Strong Positive Selection in *Aedes aegypti* and the Rapid Evolution of Insecticide Resistance

R. Rebecca Love [ID],[1] Josh R. Sikder,[1] Rafael J. Vivero,[2] Daniel R. Matute [ID],*,[3] and Daniel R. Schrider [ID]*,[1]

[1]Department of Genetics, School of Medicine, University of North Carolina, Chapel Hill, NC USA
[2]Programa de Estudio y Control de Enfermedades Tropicales, PECET, Universidad de Antioquia, Chapel Hill, NC Colombia
[3]Department of Biology, College of Arts and Sciences, University of North Carolina, Chapel Hill, NC, USA

*Corresponding authors: E-mails: dmatute@email.unc.edu; drs@unc.edu.
Associate editor: Rasmus Nielsen

Abstract

Aedes aegypti vectors the pathogens that cause dengue, yellow fever, Zika virus, and chikungunya and is a serious threat to public health in tropical regions. Decades of work has illuminated many aspects of *Ae. aegypti*'s biology and global population structure and has identified insecticide resistance genes; however, the size and repetitive nature of the *Ae. aegypti* genome have limited our ability to detect positive selection in this mosquito. Combining new whole genome sequences from Colombia with publicly available data from Africa and the Americas, we identify multiple strong candidate selective sweeps in *Ae. aegypti*, many of which overlap genes linked to or implicated in insecticide resistance. We examine the voltage-gated sodium channel gene in three American cohorts and find evidence for successive selective sweeps in Colombia. The most recent sweep encompasses an intermediate-frequency haplotype containing four candidate insecticide resistance mutations that are in near-perfect linkage disequilibrium with one another in the Colombian sample. We hypothesize that this haplotype may continue to rapidly increase in frequency and perhaps spread geographically in the coming years. These results extend our knowledge of how insecticide resistance has evolved in this species and add to a growing body of evidence suggesting that *Ae. aegypti* has an extensive genomic capacity to rapidly adapt to insecticide-based vector control.

Key words: insecticide resistance, *Ae. aegypti*, positive selection, Colombia.

Introduction

The mosquito *Aedes (Stegomyia) aegypti* (Linnaeus, 1762) vectors the arboviruses causing dengue, Zika, yellow fever, and chikungunya (Viglietta et al. 2021), the former two of which have already caused pandemics in the first two decades of the 21st century. No vaccine is publicly available for chikungunya or Zika, and treatment for all four diseases is generally supportive or symptomatic only. While considerable uncertainty exists about the actual number of annual dengue cases worldwide (Bhatt et al. 2013; Wilder-Smith and Byass 2016), one common estimate suggests approximately 400 million cases a year as of 2010 (Bhatt et al. 2013).

Because of its importance to public health, as well as its facility as a model organism, *Ae. aegypti* has been extensively studied. Two subspecies or forms have been identified, initially distinguished on the basis of morphology with corresponding differences in behavior (Mattingly 1957), though the morphological characters in question have long been recognized as subject to variation (Mattingly 1957; Brown et al. 2011). *Aedes aegypti formosus*, recognizable by its dark abdomen, generally prefers forest and other outdoor habitats and is limited to Africa; in contrast, *Ae. aegypti aegypti* has pale scales on the first abdominal tergite, is found widely across tropical regions of the world outside of Africa, is highly anthropophilic, and is found in close association in humans, often resting indoors and ovipositing in anthropogenic habitats (Mattingly 1957). The two subspecies are genetically distinct as well (Tabachnick and Powell 1979; Gloria-Soria et al. 2016; Crawford et al. 2017; Kotsakiozi et al. 2018), though both morphological and genetic distinctions break down in some parts of West Africa (Gloria-Soria et al. 2016; Crawford et al. 2017; Kotsakiozi et al. 2018). In addition, the two forms may be differentially susceptible to dengue (Sylla et al. 2009), Zika virus (Aubry et al. 2020), and yellow fever (Tabachnick et al. 1985), though the latter phenotype appears to be highly variable depending on mosquito and viral strain (Tabachnick et al. 1985; Dickson et al. 2014). A third form, *queenslandensis*, was initially recognized on the basis of morphology (Mattingly 1957); however, recent genomic analysis revealed that this group instead appears to comprise morphological variation within *Ae. aegypti aegypti* (Rašić et al. 2016).

Aedes aegypti tends to have low unaided dispersal (Harrington et al. 2005), though it may fly further if

Open Access

suitable oviposition sites or opportunities to blood feed are lacking (Reiter et al. 1995; Codeço et al. 2007). The widespread tropical distribution of *Ae. aegypti aegypti* is therefore attributed to its close association with humans (Christophers 1960). Originating in Africa (Tabachnick 1991; Brown et al. 2011; Brown et al. 2014), *Ae. aegypti* spread to the Americas (Brown et al. 2014; Gloria-Soria et al. 2016), likely due to mass transatlantic trafficking of enslaved people. It arrived in Asia more recently (Tabachnick 1991; Brown et al. 2014), possibly facilitated by the opening of the Suez Canal (Tabachnick 1991; Powell et al. 2018). The non-African domestic populations appear to result from one colonization event (Crawford et al. 2017; Rose et al. 2020), descending from a domesticated subpopulation in western Africa (Crawford et al. 2017) or possibly central Africa (Kotsakiozi et al. 2018); human preference is linked to expression and sequence variation in the *Or4* gene (McBride et al. 2014; Rose et al. 2020) as well as dry season intensity (Rose et al. 2020).

Population structure tends to be less pronounced within Africa than outside Africa (Brown et al. 2011, 2014; Gloria-Soria et al. 2016); outside Africa, specimens show strong geographic clustering (Rašić et al. 2014; Evans et al. 2015; Schmidt et al. 2020) but extensive connectivity is seen between some geographically distant populations (Goncalves da Silva et al. 2012; Monteiro et al. 2014; Gloria-Soria et al. 2016). Populations are often dynamic, showing sometimes extensive changes in genetic composition within a few years (Gloria-Soria et al. 2016) and re-invading areas where they had been eradicated (Kelly et al. 2021) or outcompeted (Brennan et al. 2021).

Overall, the global genetic structure of *Ae. aegypti* has been well-studied; however, because of the size and repetitive nature of its genome, nearly all previous work has used either sequences from a small number of genes (Bennett et al. 2016), microsatellites (Brown et al. 2011; Gloria-Soria et al. 2016), mtDNA (Goncalves da Silva et al. 2012; Bennett et al. 2016), genotyping arrays (Evans et al. 2015; Kotsakiozi et al. 2018), or reduced representation sequencing (Brown et al. 2014; Rašić et al. 2014). Exome-based population resequencing (Crawford et al. 2017; Dickson et al. 2017) and whole genome-based population resequencing have only been performed recently (Lee et al. 2019; Rose et al. 2020; Kelly et al. 2021), and the improvement of the approximately 1.3 Gb *Ae. aegypti* reference genome to a chromosomal-level assembly (Matthews et al. 2018) will greatly facilitate such efforts moving forward. Studies using a small number of loci are extremely limited in their ability to detect recent positive selection, which skews patterns of diversity across larger genomic regions (Maynard Smith and Haigh 1974; Kaplan et al. 1989; Kim and Nielsen 2004). As a consequence, we know relatively little about the extent of recent positive selection or the broad-scale genomic underpinnings of adaptation in *Ae. aegypti*, compared to our more extensive knowledge of the species' spread and structure. Positive selection has been observed on genomic regions linked to human preference (Rose et al. 2020) and

on genes that enable females to retain eggs during periods of drought (Venkataraman et al. 2022); however, with the exception of a putative selective sweep on an insecticide resistance gene (the *Vgsc* gene, discussed below) identified in a sample from southern Mexico (Saavedra-Rodriguez et al. 2019), we know little about how its genome has responded to these selective pressures at a scale beyond that of individual loci.

In some parts of the Americas, *Ae. aegypti* has been subject to insecticide-based vector control for nearly a century. In the early part of the 20th century, some of the techniques used to target *Ae. aegypti* in Latin America included fumigation of dwellings and "oiling" of water sources to kill larvae (Severo 1955). Dichlorodiphenyltrichloroethane (DDT) was used beginning in 1945 (Severo 1955) and extensively thereafter, enabling the campaign for continent-wide eradication in the Americas which began officially in 1947 (Soper 1965). The campaign was initially successful in eradicating *Ae. aegypti* from most of South and Central America (Soper 1965); however, DDT resistance had appeared in the Americas by 1964 (Soper 1965), and for this and other reasons, the campaign faltered. *Aedes aegypti* soon spread across South and Central America again (Jansen and Beebe 2010).

Resistance to DDT in *Ae. aegypti* is conferred by both increased metabolism and nonmetabolic factors (Rathor and Wood 1981; Hemingway et al. 1989; Brengues et al. 2003). The best-characterized nonmetabolic factor is *knockdown resistance* (*kdr*), first identified in houseflies, and conferring resistance to pyrethroids as well (Farnham 1977); *kdr*, caused by point mutations in the voltage-gated sodium channel gene (*Vgsc*, [Williamson et al. 1996]), has also been found in a wide variety of other insects, including *Drosophila* (Pittendrigh et al. 1997), anopheline mosquitoes (Martinez-Torres et al. 1998; Clarkson et al. 2021), tobacco budworms (Taylor et al. 1993), cockroaches (Dong and Scott 1994), and many others (Davies et al. 2007).

Unsurprisingly considering the breadth of distribution of these mutations in other insects, an extensive array of nonsynonymous point mutations associated with pyrethroid resistance have been identified in *Ae. aegypti*. (Because this resistance factor was originally identified in *Musca domestica*, the *M. domestica* amino acid numberings are often used; to avoid ambiguity, we will follow this convention.) Some, such as F1534C (Kawada et al. 2009; Harris et al. 2010), are globally distributed. Others appear limited to specific continents or regions; V1016I has so far only been recorded from the Americas (Fan et al. 2020) and Africa (Sombié et al. 2019), while V1016G has only been recorded from Asia (Fan et al. 2020). Similarly, S989P appears limited to Asia (Fan et al. 2020), while S723T has only been identified in the Americas (Fan et al. 2020), and V410L has been found in the Americas (Fan et al. 2020) and Africa (Ayres et al. 2020). A subset of the mutations associated with pyrethroid resistance, including F1534C (Hu et al. 2011; Du et al. 2013; Hirata et al.

2014; Haddi et al. 2017), V1016G (Du et al. 2013; Hirata et al. 2014), I1011M (Du et al. 2013), and V410L (Haddi et al. 2017), has been shown experimentally to directly cause pyrethroid resistance.

These mutations are often found in combination with each other; for example, V410L has been observed in conjunction with F1534C (Haddi et al. 2017; Saavedra-Rodriguez et al. 2018) and V1016I (Saavedra-Rodriguez et al. 2018), and S989P is very frequently found on a V1016G background (Kawada et al. 2014; Al Nazawi et al. 2017). Recurrent evolution (Fan et al. 2020) and recombination (Saavedra-Rodriguez et al. 2018) have been suggested as contributing factors to the combinations and distribution of these various mutations. A similar situation of recurrent mutation at the same site, as well as the presence of haplotypes comprising multiple resistance alleles, has been observed in *Anopheles gambiae* and *Anopheles coluzzii* (Clarkson et al. 2021).

Natural selection associated with insecticide resistance has been widely observed in other insects, including other disease vectors such as *An. gambiae* and *An. coluzzii* (Miles et al. 2017; Xue et al. 2021) and *Anopheles funestus* (Weedall et al. 2020); agricultural pests such as *Leptinotarsa decemlineata* (Pélissié et al. 2022) and *Helicoverpa armigera* (Walsh et al. 2018); and the model organism *Drosophila melanogaster* (Turner et al. 2008; Kolaczkowski et al. 2011; Garud et al. 2015; Schrider et al. 2016). Thus, it seems probable that positive selection for insecticide resistance would be widespread in *Aedes* as well. In addition to selection at specific insecticide resistance loci, vector control programs may cause population contractions (Urdaneta-Marquez and Failloux 2011), which can affect genome-wide change in patterns of diversity. In Colombia, vector control relies on pyrethroids and organophosphates (temephos and malathion) (Aguirre-Obando et al. 2015).

Here, we examine patterns of genetic variation in a set of 32 *Ae. aegypti* genomes collected from Colombia. We also combine these data with previously published *Aedes* genomic data to examine global patterns of genetic variation. Finally, we turn our attention toward detecting signatures of recent positive selection in *Aedes* samples from Africa and the Americas. In five of the six countries examined, we find strong signals of selection in regions containing genes that have been found to be associated with insecticide resistance in *Aedes* or other insects. We discuss the implications of these findings for adaptation to ongoing and future mosquito control efforts.

Results

Quality Control, Alignment, and Variant Genotyping
We used a stringent filtering approach to control the quality of the sequenced samples (Materials and Methods). We removed six specimens from the Colombian sample that had mean coverage below 15× or fewer than 70% of reads mapping to the reference genome, leaving a sample of 10 specimens from Cali and 24 from Río Claro (see supplementary table S1, Supplementary Material online, for complete specimen list).

In addition, we examined publicly available sequence data from the United States (US) ($n = 27$) from Lee et al. (2019), and from Santarém, Brazil ($n = 18$), Franceville, Gabon ($n = 13$), Kaya Bomu, Kenya ($n = 19$), and Ngoye, Senegal ($n = 20$), from Rose et al. (2020), for an initial total of 131 specimens in six cohorts. The Ngoye cohort comes from a population with a strong preference for human over nonhuman odor as well as a large "domestic" ancestry component; the Franceville cohort originates from a population with a strong preference for nonhuman odor and no domestic ancestry; the Kaya Bomu population is intermediate in both traits (Rose et al. 2020).

After removing the six specimens in our focal Colombian sample that failed our quality filters, mean read depth in each population ranged from 10.16 (United States) to 21.10 (Colombia), with between 94.80% (Gabon) and 98.19% (United States) of reads mapping on average (supplementary table S2, Supplementary Material online). The comparatively low number of reads mapping from the Gabonese sample may reflect divergence from the reference genome, as noted in Rose et al. (2020).

Variant calling produced an initial data set of 247,108,978 variants in 131 specimens. After filtering, we retained 79,035,096 high-quality single nucleotide polymorphisms (SNPs) with a known location on one of the three autosomes and an additional 881,515 SNPs in regions of the genome that could not be confidently assigned to chromosomes. Our analysis focused on the SNPs on the three autosomes. The number of SNPs found in each cohort, before and after applying filters based on relatedness as discussed below, is shown in supplementary table S3, Supplementary Material online.

In keeping with previous work, substantially more SNPs were called per specimen in the African sample than in the American sample (supplementary table S3, Supplementary Material online). Within the African sample, the Senegal sample had markedly fewer SNPs called per specimen than the Gabonese or Kenyan samples, perhaps reflecting that this is a human specialist population (Rose et al. 2020) and therefore may have experienced a bottleneck during the transition towards close proximity with humans (Crawford et al. 2017).

Because the presence of related individuals can confound population genetic analyses that assume a random sample of the population was taken, we also filtered genomes based on kinship coefficient after variant calling using KING-robust (Manichaikul et al. 2010), an algorithm for relationship inference in the presence of population structure. We removed one specimen each from eight pairs with high kinship coefficients (two from Colombia, two from Brazil, one from the United States, two from Kenya, and one from Senegal; see supplementary fig. S1, Supplementary Material online, for kinship coefficients

between all pairs of specimens), leaving a total of 123 remaining specimens.

Patterns of Genetic Diversity within and between Populations

To investigate the extent and patterns of population structure of our data set, we used principal component analysis. The first principal component for each chromosome explained 9–10% of the variance and generally separated samples from Africa versus samples from North and South America (figs. 1 and S2, Supplementary Material online). The second principal component, explaining ~4% of the variance, separated samples from North and South America along a cline stretching from the US sample to the Colombian sample. The two Colombian cohorts were clearly separated, with the Brazilian sample falling in between.

On each chromosome, the Senegal sample was closer in the first principal component to the North and South American samples than were the Gabon and Kenya samples. This is in keeping with previous findings (Crawford et al. 2017; Rose et al. 2020) that have suggested a Senegalese origin of the domestic populations. Overall, the three African countries exhibit greater separation from one another on chromosomes 2 and 3 than on chromosome 1, while the American countries show greater separation on chromosomes 1 and 3.

Next, we quantified the extent of genetic diversity and the prevalence of rare versus common polymorphisms by measuring nucleotide diversity (π) and Tajima's D. Nucleotide diversity was lower in the American samples than in the African samples (0.0058 vs. 0.0091 on average; supplementary fig. S3 and table S4, Supplementary

Material online), consistent with previous reports of greater allelic richness in African samples (Gloria-Soria et al. 2016). These levels of nucleotide diversity are somewhat lower than those reported elsewhere (e.g., Rose et al. [2020]), and we observed that the magnitude of π varied across the different filtering schemes that we examined. However, the rank order of π values among sampling locations remained relatively constant across filtering schemes, and we therefore conclude the comparative values between cohorts, especially the lower diversity in the Americas than Africa, is not an artifact of data quality.

Tajima's D was higher in the American cohorts, consistent with the deficit of rare variants caused by a population contraction (supplementary fig. S3 and table S4, Supplementary Material online). The Gabonese cohort also had unexpectedly high Tajima's D compared to the other African samples. In all cohorts, nucleotide diversity tended to drop sharply near the centromeres, recovering and then trending slightly downward towards the telomeres (supplementary fig. S4, Supplementary Material online). For most cohorts and chromosomes, we also observed a reduction in Tajima's D near the centromeres (supplementary fig. S4, Supplementary Material online).

Next, we quantified between-population diversity using F_{ST}. As expected, mean genome-wide F_{ST} values were lowest within continents and elevated between continents (supplementary table S5, Supplementary Material online). The decreased F_{ST} between all American populations and Senegal, compared to American populations and Gabon or Kenya, are consistent with the West African origin of domestic Ae. aegypti.

We also measured linkage disequilibrium (LD) within 500 kb windows and found that LD in the Brazilian population was noticeably elevated on all three chromosomes,
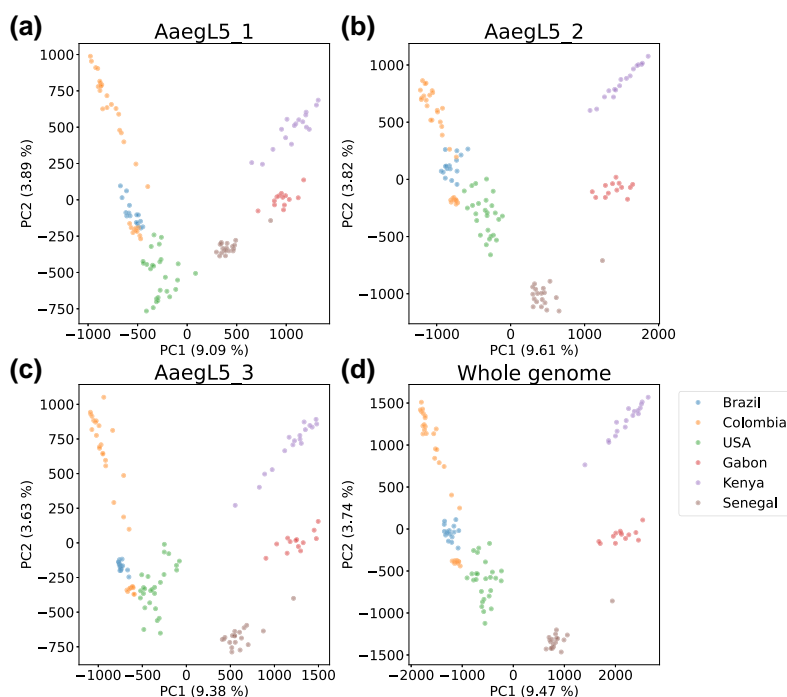

Fig. 1. Principal component analysis of all filtered SNPs, showing each chromosome separately (a–c) and combining all three chromosomes (d).

while the other five populations overlapped more closely (supplementary fig. S5 and table S6, Supplementary Material online). Indeed, the mean linkage disequilibrium is much higher in the Brazilian sample than in the other five populations—roughly twice as high in Gabon and Senegal, the two populations with the next highest mean value of r^2 (supplementary table S6, Supplementary Material online).

Strong Signatures of Selective Sweeps in Insecticide Resistance Loci

We scanned the genome for signatures of recent selective sweeps using SweepFinder2 (DeGiorgio et al. 2016), with the genome-wide empirical site frequency spectrum (SFS) for each country (supplementary fig. S6, Supplementary Material online) used as the "neutral" SFS (Materials and Methods). Strong sweep signals were observed in all populations, with the most prominent peaks in SweepFinder2's composite likelihood ratio (CLR) seen in the samples from Brazil, Colombia, Gabon, and Senegal (fig. 2). CLR peaks were also observed at the beginnings or ends of chromosomes in all populations except Senegal; the size and shape of these may be a result of the paucity of recombination and the lack of polymorphisms in these regions, respectively.

We individually examined seven peaks or groups of peaks (marked with stars in fig. 2) that contained a CLR value greater than 1,000; here, we discuss a subset of genes overlapping the regions of these peaks, with all genes listed in supplementary table S7, Supplementary Material online.

Within the American samples, two of the three prominent peaks, one each on chromosomes 2 and 3, are shared to some extent by both the Brazilian and the Colombian samples. A peak on chromosome 2 ("*a"), from approximately 351 to 352 Mb, overlaps 27 protein-coding genes and four noncoding RNAs (figs. 2b and S7a and table S7, Supplementary Material online); 18 protein-coding genes have an unspecified functional product, while the genes with specific predicted functions include an ankyrin gene (AAEL013466; discussed below) as well as a cluster of 8 glutathione transferases. Glutathione transferases have a well-known function in insecticide resistance and have been implicated in resistance to several classes of insecticides (Enayati et al. 2005).

A secondary peak is found nearby in the Brazilian sample, but not the Colombian sample, from approximately 356.3 to 356.5 Mb on chromosome 2. There are one tRNA and six protein-coding genes in the vicinity of this peak (supplementary fig. S7a and table S7, Supplementary Material online), three of which have a specific functional prediction. One, CYP4H31/AAEL002085, is a cytochrome P450 identified as upregulated in larvae selected to be resistant to the neonicotinoid insecticide imidacloprid (Riaz et al. 2013). Its expression levels are altered in larvae exposed to DDT, malathion, and temephos (El-Garj et al. 2016), and it has been implicated in deltamethrin resistance in *Culex* (Li et al. 2021).

The highest peak identified in any cohort occurs in the Brazilian sample, near 315 Mb on chromosome 3; this peak also occurs in the Colombian sample at a lesser magnitude ("*b", fig. 2c). Closer investigation shows this peak extending from approximately 314.5 to 318 Mb in the Brazilian sample (supplementary fig. S7b, Supplementary Material online). This region overlaps 36 protein-coding genes and 11 noncoding RNAs (supplementary table S7, Supplementary Material online). Of the protein-coding genes, 13 have a specific functional prediction present in VectorBase; we highlight five genes that are known from the literature. The Vanin-like protein AAEL006034 has been identified as potentially involved in the metabolism of the insecticide indoxacarb (Smith et al. 2018). The calmodulin gene AAEL012326 is overexpressed in a strain susceptible to dengue compared to one refractory to dengue (Chauhan et al. 2012) and produces a protein product that is found in extracellular vesicles produced by *Ae. aegypti* cells infected by dengue, but not uninfected control cells (Gold et al. 2020). The putative 60S ribosomal protein
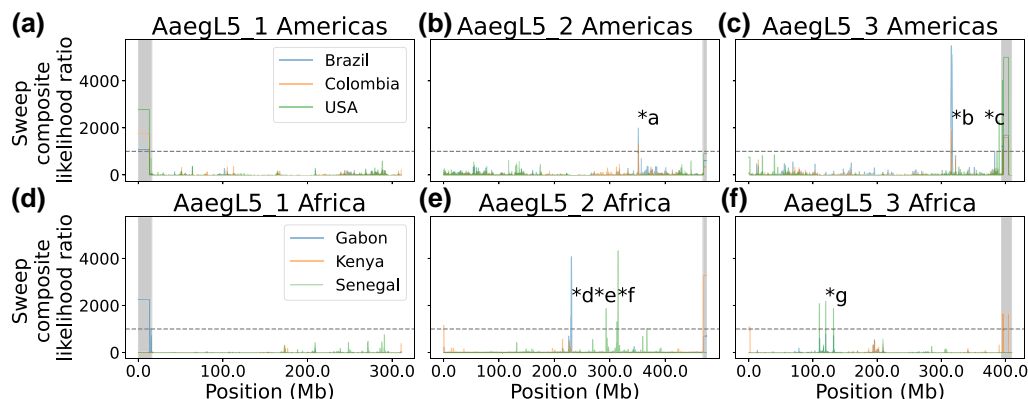
FIG. 2. Genome-wide scan of candidate selective sweeps (a-f). Shaded regions at the ends of chromosomes indicate putative technical artifacts excluded from identification of outliers. Each * with a letter marks a sweep examined in more detail in the text and in the corresponding panel of supplementary fig. S7, Supplementary Material online. Horizontal dashed lines indicate our threshold of CLR = 1,000, used in identifying sweeps examined further.

gene AAEL011587 is upregulated in response to chikungunya infection and to chikungunya and dengue coinfection (Shrinet et al. 2017). The heme peroxidase gene HPX1 (AAEL006014) is overexpressed in response to pyrethroid (permethrin) selection (Saavedra-Rodriguez et al. 2012).

Finally, this peak near 315 Mb on chromosome 3 also includes Vgsc (AAEL023266), which has been extensively associated with insecticide resistance via nonsynonymous single nucleotide changes that cause resistance to pyrethroids, DDT, or both (Hu et al. 2011; Du et al. 2013; Hirata et al. 2014; Haddi et al. 2017). In the Brazilian sample, Vgsc is roughly centered in the putative sweep, while in our Colombian sample, the region of highest sweep likelihood is upstream of Vgsc. In addition to being narrower in genomic distance, the magnitude of the putative sweep is also lower in the Colombian sample than in the Brazilian one. Additionally, one of the genes identified in the prominent sweep on chromosome 2 discussed above is ankyrin gene 2,3/unc44 (AAEL013466); ankyrin proteins are associated with, and directly interact with, voltage-gated sodium channels (Bennett and Baines 2001; Bennett and Chen 2001; Jenkins and Bennett 2001). In Ae. aegypti, variation in ankyrin domain-containing genes has been associated with pyrethroid resistance (Campbell et al. 2019); in addition, the ankyrin repeat region-containing gene AAEL011338 was identified as harboring a locus associated with pyrethroid resistance in a Brazilian sample (Cosme et al. 2022).

The US cohort contains a peak near 390 Mb on chromosome 3, as well as a smaller peak near 387 Mb ("c," figs. 2c and S7c, Supplementary Material online). The vicinity of these two peaks overlaps 49 protein-coding genes, 22 of which have specific functional predictions, and 16 noncoding RNAs (supplementary table S7, Supplementary Material online), three of which are tRNAs. The glycosyl transferase gene AAEL000559 contains nonsynonymous nucleotide variants associated with deltamethrin resistance (Faucon et al. 2015, 2017). The putative pacifastin inhibitor AAEL000551 is upregulated after dengue infection (Soares et al. 2018). The C-type lectin gene mosGCTL-1 (AAEL000563) is well-studied in terms of mosquito immunity and facilitates infection of Ae. aegypti with West Nile virus (Cheng et al. 2010). 40S ribosomal protein S21 gene RpS21 (AAEL000529) is upregulated following dengue infection (Shrinet et al. 2017), while expression of AAEL007260 is downregulated following dengue infection (Barón et al. 2010).

Next, we examined two peaks present in the Gabon and Kenya samples. On chromosome 2, a sweep signal present in these two samples stretches from approximately 227.5 to 228.2 Mb, overlapping nine genes ("*d", figs. 2e and S7d, Supplementary Material online). The two genes with specific functional predictions fully encompassed by this sweep are the mitochondrial ribosomal protein L52 gene AAEL023209 and SCRB8 (AAEL000227), a class B scavenger receptor. However, just downstream, the Gabon cohort shows another signal of a putative selective sweep, from approximately 229.5 to 231.25 Mb. This peak overlaps 39 genes, 19 of which are protein coding and 10 of which have a specific functional prediction (supplementary table S7, Supplementary Material online). Here, we discuss four known in the literature. The gene AAEL008073, predicted to encode an RNA-binding protein (Dickson et al. 2017), is upregulated in blood-fed Ae. aegypti compared to sugar-fed (Bonizzoni et al. 2011); it also displays very low heterozygosity in populations from Mexico, Thailand, and Senegal (Dickson et al. 2017). In addition, dsRNA silencing of this gene and infection with chikungunya in an Ae. aegypti cell line results in a significant decrease in chikungunya RNA levels (Dubey et al. 2022). When notch (AAEL008069), another nearby gene, is silenced, in vitro viral load of DENV-2 (dengue serotype 2) is significantly increased 6 days postinfection compared to control mosquitoes (Xiao et al. 2014). A splice variant of the gonadotropin-releasing hormone receptor AAEL011325 is downregulated in larvae exposed to brackish water compared to freshwater (Durant et al. 2021). Finally, the nicotinic acetylcholine receptor subunit gene Aaea6 (AAEL008055) is upregulated in response to posteclosion application of exogenous juvenile hormone (Zhu et al. 2010) and has also been experimentally identified as the target of the insecticide spinosad (Lan et al. 2021); more generally, nAChR genes are known as the targets of neonicotinoid insecticides (Casida 2018).

Several putative selective sweeps are detected in Senegal on chromosome 2. Two are found between 293 and 295.5 Mb ("*e", figs. 2e and S7e, Supplementary Material online). This area overlaps 38 protein-coding genes, 15 of which have specific functional predictions, and 11 noncoding RNAs. The fas-associated protein gene AAEL004287 contains a SNP highly differentiated between Bacillus thuringiensis israelensis (Bti)-resistant and susceptible lab strains (Paris et al. 2013). Expression of the ATP citrate synthase AAEL004297 is downregulated in response to both Zika infection and Wolbachia infection and further downregulated with Zika and Wolbachia coinfection compared to Wolbachia infection only (Martins et al. 2021). Expression of AAEL018319 decreases to undetectable levels following chikungunya infection (Cui et al. 2020). Expression of the ABC transporter gene AAEL016973 is significantly upregulated in a strain originating from wild-caught mosquitoes showing resistance to pyrethroids, carbamates, and DDT, compared to susceptible mosquitoes (Lien et al. 2019). Finally, expression of the deoxyuridine 5′-triphosphate nucleotidohydrolase gene AAEL012629 is consistently upregulated from 1 to 14 days after dengue infection (Bonizzoni et al. 2012).

Senegal also shows a series of sweep signatures on chromosome 2 between 312 and 316 Mb ("*f", figs. 2e and S7f, Supplementary Material online). Collectively, these peaks overlap 35 genes, 26 of which are protein coding and 11 of which have a specific predicted function (supplementary table S7, Supplementary Material online). The aldo-keto reductase AAEL007275 is upregulated after larval imidacloprid exposure (Riaz et al. 2009) and

upregulated after one generation of selection for resistance to temephos (Saavedra-Rodriguez et al. 2014). The alpha-amylases AAEL010540 and AAEL010532 contain nonsynonymous changes in putative receptors for *Bti* Cry toxins that are nearly fixed between control strains and strains selected for resistance to individual Cry toxins (Stalinski et al. 2016). Another study showed AAEL010540 underexpression in strains resistant to certain Cry toxins compared to a control strain (Després et al. 2014). ABC transporter *ABCC13* (AAEL023524) is downregulated in late dengue infection and overexpressed throughout yellow fever expression in vivo (Kumar et al. 2021).

In addition, five genes in this region (AAEL012960, AAEL013222, AAEL022016 [alternate gene ID: AAEL013219], AAEL010533, and AAEL023254 [previous gene ID: AAEL013215]) were identified as among the top 25 most differentiated genes between an urban population in Senegal versus a forest population from Senegal together with a population from Uganda (Crawford et al. 2017); AAEL012960 and AAEL023254 contain exonic SNPs highly differentiated in that comparison. This sweep also overlaps a region identified as containing many variants associated with human specialist ancestry, and as highly diverged based on the Population Branch Statistic, in these specimens (Rose et al. 2020), congruent with our identification of a putative sweep.

Finally, the cohort from Senegal also shows several putative selective sweeps on chromosome 3 between 100 and 140 Mb ("*g", figs. 2*f* and S7*g*, Supplementary Material online). The three most prominent peaks in this region overlap 25 protein coding genes, 7 of which have a specific functional prediction (supplementary table S7, Supplementary Material online).

Selective Sweep Signatures Are Enriched for Insecticide Resistance Genes in the Americas

In addition to examining individual sweeps, we looked for categories of genes disproportionately represented in sweeps across the genome. Here, we used the individual SweepFinder2 test sites, distributed approximately every 1 kb throughout the genome. We identified outliers in these test sites at the 99th percentile in each country, with enrichment testing performed in Gowinda with and without a 1-kb buffer up- and downstream of each gene (supplementary fig. S8, Supplementary Material online). We found that eight Gene Ontology (GO) terms were enriched in three populations when the 1-kb buffer was used and seven when it was not; no GO terms were enriched in four or more populations. Seven of these terms (DNA binding; DNA-templated transcription, initiation; chromosome; host cell nucleus; nucleosome; nucleosome assembly; and nucleus) related to DNA or DNA binding. Further examination of the genes underlying this enrichment showed that it was partially attributable to the presence of a cluster of histone genes at approximately 340 Mb on chromosome 3, overlapped by outlier windows in three populations (supplementary fig. S9, Supplementary

Material online). This highlighted the possibility of clusters of genes influencing the enrichment results, so we implemented a permutation test to test for significantly enriched GO terms after accounting for physical location (see Materials and Methods). After controlling for gene clustering this way, when examining genes overlapping outlier windows at the 99th percentile of SweepFinder2's CLR score, one group of related GO terms remained significantly enriched in the US sample (ribonucleoside metabolic process, purine nucleoside metabolic process, purine ribonucleoside metabolic process, purine-containing compound catabolic process, purine nucleoside catabolic process, adenosine catabolic process, nucleoside catabolic process, nucleobase-containing small molecule catabolic process, ribonucleoside catabolic process, adenosine metabolic process, purine ribonucleoside catabolic process, and glycosyl compound catabolic process) at a *q* value of 0.05; enrichment of all these terms was driven by the same three genes, AAEL003214, AAEL026165 (alternate gene ID: AAEL005672), and AAEL005676. The difference in results between our two enrichment strategies highlights the need to account for the physical clustering of genes when examining signatures of selection, which are autocorrelated across the genome (Schrider and Kern 2017; Aberg et al. 2018; Tobler et al. 2020).

We also tested whether genes potentially associated with insecticide resistance genes were enriched in outlier windows. To do this, we assembled a list of 152 cytochrome P450 genes, glutathione transferase genes, and carboxylesterase genes by searching the annotated gene file (*Ae. aegypti* LVP_AGWG AaegL5.3) available through VectorBase. We then manually added AAEL023266 (*Vgsc*) and AAEL023844 (a carboxylesterase-like gene that is part of a cluster of *CCE* genes and has been extensively associated with insecticide resistance [Cattel et al. 2021a, 2021b; Sene et al. 2021])—we note that the decision to include *Vgsc* in enrichment testing was made prior to conducting our SweepFinder2 scan. Outlier windows at the 99th percentile were significantly enriched at a *P* value of 0.05 for these genes in Colombia and the United States, but not the other four countries; however, the *P* value for enrichment in the Brazil cohort was 0.056. These results suggest that there has been strong selection for insecticide resistance genes in the Americas.

Patterns of Haplotypic Diversity in Sweeps at Insecticide Resistance Loci

As an alternative mode of characterizing selective sweep signatures, we used the multilocus genotype (MLG) statistics defined in Harris et al. (2018) to investigate two of our candidate selective sweeps. These statistics are analogous to the H12 and H2/H1 statistics defined in Garud et al. (2015), which use relative frequencies of the most frequent and second most frequent haplotypes to distinguish between soft and hard sweeps; however, unlike H12 and H2/H1, the corresponding G123 and G2/G1 can be used on MLGs concatenated from unphased data. Regions of

elevated G123 indicate a putative selective sweep; in such a region, elevated G2/G1 levels are consistent with a soft sweep, while decreased G2/G1 levels are more consistent with a hard sweep (because the most common MLG is present at much higher frequency than the second most common MLG).

We reexamined the two most conspicuous selective sweep signatures in terms of their apparent involvement in insecticide resistance, regions "a" and "b" described above, using an alternative approach for detecting and characterizing selective sweeps: the multilocus genotype statistics described by Harris et al. (2018) and Kern and Schrider (2018). Sweep "a" (figs. 2b and S7 and table S7, Supplementary Material online), in Brazil and Colombia, overlaps a cluster of glutathione transferases. Plotting G123 (the H12 analog for unphased data) and G2/G1 from 350 to 355 Mb on chromosome 2 in the samples from Brazil, Colombia, and the United States (supplementary fig. S10a, Supplementary Material online) shows elevated G123 from approximately 351 to 353.5 Mb in Brazil and Colombia; this agrees reasonably well with the CLR statistic from the same region, which shows elevated values from approximately 351 to 352.25 Mb in Brazil and Colombia. Consistent with the CLR results, the G123 statistic is not markedly elevated in this genomic region in the sample from the United States. In both South American cohorts, the G2/G1 statistic is noticeably decreased in the same region. Similar results are seen on chromosome 3 from approximately 314.5 to 318 Mb in sweep "b," which overlaps Vgsc; G123 is elevated in the Brazilian and Colombian samples, and G2/G1 is decreased (supplementary fig. S10b, Supplementary Material online). In this region, the US cohort does show elevated G123 values, but not to the extent seen in the South American cohorts. Outlier windows are also enriched in these two regions in the South American cohorts (supplementary

table S8). Thus, both the "a" and "b" sweep regions from figure 2 show signatures of positive selection according to both G123 and SweepFinder's CLR. However, we cannot be certain that the low values of G2/G1 observed in these regions should be taken as evidence of hard selective sweeps, because we do not currently have adequate information to perform sweep model selection using these statistics—this would require a well-fitting demographic model for each cohort examined (Garud et al. 2021). We also note that values of G2/G1 may not behave as expected from simulations of single-sweep scenarios when the locus in question has experienced more than one sweep in rapid succession, as appears to have been the case for Vgsc in Colombia (see Discussion).

Strong but Variable Signatures of Selection around *Vgsc* in South America

Given its important role in insecticide resistance in numerous species, we conducted a more thorough analysis of *Vgsc* and the surrounding region. We began by conducting principal component analysis of SNPs inside *Vgsc* (from 315,926,360 to 316,405,639 on chromosome 3), which showed specimens clustering into an approximate triangle (fig. 3a). One of the points of the triangle, on the right side, is anchored by a cluster of Senegalese, Gabonese, and Kenyan specimens, and the x-axis (the first principal component, which explains 20% of the variance) approximately divides African specimens from American specimens, with, however, much intermixing. The base of the triangle, on the left, comprises American specimens only; the y-axis (the second principal component, explaining 14% of the variance) shows American specimens in three distinct and roughly equidistant clusters, with some spread between them. All Colombian and Brazilian specimens can be easily assigned to one of the clusters, and a subset of
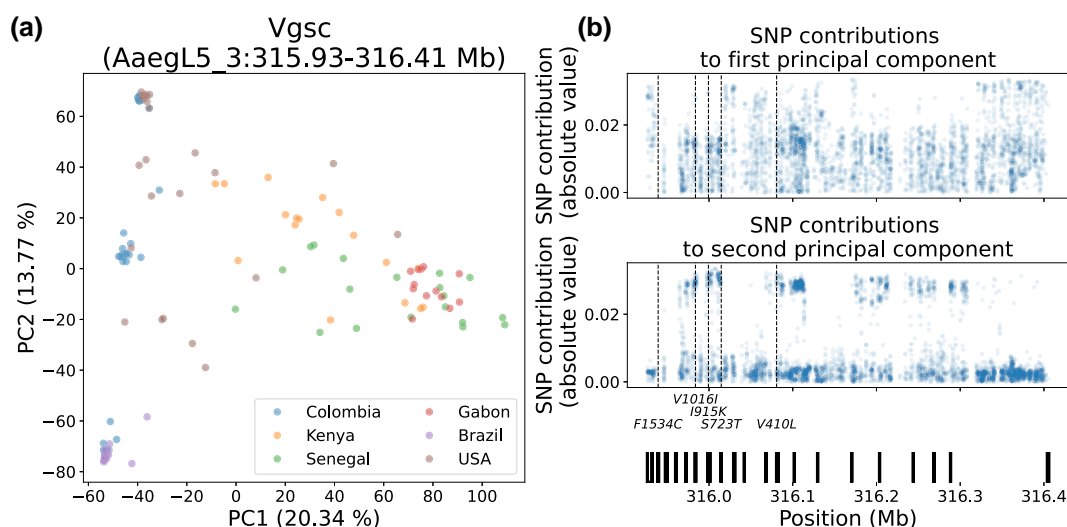

Fig. 3. Principal component analysis (a) and loadings (b) of *Vgsc*. Vertical dashed lines on the loadings panels indicate the locations of the five focal loci listed below the plot. The vertical black bars at the bottom of (b) show the location of exons in *Vgsc*.

Table 1. Genotypes at Selected Loci Previously Identified as Associated with Pyrethroid Resistance in *Vgsc*[a].

	F1534C (315939224)			V1016I (315983763)			I915K (315999297)			S723T (316014588)			V410L (316080722)			Total *n*
	F/F	F/C	C/C	V/V	V/I	I/I	I/I	I/K	K/K	S/S	S/T	T/T	V/V	V/L	L/L	
Brazil	0 (0)	0 (0)	16 (1)	14	0	0	14	0	0	15	0	0	15	0	0	16
Colombia	0 (0)	0 (0)	32 (19)	6	16	9	6	15	9	4	14	8	6	16	10	32
United States	3 (0)	0 (0)	23 (3)	5	4	14	2	3	15	8	3	12	4	5	13	26
Gabon	13 (2)	0 (0)	0 (0)	12	0	0	11	1	0	11	0	0	10	0	0	13
Kenya	15 (0)	0 (0)	2 (0)	17	0	0	4	10	2	14	0	0	16	0	0	17
Senegal	19 (7)	0 (0)	0 (0)	18	0	0	8	9	0	18	0	0	19	0	0	19

[a]Genotypes for F1534C were obtained from bcftools mpileup; the numbers in parentheses reflect the numbers remaining in each category after genotypes with qualities less than 20 were masked. Genotypes for the other four sites were derived from GATK HaplotypeCaller and were calculated after genotypes with qualities less than 20 were masked. Total numbers within a population vary between sites because not all specimens could be confidently genotyped at all sites; the total number of specimens for each population is given in the far-right column.

the specimens from the United States take intermediate positions.

We then investigated genotypes for each specimen at a number of nonsynonymous SNPs previously shown to be either causally linked or associated with pyrethroid resistance (our "focal sites"): F1534C (Kawada et al. 2009; Harris et al. 2010; Hu et al. 2011; Du et al. 2013; Hirata et al. 2014), V1016I (Saavedra-Rodriguez et al. 2007, 2008; Harris et al. 2010; Du et al. 2013), I915K (Mack et al. 2021), S723T (Saavedra-Rodriguez et al. 2019; Mack et al. 2021), V410L (Haddi et al. 2017; Saavedra-Rodriguez et al. 2019), and Q1853R (Kelly et al. 2021) (table 1; we use the *M. domestica* nomenclature). F1534C was not called in any specimen in our filtered data set, although we were later able to genotype this locus as discussed below (these genotypes were not included in the PCA but do appear in table 1); Q1853R was confidently called as invariant in 107 specimens and either called as invariant with low confidence or uncalled in the remaining 16 and thus excluded from further analysis. The remaining four sites were confidently called and were polymorphic in our data, with high LD between them in the Colombian sample. The genotypes at these four positions also correspond to the three clusters seen on the left side of figure 3a, with the bottom cluster corresponding to the specimens homozygous for the reference or wild-type alleles at the four sites, the top cluster corresponding to the specimens homozygous for alleles associated with resistance, and the middle group corresponding to specimens heterozygous at these sites.

We also looked at the contribution of each SNP to the groupings in the PCA (fig. 3b), and specifically to the second principal component, along which many American specimens cluster into discrete groups. The loadings for this principal component show several clusters of SNPs with some of the highest contributions in the PCA. When we plot the location of five associated resistance SNPs, we see that, except for F1534C (which was not included in the PCA, as previously described), they correlate with clusters of SNPs that are outliers in terms of their contribution to the second principal component. In addition to the clusters associated with the known resistance loci, there are additional clusters in the 5′ portion of the gene (note that *Vgsc* is transcribed on the reverse strand of the reference). Focusing on exonic SNPs beyond 316.1 Mb, we identified three (316,268,746; 316,268,894; and 316,269,032) with a PCA loading of greater than 0.025 that overlap two adjacent exons; all three SNPs are found in the untranslated region. *Vgsc* is extensively alternatively spliced (Lin et al. 2009), so these SNPs may be relevant in that regard or may influence expression levels; however, they may also have no functional importance and simply be in high LD with the focal polymorphisms and/or another causal polymorphism not present in our filtered data set.

Examining nucleotide diversity in the region of *Vgsc*, from 310 to 320 Mb on chromosome 3, we observe a decrease in diversity in the South American populations from approximately 315 to 318 Mb (figs. 4a and S11, Supplementary Material online). The specimens from the United States do not show as marked of a drop in this region; however, this heterogeneous cohort includes specimens with ancestry from multiple source populations (Lee et al. 2019; Kelly et al. 2021), which could give the appearance of high nucleotide diversity even if diversity is low within some ancestry groups. Relative to the Brazilian sample, the nucleotide diversity of the Colombian sample is elevated within *Vgsc* itself, and in the distal region to approximately 318 Mb, where nucleotide diversity increases in the Brazilian sample. This is in keeping with the number of clusters observed for each cohort in the PCA of the *Vgsc* region as well as the genotypes shown in table 1: one cluster for Brazil and three for Colombia, indicating multiple haplotypes in the latter but not the former.

Although the SweepFinder2 results suggest a strong skew in allele frequencies in this region, to more directly examine this, we calculated Tajima's *D*. In the American samples, Tajima's *D*, which is otherwise generally higher than in the African samples, drops sharply proximal to *Vgsc* (figs. 4b and S11, Supplementary Material online). This drop is especially pronounced in the Colombian sample. However, within *Vgsc* itself, Tajima's *D* is sharply elevated in the Colombian sample, which is consistent with the presence of multiple haplotypes in this genomic region within Colombia.
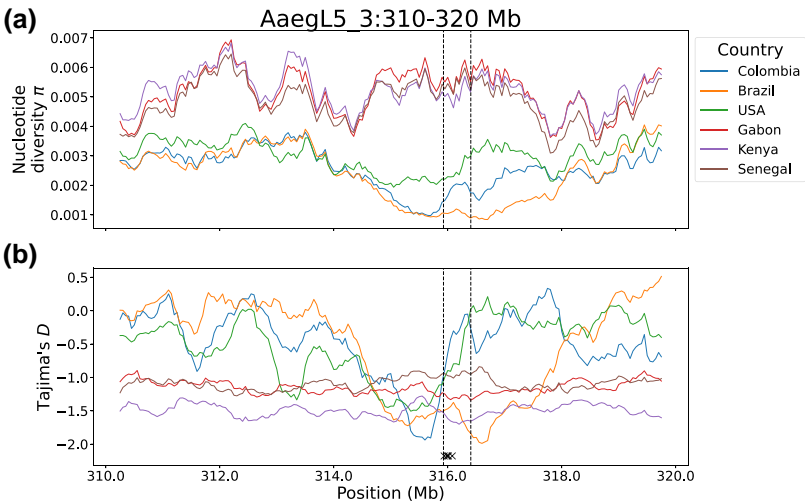
FIG. 4. Nucleotide diversity (a) and Tajima's D (b) in six cohorts, on chromosome 3 from 310 to 320 Mb. The vertical dashed lines mark the boundaries of Vgsc (315,926,360–316,405,639). The five Xs at the bottom of Tajima's D plot indicate the locations of the five focal loci.

Finally, linkage disequilibrium (r^2, calculated as described in Materials and Methods) between individual variants within *Vgsc* in the three populations from the Americas shows that regions of elevated LD are particularly prominent in the South American populations (supplementary fig. S12, Supplementary Material online). One such block occurs from approximately 315.97 to 316.25 Mb—we note that this block contains the V1016I, I915K, S723T, and V410L polymorphisms, which are all in high LD with one another in the Colombian sample (table 2; compare to US sample, supplementary table S9, Supplementary Material online). In addition, another block of elevated LD occurs from approximately 316.29 Mb to the end of the gene. The US population, in contrast, shows a more diffuse pattern of elevated LD, especially in the region between the start of the gene and approximately 316.20 Mb.

These findings are particularly striking because the Brazilian sample, though wild-type for all previously identified SNPs associated with pyrethroid resistance that could be confidently genotyped in our data, still shows a sharp drop in nucleotide diversity and Tajima's D characteristic of a selective sweep; in addition, SweepFinder2 identified a strong candidate selective sweep in both the Brazilian and Colombian samples in this region. Indeed, SweepFinder2's CLR peak was substantially higher and broader in Brazil than Colombia.

These findings, as well as the well-known role of the F1534C polymorphism in insecticide resistance, led us to reexamine this locus in our data. F1534C was also not present in the unfiltered variants, indicating that this site had not been simply filtered out. In our set of unfiltered variants, we found a total lack of coverage in all specimens in the vicinity of this SNP, spanning from approximately 315,937,705 to 315,945,225 bp. However, coverage in the original alignments appeared relatively normal (supplementary fig. S13a, Supplementary Material online). Inspection of the alignments revealed low mapping quality in this region (supplementary fig. S13b, Supplementary

Table 2. Linkage Disequilibrium between Four Known or Putative Insecticide Resistance Loci in the Colombian Sample[a].

	315983763 (V1016I)	315999297 (I915K)	316014588 (S723T)	316080722 (V410L)
315983763 (V1016I)	—	—	—	—
315999297 (I915K)	1.0	—	—	—
316014588 (S723T)	0.880309	0.885714	—	—
316080722 (V410L)	1.0	1.0	0.885714	—

[a]Linkage disequilibrium, as measured by r^2 (Materials and Methods), after masking all genotypes with a genotype quality less than 20. A fifth known insecticide resistance locus, 315939224 (F1534C), is homozygous for the alternate (resistant) allele in every specimen in this sample.

Material online), with a high proportion of reads aligning to this region also mapping to another region, scaffold NIGP01000811; reads that align equally well to multiple locations are given a mapping quality of 0 (Li et al. 2008) and thus are not considered by GATK's HaplotypeCaller using the default settings, which remove all reads with a mapping quality below 10 (https://gatk.broadinstitute.org/hc/en-us/articles/360036889192-MappingQualityReadFilter). The similarities between this scaffold and *Vgsc* have previously been noted (Itokawa et al. 2019). Aligning NIGP01000811 to chromosome 3 of AaegL5 revealed that most of the scaffold aligns to the vicinity of the *Vgsc* gene, but the first approximately 4.9 kb does not confidently align anywhere on chromosome 3 (supplementary fig. S14, Supplementary Material online).

We therefore regenotyped chromosome 3 from 315 to 317 Mb, as well as the entirety of NIGP01000811, using bcftools version 1.9 mpileup, which has a default minimum mapping quality of 0 (https://samtools.github.io/bcftools/bcftools.html). Using this method, we called putative genotypes at F1534C, revealing that every South American specimen is homozygous for the resistant allele (table 1).

In the Brazilian sample in particular, the genotypes are of very low quality, reflecting low confidence in the genotype calls. However, this is expected, as read mapping qualities influence the calculation of genotype likelihoods and therefore qualities (Danecek et al. 2021).

While one explanation for this apparent duplication is a technical artifact of assembly, we cannot rule out the presence of a genuine duplication, and therefore, we are less confident in the genotypes at F1534C than at the other putative insecticide resistance loci; however, we present these data given the established importance of this SNP on insecticide resistance and vector control. To assess the effect of using a different variant caller on clustering observed in this region, we repeated the principal component analysis using the variants genotyped by bcftools (supplementary fig. S15, Supplementary Material online); the triangular pattern observed in the calls generated by GATK is generally recapitulated, though excluding low-quality genotypes introduces a great deal of noise.

In addition, we calculated nucleotide diversity separately within two sets of South American samples: those in the top and bottom clusters of the *Vgsc* PCA (supplementary fig. S16, Supplementary Material online). This reveals a sharp valley in nucleotide diversity in the bottom cluster, which contains the specimens carrying a known resistance allele at F1534C only, and an even sharper decrease in nucleotide diversity in the top cluster, which contains specimens carrying known or putative resistance alleles at all five focal loci.

We also calculated *Vgsc*-region F_{ST} between the specimens in the top and bottom clusters of the *Vgsc* PCA. Viewing the region from 310 to 320 Mb shows two extended regions of low F_{ST} between the two clusters just proximal to the *Vgsc* gene, from approximately 314.3 to 314.5 Mb and 315.2 Mb to the proximal gene boundary of *Vgsc* (supplementary fig. S17a, Supplementary Material online), with sharp spikes of F_{ST} within *Vgsc* itself; zooming in more closely reveals the extremely abrupt changes between regions of low F_{ST} and regions of very high F_{ST} (supplementary fig. S17b, Supplementary Material online). Finally, viewing this region in conjunction with variant counts as well as areas of repetitive DNA or poorly mapping sequence (supplementary fig. S17c, Supplementary Material online) shows that such areas are distributed through the *Vgsc* region, and the F_{ST} peaks cannot be explained as artifacts of repetitive regions or regions with few variants.

Finally, drawing firm conclusions about the North American sample published in Lee et al. (2019) and reanalyzed here is difficult because of the multiple founding populations and small per-cluster sample sizes; however, when we decompose the portion of the sample originating in California into the clusters noted in Lee et al. (2019) (two clusters from central California and one cluster from southern California plus one site in central California), we see decreased nucleotide diversity upstream of *Vgsc* and increased nucleotide diversity within *Vgsc* itself in two of the three clusters (supplementary fig. S18, Supplementary Material online), consistent with the data observed in Colombia. In addition, Mack et al. (2021) observed high frequencies of the resistant allele at the five focal loci studied here in several localities in central California; data from Kelly et al. (2021) similarly suggest an increase in the frequency of resistance alleles in Californian samples between 2014 and 2018. Finally, when examining the PCA of high-quality SNPs within *Vgsc*, a subset of the US cohort clusters very closely with the Colombian specimens that carry all five focal insecticide resistance alleles (fig. 3). Thus, it appears likely that the resistance haplotype observed in Colombia is present in parts of North America as well.

Discussion

These whole-genome data from six country-level cohorts of *Ae. aegypti* on three continents allowed us to consider the history of this species in detail. In general, our findings are consistent with the known history of this species. Tajima's *D* values are either negative or close to zero in all three African cohorts, while they are positive in the three American cohorts; in addition, the Senegal cohort shows values of Tajima's *D* intermediate between the Kenyan cohort and the American cohorts. This is in keeping with previous studies of populations from these regions (Crawford et al. 2017) and is consistent with the idea of a population bottleneck associated with domestication, perhaps followed by a second bottleneck associated with expansion out of Africa (Crawford et al. 2017). The positive Tajima's D in American cohorts may reflect a relatively recent contraction due to the continent-wide eradication effort, but population genetic evidence of subsequent recovery and expansion will require larger sample sizes to detect the excess of very rare alleles produced by such a change (Tennessen et al. 2012).

Of the three American populations, Tajima's *D* is highest in the Brazilian sample; linkage disequilibrium values are also sharply elevated in this population compared to all others. Eradication of *Ae. aegypti* was claimed to be complete in Brazil (Soper 1965), but no such claim was made for Colombia. Our observations are consistent with a historical reduction in population size in Brazil, perhaps due to a founder effect during reinfestation from another country.

In addition, we present the first genome-wide scans for positive selection in *Aedes* using SNP-resolution data, identifying candidate selective sweeps in six countries. In multiple cohorts, we identify signatures of strong selective sweeps, with CLR scores over 1,000, near loci implicated in insecticide resistance, including the voltage-gated sodium channel gene *Vgsc* (Brazil and Colombia); glutathione S-transferases (Brazil and Colombia); a cytochrome P450 (Brazil); and a nicotinic acetylcholine receptor subunit gene (Gabon).

We caution that genes associated with insecticide resistance may be disproportionately well-annotated and are likely to be disproportionately well-investigated, because

of their obvious public health implications. Some sweep signatures overlap a large number of genes, many of whose functions are currently unannotated, making detailed conclusions about the targets of selection impossible; in addition, the peak of a sweep signature may not always correspond to the target of selection (Schrider et al. 2015). However, insecticide usage is known to cause strong selective pressure in many other insects (Garud et al. 2015; Miles et al. 2017; Walsh et al. 2018; Weedall et al. 2020; Xue et al. 2021; Pélissié et al. 2022), and the variety of genes implicated in insecticide resistance occurring near these sweeps is a potential sign of *Ae. aegypti*'s agility in responding to these insecticides.

The putative selective sweep that is perhaps the most interesting, because of the contrasting patterns shown in two different South American countries, is at *Vgsc*. Examining this genomic region in Colombia reveals two observations: one, somewhat unexpectedly, the putative selective sweep in this population appears to be localized upstream of the gene, not within *Vgsc* itself, based on results from SweepFinder2 as well as elevated Tajima's *D* within *Vgsc*; two, the clustering of the Colombian sample into three tight groups along principal components suggests that two long-range haplotypes linking several loci associated with insecticide resistance are present in the population, with one haplotype containing the resistant alleles at at least four previously identified insecticide resistance SNPs and the other containing the wild-type alleles. Indeed, examining the PCA loadings shows that these polymorphisms line up well with clusters of SNPs driving the separation between PCA clusters. We note that this pattern of three equidistant clusters on a PCA is the characteristic signature of a polymorphic chromosomal inversion (Love et al. 2019; Ma and Amos 2012); however, chromosomal inversions are difficult to detect de novo from short-read data, and long-read data are likely necessary to either confirm or rule out the possibility of an inversion in this region.

Examining this same region in the Brazilian sample reveals a very different picture. Here, the putative selective sweep is centered around *Vgsc*; however, PCA results, as well as genotypes at individual loci associated with resistance, suggest that the haplotype of resistant alleles is absent in this population. Indeed, all the Brazilian specimens carry the wild-type susceptible alleles at four of the five insecticide resistance loci examined. At the fifth loci, F1534C, a partial duplication of the *Vgsc* gene in the assembly prevents confident genotyping; however, if we assume that this duplication is technical in origin and that reads mapping to the duplicated region should instead be assigned to *Vgsc* itself, we find that the Brazilian sample is completely homozygous for the resistant allele at this locus (table 1). This suggests that the strong selective sweep observed in this population is driven either by F1534C, by other loci that have yet to be identified, or both, but not by V410L, S723T, I915K, or V1016I, all of which are absent from our Brazilian cohort but present in the adjacent country of Colombia.

F_{ST} values between the two haplotypes (as represented by specimens in the top and bottom clusters of the PCA) hint at a possible origin for the haplotype carrying five known resistance alleles: regions of very low F_{ST} between the two haplotypes are interspersed with sharply delineated, discrete regions of very high F_{ST}, and two such regions overlap the four focal loci (those corresponding to V1016I, I915K, S723T, and V410L) that differ between the two haplotypes. While further work is needed to draw conclusions about the origin of the haplotype carrying resistance alleles at five loci, these patterns are not suggestive of recurrent de novo mutation of the focal loci on the F1534C background.

We therefore hypothesize that the Brazilian and Colombian samples may in fact have been subject to the same selective sweep at this locus, despite the absence of the signal of such a sweep in *Vgsc* itself in the Colombian sample. Our proposed explanation for this apparent discrepancy involves the putative long-range haplotypes present in the Colombian population. The presence of two such haplotypes, both at intermediate frequency, would necessarily skew the SFS toward intermediate-frequency alleles, thereby increasing Tajima's *D* in this region. In effect, we suggest that in Colombia, a second, and ongoing, selective sweep confined to only part of the *Vgsc* gene has partially eroded the signal of an earlier sweep that encompassed the entire gene and some of the flanking regions. Alternatively, it could be that this haplotype is being maintained by some form of balancing selection in the Colombian sample, or its intermediate frequency has resulted from an influx of migrant haplotypes from another population, rather than ongoing positive selection. However, the high LD between V410L, S723T, I915K, or V1016I, as well as the lack of diversity among both the resistant and wild-type backgrounds (supplementary fig. S16, Supplementary Material online), seem to contradict these alternative explanations; we also note that V410L and V1016I have previously been noted to be in LD with each other in other locations (Vera-Maloof et al. 2015; Saavedra-Rodriguez et al. 2018; Ayres et al. 2020; Baltzegar et al. 2021). If our hypothesis is correct, the frequency of the haplotype bearing these resistance mutations may continue to increase in the coming years under continuing selective pressure from insecticide use. Indeed, the Brazilian sample is from 2012, so allele frequencies at these loci may be very different in the present day, as the resistant alleles at V1016I and F1534C are common (Linss et al. 2014; David et al. 2018; Melo Costa et al. 2020; Cosme et al. 2022) and increasing in frequency (Linss et al. 2014) in some parts of Brazil. This situation highlights the need for an increased use of temporal data in vector genomics.

An apparent sweep in the vicinity of *Vgsc* has been observed previously in a sample from southern Mexico (Saavedra-Rodriguez et al. 2019); in addition, data from North America must be interpreted with caution for reasons previously discussed, but are consistent with reduced diversity in the vicinity of *Vgsc* and an increase in

insecticide resistance alleles over time. Taken all together, these data suggest this long-range haplotype may not be limited to Colombia and could be spreading more widely across the Americas.

An earlier origin of F1534C has been posited before (Saavedra-Rodriguez et al. 2018) and would be consistent with previous observations that F1534C is more commonly found in the absence of V410L (Haddi et al. 2017; Melo Costa et al. 2020) and V1016I (Linss et al. 2014; Vera-Maloof et al. 2015; Melo Costa et al. 2020) than vice versa; in addition, in Iquitos, Peru, F1534C was directly observed to reach near fixation before V1016I became prevalent (Baltzegar et al. 2021). These findings also lead us to speculate about the catalyst for this putative earlier sweep. Pyrethroids and DDT both target the voltage-gated sodium channel, and F1534C is known to confer resistance to DDT as well as to type I pyrethroids (Harris et al. 2010; Hu et al. 2011; Du et al. 2013); cross-resistance to both insecticides has been observed in *Ae. aegypti* (Chadwick et al. 1977). Populations of *Ae. aegypti* across Colombia were still highly resistant to DDT as of 2006 (Fonseca-González et al. 2011), though its use for vector control has been banned in that country since 1993 (Fonseca-González et al. 2011). Though the F1534C allele is present at high frequency in the US sample and a subset of the sample shows reduced nucleotide diversity in this region, we do not see a putative selective sweep at this locus at anywhere near the magnitude of the other American populations. Although it is possible that heterogeneity within the US sample may be somewhat masking the signature of a strong sweep, it is also possible that the selective pressure on the sweeping allele (whether it was F1534C or otherwise), in populations in the United States, may not have been strong enough to drive one haplotype to high frequency as quickly as in South America. Unlike in South and Central America, continent-wide eradication of *Ae. aegypti* was never attempted in North America. While we can only speculate that the apparent sweep of F1534C in South America may be tied to the attempted eradication with DDT, it is worth noting that DDT, not pyrethroids, are likely to have presented the strongest anthropogenic selection pressure these populations have ever experienced.

Conclusions

Here we describe a data set of 32 *Ae. aegypti* genomes collected from Colombia. We combined these data with whole genomes from five other countries from Africa and the Americas to examine the genomic impacts of population structure, potential past demographic events, and natural selection on this important vector species. While this analysis has revealed some clues about potential recent demographic events (e.g., domestication-associated bottlenecks and recent migrations), future work with broader sampling should test these possibilities explicitly. Perhaps our clearest contribution is the strong signatures of positive selection that we identify around numerous insecticide resistance genes. These results paint a compelling picture of strong selective pressures and a population that appears to be able to rapidly adapt to these pressures in part due to a fairly large number of genes that can potentially confer resistance phenotypes; these results are consistent with previous observations that a variety of loci and pathways underlie *Ae. aegypti*'s response to insecticides in general and pyrethroids in particular (Saavedra-Rodriguez et al. 2008, 2012; Paris et al. 2013; Saavedra-Rodriguez et al. 2014). We also note that because *Ae. aegypti* has a large and expanding global population in terms of its census size, adaptation may not be mutation-limited even at individual sites. For example, two different resistance mutations at the same codon in *Vgsc*, V1016G and V1016I, are prevalent in Asia and the Americas, respectively, and at least one mutation in *Vgsc*, F1534C, appears to have occurred recurrently (Cosme et al. 2020). Moreover, the ability to occasionally travel long distances through the aid of human transportation may allow for the sharing of resistance alleles across populations. Thus, we argue that continued efforts to identify selected alleles in *Ae. aegypti* populations and to monitor their spread will be vital for informing vector control efforts.

Materials and Methods

Sample Collection

We sampled two urban locations in Medellín (Comuna 16 Belen and Comuna 13 San Javier) and a rural locality near the nature reserve area of Río Claro in the department of Antioquia. Medellín has an average altitude of 1,500 m and a mean annual temperature of 22.5 °C. Río Claro's average altitude and temperature are 1,750 m and 22 °C, respectively. We also sampled from Cali, capital of Valle del Cauca department (3.3817 N, −76.55237 W). Cali's average altitude, temperature, and annual precipitation are 1,479 m, 24 °C, and ~1,050 mm, respectively. The Cañón del Río Claro Nature Reserve is located between the municipalities of Sonsón, Puerto Triunfo, and San Luis, southwest of Antioquia (5.90 N, −74.86 W [Vivero et al. 2010]).

Our collection focused on *Ae. aegypti* larval stages, and their main potential breeding grounds (pools and tanks used to store drinking water) were inspected. Larvae were then transported in Whirl-Pak plastic bags to the insectary of the Program for the Study and Control of Tropical Diseases (PECET) of the University of Antioquia, Medellín, Colombia (27 ± 2 °C; relative humidity 80 ± 10%; photoperiod: 12 light hours), where stages III and IV were individualized and their breeding continued for entomological series, reference collection, and molecular processing of adults. Larvae, pupae, and adults were sorted following the protocols suggested by the Biosystematics Unit, Division of Entomology, Walter Reed 1997. To assign to species (*Ae. aegypti*), we used external morphological characters of larvae, pupae, and adult females. All specimen collections were performed following the guidelines of the Colombian decree nNo. 1376. No specific permits were required for this study. Mosquitoes were collected

on a private property and permission was received from landowners before sampling.

DNA Extraction and Sequencing

We extracted genomic DNA from individual wild-caught adult *Aedes* in Colombia using Genetra Puregene Tissue Kits (Qiagen, Valencia, CA, United States) and constructed barcoded libraries for sequencing using KAPA HyperPrep kits (Roche Sequencing, Pleasanton, CA) with a target fragment size of 500 bp. We assessed the size of the library inserts using a Tapestation 4150 (Agilent, Santa Clara, CA) and dsDNA screentape (Agilent, #5067–5365). Finally, we pooled the samples and sequenced them in a NovaSeq6000 S2 (Illumina) machine, generating $2 \times$ 150 bp reads. The target coverage was $15\times$ per sample. This effort yielded 40 sequenced genomes. To these specimens, we added publicly available sequence data (Lee et al. 2019; Rose et al. 2020), as previously noted.

Quality Control, Alignment, and Variant Genotyping

We used FASTQC 0.11.9 (Andrews 2010) to check initial read quality and Trimmomatic 0.39 (Bolger et al. 2014) to remove adapters and trim regions of low quality at the beginnings and ends of reads. After trimming, we rechecked read quality using FASTQC.

For newly sequenced specimens, read group information was extracted using bash scripting. Reads were aligned to AaegL5, retrieved from VectorBase (Matthews et al. 2018; Giraldo-Calderón et al. 2022), using the mem tool of bwa-mem2 version 2.1 (Vasimuddin et al. 2019), converted to BAM format, and, for specimens split across multiple lanes, merged across lanes using samtools version 1.11 (Danecek et al. 2021). We used MarkDuplicatesSpark in GATK version 4.1.9.0 (DePristo et al. 2011; Poplin et al. 2018) to mark optical duplicates and sort reads. BAM files were then indexed with samtools index. The quality of the resulting BAM files was assessed with qualimap version 2.2.2d (Okonechnikov et al. 2016) bamqc and with picard CollectMultipleMetrics.

Sites were genotyped with GATK HaplotypeCaller (Poplin et al. 2018) independently for each specimen, parallelized across chromosomes, using the EMIT_ALL_CONFIDENT_SITES flag. The resulting gVCF files were merged with picard GatherVcfs and indexed with GATK IndexFeatureFile. Quality metrics for each specimen were compiled from FASTQC, picard, and qualimap using multiqc version 1.9 (Ewels et al. 2016). Genomic VCF files for each specimen were then turned into batch variant calls using GATK. Specifically, specimens were compiled using GenomicsDBImport and genotyped using GenotypeGVCFs with the include-nonvariant-sites flag, parallelized across the genome as necessary for efficiency. The resulting VCFs were merged using picard GatherVcfs and filtered to SNPs only using the view command from bcftools 1.8 and 1.9 (Danecek et al. 2021). The quality control and alignment through individual variant calling steps, and the batch genotyping steps, were implemented as two

nextflow v. 20.10.0 (Di Tommaso et al. 2017) workflows, available on GitHub (https://github.com/rrlove/aedes_pipeline).

The merged VCFs were converted into zarr format using scikit-allel 1.3.2 (DOI:10.5281/zenodo.4759368). Variants were filtered in python using scikit-allel. For our focal sample from Colombia, we removed specimens with a mean coverage below $15\times$ or fewer than 70% of reads mapping to the reference genome. In the combined sample, we also removed invariant sites and sites with indels. Based on GATK best practices for hard filtering (https://gatk.broadinstitute.org/hc/en-us/articles/360035890471-Hard-filtering-germline-short-variants), the following six hard filtering criteria were implemented, with sites required to pass five of the six filters: 1) variant quality by depth (QD) greater than or equal to 2; 2) variant strand bias as estimated by Fisher's exact test (FS) less than 40; 3) variant strand bias as estimated by the symmetric odds ratio test (SOR) less than 4; 4) mapping quality (MQ) greater than or equal to 40; 5) mapping quality rank sum test (MQRS) between −5 and 5, inclusive; and 6) site position within reads rank sum test (RPRS) between −3 and 3, inclusive.

In addition, we implemented additional filters to retain only those sites that were called in at least 75% of specimens, in at least five of the six countries, and, further, to remove multiallelic sites. Where noted, we also masked genotypes with genotype qualities below 20.

To identify and remove closely related specimens from our sample, we calculated relatedness using KING-robust (Manichaikul et al. 2010) with the kinship flag for all three chromosomes. To remove any possible interference of the sex-defining *m* locus and surrounding region (Fontaine et al. 2017; Matthews et al. 2018) with inferred kinship coefficients, we repeated this analysis using chromosomes 2 and 3, only. We initially used kinship coefficient cutoffs of 0.20 in American specimens and 0.05 in African specimens, following Rose et al. (2020), to account for higher levels of inbreeding found in American specimens; however, we observed that while only one pair of specimens in the American sample had a kinship coefficient greater than 0.20, four additional pairs had a kinship coefficient between 0.1896 and 0.20. To be conservative, we removed one specimen from all five pairs. For all pairs, we removed the specimen with lower coverage.

Finally, for some analyses, we removed regions of the genome previously identified as repetitive (Matthews et al. 2018), as well as regions that were not uniquely mappable (i.e., a read-sized segment of the genome that matches equally well to its origin and to other areas of the genome, allowing for a certain number of mismatches). We computed mappability with GenMap (Pockrandt et al. 2020), specifying *k*mers of 150 bp and allowing up to two errors (or mismatches). Any site that was part of a *k*mer that was not uniquely mappable, or any site denoted as repetitive, was noted for exclusion from the appropriate analyses as noted below.

Patterns of Genetic Diversity within and between Populations

We visualized the genetic structure in our sample using principal component analysis on each chromosome and on the three chromosomes together. SNPs were excluded if their minor allele frequency was 5% or less in the total sample. SNPs were also excluded if they overlapped repetitive regions or nonuniquely mapping regions, and genotypes with genotype qualities below 20 were masked. Genotypes were converted to the number of alternate genotypes present (0 for homozygous reference, 1 for heterozygous, and 2 for homozygous alternate), and principal component analysis was done using scikit-allel. Plots were visualized using the first two principal components, and the percentage of variance explained by each of the first ten principal components was plotted.

We also examined patterns of genetic diversity and divergence across the genome. Nucleotide diversity (π) was calculated in nonoverlapping 500 kb windows for each of the six countries using the windowed_diversity function in scikit-allel. Tajima's D was calculated similarly, using the windowed_tajima_d function. For nucleotide diversity, we removed repetitive and nonuniquely mappable regions, and for both statistics, we masked genotypes with qualities less than 20. For visualization, both statistics were also calculated in sliding 5 Mb windows with a step size of 500 kb. Genome-wide F_{ST} between each pair of countries was calculated using scikit-allel's moving_weir_cockerham_fst function, using nonoverlapping windows of 500,000 variants, after removing repetitive and nonuniquely mapping regions.

For computational tractability, LD was calculated in parallel in 50 Mb genomic regions (or one smaller segment at the end of each chromosome). In each region, linkage disequilibrium was calculated within each cohort in 100 kb nonoverlapping windows. Linkage disequilibrium was summarized by r^2 as calculated using scikit-allel's windowed_r_squared function, modified to return nan for windows where only one variant is present. Scikit-allel's windowed_r_squared function uses the method of Rogers and Huff (2009), which does not require phasing. Before calculations, genotypes with a genotype quality below 20 were masked; then, nonsegregating variants were removed (these might occur because a position is invariant or is masked due to low genotype quality, in the subset of specimens selected for analysis).

Scanning for Selective Sweeps with SweepFinder2

Putative selective sweeps were identified in each cohort using SweepFinder2 (DeGiorgio et al. 2016), using the empirical SFS generated from all SNPs segregating in the focal cohort as the background SFS, as well as a genetic map (Matthews et al. 2018). Before calculating the site frequency spectra, we removed repetitive and nonuniquely mappable regions and masked genotypes with qualities less than 20. For computational tractability, SweepFinder2 was run in chunks of 50 Mb (or, at the end of each chromosome, one segment containing the remainder), with a grid size of 1,000, representing a test site approximately every kilobase. On each chromosome, the 50 Mb chunks overlapped by 5 Mb, and results were plotted to visually check for edge effects before merging. Finally, results for each test site were averaged within 100 and 10 kb nonoverlapping windows, for visualization at different scales.

Enrichment Analysis

We used functional enrichment analysis with Gowinda version 1.12 (Kofler and Schlötterer 2012) to look for overrepresented GO terms annotated to genes overlapping putative selective sweeps identified by Sweep Finder2. First, the loci (here a "locus" refers to an individual test site, spaced approximately every 1 kb) with sweep likelihoods in the top 1% of all loci were identified. These loci were used as the candidate site file for Gowinda, with all loci queried by SweepFinder, rounded to the nearest base pair, used as the total site file. We included 1,000 bases up- and downstream from each gene to account for possible sweeps in regulatory regions. For each cutoff, we ran 1,000,000 simulations, using the "gene" mode, with the following parameters: —min-significance 1—min-genes 1—gene-definition gene—gene-definition updownstream 1,000. The gene set file was obtained by collapsing entries in the *Ae. aegypti* LVP_AGWG AaegL5.3 gene annotation file, available through VectorBase, by GO term.

Many of the genes in the gene sets with the lowest q values were found in clusters, suggesting the potential for physical proximity and linkage disequilibrium to influence these results. Therefore, we implemented an additional enrichment test (https://github.com/SchriderLab/permEnrichmentTest) based on the approach of permuting the association between scores and tested loci. Specifically, this relationship was permuted by joining the three chromosomes into an artificial circular genome, then "rotating" all scores around the circle by a random interval. By preserving relationships between adjacent loci, this approach offers the ability to control for linkage. As before, we ran this enrichment analysis using the top 1% of all tested sites. For each cutoff, we ran 10,000 permutations of the SweepFinder2 scores before asking which GO terms were enriched in the sweep set relative to the permuted set. As in our analysis with Gowinda, we extended genes by 1,000 bp in either direction before determining which genes contained a putative sweep.

Scanning for Selective Sweeps with Garud's *H* Statistics

In addition to the site frequency spectrum-based sweep detection that we performed using SweepFinder2, we also used a haplotype-based method to detect sweeps in the vicinity of two of the most prominent peaks identified on the basis of their CLR scores. Garud's *H* statistics (Garud

et al. 2015) can also be used on unphased data with slight modifications (Harris et al. 2018). Therefore, we constructed multilocus genotypes for each specimen by summing the number of alternate alleles per specimen, per site. We then converted these to scikit-allel's HaplotypeArray format and used these arrays as input to scikit-allel's moving_garud_h function. G1, G12, G123, and G2/G1 were calculated in windows of 500 variants slid by 50 variants. The corresponding variant positions were also averaged in windows of 500 variants slid by 50 variants, to calculate the average position of each window.

Analysis of the *Vgsc* Region

Given its relevance to insecticide resistance, we examined the *Vgsc* gene and its surroundings. In addition to the genome-wide approach already detailed, we used bcftools mpileup and call, with the multiallelic variant calling model, to genotype variants present between 315 and 317Mb of AaegL5_3, and on scaffold NIGP01000811. To visualize the relationship between scaffold NIGP01000811 and chromosome 3, we aligned the two using mummer 4.0.0rc1 (Marçais et al. 2018) and visualized the alignment using Ribbon (Nattestad et al. 2021).

We performed principal component analysis as previously described on all SNPs within the bounds of *Vgsc* (AAEL023266; AaegL5_3:315,926,360–316,405,639). In addition, we visualized the "loadings" of the first two principal components, plotted against each SNP's chromosomal position. We also calculated nucleotide diversity (π) and Tajima's D in the region spanning 310–320 Mb on chromosome 3, in 500 kb windows slid by 50 kb, by country, as well as in the top and bottom clusters of the *Vgsc*-specific PCA (for South American specimens only). Finally, we divided the California portion of the US sample into three clusters as per Lee et al. (2019) (cluster 1: Menlo Park, Madera, Fresno; cluster 2: Clovis, Sanger; and cluster 3: Garden Grove, Mission Viejo, Brawley, San Diego, Exeter) and calculated π and Tajima's D within each cluster, visualizing the values in conjunction with those from Brazil and Colombia.

We also calculated F_{ST} between the haplotypes carried in the South American specimens in the top and bottom clusters of the *Vgsc* PCA, respectively, using the windowed_weir_cockerham_fst function in scikit-allel. F_{ST} was calculated in windows of 50 kb slid by 5 kb and in windows of 10 kb slid by 1 kb.

In addition, we calculated LD between individual variants, again using Rogers and Huff's (2009) estimate of r^2, within *Vgsc* and 100 kb up- and downstream after masking low-quality genotypes as described above. We calculated LD within all South American specimens, all North American specimens, and then all American specimens combined. We also calculated LD between the five focal loci; in this case, for reasons detailed previously, genotypes with qualities below 20 were masked at all loci except for 315,939,224, the position corresponding to the F1534C mutation.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Data Availability

The nextflow pipelines used to align reads, call variants, and genotype variants can be found at https://github.com/rrlove/aedes_pipeline. The code used to generate figures can be found at https://github.com/rrlove/aegypti_selection_1. Newly generated sequences from Colombia can be found at BioProject PRJNA864744.

References

Aberg KA, Shabalin AA, Chan RF, Zhao M, Kumar G, van Grootheest G, Clark SL, Xie LY, Milaneschi Y, Penninx BWJH, et al. 2018. Convergence of evidence from a methylome-wide CpG-SNP association study and GWAS of major depressive disorder. *Transl Psychiatry*. **8**(1):1–10.

Aguirre-Obando OA, Bona ACD, Duque LJE, Navarro-Silva MA. 2015. Insecticide resistance and genetic variability in natural populations of Aedes (Stegomyia) aegypti (Diptera: Culicidae) from Colombia. *Zool Curitiba*. **32**(1):14–22.

Al Nazawi AM, Aqili J, Alzahrani M, McCall PJ, Weetman D. 2017. Combined target site (kdr) mutations play a primary role in highly pyrethroid resistant phenotypes of Aedes aegypti from Saudi Arabia. *Parasit Vectors*. **10**(1):1–10.

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data [online]. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Aubry F, Dabo S, Manet C, Filipović I, Rose NH, Miot EF, Martynow D, Baidaliuk A, Merkling SH, Dickson LB, et al. 2020. Enhanced Zika virus susceptibility of globally invasive Aedes aegypti populations. *Science*. **370**(6519):991–996.

Ayres CFJ, Seixas G, Borrego S, Marques C, Monteiro I, Marques CS, Gouveia B, Leal S, Troco AD, Fortes F, et al. 2020. The V410L knockdown resistance mutation occurs in island and continental populations of Aedes aegypti in West and Central Africa. *PLoS Negl Trop Dis*. **14**(5):e0008216.

Baltzegar J, Vella M, Gunning C, Vasquez G, Astete H, Stell F, Fisher M, Scott TW, Lenhart A, Lloyd AL, et al. 2021. Rapid evolution of knockdown resistance haplotypes in response to pyrethroid selection in Aedes aegypti. *Evol Appl*. **14**(8):2098–2113.

Barón OL, Ursic-Bedoya RJ, Lowenberger CA, Ocampo CB. 2010. Differential gene expression from midguts of refractory and susceptible lines of the mosquito, Aedes aegypti, infected with Dengue-2 virus. *J Insect Sci*. **10**(1):41.

Bennett V, Baines AJ. 2001. Spectrin and ankyrin-based pathways: metazoan inventions for integrating cells into tissues. *Physiol Rev*. **81**(3):1353–1392.

Bennett V, Chen L. 2001. Ankyrins and cellular targeting of diverse membrane proteins to physiological sites. *Curr Opin Cell Biol*. **13**(1):61–67.

Bennett KL, Shija F, Linton Y-M, Misinzo G, Kaddumukasa M, Djouaka R, Anyaele O, Harris A, Irish S, Hlaing T, et al. 2016.

Historical environmental change in Africa drives divergence and admixture of *Aedes aegypti* mosquitoes: a precursor to successful worldwide colonization? *Mol Ecol*. **25**(17):4337–4354. doi:10.1111/mec.13762

Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, Drake JM, Brownstein JS, Hoen AG, Sankoh O, et al. 2013. The global distribution and burden of dengue. *Nature*. **496**(7446): 504–507.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. **30**(15): 2114–2120.

Bonizzoni M, Dunn WA, Campbell CL, Olson KE, Dimon MT, Marinotti O, James AA. 2011. RNA-seq analyses of blood-induced changes in gene expression in the mosquito vector species, *Aedes aegypti*. *BMC Genomics*. **12**(1):1–13.

Bonizzoni M, Dunn WA, Campbell CL, Olson KE, Marinotti O, James AA. 2012. Complex modulation of the *Aedes aegypti* transcriptome in response to dengue virus infection. *PLoS One*. **7**(11): e50512.

Brengues C, Hawkes NJ, Chandre F, McCarroll L, Duchon S, Guillet P, Manguin S, Morgan JC, Hemingway J. 2003. 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.

Brennan SA, Grob IC, Bartz CE, Baker JK, Jiang Y. 2021. Displacement of *Aedes albopictus* by *Aedes aegypti* in Gainesville, Florida. *J Am Mosq Control Assoc*. **37**(2):93–97.

Brown JE, Evans BR, Zheng W, Obas V, Barrera-Martinez L, Egizi A, Zhao H, Caccone A, Powell JR. 2014. Human impacts have shaped historical and recent evolution in *Aedes aegypti*, the dengue and yellow fever mosquito. *Evolution*. **68**(2):514–525.

Brown JE, McBride CS, Johnson P, Ritchie S, Paupy C, Bossin H, Lutomiah J, Fernandez-Salas I, Ponlawat A, Cornel AJ, et al. 2011. Worldwide patterns of genetic differentiation imply multiple 'domestications' of *Aedes aegypti*, a major vector of human diseases. *Proc R Soc B Biol Sci*. **278**(1717):2446–2454.

Campbell CL, Saavedra-Rodriguez K, Kubik TD, Lenhart A, Lozano-Fuentes S, Iv WCB. 2019. Vgsc-interacting proteins are genetically associated with pyrethroid resistance in *Aedes aegypti*. *PLoS One*. **14**(1):e0211497.

Casida JE. 2018. Neonicotinoids and other insect nicotinic receptor competitive modulators: progress and prospects. *Annu Rev Entomol*. **63**(1):125–144.

Cattel J, Haberkorn C, Laporte F, Gaude T, Cumer T, Renaud J, Sutherland IW, Hertz JC, Bonneville J, Arnaud V, et al. 2021a. 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*. **14**(4):1009–1022.

Cattel J, Minier M, Habchi-Hanriot N, Pol M, Faucon F, Gaude T, Gaborit P, Issaly J, Ferrero E, Chandre F, et al. 2021b. Impact of selection regime and introgression on deltamethrin resistance in the arbovirus vector *Aedes aegypti*—a comparative study between contrasted situations in New Caledonia and French Guiana. *Pest Manag Sci*. **77**(12):5589–5598.

Chadwick PR, Invest JF, Bowron MJ. 1977. An example of cross-resistance to pyrethroids in DDT-resistant *Aedes aegypti*. *Pestic Sci*. **8**(6):618–624.

Chauhan C, Behura SK, de Bruyn B, Lovin DD, Harker BW, Gomez-Machorro C, Mori A, Romero-Severson J, Severson DW. 2012. Comparative expression profiles of midgut genes in dengue virus refractory and susceptible *Aedes aegypti* across critical period for virus infection. *PLoS One*. **7**(10):e47350.

Cheng G, Cox J, Wang P, Krishnan MN, Dai J, Qian F, Anderson JF, Fikrig E. 2010. A C-type lectin collaborates with a CD45 phosphatase homolog to facilitate West Nile virus infection of mosquitoes. *Cell*. **142**(5):714–725.

Christophers SR. 1960. *Aedes aegypti: the yellow fever mosquito*: Cambridge [England]: University Press.

Clarkson CS, Miles A, Harding NJ, O'Reilly AO, Weetman D, Kwiatkowski D, Donnelly MJ; Consortium TA gambia 1000 G. 2021. The genetic architecture of target-site resistance to pyrethroid insecticides in the African malaria vectors *Anopheles gambiae* and *Anopheles coluzzii*. *Mol Ecol*. **30**(21):5303–5317.

Codeço CT, Lourenço-De-Oliveira R, Maciel-De-Freitas R. 2007. Daily survival rates and dispersal of *Aedes aegypti* females in Rio de Janeiro, Brazil. *Am J Trop Med Hyg*. **76**(4):659–665.

Cosme LV, Gloria-Soria A, Caccone A, Powell JR, Martins AJ. 2020. Evolution of kdr haplotypes in worldwide populations of *Aedes aegypti*: independent origins of the F1534C kdr mutation. *PLoS Negl Trop Dis*. **14**(4):e0008219.

Cosme LV, Lima JBP, Powell JR, Martins AJ. 2022. Genome-wide association study reveals new loci associated with pyrethroid resistance in *Aedes aegypti*. *Front Genet*. **13**. https://www.frontiersin.org/article/10.3389/fgene.2022.867231.

Crawford JE, Alves JM, Palmer WJ, Day JP, Sylla M, Ramasamy R, Surendran SN, Black WC, Pain A, Jiggins FM. 2017. 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**(1):16.

Cui Y, Liu P, Mooney BP, Franz AWE. 2020. Quantitative proteomic analysis of chikungunya virus-infected *Aedes aegypti* reveals proteome modulations indicative of persistent infection. *J Proteome Res*. **19**(6):2443–2456.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. *GigaScience*. **10**(2): giab008.

David MR, Garcia GA, Valle D, Maciel-de-Freitas R. 2018. Insecticide resistance and fitness: the case of four *Aedes aegypti* populations from different Brazilian regions. *BioMed Res Int*. **2018**:6257860.

Davies TGE, Field LM, Usherwood PNR, Williamson MS. 2007. DDT, pyrethrins, pyrethroids and insect sodium channels. *IUBMB Life*. **59**(3):151–162.

DeGiorgio M, Huber CD, Hubisz MJ, Hellmann I, Nielsen R. 2016. Sweepfinder2: increased sensitivity, robustness and flexibility. *Bioinformatics*. **32**(12):1895–1897.

DePristo M, Banks E, Poplin RE, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Publ Group*. **43**(5): 491–498.

Després L, Stalinski R, Tetreau G, Paris M, Bonin A, Navratil V, Reynaud S, David J-P. 2014. Gene expression patterns and sequence polymorphisms associated with mosquito resistance to Bacillus thuringiensis israelensis toxins. *BMC Genomics*. **15**(1):1–13.

Dickson LB, Campbell CL, Juneja P, Jiggins FM, Sylla M, Black WC IV. 2017. Exon-enriched libraries reveal large genic differences between *Aedes aegypti* from Senegal, West Africa, and populations outside Africa. *G3 GenesGenomesGenetics*. **7**(2):571–582.

Dickson LB, Sanchez-Vargas I, Sylla M, Fleming K, Iv WCB. 2014. Vector competence in West African *Aedes aegypti* is flavivirus species and genotype dependent. *PLoS Negl Trop Dis*. **8**(10):e3153.

Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. 2017. Nextflow enables reproducible computational workflows. *Nat Biotechnol*. **35**(4):316–319.

Dong K, Scott JG. 1994. Linkage of kdr-type resistance and the para-homologous sodium channel gene in German cockroaches (*Blattella germanica*). *Insect Biochem Mol Biol*. **24**(7):647–654.

Du Y, Nomura Y, Satar G, Hu Z, Nauen R, He SY, Zhorov BS, Dong K. 2013. Molecular evidence for dual pyrethroid-receptor sites on a mosquito sodium channel. *Proc Natl Acad Sci*. **110**(29): 11785–11790.

Dubey SK, Mehta D, Chaudhary S, Hasan A, Sunil S. 2022. An E3 ubiquitin ligase scaffolding protein is proviral during chikungunya virus infection in *Aedes aegypti*. Zhang B, editor. *Microbiol Spectr*. **10**(3):e00595-22.

Durant AC, Grieco Guardian E, Kolosov D, Donini A. 2021. The transcriptome of anal papillae of *Aedes aegypti* reveals their importance in xenobiotic detoxification and adds significant knowledge on ion, water and ammonia transport mechanisms. *J Insect Physiol*. **132**:104269.

El-Garj F, Wajidi MFF, Avicor SW. 2016. Allelic variants of cytochrome P450 monooxygenases: constitutive and insecticide-mediated expression in a Malaysian strain of the dengue vector, *Aedes aegypti* (Diptera: Culicidae). *Eur J Entomol*. **113**:507–515.

Enayati AA, Ranson H, Hemingway J. 2005. Insect glutathione transferases and insecticide resistance. *Insect Mol Biol*. **14**(1):3–8.

Evans BR, Gloria-Soria A, Hou L, McBride C, Bonizzoni M, Zhao H, Powell JR. 2015. A multipurpose, high-throughput single-nucleotide polymorphism chip for the dengue and yellow fever mosquito, *Aedes aegypti*. *G3 GenesGenomesGenetics*. **5**(5):711–718.

Ewels P, Magnusson M, Lundin S, Käller M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. **32**(19):3047–3048.

Fan Y, O'Grady P, Yoshimizu M, Ponlawat A, Kaufman PE, Scott JG. 2020. Evidence for both sequential mutations and recombination in the evolution of kdr alleles in *Aedes aegypti*. *PLoS Negl Trop Dis*. **14**(4):e0008154.

Farnham AW. 1977. Genetics of resistance of houseflies (*Musca domestica* L.) to pyrethroids. I. Knock-down resistance. *Pestic Sci*. **8**(6):631–636.

Faucon F, Dusfour I, Gaude T, Navratil V, Boyer F, Chandre F, Sirisopa P, Thanispong K, Juntarajumnong W, Poupardin R, et al. 2015. Identifying genomic changes associated with insecticide resistance in the dengue mosquito *Aedes aegypti* by deep targeted sequencing. *Genome Res*. **25**(9):1347–1359.

Faucon F, Gaude T, Dusfour I, Navratil V, Corbel V, Juntarajumnong W, Girod R, Poupardin R, Boyer F, Reynaud S, et al. 2017. In the hunt for genomic markers of metabolic resistance to pyrethroids in the mosquito *Aedes aegypti*: an integrated next-generation sequencing approach. *PLoS Negl Trop Dis*. **11**(4):e0005526.

Fonseca-González I, Quiñones ML, Lenhart A, Brogdon WG. 2011. Insecticide resistance status of *Aedes aegypti* (L.) from Colombia. *Pest Manag Sci*. **67**(4):430–437.

Fontaine A, Filipović I, Fansiri T, Hoffmann AA, Cheng C, Kirkpatrick M, Rašić G, Lambrechts L. 2017. Extensive genetic differentiation between homomorphic sex chromosomes in the mosquito vector, *Aedes aegypti*. *Genome Biol Evol*. **9**(9):2322–2335.

Garud NR, Messer PW, Buzbas EO, Petrov DA. 2015. Recent selective sweeps in North American *Drosophila melanogaster* show signatures of soft sweeps. *PLoS Genet*. **11**(2):e1005004.

Garud NR, Messer PW, Petrov DA. 2021. Detection of hard and soft selective sweeps from *Drosophila melanogaster* population genomic data. *PLoS Genet*. **17**(2):e1009373.

Giraldo-Calderón GI, Harb OS, Kelly SA, Rund SS, Roos DS, McDowell MA. 2022. Vectorbase.org updates: bioinformatic resources for invertebrate vectors of human pathogens and related organisms. *Curr Opin Insect Sci*. **50**:100860.

Gloria-Soria A, Ayala D, Bheecarry A, Calderon-Arguedas O, Chadee DD, Chiappero M, Coetzee M, Elahee KB, Fernandez-Salas I, Kamal HA, et al. 2016a. Global genetic diversity of *Aedes aegypti*. *Mol Ecol*. **25**(21):5377–5395.

Gloria-Soria A, Kellner DA, Brown JE, Gonzalez-Acosta C, Kamgang B, Lutwama J, Powell JR. 2016b. Temporal genetic stability of *Stegomyia aegypti* (= *Aedes aegypti*) populations. *Med Vet Entomol*. **30**(2):235–240.

Gold AS, Feitosa-Suntheimer F, Araujo RV, Hekman RM, Asad S, Londono-Renteria B, Emili A, Colpitts TM. 2020. Dengue virus infection of *Aedes aegypti* alters extracellular vesicle protein cargo to enhance virus transmission. *Int J Mol Sci*. **21**(18):6609.

Goncalves da Silva A, Cunha ICL, Santos WS, Luz SLB, Ribolla PEM, Abad-Franch F. 2012. Gene flow networks among American *Aedes aegypti* populations. *Evol Appl*. **5**(7):664–676.

Haddi K, Tomé HVV, Du Y, Valbon WR, Nomura Y, Martins GF, Dong K, Oliveira EE. 2017. Detection of a new pyrethroid resistance mutation (V410L) in the sodium channel of *Aedes aegypti*: a potential challenge for mosquito control. *Sci Rep*. **7**(1):46549.

Harrington LC, Scott TW, Lerdthusnee K, Coleman RC, Costero A, Clark GG, Jones JJ, Kitthawee S, Kittayapong P, Sithiprasasna R, et al. 2005. Dispersal of the dengue vector *Aedes aegypti* within and between rural communities. *Am J Trop Med Hyg*. **72**(2):209–220.

Harris AM, Garud NR, DeGiorgio M. 2018. Detection and classification of hard and soft sweeps from unphased genotypes by multilocus genotype identity. *Genetics*. **210**(4):1429–1452.

Harris AF, Rajatileka S, Ranson H. 2010. Pyrethroid resistance in *Aedes aegypti* from Grand Cayman. *Am J Trop Med Hyg*. **83**(2):277–284.

Hemingway J, Boddington RG, Harris J, Dunbar SJ. 1989. Mechanisms of insecticide resistance in *Aedes aegypti* (L.) (Diptera: Culicidae) from Puerto Rico. *Bull Entomol Res*. **79**(1):123–130.

Hirata K, Komagata O, Itokawa K, Yamamoto A, Tomita T, Kasai S. 2014. 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**(8):e3085.

Hu Z, Du Y, Nomura Y, Dong K. 2011. 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**(1):9–13.

Itokawa K, Sekizuka T, Maekawa Y, Yatsu K, Komagata O, Sugiura M, Sasaki T, Tomita T, Kuroda M, Sawabe K, et al. 2019. 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**(11):e0007818.

Jansen CC, Beebe NW. 2010. The dengue vector *Aedes aegypti*: what comes next. *Microbes Infect*. **12**(4):272–279.

Jenkins SM, Bennett V. 2001. Ankyrin-G coordinates assembly of the spectrin-based membrane skeleton, voltage-gated sodium channels, and L1 CAMs at Purkinje neuron initial segments. *J Cell Biol*. **155**(5):739–746.

Kaplan NL, Hudson RR, Langley CH. 1989. The "hitchhiking effect" revisited. *Genetics*. **123**(4):887–899.

Kawada H, Higa Y, Komagata O, Kasai S, Tomita T, Yen NT, Loan LL, Sánchez RAP, Takagi M. 2009. Widespread distribution of a newly found point mutation in voltage-gated sodium channel in pyrethroid-resistant *Aedes aegypti* populations in Vietnam. *PLoS Negl Trop Dis*. **3**(10):e527.

Kawada H, Oo SZM, Thaung S, Kawashima E, Maung YNM, Thu HM, Thant KZ, Minakawa N. 2014. Co-occurrence of point mutations in the voltage-gated sodium channel of pyrethroid-resistant *Aedes aegypti* populations in Myanmar. *PLoS Negl Trop Dis*. **8**(7):e3032.

Kelly ET, Mack LK, Campos M, Grippin C, Chen T-Y, Romero-Weaver AL, Kosinski KJ, Brisco KK, Collier TC, Buckner EA, et al. 2021. Evidence of local extinction and reintroduction of *Aedes aegypti* in Exeter, California. *Front. Trop. Dis*. **2**:703873.

Kern AD, Schrider DR. 2018. Diplos/HIC: an updated approach to classifying selective sweeps. *G3 GenesGenomesGenetics*. **8**(6):1959–1970.

Kim Y, Nielsen R. 2004. Linkage disequilibrium as a signature of selective sweeps. *Genetics*. **167**(3):1513–1524.

Kofler R, Schlötterer C. 2012. Gowinda: unbiased analysis of gene set enrichment for genome-wide association studies. *Bioinformatics*. **28**(15):2084–2085.

Kolaczkowski B, Kern AD, Holloway AK, Begun DJ. 2011. Genomic differentiation between temperate and tropical Australian populations of *Drosophila melanogaster*. *Genetics*. **187**(1):245–260.

Kotsakiozi P, Evans BR, Gloria-Soria A, Kamgang B, Mayanja M, Lutwama J, Goff GL, Ayala D, Paupy C, Badolo A, et al. 2018. Population structure of a vector of human diseases: *Aedes aegypti* in its ancestral range, Africa. *Ecol Evol*. **8**(16):7835–7848.

Kumar V, Garg S, Gupta L, Gupta K, Diagne CT, Missé D, Pompon J, Kumar S, Saxena V. 2021. Delineating the role of *Aedes aegypti* ABC transporter gene family during mosquito development and arboviral infection via transcriptome analyses. *Pathogens*. **10**(9):1127.

Lan J, Wang Z, Chen Z, Zhang L, Zhao J, Guan Q, Liao C, Liu N, Han Q. 2021. Identification of the *Aedes aegypti* nAChR gene family and molecular target of spinosad. *Pest Manag Sci*. **77**(4):1633–1641.

Lee Y, Schmidt H, Collier TC, Conner WR, Hanemaaijer MJ, Slatkin M, Marshall JM, Chiu JC, Smartt CT, Lanzaro GC, *et al.* 2019. Genome-wide divergence among invasive populations of *Aedes aegypti* in California. *BMC Genomics*. **20**(1):1–10.

Li X, Hu S, Yin H, Zhang H, Zhou D, Sun Y, Ma L, Shen B, Zhu C. 2021. MiR-4448 is involved in deltamethrin resistance by targeting CYP4H31 in Culex pipiens pallens. *Parasit Vectors*. **14**(1):1–13.

Li H, Ruan J, Durbin RM. 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res*. **18**(11):1851–1858.

Lien NTK, Ngoc NTH, Lan NN, Hien NT, Tung NV, Ngan NTT, Hoang NH, Binh NTH. 2019. Transcriptome sequencing and analysis of changes associated with insecticide resistance in the dengue mosquito (*Aedes aegypti*) in Vietnam. *Am J Trop Med Hyg*. **100**(5):1240–1248.

Lin W-H, Wright DE, Muraro NI, Baines RA. 2009. Alternative splicing in the voltage-gated sodium channel dmnav regulates activation, inactivation, and persistent current. *J Neurophysiol*. **102**(3):1994–2006.

Linss JGB, Brito LP, Garcia GA, Araki AS, Bruno RV, Lima JBP, Valle D, Martins AJ. 2014. Distribution and dissemination of the Val1016Ile and Phe1534Cys Kdr mutations in *Aedes aegypti* Brazilian natural populations. *Parasit Vectors*. **7**(1):1–11.

Love RR, Redmond SN, Pombi M, Caputo B, Petrarca V, della Torre A; Consortium TA gambiae 1000 G, Besansky NJ. 2019. In silico karyotyping of chromosomally polymorphic malaria mosquitoes in the *Anopheles gambiae* complex. *G3 Genes Genomes Genet*. **9**(10):3249–3262.

Ma J, Amos CI. 2012. Investigation of inversion polymorphisms in the human genome using principal components analysis. *PLoS One*. **7**(7):e40224.

Mack LK, Kelly ET, Lee Y, Brisco KK, Shen KV, Zahid A, van Schoor T, Cornel AJ, Attardo GM. 2021. Frequency of sodium channel genotypes and association with pyrethrum knockdown time in populations of Californian *Aedes aegypti*. *Parasit Vectors*. **14**(1):141.

Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M. 2010. Robust relationship inference in genome-wide association studies. *Bioinformatics*. **26**(22):2867–2873.

Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. 2018. MUMmer4: a fast and versatile genome alignment system. *PLoS Comput Biol*. **14**(1):e1005944.

Martinez-Torres D, Chandre F, Williamson MS, Darriet F, Berge JB, Devonshire AL, Guillet P, Pasteur N, Pauron D. 1998. Molecular characterization of pyrethroid knockdown resistance (kdr) in the major malaria vector *Anopheles gambiae* s.s. *Insect Mol Biol*. **7**(2):179–184.

Martins M, Ramos LFC, Murillo JR, Torres A, de Carvalho SS, Domont GB, de Oliveira DMP, Mesquita RD, Nogueira FCS, Maciel-de-Freitas R, *et al.* 2021. Comprehensive quantitative proteome analysis of *Aedes aegypti* identifies proteins and pathways involved in *Wolbachia pipientis* and Zika virus interference phenomenon. *Front Physiol*. **12**.

Matthews BJ, Dudchenko O, Kingan SB, Koren S, Antoshechkin I, Crawford JE, Glassford WJ, Herre M, Redmond SN, Rose NH, *et al.* 2018. Improved reference genome of *Aedes aegypti* informs arbovirus vector control. *Nature*. **563**(7732):501–507.

Mattingly PF. 1957. Genetical aspects of the Aëdes aegypti problem. *Ann Trop Med Parasitol*. **51**(4):392–408.

Maynard Smith J, Haigh J. 1974. The hitch-hiking effect of a favourable gene. *Genet Res*. **23**(1):23–35.

McBride CS, Baier F, Omondi AB, Spitzer SA, Lutomiah J, Sang R, Ignell R, Vosshall LB. 2014. Evolution of mosquito preference for humans linked to an odorant receptor. *Nature*. **515**(7526):222–227.

Melo Costa M, Campos KB, Brito LP, Roux E, Melo Rodovalho C, Bellinato DF, Lima JBP, Martins AJ. 2020. Kdr genotyping in *Aedes aegypti* from Brazil on a nation-wide scale from 2017 to 2018. *Sci Rep*. **10**(1):13267.

Miles A, Harding NJ, Botta G, Clarkson CS, Antao T, Kozak K, Schrider DR, Kern AD, Redmond SN, Sharakhov I, *et al.* 2017. Genetic diversity of the African malaria vector *Anopheles gambiae*. *Nature*. **552**(7683):96–100.

Monteiro FA, Shama R, Martins AJ, Gloria-Soria A, Brown JE, Powell JR. 2014. Genetic diversity of Brazilian *Aedes aegypti*: patterns following an Eradication Program. *PLoS Negl Trop Dis*. **8**(9):e3167.

Nattestad M, Aboukhalil R, Chin C-S, Schatz MC. 2021. Ribbon: intuitive visualization for complex genomic variation. *Bioinformatics*. **37**(3):413–415.

Okonechnikov K, Conesa A, García-Alcalde F. 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*. **32**(2):292–294.

Paris M, Marcombe S, Coissac E, Corbel V, David J-P, Després L. 2013. Investigating the genetics of Bti resistance using mRNA tag sequencing: application on laboratory strains and natural populations of the dengue vector *Aedes aegypti*. *Evol Appl*. **6**(7):1012–1027.

Pélissié B, Chen YH, Cohen ZP, Crossley MS, Hawthorne DJ, Izzo V, Schoville SD. 2022. Genome resequencing reveals rapid, repeated evolution in the Colorado potato beetle. *Mol Biol Evol*. **39**(2):msac016.

Pittendrigh B, Reenan R, Ffrench-Constant RH, Ganetzky B. 1997. Point mutations in the Drosophila sodium channel gene para associated with resistance to DDT and pyrethroid insecticides. *Mol Gen Genet MGG*. **256**(6):602–610.

Pockrandt C, Alzamel M, Iliopoulos CS, Reinert K. 2020. Genmap: ultra-fast computation of genome mappability. *Bioinformatics*. **36**(12):3687–3692.

Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, der Auwera GAV, Kling DE, Gauthier LD, Levy-Moonshine A, Roazen D, *et al.* 2018. Scaling accurate genetic variant discovery to tens of thousands of samples. 201178. doi:10.1101/201178. [accessed 2022 Apr 6]. http://www.biorxiv.org/content/10.1101/201178v3.

Powell JR, Gloria-Soria A, Kotsakiozi P. 2018. Recent history of *Aedes aegypti*: vector genomics and epidemiology records. *BioScience*. **68**(11):854–860.

Rašić G, Filipović I, Callahan AG, Stanford D, Chan A, Lam-Phua SG, Tan CH, Hoffmann AA. 2016. The queenslandensis and the type form of the dengue fever mosquito (*Aedes aegypti* L.) are genomically indistinguishable. *PLoS Negl Trop Dis*. **10**(11):e0005096.

Rašić G, Filipović I, Weeks AR, Hoffmann AA. 2014. Genome-wide SNPs lead to strong signals of geographic structure and relatedness patterns in the major arbovirus vector, *Aedes aegypti*. *BMC Genomics*. **15**(1):275.

Rathor HR, Wood RJ. 1981. In-vivo and in-vitro studies on DDT uptake and metabolism in susceptible and resistant strains of the mosquito *Aedes aegypti* L. *Pestic Sci*. **12**(3):255–264.

Reiter P, Amador MA, Anderson RA, Clark GG. 1995. Short report: dispersal of *Aedes aegypti* in an urban area after blood feeding as demonstrated by rubidium-marked eggs. *Am J Trop Med Hyg*. **52**(2):177–179.

Riaz MA, Chandor-Proust A, Dauphin-Villemant C, Poupardin R, Jones CM, Strode C, Régent-Kloeckner M, David J-P, Reynaud S. 2013. Molecular mechanisms associated with increased tolerance to the neonicotinoid insecticide imidacloprid in the dengue vector *Aedes aegypti*. *Aquat Toxicol*. **126**:326–337.

Riaz MA, Poupardin R, Reynaud S, Strode C, Ranson H, David J-P. 2009. Impact of glyphosate and benzo[a]pyrene on the tolerance of mosquito larvae to chemical insecticides. Role of

detoxification genes in response to xenobiotics★. *Aquat Toxicol.* **93**(1):61–69.

Rogers AR, Huff C. 2009. Linkage disequilibrium between loci with unknown phase. *Genetics.* **182**(3):839–844.

Rose NH, Sylla M, Badolo A, Lutomiah J, Ayala D, Aribodor OB, Ibe N, Akorli J, Otoo S, Mutebi J-P, et al. 2020. Climate and urbanization drive mosquito preference for humans. *Curr Biol.* **30**(18): 3570–3579.e6.

Saavedra-Rodriguez K, Campbell CL, Lenhart A, Penilla P, Lozano-Fuentes S, Black WC. 2019. Exome-wide association of deltamethrin resistance in *Aedes aegypti* from Mexico. *Insect Mol Biol.* **28**(5):591–604.

Saavedra-Rodriguez K, Maloof FV, Campbell CL, Garcia-Rejon J, Lenhart A, Penilla P, Rodriguez A, Sandoval AA, Flores AE, Ponce G, et al. 2018. Parallel evolution of vgsc mutations at domains IS6, IIS6 and IIIS6 in pyrethroid resistant *Aedes aegypti* from Mexico. *Sci Rep.* **8**(1):6747.

Saavedra-Rodriguez K, Strode C, Flores AE, Garcia-Luna S, Reyes-Solis G, Ranson H, Hemingway J, Black WC IV. 2014. Differential transcription profiles in *Aedes aegypti* detoxification genes after temephos selection. *Insect Mol Biol.* **23**(2):199–215.

Saavedra-Rodriguez K, Strode C, Flores Suarez A, Fernandez Salas I, Ranson H, Hemingway J, Black WC. 2008. Quantitative trait loci mapping of genome regions controlling permethrin resistance in the mosquito *Aedes aegypti. Genetics.* **180**(2):1137–1152.

Saavedra-Rodriguez K, Suarez AF, Salas IF, Strode C, Ranson H, Hemingway J, Black WC IV. 2012. Transcription of detoxification genes after permethrin selection in the mosquito *Aedes aegypti. Insect Mol Biol.* **21**(1):61–77.

Saavedra-Rodriguez K, Urdaneta-Marquez L, Rajatileka S, Moulton M, Flores AE, Fernandez-Salas I, Bisset J, Rodriguez M, Mccall PJ, Donnelly MJ, et al. 2007. A mutation in the voltage-gated sodium channel gene associated with pyrethroid resistance in Latin American *Aedes aegypti. Insect Mol Biol.* **16**(6):785–798.

Schmidt TL, Chung J, Honnen A-C, Weeks AR, Hoffmann AA. 2020. Population genomics of two invasive mosquitoes (*Aedes aegypti* and *Aedes albopictus*) from the Indo-Pacific. *PLoS Negl Trop Dis.* **14**(7):e0008463.

Schrider DR, Hahn MW, Begun DJ. 2016. Parallel evolution of copy-number variation across continents in *Drosophila melanogaster. Mol Biol Evol.* **33**(5):1308–1316.

Schrider DR, Kern AD. 2017. Soft sweeps are the dominant mode of adaptation in the human genome. *Mol Biol Evol.* **34**(8): 1863–1877.

Schrider DR, Mendes FK, Hahn MW, Kern AD. 2015. Soft shoulders ahead: spurious signatures of soft and partial selective sweeps result from linked hard sweeps. *Genetics.* **200**(1):267–284.

Sene NM, Mavridis K, Ndiaye EH, Diagne CT, Gaye A, Ngom EHM, Ba Y, Diallo D, Vontas J, Dia I, et al. 2021. Insecticide resistance status and mechanisms in *Aedes aegypti* populations from Senegal. *PLoS Negl Trop Dis.* **15**(5):e0009393.

Severo OP. 1955. Eradication of the Aedes aegypti mosquito from the americas 23.

Shrinet J, Srivastava P, Sunil S. 2017. Transcriptome analysis of *Aedes aegypti* in response to mono-infections and co-infections of dengue virus-2 and chikungunya virus. *Biochem Biophys Res Commun.* **492**(4):617–623.

Smith LB, Kasai S, Scott JG. 2018. Voltage-sensitive sodium channel mutations S989P + V1016G in *Aedes aegypti* confer variable resistance to pyrethroids, DDT and oxadiazines. *Pest Manag Sci.* **74**(3):737–745.

Soares TS, Rodriguez Gonzalez BL, Torquato RJS, Lemos FJA, Costa-da-Silva AL, Guimarães MDLC, Tanaka AS. 2018. Functional characterization of a serine protease inhibitor modulated in the infection of the *Aedes aegypti* with dengue virus. *Biochimie.* **144**:160–168.

Sombié A, Saiki E, Yaméogo F, Sakurai T, Shirozu T, Fukumoto S, Sanon A, Weetman D, McCall PJ, Kanuka H, et al. 2019. High frequencies of F1534C and V1016I kdr mutations and association with pyrethroid resistance in *Aedes aegypti* from Somgandé (Ouagadougou), Burkina Faso. *Trop Med Health.* **47**(1):2.

Soper FL. 1965. The 1964 status of *Aedes aegypti* eradication and yellow fever in the Americas. *Am J Trop Med Hyg.* **14**(6):887–891.

Stalinski R, Laporte F, Tetreau G, Després L. 2016. Receptors are affected by selection with each Bacillus thuringiensis israelensis Cry toxin but not with the full Bti mixture in *Aedes aegypti. Infect Genet Evol.* **44**:218–227.

Sylla M, Bosio C, Urdaneta-Marquez L, Ndiaye M, Iv WCB. 2009. Gene flow, subspecies composition, and dengue virus-2 susceptibility among *Aedes aegypti* collections in Senegal. *PLoS Negl Trop Dis.* **3**(4):e408.

Tabachnick WJ. 1991. Evolutionary genetics and arthropod-borne disease: the yellow fever mosquito. *Am Entomol.* **37**(1):14–26.

Tabachnick WJ, Powell JR. 1979. A world-wide survey of genetic variation in the yellow fever mosquito, *Aedes aegypti*. Genet Res.* **34**(3):215–229.

Tabachnick WJ, Wallis GP, Aitken TH, Miller BR, Amato GD, Lorenz L, Powell JR, Beaty BJ. 1985. Oral infection of *Aedes aegypti* with yellow fever virus: geographic variation and genetic considerations. *Am J Trop Med Hyg.* **34**(6):1219–1224.

Taylor MFJ, Heckel DG, Brown TM, Kreitman ME, Black B. 1993. Linkage of pyrethroid insecticide resistance to a sodium channel locus in the tobacco budworm. *Insect Biochem Mol Biol.* **23**(7): 763–775.

Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, McGee S, Do R, Liu X, Jun G, et al. 2012. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science.* **337**(6090):64–69.

Tobler R, Johar A, Huber C, Souilmi Y. 2020. PolyLinkR: a linkage-sensitive gene set enrichment R package. doi:10.48550/arXiv.2004.03224. [accessed 2023 Jan 25]. http://arxiv.org/abs/2004.03224.

Turner TL, Levine MT, Eckert ML, Begun DJ. 2008. Genomic analysis of adaptive differentiation in *Drosophila melanogaster. Genetics.* **179**(1):455–473.

Urdaneta-Marquez L, Failloux A-B. 2011. Population genetic structure of *Aedes aegypti*, the principal vector of dengue viruses. *Infect Genet Evol.* **11**(2):253–261.

Vasimuddin Md, Misra S, Li H, Aluru S. 2019. Efficient architecture-aware acceleration of BWA-MEM for multicore systems. In: 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS). p. 314–324.

Venkataraman K, Shai N, Lakhiani P, Zylka S, Zhao J, Herre M, Zeng J, Neal LA, Molina H, Zhao L. 2022. Rapidly evolving genes underlie *Aedes aegypti* mosquito reproductive resilience during drought. 37.

Vera-Maloof FZ, Saavedra-Rodriguez K, Elizondo-Quiroga AE, Lozano-Fuentes S, Iv WCB. 2015. Coevolution of the Ile1,016 and Cys1,534 mutations in the voltage gated sodium channel gene of *Aedes aegypti* in Mexico. *PLoS Negl Trop Dis.* **9**(12):e0004263.

Viglietta M, Bellone R, Blisnick AA, Failloux A-B. 2021. Vector specificity of arbovirus transmission. *Front Microbiol.* **12**:773211.

Vivero RJ, Muskus CE, Uribe SI, Bejarano EE, Torres C. 2010. Flebotomíneos (Diptera: Psychodidae) en la reserva natural del cañón del río Claro (Antioquia), Colombia. *Actual Biol.* **7**.

Walsh TK, Joussen N, Tian K, McGaughran A, Anderson CJ, Qiu X, Ahn S-J, Bird L, Pavlidi N, Vontas J, et al. 2018. Multiple recombination events between two cytochrome P450 loci contribute to global pyrethroid resistance in *Helicoverpa armigera. PLoS One.* **13**(11):e0197760.

Weedall GD, Riveron JM, Hearn J, Irving H, Kamdem C, Fouet C, White BJ, Wondji CS. 2020. An Africa-wide genomic evolution of insecticide resistance in the malaria vector Anopheles funestus involves selective sweeps, copy number variations, gene conversion and transposons. *PLoS Genet.* **16**(6):e1008822.

Wilder-Smith A, Byass P. 2016. The elusive global burden of dengue. *Lancet Infect Dis.* **16**(6):629–631.

Williamson MS, Martinez-Torres D, Hick CA, Devonshire AL. 1996. Identification of mutations in the houseflypara-type

sodium channel gene associated with knockdown resistance (kdr) to pyrethroid insecticides. *Mol Gen Genet MGG.* **252**(1): 51–60.

Xiao X, Liu Y, Zhang X, Wang J, Li Z, Pang X, Wang P, Cheng G. 2014. Complement-related proteins control the flavivirus infection of *Aedes aegypti* by inducing antimicrobial peptides. *PLoS Pathog.* **10**(4):e1004027.

Xue AT, Schrider DR, Kern AD; Ag1000g Consortium. 2021. Discovery of ongoing selective sweeps within anopheles mosquito populations using deep learning. *Mol Biol Evol.* **38**(3): 1168–1183.

Zhu J, Busche JM, Zhang X. 2010. Identification of juvenile hormone target genes in the adult female mosquitoes. *Insect Biochem Mol Biol.* **40**(1):23–29.