increasing across a population. To identify these selection signals, we conducted several analyses to identify regions under selection.

Tajima's D

Tajima's D (T_D) analysis was used to examine genetic diversity to distinguish alleles potentially under selection. We performed the T_D analysis using 100kb windows across the Puerto Rican samples, obtaining a median T_D of 1.70, with chromosomes 1-3 showing similar values (1: 1.68, 2: 1.69, 3: 1.71); none of these values indicate balancing selection **(Figure S7A)**. At the country level, T_D values did not significantly differ between populations at the exome level, with mean T_D values ranging from 1.02 (Kenya) to 1.70 (Puerto Rico). Across all populations, no significant differences were found in non-overlapping 100kb windows (**Figure S7B**).

However, at a more granular level, 17,627 windows showed a significant T_D value (either > 2 or < -2) indicating possible selection (100kb windows) [36]. The Puerto Rico and USA population had the highest number of significant windows (n = 3,801 and n = 3,506, respectively), while Kenya had the lowest (n = 1,037). **Table 4** shows the genomic position in the top 10 and lowest 10 Tajima's D values. Within the genes of interest, 574 genes had significant T_D values. The highest values indicating balancing selection (T_D>2) were observed in a cluster of probable cytochrome P450 genes in the USA population, including; P450 6a14 (LOC110674119, T_D = 4.8713, LOC5565578, 4.8713), P450 6a8 (LOC23687481, T_D = 4.8713), P450 4d8 (LOC5573388, T_D = 4.6521), P450 6a13 (LOC5565579, T_D = 4.4104). Several insecticide resistance-associated genes were also identified in the Puerto Rican population, including P450 b561 (T_D = 4.5337, LOC5576849), esterase FE4 (T_D = 4.3066, LOC5567206), and P450 6a14 (T_D = 4.0904, LOC5572936). The esterase FE4 gene was also highly significant in the Thailand population (T_D = 4.0683). Genes showing potential positive selection (TD < -2) included a cuticle protein 16.5 in Mexico (LOC5571813, T_D = -2.7568), two glutathione S-transferase 1 (LOC5569858, T_D = -2.5194, LOC5569859, T_D = -2.5194) and a cytochrome P450 4d8 (LOC5573388) in Thailand.

Enrichment analysis (n = 10,181 genes with available gene name conversion) showed metabolic pathways were enriched from the Tajima's D output, this covered 659 (6.5% of the inputted genes), with a Benjamini-Hochberg FDR adjusted p-value of $8.7x10^{-12}$. Gene ontology (GO) functional annotation further identified significant enrichments in genes linked to monooxygenase activity (n=155, p = $1.4x10^{-16}$), serine-type endopeptidase activity (n = 271, p = $9.3x10^{-16}$) and oxidoreductase activity (n = 131, p = 3.2x10-13). The term "ABC-transporter activity" (n=41, p = 0.077) was also enriched, a function previously linked to insecticide resistance.

Table 4. *Ten highest (**A**) and lowest (**B**) Tajima's D values across all eight populations.*

A.

Chrom.	Population	T_D	Genes	Function
2	USA	4.989	LOC5578399	mismatch repair endonuclease PMS2 isoform X1
			LOC110675918	uncharacterized protein LOC110675918
			LOC5572764	WD repeat-containing protein 89
			LOC5578401	calcium and integrin-binding family member 2
			LOC5572766	mitochondrial thiamine pyrophosphate carrier
1	USA	4.871	LOC5576390	sodium/calcium exchanger 1 isoform X1
			LOC5565578	probable cytochrome P450 6a14
			LOC110674119	probable cytochrome P450 6a14
			LOC23687481	cytochrome P450 6a8
3	USA	4.829	LOC5577739	Ras-associated protein 2-like
			LOC5576911	product=trypsin 3A1-like
			LOC110678621	uncharacterised LOC110678621
			LOC110678619	uncharacterised LOC110678619
			LOC110678620	uncharacterised LOC110678620
2	USA	4.789	LOC5572766	mitochondrial thiamine pyrophosphate carrier
			LOC5572767	uncharacterized protein LOC5572767
			LOC5572768	SAGA-associated factor 29
			LOC5566366	polyadenylate-binding protein 2
2	USA	4.757	LOC5567854	acetyl-CoA carboxylase isoform X3
			LOC110675339	tctex1 domain-containing protein 1-A-like
			LOC5567866	uncharacterized protein LOC5567866
2	Puerto Rico	4.694	LOC5574680	atrial natriuretic peptide receptor 1 isoform X1
			LOC5574679	fukutin-related protein
2	USA	4.690	LOC5565243	Nose resistant to fluoxetine protein 6
			LOC5565288	Nose resistant to fluoxetine protein 6
			LOC5578252	Nose resistant to fluoxetine protein 6
			LOC110675121	uncharacterized protein LOC110675121
3	USA	4.678	LOC5575100	39S ribosomal protein L41%2C mitochondrial
			LOC23687752	DNA topoisomerase 1
			LOC5575101	uncharacterized protein LOC5575101 isoform X1
			LOC5575098	endoribonuclease rege-1 isoform X1
2	USA	4.672	LOC5570661	Hemolymph protein
			LOC5570664	Hemolymph protein
			LOC5570673	Uncharacterised LOC5570673
			LOC5570671	Protein PBDC1
			LOC5570670	NADH dehydrogenase B18 subunit
2	USA	4.652	LOC5573386	neuroendocrine protein 7B2
			LOC5573388	cytochrome P450 4d8

B.

Chrom.	Population	T_D	Genes	Function
2	Kenya	-2.811	LOC5569978	Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1
			LOC5569977	hyperpolarization activated cyclic nucleotide gated potassium channel Ih
2	Puerto Rico	-2.770	LOC5574486	Gustatory receptor
			LOC5574489	potassium/sodium hyperpolarization-activated cyclic nucleotide-gated channel 4
			LOC110676696	Uncharacterised LOC110676696
2	Mexico	-2.763	LOC5571810	protein odr-4 homolog
			LOC5571811	uncharacterized protein LOC5571811 isoform X1
2	Mexico	-2.757	LOC5571813	cuticle protein 16.5
			LOC110677262	uncharacterized protein LOC110677262
			LOC5571814	uncharacterized protein LOC5571814
2	Thailand	-2.730	LOC110676862	Collagen alpha-1 (IV) chain
			LOC5565334	SPRY domain-containing SOCS box protein SP555
2	Mexico	-2.718	LOC5567691	sex peptide receptor
			LOC110676393	Uncharacterised LOC110676393
1	Mexico	-2.686	LOC110674038	cadherin-87A
			LOC5577039	protein D3
2	Mexico	-2.646	LOC5564742	Plexin-B
			LOC5564741	Protein-serine O-palmitoleoyltransferase por
			LOC5574639	Ca-channel protein alpha subunit D
1	Brazil	-2.645	LOC5565616	zinc finger protein 35
			LOC5565617	adipocyte plasma membrane-associated protein
			LOC5565604	adipocyte plasma membrane-associated protein
			LOC5565593	F-box only protein 11
3	Mexico	-2.642	LOC110678585	Paired box protein pax-6-like

Integrated haplotype score

To further investigate selection within each population, we performed integrated haplotype score (iHS) and haplotype heterozygosity (H12) analysis. The iHS analysis identified 341 loci across all eight populations as having significant iHS scores (-log10 p value $|iHS| \geq 4$). The number of significant loci was relatively consistent across populations, ranging from 36 to 48 **(Table 4)**. Amongst the significant loci, 14 genomic positions occurred in at least two populations, as outlined in **Table 5**. This included a neuroglian gene (LOC5570456) occurring in 4 populations (Mexico, Puerto Rico, Thailand, and USA) and an uncharacterised gene (LOC5577087) which occurred in 3 populations (Kenya, Uganda, and Burkina Faso). Analysis of significant genes indicated no genes were enriched regarding functional annotation, ontology, pathways or protein domains.

Table 5. *Genes with iHS values significant in more than one population.*

Gene	Function	No. pops	Countries
LOC5570456	neuroglian	4	Mexico, Puerto Rico, Thailand and USA
LOC5577087	Uncharacterized LOC5577087	3	Burkina Faso and Kenya
LOC5564807	histidine decarboxylase	2	Mexico and Uganda
LOC5566204	gamma-aminobutyric acid receptor subunit alpha-6	2	Puerto Rico and Kenya
LOC5569249	protein lava lamp		Brazil and Kenya
LOC5570454	nose resistant to fluoxetine protein 6	2	Mexico and Puerto Rico
LOC5572005	carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase rudimentary	2	Mexico and Thailand
LOC5572967	karyopherin beta 3	2	Brazil and Mexico
LOC5573437	facilitated trehalose transporter Tret1	2	Thailand and USA
LOC5574699	MOXD1 homolog 2	2	Burkina Faso and Uganda
LOC5576299	BTB/POZ domain-containing protein inc	2	Burkina Faso and Kenya
LOC5578908	putative fatty acyl-CoA reductase CG5065	2	Puerto Rico and Thailand
LOC5579439	Hecw ubiquitin protein ligase	2	Brazil and USA
LOC5579505	protein NPC2 homolog	2	Kenya and Uganda

No positions were identified in the four insecticide resistance genes; *vgsc, rdl, ace-1* or GSTe2. However, nine of the significant genomic positions were identified within the 756 insecticide resistance genes of interest (see **Table 6, Figure S8)**. The highest iHS values were observed in the Ugandan population in a probable cytochrome p450 gene (LOC5575901), these genes are linked to metabolic insecticide resistance [37–39]. Positions within the gamma-aminobutyric acid receptor subunit alpha (LOC5566204) were highlighted in both the Kenyan and Puerto Rican populations. This gene may encode part of the GABA receptor complex that is a target for organochlorine insecticides [40,41]. Three positions in a UDP-glucuronosyltransferase gene, which are linked to metabolic resistance [11], had significant iHS scores in the Kenyan population. The Thai population signaled a site within the sphingomyelin phosphodiesterase gene (LOC5567918).

Table 6. *Significant iHS values within insecticide resistance relevant genes.*

Population	Chrom: Position	iHS value	Gene and Role
Thailand	1:101126204	4.147	LOC5567918 sphingomyelin phosphodiesterase
Kenya	3:261441180	4.407	LOC5571109 UDP-glucuronosyltransferase
Kenya	3:361441195	4.180	
Kenya	3:261441198	4.231	
Kenya	1:28116470	4.532	LOC5566204 gamma-aminobutyric acid receptor subunit alpha
Puerto Rico	1:28116530	4.751	
Puerto Rico	1:28116545	4.274	
Puerto Rico	1:28116641	4.575	
Puerto Rico	1:28116656	4.876	
Puerto Rico	1:28116683	4.809	
Uganda	2:232716649	5.326	LOC5575901 probable cytochrome P450 303a1
Uganda	2:232716673	5.326	

Within Puerto Rico, 45 loci were identified with iHS > 4, suggesting selection pressure. The highest iHS value (5.65) was identified in a gene encoding a fatty acyl-CoA reductase (LOC5598908), in the oxidoreductase family, two significant positions were identified in this gene. Additionally, three positions in a probable protein for S-acyltransferase (LOC5569080), a transferase enzyme, exhibited high iHS values (iHS = 4.95, 5.18, 5.35). The gamma-aminobutyric acid receptor subunit alpha gene (LOC5566204) also showed high iHS values (iHS = 4.75, 4.81, 4.88) in Puerto Rico, further implicating its role in organochlorine insecticide resistance [42]. A genome-wide selection scan of Garud's H12 across the three main chromosomes did not identify any clear selection signals in any of the populations **(Figure S9).**

Cross population extended haplotype homozygosity

To examine selection across populations, cross population extended haplotype homozygosity (XP-EHH) analysis was performed. XP-EHH analysis revealed many differences between the populations. A total of 504 positions, comprising 236 genes, were identified as having significant XP-EHH scores (XP-EHH > 5 or < -5) across all population comparisons. Notably, one position, located at 313184797 in gene LOC5574943, was found to be significant in six population comparisons: Brazil vs Kenya, Brazil vs Uganda, Burkina Faso vs Kenya, Burkina Faso vs Uganda, USA vs Kenya, and USA vs Uganda. This gene is uncharacterized, limiting the functional insights that can be drawn from its selection signal. Additionally, 18 positions were found to be significant in four population comparisons.

GO term analysis revealed significant enrichment of the cellular component 'plasma membrane' (GO:0005886) in the dataset, with 29 associated genes, indicating a fold enrichment of 16.7 (adjusted p = 0.00013), suggesting a strong association of these genes with plasma membrane-related functions. Another enriched term was 'stereocilium tip' (GO:0032426), which was enriched 1.7 times, associated with three genes (adjusted p = 0.0095).

Within the 756 genes of interest, there were 14 genes with significant XP-EHH values (XP-EHH > |5|) **Table 7**. The ten highest XP-EHH values across all population comparisons were all amongst the African populations. Specifically, 18 positions were significant between Burkina Faso and Kenya, and two positions between Uganda and Kenya. These were across five genes: glutathione synthetase (LOC5571567), phosphodiesterase 12 (LOC5565275), glutathione S-transferase E14 (LOC5569853), cytochrome P450 4d1 (LOC5569662) and gamma-aminobutyric acid receptor alpha-6 (LOC5566204). Several genes had multiple positions under selection, including phosphodiesterase 12 (LOC5565275,

n = 15), cytochrome P450 4d1 (LOC5569662, n = 11), and glutathione synthetase isoform (LOC5571567, n = 7). No significant positions were identified within the four target-site resistance genes (vgsc, rdl, ace-1, and GSTe2).

Table 7. Insecticide resistance genes of interest with significant XP-EHH positions between countries.

Gene Name	Description
LOC5571567	glutathione synthetase
LOC5565275	2',5'-phosphodiesterase 12
LOC5569853	glutathione S-transferase E14
LOC5569662	cytochrome P450 4d1
LOC5567206	esterase FE4
LOC5566915	probable cytochrome P450 313a4
LOC110679003	flexible cuticle protein 12-like
LOC5564749	probable cytochrome P450 9f2
LOC5569919	cytochrome P450 4c3
LOC5567232	cuticle protein
LOC5575945	G-protein coupled receptor moody
LOC5566204	gamma-aminobutyric acid receptor subunit alpha-6
LOC5570386	cuticle protein CP14.6
LOC5564760	cytochrome P450 9e2

When comparing each population with Puerto Rico we identified 147 positions with significant XP-EHH values (>|5|). None of these positions were located in the four main target site insecticide resistance genes. Furthermore, none of these positions were detected in multiple population comparisons. All the significant values occurred when comparing the Puerto Rican population against Mexico, Thailand, or Uganda. The highest XP-EHH value (XP-EHH = 5.3) was detected at position 381006532 against the Thai population within a gene (LOC110679045) of uncharacterized function. The second-highest value (XP-EHH = 5.0) occurred at position 309385832 against the Mexican population, located in the probable pyruvate dehydrogenase E1 component subunit alpha (LOC5566309). **Figure S10** shows all population comparisons to Puerto Rico.

Across selection tests

A total of 154 genes were identified as significant in both the iHS and Tajima's D analyses. Enrichment analysis of 111 of these genes (where conversion was possible) revealed no significant trends in terms of function, ontology, or protein domains. Among the 154 genes identified in both analyses, three genes were also included in the list of potential insecticide resistance genes: gamma-aminobutyric acid receptor alpha (LOC5566204), sphingomyelin phosphodiesterase (LOC5567918), and a probable cytochrome P450 303a1 (LOC5575901). Additionally, 20 genes were significant in all three selection analyses (Tajima's D, iHS, and XP-EHH) (**Table 8**). However, enrichment analysis of these genes showed no significant link in terms of biological or molecular function or ontology.

Table 8. *Genes detected to be significant in Tajima's D, iHS and XP-EHH analysis.*

Gene	Description
LOC110676664	uncharacterized
LOC5576136	DNA-binding protein RFX2
LOC5564807	histidine decarboxylase
LOC5573900	intracellular protein transport protein uso1
LOC5565181	Actin
LOC5578603	uncharacterized
LOC5565688	fibroblast growth factor receptor homolog 1
LOC5566204	gamma-aminobutyric acid receptor subunit alpha-6
LOC5566499	putative uncharacterized protein DDB_G0285119
LOC5574807	probable chitinase 10
LOC5577087	uncharacterized LOC5577087
LOC5564637	centrosomal protein of 120 kDa
LOC5564933	protein sickie
LOC5578281	integrin alpha-PS2
LOC5565823	ornithine decarboxylase 1
LOC5568525	uncharacterized
LOC110678439	uncharacterized
LOC5569999	protein polybromo-1
LOC5573728*	Lrp4 LDL receptor related protein 4
LOC5574019*	Tbh Tyramine beta hydroxylase

*Genes with * were not included in the enrichment analysis because they were not able to be converted to DAVID format.*

Discussion

Our results have highlighted genomic differences between Puerto Rican *Ae. aegypti* populations and seven other global populations. The ancestral analysis revealed the expected alignment with geographical source, as well as previously described similarities, such as samples from Thailand grouping more closely with the Americas than with Africa. This similarity is surprising given the physical distances and oceanic barriers between continents but is consistent with the theory that *Ae. aegypti* was introduced to Asia from the Americas [2]. We also confirmed previous findings indicating that African populations have less distinct genetic differentiation between countries than the populations from the Americas, which may reflect higher levels of gene flow between these countries (**Figure S2)** [43,44].

The Puerto Rican samples clustered most closely with Mexican samples, which may reflect historical trade movements, as Puerto Rico is a territory of the USA therefore, subsequently, there are strong

ties between these regions, with 76% of exports and 50% of imports occurring with the USA in 2020 [45]. The lack of variation observed among Puerto Rican mosquitoes across different regions of the island was anticipated, likely attributable to the constrained temporal distribution and geographic dispersal of the collection sites, with maximum distances of approximately 140 km.

The Puerto Rican, USA and Mexico populations showed the highest frequency of insecticide resistance SNPs. The Puerto Rican mosquitoes had a high frequency of target site resistance SNPs (V410L, V1016I, F1534C and A302S). This has likely arisen in response to intensive spraying during outbreaks on the island, including in recent years during a Chikungunya outbreak in 2014, a Zika outbreak in 2016 and the ongoing dengue outbreak [25,29,46,47] . The V1016I and V410L were in linkage disequilibrium in Puerto Rico ($R^2 = 1$). In contrast, no insecticide resistance SNPs were detected in populations from Uganda or Kenya, which may be due to the primary focus on malaria control in these regions [48,49]. Malaria vector control efforts are mainly directed towards insecticide-treated bed nets, which the day-biting *Ae. aegypti* mosquitoes may not frequently encounter. However, insecticide resistance mutations, including F1534C, were observed in Burkina Faso, where insecticide-based responses were implemented during major dengue outbreaks in 2013 and 2017[50,51]. The F1534C codon within the *vgsc* gene exhibited one of the highest F_{ST} values in multiple population comparisons, particularly when comparing Kenyan or Ugandan populations (where the SNP was absent) with Mexican and Puerto Rican populations (where this SNP was prevalent). Although we detected the V1016I variant in multiple populations, the V1016G mutation was only identified in the Puerto Rican and Thailand populations. The V1016G variant has only been described in Asian *Ae.* aegypti populations thus far [14,52] but was also detected in amplicon sequencing on this population from Puerto Rico, suggesting the emergence of the V1016G in the Americas [26].

Several mutations were detected in the *GSTe2* gene (e.g., L111S, I150V, E178A, and A198E) which have previously been described in DDT-resistant populations from Zanzibar, but with no strong evidence of genotype-phenotype association [53]. An alternative study identified these mutations (L111S, I150V, E178A, and A198E) in a resistant *Ae. aegypti* strain (RecR) and protein modeling indicated that L111S and I150V mutations occur near the enzymatic pocket and therefore may enhance the ability to metabolise the substrate [54]. It is not clear if these mutations may confer resistance to insecticide classes other than organochlorines or if these represent a divergence from the reference sequence, therefore further functional studies are required to confirm their effect. In Puerto Rico, these variants were both detected at high frequency (85%).

A GABA a mutation in the *rdl* gene was detected in all populations at high frequency. Investigation into this position reveals that in both *Aedes albopictus* and *Anopheles gambiae* reference genomes, the equivalent amino acid is threonine; this may indicate that the *Aedes aegypti* reference (GCF_002204515.2_AaegL5.0) has diverged from the dominant form and the threonine residue may represent the ancestral form of the amino acid. A protein blast of the *Ae. aegypti rdl* protein revealed a single older sequence of the *rdl* gene (AAA68961) containing the threonine amino acid rather than serine, from 8 sequences analysed with over 90% identity. Similarly, a BLAST of *Ae. aegypti rdl* against the nucleotide database identified only one sequence with the T to A alleles to confer the change from serine to threonine (U28803.1).

Structural variant analysis of the Puerto Rican samples revealed no duplications within cytochrome P450s as seen in *Anopheles funestus* [15], but several SVs were detected in phosphodiesterase and G-protein-coupled receptor genes, which could be linked to insecticide resistance [55–57]. No significant SVs were detected in key resistance genes (vgsc, rdl, or GSTe2), although a duplication (4,948 bp) was identified upstream of the coding region of the ace-1 gene, therefore, may be unlikely to mitigate insecticide effects. This analysis would benefit from longer sequencing reads to better explore SVs, given the repetitive nature of the *Ae. aegypti* genome.

Selection analysis revealed numerous genes of interest. Notably, Tajima's D analysis indicated several genes in balancing selection, including several cytochrome 450 genes. In the USA population, balancing selection was detected in P450 6a14, P450 6a8, P450 4d8, P450 6a13, and P450 b651. A number of these P450 6a14 (CYP6N9, CYP6BB2, CYP6M9) and P450 6a13 (CYP6N12) have been associated with metabolic insecticide resistance in transcriptomic studies [58–61]. Balancing selection may indicate a fitness cost, suggesting that heterozygote advantages might drive the persistence of these genes in the population. Fitness costs relating to longevity and mating competition have been observed in *Ae. aegypti* with cytochrome P450 mediated resistance [62]. It may also be that different genotypes have advantages in different environments. This analysis also revealed other genes under positive selection, including a gustatory receptor (LOC5574486) in the Puerto Rican population. These receptors are involved in taste in the mouthparts and legs and can influence feeding, mating, biting, and egg-laying [63]. It is possible that alterations in sensing in this way could be used to avoid repellants or insecticides, there is evidence in *Drosophila* that gustatory receptors are used to detect DEET repellant [64]. Another gene highlighted in the Mexican *Ae. aegypti* population was a cuticle protein (LOC5571813). Cuticle proteins are known to be involved in insecticide resistance in multiple mosquito species [65]. These findings indicate that selection pressures might be influencing these genes to

produce beneficial phenotypes, although functional studies are needed for confirmation. Enrichment analysis of the significant Tajima's D test indicated metabolic pathways that were enriched (p-value $8.7x10^{-12}$). GO functional annotation also highlighted trends in monooxygenase activity (n=155, p = $1.4x10^{-16}$) such as cytochrome P450s implicated in insecticide resistance [14], serine-type endopeptidase activity (n = 271, p = $9.3x10^{-16}$) which can have roles in digestion, development and immunity [66] and oxidoreductase activity (n = 131, p = $3.2x10^{-13}$), which have a wide range of functions including, involvement in stress response and in the olfactory system.

Some genes implicated in metabolic resistance were also identified in iHS analysis, including cytochrome P450 (LOC5575901) in the Ugandan population (iHS = 5.326) and a UDP-glucuronosyltransferase (LOC5571109) in the Kenyan population (3 positions, iHS between 4.231 - 4.407) [39,67]. In the Thai population, a sphingomyelin phosphodiesterase was highlighted (iHS = 4.147). These enzymes are involved in producing ceramide, which has been speculated to be involved with viral resistance against dengue in the Cali-MB dengue-resistant strain of *Ae. aegypti* [68]. Furthermore, this gene has been reported to be involved in metabolic processes in pyrethroid-resistant mosquitoes, along with other lipid biosynthesis genes [57].

XP-EHH analysis identified a recurring position in the gene LOC5574943 across six population comparisons. This gene, with an uncharacterized function, warrants further investigation. Several genes potentially involved in insecticide resistance were prominently identified, especially in comparisons with African populations. Key genes include glutathione synthetase (LOC5571567), phosphodiesterase 12 (LOC5565275), glutathione S-transferase E14 (LOC5569853), cytochrome P450 4d1 (LOC5569662), and gamma-aminobutyric acid receptor alpha-6 (LOC5566204). Among these are major enzymes within families associated with metabolic insecticide resistance, including cytochrome P450 monooxygenases (P450s), esterases, UDP-glycosyl-transferases (UDPGTs) and glutathione S-transferases [15,39,54,69]. The gamma-aminobutyric acid receptor is linked to nerve action and may be part of the receptor target for organochlorine insecticides [40,70]. Gamma-aminobutyric acid receptor alpha was also highlighted in the iHS analysis. Conducting additional functional studies to explore these genes possibly under selection would be important for verifying their link to insecticide resistance or other biological functions. In the long term, this knowledge can facilitate the design of more effective vector control and surveillance measures.

Across the three selection tests (Tajima's D, iHS, and XP-EHH), 20 genes were consistently identified as significant. Enrichment analysis revealed no significant link between these genes in terms of

function, ontology, or pathway. One gene highlighted across the analysis and in different populations was the gamma-aminobutyric acid (GABA) receptor subunit alpha (LOC5566204). BLAST and ortholog analysis indicate this gene may be equivalent to the CG8916 ortholog [71], which along with *rdl* (resistance to dieldrin), *LCCH3* (ligand-gate chloride channel homolog 3), and *GRD (GABA and glycine-like receptor of Drosophila)*, may encode subunits of the ionotropic GABA receptor. **Figure S11** shows a phylogenetic tree for the similarity between these proteins of various species. The CG8916 gene has been implicated in GABA receptor function in vitro and is sensitive to several insecticides in combination with *LCCH3* [72]. This suggests that the observed selection on the GABA receptor alpha subunit may be driven by insecticide pressure, particularly from organochlorines like cyclodienes.

Further analysis of *LOC5566204* revealed that several of the positions identified by iHS in both the Kenyan and Puerto Rican populations occur within the ligand-binding region (synonymous mutations Y267, K287, R291, E324, M329, L338). A further three missense mutations were identified within the binding site (L220F, D225G, and L282F). This gene is especially noteworthy given the high prevalence (77.3%) of the *A301S* mutation in the Puerto Rican population within the *rdl* gene, which is also involved in GABAergic signalling. These findings underscore the potential role of *LOC5566204* in resistance to GABA-targeting insecticides and highlight its relevance for further functional studies to understand its role in insecticide resistance mechanisms.

Conclusion

Here, we have generated and explored *Ae. aegypti* WGS data from Puerto Rico and compared with other populations to identify genetic differences between populations, with a focus on insecticide resistance genes. A substantial number of single nucleotide polymorphisms (SNPs) and structural variants (SVs) were identified, including numerous non-synonymous SNPs in genes associated with insecticide resistance. Given the genetic similarity between Puerto Rican, USA, and Mexican Ae. aegypti populations, the mechanisms underlying resistance may be likely to overlap. Insecticide resistance is complex, and many mechanisms may exist. In this study, we have highlighted many candidate genes under selection that warrant further exploration. These genes, along with phenotypic bioassays, should be investigated to validate their roles in resistance traits. Such research could inform more effective strategies for vector control and resistance management in *Ae. aegypti*.

Methods

Sample Collection

We collected *Ae. aegypti* eggs using ovitraps across various sites in Puerto Rico, including San Juan, Dorado, Bayamon, Ponce and Guanica, in May 2022. These mosquitoes were reared in the insectary at the Puerto Rico Vector Control Unit (PRVCU). Mosquitoes were killed at 3-5 days old and stored at -20˚C. Mosquitoes were morphologically speciated. DNA was extracted using Qiagen DNAeasy kits using the standard blood and tissue protocol. The DNA of 33 *Ae. aegypti* isolates were sequenced on the Illumina MiSeq using a 2 x 250bp paired-end configuration. For this study, Puerto Rico will be treated as a separate region from the USA due to its geographical distance from the mainland, although it is a US territory.

Further publicly available samples were downloaded from the National Centre for Biotechnology Information (NCBI) [73]. Illumina genome sequences were selected from a range of geographical regions, including Mexico (n = 16), Brazil (n = 18), USA (n = 50), Thailand (n = 26), Kenya (n = 31), Burkina Faso (n = 36) and Uganda (n = 38). The collection dates of the samples were between 2012 and 2019, although some samples did not have collection dates available. The accession number of these sequences is listed (**Table S1**).

Bioinformatics

The sequence data for the Puerto Rico samples were combined with the 215 publicly available *Ae. aegypti* samples from Mexico, Brazil, USA, Thailand, Kenya, Burkina Faso, and Uganda to make a total of 248 sequences. The fastq reads were trimmed using trimmomatic (v0.39) software [74] and aligned to the NCBI reference (GCF_002204515.2_AaegL5.0) using a bed file to specify exon regions only for time and processing purposes using bowtie2 (v2.5.3) using default parameters [75].

Identification of Insecticide Resistance Associated SNPs and Structural Variants

Variants were called from the aligned files using GATKs HaplotypeCaller (V4.4.0.0) [76]. Once VCFs were created for each sample, they were normalised to remove multi-allelic sites using bcftools and combined into a multi-sample VCF using an in-house pipeline [77]. Further filtering on the final combined VCF was performed using vcftools [78] to remove minor allele frequency of >3, minimum quality of 30 and maximum missingness of 50% per site. Vcftools was used to inspect missing sites per each country population and remove sites with more than 10% missing from the combined VCF.

Gene IDs used for genes associated with insecticide resistance (*vgsc, rdl, GSTe2* and *ace-1)* were LOC5567355, LOC5570466, LOC110676855 and LOC5578456, respectively. These gene regions were used to subset SNPs identified to look for mutations associated with target site resistance. Additionally, a bed file containing a wider array of genes associated with insecticide resistance was created based on a literature search. A search for protein-coding genes involved in insecticide resistance included: "cytochrome", "voltage sodium channel", "sodium channel para", "gamma-aminobutyric", "GABA", "esterase", "glutathione", "cuticle" "G-protein coupled receptor", "salivary", "glucuronosyltransferase", "acetylcholinesterase", "cox", "ATP-binding", "ABC", "odorant-binding". A total of 606 genes in chromosomes 1, 2 and 3 were identified in this manner; these gene locations were used to identify SNPs and SVs of interest. 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 glucuronosyltransferase, among others. The bed file was then applied to the filtered multi-sample VCF using bcftools. The package snpEff (v5.1d) [79] was then used to annotate these variants, using an adapted available database Aedes_aegypti_lvpagwg. This software designates an impact rating of the variant based on the predicted impact on the amino acid by assessing the change to the transcript that alteration occurs in. High-impact variants are designated as a change that has a potential large impact on the subsequent protein for example, alterations including exon deletion, frameshift and gene duplication are deemed high impact. For a full list, see SnpEff documentation [32].

Structural variants (SVs) were investigated in the Puerto Rican samples only due to computation restrictions of mapping the whole genome. DELLY software (v1.1.8) was used to identify structural variants (SVs) [31]. Bcf files were created for each Puerto Rican sample from their bam files using DELLY, which were then merged and filtered to include only calls where they passed quality control. Break-end translocations were also filtered out due to the reliability of the use of short-read data. Filtered SVs were retained for analysis.

Enrichment analysis was performed on a list of genes that has significant results from Tajima's D, iHS and XP-EHH analysis. This analysis was performed with DAVID (Database for Annotation, Visualization and Integrated Discovery) [80,81]. Gene lists were input using Vectorbase IDs and converted to ENTREZ gene IDs with the software. Functional annotation similarities, gene ontology, pathways and protein domain enrichment were investigated, and Benjamini-adjusted p-values were assessed.

To understand further what nucleotide positions under selection corresponded to in the protein, AliView and SWISS-MODEL were used. The sequence XM_021839247.1 for the gene LOC5566402 was translated in AliView [82] to identify the amino acid codon the nucleotide base corresponded to. The amino acid was identified and reflected in the structure of the protein in SWISS-MODEL [83]; this was used for the GABA alpha receptor using model A0A6I8U9X2.

Population Genetic Analysis

A pairwise-genetic distance matrix was generated using PLINK (v1.90b6.21) [84] from the multi-sample VCF and subsequently processed in R using qqman (v0.1.9) [85] and ape (v5.7.1) [86] packages to produce principal component analysis and neighbour joining tree. The tree was visualised and annotated using iTOL. Admixture analysis was carried out using ADMIXTURE (v1.3.0) software [87]. The estimated number of ancestral populations (optimum K-value) was computed through cross-validation (K from 1-10) of eigenvalue decay (k=5) **(Figure S3)**.

Pairwise nucleotide diversity (π) was examined per population, both by site and 100kb window, using vcftools (v0.1.16) [78] to examine differences in populations. Average Tajima's D was calculated using vcftools (v0.1.16) across the genome by population using windows of 100kb to identify areas of selection. Visualisations were created using 5Mb windows **(Figure S1, S6)**. Genetic divergence was investigated with the fixation index statistic (F_{ST}). The statistics were calculated with vcftools (v0.1.16, weir cockerham) per site [78]. Selection was assessed with several metrics, including H12, iHS and XP-EHH. Garud's H_{12} was computed using scikit-allel's (v1.3.7) [88] *moving_garud_h* function using phased biallelic SNPs in windows of 1000 SNPs. The mean of 200 iterations of H12 was plotted. Similarly, iHS was calculated with scikit-allel (allel.ihs function) with the phased biallelic SNPs. The scores calculated were standardized, and p-values were calculated and plotted. XP-EHH was calculated using phased biallelic SNPs using the *allel.xpehh* function in scikit-allel. XP-EHH scores were standardised using the *allel.standardize_by_allele_count* function and plotted.

Data Availability

Whole genome sequence data for the Puerto Rican samples is available on the European Nucleotide Archive under the project number PRJEB76974.

Author Contributions

EC coordinated mosquito collection and rearing in Puerto Rico, with assistance of RM, JQM and the team at the Puerto Rico Vector Control Unit (PRVCU). DNA extraction was carried out by EC. JT downloaded and processed public data. EC designed this study and conducted all bioinformatic analysis of sequence data under the supervision of TGC and SC, and with guidance from HAF, SM, and JP. EC wrote the first draft of this manuscript. All authors reviewed and approved the final manuscript.

Funding

EC and SM are funded by Medical Research Council UK (Grant No. MR/N013638/1). TGC, SC, JT, and HAP are funded by Medical Research Council UK grants (Ref. no. MR/M01360X/1, MR/N010469/1, MR/R025576/1, MR/R020973/1, and MR/X005895/1).

References

1. Agha, S. B. & Tchouassi, D. P. Urbanization of Aedes mosquito populations and evolution of arboviral disease risk in Africa. *Curr Opin Insect Sci* **54**, 100988 (2022).
2. Powell, J. R. & Tabachnick, W. J. History of domestication and spread of Aedes aegypti--a review. *Memórias do Instituto Oswaldo Cruz* vol. 108 Preprint at https://doi.org/10.1590/0074-0276130395 (2013).
3. Crawford, J. E. *et al.* Population genomics reveals that an anthropophilic population of Aedes aegypti mosquitoes in West Africa recently gave rise to American and Asian populations of this major disease vector. *BMC Biol* **15**, 16 (2017).
4. Smith, C. The history of dengue in tropical Asia and its probable relationship to the mosquito Aedes aegypti. *J Trop Med Hyg* (1956).
5. Kraemer, M. U. G. *et al.* Past and future spread of the arbovirus vectors Aedes aegypti and Aedes albopictus. *Nature Microbiology 2019 4:5* **4**, 854–863 (2019).
6. Lessler, J. *et al.* Assessing the Global Threat from Zika Virus. *Science* **353**, aaf8160 (2016).
7. Nsoesie, E. O. *et al.* Global distribution and environmental suitability for chikungunya virus, 1952 to 2015. *Euro Surveill* **21**, (2016).
8. Bhatt, S. *et al.* The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. *Nature* **526**, 207 (2015).
9. IR MAPPER. https://www.irmapper.com/ (2024).
10. Shotkoski, F., Lee, H. -J, Zhang, H. -G, Jackson, M. B. & Ffrench-Constant, R. H. Functional expression of insecticide-resistant GABA receptors from the mosquito Aedes aegypti. *Insect Mol Biol* **3**, 283–287 (1994).
11. Pu, J. & Chung, H. New and emerging mechanisms of insecticide resistance. *Curr Opin Insect Sci* **63**, 101184 (2024).
12. Shotkoski, F., Zhang, H. G., Jackson, M. B. & ffrench-Constant, R. H. Stable expression of insect GABA receptors in insect cell lines. Promoters for efficient expression of Drosophila and mosquito Rdl GABA receptors in stably transformed mosquito cell lines. *FEBS Lett* **380**, 257–262 (1996).
13. Ffrench-Constant, R. H. The Molecular Genetics of Insecticide Resistance. *Genetics* **194**, 807 (2013).
14. Moyes, C. L. *et al.* Contemporary status of insecticide resistance in the major Aedes vectors of arboviruses infecting humans. *PLoS Negl Trop Dis* **11**, (2017).
15. 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**, (2014).
16. Spadar, A., Collins, E., Messenger, L. A., Clark, T. G. & Campino, S. Uncovering the genetic diversity in Aedes aegypti insecticide resistance genes through global comparative genomics. *Scientific Reports 2024 14:1* **14**, 1–12 (2024).
17. Faucon, F. *et al.* In the hunt for genomic markers of metabolic resistance to pyrethroids in the mosquito Aedes aegypti: An integrated next-generation sequencing approach. (2017) doi:10.1371/journal.pntd.0005526.
18. Hemingway, J., Hawkes, N. J., McCarroll, L. & Ranson, H. The molecular basis of insecticide resistance in mosquitoes. *Insect Biochem Mol Biol* **34**, 653–665 (2004).
19. Vontas, J. *et al.* Insecticide resistance in the major dengue vectors Aedes albopictus and Aedes aegypti. *Pestic Biochem Physiol* **104**, 126–131 (2012).
20. David, J. P., Ismail, H. M., Chandor-Proust, A. & Paine, M. J. I. Role of cytochrome P450s in insecticide resistance: Impact on the control of mosquito-borne diseases and use of insecticides on earth. *Philosophical Transactions of the Royal Society B: Biological Sciences* **368**, (2013).
21. Brooke, B. D., Hunt, R. H. & Coetzee, M. Resistance to dieldrin + fipronil assorts with chromosome inversion 2La in the malaria vector Anopheles gambiae. *Med Vet Entomol* **14**, 190–194 (2000).
22. Tchouakui, M. *et al.* A 6.5kb Intergenic Structural Variation Exacerbates the Fitness Cost of P450-Based Metabolic Resistance in the Major African Malaria Vector Anopheles funestus. *Genes (Basel)* **13**, (2022).
23. Redmond, S. N. *et al.* Linked-read sequencing identifies abundant microinversions and introgression in the arboviral vector Aedes aegypti. *BMC Biol* **18**, (2020).
24. Miles, A. *et al.* Genetic diversity of the African malaria vector Anopheles gambiae. *Nature 2017 552:7683* **552**, 96–100 (2017).

25. Hemme, R. R. *et al.* Rapid Screening of Aedes aegypti Mosquitoes for Susceptibility to Insecticides as Part of Zika Emergency Response, Puerto Rico - Volume 25, Number 10—October 2019 - Emerging Infectious Diseases journal - CDC. *Emerg Infect Dis* **25**, 1959–1961 (2019).
26. Collins, E. *et al.* Profiling insecticide resistance phenotypes and genotypes in Aedes aegypti populations across four regions in Puerto Rico. (2024) doi:10.21203/RS.3.RS-3943783/V1.
27. Estep, A. S., Sanscrainte, N. D., Waits, C. M., Louton, J. E. & Becnel, J. J. Resistance Status and Resistance Mechanisms in a Strain of Aedes aegypti (Diptera: Culicidae) From Puerto Rico. *J Med Entomol* **54**, 1643–1648 (2017).
28. Centres for Disease Control and Prevention. Dengue Historic Data (2010-2023). *https://www.cdc.gov/dengue/statistics-maps/historic-data.html* (2024).
29. Sharp, T. M. *et al.* Surveillance for Chikungunya and Dengue During the First Year of Chikungunya Virus Circulation in Puerto Rico. *J Infect Dis* **214**, S475–S481 (2016).
30. Pérez-Guerra, C. L. *et al.* Community perceptions on challenges and solutions to implement an Aedes aegypti control project in Ponce, Puerto Rico (USA). *PLoS One* **18**, (2023).
31. Rausch, T. *et al.* DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**, i333–i339 (2012).
32. Input & output files - SnpEff & SnpSift. https://pcingola.github.io/SnpEff/snpeff/inputoutput/.
33. Li, T. & Liu, N. Role of the G-Protein-Coupled Receptor Signaling Pathway in Insecticide Resistance. *Int J Mol Sci* **20**, (2019).
34. Muthu Lakshmi Bavithra, C., Murugan, M., Pavithran, S. & Naveena, K. Enthralling genetic regulatory mechanisms meddling insecticide resistance development in insects: role of transcriptional and post-transcriptional events. *Front Mol Biosci* **10**, 1257859 (2023).
35. Ponce-García, G. *et al.* Knockdown resistance mutations in Aedes aegypti (Diptera: Culicidae) From puerto rico. *J Med Entomol* **53**, 1410–1414 (2016).
36. Tajima, F. Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. (1989).
37. Ye, M. *et al.* The Role of Insect Cytochrome P450s in Mediating Insecticide Resistance. *Agriculture 2022, Vol. 12, Page 53* **12**, 53 (2022).
38. Epelboin, Y. *et al.* CYP450 core involvement in multiple resistance strains of Aedes aegypti from French Guiana highlighted by proteomics, molecular and biochemical studies. *PLoS One* **16**, (2021).
39. 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).
40. ffrench-Constant, R. H., Williamson, M. S., Davies, T. G. E. & Bass, C. Ion channels as insecticide targets. *J Neurogenet* **30**, 163–177 (2016).
41. Taylor-Wells, J. & Jones, A. K. Variations in the Insect GABA Receptor, RDL, and Their Impact on Receptor Pharmacology.
42. Hosie, A. M., Aronstein, K., Sattelle, D. B. & Ffrench-Constant, R. H. Molecular biology of insect neuronal GABA receptors. *Trends Neurosci* **20**, 578–583 (1997).
43. Brown, J. E. *et al.* Human impacts have shaped historical and recent evolution in Aedes aegypti, the dengue and yellow fever mosquito. *Evolution* **68**, 514–525 (2014).
44. Gloria-Soria, A. *et al.* Global Genetic Diversity of Aedes aegypti. *Mol Ecol* **25**, 5377 (2016).
45. Government of Puerto Rico Planning Board. *Puerto Rico External Trade Statistics 2020*. (2021).
46. Pérez-Guerra, C. L. *et al.* Acceptability of emergent Aedes aegypti vector control methods in Ponce, Puerto Rico: A qualitative assessment. *PLOS Global Public Health* **4**, e0002744 (2024).
47. Current Year Data (2024) | Dengue | CDC. https://www.cdc.gov/dengue/data-research/facts-stats/current-data.html.
48. World Health Organisation - African Region. *Country Disease Outlook - Uganda*. (2023).
49. World Health Organization - African Region. *Country Disease Outlook - Kenya*. (2023).
50. Im, J. *et al.* The epidemiology of dengue outbreaks in 2016 and 2017 in Ouagadougou, Burkina Faso. *Heliyon* **6**, e04389 (2020).
51. Tarnagda, Z. *et al.* Dengue Fever in Burkina Faso, 2016. *Emerg Infect Dis* **24**, 170 (2018).
52. Amelia-Yap, Z. H., Chen, C. D., Sofian-Azirun, M. & Low, V. L. Pyrethroid resistance in the dengue vector Aedes aegypti in Southeast Asia: Present situation and prospects for management. *Parasit Vectors* **11**, 1–17 (2018).

53. Kampango, A. *et al.* High DDT resistance without apparent association to kdr and Glutathione-S-transferase (GST) gene mutations in Aedes aegypti population at hotel compounds in Zanzibar. *PLoS Negl Trop Dis* **16**, (2022).
54. Helvecio, E. *et al.* Polymorphisms in GSTE2 is associated with temephos resistance in Aedes aegypti. *Pestic Biochem Physiol* **165**, (2020).
55. Li, T. & Liu, N. Role of the G-Protein-Coupled Receptor Signaling Pathway in Insecticide Resistance. *Int J Mol Sci* **20**, 4300 (2019).
56. Liu, Y. X., Hu, C., Li, Y. T., Gao, P. & Yang, X. Q. Identification of G Protein-Coupled Receptors (GPCRs) Associated with Lambda-Cyhalothrin Detoxification in Cydia pomonella. *J Agric Food Chem* **72**, 363–377 (2024).
57. Campbell, C. L. *et al.* Vgsc-interacting proteins are genetically associated with pyrethroid resistance in Aedes aegypti. *PLoS One* **14**, e0211497 (2019).
58. Dusfour, I. *et al.* Deltamethrin Resistance Mechanisms in Aedes aegypti Populations from Three French Overseas Territories Worldwide. *PLoS Negl Trop Dis* **9**, (2015).
59. Faucon, F. *et al.* Identifying genomic changes associated with insecticide resistance in the dengue mosquito Aedes aegypti by deep targeted sequencing. *Genome Res* **25**, 1347 (2015).
60. Epelboin, Y. *et al.* CYP450 core involvement in multiple resistance strains of Aedes aegypti from French Guiana highlighted by proteomics, molecular and biochemical studies. *PLoS One* **16**, e0243992 (2021).
61. Derilus, D. *et al.* Comparative Transcriptomic Analysis of Insecticide-Resistant Aedes aegypti from Puerto Rico Reveals Insecticide-Specific Patterns of Gene Expression. *Genes (Basel)* **14**, 1626 (2023).
62. Smith, L. B., Silva, J. J., Chen, C., Harrington, L. C. & Scott, J. G. Fitness costs of individual and combined pyrethroid resistance mechanisms, kdr and CYP-mediated detoxification, in Aedes aegypti. *PLoS Negl Trop Dis* **15**, e0009271 (2021).
63. Bibi, M. *et al.* In Silico Characterisation of the Aedes aegypti Gustatory Receptors. *Int J Mol Sci* **24**, 12263 (2023).
64. Lee, Y., Kim, S. H. & Montell, C. Avoiding DEET through insect gustatory receptors. *Neuron* **67**, 555 (2010).
65. Jacobs, E. *et al.* Cuticular profiling of insecticide resistant Aedes aegypti. *Scientific Reports 2023 13:1* **13**, 1–15 (2023).
66. Shakeel, M., Xu, X., De Mandal, S. & Jin, F. Role of serine protease inhibitors in insect-host-pathogen interactions. *Arch Insect Biochem Physiol* **102**, (2019).
67. Zhou, Y. *et al.* UDP-glycosyltransferase genes and their association and mutations associated with pyrethroid resistance in Anopheles sinensis (Diptera: Culicidae). *Malar J* **18**, 62 (2019).
68. Elliott, K., Caicedo, P. A., Haunerland, N. H. & Lowenberger, C. Profiling lipidomic changes in dengue-resistant and dengue-susceptible strains of Colombian Aedes aegypti after dengue virus challenge. *PLoS Negl Trop Dis* **17**, (2023).
69. Enayati, A. A., Ranson, H. & Hemingway, J. Insect glutathione transferases and insecticide resistance. *Insect Mol Biol* **14**, 3–8 (2005).
70. Insecticide Resistance Action Committee. *MODE OF ACTION CLASSIFICATION SCHEME*. https://irac-online.org/documents/moa-classification/ (2024).
71. CG8916 orthologs - NCBI. https://www.ncbi.nlm.nih.gov/gene/32553/ortholog/?scope=6656.
72. Huang, Q. T. *et al.* Functional Characteristics of the Lepidopteran Ionotropic GABA Receptor 8916 Subunit Interacting with the LCCH3 or the RDL Subunit. *J Agric Food Chem* **69**, 11582–11591 (2021).
73. NCBI. SRA - NCBI. https://www.ncbi.nlm.nih.gov/sra/ (2020).
74. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
75. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, 1–10 (2009).
76. McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297–1303 (2010).
77. fastq2matrix/scripts/merge_vcfs.py at master · LSHTMPathogenSeqLab/fastq2matrix · GitHub. https://github.com/LSHTMPathogenSeqLab/fastq2matrix/blob/master/scripts/merge_vcfs.py.
78. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
79. 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 (2012).

80. Sherman, B. T. *et al.* DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res* **50**, W216–W221 (2022).
81. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44–57 (2009).
82. Larsson, A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30**, 3276–3278 (2014).
83. Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* **46**, W296–W303 (2018).
84. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet* **81**, 559 (2007).
85. Turner, S. D. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *J Open Source Softw* **3**, 731 (2018).
86. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
87. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19**, 1655–1664 (2009).
88. Miles, A. *et al.* scikit-allel. Preprint at https://doi.org/https://doi.org/10.5281/zenodo (2024).

Supplementary Information

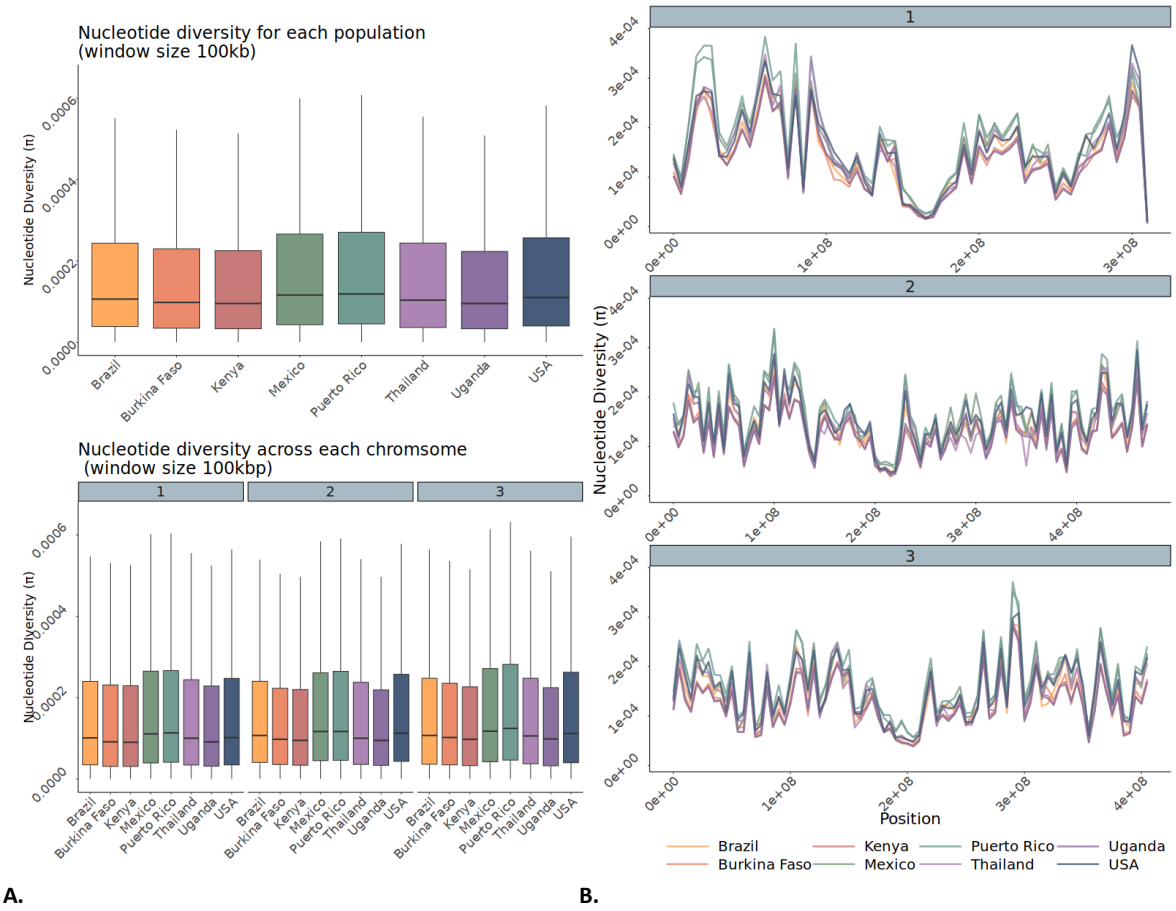
Table S1. *Accession numbers for publicly available sequences used*

ID	Country	Region	Subregion
SRR11006835	Brazil	South America	Brazil: Santarem
SRR11006836	Brazil	South America	Brazil: Santarem
SRR11006837	Brazil	South America	Brazil: Santarem
SRR11006838	Brazil	South America	Brazil: Santarem
SRR11006839	Brazil	South America	Brazil: Santarem
SRR11006840	Brazil	South America	Brazil: Santarem
SRR11006841	Brazil	South America	Brazil: Santarem
SRR11006842	Brazil	South America	Brazil: Santarem
SRR11006843	Brazil	South America	Brazil: Santarem
SRR11006846	Brazil	South America	Brazil: Santarem
SRR11006847	Brazil	South America	Brazil: Santarem
SRR11006848	Brazil	South America	Brazil: Santarem
SRR11006849	Brazil	South America	Brazil: Santarem
SRR11006850	Brazil	South America	Brazil: Santarem
SRR11006851	Brazil	South America	Brazil: Santarem
SRR11006852	Brazil	South America	Brazil: Santarem
SRR11006853	Brazil	South America	Brazil: Santarem
SRR11006854	Brazil	South America	Brazil: Santarem
SRR11006514	Burkina Faso	Africa	Burkina Faso: Ouahigouya
SRR11006515	Burkina Faso	Africa	Burkina Faso: Ouahigouya
SRR11006516	Burkina Faso	Africa	Burkina Faso: Ouahigouya
SRR11006517	Burkina Faso	Africa	Burkina Faso: Ouahigouya
SRR11006518	Burkina Faso	Africa	Burkina Faso: Ouahigouya
SRR11006519	Burkina Faso	Africa	Burkina Faso: Ouahigouya
SRR11006521	Burkina Faso	Africa	Burkina Faso: Ouahigouya
SRR11006522	Burkina Faso	Africa	Burkina Faso: Ouahigouya
SRR11006523	Burkina Faso	Africa	Burkina Faso: Ouahigouya
SRR11006524	Burkina Faso	Africa	Burkina Faso: Ouahigouya
SRR11006525	Burkina Faso	Africa	Burkina Faso: Ouahigouya
SRR11006526	Burkina Faso	Africa	Burkina Faso: Ouahigouya
SRR11006527	Burkina Faso	Africa	Burkina Faso: Ouahigouya
SRR11006528	Burkina Faso	Africa	Burkina Faso: Ouahigouya
SRR11006529	Burkina Faso	Africa	Burkina Faso: Ouahigouya
SRR11006530	Burkina Faso	Africa	Burkina Faso: Ouahigouya
SRR11006612	Burkina Faso	Africa	Burkina Faso: Ouagadougou
SRR11006613	Burkina Faso	Africa	Burkina Faso: Ouagadougou
SRR11006614	Burkina Faso	Africa	Burkina Faso: Ouagadougou
SRR11006615	Burkina Faso	Africa	Burkina Faso: Ouagadougou
SRR11006616	Burkina Faso	Africa	Burkina Faso: Ouagadougou
SRR11006617	Burkina Faso	Africa	Burkina Faso: Ouagadougou
SRR11006618	Burkina Faso	Africa	Burkina Faso: Ouagadougou
SRR11006619	Burkina Faso	Africa	Burkina Faso: Ouagadougou
SRR11006621	Burkina Faso	Africa	Burkina Faso: Ouagadougou
SRR11006622	Burkina Faso	Africa	Burkina Faso: Ouagadougou
SRR11006623	Burkina Faso	Africa	Burkina Faso: Ouagadougou
SRR11006624	Burkina Faso	Africa	Burkina Faso: Ouagadougou
SRR11006625	Burkina Faso	Africa	Burkina Faso: Ouagadougou
SRR11006626	Burkina Faso	Africa	Burkina Faso: Ouagadougou
SRR11006627	Burkina Faso	Africa	Burkina Faso: Ouagadougou
SRR11006628	Burkina Faso	Africa	Burkina Faso: Ouagadougou
SRR11006629	Burkina Faso	Africa	Burkina Faso: Ouagadougou
SRR11006630	Burkina Faso	Africa	Burkina Faso: Ouagadougou
SRR11006632	Burkina Faso	Africa	Burkina Faso: Ouagadougou
SRR11006633	Burkina Faso	Africa	Burkina Faso: Ouagadougou

SRR11006686	Kenya	Africa	Kenya: Kakamega
SRR11006687	Kenya	Africa	Kenya: Kakamega
SRR11006689	Kenya	Africa	Kenya: Kakamega
SRR11006690	Kenya	Africa	Kenya: Kakamega
SRR11006691	Kenya	Africa	Kenya: Kakamega
SRR11006692	Kenya	Africa	Kenya: Kakamega
SRR11006694	Kenya	Africa	Kenya: Kakamega
SRR11006695	Kenya	Africa	Kenya: Kakamega
SRR11006696	Kenya	Africa	Kenya: Kakamega
SRR11006699	Kenya	Africa	Kenya: Kakamega
SRR11006701	Kenya	Africa	Kenya: Kakamega
SRR11006703	Kenya	Africa	Kenya: Kakamega
SRR11006706	Kenya	Africa	Kenya: Kakamega
SRR11006887	Kenya	Africa	Kenya: Virhembe
SRR11006888	Kenya	Africa	Kenya: Virhembe
SRR11006890	Kenya	Africa	Kenya: Virhembe
SRR11006891	Kenya	Africa	Kenya: Virhembe
SRR11006892	Kenya	Africa	Kenya: Virhembe
SRR11006893	Kenya	Africa	Kenya: Virhembe
SRR11006894	Kenya	Africa	Kenya: Virhembe
SRR11006895	Kenya	Africa	Kenya: Virhembe
SRR11006896	Kenya	Africa	Kenya: Virhembe
SRR11006897	Kenya	Africa	Kenya: Virhembe
SRR11006898	Kenya	Africa	Kenya: Virhembe
SRR11006899	Kenya	Africa	Kenya: Virhembe
SRR11006901	Kenya	Africa	Kenya: Virhembe
SRR11006902	Kenya	Africa	Kenya: Virhembe
SRR11006903	Kenya	Africa	Kenya: Virhembe
SRR11006904	Kenya	Africa	Kenya: Virhembe
SRR11006905	Kenya	Africa	Kenya: Virhembe
SRR11006906	Kenya	Africa	Kenya: Virhembe
SRR11196610	Mexico	North America	Mexico: Tapachula
SRR11196611	Mexico	North America	Mexico: Tapachula
SRR11196612	Mexico	North America	Mexico: Tapachula
SRR11196613	Mexico	North America	Mexico: Tapachula
SRR11196614	Mexico	North America	Mexico: Tapachula
SRR11196615	Mexico	North America	Mexico: Tapachula
SRR11196643	Mexico	North America	Mexico: Tapachula
SRR11196644	Mexico	North America	Mexico: Tapachula
SRR11196645	Mexico	North America	Mexico: Tapachula
SRR11196646	Mexico	North America	Mexico: Tapachula
SRR11196647	Mexico	North America	Mexico: Tapachula
SRR11196648	Mexico	North America	Mexico: Tapachula
SRR11196649	Mexico	North America	Mexico: Tapachula
SRR11196650	Mexico	North America	Mexico: Tapachula
SRR11196651	Mexico	North America	Mexico: Tapachula
SRR11196652	Mexico	North America	Mexico: Tapachula
SRR11006487	Thailand	Asia	Thailand: Bangkok
SRR11006598	Thailand	Asia	Thailand: Bangkok
SRR11006660	Thailand	Asia	Thailand: Bangkok
SRR11006671	Thailand	Asia	Thailand: Bangkok
SRR11006682	Thailand	Asia	Thailand: Bangkok
SRR11006693	Thailand	Asia	Thailand: Bangkok
SRR11006704	Thailand	Asia	Thailand: Bangkok
SRR11006845	Thailand	Asia	Thailand: Bangkok
SRR11006856	Thailand	Asia	Thailand: Bangkok
SRR11006867	Thailand	Asia	Thailand: Bangkok
SRR11006878	Thailand	Asia	Thailand: Bangkok
SRR11006889	Thailand	Asia	Thailand: Bangkok

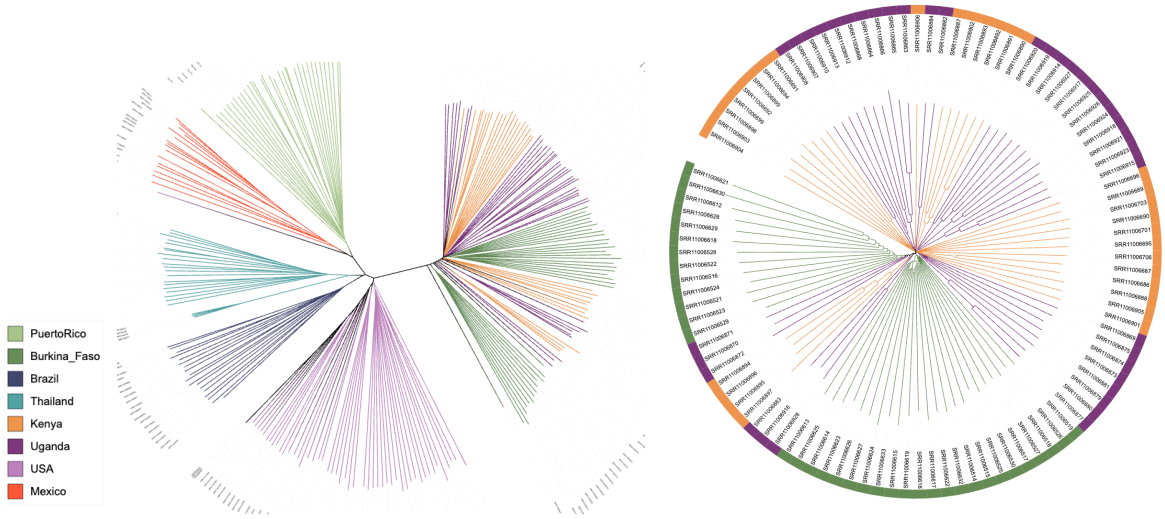
SRR11006900	Thailand	Asia	Thailand: Bangkok
SRR11006911	Thailand	Asia	Thailand: Bangkok
SRR11006922	Thailand	Asia	Thailand: Bangkok
SRR11006933	Thailand	Asia	Thailand: Bangkok
SRR11006944	Thailand	Asia	Thailand: Bangkok
SRR11006955	Thailand	Asia	Thailand: Bangkok
SRR11006956	Thailand	Asia	Thailand: Bangkok
SRR11006957	Thailand	Asia	Thailand: Bangkok
SRR1548721	Thailand	Asia	Thailand: Pai Lom
SRR1548726	Thailand	Asia	Thailand: Pai Lom
SRR1548727	Thailand	Asia	Thailand: Pai Lom
SRR1548728	Thailand	Asia	Thailand: Pai Lom
SRR9959052	Thailand	Asia	Thailand: Chiang Mai
SRR9986057	Thailand	Asia	Thailand: Chiang Mai
SRR11006863	Uganda	Africa	Uganda: Kichwamba
SRR11006864	Uganda	Africa	Uganda: Kichwamba
SRR11006865	Uganda	Africa	Uganda: Kichwamba
SRR11006866	Uganda	Africa	Uganda: Kichwamba
SRR11006868	Uganda	Africa	Uganda: Kichwamba
SRR11006869	Uganda	Africa	Uganda: Entebbe
SRR11006870	Uganda	Africa	Uganda: Entebbe
SRR11006871	Uganda	Africa	Uganda: Entebbe
SRR11006872	Uganda	Africa	Uganda: Entebbe
SRR11006873	Uganda	Africa	Uganda: Entebbe
SRR11006874	Uganda	Africa	Uganda: Entebbe
SRR11006875	Uganda	Africa	Uganda: Entebbe
SRR11006877	Uganda	Africa	Uganda: Entebbe
SRR11006879	Uganda	Africa	Uganda: Entebbe
SRR11006880	Uganda	Africa	Uganda: Entebbe
SRR11006881	Uganda	Africa	Uganda: Entebbe
SRR11006882	Uganda	Africa	Uganda: Karenga
SRR11006883	Uganda	Africa	Uganda: Karenga
SRR11006884	Uganda	Africa	Uganda: Karenga
SRR11006907	Uganda	Africa	Uganda: Kichwamba
SRR11006908	Uganda	Africa	Uganda: Kichwamba
SRR11006910	Uganda	Africa	Uganda: Kichwamba
SRR11006912	Uganda	Africa	Uganda: Kichwamba
SRR11006913	Uganda	Africa	Uganda: Kichwamba
SRR11006914	Uganda	Africa	Uganda: Bundibugyo
SRR11006915	Uganda	Africa	Uganda: Bundibugyo
SRR11006916	Uganda	Africa	Uganda: Bundibugyo
SRR11006917	Uganda	Africa	Uganda: Bundibugyo
SRR11006918	Uganda	Africa	Uganda: Bundibugyo
SRR11006919	Uganda	Africa	Uganda: Bundibugyo
SRR11006920	Uganda	Africa	Uganda: Bundibugyo
SRR11006921	Uganda	Africa	Uganda: Bundibugyo
SRR11006923	Uganda	Africa	Uganda: Bundibugyo
SRR11006924	Uganda	Africa	Uganda: Bundibugyo
SRR11006925	Uganda	Africa	Uganda: Bundibugyo
SRR11006926	Uganda	Africa	Uganda: Bundibugyo
SRR11006927	Uganda	Africa	Uganda: Bundibugyo
SRR11006928	Uganda	Africa	Uganda: Bundibugyo
SRR11100105	USA	North America	USA: Chula Vista
SRR11100109	USA	North America	USA: Calexico
SRR11100110	USA	North America	USA: Calexico
SRR11100111	USA	North America	USA: Fowler
SRR11100112	USA	North America	USA: El Centro
SRR11100113	USA	North America	USA: El Centro
SRR11100115	USA	North America	USA: Brawley/Calipatria

SRR11100119	USA	North America	USA: Cathedral City/Palm Desert
SRR11100120	USA	North America	USA: Mission Viejo
SRR11100121	USA	North America	USA: Mission Viejo
SRR11100136	USA	North America	USA: La Palma/Stanton
SRR11100137	USA	North America	USA: Riverside
SRR11100138	USA	North America	USA: Riverside
SRR11100139	USA	North America	USA: Riverside
SRR11100141	USA	North America	USA: La Habra
SRR11100142	USA	North America	USA: La Habra
SRR11100146	USA	North America	USA: Brea
SRR11100147	USA	North America	USA: Brea
SRR11100148	USA	North America	USA: Downey
SRR11100149	USA	North America	USA: Downey
SRR11100168	USA	North America	USA: Alhambra
SRR11100169	USA	North America	USA: Alhambra
SRR11100170	USA	North America	USA: Alhambra
SRR11100171	USA	North America	USA: Arvin
SRR11100172	USA	North America	USA: Arvin
SRR11100173	USA	North America	USA: Exeter
SRR11100174	USA	North America	USA: Corcoran
SRR11100176	USA	North America	USA: Tulare
SRR11100177	USA	North America	USA: Kerman
SRR11100178	USA	North America	USA: Lenmoore
SRR11100203	USA	North America	USA: Fresno
SRR11237929	USA	North America	USA: Fresno/Clovis CA
SRR11237930	USA	North America	USA: Fresno/Clovis CA
SRR11237931	USA	North America	USA: Fresno/Clovis CA
SRR11237933	USA	North America	USA: Fresno/Clovis CA
SRR11237935	USA	North America	USA: Fresno/Clovis CA
SRR11237936	USA	North America	USA: Fresno/Clovis CA
SRR11237937	USA	North America	USA: Fresno/Clovis CA
SRR11237938	USA	North America	USA: Fresno/Clovis CA
SRR11237939	USA	North America	USA: Fresno/Clovis CA
SRR11237955	USA	North America	USA: Fresno/Clovis CA
SRR11237958	USA	North America	USA: Fresno/Clovis CA
SRR11237959	USA	North America	USA: Fresno/Clovis CA
SRR11237960	USA	North America	USA: Fresno/Clovis CA
SRR11237961	USA	North America	USA: Fresno/Clovis CA
SRR11237963	USA	North America	USA: Fresno/Clovis CA
SRR11237964	USA	North America	USA: Fresno/Clovis CA
SRR11237965	USA	North America	USA: Fresno/Clovis CA
SRR11237966	USA	North America	USA: Fresno/Clovis CA
SRR11237971	USA	North America	USA: Fresno/Clovis CA



A. **B.**

Figure S1. Nucleotide Diversity Across Chromosomes and Populations A. *Nucleotide diversity across 100kb windows per population averaged and per chromosome (1, 2, 3).* ***B.*** *Nucleotide diversity across 5Mb windows for all eight populations per chromosome (1, 2, 3)*



A. **B.**

Figure S2. Phylogenetic and Genetic Diversity Analyses

A. An unrooted phylogenetic tree containing all samples showing African samples on a single branch. **B.** Phylogenetic tree with just African samples included in this study, with minimal branching observed. The African population was identified as less structured, possibly due to higher levels of gene flow. These results were reiterated by PCA analysis, admixture analysis and nucleotide diversity analysis. The nucleotide diversity indicated the least diversity in the African countries. While this trend cannot be explained by the disparity in the

number of SNPs, with African countries having more than other countries (Uganda; 1569841, Thailand; 785912, Mexico; 759572, Puerto Rico; 1010449, Kenya; 1518393, Burkina Faso; 1648271, USA; 1074961).

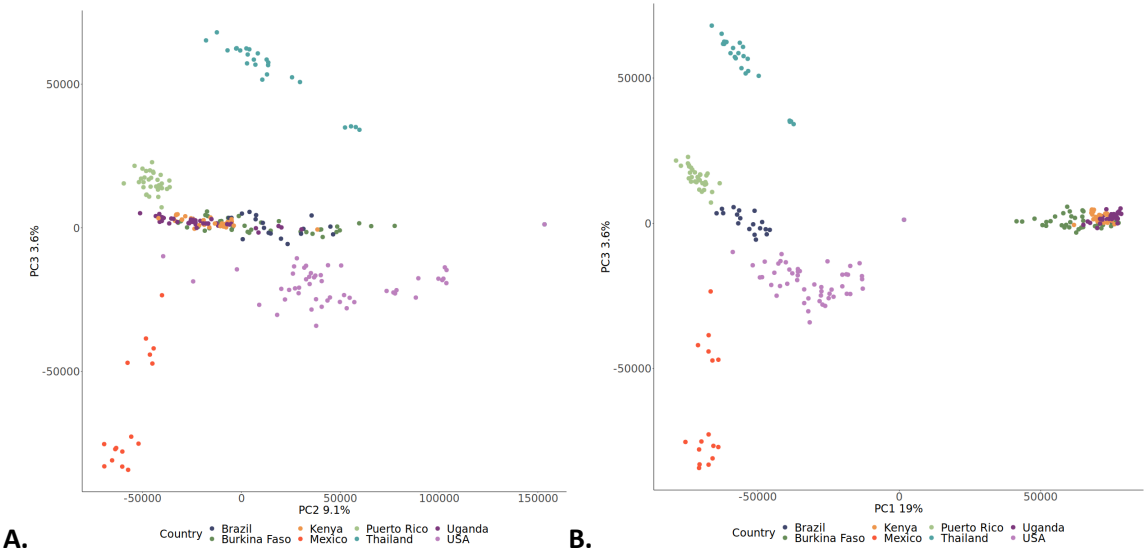


Figure S3. Principal Component Analysis Across Samples
A. Principal component analysis component 2 vs component 3 **B.** Principal component analysis component 1 vs component 3 for all autosomes for all samples.

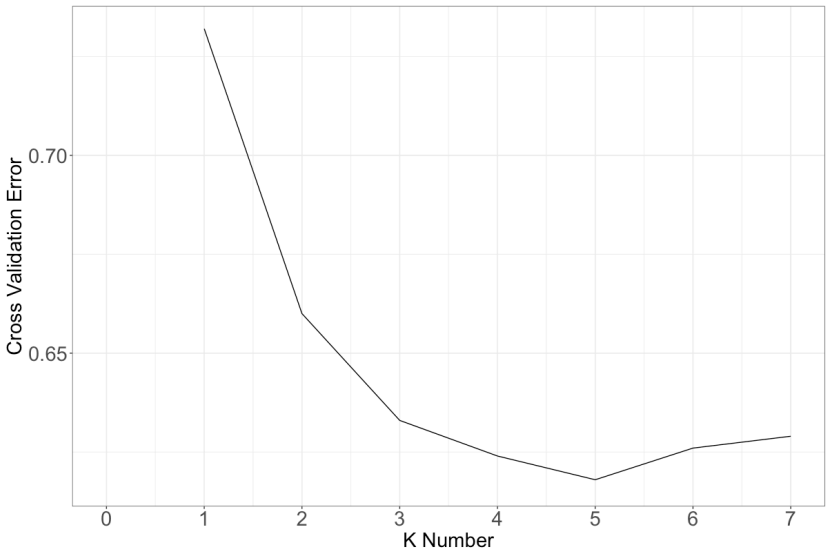


Figure S4. Cross-validation plot for K number for ancestry analysis

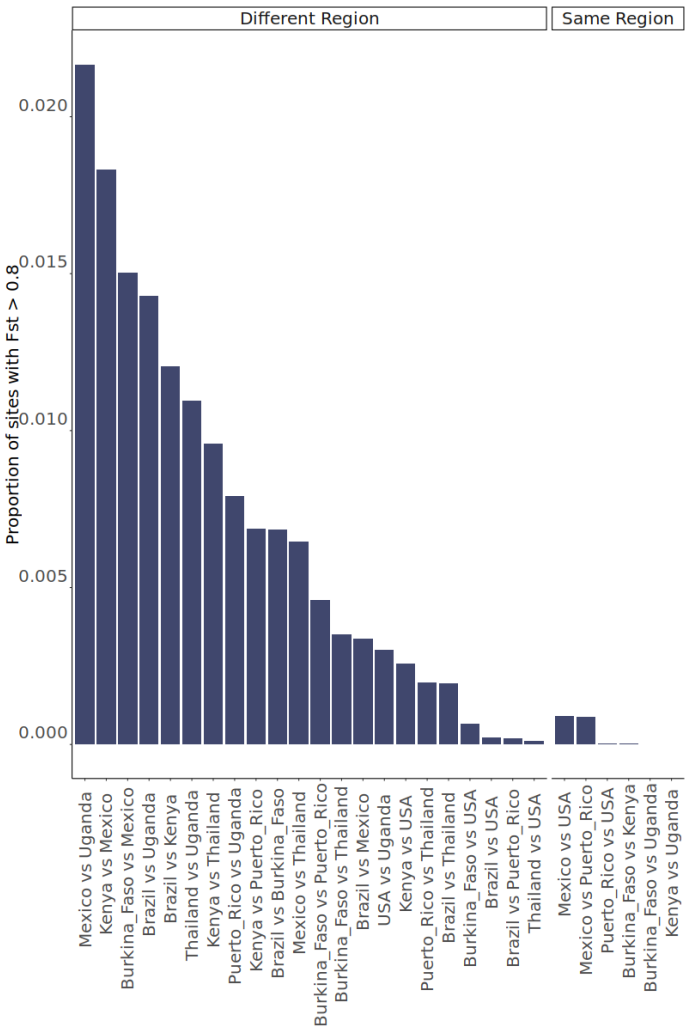


Figure S5. The proportion of sites with F_{ST} > 0.8 between different countries, in the same and across different regions.

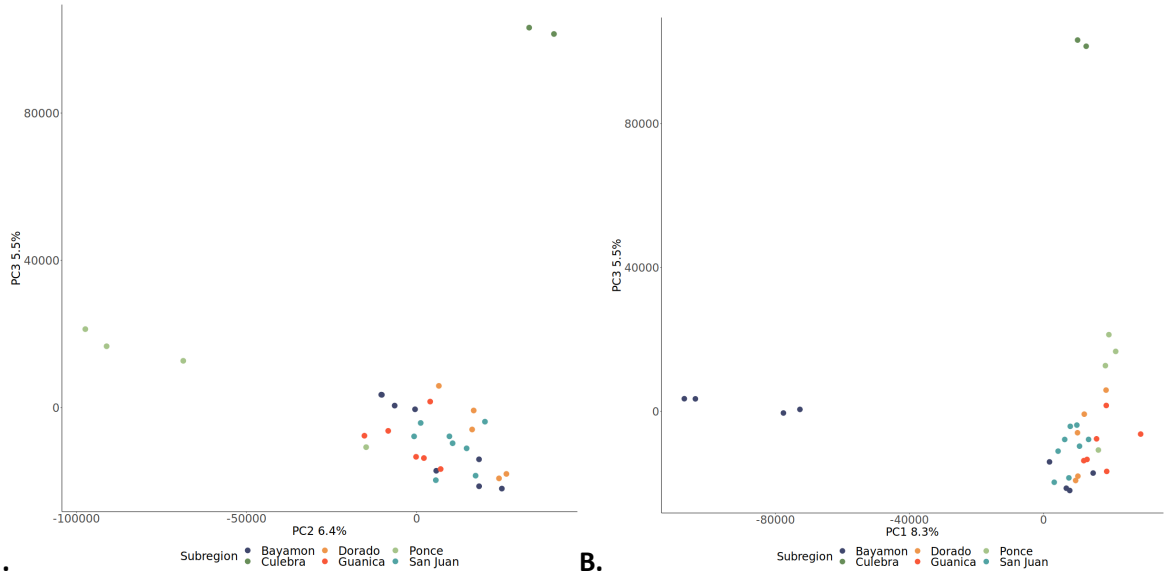


Figure S6. Principal Component Analysis Across Samples in Puerto Rico.
A. *Principal component analysis component 2 vs component 3* ***B.*** *Principal component analysis component 1 vs component 3 for all autosomes for the Puerto Rican sample.*

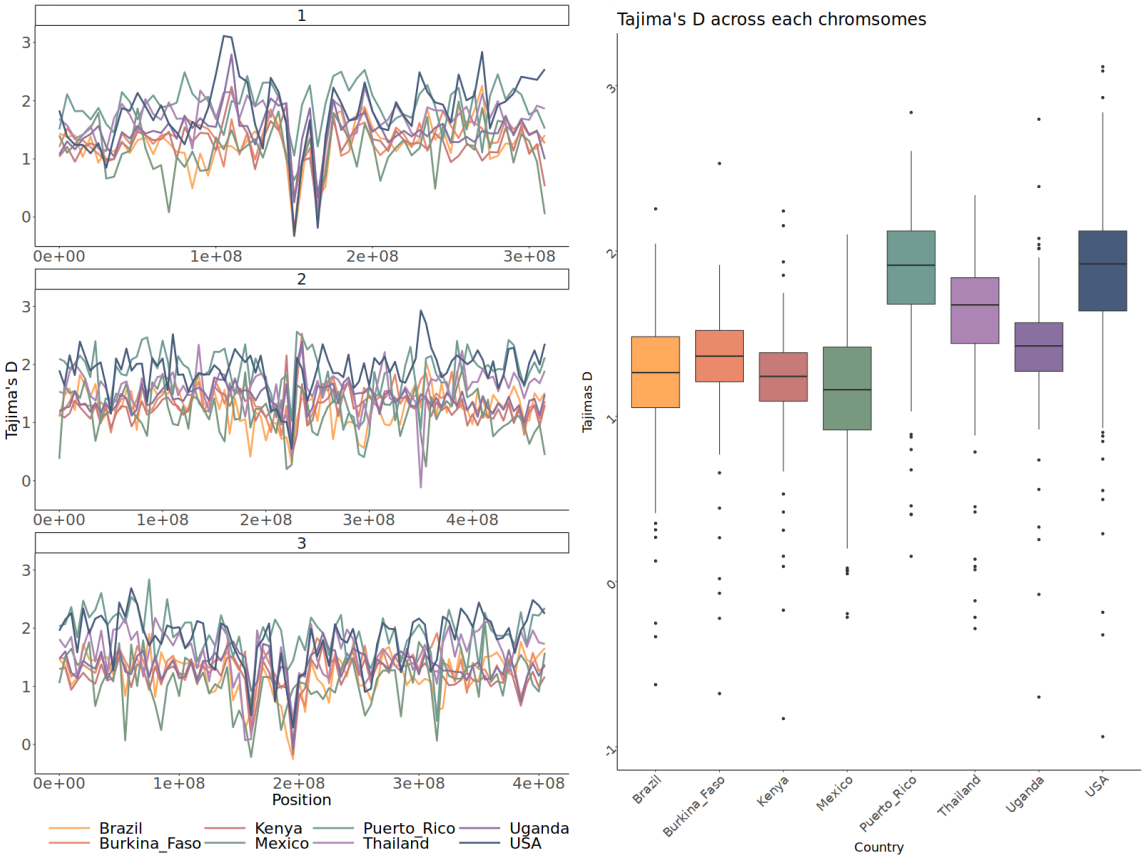


Figure S7. Tajima's D Analysis Across Populations and Chromosomes
A. *T_D by country population and per chromosome across 5Mb windows* **B.** *Distribution of T_D metric by country.*

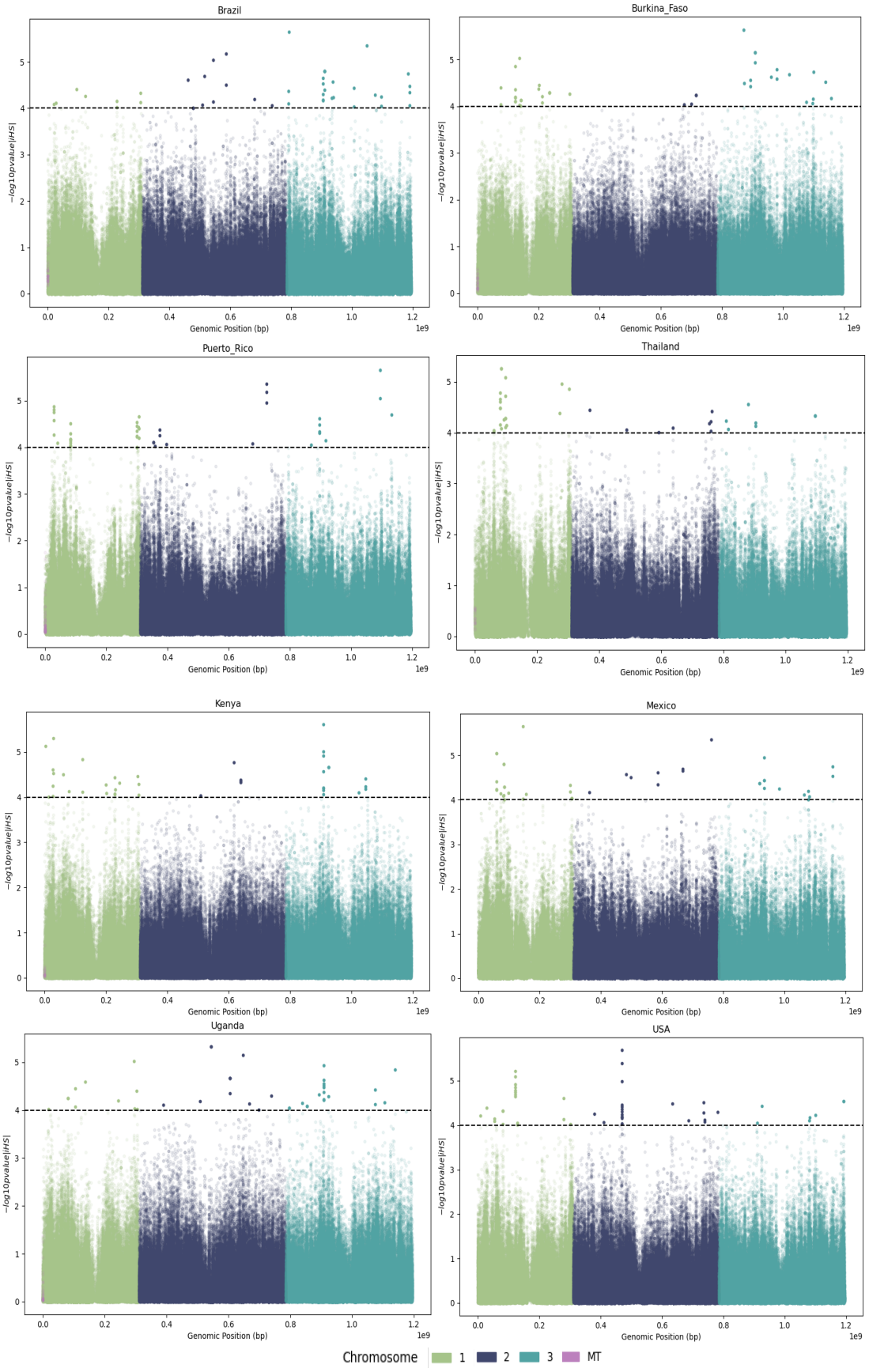
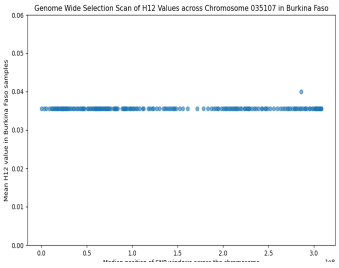
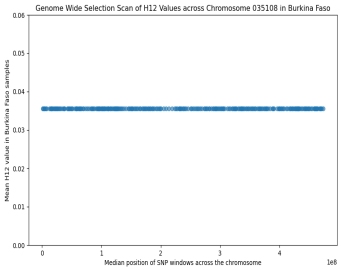
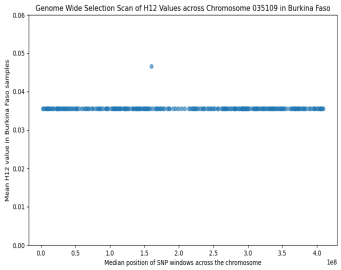
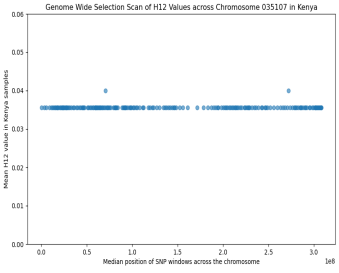


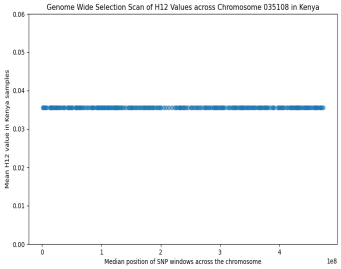
Figure S8. ***iHS plots per chromosome for each population.***

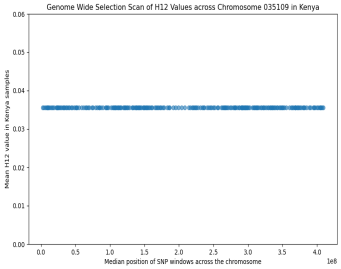

Genome Wide Selection Scan of H12 Values across Chromosome 035107 in Burkina Faso
Mean H12 value in Burkina Faso samples
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.0
0.5
1.0
1.5
2.0
2.5
3.0

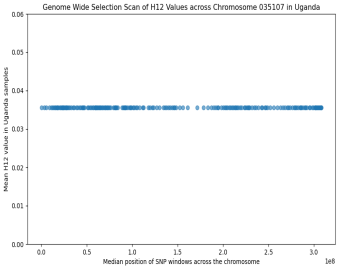

Genome Wide Selection Scan of H12 Values across Chromosome 035108 in Burkina Faso
Mean H12 value in Burkina Faso samples
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0
1
2
3
4
Median position of SNP windows across the chromosome
1e8

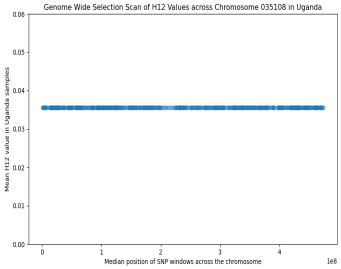

Genome Wide Selection Scan of H12 Values across Chromosome 035109 in Burkina Faso
Mean H12 value in Burkina Faso samples
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Median position of SNP windows across the chromosome
1e8

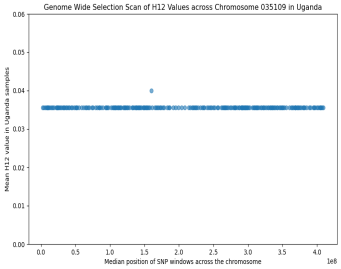

Genome Wide Selection Scan of H12 Values across Chromosome 035107 in Kenya
Mean H12 value in Kenya samples
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Median position of SNP windows across the chromosome
1e8

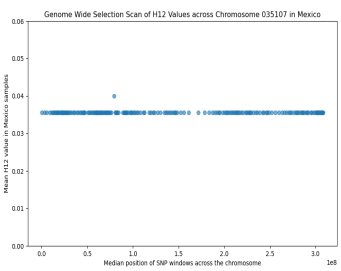

Genome Wide Selection Scan of H12 Values across Chromosome 035108 in Kenya
Mean H12 value in Kenya samples
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0
1
2
3
4
Median position of SNP windows across the chromosome
1e8

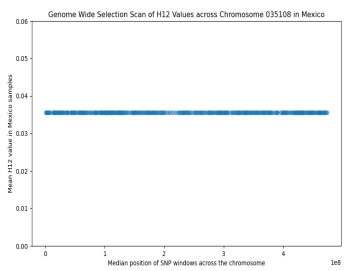

Genome Wide Selection Scan of H12 Values across Chromosome 035109 in Kenya
Mean H12 value in Kenya samples
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Median position of SNP windows across the chromosome
1e8

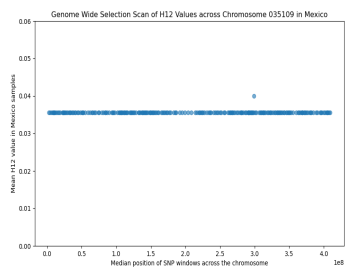

Genome Wide Selection Scan of H12 Values across Chromosome 035107 in Uganda
Mean H12 value in Uganda samples
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Median position of SNP windows across the chromosome
1e8

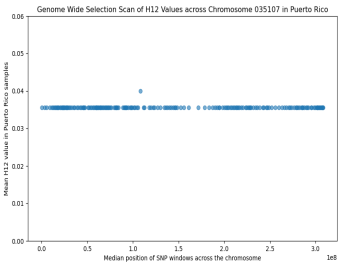

Genome Wide Selection Scan of H12 Values across Chromosome 035108 in Uganda
Mean H12 value in Uganda samples
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0
1
2
3
4
Median position of SNP windows across the chromosome
1e8

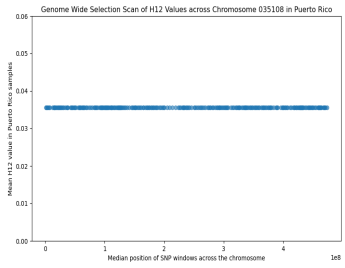

Genome Wide Selection Scan of H12 Values across Chromosome 035109 in Uganda
Mean H12 value in Uganda samples
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Median position of SNP windows across the chromosome
1e8

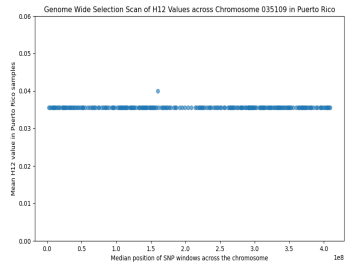

Genome Wide Selection Scan of H12 Values across Chromosome 035107 in Mexico
Mean H12 value in Mexico samples
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Median position of SNP windows across the chromosome
1e8

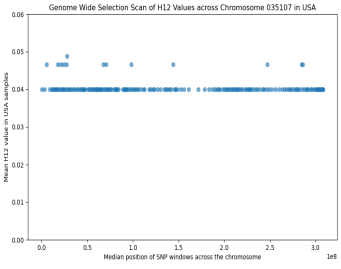

Genome Wide Selection Scan of H12 Values across Chromosome 035108 in Mexico
Mean H12 value in Mexico samples
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0
1
2
3
4
Median position of SNP windows across the chromosome
1e8

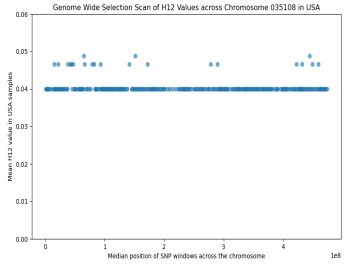

Genome Wide Selection Scan of H12 Values across Chromosome 035109 in Mexico
Mean H12 value in Mexico samples
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Median position of SNP windows across the chromosome
1e8


Genome Wide Selection Scan of H12 Values across Chromosome 035107 in Puerto Rico
Mean H12 value in Puerto Rico samples
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Median position of SNP windows across the chromosome
1e8


Genome Wide Selection Scan of H12 Values across Chromosome 035108 in Puerto Rico
Mean H12 value in Puerto Rico samples
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0
1
2
3
4
Median position of SNP windows across the chromosome
1e8


Genome Wide Selection Scan of H12 Values across Chromosome 035109 in Puerto Rico
Mean H12 value in Puerto Rico samples
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Median position of SNP windows across the chromosome
1e8


Genome Wide Selection Scan of H12 Values across Chromosome 035107 in USA
Mean H12 value in USA samples
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Median position of SNP windows across the chromosome
1e8


Genome Wide Selection Scan of H12 Values across Chromosome 035108 in USA
Mean H12 value in USA samples
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0
1
2
3
4
Median position of SNP windows across the chromosome
1e8


Genome Wide Selection Scan of H12 Values across Chromosome 035109 in USA
Mean H12 value in USA samples
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Median position of SNP windows across the chromosome
1e8

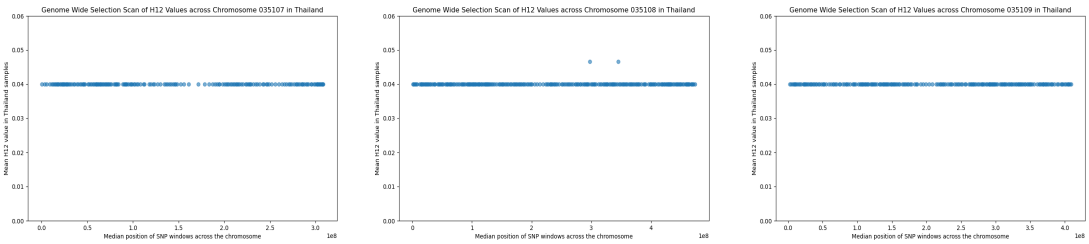


Figure S9. *Garud's H12 across each chromosome shows no signals of selection.*

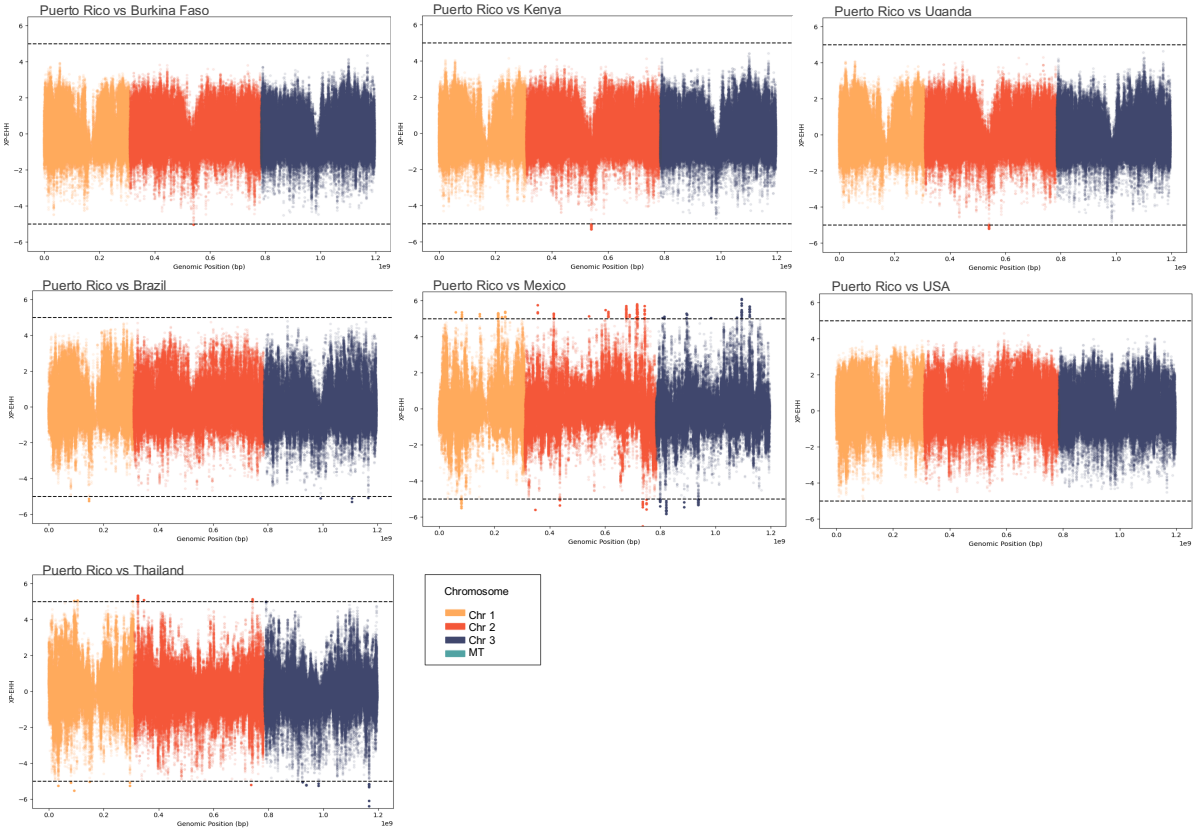


Figure S10. *XP-EHH comparisons with all populations against Puerto Rico show signals of selection between the populations. Positive values indicate selection in the Puerto Rican population, while negative values indicate selection in the other population.*

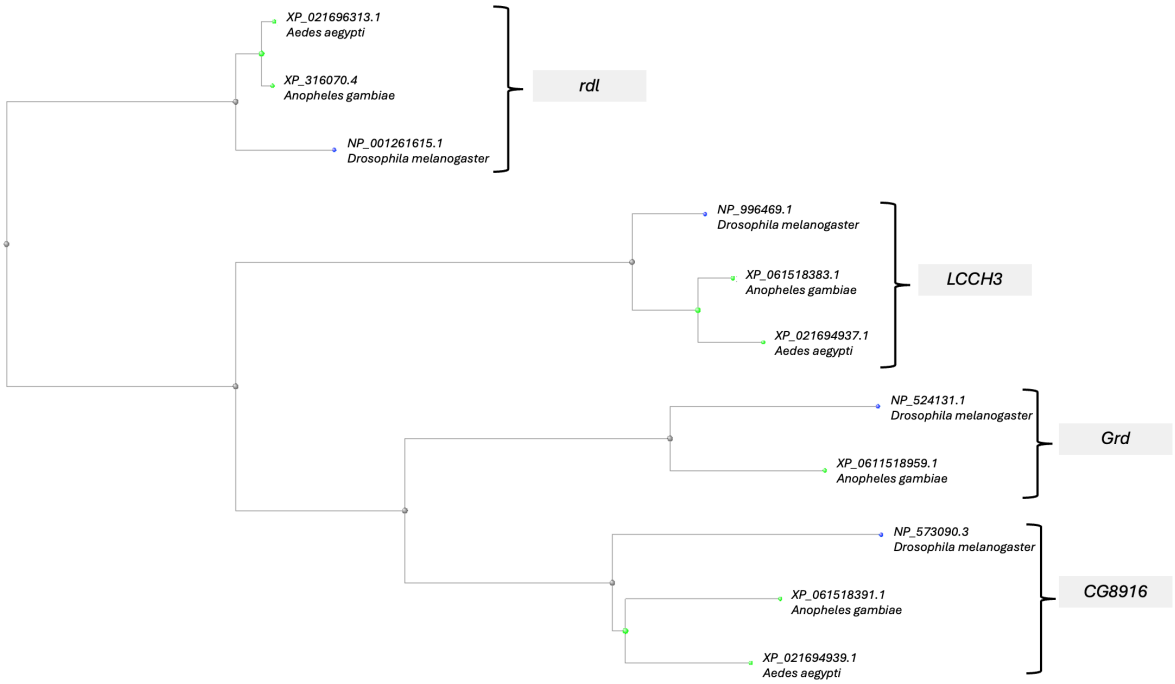


Figure S11. *Phylogenetic tree comparing protein similarity between rdl, LCCH3, Grd and CG8916 for Ae. aegypti, An. gambiae, D. melanogaster, created in COBALT (https://www.ncbi.nlm.nih.gov/tools/cobalt/cobalt.cgi?CMD=Web).*

DISCUSSION

Effective vector control remains an essential tool to combat vector-borne diseases (VBDs), especially where no effective vaccines or treatments are available, as is the case for many arboviral diseases. Although insecticides remain widely used for vector control, their efficacy is threatened by the emergence and spread of insecticide resistance, an increasing concern for vector control programmes worldwide [1]. Insecticide resistance is caused by a complex array of mechanisms [2], making it difficult to investigate. Traditional bioassays, although considered the gold standard, are both time-consuming and subjective [3,4]. As we can learn a lot from molecular information about the mechanism behind insecticide resistance, the development of molecular tools to investigate insecticide resistance is useful to complement existing phenotypic testing [4]. These additions would enhance vector control programmes by enabling monitoring of insecticide resistance status and spread in mosquito populations, as well as facilitating the discovery of potentially novel markers and mechanisms underlying insecticide resistance.

Genetic techniques have already proven to be invaluable in the field of infectious diseases, across both research and clinical settings [5–7]. Genomics, the study of an organism's whole genome, can be used to identify resistance and uncover diversity between populations [6–9]. The development of sequencing platforms with a reduction in cost per genome has made genomics-centred studies more accessible, although there are still barriers to their implementation in lower resource settings [10]. Despite these challenges, genomics has already made significant contributions to the knowledge and understanding in the field of vector-borne disease (VBDs). The *Anopheles gambiae* 1000 Genomes (Ag1000G) Project was a landmark genetic study of a medically important vector species [11,12]. The project used genomics to provide insights into both dispersal behaviour and genetic diversity across populations, as well as provide important insights into molecular mechanisms of pyrethroid resistance [13]. The *Anopheles* genomic data was made publicly available to support other researchers in their molecular studies, aiding in the identification of targets for genetic engineering of mosquito populations [14].

In this thesis, I explore insecticide resistance in *Aedes aegypti* mosquitoes through the lens of genomic techniques, aiming to add to the information bioassays can provide. The findings outlined in this thesis underscore the advantages of integrating various molecular techniques with existing approaches to explore the insecticide resistance profiles of mosquito populations. In the face of increasing insecticide resistance, an integrated approach is vital for devising more effective vector control strategies [3]. To supplement this field of study, I developed a multiplex amplicon sequencing assay to uncover

genotypes linked to insecticide resistance in several locations, including Cabo Verde **(Chapter 1)** and Puerto Rico **(Chapter 2).** This method permits high throughput generation of sequence data at reduced costs and can be used to assist vector surveillance in the future. In **Chapter 2,** I assessed various populations across Puerto Rico using intensity bioassays to ascertain their phenotypic resistance status to different concentrations of commonly used insecticides. This information was paired with the genotypic profiles to assess associations.

In addition to targeted sequencing methodologies, I utilised whole genome sequence (WGS) data to reveal diversity in genes linked to insecticide resistance in publicly available samples **(Chapter 3),** covering 15 diverse *Ae. aegypti* populations worldwide, which can be used in future genomic comparisons and investigations. Finally, I generated WGS data to demonstrate how the Puerto Rican *Ae. aegypti* population differs from other global populations, examining over 750 genes which have been understood to be implicated in insecticide resistance across a variety of vectors **(Chapter 4).**

Summary of Research Findings

This section provides a brief overview of the findings of the thesis concerning the original objectives, followed by limitations and recommendations. For a detailed discussion, please refer to the discussion section of each research Chapter.

Objective 1

Create an amplicon sequencing insecticide resistance surveillance panel to capture SNPs (single nucleotide polymorphisms) associated with insecticide resistance and apply it to a population of field Ae. aegypti samples *from Cabo Verde*

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
(Published [15])

Chapter 1 demonstrates how to identify insecticide resistance associated mutations in a population quickly and effectively using a barcoded amplicon sequencing methodology. Whilst traditional Polymerase Chain Reaction (PCR) based methods target only specific mutations, this approach targets a 500bp region, allowing the detection of both known and novel mutations in a gene region that may contribute to resistance. Additionally, barcoding allows for the pooling of multiple amplicons across many samples, reducing time and cost constraints, which is key for surveillance applications in low-income settings. I designed the amplicon sequencing panel to cover eleven regions across three genes associated with insecticide resistance for use on samples from Cabo Verde (n = 152). I hypothesised that we would identify several SNPs in this population associated with insecticide resistance due to previous dengue and Zika outbreaks in the area and subsequent intensive insecticide use prior to sample collection.

I found four non-synonymous amino-acid substitutions (V977L, K1577T, N1595T, P1612H) in the *vgsc* gene as well as the L466V mutation in the *ace-1 gene. U*nfortunately, these variants have an unknown and undocumented impact. I also identified the *rdl*-A301 mutation associated with organochlorine resistance at high frequency (98%) in the mosquitoes screened. Therefore, due to the high frequency of this mutation, we can presume that the population is at least partially resistant to organochlorines. However, as there are many genes that contribute to the phenotype, we cannot ascribe the phenotype with certainty without bioassay data.

Since there are many mechanisms involved in causing phenotypic resistance, the correlation between a single SNP and a phenotype is not always direct, therefore, this study was limited by the lack of phenotypic data associated with the samples. Unfortunately, without this data, it is not possible to link other putative SNPs identified using amplicon sequencing with the mosquito's resistance/susceptibility status, however, these SNPs can become part of a repository for candidate variants used to inform future studies.

Previous studies have found insecticide resistance in Cabo Verde. A study in 2009 identified DDT (organochlorine) resistance but pyrethroid (deltamethrin, lambda-cyhalothrin, and permethrin) and organophosphate (fenitrothion) susceptibility [16]. A later study in 2012 detected resistance to deltamethrin and cypermethrin (pyrethroids), as well as temephos (organophosphate) but not malathion (organophosphate) [17]. Given we did not detect any mutations in the *vgsc* or *ace-1* gene, this may suggest that there is inconsistent resistance to different organophosphates and that the resistance observed is mediated by metabolic resistance mechanisms.

This amplicon sequencing panel demonstrated its utility as a cost-effective method to screen a population for molecular markers of insecticide resistance. This methodology was subsequently used to detect the presence of *Ae. aegypti* larvae in water sources; no mutations were identified in this application as a lab strain was used. One of the benefits of this technique is that it has the potential to be adapted for a variety of applications with the inclusion of novel or different loci.

Objective 2

To pair an expanded amplicon sequencing insecticide resistance screening panel (created in Objective 1), with phenotypic testing on a field population of Ae. aegypti from Puerto Rico

Profiling insecticide resistance phenotypes and genotypes in *Aedes aegypti* populations across four Regions in Puerto Rico
(submitted, under peer review [18])

In **Chapter 2,** I focused on refining the methodology outlined in **Chapter 1.** As the study in **Chapter 1** was limited by a lack of phenotypic data, I prioritised the inclusion of phenotype testing in this subsequent study by using CDC bottle bioassays to test insecticide resistance in a population of *Ae. aegypti* from Puerto Rico. Collection of *Ae. aegypti* eggs were carried out using ovitraps across four sites to rear adult mosquitoes for these bioassays. Bioassays were performed to test both deltamethrin and malathion at three different concentrations to gauge the intensity of resistance. This allowed for the quantification of phenotypes in terms of percentage mortality following exposure to deltamethrin and malathion of different populations across Puerto Rico. Through this, I identified that the Puerto Rico *Ae. Aegypti* population is highly resistant to both insecticides. Mortality ranged between sites from 2-40% to deltamethrin and 18-87% to malathion at diagnostic dose, well below the <90% mortality threshold to classify as insecticide resistant [4].

Phenotypic assays were followed up by molecular screening, using the previously described barcoded amplicon sequencing technique, on 178 mosquitoes. I improved the amplicon sequencing insecticide resistance screening panel by optimising multiplex ability and increasing the number of genomic targets. This included adding amplicons that target *GSTe2* mutations (L111S, L119F), as well as speciation targets in the *cytochrome c oxidase I* gene (*cox1)*. This application highlights the adaptability of the panel to detect many mutations associated with insecticide resistance in key genes associated with target site resistance. Of the 57 SNPs identified by the panel, 14 were non-synonymous within the *vgsc, rdl* and *ace-1* genes. Five of the non-synonymous mutations had previously been linked to insecticide resistance in *Ae. aegypti* populations: V410L (*vgsc*), L978F (*vgsc*), V1012I/G (*vgsc*), F1534C (*vgsc*) and A301S (*rdl*). The remaining non-synonymous mutations identified included *vgsc*-L921I. I hypothesised that accounting for the gene homology of different mosquito species, it may be possible to identify mutations in *Ae. aegypti* that had previously been described in other species to be associated with resistance. Through this, we were able to identify that the *vgsc*-

L921I mutation may be equivalent to the L925I mutation, which has been previously linked to insecticide resistance in other medically significant vectors and agricultural pests (see **Chapter 2** results) [18–20]. This, along with the identification of the V1016G mutation at low frequency in Puerto Rico, exemplifies how genomic methodologies can be used to identify new markers and contribute to cross-species knowledge of insecticide resistance mechanisms.

I also hypothesised that there would be associations between the mutations identified and the phenotypic resistance of each population. However, the association between the resistance profiles identified and the quantitative trait phenotypes (e.g., concentrations) was less clear. As one would expect, there was a significant association between concentration and mortality ($p = 2x10^{-16}$). It was revealed that malathion had a greater impact on mortality than deltamethrin ($p = 1.46x10^{-7}$, coefficient estimate = 0.430, deltamethrin as reference). Associations between mortality and location were less convincing, with Ponce having a significantly higher mortality rate than Bayamon (coefficient estimate = 0.343, $p = 1.8x10^{-5}$), but neither San Juan nor Dorado showed significant differences ($p > 0.07$). Spearman's rank analysis between mortality rate and allele frequency for each insecticide resistance SNP identified by location indicated no correlation ($p > 0.05$). Other mechanisms of resistance may be involved, including metabolic resistance or mutations in other genes. It is also possible that it highlights the limitations of insecticide bioassays to quantify the intensity of resistance. This study could have been improved with the application of synergist assays to assess the level of metabolic resistance in the population. Additionally, following synergist assays with qPCRs or RNA sequencing to quantify the differences in the expression of certain genes linked with resistance would be highly beneficial to further understand these profiles.

Objective 3

Use public Ae. aegypti whole genome sequences to investigate the genetic diversity of key insecticide resistance-associated genes (vgsc, rdl, ace-1, GSTe2) globally

Uncovering the genetic diversity in Aedes aegypti insecticide resistance genes through global comparative genomics *(Published [8])*

In **Chapter 3,** I investigated the genetic diversity across the four genes associated with target site resistance (*vgsc, ace-1, rdl* and *GSTe2*) utilising publicly available whole genome sequence data sourced from 729 sample sequences from 5 countries. This methodology allowed SNP analysis of these genes and the ability to investigate structural variation, which has also been implicated in resistance, particularly in *An. funestus* [21,22].

I have illustrated the global diversity in these key genes that are involved in insecticide resistance, including through the identification of 1,829 genetic variants. A total of 474 variants were non-synonymous substitutions, many of which have been implicated in insecticide resistance and other mutations whose phenotypic effect is currently unknown. I also identified putative copy number variations in *GSTe2* and *vgsc*. I found that *GSTe2* was not duplicated in the majority of the 729 samples, although there are four copies in the AaegL5 reference [23]. Apart from *vgsc*, the other genes associated with insecticide resistance have, to date, been investigated to a lesser extent in *Ae. aegypti* compared to other mosquito species, and almost no genetic diversity information is available. The publication of this catalogue provides insights into these genes to inform further studies. By understanding these genes further, and in combination with other data such as bioassays and RNA data, we can start to untangle the interplay of mechanisms involved in insecticide resistance and improve control of VBD.

I endeavoured to geographically pair global insecticide resistance phenotypic data with genotypes identified. However, I found inconsistencies in profile matching. I hypothesised that at this larger scale, I might find associations between mutations identified in the gene analysis and the phenotypes described in each geographical location. However, I did not find a significant correlation in this ecological study analysis. Furthermore, establishing standardized phenotypic and genotypic profiles is crucial for enabling meaningful comparisons across studies, particularly due to significant phenotype variations observed even among proximal geographical regions. Previous research has underscored

temporal variability in phenotypes within specific regions, both increasing and decreasing susceptibility [24]. Moreover, studies have shown that phenotypes in one region can vary temporally. To gain further understanding, it would be beneficial to increase the sample size of the dataset of matched phenotypic and genotypic data and follow up with functional studies to validate the effect of the variants found.

Objective 4

Generate and analyse whole genome sequence data from a population of Ae. aegypti from Puerto Rico and examine their signals of selection and relatedness to Ae. aegypti from other geographical regions.

Genome-wide population genetics and molecular surveillance of insecticide resistance in Aedes aegypti mosquitoes from Puerto Rico *(Submitted)*

Within **Chapter 4,** I expanded the investigation into the genomic diversity of *Ae. aegypti* by sequencing the whole genome of 33 samples from Puerto Rico and combining them with 215 publicly available sequences from seven countries across the globe. Only exon sequences were analysed due to the size and complexity of the *Aedes* genome (1.3Gb), with many repetitive intergenic regions that are intensive to map to the reference genomes. However, the whole genome sequences have been added to publicly available databases and were the first sequences from Puerto Rico to be added.

A population structure analysis revealed that African samples grouped and diverged from American and Asian *Ae. aegypti*, which also formed a distinct combined group. This pattern supports previous findings that *Ae. aegypti* has spread from Africa to the Americas and subsequently to Asia [25]. The newly sequenced Puerto Rican samples grouped closely with the other samples from the Americas, including USA and Mexico. These findings were also demonstrated with phylogenetic trees and principal component analysis. Population dynamics were investigated within the Puerto Rico samples, where I found minimal genetic variation between samples. This was unsurprising given they were collected at the same time and collection sites were not very geographically dispersed. However, I only had two WGS from Culebra, the island off the mainland. It is possible that if I had more samples, I could have observed more differences.

Four known insecticide resistance mutations (*vgsc:* V410L, V1016I/G, 1534C and *rdl*: A301S) were identified across the ~250 samples. It confirmed the identification of the V1016G mutation in the Puerto Rican population, although it was identified at low frequency. I also identified multiple mutations in the *GSTe2* gene (L111S and I150V) that have been speculated by modelling analysis to be associated with DDT resistance [26]. The allele frequencies of the resistance mutations were elevated in the Puerto Rican samples, suggesting potential challenges in controlling this population with insecticides, a concern highlighted by the current ongoing dengue virus outbreak [27]. Moreover, the

profile of this population (which has paired phenotype and genomic data) may serve as a useful reference when assessing profiles of other populations.

Additionally, I investigated signals of selection across these populations and found various candidates, including multiple cytochrome P450's and gamma-aminobutyric acid receptor alpha (GABA) genes. As discussed in **Chapter 4**, the GABA receptor is a receptor that is the target for organochlorine (cyclodienes) insecticides. It is thought to be composed of a combination of genes, and recently, it has been speculated that a fourth gene is involved in GABAergic signalling [28]. This is the CG8916 gene, which is an orthologue of the GABA alpha subunit (gene: LOC5566204) [29] highlighted as being selected for in Puerto Rican and Kenyan populations. This may be due to an insecticidal environmental pressure in these populations. I also detected signals of selection in some gene families linked to metabolic resistance, including cytochrome P450s. These genes were interesting because they varied which were under between geographical regions, these differences have been reported elsewhere, with many studies producing different significant cytochrome P450s [30–37]. This may indicate differences in environmental pressure or redundancy and subsequent variation in the detoxification system.

Overall Findings and Limitations

The need for phenotypic data

A limitation of this analysis is that genomic data is not often paired with phenotypic data, preventing comprehensive investigation of the multiple mechanisms of resistance. In **Chapter 2** I address this issue by collecting phenotypic data, alongside genotypic assays, to provide a more comprehensive investigation of resistance. However, a limited number of CDC bottle bioassays were carried out due to the time-consuming nature of these experiments. Although two insecticides were tested and intensity assays were performed, this experiment could have been improved by increasing the number of replicates and testing a broader range of insecticides, with differing mechanisms of action, to gain more insight into the resistance profile. In **Chapter 3,** I used publicly available phenotypic data to complement the genetic data available. Phenotypic data and genetic data at the country level were unfortunately not geographically or temporally specific enough to generate as high of a resolution snapshot of a country's resistance profile as anticipated. Insecticide resistance profiles are generally very localised, having arisen due to different environmental pressures. The thesis underscores the complexity of insecticide resistance, emphasizing the necessity for a comprehensive approach that addresses its primary mechanisms: target site resistance and metabolic resistance. The absence of data from synergist bioassays or expression studies represents a limitation of this study, as these mechanisms typically co-occur. Without this data, a holistic understanding of the insecticide resistance profile remains incomplete. Nevertheless, this study highlights the efficacy of next-generation sequencing coupled with bioassay integration and identifies areas for improvement that could enhance future control programs.

Aedes aegypti's complex genome

In each **Chapter**, I came across difficulties with the size and characteristics of the *Ae. aegypti* genome. Both the size and repetitive nature of their genome means that it requires a lot of computational power and time to analyse *Ae. aegypti* data [23]. To accommodate this logistical problem, I have used amplicon sequencing, in which I extracted specific genes of interest and focused on exon sequences, or coding regions, to facilitate feasible analysis computation and timelines. These strategies come at the cost of potentially missing data, which may be valuable; however, they cover the most important regions of interest associated with insecticide resistance and population dynamics. Where possible, it is beneficial to use whole genome sequencing (WGS) data, either by generating it or utilising publicly

available data, as it can inform other methodologies, such as targeted amplicon sequencing or identifying candidate genes for resistance. In **Chapter 4,** I highlight the importance of generating WGS data by supplementing the globally available data with generated *Ae. aegypti* sequences from Puerto Rico. This new data will be publicly available and utilised in future studies to aid the design of tools that are appropriate for the target population.

Chapters 3 and **4** revealed that some regions in the *Ae. aegypti* reference genome (AaegL5.0) does not map well to some of the sequences I generated [23], including within the *vgsc* gene, which is important in insecticide resistance. The misaligned sequence mapping finding had been previously reported, in which, upon mapping, studies found that some sequences diverge from the reference [38,39]. The AaegL5.0 genome is based on an inbred strain of *Ae. aegypti* from West Africa, which was reared in Liverpool and is now maintained in the USA and may not be representative of other populations [40]. This highlights that future studies would likely benefit from an improved *Ae. aegypti* genome, including the production of a pan *Ae. aegypti* genome or new reference genomes from diverse geographical regions to enhance the accuracy of genomic analysis. However, it is acknowledged that this is difficult given the size and complexity of the *Ae. aegypti* genome.

Limited availability of genomic data

Chapter 4's analysis demonstrated the advantages of understanding the genomics of entire populations, as well as detecting insecticide resistance mutations and signals of selection. There are 1,407 *Ae. aegypti* genome sequences publicly available on National Centre for Biotechnology Information (NCBI) [41]. Unfortunately, one limitation of *Aedes* WGS data availability is the large size of the genome, which makes it expensive to achieve adequate coverage of the entire genome.

Not all publicly available sequences provide the metadata for the country of origin of the mosquito, however, where available, the metadata shows that the data is from 33 countries. We have identified a need for more data, specifically from Asia, with only 30 genome sequences from the entire continent, despite this region being highly affected by the diseases *Ae. aegypti* transmits [42]. Whilst *Ae. aegypti* is highly endemic to Asia, more focus has been put into *Ae. albopictus,* another common vector of arboviruses in the region.

Another advantage of increasing the publicly available repository of WGS *Ae. aegypti* data is the posited impact of climate change will have on disease transmission in more currently temperate

regions and the increasing importance of generating sequence data for these regions, such as Europe. In recent years, Zika, dengue, and Chikungunya have been detected in both France and Italy [43], while there are currently only five genome sequences from European *Ae. aegypti*. Increasing the amount of data available and understanding the resistance profile will be important in controlling the diseases these mosquitoes will inevitably import if climate change continues its current path.

Chapter 3 and **Chapter 4** showed that there is less resistance in *Ae. aegypti* in more African countries compared to American or Asian populations. However, this has highlighted countries of risk, such as Burkina Faso, where there is a strong genomic profile of resistance. Difficulties in controlling *Ae. aegypti* transmitting arboviruses in Africa is an emerging problem. There have been multiple dengue virus outbreaks in recent years in several African countries, including Burkina Faso, Ethiopia, and Cote d'Ivoire [43–46]. These outbreaks may cause further strains on public health infrastructure and stretch limited resources for vector control across the continent, and active monitoring in the coming years will be crucial.

Overall outcomes and advances

This study has made numerous incremental advances in *Ae. aegypti* vector research in the field of insecticide resistance, many of which may apply to other mosquito species.

Firstly, I developed and refined an amplicon sequencing panel to detect known SNPs associated with insecticide resistance in *Ae. aegypti*. This tool complements phenotypic bioassays and provides valuable data to surveillance programs, facilitating a deeper understanding of the mechanisms underlying insecticide resistance. Its semi-targeted approach not only identifies common SNPs but also detects previously undescribed variants, as demonstrated in **Chapter 2** [18]**.**

Further advances have been made by providing the first publicly available whole genome sequences of *Ae. aegypti* from Puerto Rico. Access to sequences from more diverse regions enables more nuanced global comparisons, offering insights into genetic diversity and the influence of environmental factors on this species. The data from these mosquitoes also has a matched phenotypic profile, which may be beneficial to future studies on insecticide resistance. This may be particularly valuable due to the highly insecticide-resistant nature of this island population of mosquitoes, comparisons to other susceptible populations may provide more power to highlight differences in key genes or variants.

To complement this, we utilised whole genome sequences for a global compare different *Ae. aegypti* populations, focusing on four genes associated with insecticide resistance. This work highlighted geographical trends in SNPs, both associated with resistance and others that indicate geographical divergence. Finally, a comparative genomics analysis was carried out utilising population genomics to demonstrate differences and similarities in global populations, as well as highlighting regions under selection within and between populations. This work identified 20 genes of interest, which were significant in three analyses, indicating alterations are under selection.

The GABA alpha subunit was one of these genes, not much is known about this gene, however, there is evidence it may form part of the GABA receptor subunit [28] and, therefore may be a target for insecticides. Several positions were identified which occur in the binding site for insecticide binding. Further examination into the role of this gene and the alterations in protein structure that these variants confer would be interesting to investigate further.

Future Perspectives

The next steps in the progression of this work would include the continuation of improving and validating the amplicon sequencing panel. With the continued discovery of new mutations attributed to insecticide resistance, it is key to keep updating the panel to capture important and emerging SNPs. To aid this, I have begun adjusting the amplicon panel for Oxford Nanopore sequencing platforms such as the MinION. These platforms allow the sequencing of longer reads, so fewer target amplicons are required, and it is possible to capture more information about the entire gene region. Adapting this potential surveillance tool to a portable platform, allowing in-field sequencing, would be highly advantageous for the application of this technique in low- or middle-resource settings. Moreover, I designed primers to be able to detect mutations in both *Ae. aegypti* and *Ae. albopictus,* as this is another important disease vector. I have not yet validated this technique with *Ae. albopictus* samples. I believe this, with the addition of a species ID amplicon such as *cox-1,* would be valuable in many regions where *Ae. aegypti* and *Ae. albopictus* are sympatric.

Investigating the putative mutations identified in both amplicon and WGS studies would provide valuable insights for vector control programmes around the world. Understanding the role of these mutations in insecticide resistance will help to elucidate their function and inform resistance prevention tactics. There are a handful of possible methodologies that could be used for this. Protein modelling to examine the likely result of the mutation may identify the candidates with roles in insecticide receptor binding. Functional investigations using genome editing, such as site-directed mutagenesis in model organisms (e.g., *Xenopus* oocytes), could also be applied. This approach has been used previously to confirm insecticide resistance mutations, including the expression of the F1534C mutation (as identified in this study) that, when inserted into cockroach *vgsc,* showed reduced sensitivity to type I pyrethroids [47,48]. These experiments are useful as they isolate the role of the specific mutation and remove confounding variables. To demonstrate this, a study utilising these techniques could be carried out to investigate the L921I (L933I, *Musca domestica)* mutation identified in the Puerto Rican *Ae. aegypti* population. A mutation at this position has previously been detected in both *Triatoma infestans* (L925I) and *Bemisia tabaci* (L925I) [19,20]. Alignment analysis indicated that amino acid 921 in *Ae. aegypti* may be the equivalent to amino acid 925 in these species. The L925I mutation has been implicated in pyrethroid resistance in *Drosophila melanogaster* by expressing the mutation in *Xenopus* oocytes. Confirming this mutation's function and its role in *Ae. aegypti,* would

be crucial in our understanding of this variant on insecticide resistance and help to investigate if it is specifically linked to pyrethroid resistance.

Further work investigating the synergistic effects of mutations would be beneficial for the overall understanding of the profile of insecticide resistance. For example, it has been described that the combination of S989P, V1016G and F1534C mutations reduces permethrin sensitivity by over 1000-fold [49]. Other combinations of mutations have also shown high linkage, which may ameliorate fitness costs the mutations otherwise incur. This theory has been proposed for the L199F mutation in the *vgsc,* which may sterically compensate for the presence of the L982W mutation [50]. Novel mutations identified in these studies may be found to have additional effects in combination with previously identified mutations, or amongst themselves and should be investigated further.

The role and influence of the 20 genes under selection should be investigated further to understand their role and possible reasons for them to be under selection. In particular, the GABA alpha subunit would be interested to learn more about, particularly its potential role in the formation of some GABA receptors. This has been illustrated in vitro expression of the *Chilo suppressalis GC8916* gene within the *Xenopus* oocyte system [28], however it would be interesting to test this with *Ae. aegypti.* As well as learning more about the sites identified and what exact role they may have in insecticide binding.

If time and funds had permitted, additional RNA studies to investigate the mechanisms of metabolic resistance would have been beneficial. By extracting RNA from both resistant and susceptible mosquitoes, as well as both exposed and unexposed to insecticides, qPCR or RNA sequencing could be used to examine differential expression profiles between the phenotypes. This would enable the identification of genes and gene families that contribute to the resistance phenotype. However, the high levels of redundancy within and across these enzyme gene families mean that multiple modifications can lead to resistance and, consequently, may result in many profiles that can lead to resistance. This complexity is difficult to disentangle without large sample sizes from multiple populations. It would be interesting to see if RNA expression studies highlight genes that were identified in the selection analysis carried out in **Chapter 4.**

As other mechanisms, such as cuticular thickening and effects of the microbiome, have also been cited as possible facilitators of resistance, they should also be investigated. These mechanisms are both linked to the over-expression of genes, so genomics can be utilised to explore this. The composition of the microbiome has been proposed to contribute to resistance by modulating the hosts detoxifying gene expression. This has been reported in multiple species, including *Ae. albopictus* and *Ae. aegypti*

[51–53]. Cuticle changes, including thickening to slow penetration of insecticide and modification of cuticle composition, have also been implicated in insecticide resistance [54]. A recent study has investigated cuticle differences between susceptible and resistant *Ae. aegypti* and found cuticle thickness of insecticide resistant *Ae. aegypti* increased over time and correlated with metabolic differences, suggesting the cuticle does have an impact on resistance profiles [55].

Overall, genomics tools can aid control programmes to confirm the presence of mutations conferring resistance and allow adaptations of programmes to ensure they are effective. I have shown that with our current understanding of insecticide resistance, phenotypic assays are still essential to fully understand the impacts on control but can be supplemented by genomic data to confirm molecular targets and streamline programme design and target specificity. In the future, with this further understanding, it would be ideal to have a panel of targets that include both target site mutations, expression of enzymes involved in metabolic resistance, and copy number variants that could capture the complex genomic components causing the phenotype. Unfortunately, the current costs associated with molecular genomics are still prohibitive to many regions where VBDs are transmitted, particularly as *Aedes* has a large genome, which contributes to higher costs. Improvements in infrastructure capacity and training would be required to make these methodologies practical to implement in many low-resource countries, although the development of platforms such as the Nanopore MinION is making sequencing in remote areas more attainable.

The future of mosquito control for *Ae. aegypti* is uncertain, and there remain limited treatments or vaccines for the diseases this species transmits. Due to this, insecticides remain the backbone of control. Insecticide resistance is a huge problem across much of the globe and is only going to deteriorate further with the continued intensive use of insecticides [3]. However, in the interim, insecticides remain essential within the array of strategies used in an integrated vector control programme. Control programmes need to maintain surveillance of resistance mechanisms and levels within populations to ensure the ongoing effectiveness of control strategies. To achieve this, control programmes should involve both phenotypic and genotypic monitoring. Developing the role of genomics for the surveillance of insecticide resistance would be highly beneficial, however, the complexities of interplaying insecticide resistance mechanisms remain unclear, and further work is required. It is important to employ strategies such as mosaics, rotations, mixes, and combinations, which can delay the emergence and spread of resistance [56].

These studies have helped to add to the catalogue of genomes and characterised genomic variants within *Ae. aegypti*, which may aid the development of new methodologies to control this vector. In

the long term, it may be that other genomic strategies for control may become more prominent. Alternative techniques, such as the sterile insect technique, including genetic modification, gene drives and the release of *Wolbachia*-infected *Ae. aegypti* to prevent dengue transmission may become more developed and utilised [57]. However, it is important to note that the resistance status of the released mosquitoes must be comparable to the wild population, so understanding this status is essential before widespread implementation.

Finally, there is a critical need to raise awareness and foster proactive responses to information regarding insecticide resistance. In recent years, there has been increased emphasis on the importance of control programs understanding both the molecular markers of insecticide resistance and the associated phenotypes within their populations, allowing for the design of more targeted and effective vector control strategies with this in mind [58]. However, studies have highlighted that control programs often lag in adjusting methods after detecting resistance [59]. Furthermore, few programs incorporate effective resistance management strategies, such as using insecticides with diverse modes of action. As well as the generation of data informing insecticide resistance status and identifying markers, the resolve and capacity to implement robust resistance management practices is crucial for the success and sustainability of vector control programs.

Conclusion

This thesis explores the use of various genomics and genetic techniques to enhance our understanding of the mosquito *Ae. aegypti*, a species that has received less attention compared to the *Anopheles* genus, with the aim of improving its control. I have demonstrated the development of a high throughput amplicon sequencing assay capable of detecting both known insecticide resistance mutations and novel mutations in key genes. This assay has been effectively paired with phenotypic assessments to enhance our understanding of resistance. Additionally, I have utilized whole genome sequencing (WGS) data to gain deeper insights into genes associated with resistance and illustrate the genomic landscape in the context of population dynamics. These methodologies can significantly contribute to the development of effective control strategies targeting mosquito populations. This thesis also presents the first full genomes of *Ae. aegypti* mosquitoes from Puerto Rico, making them the first whole genome sequences obtained from the Caribbean region. Collectively, this work provides valuable genomic resources and innovative tools that can significantly enhance the strategic management and control of *Aedes aegypti* populations, ultimately contributing to more effective and targeted public health interventions.

References

1. Corbel, V. *et al.* International workshop on insecticide resistance in vectors of arboviruses, December 2016, Rio de Janeiro, Brazil. *Parasit Vectors* **10**, 1–16 (2017).
2. Moyes, C. L. *et al.* Contemporary status of insecticide resistance in the major Aedes vectors of arboviruses infecting humans. *PLoS Negl Trop Dis* **11**, (2017).
3. WHO. *Global Vector Control Response 2017-2030*. https://apps.who.int/iris/bitstream/handle/10665/259205/9789241512978-eng.pdf?sequence=1 (2017).
4. World Health Organization. Manual for monitoring insecticide resistance in mosquito vectors and selecting appropriate interventions.
5. Riess, O. *et al.* Genomes in clinical care. *npj Genomic Medicine 2024 9:1* **9**, 1–10 (2024).
6. Miotto, O. *et al.* Genetic architecture of artemisinin-resistant Plasmodium falciparum. *Nature Genetics 2015 47:3* **47**, 226–234 (2015).
7. Lucas, E. R. *et al.* Genome-wide association studies reveal novel loci associated with pyrethroid and organophosphate resistance in Anopheles gambiae s.l. *bioRxiv* (2023) doi:10.1101/2023.01.13.523889.
8. Spadar, A., Collins, E., Messenger, L. A., Clark, T. G. & Campino, S. Uncovering the genetic diversity in Aedes aegypti insecticide resistance genes through global comparative genomics. *Scientific Reports 2024 14:1* **14**, 1–12 (2024).
9. Ruzzante, L., Reijnders, M. J. M. F. & Waterhouse, R. M. Of Genes and Genomes: Mosquito Evolution and Diversity. *Trends Parasitol* **35**, 32–51 (2019).
10. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics 2016 17:6* **17**, 333–351 (2016).
11. Ag1000g. https://www.sanger.ac.uk/tool/ag1000g/.
12. Miles, A. *et al.* Genetic diversity of the African malaria vector Anopheles gambiae. *Nature 2017 552:7683* **552**, 96–100 (2017).
13. Clarkson, C. S. *et al.* 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).
14. Kranjc, N., Crisanti, A., Nolan, T. & Bernardini, F. Anopheles gambiae Genome Conservation as a Resource for Rational Gene Drive Target Site Selection. *Insects* **12**, 1–12 (2021).
15. Collins, E. L. *et al.* A next generation targeted amplicon sequencing method to screen for insecticide resistance mutations in Aedes aegypti populations reveals a rdl mutation in mosquitoes from Cabo Verde. *PLoS Negl Trop Dis* **16**, e0010935 (2022).
16. Dia, I. *et al.* Insecticide susceptibility of Aedes aegypti populations from Senegal and Cape Verde Archipelago. *Parasit Vectors* **5**, 238 (2012).
17. Rocha, H. D. R. *et al.* Susceptibility profile of Aedes aegypti from Santiago Island, Cabo Verde, to insecticides. *Acta Trop* **152**, 66–73 (2015).
18. Collins, E. *et al.* Profiling insecticide resistance phenotypes and genotypes in Aedes aegypti populations across four regions in Puerto Rico. (2024) doi:10.21203/RS.3.RS-3943783/V1.
19. Capriotti, N., Mougabure-Cueto, G., Rivera-Pomar, R. & Ons, S. L925I Mutation in the Para-Type Sodium Channel Is Associated with Pyrethroid Resistance in Triatoma infestans from the Gran Chaco Region. *PLoS Negl Trop Dis* **8**, e2659 (2014).
20. Wei, Y., Guan, F., Wang, R., Qu, C. & Luo, C. Amplicon sequencing detects mutations associated with pyrethroid resistance in Bemisia tabaci (Hemiptera: Aleyrodidae). *Pest Manag Sci* **77**, 2914–2923 (2021).
21. Tchouakui, M. *et al.* A 6.5kb Intergenic Structural Variation Exacerbates the Fitness Cost of P450-Based Metabolic Resistance in the Major African Malaria Vector Anopheles funestus. *Genes (Basel)* **13**, (2022).
22. Leon, M. *et al.* A rapidly selected 4.3kb transposon-containing structural variation is driving a P450-based resistance to pyrethroids in the African malaria vector Anopheles funestus. *Authorea Preprints* (2023) doi:10.22541/AU.168570017.77126301/V1.
23. Matthews, B. J. *et al.* Improved reference genome of Aedes aegypti informs arbovirus vector control. *Nature 2018 563:7732* **563**, 501–507 (2018).
24. Ryan, S. J. *et al.* Seasonal and geographic variation in insecticide resistance in Aedes aegypti in southern Ecuador. *PLoS Negl Trop Dis* **13**, (2019).

25. Powell, J. R. & Tabachnick, W. J. History of domestication and spread of Aedes aegypti--a review. *Memórias do Instituto Oswaldo Cruz* vol. 108 Preprint at https://doi.org/10.1590/0074-0276130395 (2013).
26. Helvecio, E. *et al.* Polymorphisms in GSTE2 is associated with temephos resistance in Aedes aegypti. *Pestic Biochem Physiol* **165**, (2020).
27. Current Year Data (2024) | Dengue | CDC. https://www.cdc.gov/dengue/data-research/facts-stats/current-data.html.
28. Huang, Q. T. *et al.* Functional Characteristics of the Lepidopteran Ionotropic GABA Receptor 8916 Subunit Interacting with the LCCH3 or the RDL Subunit. *J Agric Food Chem* **69**, 11582–11591 (2021).
29. CG8916 orthologs - NCBI. https://www.ncbi.nlm.nih.gov/gene/32553/ortholog/?scope=6656.
30. Ye, M. *et al.* The Role of Insect Cytochrome P450s in Mediating Insecticide Resistance. *Agriculture 2022, Vol. 12, Page 53* **12**, 53 (2022).
31. Yang, T. & Liu, N. Genome analysis of cytochrome P450s and their expression profiles in insecticide resistant mosquitoes, Culex quinquefasciatus. *PLoS One* **6**, e29418 (2011).
32. Matambo, T. S., Paine, M. J. I., Coetzee, M. & Koekemoer, L. L. Sequence characterization of cytochrome P450 CYP6P9 in pyrethroid resistant and susceptible Anopheles funestus (Diptera: Culicidae). *Genet Mol Res* **9**, 554–564 (2010).
33. Ibrahim, S. S. *et al.* Allelic Variation of Cytochrome P450s Drives Resistance to Bednet Insecticides in a Major Malaria Vector. *PLoS Genet* **11**, e1005618 (2015).
34. 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).
35. Liu, N., Li, T., Reid, W. R., Yang, T. & Zhang, L. Multiple Cytochrome P450 genes: their constitutive overexpression and permethrin induction in insecticide resistant mosquitoes, Culex quinquefasciatus. *PLoS One* **6**, e23403 (2011).
36. Rodpradit, P., Boonsuepsakul, S., Chareonviriyaphap, T., Bangs, M. J. & Rongnoparut, P. Cytochrome P450 genes: molecular cloning and overexpression in a pyrethroid-resistant strain of Anopheles minimus mosquito. *J Am Mosq Control Assoc* **21**, 71–79 (2005).
37. Ibrahim, S. S., Riveron, J. M., Stott, R., Irving, H. & Wondji, C. S. The cytochrome P450 CYP6P4 is responsible for the high pyrethroid resistance in knockdown resistance-free Anopheles arabiensis. *Insect Biochem Mol Biol* **68**, 23–32 (2016).
38. Love, R. R., Sikder, J. R., Vivero, R. J., Matute, D. R. & Schrider, D. R. Strong Positive Selection in Aedes aegypti and the Rapid Evolution of Insecticide Resistance. *Mol Biol Evol* **40**, (2023).
39. Rose, N. H. *et al.* Climate and Urbanization Drive Mosquito Preference for Humans. *Curr Biol* **30**, 3570-3579.e6 (2020).
40. BEI Reagent Search. https://www.beiresources.org/Catalog/BEIVectors/MRA-735.aspx.
41. 'Aedes aegypti'[orgn:__txid7159] - SRA - NCBI. https://www.ncbi.nlm.nih.gov/sra/?term=%22Aedes+aegypti%22%5Borgn%3A__txid7159%5D.
42. Paz, S. Climate change impacts on vector-borne diseases in Europe: Risks, predictions and actions. *The Lancet Regional Health - Europe* **1**, (2021).
43. Gebremariam, T. T., Schalling, H. D. F. H., Kurmane, Z. M. & Danquah, J. B. Increasing prevalence of malaria and acute dengue virus coinfection in Africa: a meta-analysis and meta-regression of cross-sectional studies. *Malar J* **22**, 1–11 (2023).
44. Were, F. The dengue situation in Africa. *Paediatr Int Child Health* **32**, 18 (2012).
45. Nigussie, E. *et al.* A dengue virus infection in Ethiopia: a systematic review and meta-analysis. *BMC Infect Dis* **24**, 1–8 (2024).
46. Gainor, E. M., Harris, E. & Labeaud, A. D. Uncovering the Burden of Dengue in Africa: Considerations on Magnitude, Misdiagnosis, and Ancestry. *Viruses* **14**, (2022).
47. Dong, K. Insect sodium channels and insecticide resistance. *Invert Neurosci* **7**, 17 (2007).
48. Hu, Z., Du, Y., Nomura, Y. & Dong, K. A sodium channel mutation identified in Aedes aegypti selectively reduces cockroach sodium channel sensitivity to type I, but not type II pyrethroids. *Insect Biochem Mol Biol* **41**, 9 (2011).
49. Hirata, K. *et al.* A Single Crossing-Over Event in Voltage-Sensitive Na+ Channel Genes May Cause Critical Failure of Dengue Mosquito Control by Insecticides. *PLoS Negl Trop Dis* **8**, e3085 (2014).
50. Kasai, S. *et al.* Discovery of super–insecticide-resistant dengue mosquitoes in Asia: Threats of concomitant knockdown resistance mutations. *Sci Adv* **8**, eabq7345 (2022).

51. Wang, H. *et al.* Differences in the intestinal microbiota between insecticide-resistant and -sensitive Aedes albopictus based on full-length 16S rRNA sequencing. *Microbiologyopen* **10**, (2021).
52. Wang, H. *et al.* A symbiotic gut bacterium enhances Aedes albopictus resistance to insecticide. *PLoS Negl Trop Dis* **16**, (2022).
53. Arévalo-Cortés, A. *et al.* The Midgut Microbiota of Colombian Aedes aegypti Populations with Different Levels of Resistance to the Insecticide Lambda-cyhalothrin. *Insects 2020, Vol. 11, Page 584* **11**, 584 (2020).
54. Balabanidou, V., Grigoraki, L. & Vontas, J. Insect cuticle: a critical determinant of insecticide resistance. *Curr Opin Insect Sci* **27**, 68–74 (2018).
55. Jacobs, E. *et al.* Cuticular profiling of insecticide resistant Aedes aegypti. *Scientific Reports 2023 13:1* **13**, 1–15 (2023).
56. Madgwick, P. G. & Kanitz, R. What is the value of rotations to insecticide resistance management? *Pest Manag Sci* **80**, 1671–1680 (2024).
57. Alphey, L. *et al.* Genetic control of Aedes mosquitoes. *Pathogens and Global Health* vol. 107 170–179 Preprint at https://doi.org/10.1179/2047773213Y.0000000095 (2013).
58. Report on insecticide resistance in Aedes mosquitoes (Aedes aegypti, Ae. albopictus, Ae. vittatus) in WHO South-East Asia Region countries. (2022).
59. van den Berg, H. *et al.* Recent trends in global insecticide use for disease vector control and potential implications for resistance management. *Scientific Reports 2021 11:1* **11**, 1–12 (2021).